



БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ
**ИНСТИТУТ ПО ИНФОРМАЦИОННИ
И КОМУНИКАЦИОННИ ТЕХНОЛОГИИ**

Атанас Филипов Николов

**ВИДЕОСТАБИЛИЗАЦИЯ И 3D РАЗПОЗНАВАНЕ В
РЕАЛНО ВРЕМЕ**

ДИ С Е Р Т А Ц И Я

за присъждане на
образователната и научна степен „Доктор“

Докторска програма:
„Компютърни системи, комплекси и мрежи”

Професионално направление:
5.3 „Комуникационна и компютърна техника”

Научен ръководител:
доц. д-р Димо Т. Димов

София, 2016 г.

Съдържание

Списък на фигурите	5
Списък на таблиците	7
Списък на използвани съкращения и означения	8
Увод	9
Актуалност на темата.....	9
Цели и задачи на дисертационния труд	12
Методология на изследването	12
Структура на дисертацията	13
Глава 1. Обзор на основните резултати в областта	17
1.1 Видеостабилизация	17
1.1.1 Софтуерни методи за видеостабилизация	18
1.1.2 Исторически преглед на върховите постижения при софтуерната видеостабилизация.....	20
1.1.3 Хардуерни методи за видеостабилизация, използващи инерциални сензори	32
1.1.4 Интересни методи за 2D стабилизация и тяхната връзка с разработените методи по видеостабилизация в дисертационния труд	32
1.2 Биометрика.....	34
1.2.1 Традиционни методи за автентикация.....	34
1.2.2 Класификация на биометричните технологии	35
1.2.3 Исторически преглед на развитието на биометричните системи.....	35
1.2.4 Изисквания към биометричните модалности и системи	36
1.2.5 Биометрични системи на бъдещето.....	37
1.2.6 Моно- и мултибиометрични системи	38
1.2.7 Сравнителен анализ на биометричните модалности	39
1.2.8 Съвременни приложения на биометричните системи	41
1.2.9 Разпознаване на индивиди по формата на ушите	43
1.2.10 Предимства и недостатъци на биометриката по уши	44
1.2.11 Методи за разпознаване по уши.....	45
Двумерни (2D) подходи за разпознаване на уши	45
Тримерни (3D) подходи за разпознаване на уши	48
1.2.12 Публични БД от 2D изображения и 3D модели на уши.....	49
Глава 2. Методи за бърза и качествена 2D видеостабилизация.....	51
2.1 Видеостабилизация по гладка траектория	51

2.1.1 Определяне на векторите на движение	51
2.1.2 Базов векторен модел с разбиване на “3×3” области	53
2.1.3 Мултипликация на базовия векторен модел за 3×3 → 9×9 области	58
2.1.4 “Rolling shutter” разширение на базовия векторен модел.....	74
2.1.5 Експериментални резултати	76
2.1.6 Заключение и идеи за бъдеща работа	78
2.2 Видеостабилизация за индустриални високоскоростни камери.....	79
2.2.1 Въведение – необходимост и мотивация.....	80
2.2.2 Методът на Оптичния поток	80
2.2.3 Комбиниран подход “3×3OF9×9” за видеостабилизация в точка	82
2.2.4 Експериментални резултати	83
2.2.5 Заключение и идеи за бъдеща работа	86
2.3 Откриване на предварително известни рекламни клипове в TV видеопоток	86
2.3.1 Формулировка на задачата.....	88
2.3.2 Дискутирани сценарии за решаване на задачата	88
2.3.3 Описание на предложения алгоритъм по третия сценарий.....	89
2.3.4 Заключение и идеи за бъдеща работа	95
Глава 3. Методи за разпознаване на индивиди по формата на ушите.....	96
3.1 Описание на създадената в ИИКТ БД от 3D уши (3DEarDB).....	96
3.1.1 Устройства за събиране на данни	97
3.1.2 Сурови (необработени) 3D модели на уши	98
3.1.3 Обработка на суровите 3D модели	99
3.1.4 Kinect дълбочинни изображения	101
3.1.5 Full HD обхождащи видеоклипове на уши	101
3.1.6 2D интензитетни проекции	102
3.1.7 2D дълбочинни изображения	102
3.1.8 Уеб достъп до 3DEarDB	103
3.2 Представяне на 3D обекти чрез разширено Гаусово изображение (EGI)	103
3.2.1 Аналитично описание на метода за построяване на икосаедър и неговата 2D развивка	105
3.3 Експериментални резултати за разпознаване на EGI хистограми по метода на най-близкия съсед.....	107
3.3.1 Анализ на експерименталните резултати	112
3.4 Разпознаване на уши чрез подхода „Многоизгледно 3D моделиране”	113
3.5 Обещаващи методи за 3D реконструкция на обекти в реално време за целите на разпознаването по уши	114
3.6 Заключение и идеи за бъдеща работа	117

Заключение.....	119
Резюме на постигнатото и идеи за бъдеща работа.....	119
Списък на публикациите по дисертацията	121
Апробация на резултатите.....	122
Авторска справка за приносите в дисертацията.....	123
Декларация за оригиналност на резултатите.....	124
Благодарности.....	125
Литература	126
Приложения	136

Списък на фигурите¹

Фиг. 1.	Обща схема на изследванията по дисертацията.	13
Фиг. 2.	Детайлизирана схема на изследванията по дисертацията.	16
Фиг. 3.	Синтезиране на нови изгледи за качествена 3D стабилизация.	21
Фиг. 4.	Представяне на траекторията на камерата чрез статични, линейни и параболични сегменти.	24
Фиг. 5.	Оригиналните (а) и стабилизирани (б) траектории от характерни точки в пространствено времева координатна система.	28
Фиг. 6.	Сравнение между традиционната 2D стабилизация с единствена глобална траектория (а) и „снопа от локални траектории“ (б), изгледени във времето.	30
Фиг. 7.	Референтни видеоклипове от предишни успешни методи, за сравнение с текущия метод.	30
Фиг. 8.	Обща схема на верификация на една биометрична система.	34
Фиг. 9.	Идентификация по „цифровия отпечатък“ на индивида в интернет пространството.	37
Фиг. 10.	Идентификация чрез шаблона от почасови ежедневни местоположения, базиран на слеждане на GPS координатите на мобилния телефон на индивида.	37
Фиг. 11.	Най-популярните биометрични модалности за автентикация на индивиди: а) пръстови отпечатъци; б) физиономии; в) ирис.	39
Фиг. 12.	Биометрични данни за най-голямата биометрична БД в света: а) 10 пръстови отпечатъка; б) изображение на физиономията; в) изображения на двата ириса.	39
Фиг. 13.	Отпечатък на дланта и ДНК информация.	40
Фиг. 14.	Биометрични модалности, използвани в комерсиалните системи за работа в режим на верификация.	40
Фиг. 15.	Биометрични модалности в лабораторен етап на разработка.	40
Фиг. 16.	Други биометрични модалности, с възможности за използване в криминалистиката.	41
Фиг. 17.	Шестте базови възела от морфологията на ухото.	43
Фиг. 18.	Базова схема за определяне на векторите на движение: а) 3×3 разбиване на всеки кадър от видеоклипа; б) хоризонтални и вертикални акумулиращи проекции.	51
Фиг. 19.	Базов векторен модел, отчитащ трансляция и ротация между два последователни кадъра във видеото.	53
Фиг. 20.	Запис на системата уравнения от Табл. 3 в матричен вид $AX = B$	54
Фиг. 21.	Мултипликация на базовия векторен модел за разбиване на кадрите на “9×9” области, отчитащ трансляция и ротация между два последователни кадъра във видеото.	58
Фиг. 22.	Схема за съставяне на уравнения чрез обхождане по хоризонталната права между точките (5,9) и (5,1).	59
Фиг. 23.	Схема за съставяне на уравнения чрез обхождане по вертикалната права (9,5)÷(1,5).	60
Фиг. 24.	Схема за съставяне на уравнения чрез обхождане по диагонала (9,9)÷(1,1).	61
Фиг. 25.	Схема за съставяне на уравнения чрез обхождане по „основните“ диагонални прави.	62
Фиг. 26.	Схема за съставяне на уравнения чрез обхождане по „вътрешните“ диагонални прави.	64
Фиг. 27.	Симетрична схема на СЗ за построяване на уравненията по главния диагонал.	66
Фиг. 28.	Запис на системата уравнения от Табл. 10 в матричен вид $AX = B$	70
Фиг. 29.	Разширен векторен модел, отчитащ трансляция, ротация, хоризонтално накланяне и вертикално мащабиране между два кадъра.	74

¹ Евентуалните взаимствания във фигурите, от външни литературни източници, са указани към самите фигури в дисертацията.

Фиг. 30. Тестови видеоклипове, подредени във възходящ ред по степента им на трептене.	77
Фиг. 31. Идеята на метода на оптичния поток се състои в изчисляване на 2D векторното поле за всички пиксели между двойка припокриващи се изображения.	81
Фиг. 32. Стратегия за елиминиране на подвижните обекти в сцената.	83
Фиг. 33. Стабилизация на видеоклип на контролиран взрив.	84
Фиг. 34. „Стабилизирани кадри“ в „точка“ на 1-ви етап на стабилизация.	84
Фиг. 35. Стабилизирани кадри в „точка“ на 2-ри етап на стабилизацията.	85
Фиг. 36. Нашият резултат за стабилизация в статичен фон в сравнение с този от <i>Adobe After Effects CS6</i>	85
Фиг. 37. TV видео, съдържащо множество рекламни блокове, в рамките на новинарска емисия.	87
Фиг. 38. Схематично представяне на сложността на директното покадрово сравнение.	88
Фиг. 39. Диференциалното представяне на рекламен клип за различни стойности на PST-прага.	91
Фиг. 40. Детекция на сцени в диференциалното представяне на TV видеоклип и конвертирането му в целочислена редица от дължините на съставните му сцени.	92
Фиг. 41. Детекция на сцени в рекламен видеоклип (~20 сек.).	93
Фиг. 42. Целочислена редица от дължините на сцените Δl за горния рекламен видеоклип.	93
Фиг. 43. Детекция на сцени за ~15-минутен TV поток, съдържащ само един рекламен блок.	93
Фиг. 44. Илюстрация на основното правило за включване на сцена от TV поток в реклама.	94
Фиг. 45. а) картонен „шлем“; б) сканиране на участник; в) 3D скенерът VIUscan.	98
Фиг. 46. а) цветно необработено ухо; б) само повърхнината на необработеното ухо.	99
Фиг. 47. а) остри пикове; б) нови дупки; в) запълнени дупки; г) финално изглаждане.	100
Фиг. 48. а) децимация на фасетките; б) оригинална резолюция; в) изфинване на фасетките.	100
Фиг. 50. Представителни кадри за 3 хоризонтални реда на видеоклип на ухо: а) изглед отгоре; б) централен изглед; в) изглед отдолу.	101
Фиг. 49. Нормализиран 3D модел на ухо.	101
Фиг. 52. 2D изображения от даден ред на ротационната схема за 3D моделиране.	102
Фиг. 53. Дълбочинни изображения под ортографска проекция за дадена ротация на 3D ухо.	102
Фиг. 51. Схема на многоизгледно 3D моделиране на дадено ухо.	102
Фиг. 54. а) 3D мрежест модел; б) нормални вектори към фасетките на модела.	104
Фиг. 55. 3D рекурсивно разбиване, базирано на икосаедър (20-тостен) за апроксимиране на сфера чрез 20, 80, 320 и 1280 фасетки.	104
Фиг. 56. 2D представяне (по азимут и деклинация) на дадена апроксимация на сфера.	104
Фиг. 57. 3D и 2D EGI хистограми за някаква входна 3D повърхнина – във всеки EGI бин се акумулира площта на съответните триъгълните фасетки от входа.	105
Фиг. 58. Необходими параметри за определяне на координатите (x, y, z) на върховете за построяването на икосаедър и неговото 2D представяне.	106
Фиг. 59. Първоначалната БД от 11 3D модела на уши, създадена през май 2014 г.	107
Фиг. 60. а) зашумен модел на ухо с равномерен шум в даден интервал, добавен към всяка (x, y, z) координата; б) съответните 3D EGI хистограми на зашумените уши.	108
Фиг. 61. Сравнение от типа „всяко-с-всяко“ на входните зашумени модели на уши с идеалните модели от 3DEarDB.	109
Фиг. 62. Равномерни 2D проекции, извлечени от 3D модела на дясното ухо на колега от ИИКТ.	113
Фиг. 63. Вариационни методи за 3D реконструкция.	115
Фиг. 64. 3D моделиране на обекти в реално време чрез системата „CoryMe3D“.	115
Фиг. 65. 3D реконструкция на главата на <i>Meduza</i> чрез обхождащо видео.	115
Фиг. 66. 3D моделиране на сцени в голям мащаб в градски условия.	116
Фиг. 67. 3D реконструкция чрез използването на изображения от мобилни телефони.	116
Фиг. 68. 3D реконструкция на физиономии, заснети чрез предната камера (0.3 мегапиксела) на Google Nexus S смартфон.	116
Фиг. 69. Хомогенно представяне на 2D точки върху равнината $w = 1$	136
Фиг. 70. Йерархично множеството от линейни 2D трансформации.	136

Списък на таблиците

Таблица 1.	Сравнителен анализ на моно- и мултибиометричните системи.	38
Таблица 2.	Приложения на биометричните модалности в съвременното общество.	42
Таблица 3.	Система от 18 покомпонентни уравнения за центровете (i, j)	53
Таблица 4.	Система от 18 покомпонентни уравнения след 1-вия LSM пас.	56
Таблица 5.	Система от покомпонентни уравнения за обхождане по правата $(5,9) \div (5,1)$	59
Таблица 6.	Система от покомпонентни уравнения за обхождане по правата $(9,5) \div (1,5)$	60
Таблица 7.	Система от покомпонентни уравнения за обхождане по диагонала $(9,9) \div (1,1)$	61
Таблица 8.	Система от покомпонентни уравнения за обхождане по „основните“ диагонални прави.	63
Таблица 9.	Система от покомпонентни уравнения за обхождане по „вътрешните“ диагонални прави.	65
Таблица 10.	Система от 162 покомпонентни уравнения за схемата на разбиване “9×9”.	66
Таблица 11.	Систематизиране на коефициентите пред неизвестните параметри от Табл. 10.	69
Таблица 12.	Система от 18 покомпонентни уравнения за “3×3” разширения векторен модел.	75
Таблица 13.	Коефициенти пред неизвестните параметри ix и sy , за LSM функцията на Матлаб.	75
Таблица 14.	ITF оценки за 5 видеоклипа: оригиналните и техните стабилизации по 6-те подхода.	77
Таблица 15.	Грешки от стабилизацията в статичен фон след 1-ви и 2-ри етап на стабилизацията.	85
Таблица 16.	EGI резултати от разпознаването: True Recognition Rate (TRR).	110
Таблица 17.	Йерархия на линейните 2D трансформации.	139

Списък на използвани съкращения и означения

Term / Symbol (на англ.)	Термин / Означение (на бълг.)	Съкращение
3×3OF9×9	(наш) комбиниран метод за видеостабилизация	3×3OF9×9
3DEarDB	(нашата) многомоделна БД от уши	3DEarDB
Adobe After Affects	Софтуер за видеомонтаж и видеоефекти	AAE
Boundary Signal Computation	Чип за изчисляване на акумулиращи проекции	BSC
Content Base Image Retrieval	Метод за търсене на изображения в БД според тяхното съдържание	CBIR
Charge-Coupled Device	Технология на сензора на дадена видеокамера, формираща наведнъж 2D изображението	CCD
Complementary-Symmetry Metal-Oxide Semiconductor	Технология на сензора на дадена видеокамера, формираща изображението ред-по-ред	CMOS
Discrete Cosine Transform	Дискретното косинус преобразование	DCT
Degree of Freedom	Степен на свобода	DoF
Effective and Fast Content based Image or Object Retrieval	Ефективен метод за търсене на изображения или обекти в БД според тяхното съдържание	EFCIOR
Extended Gaussian Image	Разширено Гаусово изображение	EGI
Features from accelerated segment test	Бърз метод за откриване на характерни точки	FAST
Fast Fourier Transform	Бърза трансформация на Фурие	FFT
frames per second	кадри за секунда	fps
Fourier Transform	Трансформацията на Фурие	FT
Graphics Processing Unit	Видеокарта	GPU
Independent Component Analysis	Метод на независимите компоненти	ICA
Iterative Closest Point	Метод за оценка на подобие на 3D обекти	ICP
Integral Projection	Акумулираща проекция	IP
Interframe Transformation Fidelity	Мярка за подобие между два видеокадра	ITF
Kanade-Lucas-Tomasi	Метод за следене на характерни точки	KLT
Least squares method	Метод на най-малките квадрати	LSM
Optical Flow	Метод на Оптичния поток	OF
Principal Component Analysis	Метод на главните компоненти	PCA
Pixel Similarity Threshold	Праг за подобие между пикселите	PST
RANdom SAMple Consensus	Метод за елиминиране на грешни съответстващи си характерни точки	RANSAC
Sum of Absolute Differences	Сума от абсолютните разлики	SAD
Structure-from-Motion	Структура-от-Движение	SfM
Shape from Shading	Структура-от-Сянка	SfS
Scale-invariant feature transform	Детектор и дескриптор на характерни точки	SIFT
Speeded Up Robust Features	Детектор и дескриптор на характерни точки	SURF
Texas Instruments	Американска компания за електроника	TI
True Recognition Rate	Степен на разпознаваемост	TRR
Multi-View Stereo	Метод за 3D реконструкция на обекти чрез множество изображения от различни изгледи	MVS

Увод

Актуалност на темата

Настоящият дисертационен труд „Видеостабилизация и 3D разпознаване в реално време“ се базира на съвместното решаване на два проблема – този за 2D видеостабилизация в реално време и за 3D разпознаване на личността² по формата на ушите. Тези две, донякъде различни области на изследователски интерес, са съвместени в една обобщаваща схема (виж *Структура на дисертацията*, Фиг. 1 и 2). Свързващото звено между видеостабилизацията и 3D разпознаването е в 3D реконструкцията на обекти от обхождащ видеоклип с цел осигуряване на входни данни от 3D модели за БД с прецеденти. А връзката на видеостабилизацията с 3D реконструкцията е в определянето на параметрите на движението между определени области от интерес (или характерни точки) в припокриващите се кадри, с цел изчисляване на 3D координатите на камерата и на обекта. Също така, стабилизацията на заснемания обект е пряко свързана и с неговото ефективно сегментиране от фона. Последното може да се използва и за осигуряване на вход от структурирани 2D входни данни (интензитетни проекции) за други методи за разпознаване, разработени в ИИКТ.

В хронологията на развитието на изследователската работа по дисертацията се наложи прецизиране на първоначално дефинираната тема „*Стереоскопично разпознаване на 3D обекти*“ до настоящата (виж и схемата). Това бе предизвикано от породилия се интерес към разнообразните изследователски проблеми от насрещните проекти в ИИКТ (виж апробация на резултатите). Същевременно, според обзорния анализ, наличието на най-съвременни методи (с възможност за работа в реално време) в областта на 3D реконструкцията (виж Глава 3, §3.5) доведе до пренасочване на акцентите в дисертацията към вече споменатите две области - видеостабилизация и 3D разпознаване, основано на БД от 3D модели-образци, като идеята за *ушите* бе инспирирана от нарастващия интерес на биометричната общност към тях напоследък. Последното бе установено и на международния семинар по биометрика BIOMET'2014, проведен на 23-24 юни, 2014 г., БАН, София, в рамките на проекта AComIn на ИИКТ. Така, нашата основна цел е 3D разпознаването, но към него ние подхождаме и чрез видеостабилизация, доколкото основната входна информация преминава през конвенционалните 2D видео устройства, виж и примерите в *Our_VStab/Ears/...* .

Методите за видеостабилизация целят подобряване качествата на нестабилно заснето видео (любителско и/или професионално) в точка или по гладка траектория. Под стабилизация в точка се разбира стабилизацията на всички видео кадри спрямо един референтен кадър, в резултат на което се пресъздава видео, заснето от статична камера. Стабилизацията по траектория произвежда видео, при което камерата следва изгладеното оригинално движение на оператора. Известно е, че в киноиндустрията за стабилното

² В настоящия дисертационен труд взаимнозаменяемо използваме философското понятие „личност“ с „индивид“, в изразите „разпознаване на личността“ и „разпознаване на индивиди“, с цел избягване на повторения.

заснемане се използва голямо разнообразие от сложно оборудване: стативи; камери, движещи се върху релси; операторски колички; преносими стабилизиращи механизми, закрепени към тялото и др. Подобна апаратура в много случаи е скъпа за любителски цели, поради което широко се използват софтуерни методи за постигане на плавност на движението в едно любителско, не принудено заснето видео. Принципно задачата за софтуерната 2D стабилизация и дори 3D стабилизацията се смята за решена в случаите на off-line обработка. Най-известните софтуерни реализации по проблема са имплементирани в *Google YouTube Stabilizer* и *Warp Stabilizer* на *Adobe After Effects* на базата на методите, описани съответно в [1] и [2]. Но както те, така и многото други подходи за 2D, 2.5D и 3D стабилизация от последните години (виж обзора по темата в Глава 1, §1.1.2), все още не могат да достигнат достатъчно бързодействие на PC от среден клас, освен ако не се използва паралелна имплементация на GPU. Но тъй като компютърните мощности в преносимите устройства са доста по-слаби от тези на десктоп машините, за постигането на приемлива видеостабилизация определено се разчита на съвместното използване на вградените в хардуера им инерциални сензори (жироскопи и/или акселерометри) с прецизни и бързи софтуерни подходи. В тази връзка ние разработихме един бърз и ефективен софтуерен метод за 2D видеостабилизация, който е предназначен за работа в реално време върху портативни устройства и е адаптиран към BSC чипа на Texas Instruments за търсене на корелации по SAD метода между съответните 1D акумулиращи проекции по хоризонтални и по вертикали (виж Глава 2, §2.1.1). Базовият метод издържа успешно изпитанията (виж Приемо-Предавателен-Протокол). Следващи разширения и/или модификации на метода са описани в Глава 2, §2.1.3-4. В близка перспектива предвиждаме неговата адаптация като софтуерно инициализиращо средство за подобряване на точността на системата от инерциални сензори, вградени в използваното портативно устройство, на които напоследък се разчита за една ефективна видеостабилизация.

Въпреки че научните изследвания по използването на ушите в биометриката се намират в начален стадий, в последните години се обръща особено внимание на уникалността на ушите като отличителна характеристика за биометрично разпознаване на личността [91], свидетелство за което са и множеството дисертационни резултати [66, 68], книги [67, 91] и обзорни статии в областта [56, 57, 58]. От големия брой изследователски статии по темата от последните десетилетия може да се заключи, че точността, постигната чрез системите за разпознаване по уши, е съизмерима с тази на разпознаването по лица [57]. В подкрепа на последното е едно изследване през 2005 г. в Университета на Southampton, UK, относно индекса на отличимост "*decidability index*", който измерва подобие / различимостта на ушите като показател за степента на уникалност на човешките уши. Установено е, че този индекс е с един порядък по-голям за уши, отколкото за физиономии, но по-малък, отколкото при ирисовата биометрика [98]. Lei et al. показват, че ушите могат да се използват и за класификация по пол [92]. Известно е, че формата на ухото еволюира по време на ембрионалния растеж от шест развиващи се базови възела. Това обуславя уникалността на неговата структура, същевременно без тя да е напълно произволна. Засега, детайлната 3D повърхнина на човешкото ухо се разглежда като универсална и уникална характеристика, въпреки че все още не е оконча-

телно установено, че ушите на всички хора са уникални. Освен това, структурата на ушите не се променя значително с възрастта [94]. Ушите са лесни за (отдалечено) заснемане по един ненаатрапчив начин. Горните особености водят до извода, че ушите удовлетворяват изискванията (универсалност, уникалност, инвариантност във времето и лекота при регистриране), формулирани от Jain [77], и могат да се използват в биометричната идентификация и верификация на личността. Въпреки че повече от 99% от биометричните системи, достъпни на пазара, са базирани на пръстови отпечатъци, лицево или ирисово разпознаване [79], използването на ушите има и своите предимства спрямо тях [58], а именно: (i) за разлика от лицето, ушите имат относително постоянна структура, която не се променя много с възрастта и не се влияе от лицевите изражения; (ii) ухото лесно може да се заснеме от разстояние, което не изисква непременно съдействие от страна на потребителя и се счита от повечето хора за ненаатрапчиво, за разлика от заснемането на ириса; (iii) също така, безконтактното заснемане не води до проблеми с хигиената за разлика от вземането на пръстови отпечатъци.

Във връзка с разпознаването по уши в дисертацията е разработена многомоделната БД (3DEarDB), съдържаща различни представяния на ушите на ~100 индивиди (виж Глава 3, §3.1). Целта на 3DEarDB е да осигури съвместимост между различните изходни формати за представянето на ухото на един и същ човек. Тези формати включват: (i) сурови (необработени) 3D модели на уши; (ii) обработени 3D модели; (iii) Kinect дълбочинни изображения “*depth maps*”; (iv) 2D видео клипове; (v) генерирани структури от 2D интензитетни проекции на уши; (vi) генерирани структури от 2D дълбочинни изображения. Това консистентно разнообразие от изходни формати би било много полезно за научната общност по разпознаване по ушите, за тестване и сравнение на точността на алгоритмите върху различни възможни сценарии – от идеалния случай на прецизен (и статичен) 3D модел през по-реалистичния (и динамичен) случай на 2D видеоклип и до отделни 2D изображения. От нашия обзор върху публично наличните БД от 2D изображения и 3D модели на човешки уши (виж Глава 1, §1.2.12) е ясно, че до момента не съществува такова консистентно съответствие между множество от 2D и 3D изходни формати, което е явно предимство на разработената многомоделна БД.

Идеята за разработването на 3DEarDB, за нейната структура и за препоръчителния минимален брой личности, представени в нея, се оформи в дискусия на BIOMET'2014. Развитието и реализацията на тази идея стана възможно благодарение на модерния 3D лазерен скенер, закупен по проекта AComIn.

Цели и задачи на дисертационния труд

Поради естеството на изследванията по време на докторантското обучение, в настоящия дисертационен труд дефинирахме следните две основни цели:

1. Разработване на методи за ефективна стабилизация на видеоклипове, както по изгладена траектория на камерата, така и в „точка“ (статична камера).
2. Разработване на БД от 3D модели на човешки уши (за целите на биометриката) и апробирането ѝ със съвременни 2D/3D методи за разпознаване по прецеденти.

За реализацията на **първата цел** са предвидени следните задачи:

- 1.1 Разработване на метод за софтуерна 2D видеостабилизация, подходящ за вграждане в мобилни телефони, осигуряващ задоволително бързодействие (работа в реално време) и същевременно произвеждащ достатъчно прецизен резултат.
- 1.2 Адаптиране на метода за видеостабилизация за видео клипове от индустриална високоскоростна камера, с цел „перфектната стабилизация в точка“ на проследявания обект, или на фона.

За осъществяването на **втората цел** са определени следните задачи:

- 2.1 Разработване на БД от 3D модели на човешки уши, както и техни производни представяния, в насоката – тест и настройка на съвременни методи за 3D разпознаване.
- 2.2 Експериментално потвърждаване на приложимостта на известния EGI дескриптор за представяне на изпъкнало-вдлъбнати 3D обекти, като ушите, за недвусмисленото им разпознаване.

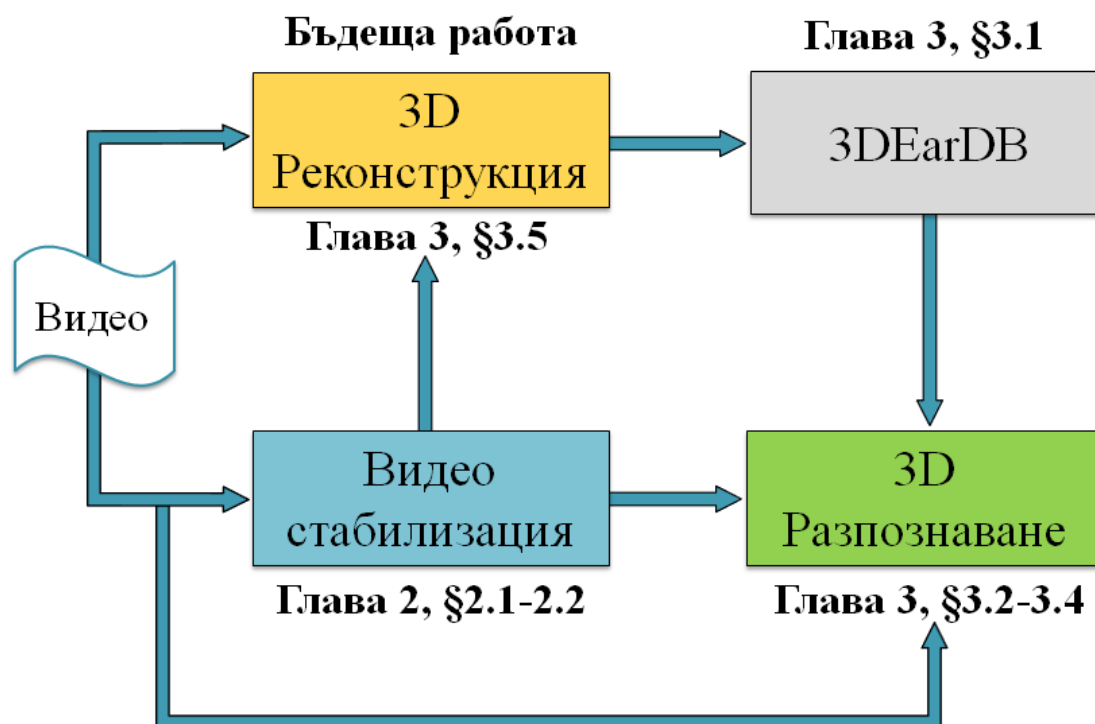
Методология на изследването

Препоръчваната методологията (в хронологията на изследванията) се основава на методи и подходи от следните области (но без ограничения само върху тях):

- *Линейна алгебра и Аналитична геометрия*: матрични уравнения, методът на най-малките квадрати; нормални вектори, скаларно и векторно произведение, собствени вектори и собствени стойности, аналитични представяния в 2D и 3D, и др.;
- *Функционален анализ*: линейна апроксимация на функция в малка околност, Fourier анализ и др.;
- *Статистика*: хистограми, равномерно и Гаусово разпределение, моменти и др.;
- *Обработка на изображения*: 2D и 3D изображения, разширено Гаусово изображение (EGI), линейни трансформации, метод на главните компоненти и др.;
- *Разпознаване на образи*: подобие между обекти, разстояния и метрики, класически методи за класификация – най-близък съсед, k -те най-близки съседа, и др.;
- *Компютърно зрение*: 3D реконструкция на обекти, стереозрение, фотометрия.

Структура на дисертацията

Настоящият дисертационен труд се базира на съвместното решаване на два проблема – този за бърза и ефективна 2D видеостабилизация и за 3D разпознаване на индивиди по формата на ушите. Тези две области на изследователски интерес са съвместени в една обобщаваща схема (Фиг. 1). Част от блоковете в схемата са предмет на изследванията в дисертацията (Видеостабилизация, 3D разпознаване и БД от 3D модели на човешки уши – 3DEarDB), а други (3D реконструкция) са оставени за бъдеща работа.



Фиг. 1. Обща схема на изследванията по дисертацията.

На Фиг. 2 е показано едно по-детайлизирано представяне на същността на методите от окрупнените блокове от Фиг. 1 и тяхната взаимовръзка, както и съответното практическо приложение в рамките на различни проекти на ИИКТ-БАН. Тези проекти са:

1. Проектът с фирма MM Solution³ на тема „Математически алгоритми за видео стабилизация и филтриране на шум от сензори за движение“ в периода май-октомври 2012 г.;

2. Европейският проект AComIn⁴ на ИИКТ, в рамките на който е използвано следното висококачествено оборудване: 3D лазерен скенер VIUscan и високоскоростна камера NAC Memrecam;

3. Мини проект за техническа изпълнимост по проблем, зададен от фирма Cross Agency Ltd.⁵, на тема „Ефективно разпознаване на видео шаблони в поточно видео“, в рамките на организираната от ИИКТ-БАН, ФМИ-СУ и ИМИ-БАН, и със съдействието

³ <https://www.mm-sol.com>

⁴ <http://www.iict.bas.bg/acomin/>

⁵ <http://parallel.bas.bg/ESGI104/Problems.html>

на Европейския консорциум за математика в индустрията, „104-та Европейска изследователска група с индустрията“ (ESGI-113), 23-27 септември 2014 г., София.

Детайлната схема от Фиг. 2 е подготвена като илюстрация на структурата на дисертацията, по глави, които са обект на разработка в хронологичен ред, както следва:

Глава 1 съдържа обзор на основните резултати в направленията видеостабилизация и биометрично разпознаване на индивиди по формата на ушите.

Раздел 1.1 дава детайлен преглед на състоянието на областта на софтуерните методи за видеостабилизация (2D, 2.5D и 3D) според тяхната актуалност в периода, в който те са заемали върхови позиции в световен мащаб. В края на раздела е разгледан един бърз и ефективен метод за 2D видеостабилизация, предназначен за работа в реално време върху портативни устройства, на чиято основа стъпват разработените в дисертацията подобрени методи за видеостабилизация.

Раздел 1.2 представя подробна справка и сравнение на предимствата и недостатъците на известните биометрични модалности, според литературните източници. Обоснована е актуалността на разпознаването на личности по формата на ушите, като са посочени съответните предимства спрямо популярните методи – по пръстови отпечатъци, физиономия (лице), ирис. Разгледани са най-известните 2D и 3D методи за разпознаване по ушите. Направен е и преглед върху наличните БД от 2D изображения и 3D модели на човешки уши, като са посочени най-високите резултати на 2D и 3D методите по детекцията на уши и разпознаването по ушите.

В **Глава 2** се разглеждат разработените по дисертацията софтуерни методи за 2D видеостабилизация, както по гладка траектория (§2.1), така и в „точка“ (§2.2), подходящи за вграждане в мобилни телефони, т.е. осигуряващи достатъчно бързодействие (работа в реално време) според наличните мощности, както и произвеждащи задоволително прецизен резултат. Също така, предложен е бърз, прецизен и относително прост (подходящ за работа в реално време) алгоритъм за локализиране на предварително известни фиксирани (рекламни) видеоклипове и/или произволни части от тях в базов TV видеопоток (§2.3).

Раздел 2.1 описва разработения базов (CCD) векторен модел с разбиване на изображенията на „3×3“ области за определяне на търсената трансформация (транслация и ротация) между двойка кадри. При стабилизацията по гладка траектория е предложено и подобрене при намирането на „векторите на движение“ чрез известния SAD подход, което е съществено за точността на резултата. Предложена е мултипликация по аналогия на базовия векторен модел, чрез разбиване на всеки кадър на „N×N“ области, $N \geq 3$, нечетно, в резултат на което се постига забележимо подобрене в качеството на стабилизирания видеоклип. За простота на изложението тук се описва случаят „9×9“. Предложено е и разширение на базовия векторен модел, който се справя успешно в корекцията на хоризонталното накланяне и вертикалното мащабиране, породени от така наречения „rolling shutter“ ефект, присъщ за камерите със CMOS сензор.

В **Раздел 2.2** е представен разработеният метод за стабилизация в „точка“, с приложение при стабилизацията на видеоклипове, произведени от индустриални високоскоростни камери, особено при необходимост от прецизни измервания на движението на динамичния обект спрямо даден статичен фон. По точно, предложен е комбиниран ме-

тод “3×3OF9×9”, съчетаващ базовия “N×N” метод с известния метод на „оптичния поток“ (OF), който при запазване на високата точност неутрализира ефекта „пълъене на сцената“, характерен и за други известни разработки по стабилизация в „точка“. Представени са експериментални резултати за разработените методи и алгоритми от §2.1 и §2.2, както самостоятелно, така и в сравнение с възможностите за стабилизация на известния комерсиален софтуер *Adobe After Effects CS6*.

В **Раздел 2.3** е описан разработеният бърз и относително прост подход (подходящ за работа в реално време) за откриване на предварително известни цели рекламни сценарии или произволни части от тях в базов TV видеопоток. Същността на алгоритъма се състои в представянето на видеозаписите (TV и реклами) като редица от цели числа, които характеризират продължителността (в брой кадри) на отделните сцени в тях, определени чрез разликата между два последователни кадъра в рамките на предварително избран праг. По този начин използваме сравнение на крайни целочислени редици вместо сравнение на последователности от JPEG изображения (кадри) от TV потока и потенциалните кандидати от множеството на рекламните видеоклипове.

В **Глава 3** е описана разработена многомоделна БД (3DEarDB), съдържаща различни моделни представяния на десните уши на ~100 индивиди. Използван е известният EGI метод за представяне на 3D обекти. Експериментирана е комбинация от възможности за разпознаването на ушите по метода на най-близкия съсед, симулиращи различна резолюция на 3D скенера, както и различна сложност на EGI представянето. Разгледан е EFCIOR методът на ИИКТ в контекста на разпознаването на уши по множество от техни 2D проекции. Представени са най-новите разработки в областта на 3D реконструкцията на обекти и сцени, работещи в реално време.

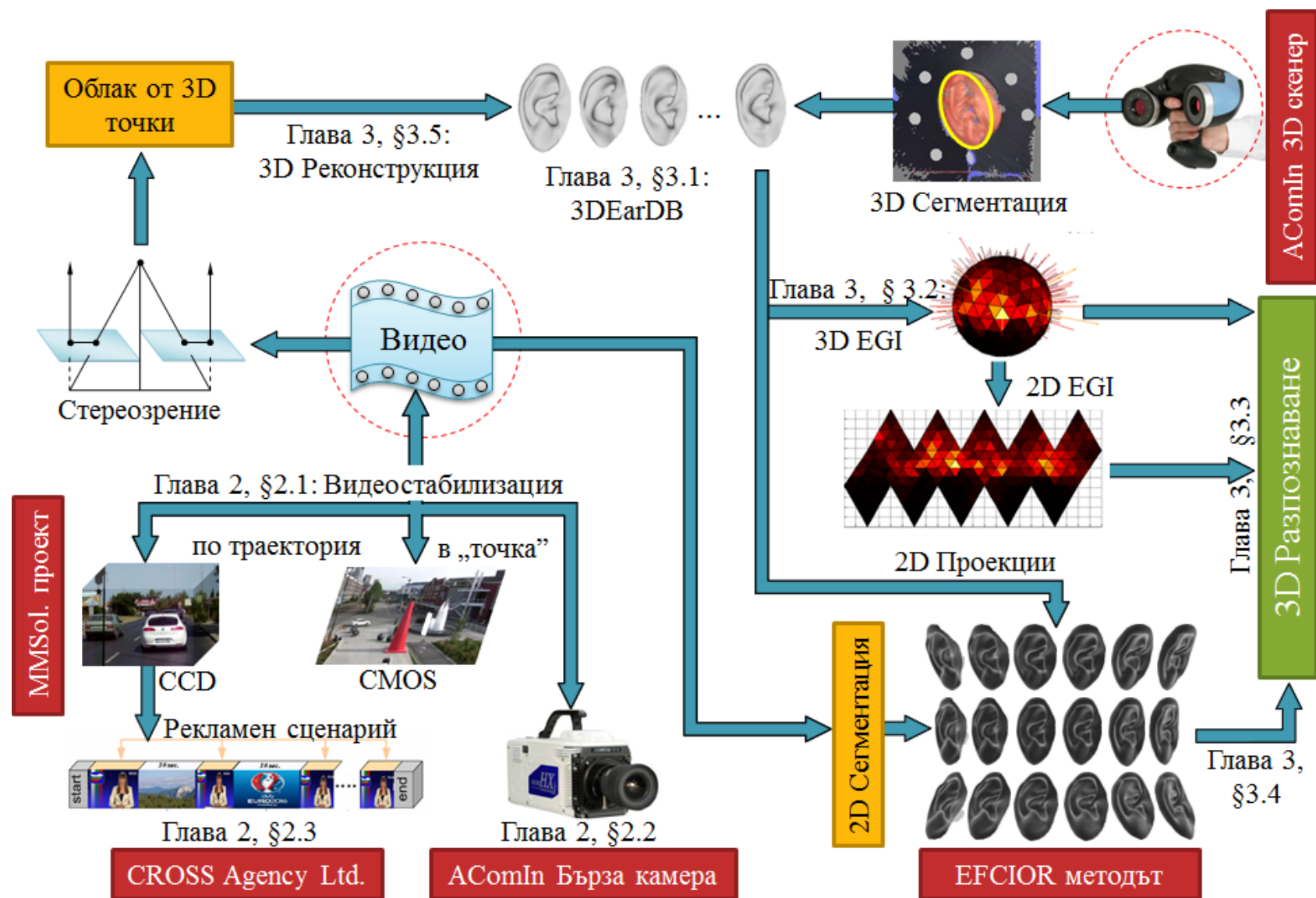
Раздел 3.1 представя създадената в ИИКТ многомоделна база данни (3DEarDB), съдържаща различни моделни представяния (2D/3D) на човешки уши, хардуерно заснети или софтуерно генерирани.

Раздел 3.2 описва същността на метода на разширеното Гаусово изображение (EGI) за представяне на 3D обекти във вид на хистограма върху единична (Гаусова) сфера.

В **раздел 3.3** са докладвани резултати от разпознаването на EGI хистограми по метода на най-близкия съсед. Извършени са експерименти върху генерирано тестово множество с различна резолюция, както на 3D скенера, така и на EGI представянето.

В **раздел 3.4** е разгледан EFCIOR методът на ИИКТ в контекста на разпознаването на уши по множество от техни 2D проекции, извлечени от 3DEarDB, като един бърз и шумоустойчив метод, представителен за областта „от 2D към 3D биометрика“.

В **раздел 3.5** са представени най-новите разработки в областта на 3D реконструкцията на обекти и сцени (работещи в реално време), докладвани на международно лятно училище по Компютърно зрение, (ICVSS'15), проведено на 12-18.07.2015г. в Италия. Очертани са насоките за бъдещата работа по дисертацията във връзка с въпроса за 3D реконструкцията на уши чрез методите на стереозрение или фотометричното стерео. Спомената е идея (в процес на експериментиране) за 3D разпознаване на уши по единствено 2D изображение чрез възстановяване на нормалните вектори на заснетия 3D обект, като разпознаването става по EGI метода, описан в §3.3.



Фиг. 2. Детайлизирана схема на изследванията по дисертацията.

Глава 1

Обзор на основните резултати в областта

1.1 Видеостабилизация

Заснетият видеоклип чрез преносимо устройство, като мобилен телефон или портативна видеокамера, страда от всички недостатъци на филмирането с нестационарна камера, защото всяко потрепване на ръката и всяко поклащане от походката се отразява като характерна и неприятна нестабилност в позицията на поредицата от кадри на видеоклипа. За справянето с проблема съществена роля играят методите за видеостабилизация, които целят подобряване качествата на нестабилно заснето видео (любителско и/или професионално) в точка или по гладка траектория. Под стабилизация в точка се разбира стабилизацията на всички видео кадри спрямо един референтен кадър, в резултат на което се имитира видео, заснето от неподвижна камера. Стабилизацията по траектория произвежда видео, при което камерата следва оригиналното движение на оператора, но в неговия изгладен вариант. В настоящия обзор ще се спрем на методите за видеостабилизация по гладка траектория, а относно „стабилизацията в точка“ отправяме към §2.2 от Глава 2.

За разлика от любителско заснетото видео, професионалните кинематографи внимателно планират движението на камерата по плавни, прости траектории, използвайки голямо разнообразие от сложно оборудване: стативи, камери, движещи се върху релси, операторски колички, преносими стабилизиращи механизми, закрепени към тялото и др. Подобна апаратура в много случаи е непрактична (или скъпа за любители), поради което широко се използват софтуерни средства за видеостабилизация, т.е. за постигане на плавност на движението в неспринудено заснето видео.

Едни от най-актуалните софтуерни реализации за (2D / 2.5D) видеостабилизация са имплементирани в свободния on-line софтуер на *Google YouTube Stabilizer* [1], както и в инструмента *Warp Stabilizer* [2] на платения професионален софтуер за видеообработка *Adobe After Effects*. Но както те, така и многото други подходи за 2D/3D стабилизация от последните години, все още не могат да работят в реално време на компютър от среден клас, да не говорим за вградени системи. Прочее, значително ускорение в бързодействието може да се постигне чрез използването на паралелна имплементация на GPU, на достатъчно бърз компютър [2, 3].

Но тъй като компютърните мощности в преносимите устройства са доста по-слаби от тези на десктоп машините, определено се разчита на използването на инерциалните сензори (жироскопи и/или акселерометри) в хардуера им, както и на съвместяването на софтуерни с хардуерни подходи за постигането на приемлива видеостабилизация.

1.1.1 Софтуерни методи за видеостабилизация

В литературата са известни множество софтуерни методи за видеостабилизация, които може да се класифицират като 2D, 3D и като така наречените комбинирани 2.5D (които се оказват доста перспективни напоследък).

A. 2D методи за видеостабилизация

Двумерните техники за видеостабилизация изчисляват 2D модел на движението (твърдотелна сцена, афинен, перспективен – виж **Приложение 1**) между съседните кадри, след което параметрите на този модел се изглаждат във времето и накрая се прилага геометрична трансформация към кадрите от оригиналното движение за получаване на кадри от изгладения модел на движението [1, 4, 5].

Известно е, че нежеланите нестабилности във видеото могат да бъдат ефикасно изгладени чрез 2D моделиране на движението между кадрите в случаите на: (i) приблизително равнинни сцени (където девиациите в дълбочината на обектите в сцената са много по-малки от средната дълбочина до сцената) с произволно движение на камерата или (ii) когато движението на камерата е чисто ротационно без значение от дълбочината в сцената. Предполагайки, че геометрията на сцената и движението на камерата попадат в тези две категории, 2D стабилизиращите методи са прецизни, работят върху целия кадър, изискват следене на малък брой характерни точки (траектории във времето) и консумират минимална изчислителна мощност. Всъщност, 2D стабилизацията е намерила широко приложение в цифровите фотоапарати и видеокамери, където е реализирана чрез механични части, които движат или лещите, или сензора на устройството [6].

В повечето случаи обаче, сцените съдържат обекти на произволна дълбочина и освен това, в не принудиено любителско видео, е невъзможно да се избегне транслационният компонент на движението на камерата. В тези случаи, 2D напасването на кадрите *не може да моделира паралакса*, породен от транслацията на гледната точка, и поради тази причина двумерното моделиране на сцената е недостатъчно на създаване на ефективна видеостабилизация. *Второто ограничение* на двумерното моделиране на движението е, че липсва информация за тримерната траектория на камерата, което прави невъзможно симулирането на идеализиран път на движението, подобно на професионалните методи за заснемане на видео.

Идеята за преобразуване на любителски заснето видео да изглежда, че все едно е получено чрез съответното кинематографско оборудване, е била осъзната за първи път от Gleicher et al. [7]. Въпреки това, техният подход е базиран на 2D напасване между кадрите и затова страда (както всички 2D подходи) от горните две ограничения.

Б. 3D методи за видеостабилизация

Тримерният подход за видеостабилизация е описан за първи път от Buehler et al. [8]. Там е приложена техниката „структура-от-движение“ (SfM), където се проследява множество от характерни точки във времето, след което се използва тяхното съответствие, за да се възстанови тримерната позиция и ориентация на камерата, както и тримерното местоположение на всяка следена характерна точка [9]. След това се използва техниката „image-based rendering“ [10, 11, 12], чрез която от даден брой входни кадри се синтезират нови кадри, все едно видени от други гледни точки, разположени по желания път на камерата с цел преобразуване на входното любителско движение в ново, плавно тримерно движение на камерата. Но е известно, че горните методи са ограничени само до статични сцени.

След като през 2009 г. Liu et al. [13] реализират за първи път 3D видеостабилизация за динамични сцени, въпреки високото качество на плавноста на движението се вижда, че нейното практическо приложение е ограничено от необходимостта за извършване на 3D реконструкция на сцената, защото използваният за целта SfM метод има сериозни проблеми с устойчивостта и общата приложимост, тъй като някои видео клипове не съдържат достатъчно информация за движението. Тези ситуации включват:

1) *Липсата на паралакс* – когато движението е чисто ротационно по x или y оста на камерата, без никаква трансляция, или когато видеото съдържа голяма плоска област.

2) *Промяна на фокусното разстояние (мащабиране)* – това е добре известен проблем за SfM, понеже методът не може да направи разлика между приближаване/отдаляване и движение напред/назад.

3) *Собствена видеостабилизация на камерата* – повечето съвременни фото- и видеокамери намаляват трептенето на камерата или оптично, или цифрово, чрез промяна на вътрешните параметри на камерата по време на работа. Обаче, повечето техники за SfM предполагат, че тези параметри са фиксирани.

4) *“Rolling shutter” ефектът* – повечето нови камери използват CMOS сензор, който не предава целия кадър наведнъж, а ред по ред във времето, което причинява (вертикално) поклатане “wobble” и (хоризонтално) накланяне “shear” на някои кадри във видеото. Въпреки това, скорошни изследвания демонстрират възможността за премахване на тези артефакти [14, 15, 16].

Производителността на SfM също е проблем, понеже изисква глобална нелинейна оптимизация. Повечето SfM реализации не са поточни, т.е. изискват произволен достъп до цялото видео, вместо само до прозорец от кадри около текущия.

В. 2.5D методи за видеостабилизация

Поради споменатите проблеми при 3D видеостабилизацията, напоследък се наблюдава тенденция към предпочитане на 2.5D подходите, които съвместяват качествата на 3D методите с производителността и устойчивостта на 2D методите [2, 17, 18, 19, 20].

Известно е, че съвременните методи за 2D видеостабилизация с претенции за постигане на резултат, еквивалентен на 3D стабилизацията, разчитат на откриването на съответстващи характерни точки между последователните кадри. Но тези точки често са неравномерно разпределени в сцената, което води до известни проблеми за модели-

ране на 2D напасването между съседните кадри. Затова се търсят подходи, които освен елиминиране на грешните съответствия, да добавят и нови съответствия на произволни (подходящи) места в сцената за постигане на равномерност на разпределението. Един такъв подход е описан в [17], където се разглежда епиполярната геометрия между съседните кадри.

Друг основен проблем на 2D методите за стабилизация е невъзможността да моделират паралакса чрез глобалното напасване на последователни кадри. Затова от съществено значение е да се приложат методи за локално напасване на отделните части от последователните кадри, след което резултатът да се обедини в един физически коректен кадър, оптимално напаснат спрямо предходния. Едва най-нови разработки в областта се справят с този проблем, в които дори отпада изискването за абсолютно равномерно разпределение на характерните точки в сцената, а празните области се стабилизират според околните им [18, 20], но всичко това на цената на неприемливо ниска скорост на обработките.

1.1.2 Исторически преглед на върховите постижения при софтуерната видеостабилизация

Тук ще разгледаме развитието във времето⁶ на върховите постижения на различните методи за видеостабилизация (2D, 3D и 2.5D), всеки един от тях, претендиращ за постигане на “state-of-the-art” резултат в сравнение с предходните методи.

A. (2009) Content-Preserving Warps for 3D Video Stabilization [13]

Това е първият метод, който успява да осъществи 3D видеостабилизация върху динамични сцени. За разлика от предишните методи за 3D видеостабилизация, тук динамиката на обектите се възпроизвежда правилно, защото всеки видео кадър се преобразува в съответния му стабилизиран кадър независимо, а не според синтезирана нова гледна точка чрез комбиниране на съдържанието от множество видео кадри, и затова в различни моменти във времето.

Алгоритъмът се състои от следните три етапа:

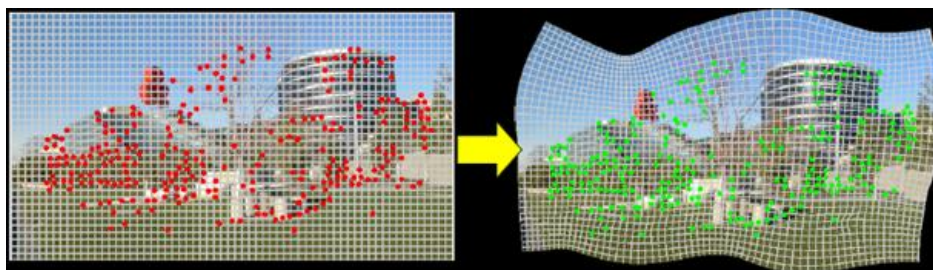
- 1) Определяне на оригиналното тримерно движение на камерата в пространството, както и позицията на разрежен облак от тримерни статични точки в сцената чрез SfM метода. За целта авторите използват свободно достъпната система Voodoo Camera Tracker⁷, която обаче изисква часове работа дори за кратка видео последователност – между 4 и 10 часа за 20 сек. видео (~30 fps). Хубавото е, че съществуват и SfM системи, които работят в реално време, но всички те изискват калибрирани (с фиксирани вътрешни параметри) видео камери [21, 22].
- 2) Потребителят интерактивно определя желания път на камерата или избира една от три възможности какъв да бъде пътят: линеен, параболичен или изгладен вариант

⁶ В периода 2009-2013г. се наблюдава засилен интерес към методите за видеостабилизация от страна на научната общност в областта, характерен със световно значимите резултати в 2D и 3D стабилизацията.

⁷ www.digilab.uni-hannover.de

на оригиналната тримерна траектория. Авторите са се фокусирали върху тези движения на камерата, защото те имат изключителен ефект откъм възприятието и са трудни за пресъздаване без специализирано оборудване.

- 3) Трансформиране на всеки входен кадър от първоначалното движение в кадър от новата траектория на камерата. За целта, авторите се възползват от метода „деформация със запазване на формата на обектите“ [23, 24], който минимизира изкривяването на локалните форми на обектите според контролни точки, зададени от потребителя. Тук обаче се деформира квадратна мрежа, наложена върху оригиналното изображение, според важността на информацията във всяка клетка, където финалното изкривяване в изображението е резултат не от желанието на потребителя, а от промяната на гледната точка в тримерното пространство (Фиг. 3). По този начин се цели моделиране на паралакса, породен от различната дълбочина до обектите в сцената, чиято изкуствена корекция е необходима за изгладения преход между съседните кадри за целите качествената 3D стабилизация.



Фиг. 3. Синтезиране на нови изгледи за качествена 3D стабилизация. Заимствано от [13].

Тъй като промяната в гледната точка, която се изисква от метода, по-принцип е малка, този физически некоректен подход за интерполиране на нови изгледи е задоволителен дори за предизвикателния случай на видео с динамични обекти.

Резултати и сравнения

За оценка на получените резултати авторите правят сравнение с други методи, като 2D стабилизацията на Matsushita et al. [5], възможността за 2D стабилизация на iMovie '09⁸ и Deshaker plugin⁹, виж и примерите в Liu et al (2009)\Additional Videos\Example 1,2,3,4.

Според демонстрираните резултати предложеният метод [13] постига по-стабилно и по-насочено движение, отколкото 2D техниките, и същевременно коректно възпроизвежда движенията на динамични обекти, като предотвратява ефекта „призрак“ при предишните 3D техники, виж Liu et al (2009)\main-demo (Content Preserving Warps).mp4.

Ограничения

Основният недостатък на предложеният метод [13] в сравнение с 2D видеостабилизацията е, че той изисква изпълнението на SfM, което представлява „тясното място“ на метода (синтезирането на нови изгледи отнема около 2 кадра в секунда). Освен това, SfM е ограничен до видео последователности, които съдържат само статични области, в

⁸ www.apple.com/ilife/imovie

⁹ www.guthspot.se/video/deshaker.htm

противен случай движението на камерата не може да се определи еднозначно. Също така, традиционният алгоритъм на SfM изисква постоянно транслираща камера. Методът също зависи от качеството на тримерната реконструкция на сцената, както и от разпределението на намерените SfM точки от сцената във всеки видео кадър. За подробности относно ограниченията на SfM виж т. Б. от §1.1.1.

Заклучение

Ключовият момент на разглеждания метод се състои в това, че малки промени на гледната точка в тримерното пространство (т.е. паралакса) могат да бъдат имитирани чрез внимателното конструиране на „деформация със запазване на значимото съдържание“, водеща до синтезиране на нови изгледи, въпреки че резултатът не е физически коректен.

Б. (2011) Subspace Video Stabilization [2]

Настоящата работа представлява първия подход за видеостабилизация, който комбинира предимствата на 2D и 3D видеостабилизацията, което означава съвместяване на високо качествения резултат на 3D методите за стабилизация (но без необходимостта от тримерна реконструкция) с производителността и устойчивостта на 2D методите. Освен това, той работи почти в реално време, изпълнява само линейни алгоритми и може да се изчисли в поточен „streaming“ режим.

Когато видеото съдържа „rolling shutter“ артефакти, това представлява сериозен проблем както за SfM, така и за традиционната 2D видеостабилизация [14, 25]. На практика, предложеният подход работи добре и с такъв тип видео, въпреки че авторите не могат да се аргументират теоретично защо това е така и смятат, че тези артефакти се проявяват като шум, към който техният алгоритъм е устойчив.

Кратък преглед на представения алгоритъм

- 1) Изчисляват се 2D траектории от характерни точки във входното видео, за чието проследяване в последователността от кадри е използван стандартният KLT подход [26].
- 2) 2D траекториите от характерни точки на сцената се асемблират в т. нар. „матрица на траекториите“, която е изключително разрежена, понеже траекториите се появяват и изчезват по време на видеото. Тази матрица впоследствие се разлага на две матрици от по-нисък ранг – матрица с коефициенти и матрица на линейно независимите траектории, наречени от авторите собствени траектории „eigen-trajectories“.
- 3) Изглаждане на собствените траектории и умножаването им с оригиналната матрицата от коефициенти, в резултат на което се получават стабилизирани траектории.
- 4) Синтезиране на нови изгледи от първоначалните кадри в съответствие с изходните изгладени траектории, използвайки метода „деформация със запазване на значимото съдържание“ на Liu et al. [13], виж т. 1 от §1.1.2.

Резултати и сравнения

Авторите са направили сравнение на техния метод с 3D стабилизацията на Liu et al. [13], както и с публично наличните 2D стабилизиращи методи на iMovie'09 и Deshaker.

За случаите, където 3D видео стабилизацията [13] се справя добре, разглежданият метод демонстрира качествено подобни или по-добри резултати, виж и примерите в Liu et al (2011)\Supplemental Video Set 1\Ex.1,2,3. Резултатите, получени от представения метод, са също по-добри от тези на iMovie и Deshaker, изследвани върху последователности с “rolling shutter” (понеже 2D напасването между кадрите не може да премахне локалните деформации), както и върху сцени с голяма разлика в дълбочината (защото 2D модели на афинна или проективна трансформация не могат да моделират паралакс), виж Liu et al (2011)\main-demo (Subspace Video Stabil).mp4.

Ограничения

Най-големият недостатък на предложения метод е, че когато не са налични достатъчен брой дълги траектории, за да покрият целия прозорец от k на брой кадъра около текущия, разлагането на собствени траектории е неуспешно, в резултат на което алгоритъмът не може да приключи и дори не може да произведе изходен резултат. Съществуват различни възможни причини за липсата на дълги траектории: (i) когато големи движещи се обекти минават през- и закриват сцената; (ii) силно замазване “blurring” от дефокусиране и/или високочестотни трептения на камерата; (iii) изключителна нестабилност на камерата, което изважда траекториите извън зрителното поле на камерата; (iv) обекти с малко текстура. Друг често срещан случай на неуспех е, когато има достатъчно на брой дълги траектории, но значителна част от тях отразяват движението на голям движещ се обект, доминиращ в сцената, виж Liu et al (2011)\Supplemental Video Set 3.

Производителност

Проведените експерименти са извършени върху компютър с процесор Intel Dual Core 3.16 GHz. Времето за изпълнения на метода, разбито в трите главни стъпки от алгоритъма, е следното: (i) 7 fps за следене на характерни точки във видеото, когато KLT тракерът е настроен да следи 500 характерни точки за кадър, с размер 640×360 пиксела; (ii) 500 fps за разлагане на матрицата с траектории; (iii) 10 fps за синтезиране на нови изгледни. Общо текущата реализация постига 4 fps, но с използването на паралелна имплементация и GPU може да се постигне и производителност в реално време. В подкрепа на това, може да се използва GPU ускорената версия¹⁰ на KLT тракера на характерни точки [27].

Софтуерна имплементация

Предложеният метод за видеостабилизация е внедрен под името Warp Stabilizer¹¹ в комерсиалния софтуер на Adobe After Effects CS5.5, който се използва за създаване на визуални ефекти в процеса на пост-продукция в правенето на филми или телевизионни предавания.

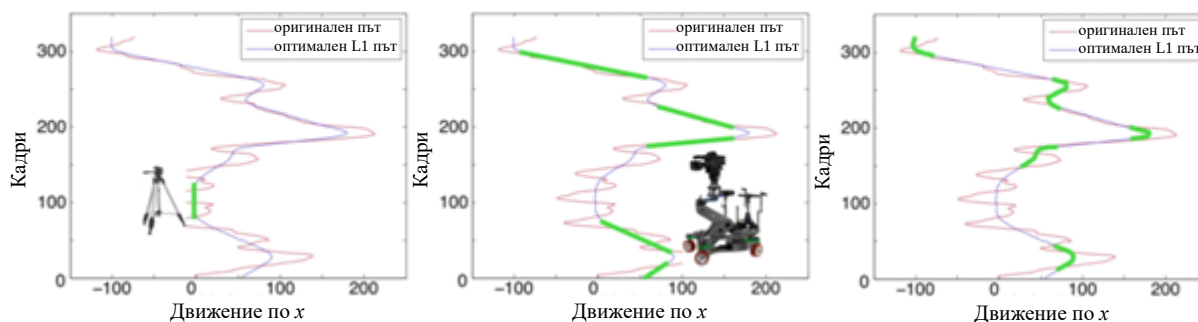
¹⁰ <http://www.ces.clemson.edu/~stb/klt/>

¹¹ www.juaw.org/techtransfer.htm

B. (2011) Auto-Directed Video Stabilization with Robust L1 Optimal Camera Paths [1]

Авторите на този метод твърдят, че са постигнали “state-of-the-art” резултат във видео стабилизацията към датата на публикуване на статията (юни, 2011). Целта на техния метод е да имитират професионално снимачно оборудване, където най-приятно усещане за плавност на видеото се получава чрез: (i) използването на статична камера; (ii) такава, която извършва единствено ротация около x или y осите на камерата; или (iii) камера, поставена върху операторска количка. За постигане на споменатия ефект авторите налагат ограничението, траекторията на камерата да бъде съставена само от константни, линейни или параболични сегменти (Фиг. 4) чрез минимизиране на първата, втората и третата производна на оптималната стабилизирана траектория:

- постоянен път (липса на движение между кадрите), представлящ статична камера, „траектория в точка”, т.е. първата производна на такъв сегмент от траекторията трябва да бъде нула: $DP(t) = 0$, където D е диференциален оператор.
- път с постоянна скорост (линейно движение), представлящ ротационно движение или видео, заснето от операторската количка, т.е. $D^2P(t) = 0$.
- път с постоянно ускорение, представлящ прехода между статична камера и камера с линейно движение, т.е. $D^3P(t) = 0$.



Фиг. 4. Представяне на траекторията на камерата чрез статични, линейни и параболични сегменти. Заимствано от [1].

Предложената оптимизация на траекторията на камерата е свързана с метода “L1 trend filtering” [28], който извършва минимизиране на втората производна чрез „L1 нормата” по метода на най-малките квадрати, като по този начин множеството от точки се апроксимира само чрез линейни сегменти.

Разглежданият алгоритъм е по-общ, защото освен с линейни, той апроксимира желаните стабилизиран път и с точкови и параболични сегменти (чрез минимизиране на първата и третата производна).

Кратък преглед на алгоритъма

- 1) Изчисляване на 2D параметричен модел (чрез трансформация на подобие и проективна трансформация, виж *Приложение 1*) между кадрите чрез следене на характерни точки, използвайки KLT подхода на Shi and Tomasi [26].
- 2) Определяне на оптималния път на камерата, който е съставен от точкови, линейни и параболични сегменти, имитиращ движението на камерата на професионалните ки-

нематографи. Предложена е L1 оптимизация за намиране на оптималния стабилен път $P(t)$, минимизирайки следната целева функция:

$$O(P) = w_1 \|D(P)\|_1 + w_2 \|D^2(P)\|_1 + w_3 \|D^3(P)\|_1, \quad (1)$$

при две ограничения: за „близост” (новият път на камерата трябва да следва преднамереното движение на оператора) и за „включване” (припокриването между оригиналните и изгладените кадри да бъде в рамките на зададен фиксиран прозорец).

- 3) Стабилизиране на видеото чрез хибриден подход за моделиране на движението между кадрите. Първоначално се изчислява трансформация на подобие между всеки 30 кадъра, което осигурява устойчивост срещу изкривявания, причинени от “rolling shutter” ефекта. След това, за вътрешните кадри от всеки отрязък, трансформацията на подобие се замества с проективна, като по този начин се изчислява действителната трансформация.

Резултати и сравнения

- авторите демонстрират, че са постигнали същия резултат като Liu et al. [13], но избягвайки тежкото изпълнение на 3D реконструкция на сцената;
- методът на Liu et al. [13] не може да се справи с ефекта на “rolling shutter”, докато по-новият им вариант (Liu et al. [2]) се справя, но по-лошо, отколкото разглежданите L1 метод;
- текущо предложеният L1 метод доста добре може да разграничи движението на обекта от движението на фона в сравнение с метода на Liu et al. [2], виж Grundmann et al (2011)\main-demo (L1 Optimization).mov.

Ограничения

Предложеният L1 метод за видеостабилизация не може да стабилизира всички видове движения, поради: (i) недостатъчен брой характерни точки за следене; (ii) прекомерно размазване, породено от изключително бързи движения; (iii) липса на неподвижни обекти в сцената. Освен тези недостатъци, изчисленият стабилизирания път е оптимален за фиксирани “cropped” размери (ширина и височина) на изходните кадри, което води до загуба на информация.

Софтуерна имплементация

Разработеният по [1] софтуер *Google YouTube Stabilizer*¹² може да бъде използван свободно on-line след регистрация на потребителя. Той работи като пост-обработка на качените видео клипове, въпреки че може да постигне до 20 fps върху видео с ниска разделителна способност.

¹² www.youtube.com/editor

Г. (2012) Video Stabilization using Epipolar Geometry [17]

В тази статия е представена една нова техника за видеостабилизация, която използва междинно ниво на моделиране (между 2D модела и пълния 3D модел на сцената), известен като проективна реконструкция [9].

Този модел не възстановява явно тримерните позиции на точките от сцената или позицията на камерата, а вместо това отчита прости геометрични зависимости между точки и епиполарни линии, което му позволява да работи върху 3D сцени с подвижни обекти. Това интерпретиране на сцената е достатъчно за получаването на физическата коректност както при пълната 3D стабилизация, но с преодоляване на нейната неустойчивост към различни ситуации, както и голямата изчислителна сложност (виж т. Б. от §1.1.1).

Кратък преглед на алгоритъма

- 1) Изчисляване на траектории от характерни точки във видео последователността чрез стандартния KLT алгоритъм за следене на точки (Shi and Tomasi [26]), реализиран в софтуера Voodoo camera tracker.
- 2) Стабилизираната траектория на камерата се изчислява чрез изглаждане на множество виртуални траектории, генерирани чрез „епиполарния трансфер на точки“ [9, 29]. Ползата от този трансфер на точки е, че не се разчита на KLT тракера за намиране на дълги траектории, които са необходими за ефективната стабилизация.
- 3) Намиране на допълнителни траектории във входното видео, за разрешаване на проблема с нееднородното разпределение на следените точки в кадрите.
- 4) Синтезиране на нови, стабилизирани кадри чрез метода „деформация със запазване на значимото съдържание“ на Liu et al. [13], виж т. 1 от §1.1.2.

Резултати и сравнения

Всички демонстрации на [17] с пълна разделителна способност са налични на уеб адреса: <http://www.cs.huji.ac.il/~raananf/projects/stab/videos/>.

- Video #1 и #2 съдържат дълбочинни сцени, за които предложеният метод се справя съпоставимо с метода за 3D видеостабилизация, предложен от Liu et al. [13].
- Във Video #3, #6, #8, #9, #10, #11 е направено сравнение с метода “*Subspace Video Stabilization*” на Liu et al. [2]. Тези видео примери потвърждават възможността на разглеждания метод да генерира резултати, качествено подобни с тези на [2], върху сцени, където: (i) липсва паралакс; (ii) е включена вградената видеостабилизация на камерата; (iii) има промяна на фокусното разстояние; и (iv) налице са слаби “*rolling shutter*” ефекти.
- Video #12 представя движение на човек, който заема голяма част от кадъра. Всъщност, методите на Liu et al. [2] и [13] отхвърлят характерните точки, принадлежащи на динамични обекти, като тези области от кадрите получават своята стабилизация от други точки, принадлежащи на статичния фон. Обаче когато разликата в дълбочината между обект и фон е значителна, не може да се получи адекватна стабилизация. За разлика от тях, в предложения метод, авторите преодоляват необходимостта да интерполират стабилизацията на движещи се обекти чрез техния статичен фон.

- На Video #16 е показано предимството от генерирането на допълнителни траектории в липсващите области от кадрите.

Ограничения

Предложеният метод за видеостабилизация силно зависи от броя и точността на следените характерни точки. Сцени с малко текстура, липса на статичен фон, значителни припокривания от големи динамични обекти в сцената, размазване от високочестотни трептения, прекомерно клатене на камерата и др., възпрепятстват успешната видеостабилизация.

Понеже методът използва епиполярната зависимост „точка-права“ между различните изгледи, за разлика от метода на Liu et al. [2], не може да се справи със силни “rolling shutter” ефекти, които нарушават горните зависимости.

Друго ограничение на метода в сравнение с 3D видеостабилизацията е невъзможността да осигури явно планиране на движението на камерата, виж Video #17.

Заклучение

Ключовите особености на разгледания метод се състоят в следното:

- постигане на робастна видеостабилизация без необходимост от 3D реконструкция;
- бързодействието може да достигне до 30 fps със забавяне от 5 до 10 кадъра;
- работи и при наличието на движещи се обекти (динамични сцени);
- освен за самостоятелно вграждане в цифрови видео камери, методът може да се използва и като част от други методи за видео обработка.

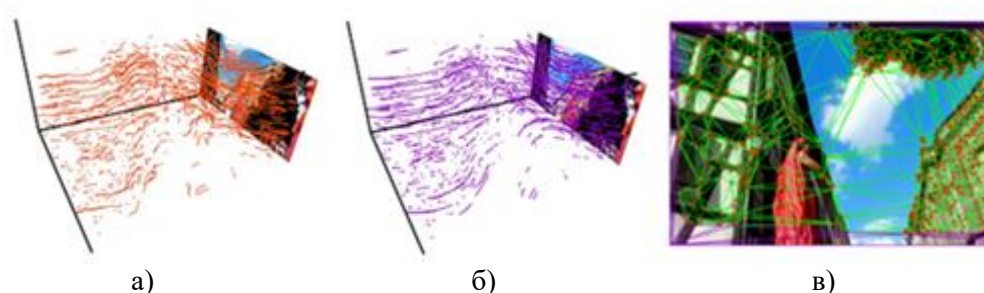
Д. (2013) Spatially and Temporally Optimized Video Stabilization [20]

Тук авторите демонстрират ефективността на техния метод чрез сравнението му с три “state-of-the-art” подхода: на Liu et al. [2], Grundmann et al. [1] и на Goldstein et al. [17], върху предизвикателно многообразие от видео материал (съдържащо голям паралакс; големи обекти, заемащи почти целия кадър; бързи движения на камерата), които затрудняват съществуващите методи за видеостабилизация. Експериментите показват, че предложената техника демонстрира качествено подобни или по-добри резултати в сравнение с предишни методи, което я нарежда на челна за времето си позиция.

След намирането на траекториите от характерни точки, новостта на метода се състои в това, че изглаждането на траекториите запазва относителното отместване между съседните точки във всеки кадър, като по този начин се постига висококачествена стабилизация без необходимостта от наличие на дълги траектории, както при Liu et al. [2] и Goldstein et al. [17], нито от 3D реконструкция на сцената (Liu et al. [13]). Изгладените траектории са представени чрез криви на Безие, което позволява на метода директно да изглади всяка траектории и по този начин да използва едновременно дълги и къси траектории за по-доброто моделиране на наличния паралакс. Друга причина за използването на криви на Безие е постигането на висока производителност, понеже всяка траектория се интерполира чрез малък брой контролни точки, вместо с всички характерни точки.

Преглед на представения алгоритъм

- 1) Изчисляване на траектории от характерни точки чрез софтуера Voodoo camera tracker, имплементиращ KLT метода на Shi and Tomasi [26]. След това, подобно на метода на Liu et al. [2], намерените траектории върху движещи се обекти се премахват.
- 2) Изглаждане на траекториите със запазване на пространствените връзки между съседните характерни точки във всеки кадър. Идеята на авторите е нестабилните траектории да се представят чрез криви на Безие, а връзките между съседните точки да се осъществят чрез метода на “*delaunay triangulation*”, като се налага ограничението всеки триъгълник да може да претърпи само Евклидова “*rigid*” трансформация в процеса на изглаждане на траекториите. Това означава, че на триъгълниците е „позволено“ само да се движат и ротират (виж *Приложение 1*).



Фиг. 5. Оригиналните (а) и стабилизирани (б) траектории от характерни точки в пространствено времева координатна система. Запазване на относителното отместване между съседните точки във всеки кадър за правилното моделиране на паралакса (в). *Заимствано от* [20].

- 3) Синтезиране на нови, стабилизирани изгледи от първоначалната видео последователност от кадри, подобно на метода на Liu et al. [13].

Резултати и сравнения

Авторите правят сравнение на техните резултати с текущите “*state-of-the-art*” техники, а именно:

- „L1 оптимизацията“ на Grundmann et al. [1];
- “Subspace” методът на Liu et al. [2];
- „Елиполярният метод“ на Goldstein et al. [17];
- 3D стабилизация на Liu et al. [13] не е включена в сравнението, защото по-късните два подхода [2] и [17], също могат да моделират паралакс и освен това демонстрират по-устойчив резултат.

Текущо разглежданият метод на Wang et al. [20] демонстрира, че може да постигне качествено подобни или по-добри резултати върху видео примерите, представени в съответните статии на предходните “*state-of-the-art*” методи, виж Liu et al (2013)/Video/CompToLiu.mp4, CompToGoldstein.mp4 и CompToGrundmann.mp4.

Относно проблема за припокривания на сцената от големи обекти, само методът на Grundman et al. [1] може да се справи с тях, докато техниките на Liu [2] и Goldstein [17] дори не могат да произведат резултат, защото изискват дълги траектории. Въпреки това, поради своята 2D същност, методът на Grundman et al. [1] е ограничен в моделира-

нето на паралакса, което се проявява като глобално изкривяване на сцената след направената стабилизация, виж *Liu et al (2013)/Video/ Main.mp4* [2:20 – 4:09].

Въпреки че методите на Liu [2] и Goldstein [17] могат добре да моделират паралакс, техните резултати не са толкова добри, когато разликата в дълбочината между много близък обект и фона е значителна, виж *Liu et al (2013)/Video/ Challenging.mp4*.

Ограничения

Представеният метод разчита единствено на следенето на характерни точки от фона на сцената (точките от траекториите на динамичните обекти се елиминират), за да стабилизира видеото. Затова, при наличие на прекомерно размазване, причинено от изключително бързи движения на камерата, или при липсата на статични обекти от фона, методът не може да стабилизира някои кадри, които се проявяват като нестабилни, виж *Liu et al (2013)/Video/ Limitation.mp4*.

Заклучение

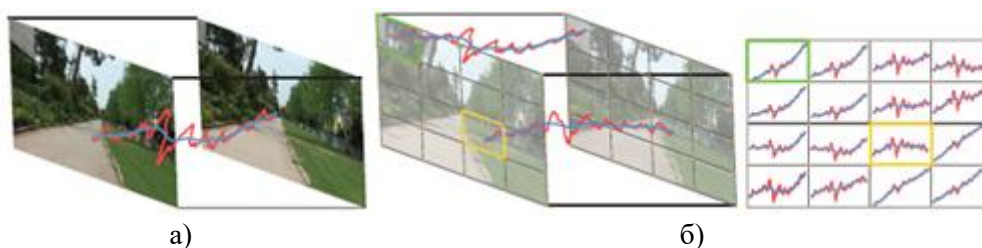
В [20] е представена най-добрата техника за видеостабилизация към момента на публикуването на статията (август, 2013). Този извод е направен на база на това, че разглежданият метод се справя успешно с много от предизвикателните случаи, затрудняващи предходните методи, и освен това продуцира качествено сравними или по-добри резултати с тях върху непредизвикателно (неекстремно) видео.

Освен това, разглежданата техника за видеостабилизация може да постигне производителност в реално време дори при стабилизирането на видео с високо качество, но при условие, че траекториите от характерни точки са изчислени предварително.

E. (2013) Bundled Camera Paths for Video Stabilization [18]

В тази разработка е предложено „по-мощно“ 2D моделиране на движението на камерата в сравнение с предишните методи с цел постигане на робастен висококачествен резултат. Предложеният модел на движението е базиран на така наречения пространствено-вариращ перспективен модел върху правоъгълна мрежа, възползващ се от последните новости в техниките за синтезиране на изгледи [13, 23, 24].

По-конкретно, авторите моделират движението чрез „сноп траектории“, които представляват множество от пространствено вариращи траектории за различни области (от правоъгълна мрежа) на кадрите от видеоклипа. С други думи – всяка предварително дефинирана клетка от видеоклипа има своя траектория, изчислена чрез конкатенация на локални перспективни трансформации във времето. След това, различните траектории се изглаждат чрез оптимизационна процедура, така че да се запази плавността между траекториите в съседните клетки с цел минимизиране на изкривяванията във финалния стабилизиран резултат.



Фиг. 6. Сравнение между традиционната 2D стабилизация с единствена глобална траектория (а) и „снопа от локални траектории“ (б), изгледени във времето. *Заимствано от [18].*

Този гъвкав локално перспективен модел се справя успешно както с нелинейни движения, причинени от паралакса (в резултат от дълбочината на сцената), така и с “rolling shutter” ефекти. В същото време, методът се възползва от устойчивостта и бързодействието на 2D методите, поради изчисляване само на линейни трансформации между кадрите, което изисква единствено намиране на съответствия между характерни точки между два последователни кадъра.

С проведените експерименти авторите твърдят, че техният метод е сравним по качество или превъзхожда другите конкурентни 2D и 3D методи за видеостабилизация.

Резултати и сравнения

За оценка на резултатите авторите предлагат на научната общност по видеостабилизация обширна база данни¹³ от 174 кратки видеоклипа (10-60 сек.), включваща известни примери от предишни статии, от интернет, а също и собствени видеоклипове.

Първото сравнение е извършено върху публични видеоклипове (Фиг. 7) на успешните резултати на “state-of-the-art” методите от предходните години:



Фиг. 7. Референтни видеоклипове от предишни успешни методи, за сравнение с текущия метод. *Заимствано от [18].*

- ✓ „3D стабилизация“ на Liu et al. [13];
- ✓ „L1 оптимизацията“ на Grundmann et al. [1];
- ✓ “Subspace” методът на Liu et al. [2];
- ✓ „Епиполярият метод“ на Goldstein et al. [17].

Критериите за оценка са стабилност, изкривяване и размери на изходното видео.

- За Видео # 2-4, 3D стабилизация на Liu et al. [13] постига най-добри резултати в стабилността и минимум на изкривяването с много малка разлика с текущо предложения метод, където пък степента на изрязване на видеото е доста по-добра.
- За Видео # 5-9 е извършено сравнение с “Subspace” метода на Liu et al. [2] и „Епиполярият метод“ на Goldstein et al. [17]. Текущият метод постига подобни резултати по отношение на стабилността, но ги превъзхожда по минимално изрязване на видеото и по-малка степен на изкривяване.

¹³ www.liushuaicheng.org/SIGGRAPH2013/index.htm

- За Видео # 10-11, текущият метод превъзхожда „L1 оптимизацията“ на Grundmann et al. [1] по всичките три показателя. А в “Subspace” метода на Liu et al. [2] докладват за Видео # 10 като провален случай, заради голямата разлика в дълбочината на обекта и фона, с която не може да се справи техният метод.

За горните експерименти виж *Shuaicheng et al(2013)/Video/main_demo_BundledPaths.mp4*

Следващото сравнение е извършено със “state-of-the-art” системите за видеостабилизация *Google YouTube Stabilizer* и *Warp Stabilizer* на *Adobe After Effects* върху собствена БД от видеоклипове, която е разделена на 7 категории според движението на камерата и вида на сцената. Това са: (i) *обикновени*; (ii) *бърза ротация*; (iii) *мащабиране*; (iv) *голям паралакс*; (v) *шофиране*; (vi) *тълпа*; (vii) *бягане*. Виж *Shuaicheng et al (2013)/Video Database/*.

И трите метода произвеждат добри резултати за категория (i), понеже трептенето там е слабо и разликата в дълбочината на сцената не е голяма. Категории (ii) и (iii) са предизвикателни за методите, изискващи дълги траектории, като например *Warp Stabilizer*, който значително изрязва размерите на изходното видео. За категория (iv) и трите системи постигат подобна стабилност, докато текущият метод [18] ясно превъзхожда другите два по отношение на изкривяването и демонстрира възможността да се справя с голям паралакс. За категории (v) и (vii) и трите системи демонстрират съизмерима стабилност, но текущата [18] ги превъзхожда по отношение на малки изкривявания и степен на изрязване на видеото.

Ограничения

Авторите наблюдават, че когато 3D реконструкцията е успешна, 3D методите често генерират най-добър резултат. Съществуват и случаи, когато техният модел на движението не може да се справи с големи припокривания в сцената, особено в комбинация с “rolling shutter” ефекти. Освен това, оптимизираната гладка траектория не следва стриктно кинематографската традиция. Методът не разглежда и отстраняването на замазването във видеото, причинено от много бързи движения.

Производителност

Бързодействието на метода е експериментирано на десктоп машина със следните параметри: четириядрен процесор Intel i7, 3.2GHz Quad-Core с 8G памет. За видеоклип с резолюция от 1280×720 пиксела, техният неоптимизиран код отнема 392 ms за обработката на един кадър (или около 2,5 fps). По-конкретно е представена следната разбивка: 300 ms за извличане на 400-600 SURF характерни точки за кадър; 50 ms за изчисляване на модела на движение; 12 ms за оптимизиране на изгладената траектория на камерата; 30 ms за произвеждане “rendering” на финалния резултат.

1.1.3 Хардуерни методи за видеостабилизация, използващи инерциални сензори

За пълнота на обзора, тук ще споменем накратко за най-известните хардуерни методи за видеостабилизация, използващи инерциални сензори. Разработените методи в дисертацията са чисто софтуерни, като евентуалната комбинация с хардуерни методи за подобряване качеството на стабилизацията е оставено за бъдеща работа.

Класически примери за 3D стабилизация чрез инерциални сензори са [30, 31, 32]. В [30] използват 3D жирокоп и 3D акселерометър, както за стабилизация на позицията на камерата, така и за неутрализиране на познатия “*rolling shutter*” ефект при CMOS камерите. От 3D акселерометъра се използва само компонентата по вертикалата (по земното ускорение), а останалите 2 компоненти се обявяват за незначими, ако мобилното устройство прави само ротации, но не и трансляции (много разпространен случай според тях). Последното се оборва в [31], като в развитие на сложния алгоритъм [30], т.е. разширен Калман филтър върху кватерниони с изглаждане, 3D-акселерометърът от [30] се замества с проследяване на характерни точки в изображенията от видеоклипа. Резултатите са доста по добри от [30] и несъществено по-добри от [32], който е опростена версия на [30], само с един 3D-жирокоп и може да коригира “*rolling shutter*” ефекти в реално време. Тези подходи, както и други подобни, страдат от незадоволителната прецизност на сензорите, използвани в съвременните мобилни устройства.

1.1.4 Интересни методи за 2D стабилизация и тяхната връзка с разработените методи по видеостабилизация в дисертационния труд

Един доказано бърз (над 100 fps за видео 1280×960) и опростен подход за 2D видеостабилизация (но само по трансляция) е описан в *White Paper* на *Texas Instruments* (TI) [33, 34]. Методът е интересен за нас, защото използва BSC (*Boundary Signal Computation*) чипа [33], от състава на DaVinci процесора (TMS320DM365) на TI [34]. BSC позволява ефективно изчисление на вертикалните и хоризонталните хистограми (акмулиращи проекции) на дадено изображение. По-точно, BSC може равномерно да разбие входното изображение на няколко части – примерно на 9 подизображения (3 по вертикал и 3 по хоризонтал), по които да изчисли съответните 18 хистограми – по 2 за всяко подизображение. Така, подходът от [33] може да се разглежда като апроксимация на класическото търсене на позицията на максимума (т.е. оптималния вектор на отместването) в 2D корелацията между текущото и референтното изображение. Но, за пестене на време, вместо 2D корелация се използват две 1D корелации, по съответните хоризонтални и вертикални проекции от BSC, за които се разчита, че съхраняват необходимата информация за оптималния вектор на отместването. Допълнително, за прецизност, вместо двете 1D корелации се използват съответните две SAD функции – хоризонтално и вертикално. Ние заимстваме този BSC/SAD подход, който позволява ефективно оценяване на оптималния вектор на отместването, без да изисква специални ускорители, например FFT, като при експериментите, емулираме BSC и модифицираме SAD.

Ако камерата е реализирана на CMOS технология (вместо CCD), то към разглежданото потрепване на видео последователността се наслагва и “*rolling shutter*” ефектът.

При хоризонтално транслиране на камерата ефектът се проявява като накланяне на изображенията по посока на трансляцията, или обратно според конструкцията на “rolling shutter”. При вертикално транслиране ефектът е вертикално мащабиране, т.е. съкращаване на вертикалния размер, когато посоката на трансляцията и на CMOS сканирането съвпадат, или разпъване, когато посоките са противоположни. Методът, описан в [35], претендира, че е първият, който третира съвместно класическата 2D стабилизация и стабилизацията спрямо “rolling shutter”.

Разработката [35] на TI, по-късно патентована като [36], се основава на равномерното разбиване на изображенията и допуска използване на различни подходи за оценка на векторите на движението на отделните подизображения, включително и BSC/SAD подхода. Същественото в метода е предложеното матрично уравнение за оценка на параметрите $(T_x, T_y, \alpha, s_x, s_y, u_x)$ на моделиращата трансформация:

$$M \cdot [T_x, T_y, \alpha, s_x, s_y, u_x]^T = [t_x, t_y]^T, \quad (2)$$

където $M = \begin{bmatrix} 1 & 0 & -y & x & 0 & y \\ 0 & 1 & x & 0 & y & 0 \end{bmatrix}$ е специфичната за метода разрежена матрица, x и y са координатите на центъра на текущото подизображение, $\hat{x}$ и $\hat{y}$ – на съответния център на референтно под-изображение, $t_x = \hat{x} - x$ и $t_y = \hat{y} - y$ са съответните измерени вектори на движението, а $[T_x, T_y, \alpha, s_x, s_y, u_x]^T$ е вектор-ред с търсените параметри на трансформацията – две трансляции, T_x и T_y ъгъл на ротация $\alpha \approx \sin(\alpha)$, две мащабирания $(1 + s_x, 1 + s_y) \approx (1, 1)$ и коефициент u_x на хоризонтално накланяне “skew”, всичките центрирани (в центъра на съответното референтно подизображение). Изображенията първоначално се разбиват на 9 области по схемата “3×3”, за които се търсят векторите на отместване чрез йерархична оценка на движението [37], използвайки отново мярката за подобие SAD. След това се прави разбиване на всяка от 9-те области на “5×5” части, за да се получат по-точни вектори на движението, чрез елиминиране на векторите, които се различават от околните, отново чрез SAD.

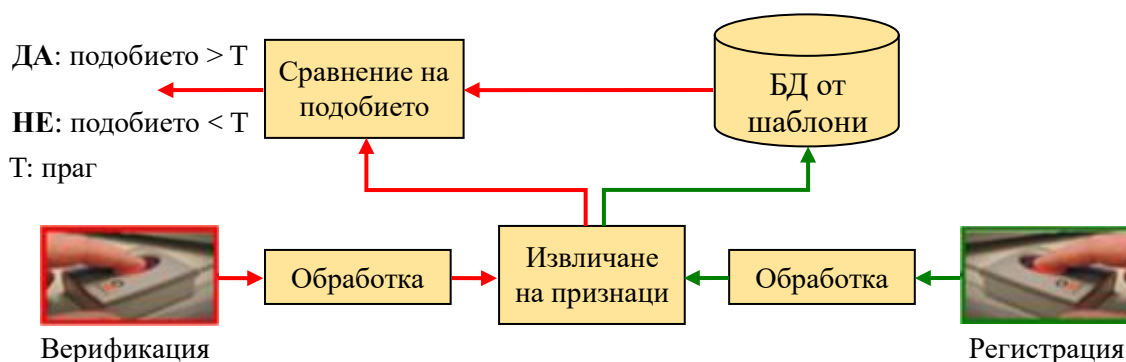
В настоящото изследване (виж Глава 2, §2.1) ние предлагаме нов метод за 2D видео стабилизация, който подобно [33, 34], е предназначен за работа в реално време върху портативни устройства и е адаптиран към BSC чипа на TI и характерния SAD подход. За разлика от базовия модел в [36], който ще наричаме тук „апроксимиращо матричен“, ние предлагаме прецизен векторен модел на измененията между два последователни кадъра. Нашият модел е нагледен и за разлика от [35] и [36] не апроксимира търсената геометрична трансформация, а освен това лесно може да се модифицира за различни комбинации на елементарните трансформации – трансляция, ротация мащабиране, накланяне и други. Ние, също, съществено прилагаме метода на най-малките квадрати (LSM). Но за разлика от [35] и [36] не игнорираме излишъка от информация, а го използваме, евентуално чрез следващи LSM стъпки, за една още по-правдоподобна оценка на търсените параметри на възстановяващата трансформация.

1.2 Биометрика

Биометриката е науката за *автентикация* (което е синоним на биометрично разпознаване, означаващо идентификация или верификация [38]) на индивида въз основа на неговите физически характеристики (като лице¹⁴, ирис, пръстов отпечатък, геометрия на ръката, уши, глас и др.) или поведенчески особености (като походка, ръкописен стил, подпис, динамика на писането чрез клавиатура и др.) [39].

1.2.1 Традиционни методи за автентикация

Една типична биометрична система се състои от подходящ сензор за регистрация на биометрична информация, извличане на признаково множество вектори от тези данни и сравняването им с информацията, съхранена в базата от данни за установяване на самоличността [40]. Една биометрична система може да работи в два различни режима: в режим на идентификация се определя идентичността на индивида, докато в режим на верификация системата потвърждава или отхвърля предявяването на определена самоличност (Фиг. 8). *Идентификацията* („Кой е той/тя?“) предполага проверка на входния биометричен шаблон (признаков вектор, модел) с всички записи от биометрични модели в БД, т.е. сравнение от вида „едно-към-много“. *Верификацията* („Аз ли съм този, за когото се представям?“) изисква сравнение на входния шаблон само с един биометричен запис в БД, т.е. сравнение от вида „едно-към-едно“ [41].



Фиг. 8. Обща схема на верификация на една биометрична система. Заимствано от Jain¹⁵.

В съвременното информационно общество определянето на самоличността е от особено значение. Традиционно, идентичността се установява чрез паспорти, лични карти, баджове, ключове, електронни пароли, персонални идентификационни номера (ПИН) и др. Тези методи са базирани на секретна информация, съхранена в лични принадлежности, електронни средства или ментално запомнена. Но притежаваните предмети могат да бъдат изгубени или откраднати, а знанието (пароли) да бъде забравено или отгътнато [42].

¹⁴ Под „лице“ ще разбираме „физиономия“, а не индивид, който ще използваме взаимнозаменяемо тук.

¹⁵ http://biometrics.cse.msu.edu/Presentations/AnilJain_21stCCC_2014.pdf

Появата на биометриката цели да адресира съпътстващите проблеми на традиционните методи за сигурност. Нейната популярност се дължи на по-голямата сигурност [43], която предлага, в сравнение с паролно-базираните системи за сигурност, както и на необходимостта от прозрачна и ненаатрапваща се биометрична идентификация в реално време, като добавка към съществуващите системи за видео наблюдение [44].

1.2.2 Класификация на биометричните технологии

Следвайки много популярна таксономия, биометричните решения могат да се класифицират като активни или пасивни. Активните системи изискват съдействие от страна на потребителя, който е съзнателен за действията си по идентификация или верификация и няма да работят, ако той откаже участие (напр. пръстови отпечатъци, ирис, геометрия и вени на ръката, отпечатък от дланта, ретина глас, подпис). От друга страна, при пасивните системи не е нужно индивидът да взема активно участие в целия процес, дори може и да не знае за неговата скрита идентификация (лица, уши, анализ на походката) [42].

Друга възможна класификация на биометричните технологии може да бъде направена в зависимост от *размерността на входния сигнал*: 1D, 2D или 3D [45]. Повечето от техниките за биометрично разпознаване са базирани на 1D сигнали – глас [46] и on-line подпис [47], или 2D изображения – off-line подпис [47], пръстови отпечатъци [48, 49], лица [50, 51], ирис [52, 53], отпечатък на дланта [54], геометрия на ръката [55], уши [56, 57, 58]. Въпреки че 2D подходите за биометрично разпознаване могат да постигнат висока точност, двумерното представяне на отличителните особености лесно може да бъде фалшифицирано, понеже голяма част от информацията, включително тримерната им структура, е загубена. Поради тези причини се наблюдава засилен интерес към изследването на нови методи за разпознаване по 3D биометрични модалности, като 3D лица [59, 60], 3D пръстови отпечатъци [61], 3D отпечатък на дланта [62], 3D уши [63, 64, 65, 66, 67, 68]. Тримерни данни могат да предложат по-богата информация за извличането на признаци и обикновено могат да се съчетаят с 2D биометричните подходи с цел повишаване на точността на разпознаващите системи.

1.2.3 Исторически преглед на развитието на биометричните системи

Първите систематични научни изследвания за идентификация на личността са положени през 1870 г., когато Alphonse Bertillon [69] въвежда използването на антропометрични измервания по тялото, физически описания и фотографии за идентификация на рецидивисти. Този метод не е имал много дълъг живот, защото много скоро след въвеждането му се установява уникалността на пръстовите отпечатъци. В началото на 20-ти век *пръстовите отпечатъци* са общоприет метод в криминалните разследвания за идентификация на заподозрени и изявени престъпници [70]. През 1936 г. офталмологът Franck Burch предлага използването на концепцията за шарките на *ириса* като метод за разпознаване на индивиди. През 60-те години на XX век се появява първата полуавтоматична система за разпознаване по *физиономията* (чрез ръчно указване на местоположението

на очите, ушите, носа и устата), разработена от Woodrow W. Bledsoe по договор с правителството на САЩ. През 1965 г. северно-американската авиация разработва първата система за разпознаване по *подписа*. През 1969 г. ФБР дава тласък за автоматизирането на разпознаването по пръстови отпечатъци. През 1974 г. е разработена първата комерсиална система за разпознаване по *геометрията на ръката*. През 1975 г. ФБР финансира разработката за прототип на скенер и *четец на пръстови отпечатъци*. През 1976 г. Texas Instruments разработват прототип на система за *разпознаване на диктори*. През 1991 г. Turk и Pentland [71] предлагат концепцията за собствените лица, водеща до възможността за *автоматичното разпознаване на лица в реално време*. През 1994г. първата автоматизирана система за разпознаване на *отпечатъци от дланта и пръстите* (RECOdermTM) е разработена от унгарската компания RECOWARE Ltd. През 1995 г. се появява първата *комерсиална система за разпознаване по ириса*. През 1998 г. ФБР създава база данни с ДНК информация за целите на съдебната експертиза. През 2000г. е описана технологията за разпознаване по вените на задната част на ръката, реализирана в първата комерсиална система за *разпознаване по кръвоносната система на човека* [72].

Първото поколение биометрични системи [70] са разработени след 2000 г. и включват разнообразие от системи за разпознаване по лица, пръстови отпечатъци и ирис, които са намерили широко приложение в множество цивилни и комерсиални системи. Такива са: системата *US-VISIT* [73] в помощ на имиграционните власти, базирана на фотографии и 10 пръстови отпечатъка на неамериканските граждани, влизащи в САЩ; системата *Privium* [74] на летището в Schiphol Amsterdam, базирана на разпознаване по ириса; системата *SmartGate* [75] на летището в Сидни, базирана на лицево разпознаване. Друг пример е системата за разпознаване по пръстови отпечатъци в увеселителните паркове на Уолт Дисни [76] с цел предотвратяване на фалшификацията на билети. Основните недостатъци в системите от първо поколение са в техните не достатъчно високи производителност (бързодействие и точност), сигурност и защита на личните данни.

1.2.4 Изисквания към биометричните модалности и системи

Според Jain et al. [77], за да бъдат използвани в биометрията, човешките физиологични или поведенчески характеристики трябва да бъдат *универсални, уникални, постоянни и лесни за регистрация (измерване)*. Това означава, че всеки човек трябва да притежава такава отличителна особеност, налична при болшинството индивиди (универсална характеристика), но различна при различните индивиди (уникална характеристика), и която е желателно да не се променя с времето (постоянна характеристика), и която да подлежи на сензорно измерване и запис (изчислима характеристика). Разбира се, биометричните системи е необходимо да отговарят и на други изисквания. Например, за *производителност* (бързодействие и точност) на системата, за *приемане от потребителите* и приложимост в ежедневието, за *степен на устойчивост срещу фалшифициране* (доколко лесно е да се заблуди системата), за възможност за *вграждане* в целевите приложения за сигурност.

Една *практическа биометрична система* трябва да осигури необходимата точност на разпознаване, бързодействие, да бъде безвредна, общоприемлива и интуитивна при използване от хората, да бъде съвместима със съответните биометрични технологии (сензори / алгоритми / производители), да бъде достатъчно устойчива на различни атаки, да осигурява защита на личните данни, да е безконфликтна към етични и здравни съображения, и културни различия. Въз основа на тези съображения *бъдещите биометрични системи* трябва да бъдат лесни за използване и вграждане, да имат ниска цена, да са надеждни и достатъчно прецизни в тяхната степен на разпознаване.

1.2.5 Биометрични системи на бъдещето

Биометричното разпознаване на бъдещето ще използва не само физически или поведенчески характеристики на личността, но и контекстуалната информация от предоставяните услуги в интернет пространството чрез мрежовата връзка между множество електронни устройства (Фиг. 9). Също, от анонимни данни за местоположението чрез проследяване на координатите на мобилния телефон всеки може да разбере „кой си ти“. Също, шаблонът на движението, нанесен върху географска карта, създава уникален отпечатък за всеки човек (Фиг. 10).



Фиг. 9. Идентификация по „цифровия отпечатък“ на индивида в интернет пространството. Заимствано от Jain¹⁶.



Фиг. 10. Идентификация чрез шаблона от почасови ежедневни местоположения, базиран на следење на GPS координатите на мобилния телефон на индивида. Заимствано от Jain¹⁶.

¹⁶http://biometrics.cse.msu.edu/Presentations/AnilJain_50YearsBiometricsResearch_SolvedUnsolvedUnexplored_ICB13.pdf

1.2.6 Моно- и мултибиометрични системи

Повечето съвременни биометрични системи типично използват единствена биометрична характеристика за установяване на идентичността на индивида, т.е. те са едномодални биометрични системи. Но множеството ограничения, които притежават (Табл. 1), водят до несигурност в системата и намалена степен на разпознаваемост [78]. За ефективното преодоляване на недостатъците на едномодалните системи особен интерес представляват мултибиометричните системи, които са в състояние да се справят с наличните проблеми (Табл. 1). Въпреки това, използването на множество биометрични модалности увеличава времето за регистриране в БД и последващата автентикация, което води до неудобства за потребителя и нарастване на цената на цялата система. За увеличаване на точността на разпознаване, една алтернатива на мултибиометриката е вграждането на така наречените „софт-биометрични” характеристики (пол, етнос, височина, цвят на очите и др.) в основната моно-биометрична система [40].

Таблица 1. Сравнителен анализ на моно- и мултибиометричните системи.

Ограничения на едномодалните биометрични системи	Предимства на мултибиометричните системи
1. Наличие на шум в биометричните данни, представени на системата, породен от недобри условия на „заснемане”, както и вариации в използваната биометрична характеристика; 2. Неуниверсалност, породена от липсата или увреждането на съответната биометрична характеристика в подмножество от потребители; 3. Наличие на горна граница на точността на разпознаване, която не може да бъде повишена чрез подобряване на методите за извличане и/или класифициране на признаковите вектори; 4. Податливост на фалшифициране на биометричните характеристики (off-line подписи, запис на глас, силиконови пръстови отпечатъци) с цел заблуждаване на системата.	1. Ефективно се справят с проблема на зашумените данни и неуниверсалността чрез използване на алтернативни биометрични модалности; 2. Устойчивост на атаки – изключително трудно фалшифициране на множество биометрични характеристики на легитимния потребител; 3. Толерантност към грешки – системата продължава да функционира дори когато даден сензор не функционира правилно или когато потребителят се бави; 4. Предлагагть многомащабна автентикация за определяне на окончателната идентичност чрез филтриране на няколко нива от потенциални кандидати; 5. Непрекъснато наблюдаване или слеене на индивиди чрез комбиниране на няколко модалности (напр. лицево разпознаване с походка).

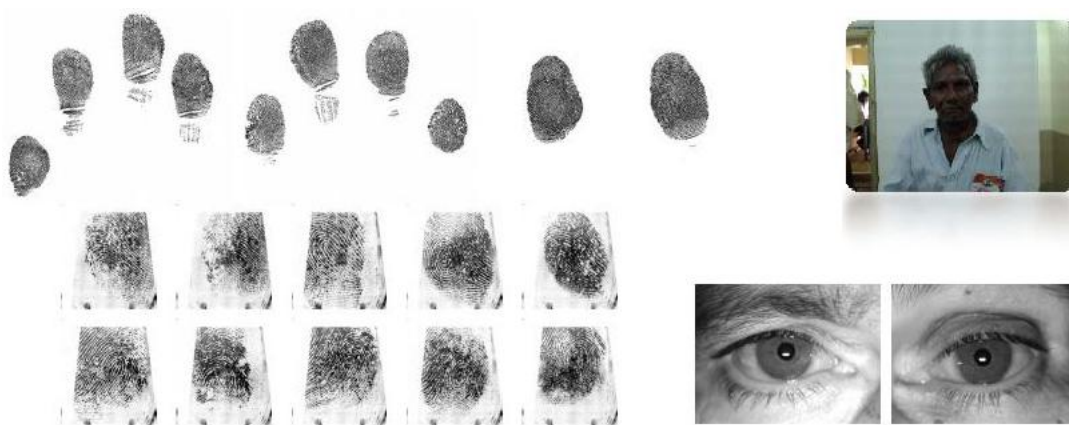
1.2.7 Сравнителен анализ на биометричните модалности

Повече от 99% от биометричните системи, налични на пазара, са базирани на пръстови отпечатащи, лицево или ирисово разпознаване (Фиг. 11) [79], което определя и тяхната голяма популярност. Поради високата точност на тези биометрични модалности и възможността за работа в реално време, системите са в състояние да работят освен в режим на верификация (1:1 сравнения) и в режим на идентификация (1:N сравнения).



Фиг. 11. Най-популярните биометрични модалности за автентикация на индивиди: а) пръстови отпечатащи; б) физиономии; в) ирис. *Заимствано от Jain*¹⁷.

Пръстовите отпечатащи и разпознаването по лица едва ли някога ще бъдат заместени от други биометрични модалности поради наличието на огромни правителствени бази данни във всяка държава. Най-голямата биометрична БД в света, състояща се от снимка на лицето, 10 пръстови отпечатаща и снимка на двата ириса (Фиг. 12), на около 966 милиона души до януари 2016 г., се разработва по индийския проект Aadhaar¹⁸,



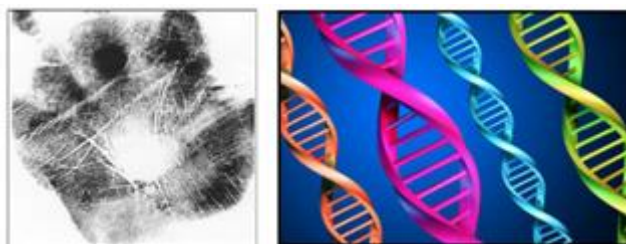
Фиг. 12. Биометрични данни за най-голямата биометрична БД в света: а) 10 пръстови отпечатаща; б) изображение на физиономията; в) изображения на двата ириса. *Заимствано от Jain*¹⁹.

Други високо надеждни биометрични модалности, подходящи за работа в режим на идентификация (1:N сравнения), са отпечатащът на дланта и ДНК информацията (виж Фиг. 13).

¹⁷ http://biometrics.cse.msu.edu/Presentations/AnilJain_Biometrics_TechnologyforHumanRecognitionNCVPRIPG13.pdf

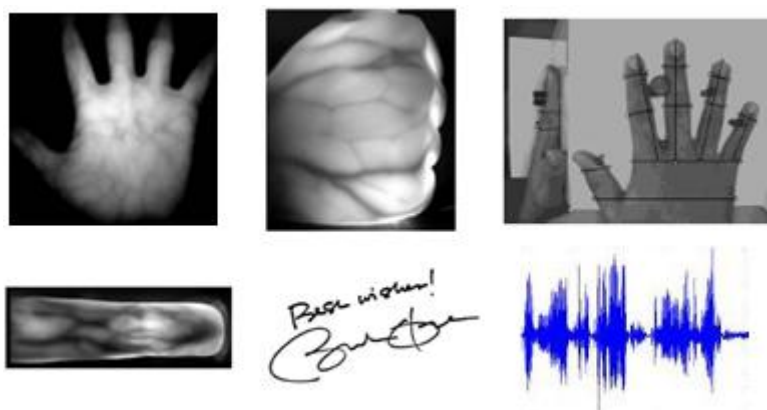
¹⁸ <https://portal.uidai.gov.in>

¹⁹ http://biometrics.cse.msu.edu/Presentations/AnilJain_21stCCC_2014.pdf



Фиг. 13. Отпечатък на дланта и ДНК информация. Заимствано от Jain²⁰.

Биометрични модалности, използвани в комерсиалните системи, но работещи основно в режим на верификация (1:1 сравнения), са: вени на дланта, ръката и пръстите; геометрия на ръката; подпис; говор (Фиг. 14).



Фиг. 14. Биометрични модалности, използвани в комерсиалните системи за работа в режим на верификация. Заимствано от Jain²⁰.

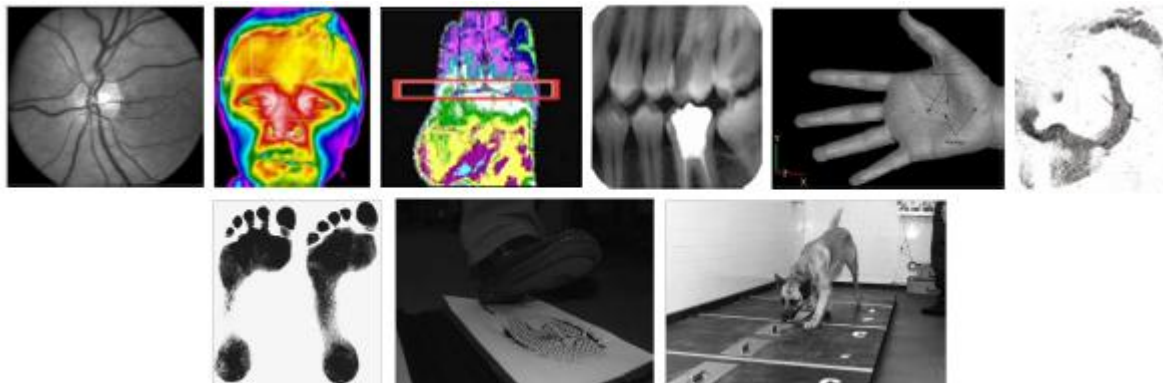
А биометрични модалности като разпознаване по походката [80], динамика на писане на клавиатура [81], кръвоносни съдове на очни дъна, кардиограма [82], околоочният контур [83], **ушите**, белези и татуировки, стави на пръстите [84], биоелектричната активност на мозъка, са все още на ниво лабораторни експерименти (Фиг. 15).



Фиг. 15. Биометрични модалности в лабораторен стадии на разработка. Заимствано от Jain²⁰.

²⁰ http://biometrics.cse.msu.edu/Presentations/AnilJain_Biometrics_TechnologyforHumanRecognitionNCVPRIPG13.pdf

За пълнота ще споменем и други биометрични модалности (виж Фиг. 16) като ретина, термограми на лицето и ръката, зъби, 3D отпечатък на дланта, отпечатък от ухото, отпечатъци от ходилата [85], отпечатък от подметката, миризма, движение на устните, стъпки, форма на носа [86], нокти на пръстите [87], предни синуси [88], генериран звук от вътрешното ухо [89].



Фиг. 16. Други биометрични модалности, с възможности за използване в криминалистиката.
Заимствано от Енциклопедията по Биометрика²¹.

В **Приложение 2** е представен сравнителен анализ на основните биометричните модалности, използвани в практиката, както и на такива в лабораторен стадий на изследване. Посочени са техни предимства и недостатъци, както и конкретни приложения²¹.

1.2.8 Съвременни приложения на биометричните системи

Бързото развитие на Биометриката като наука и практика се дължи както на нейните предимства (повишена сигурност и удобство за потребителя) спрямо традиционните методи за идентификация, така и на големия брой приложения, които тя намира в съвременното информационно общество. Тези приложения могат да се обобщят в три главни групи [41]:

1. *Комерсиални приложения:* като достъп до компютърни мрежи, интернет достъп, електронно банкиране и търговия, използване на дебитни или кредитни карти, телефонни транзакции, физически контрол на достъпа, достъп до личния компютър, сигурност на електронните данни, управление на медицински записи, електронно обучение и др.
2. *Държавно-административни приложения:* като лични карти, шофьорски книжки, обществена сигурност, международен граничен контрол и др.
3. *Приложения в криминалистиката и съдебната експертиза:* като идентификация на трупове, криминални разследвания, изчезнали хора, определяне на роднински връзки и др.

²¹ Около 90% от използваните данни са от Енциклопедията по Биометрика, (Encyclopedia of Biometrics, 2nd Edition, (Eds.) S. Z. Li and A. K. Jain. Springer, 2015).

Тези три обобщени групи от приложения могат да се разложат на по-детайлна таксономия от 12 категории за обхващане на пълния спектър на биометричните приложения (Табл. 2) [90]:

Таблица 2. Приложения на биометричните модалности в съвременното общество.

№	Приложение	Цел / Предназначение	Модалност
1	Криминалистика	идентифициране на предполагаеми престъпници	ДНК, пръстови отпечатъци
2	Проверка на съдебното минало	условие за наемане на работа	пръстови отпечатъци
3	Видео наблюдение	идентифициране на хора в дадена област от интерес	физиономии (лица), походка
4	Граничен контрол	контрол над сигурността при влизане в дадена държава	пръстови отпечатъци, ирис, лица
5	Чужда самоличност	предотвратяване на фалшификация на самоличността	пръстови отпечатъци, ирис
6	Доверен пътник	позволява на често пътуващите да избегнат по-интензивна проверка	пръстови отпечатъци, ирис
7	Физически контрол на достъпа	автентикация на самоличността с цел достъп до защитени обекти	пръстови отпечатъци, ирис, лица, геометрия на ръката, вени
8	Времево и присъствено управление на персонала	потвърждаване за присъствието на индивида в определен времеви интервал	пръстови отпечатъци, геометрия на ръката
9	Търговско разпознаване на потребителя	осъществяване на финансови транзакции и др.	пръстови отпечатъци, вени, глас
10	Отдалечена автентикация	отдалечен достъп до важна информация	глас, мобилни пръстови отпечатъци
11	Защита на цифрови документи	защита на цифрова информация от неоторизирани потребители	пръстови отпечатъци
12	Логически контрол на достъпа	позволява достъп до системи или устройства, съдържащи защитена информация	пръстови отпечатъци

1.2.9 Разпознаване на индивиди по формата на ушите

Въпреки, че научните изследвания за използването на човешкото ухо за целите на разпознаването в биометриката се намират в начален стадий, в последно време се появиха значителни дисертационни резултати [66, 68], книги [67, 91] и обзорни статии в областта [56, 57, 58].

В последните години, измежду различните физиологични биометрични модалности се обръща особено внимание и на ушите, които се смятат за надеждна отличителна характеристика за биометрично разпознаване на личността [91]. От големия брой изследователски статии по темата от последните десетилетия, може да се заключи, че точността, постигната чрез системите за разпознаване по уши, е съизмерима с тази на разпознаването по лица [57]. Съществува и случай, когато разпознаването по ушите играе съществена роля при идентификацията на заподозрян за серия от грабежи на газ станции в Холандия [93].

Известно е, че формата на ухото еволюира по време на ембрионалния растеж от шест развиващи се базови възела: i) външна спирала, ii) анти-спирала, iii) лоб, iv) трагус, v) анти-трагус и накрая vi) ушна мида, виж и Фиг. 17. Това обуславя и неговата анатомична структура, която не е напълно произволна и следователно се разглежда като универсална и уникална, и освен това не се променя значително с възрастта [94]. Ушите са лесни за (отдалечено) заснемане по ненаатрапчив начин. Горните особености водят до извода, че ушите удовлетворяват изискванията, формулирани от Jain [77], за да бъдат използвани в биометричната идентификация и верификация на личността.

За първи път идентификация на личности по морфологията на ухото се споменава през 1890 във Франция от Bertillon [95]. През 1893 Conan Doyle е публикувал статия, в която той описва особеностите на ушите на известни хора и привежда доводи, че външното ухо, точно както и лицето, отразява особеностите на характера на човека [96]. По-късно, през 1906г., чешкият лекар Imhofer забелязва, че е в състояние да направи разлика измежду 500 уши въз основа само на 4 характеристики [58]. По-нови изследвания в тази област са направени от Iannarelli през 1989 в САЩ [97]. В неговата работа ръчно е анализирано подобие то измежду 10000 уши на различни хора. Резултатът от неговия експеримент подсказва, че уникалността на човешките уши може да осигури чрез малък брой измервания (характеристики), комбинирани с информацията за раса и пол. Друго следствие от неговото изследване е, че промените във външния вид на ухото са ограничени във времето – от 8 г. до към 70 г. формата на ухото остава постоянна.

Едно по-съвременно изследване, проведено през 2005 г. в университета на Southampton, UK, също установява, че индексът на отличителност “*decidability index*” на ушите е с един порядък по-голям отколкото за физиономии, но по-малък отколкото при ирисовата биометрика [98].



Фиг. 17. Шестте базови възела от морфологията на ухото.

През юни-септември 2013 г. в централната гара на Darmstadt (Германия) е проведено изследване за оценка на вероятността за видимост на ушите на хората през деня при различни атмосферни условия [56]. Резултатите показват, че в 46% от случаите ушите са напълно видими (от 5431 наблюдения), като вероятността за закриване силно зависи от пола. При 26.03% от жените, ушите са напълно видими, докато при мъжете този процент е 69.68%.

1.2.10 Предимства и недостатъци на биометриката по уши

Въпреки разглеждането на ухото като новост в съвременната биометрика, то има своите *предимства* пред най-често използваните биометрични модалности като пръстови отпечатъци, ирис и лице [58].

- За разлика от лицето, ушите имат относително постоянна структура, която не се променя много с възрастта, те са почти неподвижни спрямо главата на човека, като не се влияят от лицеви изражения (говор, емоции) или лицеви атрибути (грим, очила, брада, мустаци). Тяхното „второ предназначение“ е единствено да поддържат обеци, рамки на очила, слухови апарати;
- За разлика от пръстовите отпечатъци и ириса, събирането на данни за ушите е по-удобно. Ухото лесно може да се заснеме от разстояние, което не изисква непременно съдействие от страна на потребителя и се счита от повечето хора за ненаатрапчиво за разлика от заснемането на ириса; Също така, безконтактното заснемане не води до проблеми с хигиената за разлика от вземането на пръстови отпечатъци;
- Също така, ушите имат по-информативна геометрична структура (свързана с тяхната анатомия), отколкото текстурата, което напоследък предизвиква интереса на изследователите в областта на 3D разпознаването на уши. 3D представянето на ушите, освен че е инвариантно по ориентация и осветление, то обогатява входните данни с информация за дълбочина, което води до увеличаване на точността на системите за разпознаване по уши.
- Изображенията на уши са „по-дискретни“ от тези на лица (физиономии), защото човешкият мозък (за разлика от изкуствения интелект) е свикнал да асоциира личността с изображения на физиономията а не толкова с тези на ухото.

Основният недостатък на използването на ушите за разпознаване се дължи на:

- Увеличена чувствителност към припокривания (от коса, шапки и др.) поради относително по-малкия размер на ухото, съчетано с ниската разделителна способност на изображенията на уши, заснети от голямо разстояние, както и вариацията в ориентацията и осветлението водят до намаляване на производителността на 2D методите за разпознаване (и детекция/откриване);
- Най-добрите резултати от разпознаването по уши са получени върху БД, съдържащи изображения на точен профил на лице при липса на закривания, които са заснети от сравнително близко разстояние, т.е. нужно е съдействие на потребителя [58].

1.2.11 Методи за разпознаване по уши

Идентификацията на личността по ушите се състои в два главни етапа: (i) детекция и (ii) разпознаване. Детекцията се занимава с откриването на уши в изображенията (и в по-тесен смисъл – със сегментирането на ушите от профила/изображението на лицето), а разпознаването – с идентификацията на индивиди на базата на сегментираните изображения на уши. Повечето от известните методи за разпознаване по уши работят върху ръчно сегментирани изображения, но съществуват и подходи за автоматична детекция на ушите от изображенията на физиономии в профил [67]. Основните предизвикателства в разпознаването по уши произлизат от: (i) недостатъчен контраст и осветление (при 2D методите) в изображенията; (ii) наличие на шум (при 2D/3D методите); (iii) неточна регистрация на входното изображение с БД от шаблони (при 2D/3D методите).

Двумерни (2D) подходи за разпознаване на уши

Когато се разглеждат приложения за работа в неконтролирана реална среда, трябва да се отчете влиянието на множество фактори върху точността на системите за разпознаване, а именно: (1) поради по-малкия им размер, закриването на ушите от други обекти е много по-сериозен проблем, отколкото при лица; (2) поради богатата 3D структура на ушите, количеството, посоката и цветността на осветлението съществено допринасят за нееднородното интензитетно представяне на ушите като 2D изображения; (3) отново поради анатомичната позиция на ушите, ротацията на главата може да затрудни разпознаването им; (4) разпознаването може да се повлияе значително от резолюцията на камерата, особено ако заснемането е от по-голямо разстояние.

От научната литература в областта е известно, че най-висока разпознаваемост се постига при използването на 3D изображения (самостоятелно или в комбинация с 2D такива), което позволява разпознаване при вариация в осветлението и ориентацията. Но в много приложения за отдалечено (без съдействие от страна на потребителя) разпознаване е необходима работа с кадри от обзорни видео клипове, което означава, че ушите най-често трябва да бъдат разпознавани от източници на 2D информация [58].

Методите за 2D разпознаване на уши могат да се групират в следните области, в зависимост от използваните техники: 1) базирани на изгледа “*appearance*” на ушите в изображенията; 2) базирани на трансформации на „силови“ полета [99]; 3) базирани на геометрични особености; 4) базирани на невронни мрежи; и 5) мултимодални методи.

1. Методи, базирани на изгледа на ушите

Техниките, основани на изгледа на ушите, използват по глобален начин (цялото изображение) или локално (части от него) изображенията на уши за разпознаване. В тази категория попадат методът на главните компоненти (PCA) [100], методът на независимите компоненти (ICA) [101], методи за обединяване на интензитетно и цветово пространство [102].

Chang et al. [100] извършват сравнение на ефективността на PCA метода, приложен към разпознаването на лица и уши, и установяват, че ушите са точно толкова подходящи

както и лицата в приложението им за биометрично разпознаване. Те докладват за разпознаемост от 71.6% при уши и 70.5% при лица (физиономии).

Zhang et al. [101] използват метода на ICA с мрежа от радиални базисни функции (RBF) за разпознаване на уши и постигат по-добра производителност от PCA метода, когато броят на признаците е значителен. Иначе, при налични по-малко признаци, PCA превъзхожда ICA подхода. Основният недостатък на PCA или ICA е, че те работят добре само когато входните изображения са заснети при контролирани условия и затова изискват много точна регистрация за постигането на добри резултати.

Един метод, базиран на обединяване на цветови пространства, е предложен от Nanni and Lumini [102]. Те избират няколко цветови пространства, след което от тях се извличат локални признаци чрез банки от Габор филтри. Подобие на признаковите вектори се извършва чрез комбинирането на изхода от множество класификатори по най-близкия съсед, построени върху различни цветови компоненти. Докладваната разпознаемост е почти 84 %.

2. Методи, базирани трансформации в силови полета

Техниките, базирани на силови полета [98, 99], преобразуват изображенията на уши в силови полета. Там всеки пиксел от изображението се разглежда като атрактор, имитирайки закона на Нютон за гравитацията, като привличането между пикселите е пропорционално на техния интензитет и обратно пропорционално на квадрата на разстоянието между тях. По този начин подобие на ушите се заменя със сравняване на силови полета. Чрез този метод е постигната разпознаемост от 99.2% [98] върху XM2VTS БД, състояща се от 252 профилни изображения. При зле регистрирани изображения, методът на силовите полета постига много по-висока точност отколкото PCA, където разпознаемостта е само 62.4% върху същата експериментална постановка. Но когато изображенията на уши са правилно сегментирани и нормализирани по мащаб и ротация, PCA достига до 98.4% точност на разпознаването.

Също е установено, че разпознаването, базирано на силови полета, е устойчиво на шум, като при добавяне на 18 dB Гаусов шум точността достига 99.6% [103].

3. Методи, базирани на геометрични особености

Mu et al. [104] използват геометричната информация на ушите, като признаковите вектори се съставят според външната форма и вътрешната структура на ухото. Те постигат 85% разпознаемост.

Методът, предложен в [105], използва SIFT характерни точки за точното регистриране на ушите чрез перспективна трансформация. Известно е, че SIFT трансформацията е устойчива на малки вариации в ориентацията и яркостта при извличането на характерните точки от изображенията. Резултатите са съизмерими с PCA метода при ръчна регистрация на изображенията. Но при използването на по-предизвикателна БД (съдържаща болшинство от реални ситуации) методът тук показва устойчивост към сложен фон, наличие на 20% припокриване и до $\pm 13^\circ$ вариация в ориентацията.

В метода на Arbab-Zavar et al [106] също се извличат SIFT характерни точки, които се филтрират с цел намаляване на техния брой при обучението на референтния модел,

защото плътността, следователно и излишъкът от SIFT точки е значителен в силно структурирани/текстурни области от изображението. Целта на това филтриране е постигане на максимална близост на признаковите вектори от тестовото и референтното ухо, които са от един и същи клас. Те постигат разпознаваемост от 91.5% върху XM2VTS БД.

В [107] е предложена техника, базирана на фрактали за класификацията на човешки уши. Изображенията на ушите се разделят на еднакъв брой по-едри области, където самоподобие се определя чрез афинна трансформация на областите в изображението. Ако дадена област е закрита, то останалите обикновено съдържат достатъчно различни особености, което прави подхода устойчив на малки промени в ориентацията / осветеността, както и на частични припокривания.

Един скорошен подход на Prakash and Gupta [108] разглежда обединяването на SURF характерни точки от различни изображения. Те предлагат използването на няколко входни изображения на един и същи човек за запис в БД с цел прилагане на множество подобряващи техники за противодействие на ефекта от нееднаквото осветление, лош контраст и шум. От друга страна, използването на SURF точки помага за правилната регистрация на изображения с налична вариация в ориентацията. Постигната разпознаваемост от 96.75% върху базата данни UND collection E, и съответно 97.35% и 98.79% за подмножествата I & II на ИТ Kanpur БД, показва, че този метод предлага значително подобряване в точността в сравнение с един от най-известните подходи в [102].

4. Методи, базирани на невронни мрежи

В [109] методът на Moreno et al. за разпознаване по уши изследва ефективността на класификацията чрез невронни мрежи. Техният признаков вектор се състои от местоположението на седем точки от ухото, които отговарят на характерните точки от работата на Iannarelli. Допълнително, те съставят и морфологичен признаков вектор, който описва ухото като цяло. Тези два признакови вектора се използват като вход за различни невронни мрежи. Най-добрият постигнат резултат е 93% чрез компресираща мрежа върху БД, съставена от ръчно изрязани 168 изображения на уши от 28 участници.

В [110] авторите съставят 6 различни признакови вектора чрез използването на 7 моментови инварианта, които са една статистическа мярка за описание на формата. Те показват, че всеки от 7-те момента е инвариантен на промени в ротацията и мащаба. Признаковите вектори се подават на невронна мрежа с обратно разпространение на грешката, която е обучена за тяхната класификация. Използвайки собствена БД от 60 изображения на уши, авторите докладват за 91.8% постигната точност.

5. Други методи

Съществува и значително разнообразие от други методи за разпознаване по изображения на уши. Те са базирани на Габор [111], Фурие [112], Хаф [113], лъчева "ray" [114], или уейвлет [115] - трансформациите; локални бинарни шаблони [116], активен модел на формата [117].

В [118] е направено сравнение между различните методи за извличане на признаковите вектори на ушите, което обхваща ефективността на Фурие дескрипторите, Габор

трансформацията, моментовите инварианти и статистически признаци. Заключение е, че най-висока разпознаваемост на ушите се постига при използването на моментовите инварианти и Габор трансформацията.

Струва си да се спомене и оригиналното предложение в [119], където е разработена система за „ушна биометрика“, базирана на акустичните свойства на ухото. Методът е основан на определянето на акустичната трансферна функция на ухото чрез стимулирането му със звукова вълна и изчисление на отразения сигнал.

6. Мултимодални подходи

Разработени са и множество мултимодални методи, разглеждащи обединението на уши с други биометрични модалности като физиономии, говор и др.

Iwano et al. [120] комбинира изображения на уши с говор от 38 участници от мъжки пол. В резултат е постигнато подобрене в производителността на системата в сравнение със самостоятелното използване само на уши, или само на говор.

Rahman et al. [121] предлага мултимодална система, използваща PCA метода върху лица и уши едновременно. Това изследване (върху 90 уши и 90 изображения на лица от 18 индивиди) докладва за подобряване в разпознаването (94.4%) в сравнение със самостоятелното използване на тези модалности.

Iwano et al. [122] предлага мултиалгоритмичен подход за разпознаване по уши, където признаковите вектори са извлечени чрез PCA и ICA методите. Това изследване също демонстрира подобряване в точността в сравнение със самостоятелно извлечените признаци по един от методите.

Тримерни (3D) подходи за разпознаване на уши

Биометричните подходи за разпознаване на уши разчитат на морфологичните свойства на ухото. Разпознаването на образи обикновено изисква решаването на два проблема: описание на обектите и правило за различаването (дискриминирането) им. Разпознаването на уши също следва тази най-обща схема: описание и извличане на признаците, както и стратегията за сравнението им.

Цветовото разпределение на ухото е почти монотонно и обособяването му като обект в тримерния свят се дължи главно на сенчестите му нюанси, като по този начин морфологията е тази, която характеризира външния му вид и отличителни черти. Поради тази причина вариациите в позицията на камерата, както и в осветлението са критични при 2D разпознаването на уши. 3D представянето на ушите, освен че е инвариантно по ориентация и осветление, то обогатява входните данни с информация за дълбочина, което води до увеличаване на точността на системите за разпознаване по уши.

Измежду най-значимите разработки в областта на тримерното разпознаване на уши са тези на Chen и Bhanu [64, 123] и Yan и Bowyer [65, 124].

Hui Chen и Bir Bhanu представят в [123] една от първите системи за детекция и разпознаване на 3D уши, която използва дълбочинната (2.5D) структура на ухото, заснета чрез дълбочинен “range” скенер. Те докладват за постигната точност на разпознаването

от 96.4% върху собствена БД, UCR [64], състояща се от 902 3D дълбочинни изображения на уши на 155 участници.

Yan и Bowyer [65] също разработват метод за автоматична сегментация и разпознаване по уши, който обединява 2D цветни изображения и изображенията от дълбочинен скенер. Ушната мида служи като отправна точка за прилагане на модела на активните контури и използване на ICP метода за намиране на подобие между тримерните обекти. Те извършват експерименти върху собствена БД, UND J2 [125], която е била най-голямата БД за уши по онова време. Постигнатата точност на разпознаването (TRR) е 97.8% за 1386 3D дълбочинни изображения на уши на 415 индивиди. Към днешна дата, UND базата данни се е превърнала в стандарт за тестване на разпознаващи алгоритми по ушите.

Zhou et. al [126] докладват, че техният подход е по-бърз и по-точен от ICP базирани алгоритми, предложени от Chen и Bhanu, Yan и Bowyer, с постигнатата разпознаваемост (TRR) от 98.6% за 830 изображения на уши на 415 индивиди.

Prakash и Gupta [127] разширяват по-ранен техен 2D подход към 3D такъв. Използвайки 3D представянето на същото множество от изображения, както в 2D анализа, те съобщават за повишаване на разпознаването (TRR) от 96.63% на 99.38%. Интересното е, че този 3D базиран подход изглежда устойчив на влиянието от случайни ротация и/или мащабиране.

Cadavid и Abdel-Mottaleb [128] са първите, които изследват точността на разпознаването на 3D уши, базирано на видео последователност. Те възстановяват 3D формата на ухото чрез метода „форма получена чрез сенките” (SfS). След това използват ICP алгоритъма за изчисляване на подобие между 3D моделите на ушите, възстановени от два видеоклипа.

В литературата са споменати и множество многомодални подходи за разпознаване, използващи ушната биометрика. Те варират от мултисензорни подходи, използващи съвместно 2D изображения и 3D модели на уши [129], до такива, които комбинират ушите с биометрични модалности като отпечатък на дланта [130], пръстови отпечатъци [131], подпис [132]. От голям интерес са модалностите, които разглеждат комбинацията между уши и лица, които варират от лице в анфас + уши; лице в профил + уши; до лице, уши + трета модалност [57]. Всички многомодални изследвания съобщават за увеличаване на степента на разпознаване, когато се използват многомодални вместо индивидуални биометрични данни [57, 58]. В [133] са представени резултати от едномодалното разпознаване по уши и физиономии, както и от тяхното комбиниране, като точността на разпознаване е следната: 2D разпознаване по лица – 81.7%; 3D разпознаване по уши – 95%; комбинация от двете – 100%.

1.2.12 Публични БД от 2D изображения и 3D модели на уши

В тази част ще бъдат разгледани публично известните до момента БД от 2D изображения и 3D модели на човешки уши. В *Приложение 3* са представени техните характерни особености по отношение на брой участници, общ брой изображения в БД, постигнати максимални резултати за разпознаване и/или сегментация, кратко описание на съдър-

жанието и предназначение им [56, 57, 58, 134]. В **Приложение 4** са показани примерни изображения от разгледаните БД от Приложение 3.

Обобщения:

- Най-популярната БД от уши, смятана за образец на достоверност на резултатите от разпознаването, е **UND-J2** БД [125], следвана от **IITK** БД [108]. В настоящия момент **UND-J2** е най-голямата БД от 3D дълбочинни изображения – 1800 на брой [135].
- Единствените БД, съдържащи реални 3D полигонални модели на уши, са **OpenHear** [136] и **SYMARE** [137], въпреки че те не са специално предназначени за биометричната общност по 3D разпознаване на уши.
- Най-добрите резултати върху двете най-представителни БД, а именно **UND-J2** и **IITK** по отношение на детекцията и разпознаването на уши, са съответно:

за 2D подходите:

- разпознаване на уши (98.79%) чрез метода на [66] върху **IITK Subset II** БД;
- детекцията на уши (96.63%) чрез метода на [138] върху **UND-J2** БД;

за 3D подходите:

- разпознаване на уши (98.30%) чрез метода на [66] върху **UND-J2** БД;
- детекцията на уши (99.38%) чрез метода на [139] върху **UND-J2** БД;
- БД, предназначени само за детекция/сегментация на уши, са: **UND-J2** БД (без изрязване или нормализация) [125], **UMIST** [140], **CAS-PEAL** [141].
- Най-реалистичната БД, съдържаща изображения на уши при вариация в осветеността, закривания от коса и обеци, е **UBEAR** [142]. Тези особености я правят особено подходяща за тестване на съвременни робастни методи за разпознаване по изображения уши, претендиращи за работа в реална работна среда.
- Неспецифични по отношение на ушната биометрика БД са: **ScFace** (предназначена за разпознаване на физиономии в неконтролирана обстановка на закрито, използвайки обзорни видео камери) [143]; **XM2VTS** (съдържа множество от изображения, звукови файлове, видео клипове, 3D модели на физиономии) [144]; **FERET** (голяма БД от изображения на физиономии – 14126 общо, заснети в полуконтролирана обстановка) [145, 146]; **OpenHear** (съдържа само 20 3D модели на главата с ушите на субектите и е предназначена за акустична симулация в помощ на дизайн на слухови устройства). **SYMARE** (включва 61 3D модела с висока резолюция на горната част от торса, главата и ушите и е предназначена за подобряване на индивидуализацията на 3D звука за бъдещи персонални аудио устройства).

Глава 2

Методи за бърза и качествена 2D видеостабилизация

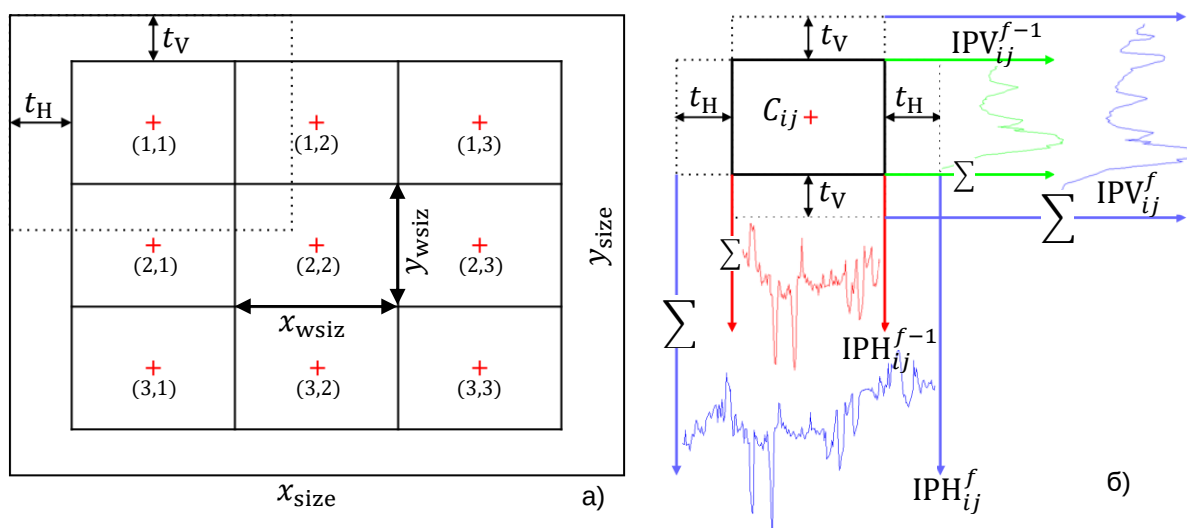
Обзорът от Глава 1, §1.1 може да се разглежда като въведение към тази глава.

2.1 Видеостабилизация по гладка траектория

Видеостабилизацията по гладка траектория произвежда видео, при което се цели изглаждане на високочестотната нестабилност на камерата със запазване на оригинално замислената от оператора траектория на движение.

Най-общо, методите за 2D видеостабилизация се основават на оценяване на параметрите на оптимална линейна трансформация на текущия кадър от видео последователността спрямо някакъв референтен кадър (да кажем предходния). Фактичестката стабилизация се реализира чрез обратната трансформация (на изчислената), с което текущият кадър се преобразува (транслира, завърта, мащабира и т.н.) до оптималното съвпадение с референтния кадър. Оценката на реалното движение на камерата, т.е. последователността от трансформации на текущия кадър към текущата референтна позиция, се основава на оценка на така наречените „вектори на движението”, които обикновено се дефинират интегрално – за някакви съответстващи области в кадрите и/или локално – за някакви съответстващи характерни точки в тях.

2.1.1 Определяне на векторите на движение



Фиг. 18. Базова схема за определяне на векторите на движение: а) 3×3 разбиване на всеки кадър от видеоклипа; б) хоризонтални и вертикални акумулиращи проекции.

Както бе споменато в увода (Глава 1, §1.1.4), тук ще се спрем на интегралното определяне на векторите на движение²², подобно на [33, 35, 36]. Целта е съвместимостта с BSC чипа на TI, откъдето следва и предпочетената класическа схема за разделяне на кадъра на “3×3” области. При този метод всеки кадър се разделя на “3×3” области [33, 34], което определя броя 9 на търсените вектори на движението $\vec{t}_{ij}$ ($i, j = 1, 2, 3$) между съответните 9 центъра на под-изображенията в текущия и предходния кадър, Фиг. 18а.

Ние също използваме идеята за оценка на 9-те вектора на движението чрез SAD подхода [33], описан накратко в Глава 1, §1.1.4, а именно:

- Изчисляват се *хоризонтални* (по колони) и *вертикални* (по редове) акумулиращи проекции (сумирания) IPH_{ij}^f и IPV_{ij}^f за всяка от 9-те подобласти с център C_{ij} за текущия кадър (f) и предходния кадър ($f - 1$), виж Фиг. 18б. Тук (x_{size}, y_{size}) са оригиналните размери на кадъра, а (x_{wsiz}, y_{wsiz}) са размерите на работната област C_{ij} от текущия кадър: $x_{wsiz} = (x_{size} - 2t_H)/3$, $0 \leq t_H < x_{size}/8$ и $y_{wsiz} = (y_{size} - 2t_V)/3$, $0 \leq t_V < y_{size}/8$, където t_H и t_V приемаме за допустими граници за максималните x, y компоненти на търсените вектори на движението $\vec{t}_{ij}$. Тези граници биха могли да се определят предварително на базата на предварителни експерименти, при които се търси корелация само на една централна област между два последователни кадъра. В резултат на статистически анализ, t_H и t_V се определят, така че да покриват максималните налични отмествания във видеото, доколкото е възможно.
- Съответните вектори на движението $\vec{t}_{ij} = (t_{xij}^{(f)}, t_{yij}^{(f)})$ се получават като минимум по SAD подхода:

$$t_{xij}^{(f)} = \underset{(\tau)}{\operatorname{argmin}} \{SADH_{ij}^f(\tau)\}; \quad SADH_{ij}^f(\tau) = \sum_{x=-x_{wsiz}/2}^{x_{wsiz}/2} |IPH_{ij}^f(x + \tau) - IPH_{ij}^{f-1}(x)|, \quad -t_H < \tau < t_H; \quad (3a)$$

$$t_{yij}^{(f)} = \underset{(\tau)}{\operatorname{argmin}} \{SADV_{ij}^f(\tau)\}; \quad SADV_{ij}^f(\tau) = \sum_{y=-y_{wsiz}/2}^{y_{wsiz}/2} |IPV_{ij}^f(y + \tau) - IPV_{ij}^{f-1}(y)|, \quad -t_V < \tau < t_V. \quad (3b)$$

- **Нашето подобрение** на SAD подхода се състои в това, че вместо класическите проекции IPH_{ij}^f и IPV_{ij}^f , ние провеждаме горното изчисление върху нормализираните по средна стойност проекции $\bar{IPH}_{ij}^f$ и $\bar{IPV}_{ij}^f$, при това – средната стойност е „плаваща“, както следва:

- по хоризонтали:

$$\bar{IPH}_{ij}^f(x, \tau) = IPH_{ij}^f(x) - \bar{IPH}_{ij}^f(\tau), \quad \bar{IPH}_{ij}^f(\tau) = \frac{1}{x_{wsiz}} \sum_{p=-x_{wsiz}/2}^{x_{wsiz}/2} IPH_{ij}^f(\tau + p); \quad (4a)$$

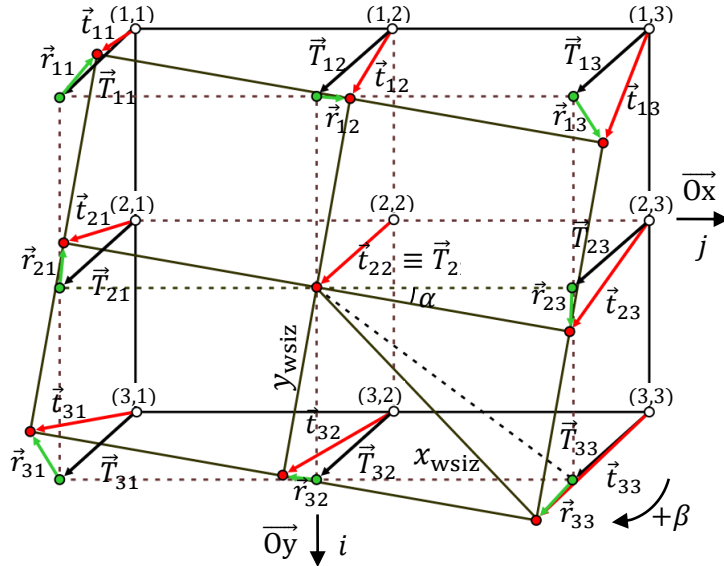
- по вертикали:

$$\bar{IPV}_{ij}^f(y, \tau) = IPV_{ij}^f(y) - \bar{IPV}_{ij}^f(\tau), \quad \bar{IPV}_{ij}^f(\tau) = \frac{1}{y_{wsiz}} \sum_{p=-y_{wsiz}/2}^{y_{wsiz}/2} IPV_{ij}^f(\tau + p). \quad (4b)$$

²² Въпреки че във видеофайловете с прег компресия се съдържа карта на векторите на движение, които може да се използват директно, формулировката на задачата от фирмата заявител изискваше изчисляването на тези вектори чрез емулиране на BSC чипа на TI. Виж и раздел §2.1.6 от тази глава.

2.1.2 Базов векторен модел с разбиване на “3×3” области

В предлагания базов векторен модел, всеки оценен по SAD вектор на движение $\vec{t}_{ij}$ между два кадъра се разлага на сума от два вектора: вектор на трансляция $\vec{T}_{ij}$ и вектор на ротация $\vec{r}_{ij}$, т.е. $\vec{t}_{ij} = \vec{T}_{ij} + \vec{r}_{ij}$ (виж Фиг. 19).



Фиг. 19. Базов векторен модел, отчитащ трансляция и ротация между два последователни кадъра във видеото.

По този начин формираме 9 векторни уравнения, по едно за всяка двойка центрове (C_{ij}^f, C_{ij}^{f-1}) , $(i = 1,2,3)$, $(j = 1,2,3)$, в дадени два (не обезателно последователни) кадъра $(f-1)$ и (f) , $f = 1,2,3, \dots$ от клипа. Неизвестните параметри са: векторът на трансляция $\vec{T} = (T_x, T_y)$, който е един и същ ($\vec{T}_{ij} = \vec{T}$) за всяка двойка центрове, и референтният вектор на ротация $\vec{r} = (r_x, r_y)$, избран като $\vec{r} = \vec{r}_{23} = (r_{x23}, r_{y23})$, чрез който се изразяват останалите вектори $\vec{r}_{ij}$ на ротация (на един и същи ъгъл α). Или казано по друг начин – съставяме система от 18 покомпонентни уравнения (виж Табл. 3), в която имаме 4 неизвестни параметъра (T_x, T_y, r_x, r_y) .

Таблица 3. Система от 18 покомпонентни уравнения за центровете (i, j) .

(i, j)	$\vec{Ox}$	$\vec{Oy}$
(1,1)	$T_x - r_x + kr_y = t_{x11}$	$T_y - kr_x - r_y = t_{y11}$
(1,2)	$T_x + kr_y = t_{x12}$	$T_y - kr_x = t_{y12}$
(1,3)	$T_x + r_x + kr_y = t_{x13}$	$T_y - kr_x + r_y = t_{y13}$
(2,1)	$T_x - r_x = t_{x21}$	$T_y - r_y = t_{y21}$
(2,2)	$T_x = t_{x22}$	$T_y = t_{y22}$
(2,3)	$T_x + r_x = t_{x23}$	$T_y + r_y = t_{y23}$
(3,1)	$T_x - r_x - kr_y = t_{x31}$	$T_y + kr_x - r_y = t_{y31}$
(3,2)	$T_x - kr_y = t_{x32}$	$T_y + kr_x = t_{y32}$
(3,3)	$T_x + r_x - kr_y = t_{x33}$	$T_y + kr_x + r_y = t_{y33}$

Горната преопределена линейна система можем да решим по класическия диференциален подход за оптимизация по квадратичен критерий (минимум на средно квадратичната грешка). При линейни системи, каквито са разглежданите тук случаи, този подход е еквивалентен с този по метода на най-малките квадрати (LSM), известен от линейната алгебра, както и от статистическата оптимизация при допускането за нормално разпределение на грешките. За нагледност и бързина при извеждане на прякото решение на получаваната система от линейни уравнения, тук използваме идеята от линейната алгебра за представянето ѝ като матрично уравнение $AX = B$ (Фиг. 20), чието решение X по метода на най-малките квадрати е следното:

$$(A^T A)X = A^T B \Rightarrow X = (A^T A)^{-1} A^T B, \quad (5)$$

(виж и §2.2.2 Метод на Оптичния поток).

$$\begin{array}{c} \overrightarrow{Ox} \\ \overrightarrow{Oy} \end{array}
 \begin{bmatrix} A_1 & A_2 & A_3 & A_4 \end{bmatrix} \cdot X = B$$

$$\begin{bmatrix} 1 & 0 & -1 & k \\ 1 & 0 & 0 & k \\ 1 & 0 & 1 & k \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & -k \\ 1 & 0 & 0 & -k \\ 1 & 0 & 1 & -k \\ 0 & 1 & -k & -1 \\ 0 & 1 & -k & 0 \\ 0 & 1 & -k & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & k & -1 \\ 0 & 1 & k & 0 \\ 0 & 1 & k & 1 \end{bmatrix} \cdot \begin{bmatrix} T_x \\ T_y \\ r_x \\ r_y \end{bmatrix} = \begin{bmatrix} t_{x11} \\ t_{x12} \\ t_{x13} \\ t_{x21} \\ t_{x22} \\ t_{x23} \\ t_{x31} \\ t_{x32} \\ t_{x33} \\ t_{y11} \\ t_{y12} \\ t_{y13} \\ t_{y21} \\ t_{y22} \\ t_{y23} \\ t_{y31} \\ t_{y32} \\ t_{y33} \end{bmatrix}$$

Коефициенти пред неизвестните параметри
(T_x , T_y , r_x , r_y) на системата от Табл. 3.

A_1	$T_{xx} (T_x \text{ по } \overrightarrow{Ox})$			$T_{xy} (T_x \text{ по } \overrightarrow{Oy})$		
i,j	1	2	3	1	2	3
1	1	1	1	0	0	0
2	1	1	1	0	0	0
3	1	1	1	0	0	0

A_2	$T_{yx} (T_y \text{ по } \overrightarrow{Ox})$			$T_{yy} (T_y \text{ по } \overrightarrow{Oy})$		
i,j	1	2	3	1	2	3
1	0	0	0	1	1	1
2	0	0	0	1	1	1
3	0	0	0	1	1	1

A_3	$R_{xx} (R_x \text{ по } \overrightarrow{Ox})$			$R_{xy} (R_x \text{ по } \overrightarrow{Oy})$		
i,j	1	2	3	1	2	3
1	-1	0	1	-k	-k	-k
2	-1	0	1	0	0	0
3	-1	0	1	k	k	k

A_4	$R_{yx} (R_y \text{ по } \overrightarrow{Ox})$			$R_{yy} (R_y \text{ по } \overrightarrow{Oy})$		
i,j	1	2	3	1	2	3
1	k	k	k	-1	0	1
2	0	0	0	-1	0	1
3	-k	-k	-k	-1	0	1

$$A = \{A_{ij}\}_{18 \times 4} = [A_1 \ A_2 \ A_3 \ A_4] = \left[\begin{pmatrix} T_{xx} \\ T_{xy} \end{pmatrix}, \begin{pmatrix} T_{yx} \\ T_{yy} \end{pmatrix}, \begin{pmatrix} R_{xx} \\ R_{xy} \end{pmatrix}, \begin{pmatrix} R_{yx} \\ R_{yy} \end{pmatrix} \right]$$

Фиг. 20. Запис на системата уравнения от Табл. 3 в матричен вид $AX = B$.

За решаване на матрично уравнение $AX = B$ от Фиг. 20, първо го представяме във вида $(A^T A)X = A^T B$:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \cdot \begin{bmatrix} T_x \\ T_y \\ r_x \\ r_y \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}, \quad (5a)$$

$$(A^T A)_{[4 \times 4]} \cdot X_{[4 \times 1]} = (A^T B)_{[4 \times 1]}$$

където матрицата $(A^T A)_{[4 \times 4]}$ се получава *диагонална*:

$$A^T A = \begin{bmatrix} 9 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 \\ 0 & 0 & 6 + 6k^2 & 0 \\ 0 & 0 & 0 & 6 + 6k^2 \end{bmatrix}, \quad (5b)$$

където $a_{ij} = 0$, $i \neq j$, $(i, j = 1 \div 4)$, а за коефициентите на вектора-стълб $A^T B = (a, b, c, d)^T$, след еквивалентни преобразувания, получаваме:

$$\begin{aligned} a &= \sum_{i=1}^3 \sum_{j=1}^3 t_{xij}; \quad b = \sum_{i=1}^3 \sum_{j=1}^3 t_{yij}; \\ c &= k \sum_{j=1}^3 (t_{y3j} - t_{y1j}) + \sum_{i=1}^3 (t_{xi3} - t_{xi1}); \quad d = k \sum_{j=1}^3 (t_{x1j} - t_{x3j}) + \sum_{i=1}^3 (t_{yi3} - t_{yi1}). \end{aligned} \quad (5b)$$

Така, за търсеното LSM решение $X = [T_x, T_y, r_x, r_y]^T$, финално получаваме:

$$T_x = \frac{a}{9}, \quad T_y = \frac{b}{9}, \quad r_x = \frac{c}{6 + 6k^2}, \quad r_y = \frac{d}{6 + 6k^2}, \quad (5g)$$

където $k = y_{wsiz} / x_{wsiz}$, а ъгълът на ротация α се определя чрез компонентите на референтния вектор $\vec{r}$, като $\alpha = 2\arctg(r_x/r_y)$.

За повишаване на достоверността на намереното решение можем да приложим LSM втори път, но върху по-малък брой уравнения за търсените неизвестни (T_x, T_y, r_x, r_y). За целта е достатъчно да въведем новите параметри e_{ij} , $(i = 1,2,3), (j = 1,2,3)$, както следва: $e_{ij} = 0$, ако ще елиминираме уравнението (i, j) , иначе $e_{ij} = 1$.

При експериментите по предложения метод ние предполагаме Гаусово разпределение на отклоненията ε_{xij} и ε_{yij} , $(i = 1,2,3), (j = 1,2,3)$, получени при първия LSM пас. Така, оценяваме коефициентите e_{xij} и e_{yij} , както следва:

$$e_{xij} = \begin{cases} 0, & \text{if } |\varepsilon_{xij}| > \eta \sigma_x \\ 1, & \text{if } |\varepsilon_{xij}| \leq \eta \sigma_x \end{cases}, \text{ подобно и за } e_{yij}, \text{ където } \sigma_x \text{ и } \sigma_y \text{ са средно-квадратичните}$$

отклонения: $(\sigma_x)^2 = \frac{1}{9} \sum_{i,j} (\varepsilon_{xij} - \mu_x)^2$ и $(\sigma_y)^2 = \frac{1}{9} \sum_{i,j} (\varepsilon_{yij} - \mu_y)^2$, спрямо съответните средни стойности μ_x и μ_y , а η е експертен коефициент: $1 < \eta < 3$. Т.е., идеята е да се елиминират уравненията, даващи относително големи отклонения при първия LSM пас. Така, за втория LSM имаме следната резултантна система от 18 уравнения, виж Табл. 4.

Таблица 4. Система от 18 покомпонентни уравнения след 1-вия LSM пас.

(i, j)	$\vec{Ox}$	$\vec{Oy}$
(1,1)	$(T_x - r_x + kr_y = t_{x11})e_{x11}$	$(T_y - kr_x - r_y = t_{y11})e_{y11}$
(1,2)	$(T_x + kr_y = t_{x12})e_{x12}$	$(T_y - kr_x = t_{y12})e_{y12}$
(1,3)	$(T_x + r_x + kr_y = t_{x13})e_{x13}$	$(T_y - kr_x + r_y = t_{y13})e_{y13}$
(2,1)	$(T_x - r_x = t_{x21})e_{x21}$	$(T_y - r_y = t_{y21})e_{y21}$
(2,2)	$(T_x = t_{x22})e_{x22}$	$(T_y = t_{y22})e_{y22}$
(2,3)	$(T_x + r_x = t_{x23})e_{x23}$	$(T_y + r_y = t_{y23})e_{y23}$
(3,1)	$(T_x - r_x - kr_y = t_{x31})e_{x31}$	$(T_y + kr_x - r_y = t_{y31})e_{y31}$
(3,2)	$(T_x - kr_y = t_{x32})e_{x32}$	$(T_y + kr_x = t_{y32})e_{y32}$
(3,3)	$(T_x + r_x - kr_y = t_{x33})e_{x33}$	$(T_y + kr_x + r_y = t_{y33})e_{y33}$

По аналогичен начин, както при 1-вия LSM пас, решаваме системата (от Табл. 4) по познатия матричен начин $(A^T A)X = A^T B$, но тук $A^T A$ има следния по-общ вид:

$$A^T A = \begin{bmatrix} a_{11} & 0 & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & 0 & a_{44} \end{bmatrix}, \quad (6)$$

където:

$$a_{11} = \sum_{i=1}^3 \sum_{j=1}^3 T_{xxij} e_{xij}; \quad a_{22} = \sum_{i=1}^3 \sum_{j=1}^3 T_{yyij} e_{yij}; \quad (6a)$$

$$a_{13} = \sum_{i=1}^3 \sum_{j=1}^3 R_{xxij} e_{xij}; \quad a_{24} = \sum_{i=1}^3 \sum_{j=1}^3 R_{yyij} e_{yij}; \quad (6б)$$

$$a_{14} = \sum_{i=1}^3 \sum_{j=1}^3 R_{yxij} e_{xij}; \quad a_{23} = - \sum_{i=1}^3 \sum_{j=1}^3 R_{xyij} e_{yij}; \quad (6в)$$

$$a_{33} = \sum_{i=1}^3 \sum_{j=1}^3 ((R_{xxij} e_{xij})^2 + (R_{xyij} e_{yij})^2); \quad a_{44} = \sum_{i=1}^3 \sum_{j=1}^3 ((R_{yyij} e_{yij})^2 + (R_{yxij} e_{xij})^2); \quad (6г)$$

$$a_{12} = a_{21} = a_{34} = a_{43} = 0. \quad (6д)$$

Отбелязаните равенства за съответните коефициенти от (6а-6г) се получават поради следното (макар и необязателно) ограничение, с което печелим простота на изразите при разширението на алгоритъма (с втори LSM пас), а именно:

$$e_{xij} = e_{yij} \in \{0,1\}, \forall(i, j), \quad (6е)$$

т.е. игнорираме и двете уравнения (по Ox и по Oy) на дадена под-област с център (i, j) . Освен това, от таблиците на Фиг. 20, покоординатно представящи вектор-стълбовете

A_1, A_2, A_3 и A_4 , имаме: $T_{xxij} = T_{yyij}$, $R_{xxij} = R_{yyij}$, $R_{yxij} = -R_{xyij}$, $\forall(i,j)$. Така, като възприемем ограничението (6е), т.е. $a_{11} = a_{22}$, $a_{13} = a_{24}$, $a_{14} = a_{23}$, $a_{33} = a_{44}$, то матричното уравнение $(A^T A)X = A^T B$, приема вида:

$$\begin{bmatrix} a_{11} & 0 & a_{13} & a_{14} \\ 0 & a_{11} & -a_{14} & a_{13} \\ a_{13} & -a_{14} & a_{33} & 0 \\ a_{14} & a_{13} & 0 & a_{33} \end{bmatrix} \cdot \begin{bmatrix} T_x \\ T_y \\ r_x \\ r_y \end{bmatrix} = \begin{bmatrix} \mathcal{A} \\ \mathcal{B} \\ \mathcal{C} \\ \mathcal{D} \end{bmatrix}, \quad (7)$$

където:

$$\begin{aligned} \mathcal{A} &= \sum_{i=1}^3 \sum_{j=1}^3 t_{xij} e_{xij}; \quad \mathcal{B} = \sum_{i=1}^3 \sum_{j=1}^3 t_{yij} e_{yij}; \\ \mathcal{C} &= k \sum_{j=1}^3 (t_{y3j} e_{y3j} - t_{y1j} e_{y1j}) + \sum_{j=1}^3 (t_{xi3} e_{xi3} - t_{xi1} e_{xi1}); \\ \mathcal{D} &= k \sum_{j=1}^3 (t_{x1j} e_{x1j} - t_{x3j} e_{x3j}) + \sum_{i=1}^3 (t_{yi3} e_{yi3} - t_{yi1} e_{yi1}); \end{aligned} \quad (7a)$$

Решавайки резултантната линейна система

$$\begin{cases} a_{11}T_x + 0 & +a_{13}r_x & +a_{14}r_y & = \mathcal{A} \\ 0 & +a_{11}T_y & -a_{14}r_x & +a_{13}r_y & = \mathcal{B} \\ a_{13}T_x & -a_{14}T_y & +a_{33}r_x & +0 & = \mathcal{C} \\ a_{14}T_x & +a_{13}T_y & +0 & +a_{33}r_y & = \mathcal{D} \end{cases} \quad (7b)$$

за неизвестните параметри T_x , T_y , r_x , r_y , финално получаваме:

$$\begin{aligned} T_x &= \frac{-a_{13}\mathcal{C} - a_{14}\mathcal{D} + a_{33}\mathcal{A}}{-a_{13}^2 - a_{14}^2 + a_{11}a_{33}}; & T_y &= \frac{-a_{13}\mathcal{D} + a_{14}\mathcal{C} + a_{33}\mathcal{B}}{-a_{13}^2 - a_{14}^2 + a_{11}a_{33}}; \\ r_x &= \frac{a_{11}\mathcal{C} - a_{13}\mathcal{A} + a_{14}\mathcal{B}}{-a_{13}^2 - a_{14}^2 + a_{11}a_{33}}; & r_y &= \frac{a_{11}\mathcal{D} - a_{13}\mathcal{B} - a_{14}\mathcal{A}}{-a_{13}^2 - a_{14}^2 + a_{11}a_{33}}, \end{aligned} \quad (7b)$$

където ъгълът на ротация $\alpha = 2\arctg(r_x/r_y)$.

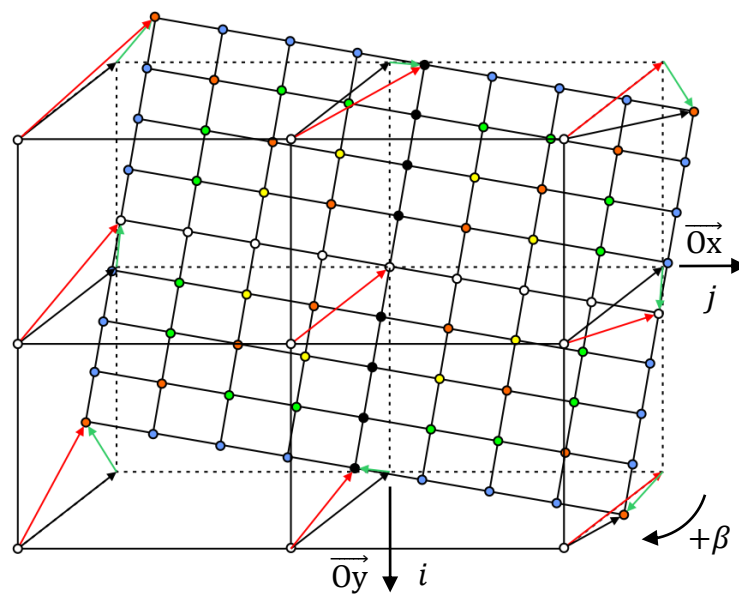
По този начин изведохме прякото решение за търсените параметри на движението (глобална трансляция (T_x, T_y) и ъгъл на ротация α), което е особено подходящо за имплементации в реално време върху портативни устройства с ограничени изчислителни мощности. Това решение зависи изцяло от прецизността на намерените вектори на движението по SAD подхода. Затова, в следващия раздел предлагаме едно разширение на метода “3×3” към разбиване на кадъра на “9×9” подобласти, разчитайки че с увеличаване на броя на търсените векторите на движение, ще постигнем едно още по-прецизно решение.

2.1.3 Мултипликация на базовия векторен модел за $3 \times 3 \rightarrow 9 \times 9$ области

Тук ще разгледаме извеждането на прякото решение за търсените параметри (транслация и ротация) на разширения “9×9” векторен модел, в резултат на което се постига значително подобрене в качеството на изходния видеоклип.

По аналогия на векторния модел с “3×3” разбиване, съставяме система от 162 по-компонентни уравнения (по $\vec{Ox}$ и $\vec{Oy}$), в която неизвестните параметри (T_x , T_y , r_x , r_y) се намират отново по матричния LSM метод (виж §2.1.2).

Поради големия брой уравнения, които трябва да се съставят, ще разгледаме тяхното поетапно систематизиране, базирано на специфично симетрично обхождане по схемата на разширения CCD векторен модел със схема на разбиване “9×9” (виж Фиг. 21). Систематизирането се изчерпва със следните разгледани по-долу случаи: C1 ÷ C5.



Фиг. 21. Мултипликация на базовия векторен модел за разбиване на кадрите на “9×9” области, отчитащ транслация и ротация между два последователни кадъра във видеото.

Основни параметри за всички схеми на обхождане:

$$k = \frac{y_{size}}{x_{size}} = \frac{d_{95}}{d_{59}} - \text{главен коефициент;}$$

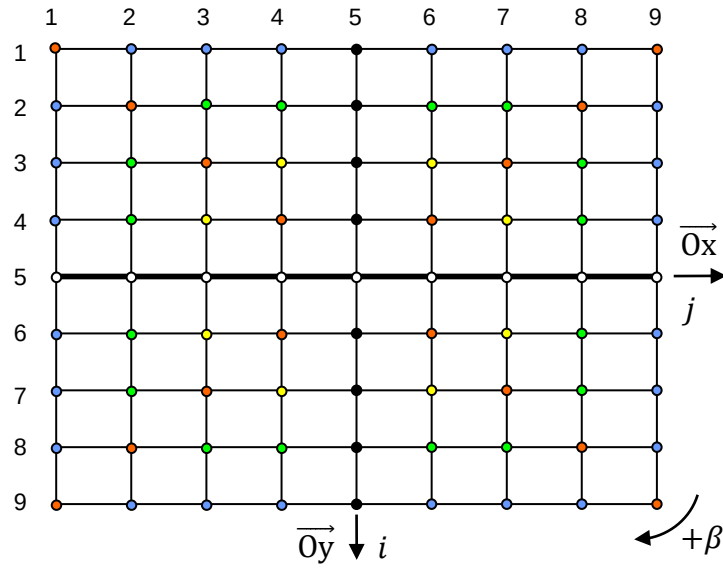
d_{ij} – разстояние от центъра на схемата (5,5) до произволна точка (i, j) , виж и Фиг. 22;

β_{ij} – ъгъл на ротация от оста $\vec{Ox}$ до точка (i, j) ;

$(r_{x59}, r_{y59}) = (r_x, r_y)$ е референтният вектор на ротация.

Допълнителни съотношения:

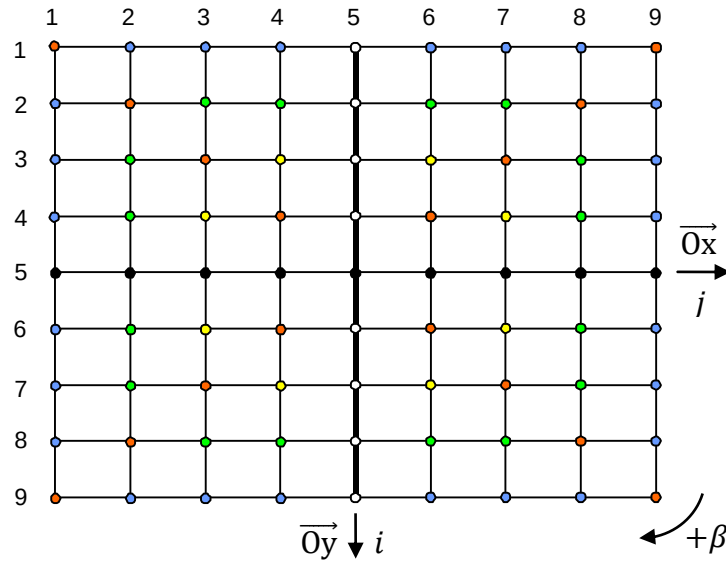
$$\frac{d_{56}}{d_{59}} = \frac{1}{4}, \quad \frac{d_{57}}{d_{59}} = \frac{1}{2}, \quad \frac{d_{58}}{d_{59}} = \frac{3}{4}, \quad \frac{d_{65}}{d_{59}} = \frac{k}{4}, \quad \frac{d_{75}}{d_{59}} = \frac{k}{2}, \quad \frac{d_{85}}{d_{59}} = \frac{3}{4}k, \quad \frac{d_{95}}{d_{59}} = k \quad (8)$$

С1) Уравнения по хоризонталната права $(5,9) \div (5,1)$ 

Фиг. 22. Схема за съставяне на уравнения чрез обхождане по хоризонталната права между точките $(5,9)$ и $(5,1)$.

Таблица 5. Система от покомпонентни уравнения за обхождане по правата $(5,9) \div (5,1)$.

център (i, j)	LSM on $\overrightarrow{Ox}$	LSM on $\overrightarrow{Oy}$
(5,5)	$T_x - t_{x55} = \varepsilon_{x55}$	$T_y - t_{y55} = \varepsilon_{y55}$
(5,9) +	$T_x + \mathbf{r}_x - t_{x59} = \varepsilon_{x59}$	$T_y + \mathbf{r}_y - t_{y59} = \varepsilon_{y59}$
(5,1) -	$T_x - \mathbf{r}_x - t_{x51} = \varepsilon_{x51}$	$T_y - \mathbf{r}_y - t_{y51} = \varepsilon_{y51}$
(5,8) +	$T_x + \frac{3}{4}\mathbf{r}_x - t_{x58} = \varepsilon_{x58}$	$T_y + \frac{3}{4}\mathbf{r}_y - t_{y58} = \varepsilon_{y58}$
(5,2) -	$T_x - \frac{3}{4}\mathbf{r}_x - t_{x52} = \varepsilon_{x52}$	$T_y - \frac{3}{4}\mathbf{r}_y - t_{y52} = \varepsilon_{y52}$
(5,7) +	$T_x + \frac{1}{2}\mathbf{r}_x - t_{x57} = \varepsilon_{x57}$	$T_y + \frac{1}{2}\mathbf{r}_y - t_{y57} = \varepsilon_{y57}$
(5,3) -	$T_x - \frac{1}{2}\mathbf{r}_x - t_{x53} = \varepsilon_{x53}$	$T_y - \frac{1}{2}\mathbf{r}_y - t_{y53} = \varepsilon_{y53}$
(5,6) +	$T_x + \frac{1}{4}\mathbf{r}_x - t_{x56} = \varepsilon_{x56}$	$T_y + \frac{1}{4}\mathbf{r}_y - t_{y56} = \varepsilon_{y56}$
(5,4) -	$T_x - \frac{1}{4}\mathbf{r}_x - t_{x54} = \varepsilon_{x54}$	$T_y - \frac{1}{4}\mathbf{r}_y - t_{y54} = \varepsilon_{y54}$

С2) Уравнения по вертикалната права (9,5)÷(1,5)**Фиг. 23.** Схема за съставяне на уравнения чрез обхождане по вертикалната права (9,5)÷(1,5).

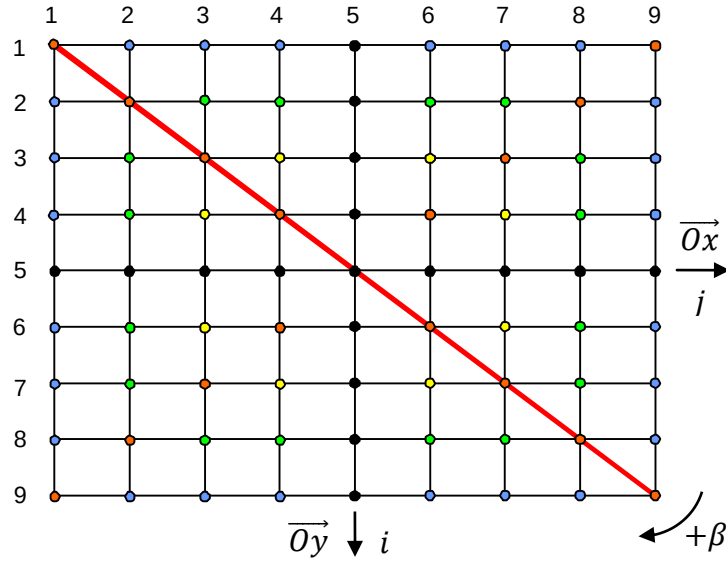
Допълнителни параметри за случая С2 (виж Фиг. 23):

$$\mathcal{R}(\beta_{95}) = \begin{bmatrix} \cos(\beta_{95}) & \sin(\beta_{95}) \\ -\sin(\beta_{95}) & \cos(\beta_{95}) \end{bmatrix} \equiv \mathcal{R}\left(\frac{\pi}{2}\right) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad (9a)$$

$$(r_{x95}, r_{y95}) = k(r_x, r_y) \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = k(-r_y, r_x) \quad (9b)$$

Таблица 6. Система от покомпонентни уравнения за обхождане по правата (9,5)÷(1,5).

център (i, j)	LSM on $\vec{Ox}$	LSM on $\vec{Oy}$
(9,5) +	$T_x - kr_y - t_{x95} = \varepsilon_{x95}$	$T_y + kr_x - t_{y95} = \varepsilon_{y95}$
(1,5) -	$T_x + kr_y - t_{x15} = \varepsilon_{x15}$	$T_y - kr_x - t_{y15} = \varepsilon_{y15}$
(8,5) +	$T_x - \frac{3}{4}kr_y - t_{x85} = \varepsilon_{x85}$	$T_y + \frac{3}{4}kr_x - t_{y85} = \varepsilon_{y85}$
(2,5) -	$T_x + \frac{3}{4}kr_y - t_{x25} = \varepsilon_{x25}$	$T_y - \frac{3}{4}kr_x - t_{y25} = \varepsilon_{y25}$
(7,5) +	$T_x - \frac{1}{2}kr_y - t_{x75} = \varepsilon_{x75}$	$T_y + \frac{1}{2}kr_x - t_{y75} = \varepsilon_{y75}$
(3,5) -	$T_x + \frac{1}{2}kr_y - t_{x35} = \varepsilon_{x35}$	$T_y - \frac{1}{2}kr_x - t_{y35} = \varepsilon_{y35}$
(6,5) +	$T_x - \frac{1}{4}kr_y - t_{x65} = \varepsilon_{x65}$	$T_y + \frac{1}{4}kr_x - t_{y65} = \varepsilon_{y65}$
(4,5) -	$T_x + \frac{1}{4}kr_y - t_{x45} = \varepsilon_{x45}$	$T_y - \frac{1}{4}kr_x - t_{y45} = \varepsilon_{y45}$

С3) Уравнения по диагонала (9,9)÷(1,1)**Фиг. 24.** Схема за съставяне на уравнения чрез обхождане по диагонала (9,9)÷(1,1).

Допълнителни параметри за случая С3 (виж Фиг. 24):

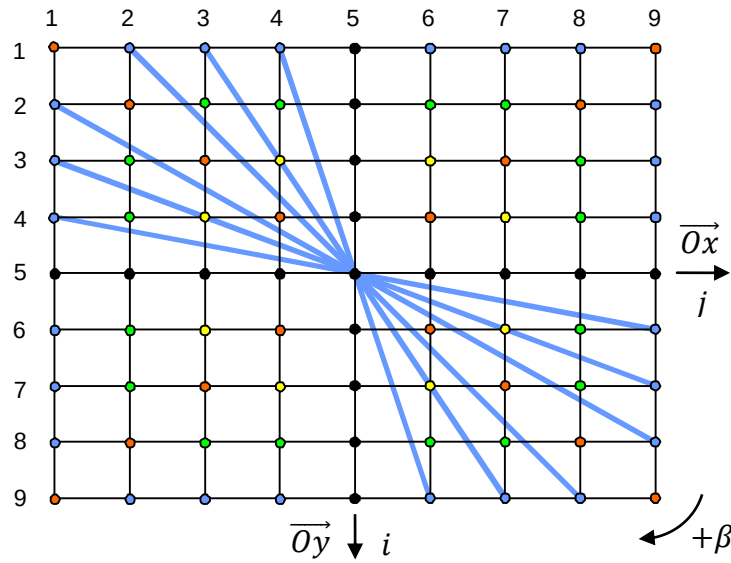
$$\mathbf{R}(\beta_{99}) = \begin{bmatrix} \cos(\beta_{99}) & \sin(\beta_{99}) \\ -\sin(\beta_{99}) & \cos(\beta_{99}) \end{bmatrix}, \quad \cos(\beta_{99}) = \frac{d_{59}}{d_{99}}, \quad \sin(\beta_{99}) = \frac{d_{95}}{d_{99}} \quad (10a)$$

$$(r_{x99}, r_{y99}) = \frac{d_{99}}{d_{59}} (r_x, r_y) \mathbf{R}(\beta_{99}) = (r_x, r_y) \begin{bmatrix} 1 & \frac{d_{95}}{d_{59}} \\ -\frac{d_{95}}{d_{59}} & 1 \end{bmatrix} = (r_x - kr_y, kr_x + r_y) \quad (10b)$$

Таблица 7. Система от покомпонентни уравнения за обхождане по диагонала (9,9)÷(1,1).

център (i, j)	LSM on $\overrightarrow{Ox}$	LSM on $\overrightarrow{Oy}$
(9,9) +	$T_x + r_x - kr_y - t_{x99} = \varepsilon_{x99}$	$T_y + kr_x + r_y - t_{y99} = \varepsilon_{y99}$
(1,1) -	$T_x - r_x + kr_y - t_{x11} = \varepsilon_{x11}$	$T_y - kr_x - r_y - t_{y11} = \varepsilon_{y11}$
(8,8) +	$T_x + \frac{3}{4}r_x - \frac{3}{4}kr_y - t_{x88} = \varepsilon_{x88}$	$T_y + \frac{3}{4}kr_x + \frac{3}{4}r_y - t_{y88} = \varepsilon_{y88}$
(2,2) -	$T_x - \frac{3}{4}r_x + \frac{3}{4}kr_y - t_{x22} = \varepsilon_{x22}$	$T_y - \frac{3}{4}kr_x - \frac{3}{4}r_y - t_{y22} = \varepsilon_{y22}$
(7,7) +	$T_x + \frac{1}{2}r_x - \frac{1}{2}kr_y - t_{x77} = \varepsilon_{x77}$	$T_y + \frac{1}{2}kr_x + \frac{1}{2}r_y - t_{y77} = \varepsilon_{y77}$
(3,3) -	$T_x - \frac{1}{2}r_x + \frac{1}{2}kr_y - t_{x33} = \varepsilon_{x33}$	$T_y - \frac{1}{2}kr_x - \frac{1}{2}r_y - t_{y33} = \varepsilon_{y33}$
(6,6) +	$T_x + \frac{1}{4}r_x - \frac{1}{4}kr_y - t_{x66} = \varepsilon_{x66}$	$T_y + \frac{1}{4}kr_x + \frac{1}{4}r_y - t_{y66} = \varepsilon_{y66}$
(4,4) -	$T_x - \frac{1}{4}r_x + \frac{1}{4}kr_y - t_{x44} = \varepsilon_{x44}$	$T_y - \frac{1}{4}kr_x - \frac{1}{4}r_y - t_{y44} = \varepsilon_{y44}$

С4) Уравнения по правите $(6,9) \div (4,1)$, $(7,9) \div (3,1)$, $(8,9) \div (2,1)$, $(9,8) \div (1,2)$, $(9,7) \div (1,3)$, $(9,6) \div (1,4)$



Фиг. 25. Схема за съставяне на уравнения чрез обхождане по „основните“ диагонални прави.

Допълнителни параметри за случая С4 (виж Фиг. 25):

$$\mathcal{R}(\beta) = \begin{bmatrix} \cos(\beta) & \sin(\beta) \\ -\sin(\beta) & \cos(\beta) \end{bmatrix} \quad (11a)$$

$$\begin{aligned} \cos(\beta_{69}) &= \frac{d_{59}}{d_{69}}, & \sin(\beta_{69}) &= \frac{d_{65}}{d_{69}}, \\ (r_{x69}, r_{y69}) &= \frac{d_{69}}{d_{59}} (r_x, r_y) \mathcal{R}(\beta_{69}) = (r_x, r_y) \begin{bmatrix} 1 & \frac{d_{65}}{d_{59}} \\ -\frac{d_{65}}{d_{59}} & 1 \end{bmatrix} = (r_x - \frac{1}{4}kr_y, \frac{1}{4}kr_x + r_y) \end{aligned} \quad (116)$$

$$\begin{aligned} \cos(\beta_{79}) &= \frac{d_{59}}{d_{79}}, & \sin(\beta_{79}) &= \frac{d_{75}}{d_{79}}, \\ (r_{x79}, r_{y79}) &= \frac{d_{79}}{d_{59}} (r_x, r_y) \mathcal{R}(\beta_{79}) = (r_x, r_y) \begin{bmatrix} 1 & \frac{d_{75}}{d_{59}} \\ -\frac{d_{75}}{d_{59}} & 1 \end{bmatrix} = (r_x - \frac{1}{2}kr_y, \frac{1}{2}kr_x + r_y) \end{aligned} \quad (11b)$$

$$\begin{aligned} \cos(\beta_{89}) &= \frac{d_{59}}{d_{89}}, & \sin(\beta_{89}) &= \frac{d_{85}}{d_{89}}, \\ (r_{x89}, r_{y89}) &= \frac{d_{89}}{d_{59}} (r_x, r_y) \mathcal{R}(\beta_{89}) = (r_x, r_y) \begin{bmatrix} 1 & \frac{d_{85}}{d_{59}} \\ -\frac{d_{85}}{d_{59}} & 1 \end{bmatrix} = (r_x - \frac{3}{4}kr_y, \frac{3}{4}kr_x + r_y) \end{aligned} \quad (11r)$$

$$\cos(\beta_{98}) = \frac{d_{58}}{d_{98}}, \quad \sin(\beta_{98}) = \frac{d_{95}}{d_{98}},$$

$$(r_{x98}, r_{y98}) = \frac{d_{98}}{d_{59}} (r_x, r_y) \mathcal{R}(\beta_{98}) = (r_x, r_y) \begin{bmatrix} \frac{d_{58}}{d_{59}} & \frac{d_{95}}{d_{59}} \\ -\frac{d_{95}}{d_{59}} & \frac{d_{58}}{d_{59}} \end{bmatrix} = \left(\frac{3}{4}r_x - kr_y, kr_x + \frac{3}{4}r_y\right) \quad (11д)$$

$$\cos(\beta_{97}) = \frac{d_{57}}{d_{97}}, \quad \sin(\beta_{97}) = \frac{d_{95}}{d_{97}},$$

$$(r_{x97}, r_{y97}) = \frac{d_{97}}{d_{59}} (r_x, r_y) \mathcal{R}(\beta_{97}) = (r_x, r_y) \begin{bmatrix} \frac{d_{57}}{d_{59}} & \frac{d_{95}}{d_{59}} \\ -\frac{d_{95}}{d_{59}} & \frac{d_{57}}{d_{59}} \end{bmatrix} = \left(\frac{1}{2}r_x - kr_y, kr_x + \frac{1}{2}r_y\right) \quad (11е)$$

$$\cos(\beta_{96}) = \frac{d_{56}}{d_{96}}, \quad \sin(\beta_{96}) = \frac{d_{95}}{d_{96}},$$

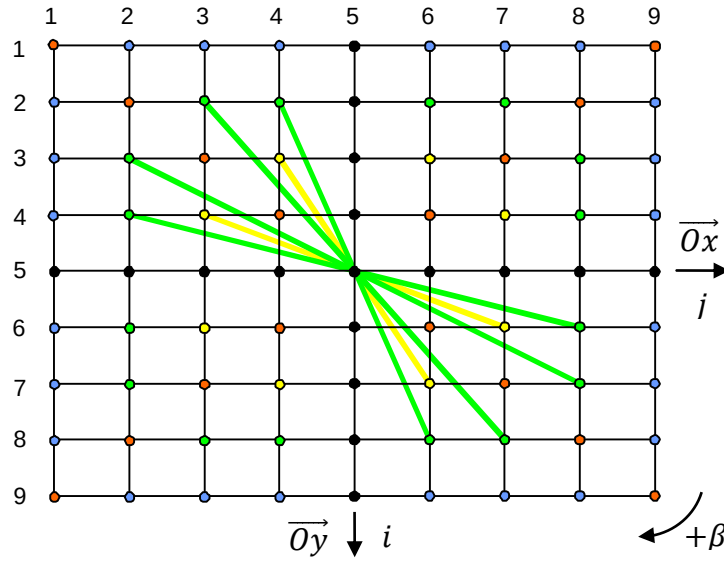
$$(r_{x96}, r_{y96}) = \frac{d_{96}}{d_{59}} (r_x, r_y) \mathcal{R}(\beta_{96}) = (r_x, r_y) \begin{bmatrix} \frac{d_{56}}{d_{59}} & \frac{d_{95}}{d_{59}} \\ -\frac{d_{95}}{d_{59}} & \frac{d_{56}}{d_{59}} \end{bmatrix} = \left(\frac{1}{4}r_x - kr_y, kr_x + \frac{1}{4}r_y\right) \quad (11ж)$$

Таблица 8. Система от покомпонентни уравнения за обхождане по „основните“ диагонални прави.

център (i, j)	LSM on $\overrightarrow{Ox}$	LSM on $\overrightarrow{Oy}$
(6,9) +	$T_x + r_x - \frac{1}{4}kr_y - t_{x69} = \varepsilon_{x69}$	$T_y + \frac{1}{4}kr_x + r_y - t_{y69} = \varepsilon_{y69}$
(4,1) -	$T_x - r_x + \frac{1}{4}kr_y - t_{x41} = \varepsilon_{x41}$	$T_y - \frac{1}{4}kr_x - r_y - t_{y41} = \varepsilon_{y41}$
(7,9) +	$T_x + r_x - \frac{1}{2}kr_y - t_{x79} = \varepsilon_{x79}$	$T_y + \frac{1}{2}kr_x + r_y - t_{y79} = \varepsilon_{y79}$
(3,1) -	$T_x - r_x + \frac{1}{2}kr_y - t_{x31} = \varepsilon_{x31}$	$T_y - \frac{1}{2}kr_x - r_y - t_{y31} = \varepsilon_{y31}$
(8,9) +	$T_x + r_x - \frac{3}{4}kr_y - t_{x89} = \varepsilon_{x89}$	$T_y + \frac{3}{4}kr_x + r_y - t_{y89} = \varepsilon_{y89}$
(2,1) -	$T_x - r_x + \frac{3}{4}kr_y - t_{x21} = \varepsilon_{x21}$	$T_y - \frac{3}{4}kr_x - r_y - t_{y21} = \varepsilon_{y21}$
(9,8) +	$T_x + \frac{3}{4}r_x - kr_y - t_{x98} = \varepsilon_{x98}$	$T_y + kr_x + \frac{3}{4}r_y - t_{y98} = \varepsilon_{y98}$
(1,2) -	$T_x - \frac{3}{4}r_x + kr_y - t_{x12} = \varepsilon_{x12}$	$T_y - kr_x - \frac{3}{4}r_y - t_{y12} = \varepsilon_{y12}$
(9,7) +	$T_x + \frac{1}{2}r_x - kr_y - t_{x97} = \varepsilon_{x97}$	$T_y + kr_x + \frac{1}{2}r_y - t_{y97} = \varepsilon_{y97}$
(1,3) -	$T_x - \frac{1}{2}r_x + kr_y - t_{x13} = \varepsilon_{x13}$	$T_y - kr_x - \frac{1}{2}r_y - t_{y13} = \varepsilon_{y13}$
(9,6) +	$T_x + \frac{1}{4}r_x - kr_y - t_{x96} = \varepsilon_{x96}$	$T_y + kr_x + \frac{1}{4}r_y - t_{y96} = \varepsilon_{y96}$
(1,4) -	$T_x - \frac{1}{4}r_x + kr_y - t_{x14} = \varepsilon_{x14}$	$T_y - kr_x - \frac{1}{4}r_y - t_{y14} = \varepsilon_{y14}$

С5) Уравнения по правите

(6,8)÷(4,2), (7,8)÷(3,2), (8,7)÷(2,3), (8,6)÷(2,4), (6,7)÷(4,3), (7,6)÷(3,4)

**Фиг. 26.** Схема за съставяне на уравнения чрез обхождане по „вътрешните“ диагонални прави.*Допълнителни параметри за случая С5 (виж Фиг. 26):*

$$\mathcal{R}(\beta) = \begin{bmatrix} \cos(\beta) & \sin(\beta) \\ -\sin(\beta) & \cos(\beta) \end{bmatrix} \quad (12a)$$

$$\cos(\beta_{68}) = \frac{d_{58}}{d_{68}}, \quad \sin(\beta_{68}) = \frac{d_{65}}{d_{68}},$$

$$(r_{x68}, r_{y68}) = \frac{d_{68}}{d_{59}} (r_x, r_y) \mathcal{R}(\beta_{68}) = (r_x, r_y) \begin{bmatrix} \frac{d_{58}}{d_{59}} & \frac{d_{65}}{d_{59}} \\ -\frac{d_{65}}{d_{59}} & \frac{d_{58}}{d_{59}} \end{bmatrix} = \left(\frac{3}{4}r_x - \frac{1}{4}kr_y, \frac{1}{4}kr_x + \frac{3}{4}r_y \right) \quad (126)$$

$$\cos(\beta_{78}) = \frac{d_{58}}{d_{78}}, \quad \sin(\beta_{78}) = \frac{d_{75}}{d_{78}},$$

$$(r_{x78}, r_{y78}) = \frac{d_{78}}{d_{59}} (r_x, r_y) \mathcal{R}(\beta_{78}) = (r_x, r_y) \begin{bmatrix} \frac{d_{58}}{d_{59}} & \frac{d_{75}}{d_{59}} \\ -\frac{d_{75}}{d_{59}} & \frac{d_{58}}{d_{59}} \end{bmatrix} = \left(\frac{3}{4}r_x - \frac{1}{2}kr_y, \frac{1}{2}kr_x + \frac{3}{4}r_y \right) \quad (12b)$$

$$\cos(\beta_{87}) = \frac{d_{57}}{d_{87}}, \quad \sin(\beta_{87}) = \frac{d_{85}}{d_{87}},$$

$$(r_{x87}, r_{y87}) = \frac{d_{87}}{d_{59}} (r_x, r_y) \mathcal{R}(\beta_{87}) = (r_x, r_y) \begin{bmatrix} \frac{d_{57}}{d_{59}} & \frac{d_{85}}{d_{59}} \\ -\frac{d_{85}}{d_{59}} & \frac{d_{57}}{d_{59}} \end{bmatrix} = \left(\frac{1}{2}r_x - \frac{3}{4}kr_y, \frac{3}{4}kr_x + \frac{1}{2}r_y \right) \quad (12r)$$

$$\cos(\beta_{86}) = \frac{d_{56}}{d_{86}}, \quad \sin(\beta_{86}) = \frac{d_{85}}{d_{86}},$$

$$(r_{x86}, r_{y86}) = \frac{d_{86}}{d_{59}} (r_x, r_y) \mathcal{R}(\beta_{86}) = (r_x, r_y) \begin{bmatrix} \frac{d_{56}}{d_{59}} & \frac{d_{85}}{d_{59}} \\ -\frac{d_{85}}{d_{59}} & \frac{d_{56}}{d_{59}} \end{bmatrix} = \left(\frac{1}{4}r_x - \frac{3}{4}kr_y, \frac{3}{4}kr_x + \frac{1}{4}r_y \right) \quad (12д)$$

$$\cos(\beta_{67}) = \frac{d_{57}}{d_{67}}, \quad \sin(\beta_{67}) = \frac{d_{65}}{d_{67}},$$

$$(r_{x67}, r_{y67}) = \frac{d_{67}}{d_{59}} (r_x, r_y) \mathcal{R}(\beta_{67}) = (r_x, r_y) \begin{bmatrix} \frac{d_{57}}{d_{59}} & \frac{d_{65}}{d_{59}} \\ -\frac{d_{65}}{d_{59}} & \frac{d_{57}}{d_{59}} \end{bmatrix} = \left(\frac{1}{2}r_x - \frac{1}{4}kr_y, \frac{1}{4}kr_x + \frac{1}{2}r_y \right) \quad (12е)$$

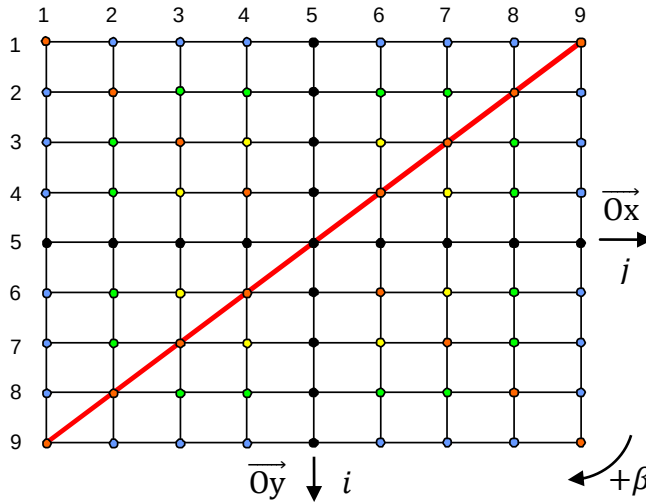
$$\cos(\beta_{76}) = \frac{d_{56}}{d_{76}}, \quad \sin(\beta_{76}) = \frac{d_{75}}{d_{76}},$$

$$(r_{x76}, r_{y76}) = \frac{d_{76}}{d_{59}} (r_x, r_y) \mathcal{R}(\beta_{76}) = (r_x, r_y) \begin{bmatrix} \frac{d_{56}}{d_{59}} & \frac{d_{75}}{d_{59}} \\ -\frac{d_{75}}{d_{59}} & \frac{d_{56}}{d_{59}} \end{bmatrix} = \left(\frac{1}{4}r_x - \frac{1}{2}kr_y, \frac{1}{2}kr_x + \frac{1}{4}r_y \right) \quad (12ж)$$

Таблица 9. Система от покомпонентни уравнения за обхождане по „вътрешните“ диагонални прави.

център (i, j)	LSM on $\vec{Ox}$	LSM on $\vec{Oy}$
(6,8) +	$T_x + \frac{3}{4}r_x - \frac{1}{4}kr_y - t_{x68} = \varepsilon_{x68}$	$T_y + \frac{1}{4}kr_x + \frac{3}{4}r_y - t_{y68} = \varepsilon_{y68}$
(4,2) -	$T_x - \frac{3}{4}r_x + \frac{1}{4}kr_y - t_{x42} = \varepsilon_{x42}$	$T_y - \frac{1}{4}kr_x - \frac{3}{4}r_y - t_{y42} = \varepsilon_{y42}$
(7,8) +	$T_x + \frac{3}{4}r_x - \frac{1}{2}kr_y - t_{x78} = \varepsilon_{x78}$	$T_y + \frac{1}{2}kr_x + \frac{3}{4}r_y - t_{y78} = \varepsilon_{y78}$
(3,2) -	$T_x - \frac{3}{4}r_x + \frac{1}{2}kr_y - t_{x32} = \varepsilon_{x32}$	$T_y - \frac{1}{2}kr_x - \frac{3}{4}r_y - t_{y32} = \varepsilon_{y32}$
(8,7) +	$T_x + \frac{1}{2}r_x - \frac{3}{4}kr_y - t_{x87} = \varepsilon_{x87}$	$T_y + \frac{3}{4}kr_x + \frac{1}{2}r_y - t_{y87} = \varepsilon_{y87}$
(2,3) -	$T_x - \frac{1}{2}r_x + \frac{3}{4}kr_y - t_{x23} = \varepsilon_{x23}$	$T_y - \frac{3}{4}kr_x - \frac{1}{2}r_y - t_{y23} = \varepsilon_{y23}$
(8,6) +	$T_x + \frac{1}{4}r_x - \frac{3}{4}kr_y - t_{x86} = \varepsilon_{x86}$	$T_y + \frac{3}{4}kr_x + \frac{1}{4}r_y - t_{y86} = \varepsilon_{y86}$
(2,4) -	$T_x - \frac{1}{4}r_x + \frac{3}{4}kr_y - t_{x24} = \varepsilon_{x24}$	$T_y - \frac{3}{4}kr_x - \frac{1}{4}r_y - t_{y24} = \varepsilon_{y24}$
(6,7) +	$T_x + \frac{1}{2}r_x - \frac{1}{4}kr_y - t_{x67} = \varepsilon_{x67}$	$T_y + \frac{1}{4}kr_x + \frac{1}{2}r_y - t_{y67} = \varepsilon_{y67}$
(4,3) -	$T_x - \frac{1}{2}r_x + \frac{1}{4}kr_y - t_{x43} = \varepsilon_{x43}$	$T_y - \frac{1}{4}kr_x - \frac{1}{2}r_y - t_{y43} = \varepsilon_{y43}$
(7,6) +	$T_x + \frac{1}{4}r_x - \frac{1}{2}kr_y - t_{x76} = \varepsilon_{x76}$	$T_y + \frac{1}{2}kr_x + \frac{1}{4}r_y - t_{y76} = \varepsilon_{y76}$
(3,4) -	$T_x - \frac{1}{4}r_x + \frac{1}{2}kr_y - t_{x34} = \varepsilon_{x34}$	$T_y - \frac{1}{2}kr_x - \frac{1}{4}r_y - t_{y34} = \varepsilon_{y34}$

Останалите уравнения се съставят по аналогичен начин на **С3**, **С4**, **С5**, с единствената разлика в знака на завъртането на референтния вектор на ротация, на ъгъл $(-\beta)$.



Фиг. 27. Симетрична схема на **С3** за построяване на уравненията по главния диагонал.

Например за симетричния случай **С3** (виж Фиг. 27), получаваме:

$$\mathcal{R}(-\beta_{99}) = \begin{bmatrix} \cos(\beta_{99}) & -\sin(\beta_{99}) \\ +\sin(\beta_{99}) & \cos(\beta_{99}) \end{bmatrix}$$

$$\cos(\beta_{99}) = \frac{d_{59}}{d_{99}}, \quad \sin(\beta_{99}) = \frac{d_{95}}{d_{99}}$$

$$(r_{x19}, r_{y19}) = \frac{d_{99}}{d_{59}} (r_x, r_y) \mathcal{R}(-\beta_{99})$$

$$= (r_x, r_y) \begin{bmatrix} 1 & -\frac{d_{95}}{d_{59}} \\ \frac{d_{95}}{d_{59}} & 1 \end{bmatrix}$$

$$= (r_x + kr_y, -kr_x + r_y)$$

В Табл. 10 са дадени уравненията за всички центрове (i, j) , сортирани за случая на обхождане на схемата “9×9” по редове, с цел извличане на определени нагледни зависимости между коефициентите пред неизвестните параметри (виж Табл. 11).

Таблица 10. Система от 162 покомпонентни уравнения за схемата на разбиване “9×9”.

(i, j)	$\vec{Ox}$	$\vec{Oy}$
(1,1)	$T_x - r_x + kr_y = t_{x11}$	$T_y - kr_x - r_y = t_{y11}$
(1,2)	$T_x - (3/4)r_x + kr_y = t_{x12}$	$T_y - kr_x - (3/4)r_y = t_{y12}$
(1,3)	$T_x - (1/2)r_x + kr_y = t_{x13}$	$T_y - kr_x - (1/2)r_y = t_{y13}$
(1,4)	$T_x - (1/4)r_x + kr_y = t_{x14}$	$T_y - kr_x - (1/4)r_y = t_{y14}$
(1,5)	$T_x + kr_y = t_{x15}$	$T_y - kr_x = t_{y15}$
(1,6)	$T_x + (1/4)r_x + kr_y = t_{x16}$	$T_y - kr_x + (1/4)r_y = t_{y16}$
(1,7)	$T_x + (1/2)r_x + kr_y = t_{x17}$	$T_y - kr_x + (1/2)r_y = t_{y17}$
(1,8)	$T_x + (3/4)r_x + kr_y = t_{x18}$	$T_y - kr_x + (3/4)r_y = t_{y18}$
(1,9)	$T_x + r_x + kr_y = t_{x19}$	$T_y - kr_x + r_y = t_{y19}$
(2,1)	$T_x - r_x + (3/4)kr_y = t_{x21}$	$T_y - (3/4)kr_x - r_y = t_{y21}$
(2,2)	$T_x - (3/4)r_x + (3/4)kr_y = t_{x22}$	$T_y - (3/4)kr_x - (3/4)r_y = t_{y22}$
(2,3)	$T_x - (1/2)r_x + (3/4)kr_y = t_{x23}$	$T_y - (3/4)kr_x - (1/2)r_y = t_{y23}$
(2,4)	$T_x - (1/4)r_x + (3/4)kr_y = t_{x24}$	$T_y - (3/4)kr_x - (1/4)r_y = t_{y24}$
(2,5)	$T_x + (3/4)kr_y = t_{x25}$	$T_y - (3/4)kr_x = t_{y25}$
(2,6)	$T_x + (1/4)r_x + (3/4)kr_y = t_{x26}$	$T_y - (3/4)kr_x + (1/4)r_y = t_{y26}$

(2,7)	$T_x + (1/2)r_x + (3/4)kr_y = t_{x27}$	$T_y - (3/4)kr_x + (1/2)r_y = t_{y27}$
(2,8)	$T_x + (3/4)r_x + (3/4)kr_y = t_{x28}$	$T_y - (3/4)kr_x + (3/4)r_y = t_{y28}$
(2,9)	$T_x + r_x + (3/4)kr_y = t_{x29}$	$T_y - (3/4)kr_x + r_y = t_{y29}$
(3,1)	$T_x - r_x + (1/2)kr_y = t_{x31}$	$T_y - (1/2)kr_x - r_y = t_{y31}$
(3,2)	$T_x - (3/4)r_x + (1/2)kr_y = t_{x32}$	$T_y - (1/2)kr_x - (3/4)r_y = t_{y32}$
(3,3)	$T_x - (1/2)r_x + (1/2)kr_y = t_{x33}$	$T_y - (1/2)kr_x - (1/2)r_y = t_{y33}$
(3,4)	$T_x - (1/4)r_x + (1/2)kr_y = t_{x34}$	$T_y - (1/2)kr_x - (1/4)r_y = t_{y34}$
(3,5)	$T_x + (1/2)kr_y = t_{x35}$	$T_y - (1/2)kr_x = t_{y35}$
(3,6)	$T_x + 1/4r_x + (1/2)kr_y = t_{x36}$	$T_y - (1/2)kr_x + 1/4r_y = t_{y36}$
(3,7)	$T_x + (1/2)r_x + (1/2)kr_y = t_{x37}$	$T_y - (1/2)kr_x + (1/2)r_y = t_{y37}$
(3,8)	$T_x + (3/4)r_x + (1/2)kr_y = t_{x38}$	$T_y - (1/2)kr_x + (3/4)r_y = t_{y38}$
(3,9)	$T_x + r_x + (1/2)kr_y = t_{x39}$	$T_y - (1/2)kr_x + r_y = t_{y39}$
(4,1)	$T_x - r_x + (1/4)kr_y = t_{x41}$	$T_y - (1/4)kr_x - r_y = t_{y41}$
(4,2)	$T_x - 3/4r_x + (1/4)kr_y = t_{x42}$	$T_y - (1/4)kr_x - 3/4r_y = t_{y42}$
(4,3)	$T_x - 1/2r_x + (1/4)kr_y = t_{x43}$	$T_y - (1/4)kr_x - 1/2r_y = t_{y43}$
(4,4)	$T_x - (1/4)r_x + (1/4)kr_y = t_{x44}$	$T_y - (1/4)kr_x - (1/4)r_y = t_{y44}$
(4,5)	$T_x + (1/4)kr_y = t_{x45}$	$T_y - (1/4)kr_x = t_{y45}$
(4,6)	$T_x + (1/4)r_x + (1/4)kr_y = t_{x46}$	$T_y - (1/4)kr_x + (1/4)r_y = t_{y46}$
(4,7)	$T_x + 1/2r_x + (1/4)kr_y = t_{x47}$	$T_y - (1/4)kr_x + 1/2r_y = t_{y47}$
(4,8)	$T_x + 3/4r_x + (1/4)kr_y = t_{x48}$	$T_y - (1/4)kr_x + 3/4r_y = t_{y48}$
(4,9)	$T_x + r_x + (1/4)kr_y = t_{x49}$	$T_y - (1/4)kr_x + r_y = t_{y49}$
(5,1)	$T_x - r_x = t_{x51}$	$T_y - r_y = t_{y51}$
(5,2)	$T_x - (3/4)r_x = t_{x52}$	$T_y - (3/4)r_y = t_{y52}$
(5,3)	$T_x - (1/2)r_x = t_{x53}$	$T_y - (1/2)r_y = t_{y53}$
(5,4)	$T_x - (1/4)r_x = t_{x54}$	$T_y - (1/4)r_y = t_{y54}$
(5,5)	$T_x = t_{x55}$	$T_y = t_{y55}$
(5,6)	$T_x + (1/4)r_x = t_{x56}$	$T_y + (1/4)r_y = t_{y56}$
(5,7)	$T_x + (1/2)r_x = t_{x57}$	$T_y + (1/2)r_y = t_{y57}$
(5,8)	$T_x + (3/4)r_x = t_{x58}$	$T_y + (3/4)r_y = t_{y58}$
(5,9)	$T_x + r_x = t_{x59}$	$T_y + r_y = t_{y59}$
(6,1)	$T_x - r_x - (1/4)kr_y = t_{x61}$	$T_y + (1/4)kr_x - r_y = t_{y61}$
(6,2)	$T_x - (3/4)r_x - (1/4)kr_y = t_{x62}$	$T_y + (1/4)kr_x - (3/4)r_y = t_{y62}$
(6,3)	$T_x - (1/2)r_x - (1/4)kr_y = t_{x63}$	$T_y + (1/4)kr_x - (1/2)r_y = t_{y63}$
(6,4)	$T_x - (1/4)r_x - (1/4)kr_y = t_{x64}$	$T_y + (1/4)kr_x - (1/4)r_y = t_{y64}$
(6,5)	$T_x - (1/4)kr_y = t_{x65}$	$T_y + (1/4)kr_x = t_{y65}$
(6,6)	$T_x + (1/4)r_x - (1/4)kr_y = t_{x66}$	$T_y + (1/4)kr_x + (1/4)r_y = t_{y66}$

(6,7)	$T_x + (1/2)r_x - (1/4)kr_y = t_{x67}$	$T_y + (1/4)kr_x + (1/2)r_y = t_{y67}$
(6,8)	$T_x + (3/4)r_x - (1/4)kr_y = t_{x68}$	$T_y + (1/4)kr_x + (3/4)r_y = t_{y68}$
(6,9)	$T_x + r_x - (1/4)kr_y = t_{x69}$	$T_y + (1/4)kr_x + r_y = t_{y69}$
(7,1)	$T_x - r_x - (1/2)kr_y = t_{x71}$	$T_y + (1/2)kr_x - r_y = t_{y71}$
(7,2)	$T_x - (3/4)r_x - (1/2)kr_y = t_{x72}$	$T_y + (1/2)kr_x - (3/4)r_y = t_{y72}$
(7,3)	$T_x - (1/2)r_x - (1/2)kr_y = t_{x73}$	$T_y + (1/2)kr_x - (1/2)r_y = t_{y73}$
(7,4)	$T_x - (1/4)r_x - (1/2)kr_y = t_{x74}$	$T_y + (1/2)kr_x - (1/4)r_y = t_{y74}$
(7,5)	$T_x - (1/2)kr_y = t_{x75}$	$T_y + (1/2)kr_x = t_{y75}$
(7,6)	$T_x + (1/4)r_x - (1/2)kr_y = t_{x76}$	$T_y + (1/2)kr_x + (1/4)r_y = t_{y76}$
(7,7)	$T_x + (1/2)r_x - (1/2)kr_y = t_{x77}$	$T_y + (1/2)kr_x + (1/2)r_y = t_{y77}$
(7,8)	$T_x + (3/4)r_x - (1/2)kr_y = t_{x78}$	$T_y + (1/2)kr_x + (3/4)r_y = t_{y78}$
(7,9)	$T_x + r_x - (1/2)kr_y = t_{x79}$	$T_y + (1/2)kr_x + r_y = t_{y79}$
(8,1)	$T_x - r_x - (3/4)kr_y = t_{x81}$	$T_y + (3/4)kr_x - r_y = t_{y81}$
(8,2)	$T_x - (3/4)r_x - (3/4)kr_y = t_{x82}$	$T_y + (3/4)kr_x - (3/4)r_y = t_{y82}$
(8,3)	$T_x - (1/2)r_x - (3/4)kr_y = t_{x83}$	$T_y + (3/4)kr_x - (1/2)r_y = t_{y83}$
(8,4)	$T_x - (1/4)r_x - (3/4)kr_y = t_{x84}$	$T_y + (3/4)kr_x - (1/4)r_y = t_{y84}$
(8,5)	$T_x - (3/4)kr_y = t_{x85}$	$T_y + (3/4)kr_x = t_{y85}$
(8,6)	$T_x + (1/4)r_x - (3/4)kr_y = t_{x86}$	$T_y + (3/4)kr_x + (1/4)r_y = t_{y86}$
(8,7)	$T_x + (1/2)r_x - (3/4)kr_y = t_{x87}$	$T_y + (3/4)kr_x + (1/2)r_y = t_{y87}$
(8,8)	$T_x + (3/4)r_x - (3/4)kr_y = t_{x88}$	$T_y + (3/4)kr_x + (3/4)r_y = t_{y88}$
(8,9)	$T_x + r_x - (3/4)kr_y = t_{x89}$	$T_y + (3/4)kr_x + r_y = t_{y89}$
(9,1)	$T_x - r_x - kr_y = t_{x91}$	$T_y + kr_x - r_y = t_{y91}$
(9,2)	$T_x - (3/4)r_x - kr_y = t_{x92}$	$T_y + kr_x - (3/4)r_y = t_{y92}$
(9,3)	$T_x - (1/2)r_x - kr_y = t_{x93}$	$T_y + kr_x - (1/2)r_y = t_{y93}$
(9,4)	$T_x - (1/4)r_x - kr_y = t_{x94}$	$T_y + kr_x - (1/4)r_y = t_{y94}$
(9,5)	$T_x - kr_y = t_{x95}$	$T_y + kr_x = t_{y95}$
(9,6)	$T_x + (1/4)r_x - kr_y = t_{x96}$	$T_y + kr_x + (1/4)r_y = t_{y96}$
(9,7)	$T_x + (1/2)r_x - kr_y = t_{x97}$	$T_y + kr_x + (1/2)r_y = t_{y97}$
(9,8)	$T_x + (3/4)r_x - kr_y = t_{x98}$	$T_y + kr_x + (3/4)r_y = t_{y98}$
(9,9)	$T_x + r_x - kr_y = t_{x99}$	$T_y + kr_x + r_y = t_{y99}$

Таблица 11. Систематизиране на коефициентите пред неизвестните параметри от Табл. 10.

A_1	$T_{xx} (T_x \text{ по } \vec{Ox})$									$T_{xy} (T_x \text{ по } \vec{Oy})$								
(i, j)	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
2	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
3	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
4	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
5	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
6	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
7	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
8	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
9	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0

A_2	$T_{yx} (T_y \text{ по } \vec{Ox})$									$T_{yy} (T_y \text{ по } \vec{Oy})$								
(i, j)	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
2	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
3	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
4	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
5	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
6	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
7	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
8	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
9	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1

A_3	$R_{xx} (R_x \text{ по } \vec{Ox})$									$R_{xy} (R_x \text{ по } \vec{Oy})$								
(i, j)	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
1	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1	-k	-k	-k	-k	-k	-k	-k	-k	-k
2	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1	-3k/4	-3k/4	-3k/4	-3k/4	-3k/4	-3k/4	-3k/4	-3k/4	-3k/4
3	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1	-k/2	-k/2	-k/2	-k/2	-k/2	-k/2	-k/2	-k/2	-k/2
4	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1	-k/4	-k/4	-k/4	-k/4	-k/4	-k/4	-k/4	-k/4	-k/4
5	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1	0	0	0	0	0	0	0	0	0
6	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1	k/4	k/4	k/4	k/4	k/4	k/4	k/4	k/4	k/4
7	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1	k/2	k/2	k/2	k/2	k/2	k/2	k/2	k/2	k/2
8	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1	3k/4	3k/4	3k/4	3k/4	3k/4	3k/4	3k/4	3k/4	3k/4
9	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1	k	k	k	k	k	k	k	k	k

A_4	$R_{yx} (R_y \text{ по } \vec{Ox})$									$R_{yy} (R_y \text{ по } \vec{Oy})$								
(i, j)	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
1	k	k	k	k	k	k	k	k	k	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1
2	3k/4	3k/4	3k/4	3k/4	3k/4	3k/4	3k/4	3k/4	3k/4	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1
3	k/2	k/2	k/2	k/2	k/2	k/2	k/2	k/2	k/2	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1
4	k/4	k/4	k/4	k/4	k/4	k/4	k/4	k/4	k/4	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1
5	0	0	0	0	0	0	0	0	0	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1
6	-k/4	-k/4	-k/4	-k/4	-k/4	-k/4	-k/4	-k/4	-k/4	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1
7	-k/2	-k/2	-k/2	-k/2	-k/2	-k/2	-k/2	-k/2	-k/2	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1
8	-3k/4	-3k/4	-3k/4	-3k/4	-3k/4	-3k/4	-3k/4	-3k/4	-3k/4	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1
9	-k	-k	-k	-k	-k	-k	-k	-k	-k	-1	-3/4	-1/2	-1/4	0	1/4	1/2	3/4	1

За решаването на преопределена система от 162 линейни уравнения (виж Табл. 10) отново използваме идеята за записване на системата в матричен вид $(A^T A)X = A^T B \Rightarrow X = (A^T A)^{-1} A^T B$, след което я решаваме за неизвестния вектор X , съдържащ търсените параметри на движението.

$$\begin{array}{c}
 \overrightarrow{Ox} \\
 \overrightarrow{Oy}
 \end{array}
 \cdot
 \begin{bmatrix}
 A_1 & A_2 & A_3 & A_4 \\
 1 & 0 & -1 & k \\
 1 & 0 & -\frac{3}{4} & k \\
 1 & 0 & -\frac{1}{2} & k \\
 \vdots & \vdots & \vdots & \vdots \\
 \vdots & \vdots & \vdots & \vdots \\
 1 & 0 & \frac{1}{2} & -k \\
 1 & 0 & \frac{3}{4} & -k \\
 1 & 0 & 1 & -k \\
 0 & 1 & -k & -\frac{1}{4} \\
 0 & 1 & -k & -\frac{3}{4} \\
 0 & 1 & -k & -\frac{1}{2} \\
 \vdots & \vdots & \vdots & \vdots \\
 \vdots & \vdots & \vdots & \vdots \\
 0 & 1 & k & \frac{1}{2} \\
 0 & 1 & k & \frac{3}{4} \\
 0 & 1 & k & 1
 \end{bmatrix}
 \cdot
 \begin{bmatrix}
 T_x \\
 T_y \\
 r_x \\
 r_y
 \end{bmatrix}
 =
 \begin{bmatrix}
 t_{x11} \\
 t_{x12} \\
 t_{x13} \\
 \vdots \\
 \vdots \\
 t_{x97} \\
 t_{x98} \\
 t_{x99} \\
 t_{y11} \\
 t_{y12} \\
 t_{y13} \\
 \vdots \\
 \vdots \\
 \vdots \\
 t_{y97} \\
 t_{y98} \\
 t_{y99}
 \end{bmatrix}$$

$$A_{[162 \times 4]} \cdot X_{[4 \times 1]} = B_{[162 \times 1]}$$

$$A = \{A_{ij}\}_{162 \times 4} = [A_1 \ A_2 \ A_3 \ A_4] = \left[\begin{pmatrix} T_{xx} \\ T_{xy} \end{pmatrix}, \begin{pmatrix} T_{yx} \\ T_{yy} \end{pmatrix}, \begin{pmatrix} R_{xx} \\ R_{xy} \end{pmatrix}, \begin{pmatrix} R_{yx} \\ R_{yy} \end{pmatrix} \right]$$

Фиг. 28. Запис на системата уравнения от Табл. 10 в матричен вид $AX = B$.

Матричното уравнение от Фиг. 28, (виж 5а) представяме във вида $(A^T A)X = A^T B$:

$$\begin{bmatrix}
 a_{11} & a_{12} & a_{13} & a_{14} \\
 a_{21} & a_{22} & a_{23} & a_{24} \\
 a_{31} & a_{32} & a_{33} & a_{34} \\
 a_{41} & a_{42} & a_{43} & a_{44}
 \end{bmatrix}
 \cdot
 \begin{bmatrix}
 T_x \\
 T_y \\
 r_x \\
 r_y
 \end{bmatrix}
 =
 \begin{bmatrix}
 a \\
 b \\
 c \\
 d
 \end{bmatrix}, \quad (13)$$

$$(A^T A)_{[4 \times 4]} \cdot X_{[4 \times 1]} = (A^T B)_{[4 \times 1]}$$

където, където матрицата $(A^T A)_{[4 \times 4]}$ се получава *диагонална* (подобно на 5б):

$$A^T A = \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ 0 & 0 & a_{33} & 0 \\ 0 & 0 & 0 & a_{44} \end{bmatrix}, \quad (13a)$$

т.е. $a_{ij} = 0$, $i \neq j$, $(i, j = 1 \div 4)$ и след еквивалентни преобразования, за нейните елементи $(a_{11}, a_{22}, a_{33}, a_{44})$ получаваме:

$$a_{11} = \sum_{i=1}^9 \sum_{j=1}^9 T_{xxij}^2 = 81; \quad a_{22} = \sum_{i=1}^9 \sum_{j=1}^9 T_{yyij}^2 = 81; \quad (13б)$$

$$a_{33} = \sum_{i=1}^9 \sum_{j=1}^9 (R_{xxij}^2 + R_{xyij}^2) = 18 \left(1^2 + \left(\frac{3}{4}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{4}\right)^2 \right) + 18k^2 \left(1^2 + \left(\frac{3}{4}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{4}\right)^2 \right) = \frac{135}{4} + \frac{135}{4}k^2; \quad (13в)$$

$$a_{44} = \sum_{i=1}^9 \sum_{j=1}^9 (R_{yxij}^2 + R_{yyij}^2) = \frac{135}{4} + \frac{135}{4}k^2; \quad (13г)$$

а за коефициентите на вектора-стълб $A^T B = (a, b, c, d)^T$ имаме:

$$\mathbf{a} = \sum_{i=1}^9 \sum_{j=1}^9 t_{xij}; \quad \mathbf{b} = \sum_{i=1}^9 \sum_{j=1}^9 t_{yij}; \quad (13д)$$

$$\begin{aligned} \mathbf{c} = & k \sum_{j=1}^9 (t_{y9j} - t_{y1j}) + \frac{3}{4}k \sum_{j=1}^9 (t_{y8j} - t_{y2j}) + \frac{1}{2}k \sum_{j=1}^9 (t_{y7j} - t_{y3j}) + \frac{1}{4}k \sum_{j=1}^9 (t_{y6j} - t_{y4j}) \\ & + \sum_{i=1}^9 (t_{xi9} - t_{xi1}) + \frac{3}{4} \sum_{i=1}^9 (t_{xi8} - t_{xi2}) + \frac{1}{2} \sum_{i=1}^9 (t_{xi7} - t_{xi3}) + \frac{1}{4} \sum_{i=1}^9 (t_{xi6} - t_{xi4}); \end{aligned} \quad (13е)$$

$$\begin{aligned} \mathbf{d} = & k \sum_{j=1}^9 (t_{x1j} - t_{x9j}) + \frac{3}{4}k \sum_{j=1}^9 (t_{x2j} - t_{x8j}) + \frac{1}{2}k \sum_{j=1}^9 (t_{x3j} - t_{x7j}) + \frac{1}{4}k \sum_{j=1}^9 (t_{x4j} - t_{x6j}) \\ & + \sum_{i=1}^9 (t_{yi9} - t_{yi1}) + \frac{3}{4} \sum_{i=1}^9 (t_{yi8} - t_{yi2}) + \frac{1}{2} \sum_{i=1}^9 (t_{yi7} - t_{yi3}) + \frac{1}{4} \sum_{i=1}^9 (t_{yi6} - t_{yi4}); \end{aligned} \quad (13ж)$$

Така, за търсеното LSM решение $\mathbf{X} = [T_x, T_y, r_x, r_y]^T$, финално получаваме:

$$T_x = \frac{a}{81}, \quad T_y = \frac{b}{81}, \quad r_x = \frac{4c}{135 + 135k^2}, \quad r_y = \frac{4d}{135 + 135k^2}, \quad (13з)$$

където $k = y_{wsiz} / x_{wsiz}$, а ъгълът на ротация α се определя чрез компонентите на референтния вектор $\vec{r}$, като $\alpha = 2\arctg(r_x/r_y)$.

Отново (както при метода “3×3”) прилагаме LSM втори път с цел редуциране на влиянието върху точността на решението от некоректно изчислените вектори на движението чрез SAD метода. За целта елиминираме (по познатия Гаусов подход, виж §2.1.2) определени уравнения (от Табл. 10) в зависимост от стойността на параметрите e_{xij} и e_{yij} , ($i = 1, 2, \dots, 9$), ($j = 1, 2, \dots, 9$), като налагаме ограничението за едновременно елиминироване на двойка уравнения за съответния център (i, j) , т.е. $e_{xij} = e_{yij} \in \{0, 1\}$.

По подобен начин, както при 1-вия LSM пас (виж Фиг. 28), решаваме модифицираната система (с параметрите e_{ij}) от Табл. 10 (аналогична на Табл. 4) по познатия матричен подход $(A^T A)X = A^T B$, като $A^T A$ има следния по-общ вид:

$$A^T A = \begin{bmatrix} a_{11} & 0 & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & 0 & a_{44} \end{bmatrix}, \quad (14)$$

където съответните коефициенти са:

$$a_{11} = \sum_{i=1}^9 \sum_{j=1}^9 T_{xxij} e_{xij} ; \quad a_{22} = \sum_{i=1}^9 \sum_{j=1}^9 T_{yyij} e_{yij}; \quad (14a)$$

$$a_{13} = \sum_{i=1}^9 \sum_{j=1}^9 R_{xxij} e_{xij} ; \quad a_{24} = \sum_{i=1}^9 \sum_{j=1}^9 R_{yyij} e_{yij}; \quad (14б)$$

$$a_{14} = \sum_{i=1}^9 \sum_{j=1}^9 R_{yxij} e_{xij} ; \quad a_{23} = - \sum_{i=1}^9 \sum_{j=1}^9 R_{xyij} e_{yij}; \quad (14в)$$

$$a_{33} = \sum_{i=1}^9 \sum_{j=1}^9 \left((R_{xxij} e_{xij})^2 + (R_{xyij} e_{yij})^2 \right) ; \quad a_{44} = \sum_{i=1}^9 \sum_{j=1}^9 \left((R_{yyij} e_{yij})^2 + (R_{yxij} e_{xij})^2 \right); \quad (14г)$$

$$a_{12} = a_{21} = a_{34} = a_{43} = 0. \quad (14д)$$

След възприемане на ограничението $e_{xij} = e_{yij} \in \{0, 1\}$, т.е. $a_{11} = a_{22}$, $a_{13} = a_{24}$, $a_{14} = a_{23}$, $a_{33} = a_{44}$, и вземайки предвид, че: $T_{xxij} = T_{yyij}$, $R_{xxij} = R_{yyij}$, $R_{yxij} = -R_{xyij}$ (виж Табл. 11), то матричното уравнение $(A^T A)X = A^T B$, приема вида:

$$\begin{bmatrix} a_{11} & 0 & a_{13} & a_{14} \\ 0 & a_{11} & -a_{14} & a_{13} \\ a_{13} & -a_{14} & a_{33} & 0 \\ a_{14} & a_{13} & 0 & a_{33} \end{bmatrix} \cdot \begin{bmatrix} T_x \\ T_y \\ r_x \\ r_y \end{bmatrix} = \begin{bmatrix} \mathcal{A} \\ \mathcal{B} \\ \mathcal{C} \\ \mathcal{D} \end{bmatrix}, \quad (15)$$

където:

$$\mathcal{A} = \sum_{i=1}^9 \sum_{j=1}^9 t_{xij} e_{xij} ; \quad \mathcal{B} = \sum_{i=1}^9 \sum_{j=1}^9 t_{yij} e_{yij} ; \quad (15a)$$

$$\begin{aligned}
\mathcal{C} = & k \sum_{j=1}^9 (t_{y9j}e_{y9j} - t_{y1j}e_{y1j}) + \frac{3}{4}k \sum_{j=1}^9 (t_{y8j}e_{y8j} - t_{y2j}e_{y2j}) + \frac{1}{2}k \sum_{j=1}^9 (t_{y7j}e_{y7j} - t_{y3j}e_{y3j}) \\
& + \frac{1}{4}k \sum_{j=1}^9 (t_{y6j}e_{y6j} - t_{y4j}e_{y4j}) + \sum_{i=1}^9 (t_{xi9}e_{xi9} - t_{xi1}e_{xi1}) + \frac{3}{4} \sum_{i=1}^9 (t_{xi8}e_{xi8} - t_{xi2}e_{xi2}) \\
& + \frac{1}{2} \sum_{i=1}^9 (t_{xi7}e_{xi7} - t_{xi3}e_{xi3}) + \frac{1}{4} \sum_{i=1}^9 (t_{xi6}e_{xi6} - t_{xi4}e_{xi4});
\end{aligned} \tag{156}$$

$$\begin{aligned}
\mathcal{D} = & k \sum_{j=1}^9 (t_{x1j}e_{x1j} - t_{x9j}e_{x9j}) + \frac{3}{4}k \sum_{j=1}^9 (t_{x2j}e_{x2j} - t_{x8j}e_{x8j}) + \frac{1}{2}k \sum_{j=1}^9 (t_{x3j}e_{x3j} - t_{x7j}e_{x7j}) \\
& + \frac{1}{4}k \sum_{j=1}^9 (t_{x4j}e_{x4j} - t_{x6j}e_{x6j}) + \sum_{i=1}^9 (t_{yi9}e_{yi9} - t_{yi1}e_{yi1}) + \frac{3}{4} \sum_{i=1}^9 (t_{yi8}e_{yi8} - t_{yi2}e_{yi2}) \\
& + \frac{1}{2} \sum_{i=1}^9 (t_{yi7}e_{yi7} - t_{yi3}e_{yi3}) + \frac{1}{4} \sum_{i=1}^9 (t_{yi6}e_{yi6} - t_{yi4}e_{yi4});
\end{aligned} \tag{15B}$$

За намиране на неизвестните, решаваме (чрез заместване) резултантната линейна система:

$$\begin{cases} a_{11}T_x & + 0 & + a_{13}r_x & + a_{14}r_y & = \mathcal{A} \\ 0 & + a_{11}T_y & - a_{14}r_x & + a_{13}r_y & = \mathcal{B} \\ a_{13}T_x & - a_{14}T_y & + a_{33}r_x & + 0 & = \mathcal{C} \\ a_{14}T_x & + a_{13}T_y & + 0 & + a_{33}r_y & = \mathcal{D} \end{cases} \tag{15Г}$$

От първите две уравнения изразяваме T_x и T_y :

$$T_x = \frac{\mathcal{A} - a_{13}r_x - a_{14}r_y}{a_{11}}, \quad T_y = \frac{\mathcal{B} - a_{13}r_y + a_{14}r_x}{a_{11}} \tag{15Д}$$

Замествайки T_x и T_y в последните две уравнения на системата, получаваме:

$$a_{13} \frac{\mathcal{A} - a_{13}r_x - a_{14}r_y}{a_{11}} - a_{14} \frac{\mathcal{B} - a_{13}r_y + a_{14}r_x}{a_{11}} + a_{33}r_x = \mathcal{C} \tag{15Е}$$

$$\begin{aligned}
& a_{14} \frac{\mathcal{A} - a_{13}r_x - a_{14}r_y}{a_{11}} + a_{13} \frac{\mathcal{B} - a_{13}r_y + a_{14}r_x}{a_{11}} + a_{33}r_y = \mathcal{D} \\
\Rightarrow \quad r_x &= \frac{a_{11}\mathcal{C} - a_{13}\mathcal{A} + a_{14}\mathcal{B}}{-a_{13}^2 - a_{14}^2 + a_{11}a_{33}}, \quad r_y = \frac{a_{11}\mathcal{D} - a_{13}\mathcal{B} - a_{14}\mathcal{A}}{-a_{13}^2 - a_{14}^2 + a_{11}a_{33}}
\end{aligned} \tag{15Ж}$$

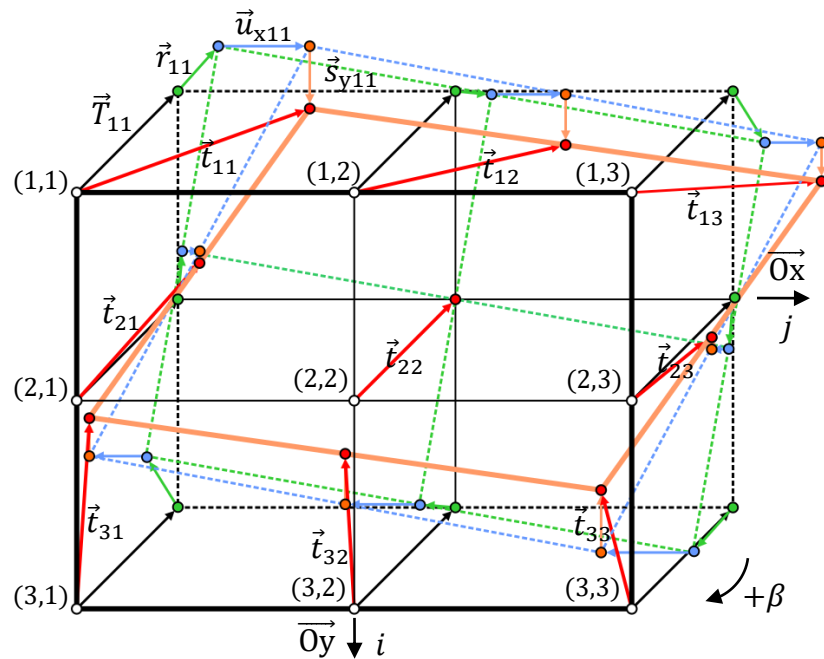
Така, за неизвестните параметри T_x , T_y , r_x , r_y , финално получаваме:

$$\begin{aligned}
T_x &= \frac{-a_{13}\mathcal{C} - a_{14}\mathcal{D} + a_{33}\mathcal{A}}{-a_{13}^2 - a_{14}^2 + a_{11}a_{33}}; \quad T_y = \frac{-a_{13}\mathcal{D} + a_{14}\mathcal{C} + a_{33}\mathcal{B}}{-a_{13}^2 - a_{14}^2 + a_{11}a_{33}}; \\
r_x &= \frac{a_{11}\mathcal{C} - a_{13}\mathcal{A} + a_{14}\mathcal{B}}{-a_{13}^2 - a_{14}^2 + a_{11}a_{33}}; \quad r_y = \frac{a_{11}\mathcal{D} - a_{13}\mathcal{B} - a_{14}\mathcal{A}}{-a_{13}^2 - a_{14}^2 + a_{11}a_{33}}.
\end{aligned} \tag{15З}$$

2.1.4 “Rolling shutter” разширение на базовия векторен модел

Известно е, че камерите, базирани на CMOS технология, страдат от т.нар. “rolling shutter” ефект, поради начина на формиране изображения във времето, а именно ред по ред, а не целия 2D масив едновременно, както е при CCD технологията. Т.е. при хоризонтално транслиране на камерата ефектът се проявява като накланяне на изображенията, а при вертикално транслиране ефектът е вертикално мащабиране. Подобен ефект се наблюдава и при статична камера, но при наличие на подвижни обекти в заснеманата сцена. Разбира се, колкото по надълбоко в сцената е обектът, толкова влиянието от движението му върху ефекта “rolling shutter” намалява (за справка виж Глава 1, §1.1.4). По тази причина, освен към обичайните трансляция и ротация, в резултат от потрепването на камерата се наслагват още две съставки, които трябва да се компенсират за осигуряване на оригиналната твърдотелност на сцената.

Вземайки предвид горните особености на CMOS камерите, тук предлагаме аналогично на базовия CCD векторен модел (виж §2.1.2), базов CMOS векторен модел. В него, всеки оценен по SAD вектор на движение $\vec{t}_{ij}$ между два кадъра се разлага на сума от четири вектора: вектор на трансляция $\vec{T}_{ij}$, вектор на ротация $\vec{r}_{ij}$, вектор на хоризонтален наклон $\vec{u}_{xij}$ и вектор на вертикално мащабиране $\vec{s}_{yij}$ т.е. $\vec{t}_{ij} = \vec{T}_{ij} + \vec{r}_{ij} + \vec{u}_{xij} + \vec{s}_{yij}$ (виж Фиг. 29).



Фиг. 29. Разширен векторен модел, отчитащ трансляция, ротация, хоризонтално накланяне и вертикално мащабиране между два кадъра по зависимостта: $\vec{t}_{ij} = \vec{T}_{ij} + \vec{r}_{ij} + \vec{u}_{xij} + \vec{s}_{yij}$.

	$S_{yx} (S_y \text{ on } \vec{Ox})$									$S_{yy} (S_y \text{ on } \vec{Oy})$								
(i, j)	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
1	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1
2	0	0	0	0	0	0	0	0	0	3/4	3/4	3/4	3/4	0	3/4	3/4	3/4	3/4
3	0	0	0	0	0	0	0	0	0	1/2	1/2	1/2	1/2	0	1/2	1/2	1/2	1/2
4	0	0	0	0	0	0	0	0	0	1/4	1/4	1/4	1/4	0	1/4	1/4	1/4	1/4
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	-1/4	-1/4	-1/4	-1/4	0	-1/4	-1/4	-1/4	-1/4
7	0	0	0	0	0	0	0	0	0	-1/2	-1/2	-1/2	-1/2	0	-1/2	-1/2	-1/2	-1/2
8	0	0	0	0	0	0	0	0	0	-3/4	-3/4	-3/4	-3/4	0	-3/4	-3/4	-3/4	-3/4
9	0	0	0	0	0	0	0	0	0	-1	-1	-1	-1	0	-1	-1	-1	-1

2.1.5 Експериментални резултати

За количествена оценка на качеството на предложените алгоритми (виж §2.1.2 и §2.1.3) за видеостабилизация по гладка траектория ние използваме мярката *Interframe Transformation Fidelity* (ITF), предложена в [147] и приложена също в [32], която модифицираме за граничните случаи:

$$\text{ITF} = \frac{1}{N-1} \sum_{k=1}^{N-1} \text{PSNR}((k), (k-1)), \quad (16)$$

където PSNR^{23} е отношението сигнал-шум “*peak signal-to-noise ratio*” между два последователни кадъра $(k-1)$ и (k) , а N е броят на кадрите във видеоклипа. Ние не използваме класическата PSNR оценка, а нейна модификация, предназначена именно за оценка на резултати от видеостабилизация, според специализираната литература [147]. Тя разглежда подобие между съответните пиксели от съседните кадри в рамките на предварително дефиниран праг за подобие между съответните пиксели от кадрите, което я прави подходяща за оценка на близостта между кадрите от един тематично свързан видеоклип (шот):

$$\text{PSNR}((k), (k-1)) = 10 \left| \log_{10} \left(\text{MSE}((k), (k-1)) \right) \right|, \quad (16a)$$

където:

$$\text{MSE}((k), (k-1)) = \frac{1}{x_{\text{size}} y_{\text{size}}} \left(1 + \sum_{y=0}^{y_{\text{size}}-1} \sum_{x=0}^{x_{\text{size}}-1} D(I_k(x, y), I_{k-1}(x, y)) \right). \quad (16b)$$

$I_k(x, y)$ е стойността за пиксел (x, y) в кадър (k) ,

$$D(a, b) = \begin{cases} 1, & \text{if } |a - b| > t \\ 0, & \text{otherwise} \end{cases}, 0 \leq a, b \leq 255, \quad (16b)$$

t е праг за компенсиране на ефекта от компресията и/или леки замазвания, както и за компенсиране на малки отмествания между стабилизираните кадри, $0 \leq t \leq 255$.

²³ Поради относителната простота на формулите, тук използваме утвърдените в Обработката на сигнали аббревиатури за обозначаване на съответните величини, въпреки че това е необичайно от математична гледна точка.

Нашата дефиниция на PSNR и MSE леко се различава от [147], с цел избягване на неудобството при граничния случай ($\max \{MSE((k), (k-1))\} = 0 \Leftrightarrow PSNR((k), (k-1)) \rightarrow \infty$), т.е. при пълно съвпадение на две изображения в рамките на избрания праг t .

Ние дефинираме експериментално стойността на прага $t = 32$, че да е подходящ за нашите тестови видеоклипове. Целта е ITF мярката да дава достатъчно различим резултат между оригиналното и стабилизираното видео според различните методи, предмет на настоящото изследване.

Експериментите са проведени върху 5 видеоклипа (от различни шотове) с размер 640×480 пиксела, с различна степен на трептене, оценена по скала от 1 (леко) до 5 (силно), Фиг. 30, виж и примерите в *Our_VStab/Trajectory/...*.



Фиг. 30. Тестови видеоклипове, подредени във възходящ ред по степента им на трептене.

Съответните ITF стойности за качеството на експериментите върху тестовите видеоклипове за стабилизация са дадени в Табл. 14. По-голяма ITF стойност означава по-добра стабилизация. Кратко описание на изследваните методи е представено в индексите към таблицата.

Таблица 14. ITF оценки за 5 видеоклипа: оригиналните и техните стабилизации по 6-те подхода.

метод видео	24 original	25 3x3/noAVG/I	26 3x3/AVG/I	27 3x3/noAVG/II	28 3x3/AVG/II	29 9x9/AVG/II	30 FAST
1	11.85	16.57	17.45	17.68	17.99	17.99	18.09
2	11.09	14.36	16.06	16.01	16.43	16.29	16.37
3	10.01	14.02	14.79	14.84	15.30	15.51	15.70
4	9.61	13.05	13.66	13.85	14.03	14.31	14.32
5	9.40	12.58	13.07	13.17	13.29	13.32	13.40
BVM* [ms]	-	2	2	2	2	15	(130)
Общо** [ms]	33	42	42	42	42	55	170
[fps]	30	23.8	23.8	23.8	23.8	18.2	5.88

Забележки (*) и (**) са дадени на следващата страница.

²⁴ (original): → в тази колона са дадени ITF стойностите на оригиналните (нестабилизирани) клипове;

²⁵ (3x3/noAVG/I): → използва 3x3 схема на разделяне, без IP нормализации и прилагане на LSM само веднъж;

²⁶ (3x3/AVG/I): → същото като горното, но с нормализация на акумулираните проекции (IP);

²⁷ (3x3/noAVG/II): → използва 3x3 схема на разделяне, без IP нормализации, но прилагане на LSM два пъти;

²⁸ (3x3/AVG/II): → същото като горното, но с нормализация на акумулираните проекции (IP);

²⁹ (9x9/AVG/II): → използва 9x9 схема на разделяне, с IP нормализации и втори пас на LSM (LSM-II);

³⁰ (FAST): → FAST методът [148] (имплементиран в Computer Vision Toolbox на Matlab) се използва за намиране на съответстващи характерни точки в два припокриващи се кадъра с цел изчисляване на параметрите на желаната (твърдотелна) трансформация (виж Приложение I). За подобряване на точността на намерените съответни точки ние използваме външен код за Матлаб, имплементиращ RANSAC алгоритъма [149].

***Заб.:** BVM е съкращение от “*Basic Vector Model*” или базов векторен модел, както е наречен нашият векторен модел за съответното разбиване на кадрите (“3×3” или “9×9”).

****Заб.:** Необходимо е да се отбележи, че за Матлаб реализацията на изследваните алгоритми, в таблицата са посочени „чистите“ времена за изпълнение на същинските алгоритми за видеостабилизация. Тук не са включени външните операции за алгоритмите, като: (i) *време за четене* ≈ 10 ms/frame; (ii) *обратна трансформация*³¹ ≈ 20 ms/frame; (iii) *запис на стабилизирани кадри* ≈ 10 ms/frame. Очаква се, че тези външни операции могат да се изпълнят много по-ефективно върху специализирания хардуер на съответното портативно филмиращо устройство.

Експериментите са проведени на десктоп компютър със следните характеристики: *процесор* Intel Core i5-2400 3.10GHz; *оперативна памет* 8GB DDR3; *видеокарта* NVIDIA Quadro 600 1GB; *твърд диск* WD Black SATA 7200rpm.

Въз основа експерименталните резултати могат да се направят следните изводи:

- ◆ ITF оценките отразяват степента на потрепване на камерата, адекватно на нашата визуална класификация (т.е., по-високо ITF $\Leftrightarrow$ по-добре стабилизирани видеоклипи);
- ◆ Най-лоши ITF резултати за видеоклиповете се получават при **(3×3/noAVG/I)** подхода, т.е. когато не се използва нито нормализиране на акумулираните проекции (integral projections, IP), нито втори пас на LSM (LSM-II);
- ◆ Стабилизацията по **(3×3/noAVG/II)** е почти толкова неприемлива, колкото и по **(3×3/AVG/I)** метода, но според ITF оценката, използването на LSM-II оказва по-добро влияние върху качеството на изходния резултат, отколкото IP нормализирането;
- ◆ Качеството на резултата (визуално и по ITF) значително се повишава при **(3×3/AVG/II)**, т.е. когато имаме комбинация между IP нормализиране и LSM-II;
- ◆ С разбиването на повече области, сравни **(9×9/AVG/II)** с **(3×3/AVG/II)**, стабилизацията се отразява по-добре на видеоклиповете с ниско качество (по-малко ITF), отколкото при тези с по-добро качество (по-високо ITF);
- ◆ Визуално най-приятни резултати се получават при **(FAST)** метода, което е особено забележимо при видеоклиповете с лошо качество (висока степен на трептене);
- ◆ Забелязва се, че резултатите от метода с разбиване на 9×9 области се доближават до тези на метода с характерни точки **(FAST)**. Т.е. равномерното разбиване на повече области се оказва доста обещаващо, като един алтернативен и бърз метод в сравнение с общоприетия и използван в общността по видеостабилизация метод с характерни точки.

2.1.6 Заключение и идеи за бъдеща работа

Предложен е нов ефективен и робастен метод за 2D видеостабилизация по гладка траектория, подходящ за работа в реално време на преносими устройства.

Тук съществено се емулира BSC чипа на Texas Instruments за намиране на „векторите на движение” чрез 1D корелации по SAD метода (суми от абсолютните разлики) между съответните 1D акумулиращи проекции – по хоризонтали и по вертикали. Предложено е подобрене (наричаме го нормализация на SAD хистограмите) при намиране-

³¹ При извършването на обратното възстановяване от оригиналните към стабилизирани кадри, последователността на трансформации е в обратен ред на изчислените, за да се избегне антикомутативността.

то на „векторите на движение“, което е съществено за точността на резултата от стабилизацията. Въпреки че „векторите на движението“ могат да се използват директно от трег стандарта, ние целим директен запис на стабилизираното видео в реално време върху даденото портативно устройство, а не стабилизация като постобработка върху предварително заснето видео. По отношение на бързодействието, точността, както и максималната дължина на намерените вектори на движение по нашия подход в сравнение с тези от трег видео, би било интересно да се проведат допълнителни изследвания, които оставяме за бъдеща работа.

Разработен е оригинален векторен модел “ $N \times N$ ” ($N \geq 3$, нечетно) за представяне на (случайни или преднамерени, но неизвестни) линейни трансформации между последователни кадри от видеопоток, разглеждайки съответните сцени като „несвидаеми, макар и с динамична структура“. От предварително оценените входни „вектори на движението“, моделът предоставя „векторни описатели“ за центровете на припокриващите се правоъгълници (N^2 на брой под-кадри), равномерно разбиващи всеки кадър, за всеки един разглеждан случай на трансформация между кадрите (транслация, ротация и др.). Моделът позволява 2D моделиране на нарастващи по сложност интерпретации на движението, като мащабиране, “rolling shutter” ефекти, перспективни изкривявания и др. Интерпретациите са интуитивни, несложни и водят до алгоритми, изпълними в реално време в компютърен ресурс от невисок клас, характерен за болшинството портативни устройства, като умни телефони, таблети и др.

Методът се предвижда като софтуерно инициализиращо средство за подобряване на точността на системата от инерциални сензори (жироскопи, акселерометри), вградени в използваното портативно устройство, отговорни за същинската хардуерна видеостабилизация. Извършени са множество експерименти за стабилизация по гладка траектория върху видеоклипове с нарастваща сложност на трептенето, в комбинация с различни схеми на разбиване (3×3 / 9×9), с и без нормализация на SAD хистограмите, тегловна оценка на уравненията от 1-я пас на LSM метода и потискане на грубите отклонения на следващ 2-ри пас на LSM. Проведените експерименти са обещаващи за бъдещото развитие на метода в насока за неговата евентуална комбинация с (известните) методи за видеостабилизация, използващи инерциални сензори.

2.2 Видеостабилизация за индустриални високоскоростни камери

В тази глава от дисертацията се разглежда проблемът за видеостабилизация в „точка“, който цели стабилизация на всички видео кадри спрямо един избран референтен кадър за произвеждане на ново видео, все едно заснето от статично фиксирана камера. Важността на задачата се състои в осигуряването на статичен фон във видео последователността, необходим за провеждане на коректни измервания в кадрите при изследването на определени характеристики на динамичните обекти във видеоклипа.

2.2.1 Въведение – необходимост и мотивация

В настоящото изследване се предлага разработване и тестване на софтуерни алгоритми за видеостабилизация на клипове от високоскоростната камера NAC Memrecam HX-6³², закупена по проекта AComIn на ИИКТ-БАН. Необходимостта от видеостабилизация при подобен вид камери възниква, когато е невъзможно да се избегне динамиката на средата, т.е. камерата не може да бъде статично фиксирана. От друга страна, стабилният фон във видеоклипа е съществено изискване за прецизното измерване/изчисление на определени характеристики и свойства на изследваните динамични обекти и/или процеси (разстояния, скорости, ускорения, посока и вид на движението, еластичност, якост и др.), заснети с високоскоростната камера. Но в много случаи не може да се избегне трептенето във видеото, причинено от нестабилности на фундамента за статива на камерата, или пряко от ръката, държаща камерата, например: (1) заснемане на бързи процеси във високо вибрираща среда (в заводски условия); (2) снимане на контролирани взривове (масивни сътресения в близост до камерата); (3) снимане на ръка от превозно средство; (4) снимане от ръка „на момента”, в „полеви условия” и др.

Поради спецификата на камерата – скорости на заснемане от 50 до 370330 fps, видеостабилизацията се провежда в off-line режим, т.е. бързодействие на алгоритмите по принцип тук не се търси. Най-важна е „перфектната” стабилизация, особено в случаите на така наречената „стабилизация в точка”, т.е. изискване за пълна неподвижност на сцената в стабилизирания видеоклип. Тук именно възниква предизвикателството за елиминиране или поне редуциране на ефекта „пъзене на сцената” (виж. §2.2.3 и §2.2.4), което е характерен проблем и при болшинството известни алгоритми по въпроса (например *Warp Stabilizer* на *Adobe After Effects CS6*). Но при тези алгоритми доминиращото изискване е стабилизация по гладка траектория на движение на камерата и/или на обектите в заснеманата сцена, при което ефектът на пъзене не се забелязва от потребителя.

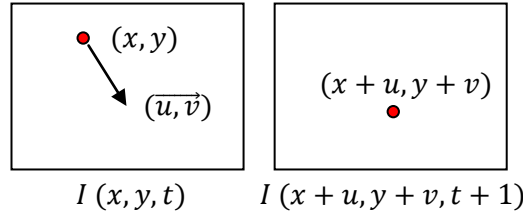
2.2.2 Методът на Оптичния поток

В този раздел ще разгледаме същността на метода на оптичния поток, чието решение е напълно съвместимо като вход за предложените векторни модели (CCD или CMOS) за намиране на съответната трансформация между два (последователни или достатъчно припокриващи се) кадъра, виж. §2.1.3 и §2.1.4. Неговото предимство спрямо SAD метода за намиране на векторите на движението в конкретна схема на разбиване (“3×3” или “9×9”) се състои в значително по-високата му прецизност, както и произвеждане на резултат, т.е. вектор на движение за всеки пиксел с подпикселна точност.

Задачата на метода се състои в изчисляването на 2D векторно поле на движението на всеки пиксел между двойка кадри чрез пространствено-времевите вариации на яркостта (виж Фиг. 31). За определяне на вектора на отместване (u , v) за всяка точка от кадрите трябва да се вземат предвид следните ключови допускания за правилната работа на метода:

³² <http://www.nacinc.com/products/memrecam-high-speed-digital-cameras/hx-6/>

1. *константна яркост*: проекцията на една 3D точка изглежда еднакво в кадрите;
2. *времева устойчивост*: движението е плавно с малки отмествания между кадрите;
3. *пространствена кохерентност*: движението на дадена точка е същото както на нейните съседи.



Фиг. 31. Идеята на метода на оптичния поток се състои в изчисляване на 2D векторното поле (отмествания) за всички пиксели между двойка припокриващи се изображения.

Първото допускане води до равенството:

$$I(x + u, y + v, t + 1) = I(x, y, t) \quad (17)$$

При малки отмествания можем да апроксимираме линейно функцията $I(x + u, y + v, t + 1)$ в малка околност на (x, y, t) чрез ред на Тейлор:

$$I(x + u, y + v, t + 1) \approx I(x, y, t) + \frac{\partial I}{\partial x}(x, y, t) \cdot u + \frac{\partial I}{\partial y}(x, y, t) \cdot v + \frac{\partial I}{\partial t}(x, y, t) \quad (17a)$$

Замествайки (17a) в (17), получаваме основното уравнение на оптичния поток:

$$\begin{aligned}
 I(x, y, t) + I_x \cdot u + I_y \cdot v + I_t - I(x, y, t) &= 0 \Leftrightarrow I_x \cdot u + I_y \cdot v + I_t = 0 \\
 \Leftrightarrow \nabla I [\mathbf{u} \ \mathbf{v}]^T + I_t &= 0,
 \end{aligned} \quad (18)$$

където: I_x и I_y са пространствените производни в момента от време t ; I_t е времевата производна в (x, y, t) , т.е. $I_t = I(x, y, t + 1) - I(x, y, t)$; и $\nabla I = [I_x \ I_y]$ е векторът на градиента в точката (x, y, t) .

За еднозначното решаване на основното уравнение (18) са необходими поне 2 линейно независими уравнения. Така, базирайки се на допускането за пространствена кохерентност, Lukas & Kanade [150], предлагат разглеждане на съседство от пиксели в област 5×5 :

$$\begin{bmatrix} I_x(p_1) & I_y(p_1) \\ I_x(p_2) & I_y(p_2) \\ \vdots & \vdots \\ I_x(p_{25}) & I_y(p_{25}) \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = - \begin{bmatrix} I_t(p_1) \\ I_t(p_2) \\ \vdots \\ I_t(p_{25}) \end{bmatrix} \Rightarrow \begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_y I_x & \sum I_y I_y \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = - \begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix} \quad (19)$$

Така, решението на преопределената система $A\vec{x} = \vec{b}$ в LSM-смисъл може да се намери чрез решаване на матричното уравнение $(A^T A)\vec{x} = A^T \vec{b} \Rightarrow \vec{x} = (A^T A)^{-1} A^T \vec{b}$.

Това матрично уравнение е било предложено за първи път от Lukas & Kanade, чието решение е възможно, когато: (1) $A^T A$ е обратима матрица; (2) собствените стойности λ_1 и λ_2 на $A^T A$ не трябва да бъдат твърде малки, $\det(A^T A) = \lambda_1 \lambda_2 \neq 0$.

Методът на Lukas & Kanade в този си вид е приложим само за малки междукadroви отмествания (в рамките на 1 пиксел). За преодоляване на това ограничение се използва итеративен подход за изчисляване на оптичния поток в Гаусова пирамида от изображения на различен мащаб [151].

2.2.3 Комбиниран подход “3×3OF9×9” за видеостабилизация в точка

За решаването на поставената задача за стабилизация в „точка“ никой от трите подхода (“3×3”, “9×9” или „оптичния поток“) не би могъл да се справи самостоятелно със задоволителна точност. Могат да се отбележат следните техни недостатъци:

- методът “3×3” е най-чувствителен към грешки, поради използвания малък брой вектори на движението, чрез които се определят търсената трансляция и ротация.
- методът “9×9” е сравнително устойчив към погрешно определени вектори на движението по SAD метода, поради по-големия брой области, разпределени равномерно в кадрите. Но както и при “3×3” метода, той страда от своето целочислено определяне на векторите на движение (при чиста трансляция) по метода SAD, което води до трептене в рамките на 1 пиксел при обратната трансформация за възстановяване (това е особено видно при стабилизиране в точка, а не по траектория).
- методът на „оптичния поток“ в своя най-олекотен (бърз и непрецизен) вариант [152], работи добре само за сравнително малки отмествания между кадрите. Чувствителен е към неравномерна промяна на осветеността в изображенията (сенки), както и към рязката промяна на границата в дълбочината между обектите.

За преодоляване на недостатъците на трите метода поотделно, разглеждайки ги независимо, е предложен оригинален комбиниран подход, наречен “3×3OF9×9”, за видеостабилизация в „точка“. Той се състои от два последователни етапа на стабилизация.

Първи етап:

1. извършва се груба стабилизация спрямо избран референтен кадър чрез “3×3” метода (виж §2.1.2), който е в състояние да компенсира бързо и ефективно най-големите отмествания между кадрите;
2. върху относително стабилизиранияте двойки кадри се използва известният метод на „оптичния поток“, който определя векторите на движението с подпикселна точност за всеки пиксел между сравняваната двойка кадри;
3. прилага се “9×9” методът (виж §2.1.3) върху съответните 9×9 центъра на съответните области от векторното поле на „оптичния поток“ за прецизирането на грубата стабилизация от “3×3” метода.

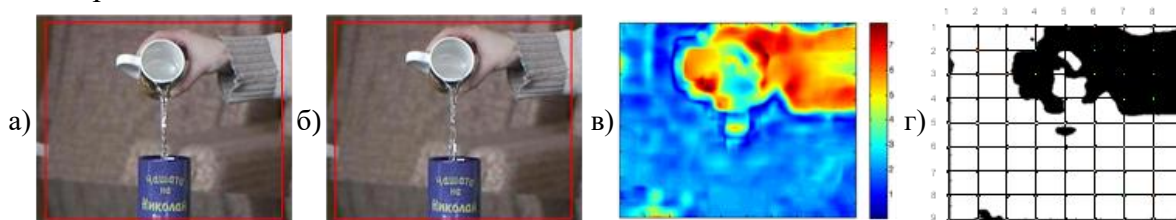
Резултатът от първия етап на стабилизацията в точка според “3×3OF9×9” метода води до типичния ефект „пълзене на сцената“. Този ефект се дължи на натрупване във времето на подпикселна грешка в изчислените трансформации между двойка кадри, акумулирани в глобална траектория с начало избран референтен кадър. Допълнително, наличието на масивен динамичен обект, както и на обекти на различна дълбочина води до големи грешки в изчислените параметри на движението, които в текущите експерименти са преобладаващи трансляции.

Впрочем, този първи етап на стабилизация по метода “3×3OF9×9” може да бъде използван самостоятелно за прецизна 2D „твърдотелна“ стабилизация (транслация + ротация) по гладка траектория на цената на допълнително забавяне.

Втори етап:

За значителното редуциране на акумулираната грешка в глобалната референтна траектория, която често води до ефекта „пълзене на сцената“, се прилага следната стратегия:

1. Между всеки два последователни кадъра, в резултат от *първия етап* на стабилизацията се изчислява „карта“ с големините на векторите на оптичния поток. Върху тази карта се прилага бинарна маска (изчислена по метода на Отсу) с цел изолиране на влиянието на динамичните обекти (виж Фиг. 32) чрез предварителното изключване на определени вектори на движението, участващи в схемата на “9×9” метода, и по този начин подобряване на точността на стабилизацията на статичния фон;
2. През достатъчно голям брой кадри (например интервал от 100) се изчисляват векторите на движение (по “9×9” схемата), така че грешката в акумулираната траектория да бъде равномерно разпределена между кадрите в интервал. По този начин грешката в глобалната траектория се коригира всеки път в избрания интервал от кадри.



Фиг. 32. Стратегия за елиминиране на подвижните обекти в сцената: а) първи контролен кадър; б) втори контролен кадър; в) карта на големините на векторите от оптичния поток; г) бинаризирана карта по метода на Отсу [153], приложена към “9×9” векторите на движение за подобряване на точността в стабилизацията на статичния фон.

2.2.4 Експериментални резултати

Извършените експерименти за стабилизация в статичен фон са върху два типа видео:

- Видеоклип, свален от интернет страницата на производителя на камерата NAC Memecam NX-6, съдържащ запис на контролиран взрив, където е налице огромен динамичен обект (облаци с пушек от взрива). Видеоето е с разделителна способност 1280×720 пиксела и служебна скорост от 25 кадъра в секунда.
- Видеоклипове, заснети от ръка с нарочно (симулирано) трепкане, съдържащи запис на вода/капки в резултат от преливането от чаша в чаша. Видеоклиповете са с разделителна способност 640×480 пиксела и скорост от 50, 100 и 500 кадъра в секунда.

И при двата вида експерименти първоначално е приложен само първият етап на “3×3OF9×9” метода за нагледна демонстрация на ефекта „пълзене на сцената“.

1. Експеримент с контролиран взрив

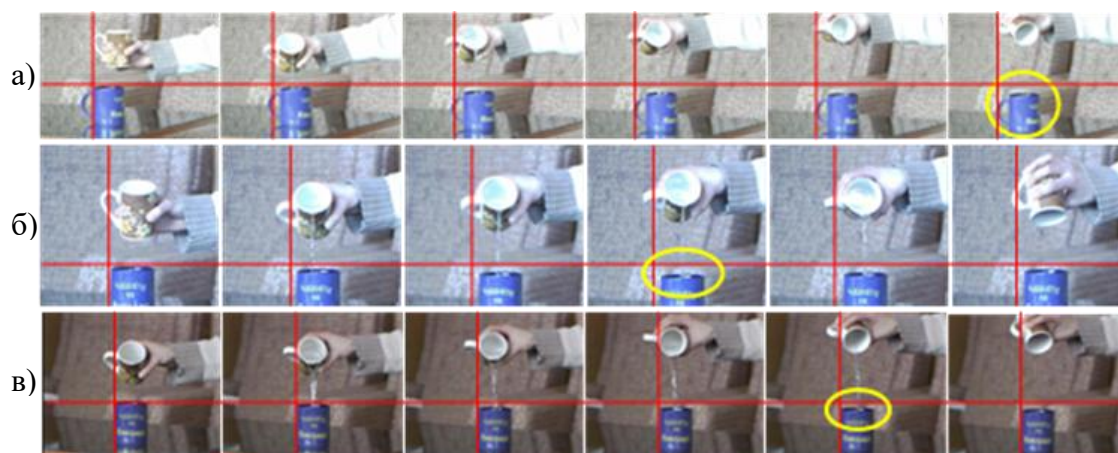
Наличието на преобладаващ подвижен обект в сцената „увлича“ алгоритъма, който е с фиксирана схема на разбиване $3 \times 3 / 9 \times 9$, по посока на движението на обекта, което се изразява в характерното „пъзене на сцената“, т.е. фонът се движи в обратна посока на движението на обекта (Фиг. 33б). По тази причина стабилизираме по траекторията на масивния обект (Фиг. 33в), при което ефектът на пъзене е игнориран и е постигнат желаният статичен фон, виж и примера *Our_VStab/Point/Explosion.mp4*.



Фиг. 33. Стабилизация на видеоклип на контролиран взрив: а) първи кадър; б) последен кадър след стабилизация в точка (пъзене на сцената); в) последен кадър след стабилизацията по траектория (фонът е статичен).

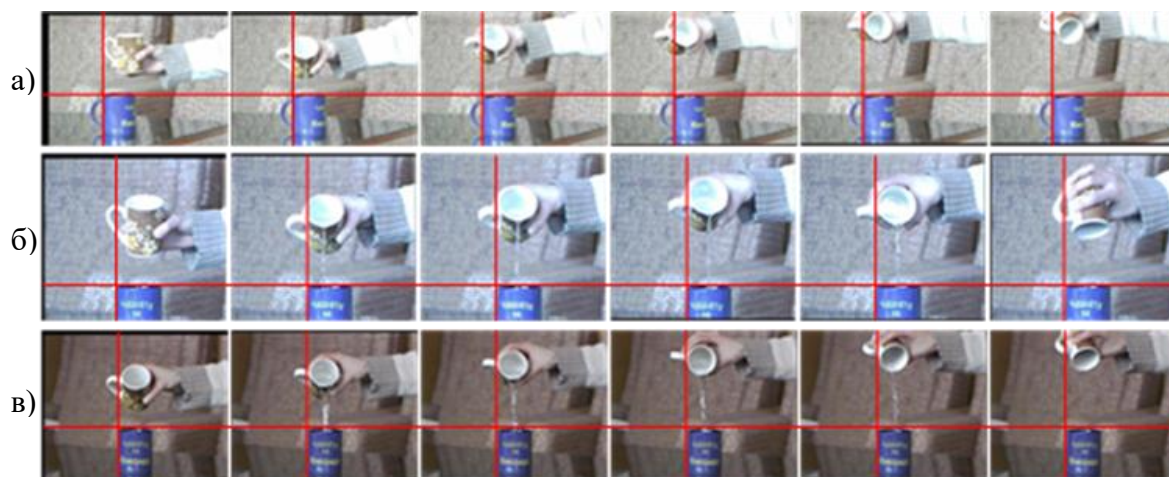
2. Експеримент с преливане на вода

При този експеримент, и за трите видео клипа с различни скорости и съответно различни амплитуди на трептене, също се констатира „пъзене на сцената“ (виж Фиг. 34), но в много по-малки граници в сравнение с контролирания взрив. Това позволява да се приложи нашата стратегия за редуциране на грешката от „пъзене на сцената“ (Фиг. 35) чрез изолиране (сегментиране) на подвижния обект, виж и примерите в *Our_VStab/Point/Water/1st_pass/...*



Фиг. 34. „Стабилизирани кадри“ в „точка“ на 1-ви етап на стабилизация: а) 50 fps; б) 100 fps и в) 500 fps.

Грешките от „пъзненето на сцената“ на 1-ви етап на стабилизацията (Фиг. 34) значително са редуцирани на 2-рия етап, като те сега са $<1\%$ от размера на кадрите (Табл. 15), виж и примерите в *Our_VStab/Point/Water/2nd_pass/...*



Фиг. 35. Стабилизирани кадри в „точка“ на 2-ри етап на стабилизацията: а) 50 fps; б) 100 fps и в) 500 fps.

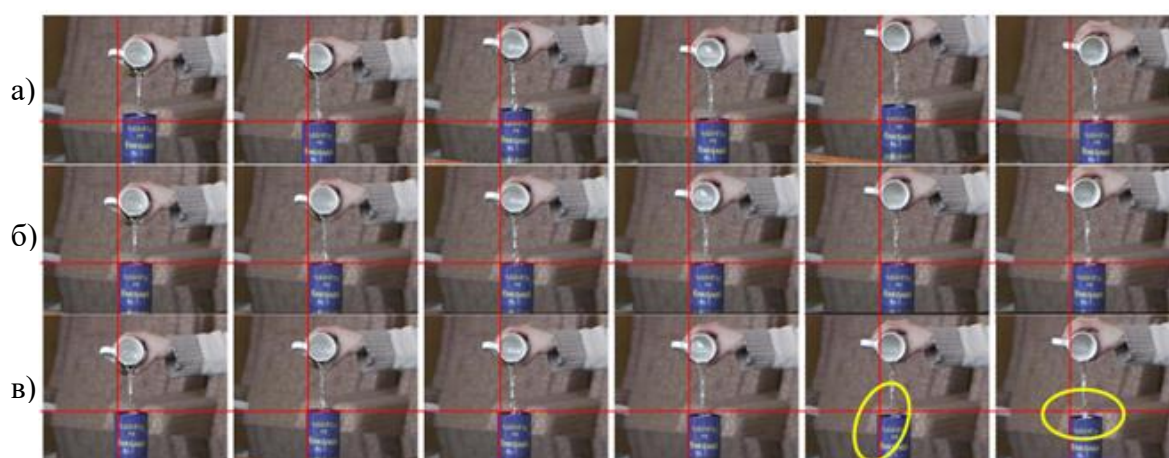
В Таблица 15 са дадени максималните грешки по Ох и Оу посоките за съответните скорости на видеоклиповете с преливане на вода за двата етапа на стабилизацията.

Таблица 15. Грешки от стабилизацията в статичен фон след 1-ви и 2-ри етап на стабилизацията.

Резолюция 640×480	Видео 1 (50 fps)		Видео 2 (100 fps)		Видео 3 (500 fps)	
	1-етап	2-ри етап	1-етап	2-ри етап	1-етап	2-ри етап
max Ох грешка	3.44 %	0.16 %	1.09 %	0.47 %	1.09 %	0.16 %
max Оу грешка	5.00 %	1.04 %	3.75 %	0.63 %	4.17 %	0.42 %

3. Сравнение с резултатите от Adobe After Effects CS6

На Фиг. 36 е показано визуално сравнение (нашата стабилизация срещу тази на ААЕ) за стабилизация в „точка“ за видеоклип с преливане на вода с 500 fps. На този експеримент ефектът от бавното „плуване“ на сцената е забележим при резултата от ААЕ, въпреки че в софтуера им явно е указана опцията за стабилизация в статична камера (фон), виж примерите в *Our_VStab/Point/Water/AAE/...*.



Фиг. 36. Нашият резултат за стабилизация в статичен фон в сравнение с този от Adobe After Effects CS6: а) оригинални видео кадри; б) нашият резултат чрез “3×3OF9×9” метода; в) ААЕ стабилизацията.

2.2.5 Заключение и идеи за бъдеща работа

Предложен е комбиниран метод “3×3OF9×9”, съчетаващ базовия “N×N” метод с известния метод на „оптичния поток“, с детекция и отстраняване на подвижните обекти в сцената за неутрализиране на ефекта „пъзене на сцената“. Методът се отличава с достатъчно бързодействие и прецизност на стабилизацията. Той е предназначен за видео експерименти (например с високоскоростни камери), където обикновено се цели неподвижност на фона. Постигнатите резултати се оценяват като много добри, съблюдавайки изискването за статичен фон на сцената или за статичен преобладаващ обект в сцената, което позволява провеждане на коректни измервания във видеокадрите, игнорирайки в значима степен динамиката на средата, която предизвиква трептенето във видеото. Характерният ефект на „пъзене на сцената“, който е налице дори и при комерсиалните софтуерни решения, е редуциран до под 1% от размера (640×480 пиксела) на кадрите. Чистото време за изпълнение на алгоритъма е около 30 fps. Бъдещите усилия ще бъдат насочени към разработване на по-ефективна сегментация на големи подвижни обекти в контекста на предложения метод.

2.3 Откриване на предварително известни рекламни клипове в TV видеопоток

Детекцията и разпознаването на рекламни блокове в TV видеопоток има ключово значение при анализа на медийното присъствие на редица компании, организации и агенции, осигуряващо им значителни преимущества в среда на висока конкуренция. Компаниите плащат огромни суми за излъчване на техните реклами в определени TV канали в подходящ времеви интервал с цел привличане на внимание към определени продукти и увеличаване на продажбите им. От една страна, рекламирането е основно средство за издръжката на болшинството телевизии, а от друга – силен маркетингов инструмент за платежоспособните компании. Компаниите-рекламодатели обикновено плащат на други компании да следят дали техните реклами действително са излъчвани според условията на договора с дадената телевизия. Текущо, тази задача се осъществява от хора (оператори), които наблюдават различни TV канали в съответните предварително фиксирани времеви интервали за излъчване на реклами. За намаляване на разходите по заплащане на такива служители, както и за осигуряване на по-стриктен и достоверен контрол върху качеството на извършваната работа, компаниите определено биха желали да автоматизират тази задача чрез използването на компютризирани системи за детекция и разпознаване на реклами. Такива системи биха записвали прецизно часа и датата на излъчване на следената реклама, както и идентификатора на съответния канал. Освен тази информация, компаниите също биха се възползвали и от възможността за телевизионното наблюдение на маркетинговата дейност на своите конкуренти.

Освен комерсиално погледнато, TV рекламите представляват интерес и от психологическа гледна точка – от една страна, за изследване на въздействието им върху покупателните навици на населението, а от друга, за предпазване на определена група хора (деца и др.) от тяхното влияние.

Известно е, че рекламите се излъчват в рекламни блокове по време или в промежудутъка между различни TV предавания (виж. Фиг. 37).



Фиг. 37. TV видео, съдържащо множество рекламни блокове, в рамките на новинарска емисия. *Заимствано от Google, като идея.*

Важно е да се отбележи, че за да бъде допусната дадена реклама за излъчване в ефир, е необходимо спазването на определени изисквания – освен към смисъла на нейното съдържание, така и към нейната видеоструктура от техническа гледна точка. Тези отличителни особености играят съществена роля в множество методи за детекция и разпознаване на рекламно съдържание и могат да се класифицират като директно и индиректно измерими [154, 155]. Към първата група принадлежат следните измерими особености:

- повторяема видео последователност;
- ограничен времеви интервал – най-често между 10-60 сек. за единична реклама и до 6-30 мин. за рекламен блок;
- липса на корелация между съседни сцени;
- увеличаване на силата на звука по време на рекламите и намаляване след това.

Към групата на индиректно измеримите характеристики на рекламите се отнасят:

- висока степен на динамика и сензация на съдържанието, т.е. висока честота на преходите между сцените и драстична смяна в цветовете;
- премахване на емблемата на TV канала по време на рекламния блок;
- последната сцена от дадена реклама е често статично изображение, представящо продукт и/или компания;
- наличие на текст за име на продукт или компания, представляващ полезна семантична информация, която може да бъде идентифицирана и оценена [156].

2.3.1 Формулировка на задачата

Вход:

1. Множество от 30 мин. записи в MPEG-4 формат (25 fps, с размер 448×336 пиксела) на TV програма на даден канал в рамките на 24 часа.
2. Множество от рекламни видеоклипове (около 30 сек. всеки) в същия формат.

Изход:

Времето локализиране на даден рекламен видео шаблон в TV поток.

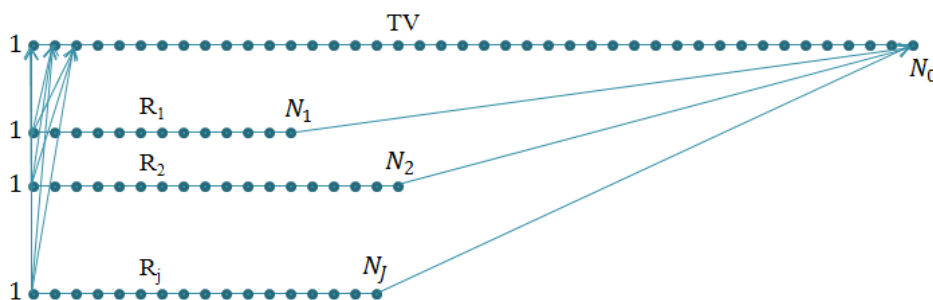
Ограничения:

Достатъчно бързодействие (поне в реално време) на алгоритъма, съобразено с наличните компютърни мощности.

2.3.2 Дискутирани сценарии за решаване на задачата

1) Първи сценарий - директно сравнение кадър по кадър

Най-лесният и най-бавен начин за проверка дали даден кадър от едно шаблонно видео (реклама) се съдържа в друг видеопоток, може да бъде извършен чрез директно сравнение на пикселите от съответните кадри на двата клипа. Така, максималният брой сравнения ще бъде $N_0(\sum_{j=1}^J N_j)$, където N_0 е броят на кадрите в TV видеопотока; N_j е броят на кадрите в j -тата реклама; J е броят на търсените рекламни видеоклипове. На Фиг. 38 схематично е показана сложността на алгоритъма, който от гледна точка на бързодействието не е подходящ за решаването на поставената задача.



Фиг. 38. Схематично представяне на сложността на директното покадрово сравнение.

2) Втори сценарий – ускорение чрез DCT коефициентите на JPEG кадрите

Използвайки коефициентите на дискретното косинус преобразование (DCT) за област 8×8 от JPEG кадрите, може да се постигне известно ускорение при сравнението на кадрите. Това е възможно поради факта, че много от DCT коефициентите в 8×8 областта са ~ равни на нула и по този начин ще сравняваме по-малко числа отколкото пиксели в същата област. Но въпреки това, максималният брой сравнения е от същия порядък, както при сценарий 1. Освен това, използването на DCT коефициентите би имало смисъл само ако ги извлечем директно от JPEG кадрите, а не да ги изчисляваме отново.

3) Трети сценарий – търсене на съответствия между сцените в рекламите и основния телевизионен поток според дължината (в брой кадри)

За удовлетворяване на изискването за достатъчно бързодействие на алгоритъма, *тук предлагаме* сравнението „кадър-по-кадър“ да се замени със сравнение „сцена-по-сцена“. Този алгоритъм има предимство в случая, когато броят на рекламните видеоклипове е голям.

Същността на алгоритъма се състои в представянето на видеозаписите като редица от цели числа, които характеризират продължителността (в брой кадри) на отделните сцени в тях. Обикновено понятието „сцена“ се разглежда като тематично цялостен фрагмент от видеото. За нашите цели, сцена е всяка част от видеото, където финално имаме рязка промяна между два (или няколко) последователни кадъра, като указание за следваща сцена. Като изключение, първата сцена стартира с началото на видеото, а последната приключва с края му. По този начин първоначално можем да сравняваме сцените по дължина (продължителност в брой кадри) и само за „подозрителните“ сцени (т.е. с близки дължини) извършваме покадрови сравнения от типа „едно-към-много“. Предимството на тази стратегия е в много по-малкия брой покадрови сравнения, отколкото при предходните сценарии.

Проблемът, който възниква тук, е определянето на оптималния брой на сцените, за да няма твърде много фалшиви или пропуснати сцени. Т.е. не е необходимо да бъдат открити точно действителните (отговарящи на визуалното възприятие или тематичните) граници на сцените – важно е алгоритъмът да бъде приложен по един и същ начин както за TV потока, така и за търсените в него рекламни фрагменти.

Един възможен подход за намиране на точните граници на сцените може да се базира на свойството инвариантност по трансляция на трансформацията на Фурие (FT). Това означава, че модулът (мощността) на FT спектъра ще бъде един и същ за две транслирани изображения. При дадена скорост на видеото, например, стандартната 25 fps, можем да разглеждаме кадрите от една и съща сцена като доминиращ общ фон, който е статичен или се променя много плавно от кадър към кадър, в резултат на шум или от произволните, но малки движения на обектите (от кадър към кадър) в сцената. Тогава, разликата между мощностните спектри на тези кадри ще бъде достатъчно малка, приблизително нула, което можем да използваме като критерий за принадлежност на кадрите към една и съща сцена. Този подход все още не е практически приложен от нас за детекция на сцени, въпреки че изглежда много обещаващ за желаното качество на резултата. Подходът може да се разглежда и в аналогия с корелационния подход (споменат в Глава 1, §1.1.4 за целите на видеостабилизацията), където за бързодействие може да работим в спектралното пространство по Фурие, т.е. с производението на кадровите спектри, а не просто с разликата между мощностите им.

2.3.3 Описание на предложения алгоритъм по третия сценарий

Понеже се стремим към работа в реално време (или по-бързо), за детекцията на сцени ние предлагаме един много прост и бърз подход, основан на разликата между два последователни кадъра в рамките на предварително избран праг. Този подход осигурява

достатъчно надеждна детекция на цели реклами и/или произволни части от тях, в даден TV видеопоток. Нашият алгоритъм се състои от 3 стъпки:

1. Конвертиране на видеозаписите (TV и реклами) в така наречените тук „диференциални представяния“.
2. Разделяне на видеото на последователност от сцени и представянето му като низ от цели числа, всяко едно даващо продължителността (в брой кадри) на всяка сцена.
3. Сравнение на целочислени редици (вместо последователности от JPEG кадри) между TV потока и потенциалните кандидати от сцени от множеството на рекламните видеоклипове.

Ще опишем тези стъпки по-подробно в следващите 3 точки C1, C2 и C3.

C1. Изчисляване на диференциалните представяния (за TV потока и рекламите)

C1.1 Диференциално представяне за TV потока:

$$v(n) = \frac{1}{X \times Y} \sum_{\forall(x,y)} f(I_n, I_{n+1}), \quad (20)$$

където:

n е последователният номер на (даден) кадър от TV потока, $n = 1, \dots, (N - 1)$;

$I_n(x, y)$ е интензитетът на пиксел (x, y) от n -тия кадър, $I_n(x, y) = 0, \dots, 255$;

$X \times Y$ е размерът на всеки един кадър, в пиксели;

$f(I_n, I_{n+1})$ е функция с Булев изход $\{0,1\}$, дефинирана върху разликата между два последователни кадъра $I_n(x, y)$ и $I_{n+1}(x, y)$, $(x, y) \in X \times Y$. Най-просто е да се използват абсолютните попикселеви разлики на интензитетите (т.е. диференциалният, разликният видеопоток), което ще означим междинно като:

$$f_n(x, y) = |I_{n+1}(x, y) - I_n(x, y)| \quad (20a)$$

За по-реалистично откриване на сцените при наличие на сравнително голям подвижен обект или при малки разлики в интензитета на пикселите на основния TV поток и/или на рекламите в него (поради различия в компресията), тук прилагаме така наречения „праг за подобие между кадрите“ (PST) [147, 157]. Така, окончателно за функцията $f(I_n, I_{n+1})$ получаваме:

$$f(I_n, I_{n+1}) = \begin{cases} 1, & f_n(x, y) > \text{PST} \in \{0,1, \dots, 255\} \\ 0, & \text{otherwise} \end{cases} \quad (20b)$$

а за диференциалното представяне v имаме: $v(n) \in [0,1]$, $n = 1, \dots, (N - 1)$.

С1.2 Диференциални представления на рекламни видеоклипове

Дефинираме го по аналогия с това на основния TV видеоклип:

$$r^j(l) = \frac{1}{X \times Y} \sum_{\forall(x,y)} |R_{l+1}^j(x,y) - R_l^j(x,y)|, \quad (21)$$

където:

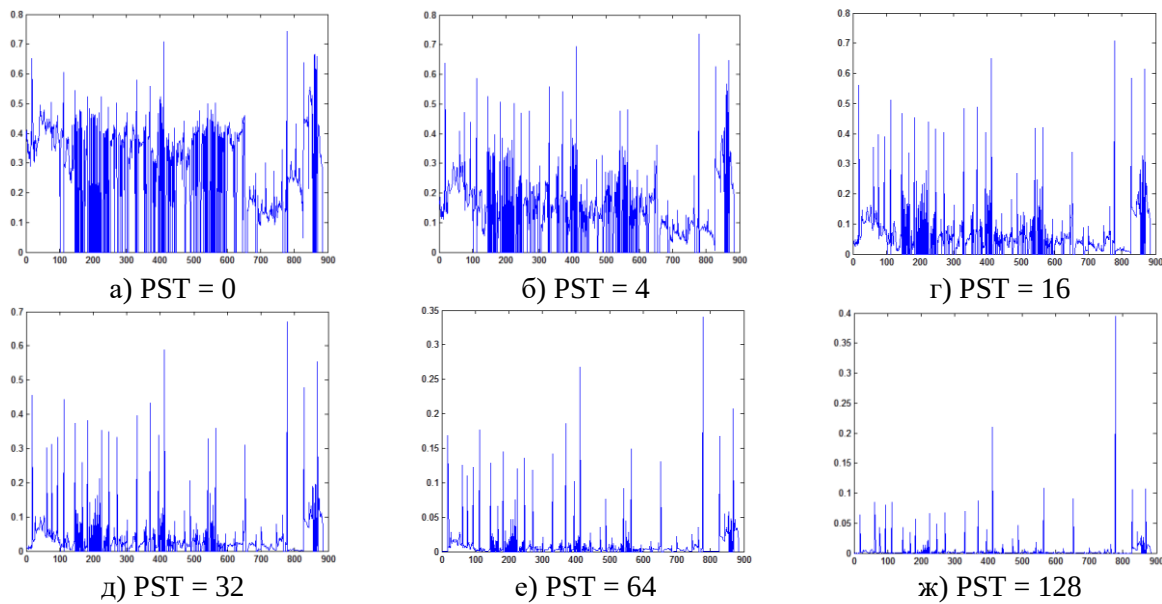
- $l = 1, \dots, (N_j - 1)$ са последователният номер на (даден) кадър от j -тата реклама;
- $j = 1, \dots, J$; J е общият брой на рекламите (които се търсят в основния видеоклип);
- а $R_l^j(x,y)$ е интензитетът в l -тия кадър от j -тата реклама.

След изчисляването по (20) и (21), за диференциалния TV видеопоток и за рекламните видеоклипове, получаваме съответните 1D последователности:

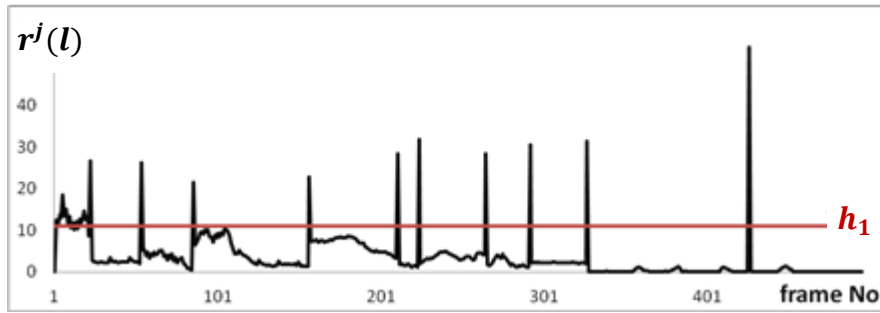
- за TV видеопотока: $v(1), \dots, v(N - 1)$;
- за j -тата реклама, съответно: $r^j(1), \dots, r^j(N_j - 1)$, $j = 1, \dots, J$.

Подобието в дефинициите (20) и (21), и използването на един и същи PST-праг там гарантира определянето на сцените по еднакъв начин, както за TV потока, така и за рекламните видеоклипове.

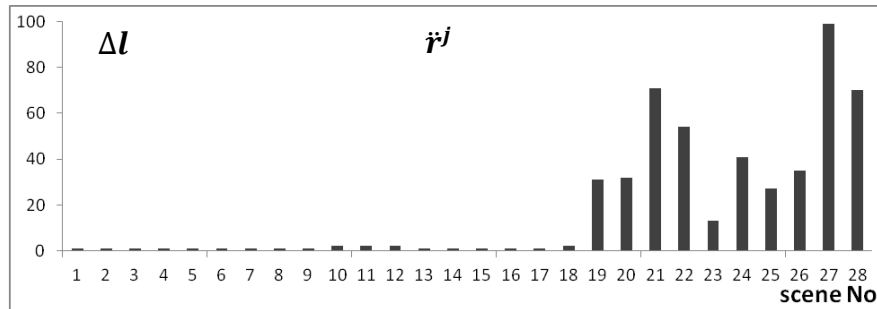
На Фиг. 39 е дадено диференциалното представяне за даден рекламен видеоклип, с дължина 900 кадъра, за различни стойности на PST-прага. Както се вижда, когато PST-прагът е много малък, различимостта между сцените е ниска, защото значителен шум от статичния фон се смесва с плавното движение на обекта. От друга страна, когато прагът е твърде висок, информацията от подвижния обект може да се загуби, понеже алгоритъмът го интерпретира като фон. Целта е да се намери такъв PST-праг, при който само голяма разлика в болшинството пиксели между двата кадъра да се интерпретира като нова сцена.



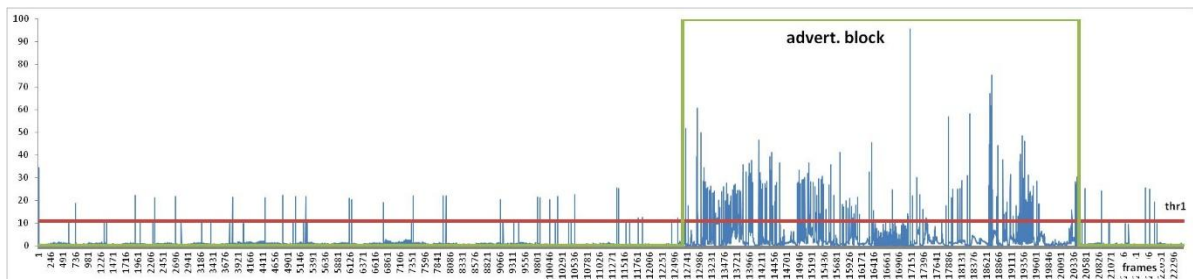
Фиг. 39. Диференциалното представяне на рекламен клип за различни стойности на PST-прага.



Фиг. 41. Детекция на сцени в рекламен видеоклип (~20 сек.)

Фиг. 42. Целочислена редица $\tilde{r}^j = (\Delta l_1^j, \Delta l_2^j, \dots, \Delta l_{s_j}^j)$ от дължините на сцените Δl за горния рекламен видеоклип.

На Фиг. 43 е представен резултат за детекция на сцени в 15-минутен TV клип на „толк-шоу“ предаване. Особено забележимо е увеличаването на честотата на сцените в рекламния блок.



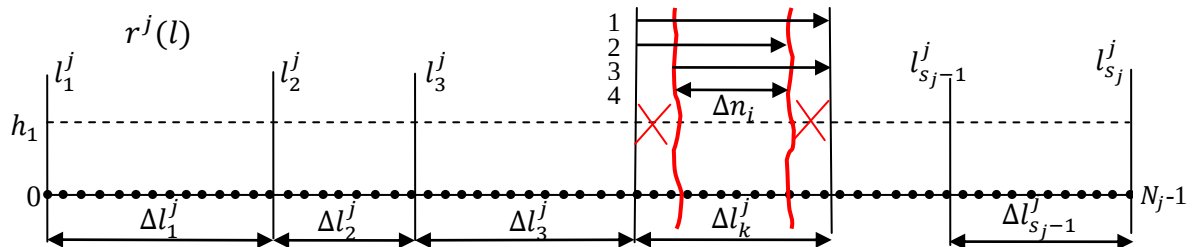
Фиг. 43. Детекция на сцени за ~15-минутен TV поток, съдържащ само един рекламен блок.

Въпросът за намирането на алгоритъм за оптималната стойност на PST-прага (постоянен или променлив) за детекция на сцените остава актуален. Съществено за алгоритъма за разбиването по сцени е детекцията на сцените да става по еднакъв начин (с един и същ праг) в TV потока и в рекламните клипове. По този начин успяваме да заменим обемистите сравнения от типа „кадър-по-кадър“ с целочислени сравнения от типа „дължина-по-дължина“ на сцени. Очевидно, това редуциране на сравняваните обеми от информация (редици от изображения $\rightarrow$ дължини в брой изображения) се очаква драстично да намали сложността на поставената задача, като я привежда към по-проста алгебрична такава.

С3. Детекция на цялостни и/или частични сцени от реклами в TV поток

Предвид спецификата на поставената на задачата, можем да дефинираме следното:

Основно правило: Една сцена от TV видеоклип може да бъде част от сцена от дадена реклама (т.е. да съвпада или да е по-къса), но сцена от рекламен видеоклип може само точно да съвпада със съответната сцена от TV видеоклипа (виж и Фиг. 44).



Фиг. 44. Илюстрация на основното правило за включване на сцена от TV поток в рекламата.

Това правило се основава на следните (практични) съображения:

- Рекламните клипове, предоставени от клиента-рекламодател, могат да се смятат за постоянни (непроменливи) по дължина и фиксирани (като тематична последователност от кадри); в противен случай бихме имали нов рекламен видеоклип.
- Имплементирането на рекламните видеоклипове в даден TV поток може да е обект на редакторска намеса, например, отсичания от началото и/или края на рекламната, и/или съкращения на (тематично) цели сцени от нея, за вместване в даден наличен интервал от телевизионното време. Тези отсичания могат да се ограничат до нашата трактовка на понятието „сцена“. Така, колкото по-дълги са целочислените редици на рекламните, толкова по-голяма е и вероятността за разпознаване на рекламната в TV потока, макар и да е частично съкратена.

От тези съображения произлиза още едно правило, което е тясно свързано с алгоритъма на търсене, а именно:

Второ правило: Началното търсене (ускореното, според съответните целочислени редици от дължини на сцени) провеждаме в посока от TV клипа към рекламните, иначе (в обратната посока) рискуваме да пропуснем съответствия на сцени.

Нашият алгоритъм покрива следните 4 случая на включване на реклама в TV поток:

1. TV сцената Δn_i точно съвпада със сцената Δl_k^j от j -тата реклама;
2. Само началото на TV сцената Δn_i съвпада с началото на Δl_k^j сцената;
3. Само края на TV сцената Δn_i съвпада с края на Δl_k^j сцената;
4. Началото и края на TV сцената Δn_i са вътрешни за Δl_k^j сцената.

Това дърво на решения изчерпва най-често срещаните ситуации в телевизионното рекламодаване (според информация от професионални източници). Дървото е приложимо също за случаите, когато повече от една TV-сцена Δn_i е вътрешна за дадена Δl_k^j сцена.

В последния етап на алгоритъма, при откриване на съвпадение или включване на сцена от TV-потока в сцена от някоя реклама, допълнително се извършва покадрово

сравнение (пиксел-по-пиксел) между даден представителен кадър от TV-потокa с всички кадри на кандидат-сцената от рекламата.

Псевдокод на предложения алгоритъм

```

for  $\forall \Delta n_i \in \tilde{v}$            // дължината на  $i$ -тата TV сцена
  for  $\forall j \in J$              // брой на рекламните видеоклипове
    for  $\forall k$                  //  $k$ -тата сцена в  $j$ -тата реклама
      if  $\Delta n_i \leq \Delta l_k^j$ 
        for  $\forall l \in \Delta l_k^j$  // кадри от  $k$ -тата сцена в  $j$ -тата реклама
          { if  $\left( \frac{1}{X \times Y} \sum_{v(x,y)} |\overline{I_{\Delta n_i}(x,y)} - R_l^j(x,y)| \right) < h_2^*$  }
          // сцена  $\Delta n_i \subseteq \Delta l_k^j$  от  $j$ -тата реклама е намерена в TV поток
        end
      end
    end
  end
end

```

***Заб.:** h_2 е оптимален праг за близост между два визуално еднакви кадъра, от TV потока и от j -тата реклама, които са минимално различни по съдържание, основно поради наличие на шум при компресията на кадрите; $\overline{I_{\Delta n_i}(x,y)}$ е средният кадър в сцената Δn_i от TV видеоклипа.

2.3.4 Заключение и идеи за бъдеща работа

Предложен е относително прост и бърз (предвид работа в реално време) алгоритъм за локализиране на предварително известни рекламни видеоклипове или произволни части от тях в TV видеопоток. Алгоритъмът работи в обратната посока от TV клипа към множеството от рекламни клипове. Същността на алгоритъма се състои в разделяне на видеоклиповете на последователност от сцени (според нашата дефиниция за сцена) и представянето им като крайни редици от цели числа, представляващи продължителността (т.е. дължината в брой кадри) на всяка сцена. По този начин използваме сравнение на крайни целочислени редици вместо сравнение на последователности от JPEG изображения (кадри) от TV потока и потенциалните кандидати от множеството на рекламните видеоклипове. Т.е. конвенционалното 2D сравнение *кадър-по-кадър* се заменя с 1D сравнението на *дължина-по-дължина* на сцени и само при „подозрителните“ сцени (т.е. с близки дължини) извършваме покадрови сравнения, от типа „едно-към-много“. Този алгоритъм изисква много по-малък брой 2D покадрови сравнения (в сравнение с другите разгледаните методи), което е съществено предимство в реалните случаи на значителен брой реклами за търсене в TV потока.

Експериментиран е прост подход за детекция на сцени, базиран на разликата между два последователни кадъра в рамките на предварително дефиниран праг за подобие между съответните пиксели от кадрите (PST). Целта е да се намери такъв PST-праг, при който само голяма разлика в болшинството пиксели между два кадъра да се интерпретира като нова сцена (т.е. да бъде толерантен към неголеми подвижни обекти в рамките на една тематична сцена), за да няма твърде много фалшиви сцени. Т.е. въпросът за намирането на алгоритъм за автоматичното определяне на оптималната стойност на прага (постоянен или променлив) за детекция на сцените остава актуален за бъдещата работа по метода.

Глава 3

Методи за разпознаване на индивиди по формата на ушите

Обзорът от Глава 1, §1.2 може да се разглежда като въведение към тази глава.

3.1 Описание на създадената в ИИКТ БД от 3D уши (3DEarDB)

Първият вариант на създадената в ИИКТ БД от 3D уши (събрана през май, 2014 г.) съдържаше 3D модели на дясно ухо на 11 участници и беше предназначена за първоначални експерименти с двата метода, имащи връзка с настоящата дисертация (виж §3.2 и §3.4), за разпознаване на индивиди по формата на ушите [42, 159].

Разширеният вариант на базата данни от 3D модели на уши, която нарекохме 3DEarDB, беше събрана главно през средата на 2015 г. в ИИКТ-БАН в рамките на проекта AComIn³³. Ние сканирахме повече от 100 прецизни 3D модела на десните уши (без текстурна/цветова информация) на участниците (66 мъже и 34 жени) на възраст между 25 и 65 г. Прецизността на сканирането се дължи на използвания 3D лазерния скенер на ИИКТ (виж §3.1.1), който настроихме за 1 mm резолюция на сканиране (разстояние между съседните 3D точки), при точност от 0.05 mm за всяка 3D точка. Това не е най-високата възможна резолюция при този тип скенер, но за по-голямо бързодействие на сканирането, както и за постигане на по-малък размер на изходния файл, смятаме тази резолюция за достатъчна за текущите изследвания.

Целта на 3DEarDB е да осигури съвместимост между различните изходни формати за представянето на ухото на един и същ човек. Тези формати включват: (i) сурови (необработени) 3D модели на уши; (ii) обработени 3D модели; (iii) Kinect дълбочинни изображения “*depth maps*”; (iv) 2D видео клипове; (v) генерирани структури от 2D интензитетни проекции на уши; (vi) генерирани структури от 2D дълбочинни изображения на уши. Това консистентно разнообразие от изходни формати би било много полезно за научната общност по биометрика за тестване и сравнение на точността на алгоритмите върху различни възможни сценарии – от идеалния случай на прецизен (и статичен) 3D модел, до по-реалистичния (и динамичен) случай на 2D видеоклип или изображения.

Според наличната информация, (виж също Глава 1, §1.2.12) измежду съществуващите бази данни от уши, единствената БД, която осигурява съответстващи 2D и 3D данни за ушите на един и същ човек е UND Collections F, G, and J2, [160]. Важно е да се

³³ <http://www.iict.bas.bg/acomin/>

отбележи, че UND 3D данните не представляват фасетъчни (триангулационни) 3D модели, а просто 2D дълбочинни изображения. Освен това, там липсват съответните видеоклипове, които биха могли да се използват за 3D реконструкция на уши. Тази реконструкция може да се разглежда като алтернатива на 3D информацията от 2D дълбочинните изображения, чието заснемане обаче не е толкова разпространено в реална работна среда, за разлика от изобилието на видеоклиповете от популярните обзорни 2D видеокамери.

Едва най-новите БД OpenHear [136] и SYMARE [137] съдържат реални 3D модели на уши, но те не са специално предназначени за биометрика по ушите. Освен това, нито OpenHear (само 20 модел на физиономии), нито SYMARE (с 61 участници) могат да се разглеждат като достатъчно статистически представителни.

Съществено изискване за общността по биометрика е една такава БД от 3D уши да обхваща 100 или повече заснети индивиди. От особено значение е и наличието на конвенционални видео данни, които са най-реалистичният случай при съвременните технологии – за вграждане на алгоритми за разпознаване по ушите в обзорните видеокамери, в потребителската електроника и др. По тази причина е полезно да се осигурят прецизни 3D модели на уши, които да бъдат използвани като референтни модели за изчисление на грешката от 3D видео реконструкцията за сравнение между идеалната (3D модели) и реалната (3D реконструкция) разпознаваемост на изследваните дескриптори и класификатори. Цветността при ушите е не (достатъчно) информативна за нашите изследвания, поради което не я включваме при 3D сканирането, най-вече от съображения за пестене на компютърни ресурси. Съответната цвятова информация се съхранява в придружаващите видеоклипове.

Следващият раздел съдържа едно по-детайлно описание на нашата многомоделна БД от уши, разглеждайки два основни типа данни – хардуерно заснети и софтуерно генерирани. Хардуерно заснетите данни се състоят от: сурови и дообработени 3D модели на уши (от лазерния скенер), 3D дълбочинни изображения (от Kinect сензор) и 2D видеоклипове (от цифрова фото камера). Софтуерно генерираните представяния на уши чрез 3D моделите са два типа: (i) структури от изображения, т.е. 2D интензитетни проекции с различно осветление и/или ориентации (чрез MeshLab³⁴) и (ii) структури от 2D дълбочинни изображения със същата ориентация както интензитетните проекции (чрез Wolfram Mathematica³⁵).

3.1.1 Устройства за събиране на данни

За заснемането на уши се използват 3 вида устройства, описани по-долу. Засега 3DEarDB съдържа модели само на десни уши, по един 3D модел на индивид.

VIUscan 3D лазерен скенер

Този портативен скенер на Creaform (Фиг. 45в) беше закупен чрез проекта AComIn за „Умната лаборатория“ (SmartLab) на ИИКТ-БАН, в края на 2013 г. Той може да възп-

³⁴ <http://meshlab.sourceforge.net/>

³⁵ <https://www.wolfram.com/mathematica/>

произвежда 3D фасетъчен (триангулачен) модел на сканирания обект, както и неговата текстура и цветове. Въпреки че не използваме неговата максимална резолюция (0.1 mm), нито предоставяната цветова информация, те могат да бъдат много полезни при други приложения, където 3D обектите имат разнообразна текстура с фини структурни детайли [161].

Съществена особеност на този вид скенери е, че изискват поставянето на специфични ретро-рефлексивни маркери върху или около обекта за сканиране. По време на сканирането скенерът трябва да „вижда“ поне 4 маркера, които трябва да са фиксирани спрямо обекта. За да облекчим нашата работа, ние направихме специален „шлем“ от картон с достатъчно залепени маркери по него, който се поставя на главата, около ухото на субекта преди сканиране (виж Фиг. 45а, б). Изключването на цветовата информация прави процедурата по сканиране по-бърза, до 10 мин. за ухо, както и по-удобна, поради липсата на необходимост от специално осветление.



Фиг. 45. а) картонен „шлем“; б) сканиране на участник; в) 3D скенерът VIUscan.

Kinect Xbox One сензор³⁶

Този дълбочинен сензор на Майкрософт е подобрен вариант на своя предшественик Xbox 360 и е наличен от октомври 2014 г. Той има масив от инфрачервени излъчватели и 512×424 пикселова “time-of-flight” камера, които възстановяват дълбочината в сцената с цел следене на движения и разпознаване на жестове. Тази нова версия на Kinect също включва Full HD (1920×1080) видеокамера с увеличено поле на видимост.

Ние планираме да използваме Kinect за получаване на реални дълбочинни изображения на уши, а също и да използваме придружаващия го софтуер за 3D реконструкция чрез дълбочинни видеоклипове или изображения.

Olympus Photo Camera³⁷

За заснемането на обхождащи Full HD (1920×1080) видеоклипове на ушите на участниците беше използвана Olympus SH-21 фотокамера с 16 MP CMOS 1/2.3" сензор, произвеждаща цветно видео в MP4 формат.

3.1.2 Сурови (необработени) 3D модели на уши

Един необработен 3D модел на ухо, заснет чрез VXelements софтуера на VIUscan скенера [161], е представен на Фиг. 46б. Първоначалният изходен файлов формат е CSF, чий-

³⁶ https://en.wikipedia.org/wiki/Kinect_for_Xbox_One#Specifications

³⁷ <http://www.olympus-global.com/en/news/2011b/nr111110sh21e.jsp>

то осреднен обем (върху всички уши от нашата БД) е около 64MB на ухо. На Фиг. 46а е показано цветно сканирано ухо само за илюстрация като как изглежда, тъй като не съхраняваме цветността засега. Разбира се, цветовите данни могат успешно да се използват за автоматична сегментация на 3D ушите, което е предмет на бъдеща работа.



Фиг. 46. а) цветно необработено ухо; б) само повърхнината на необработеното ухо.

3.1.3 Обработка на суровите 3D модели

За създаване на завършен и подходящо изгладен 3D модел за всяко ухо, тук ще опишем пост-обработката в 6 стъпки, където съществено използваме софтуерните продукти VXelements [161] и MeshLab [162].

Стъпка 1: Груба сегментация (чрез VXelements)

- Използване на филтъра *Remove Isolated Patches* за премахване на изолираните 3D фасетки от входните CSF данни, получени от отблясъци по време на сканирането.
- Извършване на груба ръчна сегментация на повърхнината на ухото от заобикалящия го фон, използвайки инструментите *Brush Selection*, *Reverse Selection* и *Delete Facets*.

Стъпка 2: Запълване на случайни празнини/дупки (чрез VXelements)

- Стартиране на *Optimize Surface* алгоритъма за реконструкция всеки път, когато се избира различен размер на дупките в ухото, които да бъдат запълнени. Това е процедурата, която отнема най-много време, защото задоволителните резултати са трудно предвидими и се достигат чрез множество полуавтоматични експерименти.
- След запълването на съответните празнини резултатът се записва в CSF формат, чийто размер тук е около 49 MB за ухо. След това CSF файлът се конвертира в OBJ файл (с размер около 600 KB за ухо) за последваща обработка в MeshLab.

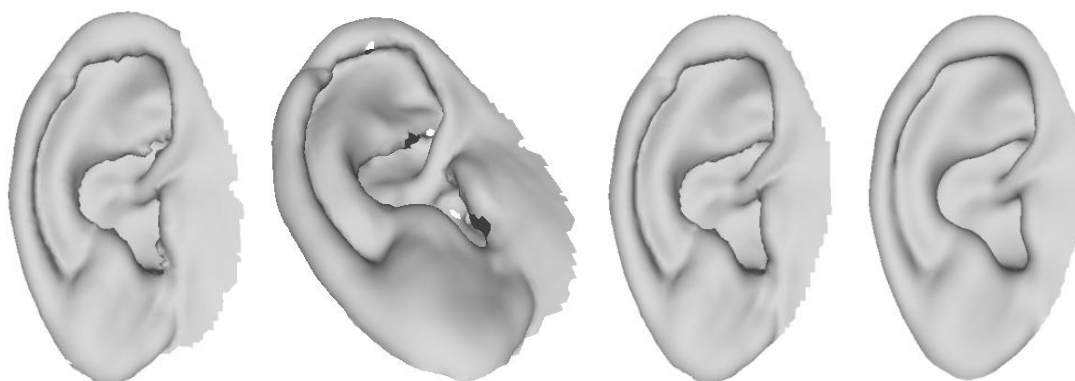
Стъпка 3: Фина обработка на 3D фасетките на ушите (чрез MeshLab)

Това включва по-фино сегментиране на фона, както и премахване на острите пикове (виж Фиг. 47а) в 3D модела, получен от *Optimize Surface* алгоритъма на предходната

стъпка. Разбира се, премахването на пикове води до необходимостта от запълване на нови дупки (Фиг. 47в), но с много по-малък размер (Фиг. 47б), което обикновено не е проблем за MeshLab.

Стъпка 4: Допълнително изглаждане на модела (чрез MeshLab)

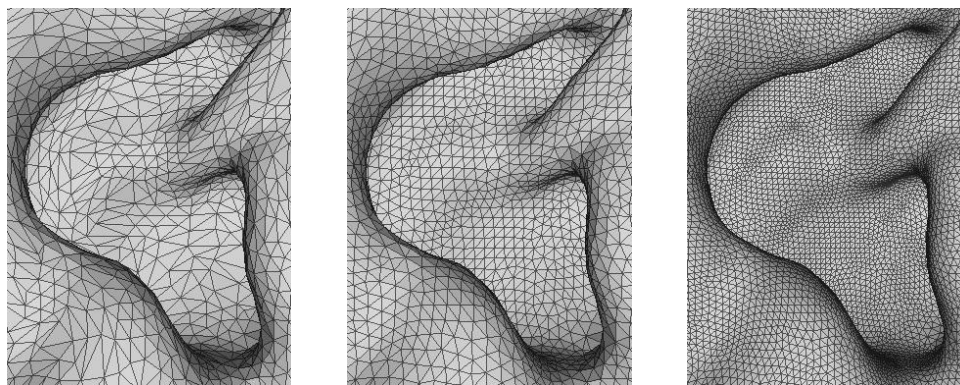
След запълването на празнините (Фиг. 47в), финалната стъпка се състои в изглаждане на завършения 3D модел (Фиг. 47г). MeshLab функцията, която използваме за тази цел, е *HC Laplacian Smooth*, базирана на статията на Vollmer et al. [163]. На този финален етап от обработката на ушите всеки 3D модел се състои от около 6-8 хиляди (триъгълни) фасетки, определени от около 3-4 хиляди върхове (3D точки). Като изключим данните за нормалните вектори, които разглеждаме като производни и излишни, осредненият размер на OBJ файловете се намалява до около 240 KB за ухо.



Фиг. 47. а) остри пикове; б) нови дупки; в) запълнени дупки; г) финално изглаждане.

Стъпка 5: Огрубяване (decimation) и изфинване (subdivision) (чрез MeshLab)

Тези манипулации на гъстотата на триангулационната мрежа на 3D моделите е необходима при генерирането на тестови данни за нашия EGI метод за класификация [42]. Същия използваме тук, за да проверим експериментално функционалността на нашата 3DEarDB, виж §3.2. MeshLab функцията, чрез която изфинваме мрежата, т.е. увеличаваме “*subdivision*” броя на фасетките (Фиг. 48в), е *Subdivision Surfaces: LS3 Loop*, базирана на [164], а функцията за огрубяване на мрежата, т.е. редуциране (децимация) на този брой (Фиг. 48а), е *Quadratic Edge Collapse Decimation*.

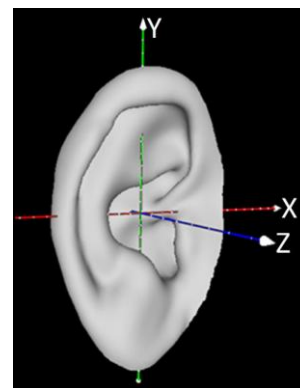


Фиг. 48. а) децимация на фасетките; б) оригинална резолюция; в) изфинване на фасетките.

Стъпка 6: Геометрична нормализация (в MATLAB)

Това включва нормализация по трансляция, ориентация и мащаб за всеки 3D модел:

- Транслиране на центъра на координатната система в центъра на тежестта на модела, т.е. в осреднените (x, y, z) координати на всички 3D точки (върхове) на модела. След изваждането му от всички върхове новият център на тежестта става $(0, 0, 0)$.
- Ротация на главните оси, т.е. собствените вектори на ковариационната матрица върху целия модел (всички върхове). За нормализация по ротация, върховете се ротират обратно към вече централизираната декартова координатна система (виж Фиг. 49).
- Мащабиране: трите собствени стойности (асоциирани с главните оси) се използват за нормализация на модела по мащаб, така че минималният ограждащ паралелепипед (или съответният еквивалентен елипсоид) да достигне до предефинирани размери. Трите мащабиращи коефициента (реципрочни на собствените стойности) за всеки модел се запазват, ако реалният размер на всяко ухо би имал допълнително значение.



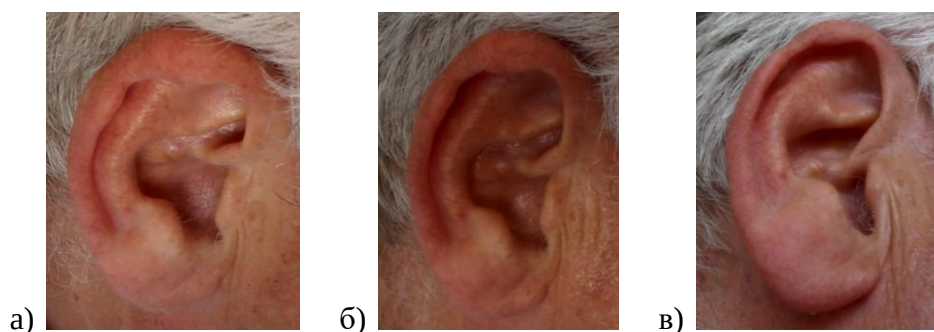
Фиг. 49. Нормализиран 3D модел на ухо (в MeshLab).

3.1.4 Kinect дълбочинни изображения

За сега не разполагаме с Kinect 3D данни за ушите, поради което ги генерираме софтуерно – 2D дълбочинни изображения от 3D моделите в 3DEarDB, виж и §3.1.7.

3.1.5 Full HD обхождащи видеоклипове на уши

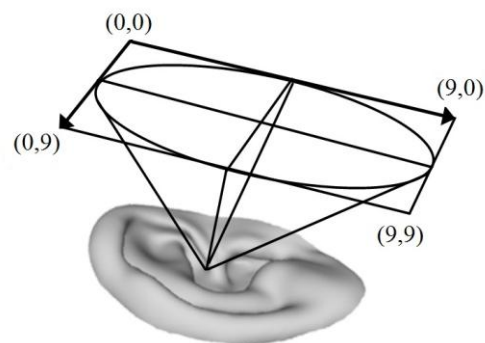
За болшинството от ушите са направени видеоклипове с разделителна способност 1920×1080 , равномерно заснети по азимут от -80° до $+80^\circ$ чрез 3 хоризонтални обхождания (с различна деклинация – горе, център и долу) спрямо центъра на фронталния изглед (Фиг. 50) в същата лаборатория, непосредствено след 3D сканирането. Чрез пресъздаването на почти реална ситуация на заснемане, тези видеоклипове се явяват „мост“ между нашата 3DEarDB и другите БД и позволяват комбинирани тестове за разпознаване на 2D реални входни данни и съответните им 3D модели. Всеки видеоклип е с дължина около 20 сек., при 30 fps и с размер около 45 MB в MP4 формат.



Фиг. 50. Представителни кадри за 3 хоризонтални реда на видеоклип на ухо: а) изглед отгоре; б) централен изглед; в) изглед отдолу.

3.1.6 2D интензитетни проекции

Множеството от 2D проекции от даден 3D модел на ухо са произведени в MeshLab чрез за-reждане на определен брой слоеве, всеки един съдържащ съответна ротация на 3D ухото. След това чрез функцията *Save snapshot* се записват 2D изображения на всички слоеве в JPEG формат. Осветлението е тривиалното (по подразбиране), т.е. фронтално и кохерентно. 2D интензитетните проекции се генерират според избрана схема за 3D ротация – в 100 посоки, равномерно разпределени спрямо центъра на тежестта на ухото, т.е. по 10 деклинации и 10 азимута, равномерно избрани в интервала $(-45^\circ, +45^\circ)$, виж Фиг. 51. Разбира се, ъгловата стъпка може да бъде по-малка или по-голяма с цел манипулиране на плътността на резултатното множество от 2D проекции. Този вид 3D представяне на ушите, наречен „Многоизгледно 3D моделиране“, беше разработено за наши експерименти по 3D разпознаване на уши чрез множество от техни 2D проекции [159]. Илюстрация на 10 двумерни (2D) проекции на дадено ухо, произведени от негов тримерен (3D) модел (за даден централен ред от схемата за ротация), са показани на Фиг. 52.



Фиг. 51. Схема на многоизгледно 3D моделиране на дадено ухо.



Фиг. 52. 2D изображения от даден ред на ротационната схема за 3D моделиране, виж и Фиг. 51.

3.1.7 2D дълбочинни изображения

За генерирането на 2D дълбочинни изображения от 3D модел на ухо е използвана вградена функция на софтуера Wolfram Mathematica, като вместо интензитетни стойности в двумерната матрица на изображението се записват z-координатите на тримерните точки от 3D модела (виж Фиг. 53). С цел съвместимост между 2D интензитетните проекции и дълбочинните изображения, и двете множества изображения са генерирани чрез една и съща схема на ротация.



Фиг. 53. Дълбочинни изображения под ортографска проекция за дадена ротация на 3D ухо.

3.1.8 Уеб достъп до 3DEarDB

Текущата версия на тримерната база данни от уши 3DEarDB предстои да бъде предоставена на изследователската общност по биометрика за свободно ползване за академични и некомерсиални цели. Подробно описание на нейната структура, вградени функции и други възможности, както и лицензионно споразумение за ползване ще бъдат достъпни на уеб-сайта на ИИКТ-БАН в най-скоро време.

3.2 Представяне на 3D обекти чрез разширено Гаусово изображение (EGI)

Разширеното Гаусово изображение (EGI) първоначално е предложено от В.К.Р. Horn [165], през 1984 г., за недвусмислено представяне на изпъкнала 3D повърхност във вид на хистограма на ориентациите върху сфера. Този модел е разширен (Complex-EGI) от К. Ikeuchi [166] през 1991 г., за преодоляване на проблемите, породени от неизпъкналите области на 3D обектите. Възникват и други подобрения на първоначалната идея за EGI, всичките свързани с подобрения върху недвусмисленото представяне и на вдлъбнатите части на 3D обектите. Това са: More Extended Gaussian Image (MEGI) през 1994 г. [167], Multi-Shell Extended Gaussian Image (MSEGI) и Adaptive Volumetric Extended Gaussian Image (AVEGI) през 2007 г. [168], Enriched Complex Extended Gaussian Image (EC-EGI) през 2010 г. [169].

Обикновено, повърхнината (обвивката) на даден 3D модел се представя чрез 3D дискретна мрежа, съставена от множество триъгълни фасетки (виж. Фиг. 54а):

$$T = \{T_1, \dots, T_N\}, T_i \subset R^3, \quad (23)$$

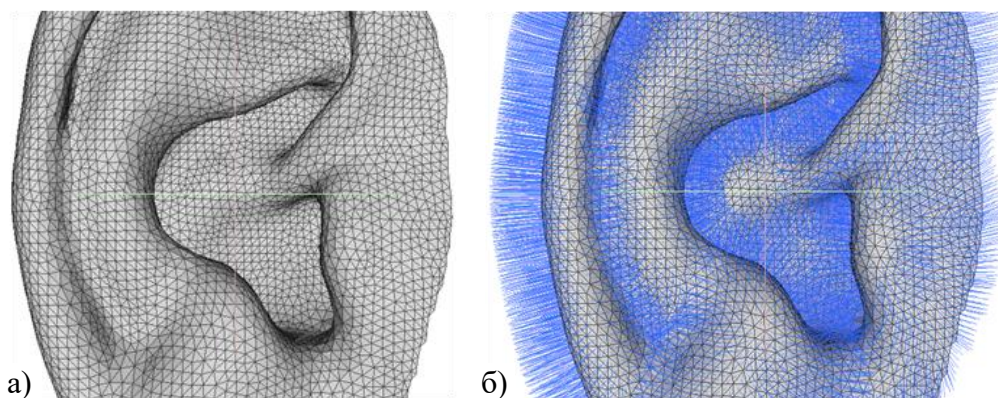
където N е броят на фасетките (триъгълниците) на 3D модела, а всяка фасетка T_i се представя със своите три върха:

$$T_i = \{P_{A_i}, P_{B_i}, P_{C_i}\}, i = 1, \dots, N. \quad (23a)$$

Означавайки с w_i и $\bar{n}_i$ съответно (меди-) центъра и нормалата на даден триъгълник T_i (виж Фиг. 54б), площта му A_i , както и площта A на 3D модела, се изчисляват като:

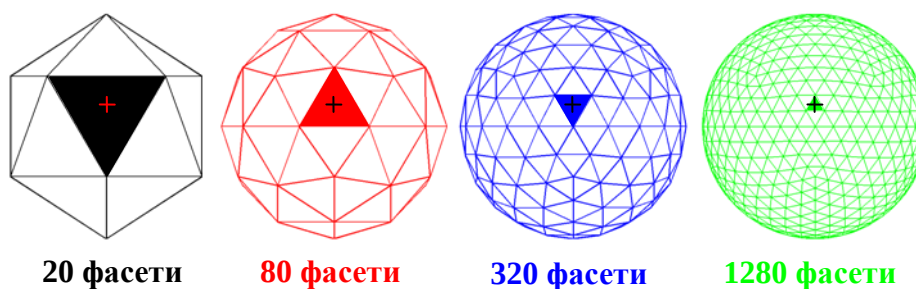
$$w_i = \frac{P_{A_i} + P_{B_i} + P_{C_i}}{3}, \quad \bar{n}_i = (P_{C_i} - P_{A_i}) \times (P_{B_i} - P_{C_i}), \quad (24a)$$

$$A_i = \frac{1}{2} |(P_{C_i} - P_{A_i}) \times (P_{B_i} - P_{C_i})|, \quad A = \sum_{i=1}^N A_i. \quad (24b)$$

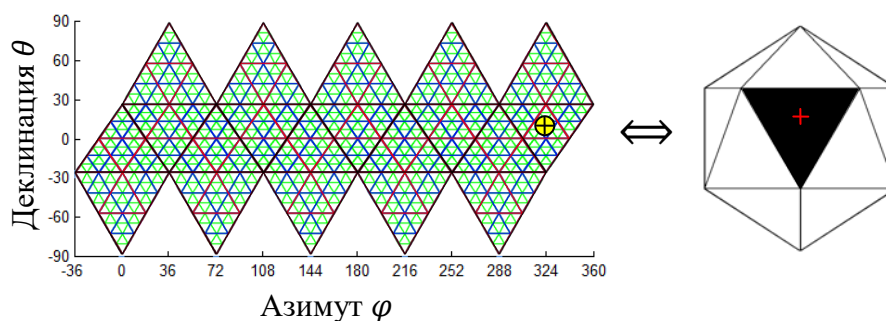


Фиг. 54. а) 3D мрежест (триангулачен) модел; б) нормални вектори към фасетките на модела.

Най-общо казано, EGI представянето на една (дискретизирана) 3D повърхнина представлява хистограма от всички ориентации на повърхнината върху единична (Гаусова) сфера. На практика, Гаусовата сфера също се представя в дискретен вид, започвайки на най-ниското ниво на дискретизация от икосаедър (правилен многостен с 20 триъгълни страни) и достигайки до все по-прецизна апроксимация на сферата на следващите нива чрез рекурсивно разбиване всяка триъгълна стена (от дадено ниво) на 4 по-малки триъгълника (Фиг. 55). Т.е. в зависимост от избраното ниво n на дискретизация, броят на триъгълните фасетки на сферата е равен на $m = 4^n 20$, $n = 0, 1, \dots$.



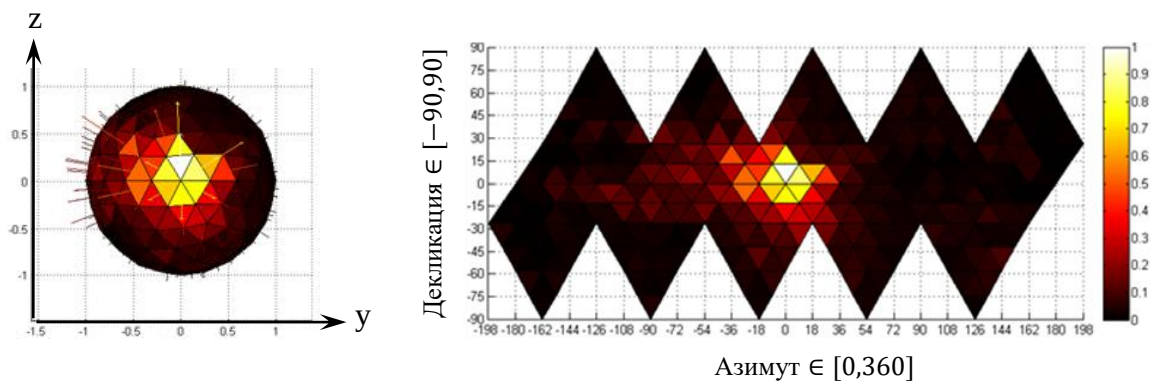
Фиг. 55. 3D рекурсивно разбиване, базирано на икосаедър (20-тостен) за апроксимиране на сфера чрез 20, 80, 320 и 1280 фасетки.



Фиг. 56. 2D представяне (по азимут и деклинация) на дадена апроксимация на сфера чрез съответното разрязване и разгъване (при минимални „топологични“ изкривявания на базовите 20 сферични триъгълника) по базовия икосаедър (вдясно). Четирите апроксимации от Фиг. 55 са съвместено представени тук (в съответния им цвят) за илюстрация на вложеността на 2D представянията. Допълнително е отбелязано мястото (центъра) на челната фасетка.

Така, според ориентацията на нормалния вектор към всяка фасетка на входната 3D повърхнина (Фиг. 54б), площта на фасетката се акумулира към съответния „бин“ (т.е. сферична триъгълна фасетка) от дискретната Гаусова сфера (Фиг. 57). Така, бинът (като сферична фасетка) се определя с прободната точка на дадена входна нормала и сферата. Но, за бина като индекс е достатъчно да се намери максимумът от всички скаларни произведения на нормалите на входния обект и бин-нормалите на сферата. Последните обикновено се дефинират като нормала към прилежащия (равнинен) триъгълник на дадена сферична фасетка. Т.е., теглото на всеки бин от EGI сферата е равно на сумарната площ на всички фасетки от 3D модела, с посоки на нормалните вектори, попадащи в пространствения ъгъл, определен от съответната EGI-фасетка (т.е. EGI триъгълник, бил той равнинен или сферичен).

Без ограничение на общността на разглежданията, очевидно можем да интерпретираме EGI представянето като сферична (3D) хистограма, а също и като съответната 2D хистограма, виж Фиг. 56 и 57.



Фиг. 57. 3D и 2D EGI хистограми за някаква входна 3D повърхнина – във всеки EGI бин се акумулира площта на съответните триъгълните фасетки от входа.

3.2.1 Аналитично описание на метода за построяване на икосаедър и неговата 2D развивка

Накратко, процесът на построяване на икосаедър и неговото 2D представяне (чрез съответното разрязване и разгъване) се състои в следното (виж Фиг. 58):

- определяне на азимута φ_i за всеки връх от икосаедъра;
- изчисляване на радиуса ρ (до върховете, лежащи на сеченията с деклинации $\pm\theta_1$) чрез известната зависимост между дължината на ръба a на икосаедъра и радиуса r на описаната около него сфера;
- изразяване на всяка (x, y) координата на върховете на икосаедъра чрез радиуса ρ и азимута φ_i до съответните върхове;
- изчисляване на z координатите на върховете на икосаедъра чрез радиусите ρ и r , ($r = 1$);
- определяне на двете нива на деклинации $+\theta_1$ и $-\theta_1$, необходими за 2D представянето на икосаедъра.

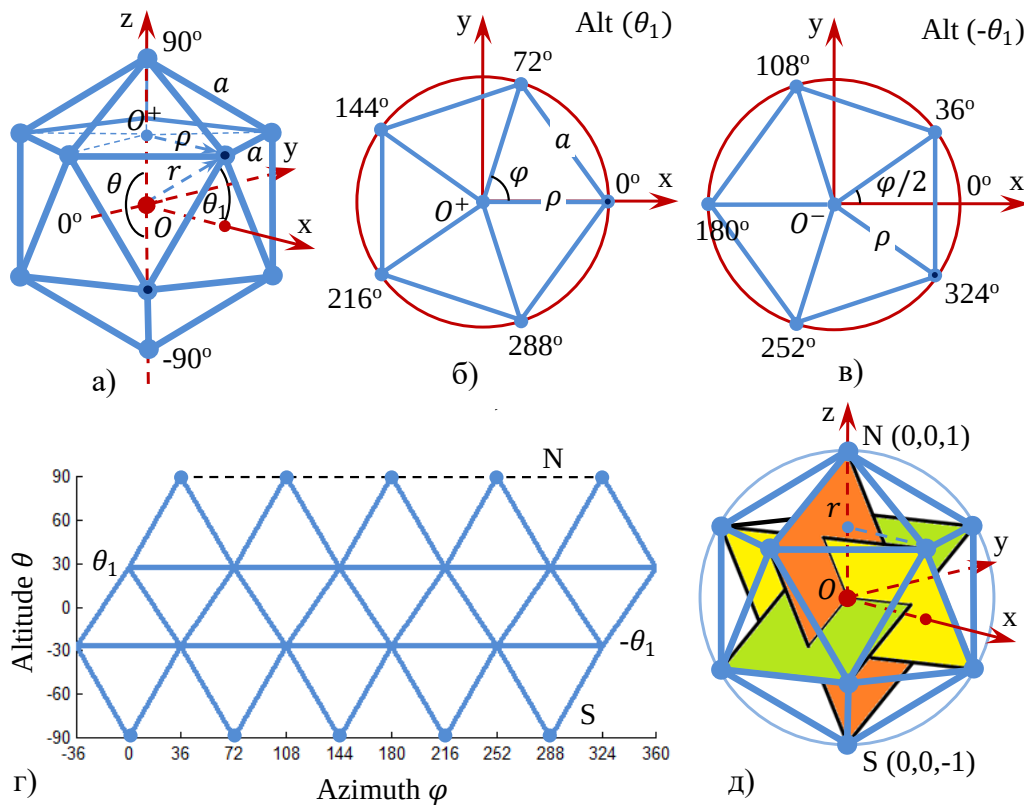
Едно по-детайлно описание на горните стъпки е дадено в следващите точки:

1. Азимутът на върховете с деклинация на ниво $+\theta_1$ (Фиг. 58б) е равен на: $\varphi_i^+ = i \Delta\varphi = i 2\pi/5$, и съответно за върховете с деклинация $-\theta_1$ (Фиг. 58в), азимутът е: $\varphi_i^- = i \Delta\varphi - \Delta\varphi/2 = \pi/5(2i - 1)$, където $i = 0 \div 4$. Азимутът на крайните върхове, N (0,0,1) и S (0,0,-1) е нула (Фиг. 58д).
2. Известно е, че ако дължината на ръба на икосаедъра е a , то радиусът r на описаната сфера около икосаедъра е: $r = a \sin\left(\frac{2\pi}{5}\right)$, виж. Фиг. 58а.

В нашия случай ($r = 1$) $\Rightarrow a = \frac{1}{\sin\left(\frac{2\pi}{5}\right)} \approx 1.0515$, а радиусът ρ е:

$$\rho = a \left(2 \sin\left(\frac{\Delta\varphi}{2}\right)\right)^{-1} = \left(2 \sin\left(\frac{\pi}{5}\right) \sin\left(\frac{2\pi}{5}\right)\right)^{-1} \approx 0.8944, \text{ виж. Фиг. 58а.}$$

3. Координатите (x, y) на върховете, лежащи на деклинация $+\theta_1$ са: $x_i^+ = \rho \cos(\varphi_i^+)$, $y_i^+ = \rho \sin(\varphi_i^+)$, Фиг. 58б. Съответно за върховете на $-\theta_1$ деклинация, имаме: $x_i^- = \rho \cos(\varphi_i^-)$, $y_i^- = \rho \sin(\varphi_i^-)$, $i = 0 \div 4$ (Фиг. 58в). Координатите (x, y) на горния (N) и долния (S) връх са (0, 0), виж Фиг. 58д.



Фиг. 58. Необходими параметри за определяне на координатите (x, y, z) на върховете за построяването на икосаедър и неговото 2D представяне: а) радиус r на описаната единична сфера; дължина на ръба на икосаедъра a ; деклинация $\theta \in [-90^\circ, 90^\circ]$; радиус ρ на O_{xy}^+ и O_{xy}^- 2D сеченията; б) и в) азимут $\varphi \in [0^\circ, 360^\circ]$; деклинации $+\theta_1$ и $-\theta_1$ на върховете, принадлежащи на двете 2D сечения на икосаедъра; г) същият икосаедър, разгънат върху 2D равнина, представена чрез азимут φ и деклинация θ на неговите триъгълни страни; д) икосаедър, вписан в единична сфера, образуван от три взаимно перпендикулярни „златни“ правоъгълници, чиито отношение на страните е равно на златното сечение $(1 + \sqrt{5})/2 \approx 1.6180$.

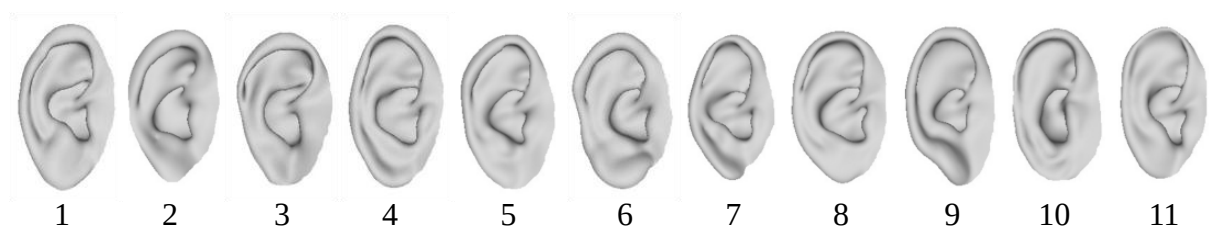
4. Разстоянието $\overline{OO_+} = \sqrt{r^2 - \rho^2} = \sqrt{1 - \rho^2} \approx 0.4472$, което всъщност е $\rho/2$. Така, z координатите на върховете, лежащи на $+\theta_1$ и $-\theta_1$ нивата са $\pm\rho/2$ (Фиг. 58а). Крайните върхове (N и S) имат $z = \pm 1$ (Фиг. 58д).
5. Нивата на деклинации $\pm\theta_1$, могат да се изчислят като: $\pm\theta_1 = \arccos\left(\frac{\sqrt{x^2+y^2}}{r}\right) \approx 26.57^\circ$, за всяка (x, y) точка върху тези нива, където $r = 1$, (Фиг. 58а).
6. За следващите нива на дискретизация на Гаусовата сфера (след началния икосаедър) координатите на върховете могат да се изчислят по аналогичен начин (както бе споменато в текста към Фиг. 55) – след разбиване на всеки триъгълник от дадено ниво (чрез средите на трите му страни) на 4 по-малки триъгълника и възприемане на съответните 4 проекции (на тези 4 триъгълника) върху Гаусовата сфера.

3.3 Експериментални резултати за разпознаване на EGI хистограми по метода на най-близкия съсед

За тестване на функционалността на текущата 3DEarDB използваме нашия EGI базиран подход за разпознаване на уши [42] и експериментите се оказват доста обещаващи.

Разбира се, EGI представянето е еднозначно само за изцяло изпъкнали (или изцяло вдлъбнати) повърхности. Ние приложихме EGI за изпъкнало-вдлъбнати обекти, каквито са ушите, въпреки очакванията за (някаква) нееднозначност. Още в нашия първоначален експеримент (с малка БД, само 11 3D модела на уши, Фиг. 59) беше установено, че разликите между два 3D модела на уши (т.е. на двама индивиди) се оказват достатъчно големи, така че нееднозначността от изпъкналости и вдлъбнатости да не влияе върху разпознаваемостта, разбира се до определени граници на зашумяването, което прилагаме. Нашето обяснение за този позитивен експеримент е, че ухото е по-скоро плосък 3D обект.

Така, ние използваме разширеното Гаусово изображение, защото това е бърз, удобен и достатъчно дискриминативен подход за представянето на 3D обекти в дискретния случай (т.е. заснети чрез 3D скенер) във вид на хистограма от ориентациите на нормалните вектори на входния обект върху (единична) сфера. На практика, ние работим със съответната 1D хистограма, получена чрез определено обхождане на биновете на 3D хистограмата.



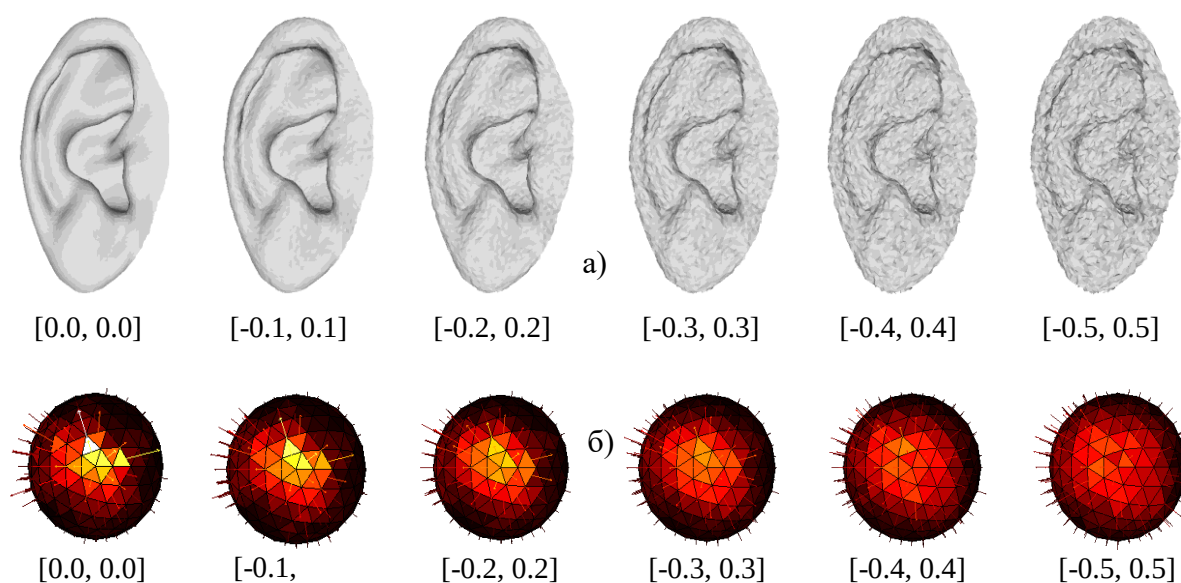
Фиг. 59. Първоначалната БД от 11 3D модела на уши, създадена през май 2014 г.

Текущата версия на нашата БД от 3D модели на уши (към юни 2015 г.) съдържа повече от 100 модела, което според нас е достатъчно представителна извадка от статистическа гледна точка за надеждността на получените резултати.

Целта на текущите експерименти (върху цялата БД от 100 3D модела на уши) е отново да се демонстрира възможността за различимост на ушите, бидейки неизпъкнали обекти, използвайки конвенционалния EGI дескриптор.

Така, за всяко ухо в БД се построява съответното му EGI представяне (виж §3.2). EGI преобразува данните от входния 3D модел на обекта във вид на ориентирана хистограма върху сфера, която може да бъде визуализирана и като 2D изображение (Фиг. 57) или дори като 1D хистограма чрез подходящо обхождане на фасетките, например по спирала. Двумерното (2D) EGI представяне е интересно не само от гледна точка на по-добрата визуализация на резултатната хистограма, но и като възможност за прилагане на добре известни или нови подходи за обработка и разпознаване на изображения.

За имитиране на по-реални условия на експеримента, *ушите се зашумяват адитивно*, преди да се използват за тестване на разпознаемостта (при сравнението им с идеалните 3D уши от 3DEarDB). Този шум е предназначен за симулиране на различна (по-ниска) от декларираната точност (0.05 mm) на 3D скенера. За тази цел се генерира равномерен шум в съответните интервали (± 0.05 mm), ..., (± 0.4 mm), който се добавя към всяка от координатите (x, y, z) на върховете на сканирания обект (Фиг. 60а). Забелязва се, че при по-високи нива на шума, EGI хистограмата на даденото зашумено ухо започва да се „размива” и уеднаквява, т.е. губи от своята дискриминативност (Фиг. 60б).



Фиг. 60. а) зашумен модел на ухо с равномерен шум в даден интервал, добавен към всяка (x, y, z) координата; б) съответните 3D EGI хистограми на зашумените уши.

За изчисляване на *подобие* между EGI хистограмите (на входните зашумени модели от БД спрямо тези на идеалните модели в БД) използваме следните две геометрични оценки с цел определяне, коя е по-устойчивата за съответното ниво на шум при разпознаването:

- Евклидово разстояние:

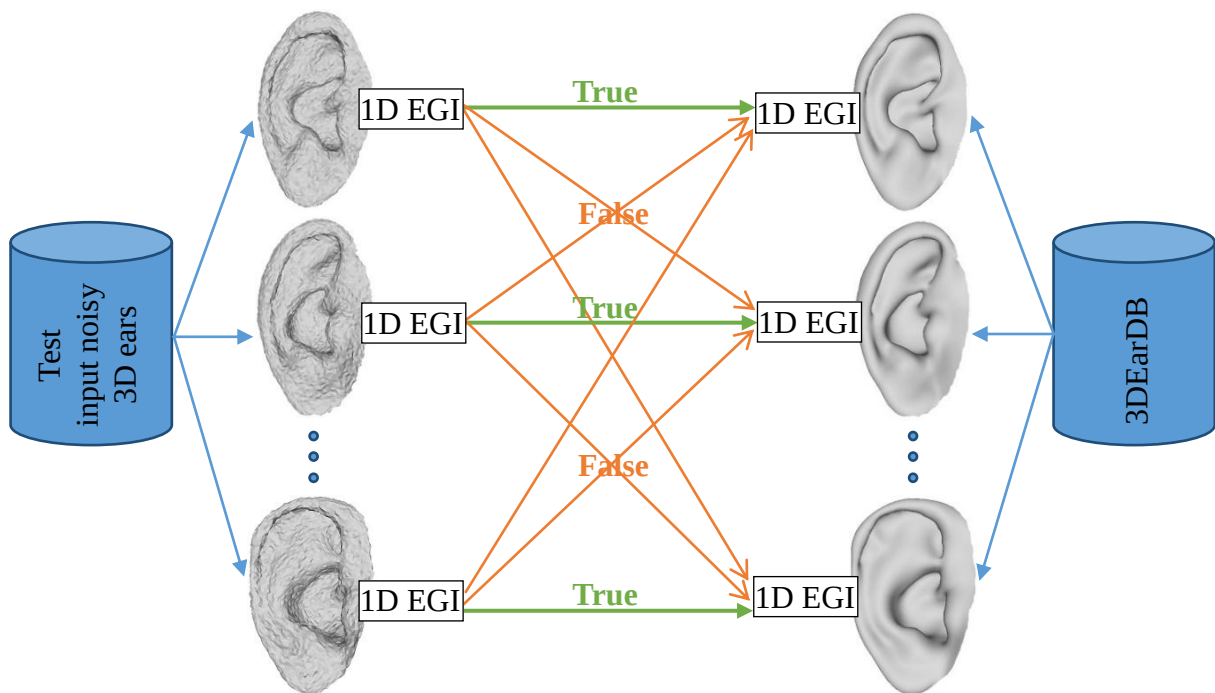
$$E_2 = \sqrt{\sum_{i=1}^m (M_i - S_i)^2}; \quad (25a)$$

- Bray-Curtis разстоянието [170]:

$$E_{BC} = \frac{\sum_{i=1}^m |M_i - S_i|}{\sum_{i=1}^m (M_i + S_i)}, \quad 0 \leq E_{BC} \leq 1, \quad (256)$$

където M_i и S_i са съответните бинове от сравняваните хистограми (на модела и на входния обект), $i = 1, 2 \dots m$; $m = 20, 80, 320, 1280$, виж Табл. 16.

След изчисляване на разстоянията между EGI хистограмите на всички входни зашумени уши с всички идеални модели от БД (виж Фиг. 61), *класификацията* се извършва чрез простия и бърз класификатор „по най-близкия съсед”. Този метод за класификация е достатъчно надежден в нашия случай, защото знаем разпределението на класове, където всеки модел на ухо се разглежда като център на клас, т.е. имаме 100 класа. За всеки център на клас се генерират изкуствено достатъчен брой погрешни (зашумени) версии, с което спестяваме значителния човешки и компютърен ресурс, необходим за съответните реални експерименти.



Фиг. 61. Сравнение от типа „всяко-с-всяко“ на входните зашумени модели на уши с идеалните модели от 3DEarDB за съответната резолюция на сканиране (0.5 mm, 1.0 mm и 1.4 mm), и съответната апроксимация на EGI представянето (20, 80, 320 и 1280).

Тествани са 3 варианта на 3DEarDB за 3 резолюции на сканиране: 1.0 mm е оригиналната резолюция, а 0.5 mm (4 пъти повече фасетки) и 1.4 mm (2 пъти по-малко фасетки) са софтуерно преизчислени резолюции от оригиналната (виж Стъпка 5 в §3.1.3). Освен това, са тествани и 4 нива за EGI представянето на моделите на ушите – при апроксимиране на Гаусовата сфера с 20, 80, 320 и 1280 фасетки (виж Табл. 16).

Таблица 16. EGI резултати от разпознаването: True Recognition Rate (TRR), виж (26).

TRR [%]		0.5 mm резолюция (пре-изчислена по-висока рез.)			1.0 mm 3D резолюция на сканиране (оригинална)			1.4 mm резолюция (пре-изчислена по-ниска рез.)		
шум [mm] ³⁸		0.05	0.10	0.15	0.10	0.20	0.30	0.20	0.30	0.40
~ % ширина		0.16	0.31	0.47	0.31	0.62	0.93	0.62	0.93	1.24
~ % височина		0.10	0.20	0.30	0.20	0.40	0.60	0.40	0.60	0.80
~ % дълбочина		0.38	0.76	1.14	0.76	1.52	2.27	1.52	2.27	3.03
20 фасети	E_2	100	98	70	100	98	70	99	90	63
	E_{BC}	100	100	92	100	100	89	100	99	80
80 фасети	E_2	100	93	17	100	92	16	100	88	40
	E_{BC}	100	100	67	100	100	70	100	100	80
320 фасети	E_2	100	84	25	100	83	27	100	80	44
	E_{BC}	100	100	80	100	100	66	100	100	80
1280 фасети	E_2	100	63	13	100	52	16	100	56	30
	E_{BC}	100	100	62	100	95	48	100	96	78

Очевидно, при някакво достатъчно високо ниво на входния шум, БД ще започне да греша. Целта на предложения тук подход (от типа „всеки-към-всеки”) за тестване на нашата 3DEarDB е да се оцени максимално допустимата стойност на входния шум, с което косвено да се оцени и пригодността на използвания 3D скенер за набиране на 3D модели на обекти от даден тип, в нашия случай – човешки уши.

Ще коментираме подхода в термините на класическите методи за оценка на класификатори:

- Поради увереността, че повърхността на ухото е достатъчно прецизен идентификатор на индивида (Глава 1, §1.2.9) и поради декларираната точност на използвания 3D скенер, която оценяваме експертно като (пре)достатъчна за случая (Глава 3, §3.1), можем да разглеждаме 3DEarDB като „памет” на класификатора от типа „най-близкия съсед”, при дадена дефиниция за разстояние, например (25a), (25б) или друго [42].

- С други думи, смятаме, че процесът на обучение на класификатора 3DEarDB е приключил със зареждането на неговата памет 3DEarDB. Както се вижда от структурата

³⁸ За съвместимост с други (бъдещи) експерименти, интервалите на зашумяване са изразени в проценти от средните размери на ушите, според собствените стойности на главните оси на обхващащия елипсоид: ширина (32.3 mm по O_x), височина (50.3 mm по O_y), дълбочина (13.2 mm по O_z).

на експеримента (Фиг. 61), всеки един 3D модел M_i от БД се разглежда като център на клас (M_i), $i = 1, \dots, Q$, където Q е мощността на 3DEarDB, в брой 3D модели. В случая сме избрали $Q = 100$, за целочисленост на резултатите в % (виж Табл. 16).

- Така, ако разглеждаме нашия подход „всеки-към-всеки” като произведен от (подобен на) т.нар. “Leave One Out Cross Validation”³⁹, т.е. като „ N -кратна крос-валидация на класификатора”, където $N = Q$, то ще трябва да отчитаме само случаите, когато всички обучаващи примери ($Q-1$ на брой) се разпознават позитивно (защото обучението е вече фиксирано). От друга страна, ние се интересуваме точно от тези случаи, защото търсим максимално допустимата стойност на входния шум.

- От друга страна, ако разглеждаме нашия подход „всеки-към-всеки” в термините на оценка на класификатор с 2 състояния, то ще достигнем до следния извод за т.нар. „матрица на класификацията” (Confusion matrix)⁴⁰:

$$FN = (\text{брой False Negative}) = (\text{брой False Positive}) = FP = 0;$$

$$TN = (\text{брой True Negative}) = (Q-1) \cdot (TP) = Q-1, TP = (\text{брой True Positive}) = 1.$$

Наистина, за всеки клас (M_i) ние генерираме единствен, но случаен, тестов пример:

$$S_i = M_i + \varepsilon = M(x_i, y_i, z_i) + \varepsilon(x, y, z), i = 1, \dots, Q,$$

където ε е случаен 3D вектор с дадено „псевдо-равномерно” разпределение, с отклонение Δ по всяка една от трите координати. Ние търсим максималното допустимо Δ , където Δ съответства на използваната точност на сканиране на входните обекти (за разпознаване и/или зареждане (фиксирано обучение) на 3DEarDB).

В същия дух, за всеки един клас (M_i), $i = 1, \dots, Q$, можем да запишем:

$$(TP + FN) = P = (\text{брой Positive}) = |\{S_i | S_i = M_i + \varepsilon\}| = |\{\varepsilon\}| = 1, \text{ и}$$

$$(TN + FP) = N = (\text{брой Negative}) = |\bigcup_{j \neq i} \{S_j | S_j = M_j + \varepsilon\}| = (Q - 1) |\{\varepsilon\}| = (Q - 1).$$

На практика, за да оценим „поведението” на оптималната (допустимата максимална) стойност на Δ , виж и Табл. 16, ние варираме основните параметри на експеримента – за двата вида разстояние (25а) и (25б), за 3 различни резолюции на 3D моделите и за 4 последователни нива на EGI дискретизация. Същевременно, вариациите за Δ организираме по принципа „увеличаваме Δ до първо срещане на $FN > 0$ или $FP > 0$ ”, плюс няколко опита наоколо. Така, поради споменатите вариации, можем да смятаме, че за всеки клас (M_i) генерираме тестово множество

$$\{S_i\} = \{S_i | S_i = M(x_i, y_i, z_i) + \varepsilon(x, y, z)\}, \text{ с мощност } |\{S_i\}| = C, i = 1, \dots, Q,$$

където константата C отразява броя на комбинациите от вариации на параметрите, които се обединяват (сумират), но очевидно без бройката на използваните видове разстояния, т.е. $1 \leq C \leq 36 = 3 * 4 * (2 + 1)$. Така, например, за всяка една клетка от Табл. 16, имаме $C = 1$, за всеки ред от нея $C = 9$, докато за двете половинки (т.е. всички четни и всички нечетни редове) на Табл. 16 ще имаме $C = 36$.

³⁹ [https://en.wikipedia.org/wiki/Cross-validation_\(statistics\)](https://en.wikipedia.org/wiki/Cross-validation_(statistics))

⁴⁰ https://en.wikipedia.org/wiki/Confusion_matrix

Или сумарно, по всички класове, поради симетричността, можем да запишем:

$$P = (TP + FN) = C,$$

$$N = (TN + FP) = C(Q-1).$$

Така, използваната от нас (единствена и достатъчна в нашия случай) характеристика TRR (True Recognition Rate), виж Табл. 16, може да се изрази с популярната напоследък характеристика “accuracy”:

$$TRR = \frac{\text{брой правилно разпознати изходи}}{\text{брой експерименти}} = \frac{\sum_{i=1}^Q (TP_i + TN_i)}{\sum_{i=1}^Q C \cdot Q} = \frac{\sum_{i=1}^Q TP_i + \sum_{i=1}^Q TN_i}{C Q^2} = \frac{TP + TN}{P + N} = \text{accuracy}, \quad (26)$$

където TP_i и TN_i са съответните стойности за под-експериментите (C на брой) за даден клас (M_i), $i = 1, \dots, Q$, а $TP = \sum_{i=1}^Q TP_i$ и $TN = \sum_{i=1}^Q TN_i$ са сумарните стойности за целия експеримент, който се състои от $P + N = C Q^2$ единични под-експеримента.

В нашата интерпретация на абревиатурата TRR, се акцентира върху Recognition (разпознаването), за разлика от други източници, където TRR означава True Reject Rate.

3.3.1 Анализ на експерименталните резултати

- ❖ Проведените експерименти върху текущата 3DEarDB (от 100 3D модела на уши) потвърждават възможността за използване на EGI представянето за недвусмислена идентификация на уши, независимо че тяхната повърхнина е съставена от комбинация между изпъкнали и вдлъбнати области. Това се потвърждава от резултатите за изследваните граници на зашумяване, които надвишават декларираната прецизност от 0.05 mm на 3D скенера VIUscan, виж най-левите колони на Табл. 16, където $TRR = 100\%$ за всяка от трите експериментирани резолюции.
- ❖ Както се очаква, базирайки се на предишни експерименти върху БД от 11 уши, Bray-Curtis разстоянието (E_{BC}) е по-устойчиво за съответното ниво на шум от Евклидовото разстояние (E_2), в резултат на което се постига по-висок TRR.
- ❖ За останалите резултати, при които $TRR < 100\%$, се наблюдава един „феномен“ – при ниво на шума по-високо от граничното (виж средните и най-десните колони от Табл. 16) се наблюдава неочакван спад на TRR при подобряване на EGI представянето ($20 \rightarrow 80 \rightarrow 320 \rightarrow 1280$) или на резолюцията на сканиране ($0.5 \leftarrow 1.0 \leftarrow 1.4$) за съответните нива на зашумяване.
- ❖ Този „феномен“ за TRR трябва да се разглежда извън основните положителни резултати за функционалността на 3DEarDB. Той може да се обясни освен с наличието на изпъкнали и вдлъбнати области в 3D моделите, още и с комбинацията на други нелинейности като: (i) неравномерната триангулация на 3D моделите; (ii) неравномерността на EGI апроксимацията; (iii) изглаждащи ефекти при софтуерната манипулация на резолюцията и др.
- ❖ Поради необходимостта за работа на алгоритъма за разпознаване на уши в реално време, си струва да се обърне по-сериозно внимание на споменатия „феномен“, защото намаляването на геометричната резолюция на 3D сканирането и намаляването на сложността на EGI представянето са необходими предпоставки за работата в реално време.

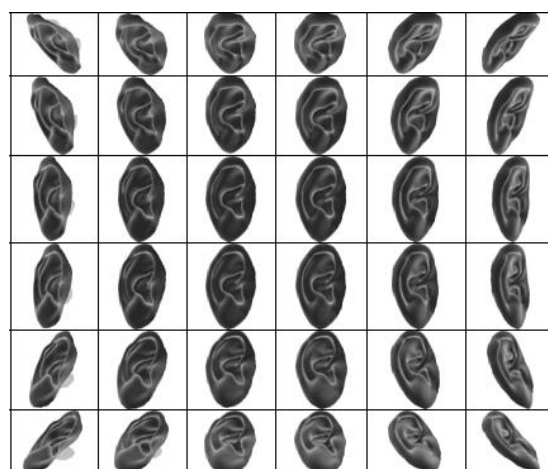
3.4 Разпознаване на уши чрез подхода „Многоизгледно 3D моделиране“⁴¹

Тримерното разпознаване на 3D обекти чрез т.нар. подход „Многоизгледно 3D моделиране“ (виж и §3.1.6) се базира на оригинален CBIR метод за достъп до БД с прецеденти [171]. Разработката, авторско дело на доц. д-р Димо Димов, е тествана в неговата експериментална софтуерна среда, наречена EFCIOR (Effective and Fast Content based Image or Object Retrieval), също резултат на дългогодишните му изследвания в областта. Съществена особеност на EFCIOR е нейното бързодействие, т.е. потенциална възможност за работа в „реално време“, защото времето за сравнение на входното изображение с всички записи от БД с изображения е от порядъка на $t \cdot \log_2 N$, където N е размерът на БД (в брой изображения), а t е средното време за достъп до HDD на компютъра.

В предишни версии на системата, а именно EFIRS, FANTIR и др., разработеният CBIR подход е бил прилаган успешно към разпознаването на жестове от азбуката на слухово увредени хора и към разпознаването на физиономии [172, 173]. В тези случаи, БД от изображения е била организирана чрез извличане на равномерно разположени 2D проекции от кратки видеоклипове, проследяващи 3D обекта в желания пространствен ъгъл [174]. За по-точното пространствено-времево „сканиране“ на представяните в БД обекти е била създадена специална въжено-механична конструкция.

Системата EFCIOR бе приложена успешно и за 3D разпознаване на уши [159], където проекциите на обектите (уши) са извлечени от техните 3D модели (получени с помощта на 3D скенера VIUscan на ИИКТ), което значително опрости процедурата по събиране на необходимите изображения (2D проекции) за БД. В същото време същинското разпознаване си остава базирано на входни изображения от конвенционални 2D камери.

Всеки 3D обект за разпознаване се представя в БД чрез достатъчен брой 2D проекции (изображения) на обекта от различни гледни точки. Необходимите представителни 2D проекции на 3D обекта от интерес трябва да са направени равномерно в експерименталния пространствен ъгъл на видимост на обекта, което е илюстрирано на Фиг. 62 (виж и Фиг. 51), за случая на човешки уши. Важно е да се отбележи, че броят на проекциите за всеки 3D обект, представен в БД, може да бъде значително намален от няколко стотин до няколко десетки изображения в зависимост от нивото на шумоустойчивост на реализирания CBIR подход. Реализираните CBIR подходи са 3, с по няколко версии все-



Фиг. 62. Равномерни 2D проекции, извлечени от 3D модела на дясното ухо на колега от ИИКТ.

⁴¹ Разглежданият метод не е разработен в рамките на текущата дисертация, но се разглежда тук от гледна точка на 3DEarDB, от която се генерират входните данни за зареждане на специфичната БД на EFCIOR.

ки. Шумоустойчивостта на системата покрива така наречените „артефактен“ и „регулярен“ шум в изображенията, а също и шума от произволни линейни трансформации на входа (транслация, ротация и мащабиране).

Връзката на системата EFCIOR с изследванията по дисертацията (виж Фиг. 2 от Увода) се състои в практическото използване на един от възможните изходни формати (2D проекции на уши) в създадената многомоделна 3DEarDB (виж §3.1) за разпознаване на хора по формата на ушите. Проведените експерименти с малка БД (с проекциите на 3D модели на уши на 10 човека) [159], докладват за постигнат TRR ~ 97-98%, което се смята за много обещаващо за бъдещото развитие на метода, представителен за областта „от 2D към 3D биометрика“.

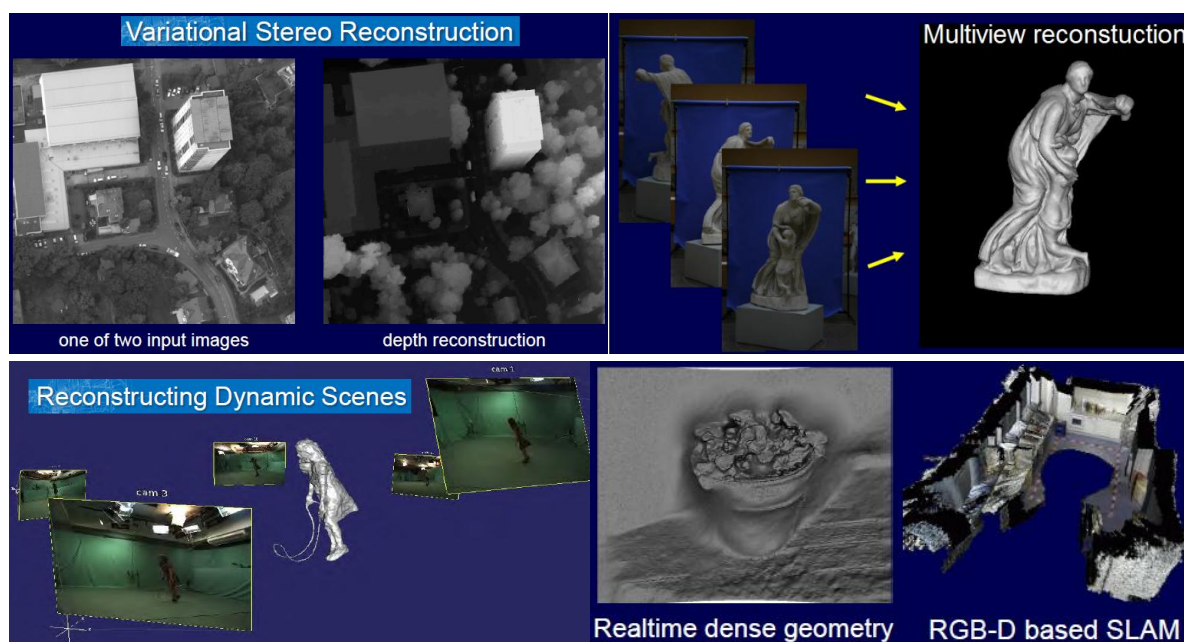
3.5 Обещаващи методи за 3D реконструкция на обекти в реално време за целите на разпознаването по уши

Най-често използваните методи за 3D реконструкция на обекти и сцени включват многоизгледно стерео (MVS) [175], сканиране чрез структурирана светлина [176], лазерно сканиране [177]. Най-общо казано, MVS подходите са нискобюджетни, но предлагат най-ниска точност в сравнение с другите: ~ 1 mm за 20 cm широк обект [175, 178]. Методите, използващи структурирана светлина, могат да постигнат по-висока точност – до 0.05 mm [179, 180] за относително кратко време. Лазерните скенери за близки разстояния могат да достигнат точност до 0.03 mm [177], но на цената на скъпо оборудване и дълго време за сканиране. Едва по-новите подходи за 3D реконструкция чрез MVS в комбинация с форма от засенчването (SfS) методи, демонстрират резултати, сравними с лазерното сканиране [181]. С развитието на 3D технологиите в последно време дори е възможно извършването на относително прецизна 3D реконструкция и в реално време [182, 183, 184].

Най-нови разработки в областта на 3D реконструкцията на обекти и сцени бяха докладвани на международно лятно училище по Компютърно зрение (ICVSS'15), проведено на 12-18.07.2015 г. в гр. Сампиери, Сицилия, Италия. Целта на училището е да осигури ясна представа и задълбочен анализ на новостите „state-of-the-art“ в областта на Компютърното зрение (Computer Vision) и Машинното обучение (Machine Learning). Лекторите са световно известни експерти, както от академичния, така и от индустриалния сектор в областта.

В рамките на две лекции (на *Daniel Cremers* от университета в Munich и на *Marc Pollefeys* от университета на ETH Zurich) бяха представени резултати от реално работещи системи за 3D реконструкция, както на PC, така и на мобилни телефони.

Daniel Cremers демонстрира възможностите на вариационните методи (като алтернатива на методите, използващи съответствия между характерни точки в изображения) за 3D реконструкция: (i) от стерео изображения, Фиг. 63а [185]; (ii) от много изгледи, Фиг. 63б [186]; (iii) на динамични сцени, Фиг. 63в [187]; (iv) на плътни сцени в реално време (24 fps за резолюция 480×360 пиксела), Фиг. 63г [188]; (v) на сцени в голям мащаб, Фиг. 63д [189]; (vi) на сцени в реално време, заснети от движещ се автомобил [190].



Фиг. 63. Вариационни методи за 3D реконструкция: а) от стерео изображения; б) от множество изгледи; в) на динамични сцени; г) на плътни обекти; д) на големи сцени.

Той също демонстрира и тяхната система “*CopyMe3D*” (Фиг. 64) за 3D моделиране на обекти в реално време, работеща на PC с помощта на Kinect дълбочинен сензор [191], където сканирането на обекта отнема около 10 сек. (за 300 кадъра), докато пълната обработка и извеждането на резултата типично отнемат около 1 минута на PC от среден клас.



Фиг. 64. 3D моделиране на обекти в реално време чрез системата “*CopyMe3D*”.

Marc Pollefeys представи техният резултат (Фиг. 65) на една от първите в света системи за автоматично извличане на детайлни 3D модели от видеоклип (множество изображения от различни гледни точки). Първата им имплементация била готова още през 1997 г., но през 2004 г. публикуват резултатите след направени подобрения [192]. Реконструкцията там се базира на класическия подход със следене на характерни точки в множество кадри за възстановяване на 3D дълбочината на точките, както и на позицията и ориентацията на каме-



Фиг. 65. 3D реконструкция на главата на Medusa чрез обхождащо видео.

рата в пространството. Беше демонстрирано и моделиране на градове (Фиг. 66) в реално време (30 fps) от видео клип, чрез GPU имплементация на PC [193].



Фиг. 66. 3D моделиране на сцени в голям мащаб в градски условия.

През 2013 г. техният колектив представя разработка за превръщането на мобилните телефони в 3D скенери [194], с приложение към заснемане на културно наследство, идеята на 3D „селфита“ и др. (виж Фиг. 67).



Фиг. 67. 3D реконструкция чрез използването на изображения или видео от мобилни телефони.

Последни разработки на неговата група са 3D картиране на маршрута за автономно шофиране [195], автономна навигация и картиране за летящ дрон – без лазер, GPS или мрежова връзка [196].

Съществуват и други системи за 3D възстановяване на обекти с конкретно приложение в биометриката, като мобилното приложение за 3D реконструкция на лица от видеоклип (Фиг. 68) за многомодална идентификация [197].



Фиг. 68. 3D реконструкция на физиономии, заснети чрез предната камера (0.3 мегапиксела) на Google Nexus S смартфон: а) многоизгледна 3D реконструкция на лице; б) детекция и сегментация на уши от 3D профил на лице.

Поради недостижимия бум в областта на плътната 3D реконструкция на обекти и сцени, работеща в реално време на РС и дори на мобилни телефони, разработваният от нас метод, базиран на стереозрение, за определяне на дълбочина до определени точки от обекти [198], морално остаря и не беше доразвит в рамките на текущата дисертация. Споменаваме го във връзка с 3D реконструкцията на уши от обхождащ видеоклип (виж. Фиг. 2 от Увода), която не е обезателно необходима да бъде плътна, а само набор от 3D координати на разрежено множество от характерни точки би било достатъчно за разпознаването чрез изследвания в дисертацията EGI метод. Тази разработка е оставена за бъдеща работа.

Друг евентуален подход за разпознаване на уши в контекста на EGI дескриптора са методите, базирани на фотометрично стерео [199, 200, 201], където се възстановяват 3D нормалите към 2D координатите на дадено изображение. Известно е, че за недвусмисленото определяне на нормалите са необходими поне 3 статични (съвпадащи) изображения, заснети с осветление от различна гледна точка, което трябва да бъде ясно дефинирано като посока и интензитет. В тази връзка (фотометричното стерео) съществува и разработено мобилно приложение за Apple iPhone и iPad, наречено *Trimensional*⁴², за груба 3D реконструкция на обекти (физиономии). Методът създава 3D модели чрез осветяване на физиономията (на тъмно), на разстояние около 20 см, от различни позиции чрез вградени светодиоди в екрана на устройството и синхронно заснемане чрез предната камера.

Недостатък на подходите, използващи принципите на фотометричното стерео, е необходимостта от прецизна информация за източника на светлина, както и невъзможността за работа на дневна светлина. Един по-практичен подход би бил възстановяване на желаните нормали от единствено изображение, които ще се използват за вход към EGI метода за разпознаване. Такава идея също е предмет на по-задълбочени бъдещи изследвания – теоретичният модел е разработен, но предстои да бъде експериментиран.

3.6 Заключение и идеи за бъдеща работа

В текущата Глава 3 от дисертацията е описана нашата многомоделна БД от уши, наречена 3DEarDB, събрана в ИИКТ-БАН. Тя е съставена от различни съответстващи си множества от представяния на ушите на около 100 индивиди, продуцирани чрез различни устройства за заснемане: 3D лазерен скенер, Kinect сензор (за *бъдеща работа*) и цифрова фотокамера. 3DEarDB се отличава от известните подобни БД със своята пълнота от различни изходни формати на заснетите уши: 3D мрежести (триангулационни) модели, 2D дълбочинни изображения, 2D видеоклипове, 2D интензитетни проекции. По тази причина, 3DEarDB би могла да бъде полезна за общността по биометрика за сравнителен анализ на голямото разнообразие от 2D/3D подходи за разпознаване, както и на нови такива, базирани на реалната 3D информация от триангулационните модели.

Резултатите от проведените експерименти върху собствената БД от 100 3D модела на уши (3DEarDB) потвърждават възможността за използване на класическото разширено Гаусово изображение (EGI) за недвусмислена идентификация на уши, независимо

⁴² <http://www.trimensional.com/>

че тяхната повърхнина е съставена от изпъкнали и вдлъбнати области, което можем да обясним с факта, че ушите са достатъчно „сплеснати“ 3D обекти.

С цел ускоряване на процедурата по заснемане (текущо отнема около 10 мин. на ухо) за в близко бъдеще, освен използването на Kinect устройство, планираме използването на 3D скенер, работещ на принципа на структурирана светлина, може би на цената по-ниска прецизност.

Разгледана е връзката с EFCIOR системата в контекста на разпознаването на обекти (уши) по множество от техни 2D проекции (един от изходните формати на многомоделната 3DEarDB), като един бърз и шумоустойчив метод, представителен за областта „от 2D към 3D биометрика“.

Направен е преглед на най-съвременните разработки в областта на 3D реконструкцията на обекти и сцени (работещи в реално време). Очертани са насоките на *бъдещата работа* по дисертацията във връзка със стереоскопичното определяне на дълбочината до определени (характерни) точки от обекта на интерес. Също се предвиждат и *бъдещи* разработки по възстановяване на нормалите към пикселите на 2D изображения, било чрез методите на фотометричното стерео или (за предпочитане) от единствено изображение с цел осигуряване на вход за EGI подхода за разпознаване на 3D обекти.

Заклучение

Резюме на постигнатото и идеи за бъдеща работа

Предложен е нов ефективен и робастен софтуерен метод за 2D видеоостабилизация по гладка траектория, подходящ за работа в реално време на преносими устройства (виж Глава 2, §2.1). Тук съществено се емулира BSC чипът на Texas Instruments за търсене на SAD корелации между съответните хоризонтални и вертикални 1D акумулиращи проекции между два кадъра. Методът е базиран на прецизен векторен модел, позволяващ нарастващи по сложност интерпретации на движението, от твърдотелни модели на трансформация между кадрите – до модели на *“rolling shutter”* ефекта, характерен за болшинството портативни устройства като „умни“ телефони, таблети и др. Извършените експериментални изследвания върху множество видеоклипове с различна степен на потрепване доказват ефективността на предложените алгоритми в зависимост от търсеното бързодействие (включвайки и времената на външните за методите операции), което варира от 18 до 24 fps (виж Глава 2, §2.1.5). За бъдещото развитие на метода се предвижда използването му (за подобряване на точността), синхронизирано със системата от инерциални сензори (жироскопи, акселерометри), вградени в съответното портативно устройство, на които се разчита за една ефективна хардуерна видеоостабилизация.

Разработен е комбиниран метод $3 \times 3 \text{OF9} \times 9$, предназначен за видеоостабилизация в „точка“ на видеоклипове от високоскоростни индустриални камери (виж Глава 2, §2.2). Целта на метода е осигуряването на статичен фон, необходим за провеждането на коректни измервания на определени характеристики на динамичните обекти във видеоклипа. Предложеният комбиниран подход с детекция и отстраняване на подвижните обекти в сцената е експериментиран успешно върху публични и собствени видеоклипове и се отличава с достатъчно бързодействие и прецизност на стабилизацията. Характерният ефект на „пъзлене на сцената“, който е налице дори и при комерсиалните софтуерни решения, е редуциран до под 1% от размера (640×480 пиксела) на кадрите, докато чистото време за изпълнение на алгоритъма е в рамките на около 30 fps. Бъдещото усъвършенстване на метода ще бъде насочено към разработването на по-ефективна сегментация за отстраняване на влиянието на големи подвижни обекти в заснеманата сцена.

Предложен е относително прост и бърз (предвид работа в реално време) алгоритъм за локализиране на предварително известни рекламни видеоклипове или произволни части от тях в TV видеопоток. Същността на алгоритъма се състои в разделяне на видеоклиповете на последователност от сцени (според нашата дефиниция за сцена) и представянето им като крайни редици от цели числа, представляващи продължителността (т.е. дължината в брой кадри) на всяка сцена. По този начин използваме сравнение на крайни целочислени редици вместо сравнение на последователности от JPEG изображения (кадри) от TV потока и потенциалните кандидати от множеството на рекламните видеоклипове. Т.е. конвенционалното 2D сравнение *кадър-по-кадър* се заменя със сравнението на *дължина-по-дължина* на сцени и само при „подозрителните“ сцени извърш-

ваме покадрови сравнения, от типа „едно-към-много“. Този алгоритъм изисква много по-малък брой 2D покадрови сравнения (в сравнение с другите разгледаните методи), което е съществено предимство в реалните случаи на значителен брой реклами за търсене в TV потока.

Разработена е многомодалната БД (3DEarDB), съдържаща различни съответстващи си моделни представяния от изходни формати на ушите на ~ 100 индивиди (виж Глава 3, §3.1). 3DEarDB е предназначена за научната общност по разпознаване на уши с цел тестване и сравнение на точността на алгоритмите върху различни възможни сценарии – от идеалния случай на прецизен (и статичен) 3D модел до по-реалистичния (и динамичен) случай на 2D видеоклип, и до отделни 2D изображения. Прецизността на наличните 3D модели на уши се дължи на използването на 3D лазерния скенер на ИИКТ, с 1 mm резолюция на сканиране между съседните 3D точки и с точност от 0.05 mm за всяка 3D точка. Предимството на 3DEarDB спрямо публично известните БД от 2D/3D изображения на човешки уши се състои в наличието на прецизни мрежести (триангулационни) 3D модели на уши, а не само дълбочинни 2D изображения, както и в консистентното съответствие между множеството от 2D и 3D изходни формати. Проведените експерименти (върху 3DEarDB със 100 3D модела на уши) потвърждават възможността за недвусмисленото разпознаване на ушите (бидейки не изцяло изпъкнали обекти) чрез използването на конвенционалния EGI дескриптор за тяхното представяне във вид на хистограми. Бъдещата работа тук е насочена към разработването и експериментирането на една фотометрична идея за 3D разпознаване на уши (чрез EGI метода), но по единствено 2D изображение, чрез възстановяване на реалните нормални вектори към заснетия 3D обект.

Списък на публикациите по дисертацията

Списък на отпечатаните научни публикации:

1. Nikolov, A., D. Dimov: *2D Video Stabilization for Industrial High-Speed Cameras*, In: Cybernetics and Information Technologies, Special Issue on Information Fusion, Bulgarian Academy of Sciences, Vol. 15, No. 7, pp. 23-34, 2015, Sofia, DOI: 10.1515/cait-2015-0086, Print ISSN: 1311-9702, Online ISSN: 1314-4081.
2. Dimov, D., A. Nikolov: *Real Time Video Stabilization for Handheld Devices*, In: Rachev, B., A. Smrikarov (Eds.) Proceedings of CompSysTech'14, June 27, 2014, Ruse, Bulgaria, ACM ICPS, Vol. 883, pp. 124-133, ISBN: 978-1-4503-2753-4.
3. Cantoni, V., D. T. Dimov, and A. Nikolov: *3D Ear Analysis by an EGI Representation*, In: Cantoni, V., D. T. Dimov, and M. Tistarelli (Eds.) Proceedings of First International Workshop on Biometrics, BIOMET'2014, June 23-24, 2014, Sofia, Bulgaria, Biometric Authentication, LNCS, Vol. 8897, pp. 136-150, Springer, ISBN: 978-3-319-13385-0.
4. Nikolov, A., D. Dimov, V. Kolev, M. Ivanov, K. Ivanova, O. Kounchev, M. Bojkova, P. Mateev: *Effective recognition of the video pattern in a recorded video stream*, In: 104-th European Study Group with Industry (ESGI'104) - problems & final reports, Sept. 23-27, 2014, Sofia, Bulgaria, pp. 35-47, ISBN: 978-954-9526-87-5.
5. Nikolova, I., A. Nikolov, G. Zapryanov: *Depth estimation using shifted digital still camera*, In: Proc. of the 12th Int. Conference on Computer Systems and Technologies (CompSysTech'11), June 16-17, 2011, Vienna, Austria, ACM ICPS Vol. 578, pp. 234-240, ISBN: 978-1-4503-0917-2.

Списък на приетите за печат научни публикации:

6. Nikolov, A., V. Cantoni, D. Dimov, A. Abate, S. Ricciardi: *Multimodel Ear Database for Biometric Applications*, In: Proc. of International Conference Advanced Computing for Innovation, Nov. 10-11, 2015, Sofia, Bulgaria, ... (to be published as a special volume of Springer Studies in Computational Intelligence).

Цитирания на направените публикации по дисертацията:

1. Публикация №1: "*Real Time Video Stabilization for Handheld Devices*" е цитирана в: Aguilar, W. G., C. Angulo: *Real-Time Model-Based Video Stabilization for Microaerial Vehicles*, In Journal of Neural Processing Letters, pp. 1-19, 2015, Springer US, DOI: 10.1007/s11063-015-9439-0, Print ISSN: 1370-4621, Online ISSN: 1573-773X.
2. Публикация № 2: "*3D Ear Analysis by an EGI Representation*" е цитирана в: Pflug, A.: *Ear Recognition - Biometric Identification using 2- and 3-Dimensional Images of Human Ears*. PhD Thesis. Faculty of Computer Science and Media Technology, Gjøvik University College, Norway, 2015.
3. Публикация № 4: "*Depth estimation using shifted digital still camera*" е цитирана в: Duan, Y., H. Deng, F. Wang: *Depth Camera in Human-Computer Interaction: An Overview*, In: Proc. of Fifth International Conference on Intelligent Networks and Intelligent Systems (ICINIS), 1-3 Nov. 2012, Tianjin, pp. 25-28, DOI: 10.1109/ICINIS.2012.44, Print ISBN: 978-1-4673-3083-1.

Апробация на резултатите

Разработените алгоритми по стабилизация на видеоклипове и по откриване на рекламни сценарии в телевизионен видеопоток (описани в Глава 2), както и разработената БД от 3D модели на човешки уши (Глава 3, §3.1) са резултат от работата по следните проекти на ИИКТ-БАН:

1. Европейският проект **AComIn (Advanced Computing for Innovation) на ИИКТ** (по договор 316087, финансиран от FP7 Capacity Programme „Research Potential of Convergence Regions”), в рамките на който е използвано висококачественото оборудване от „Умната лаборатория“ (SmartLab), а именно: 3D лазерният скенер VIUScan и високоскоростната камера NAC Memrecam; както и многостранна подкрепа при разработването на тази дисертация; както и следващият:
2. Подпроект „Стабилизация на видео от високоскоростна камера“ по договор ДО1-192 на МОН „Съфинансиране на проект АКОМИН 316087 по 7РП на ЕС”; а също:
3. „Математически алгоритми за видео стабилизация и филтриране на шум от сензори за движение“. Договор за изследвания и разработка (№ 210214) между фирма „ММ Солушънс“ АД и ИИКТ от 07.05.2012 г. до 02.10.2012 г.;
4. Проект № BG161PO003-1.1.06-0038-C0001/07.12.2012г. „Експериментално развитие на софтуерни библиотеки за подобряване на качеството на изображението и стабилизация с използване на инерционни сензори”, Бенефициент: „ММ Солушънс“ АД, София, ул. Тинтява 15, Партньор: ИИКТ-БАН;
5. Мини проект за техническа изпълнимост по проблем, зададен от фирма Cross Agency Ltd., на тема: „Ефективно разпознаване на видео шаблони в поточно видео“, в рамките на организираната от ИИКТ-БАН, ФМИ-СУ и ИМИ-БАН, със съдействието на Европейския консорциум за математика в индустрията, „104-та Европейска изследователска група с индустрията“ (ESGI-113), 23-27 септ. 2014 г., София.

Участие в други проекти, подпомогнали разработването на дисертацията:

- Двустранен договор по програмата Lifelong Learning Programme: HIGHER EDUCATION (ERASMUS) за академичната година 2012/2013, между БАН (BG Sofia 30) и Университета на Павия, Италия (I Pavia 01);
- „Изграждане и развитие на млади висококвалифицирани изследователи за ефективно прилагане на биомедицинските изследвания за подобряване качеството на живот” по ОП „Развитие на човешките ресурси” (Европейски Социален Фонд 2007-2013) на МОН, договор BG051PO001-3.3.06-0048/04.10.2012 с координатор Институт по експериментална морфология, патология и антропология с музей – БАН;
- „Подкрепа на творческото развитие на докторанти, пост-докторанти и млади учени в областта на компютърните науки” по ОП „Развитие на човешките ресурси” (Европейски Социален Фонд 2007-2013) на МОН, договор BG051PO001-3.3.04/13 с координатор Великотърновският университет „Св.св. Кирил и Методий“.

Авторска справка за приносите в дисертацията

✓ Научни приноси

1. Разработен е оригинален векторен модел " $N \times N$ " ($N \geq 3$, нечетно) за представяне на случайни линейни трансформации между последователните кадри от видеопоток, с избор и настройка за сложността на интерпретацията, (Глава 2, §2.1.2 и §2.1.6).
2. Предложен е метод за преобразуване на видеопотоци от 2D интензитетни изображения до редици от цели числа, представляващи продължителността в брой кадри на всяка сцена, което значително опростява търсенето на подобия (Глава 2, §2.3.3-4).

✓ Научно-приложни приноси

1. Предложен е ефективен (бърз и прецизен) метод, основан на оригиналния векторен модел, за 2D видеостабилизация по гладка траектория, подходящ за вграждане в мобилни устройства, (Глава 2, §2.1.2-3).
2. Предложено е разширение на базовия векторен модел за корекцията на хоризонталното накланяне и вертикалното мащабиране, породени от така наречения "*rolling shutter*" ефект, присъщ за камерите със CMOS сензор, (Глава 2, §2.1.4).
3. Предложен е комбиниран метод " $3 \times 3 \text{ OF } 9 \times 9$ ", съчетаващ базовия " $N \times N$ " метод с известния метод на „оптичния поток“ за прецизна стабилизация в „точка“ с приложение при стабилизацията на клипове от високоскоростни камери, (Глава 2, §2.2.3).
4. Предложен е прецизен, относително прост и бърз (за работа в реално време) алгоритъм за локализиране на предварително известни рекламни клипове и/или произволни части от тях в базов TV видеопоток, (Глава 2, §2.3.3).
5. Разработена е структурата на многомоделна база данни, наречена 3DEarDB, съдържаща 3D модели, както и производни представяния на дясното ухо на индивиди за целите на биометричната автентикация, (Глава 3, §3.1).

✓ Приложни приноси

1. Извършени са множество експерименти за стабилизация по гладка траектория върху видеоклипове с нарастваща нестабилност на видеокамерата, (Глава 2, §2.1.5).
2. Извършени са множество от експерименти за стабилизация в „точка“ върху публични и собствени видеоклипове, (Глава 2, §2.2.4).
3. Експериментиран е прост подход за детекция на сцени, базиран на диференциалното представяне на видеоклиповете, (Глава 2, §2.3.3).
4. Експериментално е потвърдена възможността за използване на конвенционалното EGI представяне на 3D обекти, за недвусмисленото разпознаване на изпъкнало-вдлъбнати обекти, за случая на човешки уши, (Глава 3, §3.2).
5. Потвърдена е функционалността и е оценена шумоустойчивостта на разпознаване чрез 3DEarDB по метода на най-близкия съсед върху тестово множество от зашумени версии на 3D моделите в 3DEarDB, чийто брой ~ 100 се смята експертно за статистически значим, (Глава 3, §3.3).

Декларация за оригиналност на резултатите

Декларирам, че настоящата дисертация съдържа оригинални резултати, получени при проведени от мен научни изследвания {с подкрепата и съдействието на научния ми ръководител}. Резултатите, които са получени, описани и/или публикувани от други учени, са надлежно и подробно цитирани в библиографията.

Настоящата дисертация не е прилагана за придобиване на научна степен в друго висше училище, университет или научен институт.

Подпис:

Благодарности

Изказвам своята дълбока благодарност и признателност към моя научен ръководител доц. д-р Димо Димов за непрекъснатото и неуморно провокиране на научните ми интереси в сферата на Компютърното зрение и математиката и в частност на Обработката на сигнали (изображения) и разпознаването на образи. През годините на докторантското ми обучение (в бившата секция ОСРО на ИИКТ-БАН) доц. Димов прояви завидно търпение, упоритост и настойчивост при разширяването на миогледа ми относно компютърната и инженерната наука. Значим е и опитът, който придобих по време на участието ми в най-разнообразни научно-изследователски проекти и международни конференции. Също така доц. Димов допринесе за създаването на международните ми контакти при сътрудничеството ни с проф. Вирджинио Кантони и неговия екип от лабораторията по Компютърно зрение (CVML) на университета на Павия, Италия.

Благодаря на колегите Николай Стоименов и Иван Георгиев за оказаното съдействие при провеждането на експериментите по сканирането на уши с 3D лазерния скенер, както и на колегата Станислав Гьошев за заснемането на клипове с високоскоростната камера, закупени за „умната лаборатория“ по проекта AComIn на ИИКТ-БАН. Благодаря и на всички колеги (от ИИКТ, ИМИ, СУ) за участието им в създаването на БД от 3D модели на уши.

Благодаря и на проф. Галя Ангелова (координатор на проекта AComIn) за предоставената възможност за работа по подпроекта „Стабилизация на видео от високоскоростна камера“ по договор ДО1-192 на МОН „Съфинансиране на проект АКОМИН 316087 по 7РП на ЕС“. Също така, благодарение на проекта AComIn, бях командирован на два пъти в университета на Павия, където установих изключително ползотворни приятелски и научни контакти с колектива от CVML Lab. От значение е и участието ми в престижното международно училище по Компютърно зрение (ICVSS'15) в Италия, където придобих най-актуални знания за съвременното състояние на областта по разпознаването на образи и 3D реконструкцията на обекти и сцени.

Благодаря и за участието ми в двата проекта по Оперативната програма „Развитие на човешките ресурси“ с координатори съответно: Великотърновският университет и Институтът по експериментална морфология, патология и антропология с музей – БАН.

Изобщо, годините, прекарани в ИИКТ-БАН, въпреки многообразието от трудности и препятствия от най-различно естество, се отразиха изключително въздействащо и ползотворно върху представите и разбиранията ми относно научната работа и перспективите, които тя дава на истински отдадения последовател.

Благодаря и на моето семейство: на М. Момчилова, на Ф. Николов и на Д. Момчилов, за подкрепата в моментите на изпитания по време на докторантското ми следване. Също, благодаря и на М. Петрова и К. Петров за отзивчивостта и добротата.

Настоящия дисертационен труд посвещавам на доц. д-р Димчо Момчилов.

Литература

- [1] Grundmann, M., Kwatra, V., Essa, I.: Auto-Directed Video Stabilization with Robust L1 Optimal Camera Paths. Proc. of Conference on Computer Vision and Pattern Recognition (CVPR'11), pp. 225-232 (2011).
- [2] Liu, F., Gleicher, M., Wang, J., Jin, H., Agarwala, A.: Subspace Video Stabilization. ACM Transactions on Graphics, Vol. 30, Issue 1, pp. 4:1-4:10 (2011).
- [3] Wang, Y.-S., Liu, F., Hsu, P.-S., Lee, T.-Y.: Spatially and Temporally Optimized Video Stabilization. IEEE Transactions on Visualization and Computer Graphics, Vol. 19, Issue 8, pp. 1354-1361 (2013).
- [4] Morimoto, C., Chellappa, R.: Evaluation of Image Stabilization Algorithms. In DARPA Image Understanding Workshop DARPA97, pp. 295-302 (1997).
- [5] Matsushita, Y., Ofek, E., Ge, W., Tang, X., Shum, H.-Y.: Full-Frame Video Stabilization with Motion Inpainting. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, Issue 7, pp. 1150-1163 (2006).
- [6] Lukac, R.: Single-Sensor Imaging: Methods and Applications for Digital Cameras. CRC Press, Inc., Boca Raton, FL, USA, 2008.
- [7] Gleicher, M. L., Liu, F.: Re-Cinematography: Improving the Camerawork of Casual Video. ACM Transactions on Multimedia, Vol. 5, Issue 1, pp. 1-28 (2008).
- [8] Buehler, C., Bosse, M., Mcmillan, L.: Nonmetric Image-Based Rendering for video Stabilization. Proc. of Conf. on Comp. Vision and Pattern Recog. (CVPR), pp. 609-614 (2001).
- [9] Hartley, R. I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, 2000.
- [10] Fitzgibbon, A., Wexler, Y., Zisserman, A.: Image-Based Rendering using Image-Based Priors. Int. Journal of Computer Vision, Vol. 63, Issue 2, pp. 141-151 (2005).
- [11] Bhat, P., Zitnick, C. L., Snavely, N., Agarwala, A., Agrawala, M., Cohen, M., Curless, B., Kang, S. B.: Using Photographs to Enhance Videos of a Static Scene. In Rendering Techniques 2007: 18th Eurographics Workshop on Rendering, pp. 327-338 (2007).
- [12] Zhang, G., Hua, W., Qin, X., Shao, Y., Bao, H.: Video Stabilization based on a 3D Perspective Camera Model. The Visual Computer, Vol. 25, Issue 11, pp. 997-1008 (2009).
- [13] Liu, F., Gleicher, M., Jin, H., Agarwala, A.: Content-Preserving Warps for 3D Video Stabilization. ACM Transactions on Graphics, Vol. 28, Issue 3, pp. 44:1-44:9 (2009).
- [14] Liang, C. K., Chang, L. W., Chen, H. H.: Analysis and Compensation of Rolling Shutter Effect. IEEE Transactions on Image Processing, Vol. 17, Issue 8, pp. 1323-1330 (2008).
- [15] Baker, S., Bennett, E., Kang, S. B., Szeliski, R.: Removing Rolling Shutter Wobble. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2392-2399 (2010).
- [16] Forssen, P.-E., Ringaby, E.: Rectifying Rolling Shutter Video from Hand-Held Devices. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 507-514 (2010).
- [17] Goldstein, A., Fattal, R.: Video Stabilization using Epipolar Geometry. ACM Transactions on Graphics, Vol. 31, Issue 5, pp. 126:1-126:10 (2012).
- [18] Shuaicheng, L., Yuan, L., Tan, P., Sun, J.: Bundled Camera Paths for Video Stabilization. ACM Transactions on Graphics, Vol. 32, Issue 4, Article 78, 10 pages (2013).
- [19] Shuaicheng, L., Yuan, L., Tan, P., Sun, J.: SteadyFlow: Spatially Smooth Optical Flow for Video Stabilization. IEEE Conference on CVPR, 23-28 June, pp. 4209-4216 (2014).
- [20] Wang, Y.-S., Liu, F., Hsu, P.-S., Lee, T.-Y.: Spatially and Temporally Optimized Video Stabilization. IEEE Transactions on Visualization and Computer Graphics, Vol. 19, Issue 8, pp. 1354-1361 (2013).
- [21] Nister, D., Naroditsky, O., Bergen, J.: Visual Odometry. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 652-659 (2004).
- [22] Davison, A. J., Reid, I. D., Molton, N. D., Stasse, O.: MonoSLAM: Real-Time Single Camera SLAM. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, Issue 6, pp. 1052-1067 (2007).

- [23] Igarashi, T., Moscovich, T., Hughes, J. F.: As-rigid-as-possible Shape Manipulation. *ACM Transactions on Graphics*, Vol. 24, Issue 3, pp. 1134-1141 (2005).
- [24] Schaefer, S., Mcphail, T., Warren, J.: Image Deformation using Moving Least Squares. *ACM Transactions on Graphics*, Vol. 25, Issue 3, pp. 533-540 (2006).
- [25] Meingast, M., Geyer, C., Sastry, S.: Geometric Models of Rolling-Shutter Cameras. *Proc. of the 6th International workshop on Omnidirectional vision, Camera networks, and non-classical cameras*, pp. 12-19 (2005).
- [26] Shi, J., Tomasi, C.: Good Features to Track. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593-600 (1994).
- [27] Sinha, S., Frahm, J.-M., Pollefeys, M., Genc, Y.: GPU Based Video Feature Tracking and Matching. *Proc. of the Workshop on Edge Comp. Using New Commodity Architectures*, 2006.
- [28] Kim, S.-J., Koh, K., Boyd, S., Gorinevsky, D. L1 Trend Filtering. *SIAM Review*, 2009.
- [29] Laveau, S., Faugeras, O.: 3-d Scene Representation as a Collection of Images. *Proc. of the 12th IAPR International Conference on Pattern Recognition*, Vol. 1, pp. 689-691 (1994).
- [30] Hanning, G., Forslow, N., Forssen, P.-E., Ringaby, E., Tornqvist, D., Callmer, J.: Stabilizing Cell Phone Video using Inertial Measurement Sensors. *IEEE Int. Conf. on Computer Vision Workshops*, pp. 1-8 (2011).
- [31] Jia, C., Evans, B. L.: Probabilistic 3-D Motion Estimation for Rolling Shutter Video Rectification from Visual and Inertial Measurements. *Proc. of IEEE Multimedia Signal Processing Workshop*, The University of Texas at Austin, 2012.
- [32] Karpenko, A., D. Jacobs, J. Baek, and M. Levoy, Digital Video Stabilization and Rolling Shutter Correction using Gyroscopes, *Stanford University Computer Science Tech Report 2011-03*.
- [33] Fitzgerald, J. A., Translational Motion Stabilization for Embedded Video Applications, *White Paper SPRAAX8*, Texas Instruments, 2008.
- [34] TMS320DM365 Video Stabilization Algorithm Datasheet, Texas Instruments, January 2010.
- [35] Hong, W., Wei, D., Batur, A.U.: Video Stabilization and Rolling Shutter Distortion Reduction. *Proc. of 17th IEEE Int. Conf. on Image Processing (ICIP)*, pp. 3501-3504 (2010).
- [36] Hong, W., Wei, D., Batur, A.U.: Video Stabilization and Reduction of Rolling Shutter Distortion, *US Patent Application Publication*, No.: US 2011/0176014 A1, Date: Jul. 21, 2011.
- [37] Batur, A. U., Flinchbaugh, B.: Video Stabilization with Optimized Motion Estimation Resolution. *Proc. of IEEE Int. Conf. on Image Processing*, pp. 465-468 (2006).
- [38] Wayman, J. L.: Biometric Verification / Identification / Authentication / Recognition: The Terminology. In: Li, S. Z., Jain, A. K. (Eds.) *Encyclopedia of Biometrics*, pp. 153-157. Springer, Heidelberg (2009).
- [39] Ross, A., Jain, A. K.: Biometrics, Overview. In: Li, S. Z., Jain, A. K. (Eds.) *Encyclopedia of Biometrics*, pp. 168-172. Springer, Heidelberg (2009).
- [40] Jain, A. K., Nandakumar, K.: Biometric System Design, Overview. In: Li, S. Z., Jain, A. K. (Eds.) *Encyclopedia of Biometrics*, 2nd Edition, pp. 229-235. Springer, Heidelberg (2015).
- [41] Jain, A. K., Ross, A., Nandakumar, K.: *Introduction to Biometrics*. Springer, 2011.
- [42] Cantoni, V., Dimov, D. T., Nikolov, A.: 3D Ear Analysis by an EGI Representation, In: Cantoni, V., D. T. Dimov, and M. Tistarelli (Eds.) *Proceedings of First International Workshop on Biometrics, BIOMET'2014*, June 23-24, 2014, Sofia, Bulgaria, Biometric Authentication, LNCS, Vol. 8897, pp. 136-150. Springer (2014).
- [43] Adler, A., Schuckers, S. A. C.: Security and Liveness, Overview. In: Li, S. Z., Jain, A. K. (Eds.) *Encyclopedia of Biometrics*, 2nd Edition, pp. 1335-1341. Springer, Heidelberg (2015).
- [44] Chellappa, R., Sankaranarayanan, A. C.: Surveillance. In: Li, S. Z., Jain, A. K. (Eds.) *Encyclopedia of Biometrics*, 2nd Edition, pp. 1511-1515. Springer, Heidelberg (2015).
- [45] Zhang, D., Lu, G.: *3D Biometrics - Systems and Applications*. Springer, 2013.
- [46] Barbu, T.: Comparing Various Voice Recognition Techniques. *Proc. of the 5th Conf. on Speech Technology and Human-Computer Dialogue*, pp. 1-6 (2009).
- [47] Houmani, N., et al.: BioSecure Signature Evaluation Campaign (BSEC'2009): Evaluating Online Signature Algorithms Depending on The Quality of Signatures. *Journal of Pattern Recognition*, Vol. 45, Issue 3, pp. 993-1003 (2012).
- [48] Maltoni, D., Maio, D., Jain, A. K., Prabhakar, S.: *Handbook of Fingerprint Recognition*. 2nd Edition. Springer, 2009.

- [49] Yoon, S.: Fingerprint Recognition: Models and Applications, Ph.D. Thesis, 2014.
- [50] Li, S. Z., Jain, A. K.: Handbook of Face Recognition. 2nd Edition. Springer, 2011.
- [51] Rowden, L. B., Jain, A. K.: A Longitudinal Study of Automatic Face Recognition. ICB, Phuket, Thailand, May 19-22, 2015.
- [52] Bowyer, K.W., Hollingsworth, K., Flynn, P.J.: Image Understanding for Iris Biometrics: a Survey. Computer Vision and Image Underst, Vol. 110, Issue 2, pp. 281-307 (2008).
- [53] Rathgeb, Ch., Uhl, A., Wild, P.: Iris Biometrics - from Segmentation to Template Security. Springer, 2013.
- [54] Zhang, D., Kong, W., You, J., Wong, M.: Online Palmprint Identification. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 25, Issue 9, pp. 1041-1050 (2003).
- [55] Sanchez-Reillo, R., Sanchez-Avilla, C., Gonzalez-Marcos, A.: Biometric Identification Through Hand Geometry Measurements. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 22, Issue 10, pp. 1168-1171 (2000).
- [56] Pflug, A., Busch, C.: Ear Biometrics: A Survey Of Detection, Feature Extraction and Recognition Methods. IET Biometrics, Vol. 1, Issue 2, pp. 114-129 (2012).
- [57] Abaza, A., Ross, A., Hebert, Ch., Harrison, M., Nixon, M.S.: A Survey on Ear Biometrics. Journal of ACM Computing Surveys, Vol. 45, Issue 2, Article No. 22, 35 pages (2013).
- [58] Barra, S., Marsico, M. D., Nappi, M., Riccio, D.: Unconstrained Ear Processing: What is Possible and What Must Be Done. In: Scharcanski, J., H. Proença, and E. Du (Eds.) Signal and Image Processing for Biometrics. LNEE, Vol. 292, pp. 129-190. Springer (2014).
- [59] Kakadiaris, I.A., Passalis, G., Toderici, G., Murtuza, M.N., Lu, Y.L., Karampatziakis, N., Theoharis, T.: Three-Dimensional Face Recognition in The Presence of Facial Expressions: An Annotated Deformable Model Approach. IEEE Trans. on Pattern Analysis and Machine Intell., Vol. 29, Issue 4, pp. 640-649 (2007).
- [60] Abate, A. F., Nappi, M., Riccio, D., Sabatino, G.: 2D and 3D Face Recognition: A Survey. Pattern Recognit Letters, Vol. 28, Issue 14, pp. 1885-1906 (2007).
- [61] Woodard, D.L., Faltemier, T.C., Yan, P., Flynn, P.J., Bowyer, K.W.: A Comparison of 3D Biometric Modalities. Conf. on Comp. Vision and Pattern Recognition Workshop, pp. 57-60 (2006).
- [62] Zhang, D., Lu, G., Li, W., Zhang, L., Luo, N.: Palmprint Recognition Using 3-D Information. IEEE Trans. on Systems, Man, and Cybernetics, Part C, Vol. 39, Issue 5, pp. 505-519 (2009).
- [63] Chen, H., Bhanu, B.: Contour Matching for 3D Ear Recognition. Proc. of 7th IEEE Workshops on Applications of Computer Vision, Vol. 1, pp. 123-128 (2005).
- [64] Chen, H., Bhanu, B.: Human Ear Recognition in 3D. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 29, Issue 4, pp. 718-737 (2007).
- [65] Yan, P., Bowyer, K.W.: Biometric Recognition Using 3D Ear Shape. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 29, Issue 8, pp. 1297-1308 (2007).
- [66] Prakash, S.: Human Recognition using 2D and 3D Ear Images. PhD Thesis. Department of Computer Science and Engineering, Indian Institute of Technology, Kanpur, India, 2012.
- [67] Prakash, S., Gupta, P.: Ear Biometrics in 2D and 3D. Springer, 2015.
- [68] Pflug, A.: Ear Recognition - Biometric Identification using 2- and 3-Dimensional Images of Human Ears. PhD Thesis. Faculty of Computer Science and Media Technology, Gjøvik University College, Norway, 2015.
- [69] Rhodes, H. T. F.: Alphonse Bertillon: Father of Scientific Detection. Abelard-Schuman, New York, 1956.
- [70] Jain, A. K., Kumar, A.: Biometric Recognition: An Overview. In: Mordini, E., Tzovaras, D. (Eds.) Second Generation Biometrics: The Ethical, Legal and Social Context. Springer, 2012.
- [71] Turk M., Pentland, A.: Eigen Faces for Recognition. Journal of Cognitive Neuroscience, Vol. 3, No.1, MIT, pp. 71-86 (1991).
- [72] Biometrics History. National Science and Technology Council (NSTC), 7 August, 2006, <http://www.biometrics.gov/ReferenceRoom/Introduction.aspx>
- [73] New Biometric Technology Improves Security and Facilitates U.S. Entry Process for International Travelers. US Department of Homeland Security, https://www.dhs.gov/xlibrary/assets/usvisit/usvisit_edu_10-fingerprint_consumer_friendly_content_1400_words.pdf
- [74] Fast Border Passage with Iris Scan. Schiphol, <http://www.schiphol.nl/Travellers/AtSchiphol/Privium/PriviumBenefits/FastBorderPassage.htm>

- [75] Introducing SmartGate. Australian Government, http://www.customs.gov.au/webdata/resources/files/BR_introSmrtGt0409.pdf
- [76] Walt Disney World Resort, <http://disneyworld.disney.go.com/>
- [77] Jain, A. K., Ross, A., Prabhakar, S.: An Introduction to Biometric Recognition. IEEE Trans. on Circuits and Systems for Video Technology, Vol. 14, pp. 4-20 (2004).
- [78] Ross, A.: An Introduction to Multibiometrics. Proc. of the 15th European Signal Processing Conference (EUSIPCO), Poznan, Poland (2007).
- [79] Jain, A. K.: Biometrics - Technology for Human Recognition. Noblis Technology Tuesdays Special Presentation, Noblis, Falls Church, VA, November 18th, 2014, <http://biometrics.cse.msu.edu/pres/index.html>
- [80] Wang, J., She, M., Nahavandi, S., Kouzani, A.: A Review of Vision-based Gait Recognition Methods for Human Identification. Digital Image Computing: Techniques and Appl. (2010).
- [81] Bender, S., Postley, H.: Key sequence rhythm recognition system and method. US Patent No. 7206938, 2007.
- [82] Irvine, J. M., Israel, S. A., Scruggs, W.T., Worek, W. J.: Eigenpulse: Robust Human Identification from Cardiovascular Function. Pattern Recog., Vol. 41, No. 11, pp. 3427-3435 (2008).
- [83] Park, U., Ross, A., Jain, A. K.: Periocular Biometrics in the Visible Spectrum: A Feasibility Study. Proc BTAS, Washington DC, 2009.
- [84] Kumar, A., Ravikanth, C.: Personal Authentication using Finger Knuckle Surface. IEEE Trans. on Inf. Forensics & Security, Vol. 4, No. 1, pp. 98-110 (2009).
- [85] Nakajima, K., Mizukami, Y., Tanka, K., Tamura, T.: Footprint-Based Personal Recognition. IEEE Trans. on Biomed. Eng., Vol. 47, No. 11, pp. 1534-1537 (2000)
- [86] Song, S., Ohnuma, K., Liu, Z., Mei, L., Kawada, A., Monma, T.: Novel Biometrics based on Nose Pore Recognition. Optical Eng., Vol. 48, No. 5 (2009).
- [87] Topping, A., Kuperschmidt, V., Gormley, A.: Method and Apparatus for The Automated Identification of Individuals by Their Beds of Their Fingernails. U.S. Patent No. 5751835, 1998.
- [88] Falguera, J. R., Falguera, F. P. S., Marana, A. N.: Frontal Sinus Recognition for Human Identification. Proc. of SPIE, Vol. 6944, Orlando, Florida: 69440S (2008).
- [89] Swabey, M. A., Beeby, S. P., Brown, A. D., Chad, J. E.: Otoacoustic Emission as Biometric. Proc. of ICBA, LNCS, Vol. 3072, pp. 1-34 (2004).
- [90] Day, D.: Biometric Applications, Overview. In: Li, S. Z., Jain, A. K. (Eds.) Encyclopedia of Biometrics, 2nd Edition, pp. 169-174. Springer, Heidelberg (2015).
- [91] Bhanu, Bir., Chen, Hui.: Human Ear Recognition by Computer. Springer, 2008.
- [92] J. Lei, Zhou, J., Abdel-Mottaleb, M.: Gender Classification using Automatically Detected and Aligned 3D Ear Range Data. Int. Conf. on Biometrics (ICB), pp. 1-7 (2013).
- [93] Hoogstrate, A., Heuvel, H. V. D., Huyben, E.: Ear Identification based on Surveillance Camera Images. Science & Justice, Vol. 41, Issue 3, pp. 167-172 (2001).
- [94] Sforza, C., Grandi, G., Binelli, M., Tommasi, D. G., Rosati, R., Ferrario, V. F.: Age- and Sex-Related Changes in The Normal Human Ear. Forensic Science International, Vol. 187 (1-3), pp. 110e1-110e7 (2009).
- [95] Bertillon, A.: La Photographie Judiciaire: Avec Un Appendice Sur La Classification Et L'Identification Anthropometriques. Gauthier-Villars, Paris, 1890.
- [96] Doyle, A. C.: A Chapter on Ears. Solis Press, 2012.
- [97] Iannarelli, A. V.: Ear Identification. Paramount Publishing Company, USA, 1989.
- [98] Hurley, D. J., Nixon, M. S., Carter, J. N.: Force Field Feature Extraction for Ear Biometrics. Computer Vision and Image Understanding, Vol. 98, Issue 3, pp. 491-512 (2005).
- [99] Hurley, D. J., Nixon, M. S., Carter, J. N.: Automatic Ear Recognition by Force Field Transformations. Proc. of IEE Colloquium: Visual Biometrics, 7/1-7/5 (2000).
- [100] Chang, K., Bowyer, K. W., Sarkar, S., Barnabas, V.: Comparison and Combination of Ear and Face Images in Appearance-Based Biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, Issue 9, pp. 1160-1165 (2003).
- [101] Zhang, H. J., Mu, Z. C., Qu, W., Liu, L. M., Zhang, C. Y.: A Novel Approach for Ear Recognition based on ICA and RBF Network. Proc. of 4th International Conference on Machine Learning and Cybernetics (CMLC'05), pp. 4511-4515 (2005).

- [102] Nanni, L., Lumini, A.: Fusion of Color Spaces for Ear Authentication. *Pattern Recognition*, Vol. 42, Issue 9, pp. 1906-1913 (2009).
- [103] Hurley, D. J., Nixon, M. S., Carter, J. N.: Ear Biometrics by Force Field Convergence. *Proc. of Int. Conf. on Audio- and Video-based Biometric Person Auth.*, LNCS, Vol. 3546, pp. 119-128 (2005).
- [104] Mu, Z., Yuan, L., Xu, Z., Xi, D., Qi, S.: Shape and Structural Feature based Ear Recognition. *Proc. of Advances in Biometric Person Authentication*. LNCS, Vol. 3338, pp. 663-670 (2004).
- [105] Bustard, J. D., Nixon, M. S.: Robust 2D Ear Registration and Recognition based on SIFT Point Matching. *Proc. of Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS'08)*, pp. 1-6 (2008).
- [106] Arbab-Zavar, B., Nixon, M., Hurley, D.: On Model-Based Analysis of Ear Biometrics. In *First IEEE International Conference on Biometrics: Theory, Applications, and Systems, (BTAS'07)*, pp. 1-5 (2007).
- [107] De Marsico, M., Nappi, M., Riccio, D.: HERO: Human Ear Recognition Against Occlusions. *IEEE Computer Society Conf. on Comp. Vis. and Patt. Recog. Workshops*, pp. 178-183, (2010).
- [108] Prakash, S., Gupta, P.: An Efficient Ear Recognition Technique Invariant to Illumination and Pose. *Telecommunication Systems Journal, Special Issue on Signal Processing Applications in Human Computer Interaction*, Vol. 30, pp. 38-50 (2011).
- [109] Moreno, B., Sanchez, A., Velez, J. F.: On The Use of Outer Ear Images for Personal Identification in Security Applications. *Proc. of IEEE Conf. on Sec. Tech.*, pp. 469-476 (1999).
- [110] Wang, X. Q., Xia, H. Y., Wang, Z. L.: The Research of Ear Identification based on Improved Algorithm of Moment Invariants. *Proc. of 3rd Int. Conf. on Inf. & Computing*, p. 58 (2010).
- [111] Zhang, B., Mu, Z., Zeng, H., Luo, S.: Robust Ear Recognition via Nonnegative Sparse Representation of Gabor Orientation Information. *The Scientific World Journal*, 2014.
- [112] Abate, A. F., Nappi, M., Riccio, D., Ricciardi, S.: Ear Recognition by Means of a Rotation Descriptor. *Proc. of 18th Int. Conf. on Pattern Recognition*, Vol. 4, pp. 437-440 (2006).
- [113] Arbab-Zavar, B., Nixon, M. S.: On Shape-Mediated Enrolment in Ear Biometrics. *Proc. of the 3rd Int. Conf. on Advances in Visual Computing*, Vol. Part II, pp. 549-558 (2007).
- [114] Alistair, H., Cummings, A.H., Nixon, M.S., Carter, J.N.: A Novel Ray Analogy for Enrolment of Ear Biometrics. *Proc of 4th IEEE Int. Conf. on Biometrics: Theory Appl. Sys.*, pp. 1-6 (2010).
- [115] Sana, P., Gupta, A.: Ear Biometrics: a New Approach. *Adv. in Patt. Recognition*, pp. 1-5 (2007).
- [116] Ojala, T., Pietikäinen, M., Harwood, D.: Performance Evaluation of Texture Measures with Classification Based on Kullback Discrimination of Distributions. *Proc. of the 12th IAPR Int. Conf. on Pattern Recognition (ICPR'94)*, Vol. 1, pp. 582-585 (1994).
- [117] Lu, L., Xiaoxun, Z., Youdong, Z., Yunde, J.: Ear Recognition based on Statistical Shape Model. *Proc. of 1st Int. Conf. on Innovative Computing, Information and Control*, pp. 353-356 (2006).
- [118] Wang, X., Yuan, W.: Human Ear Recognition based on Block Segmentation. *Int. Conf. on Cyber-Enabled Distributed Computing and Knowledge Discovery*, pp. 262-266 (2009).
- [119] Akkermans, A.H.M., Kevenaar, T.A.M., Schobben, D.W.E.: Acoustic Ear Recognition for Person Identification. *4th IEEE Workshop on Automatic Identif. Adv. Technologies*, pp. 219-223 (2005).
- [120] Iwano, K., Hirose, T., Kamibayashi, E., Furui, S.: Audio-Visual Person Authentication using Speech and Ear Images. *Proc. of Workshop on Multimodal User Authentic.*, pp. 85-90 (2003).
- [121] Rahman, M. M., Ishikawa, S.: Proposing a Passive Biometric System for Robotic Vision. *Proc. of 10th International Symposium on Artificial Life and Robotics (AROB'05)*, pp. 4-6 (2005).
- [122] Iwano, K., Miyazaki, T., Furui, S.: Multimodal Speaker Verification using Ear Image Features Extracted by PCA and ICA. *Proc. of Int. Conf. on Audio and Video Based Biometric Person Authentication*. LNCS, Vol. 3546, pp. 588-5996 (2005).
- [123] Chen, H., Bhanu, B.: Contour Matching for 3D Ear Recognition. *Proc. of the 7th IEEE Workshop on Applications of Computer Vision (WACV/MOTION)*, 2005.
- [124] Yan, P., Bowyer, K. W.: ICP-Based Approaches for 3D Ear Recognition. *Proc. of SPIE*, Vol. 5779, pp. 282-291 (2005).
- [125] Yan, P., Bowyer, K. W.: Biometric Recognition using Three-Dimensional Ear Shape. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 29, Issue 8, pp. 1297-1308 (2007).

- [126] Zhou, J., Cadavid, S., Abdel-Mottaleb, M.: A Computationally Efficient Approach to 3D Ear Recognition Employing Local and Holistic Features. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 98-105 (2011).
- [127] Gupta, P., Prakash, S.: An Efficient Technique for Ear Detection in 3D: Invariant to Rotation and Scale. 5th IAPR Int. Conf. on Biometrics, pp. 97-102 (2012).
- [128] Cadavid, S., Abdel-Mottaleb, M.: 3D Ear Modeling and Recognition from Video Sequences using Shape from Shading. IEEE Trans. on Inf. Foren. and Security, Vol. 3, No. 4, pp. 709-718 (2008).
- [129] Yan, P., Bowyer, K. W.: Multi-Biometrics 2D and 3D Ear Recognition. In: Kanade, T., Jain, A., Ratha, N. K. (Eds.) Proc. of the Int. Conf. on Audio- and Video-Based Person Authentication (AVBPA'05). LNCS, Vol. 3546, pp 503-512 (2005).
- [130] Faez, K., Motamed, S., Yaqubi, M.: Personal Verification using Ear and Palm-Print Biometrics. Proc. of IEEE Int. Conf. on Systems, Man and Cybernetics (SMC'08), pp 3727-3731 (2008).
- [131] Gnanasivam, P., Muttan, S.: Ear and Fingerprint Biometrics for Personal Identification. Proc. of The Int. Conf. on Signal Processing, Communication, Computing and Networking Technologies (ICSCCN'11), pp 347-352 (2011).
- [132] Monwar, M., Gavrilova, M.: FES: A System for Combining Face, Ear and Signature Biometrics using Rank Level Fusion. Proc. of the 5th International Conference on Information Technology: New Generations (ITNG'08), pp 922-927 (2008).
- [133] Cadavid, S., Mahoor, M. H., Abdel-Mottaleb, M.: Multi-Modal Biometric Modeling and Recognition of The Human Face and Ear. Proc. of the IEEE Int. Workshop on Safety, Security and Rescue Robotics (SSRR'09), pp 1-6 (2009).
- [134] Kandgaonkar, T. V., Mente, R. S., Shinde, A. R., Raut, S. D.: Ear Biometrics: A Survey on Ear Image Databases and Techniques for Ear Detection and Recognition. IBMRD's Journal of Management and Research, Vol. 4, Issue 1 (2015).
- [135] Zhang, L., Ding, Z., Li, H., Shen, Y.: 3D Ear Identification Based on Sparse Representation. PLoS ONE, Vol. 9, Issue 4 (2014).
- [136] Open Hear database, <http://www2.imm.dtu.dk/projects/OpenHear/>
- [137] SYMARE database, <http://www.ee.usyd.edu.au/carlab/symare.htm>
- [138] Prakash, S., Gupta, P.: An Efficient Ear Localization Technique. Image and Vision Computing, Vol. 30, pp. 38-50 (2012).
- [139] Prakash, S., Gupta, P.: An Efficient Technique for Ear Detection In 3D: Invariant to Rotation and Scale. In the 5th IAPR International Conference on Biometrics (ICB), 2012.
- [140] Graham, D. B., Allinson, N. M.: Characterizing Virtual Eigensignatures for General Purpose Face Recognition. In: Face Recognition: From Theory to Applications, H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman-Soulie and T. S. Huang (Eds.) NATO ASI Series F, Computer and Systems Sciences, Vol. 163, pp 446-456, 1998, <http://www.sheffield.ac.uk/eee/research/iel/research/face#>
- [141] Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., and Zhao, D.: CAS-PEAL - Large-Scale Chinese Face Database and Baseline Evaluations. IEEE Trans. Syst. Man Cybernet. Part A Syst. Hum., Vol. 38, Issue 1, pp. 149-161, 2008, <http://www.jdl.ac.cn/peal/home.htm>
- [142] Raposo, R., Hoyle, E., Peixinho, A., Proenca, H.: UBEAR: A Dataset of Ear Images Captured On-The-Move in Uncontrolled Conditions. IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM'11), pp. 84-90, 2011, <http://ubear.di.ubi.pt/ubear.html>
- [143] Grgic, M., Delac, K., Grgic, S.: Scface – Surveillance Cameras Face Database. Multimedia Tools Appl., Vol. 51, Issue 3, pp. 863-879, 2011, <http://www.scface.org/>
- [144] Messer, K., Matas, J., Kittler, J., Luettin, J., Maitre, G.: XM2VTSDB: The extended M2VTS Database. Proc. of the 2nd Int. Conf. on Audio and Video-Based Biometric Person Authentication, 1999, <http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/>
- [145] Phillips, P. J., Rauss, P. J., Der, S. Z.: FERET (Face Recognition Technology) Recognition Algorithm Development and Test Results. Army Research Lab technical report 995, 1996. <http://www.itl.nist.gov/iad/humanid/feret/>

- [146] Phillips, P. J., Moon, H., Rauss, P. J., Rizvi, S.: The FERET Evaluation Methodology For Face Recognition Algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 10, 2000, <http://www.itl.nist.gov/iad/humanid/feret/>
- [147] Marcenaro, L., Vernazza, G., Regazzoni, C. S.: Image Stabilization Algorithms For Video-Surveillance Applications. IEEE Int. Conf. on Image Processing, Vol. 1, pp. 349-352 (2001).
- [148] Rosten, E., Porter, R., Drummond, T.: FASTER and better: A Machine Learning Approach to Corner Detection. IEEE Transaction on PAMI, Vol. 32, pp. 105-119 (2010).
- [149] Kovesi, P., MATLAB and Octave Functions for Computer Vision and Image Processing, <http://www.csse.uwa.edu.au/~pk/research/matlabfns/>
- [150] Lucas, B., Kanade, T.: An Iterative Image Registration Technique with An Application To Stereo Vision. Proc. of the Int. Joint Conf. on Artificial Intelligence, pp. 674-679 (1981).
- [151] Bouguet, J.: Pyramidal Implementation of the Lucas-Kanade Feature Tracker: Description of The Algorithm. Technical report, OpenCV Document, Intel Microprocessor Res. Labs, 2000.
- [152] Piotr's Computer Vision Matlab Toolbox, <http://vision.ucsd.edu/~pdollar/toolbox/doc/>
- [153] Otsu, N. A.: Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, Issue 1, pp. 62-66 (1979).
- [154] Lienhart, R., Kuhmünch, C., Effelsberg W.: On the Detection and Recognition of Television Commercials. Reihe Informatik 16 (1996).
- [155] Tanwer, A., Reel, P. S.: Effects of Threshold of Hard Cut Based Technique for Advertisement Detection in TV Video Streams. Proc. of the IEEE Students Technology Symposium, Kharagpur, India, 03-04 April, 2010.
- [156] Lienhart, R., Stuber, F.: Automatic Text Recognition in Digital Videos. In Image and Video Processing IV, Vol. 2666, No. 20, pp. 180-188 (1996).
- [157] Dimov, D., Nikolov, A.: Real Time Video Stabilization for Handheld Devices, In: Rachev, B., A. Smrikarov (Eds.) Proc. of CompSysTech'14, June 27, 2014, Ruse, Bulgaria, ACM ICPS, Vol. 883, pp. 124-133 (2014).
- [158] Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufman Publisher, Elsevier, 2006.
- [159] Dimov, D. T., Cantoni, V.: Appearance-Based 3D Object Approach to Human Ears Recognition. In: Cantoni, V., Dimov, D. T., Tistarelli, M. (Eds.) Proc. of the 1st Int. Workshop on Biometrics, BIOMET June 23-24, 2014, Sofia, Bulgaria. Biometric Authentication, LNCS, vol. 8897, pp. 121-135. Springer, Heidelberg (2014).
- [160] CVRL DATA SETS (University of Notre Dame UND database), http://www.cse.nd.edu/~cvrl/CVRL/Data_Sets.html
- [161] HANDY SCAN 3D: The Portable 3D Scanners for Industrial Application, http://www.creaform3d.com/sites/default/files/assets/brochures/files/handyscan/Handyscan3D_Brochure_EN_HQ_22052012.pdf
- [162] Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G.: MeshLab: an Open-Source Mesh Processing Tool. Proc. of Eurographics Italian Chapter Conf., pp. 129-136 (2008).
- [163] Vollmer, J., Mencl, R., Müller, H.: Improved Laplacian Smoothing of Noisy Surface Meshes. Int. Conf. on EUROGRAPHICS, Vol. 18, Issue 3, pp. 131-138 (1999).
- [164] Boyé, S., Guennebaud, G., Schlick, C.: Least Squares Subdivision Surfaces. Computer Graphics Forum, Vol. 29, Issue 7, pp. 2021-2028 (2010).
- [165] Horn, B.K.P.: Extended Gaussian images. Proc. of the IEEE, Vol. 72, pp. 1671-1686 (1984).
- [166] Kang, S. B., Ikeuchi, K.: Determining 3-D Object Pose using the Complex Extended Gaussian Image. IEEE Computer Society Conf. on Comp. Vision and Pattern Recog., pp. 580-585 (1991).
- [167] Matsuo, H., Iwata, A.: 3-D Object Recognition Using MEGI Model from Range Data. Proc. of 12th Int. Conf. Pattern Recognition, pp. 843-846 (1994).
- [168] Wang, D., Zhang, J., Wong, H.-S., Li, Y.: 3D Model Retrieval Based on Multi-Shell Extended Gaussian Image. In: Qiu, G., Leung, C., Xue, X.-Y., Laurini, R. (Eds.) VISUAL 2007. LNCS, Vol. 4781, pp. 426-437. Springer, Heidelberg (2007).
- [169] Hu, Z., Chung, R., Fung, K.S.M.: EC-EGI: Enriched Complex EGI for 3D Shape Registration. Machine Vision and Applications, Vol. 21, pp. 177-188 (2010).

- [170] Bray, J. R., Curtis, J. T.: An Ordination of Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, Vol. 27, pp. 325-349 (1957).
- [171] Dimov, D.: A Polar-Fourier-Wavelet Transform for Effective CBIR. In: Morzy, T., Morzy, M., Nanopoulos, A. (Eds.) *ADMKD 2007, 11th ADBIS'07*, Varna, BG, pp. 107-118 (2007).
- [172] Dimov D., Marinov, A., Zlateva, N.: CBIR Approach to the Recognition of a Sign Language Alphabet. *Proc. of CompSysTech'2007*, June 14-15, Rousse, 2007, Vol. 285, pp.V.2.1-9 (2007).
- [173] Dimov D. T., Zlateva, N., Marinov, A.: CBIR Approach to Face Recognition. *Proc. of A&I'08 Conf.*, Sofia, Workshop on Multisensor Signal, Image and Data Processing, pp. IV.21-26 (2008).
- [174] Dimov, D., Zlateva, N., Marinov, A.: CBIR Over Multiple Projections of 3D Objects. In: Fierrez, J., Ortega-Garcia, J., et al. (Eds.) *Proc. of the Joint COST 2101 & 2102 BioID_MultiComm'09*, LNCS, Vol. 5707, pp. 146-153 (2009).
- [175] Furukawa, Y., Ponce, J.: Accurate, Dense, and Robust Multiview Stereopsis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 32, Issue 8, pp. 1362-1376 (2010).
- [176] Geng, J.: Structured-Light 3D Surface Imaging: a Tutorial. *Advances in Optics and Photonics*, Vol. 3, pp. 128-160 (2011).
- [177] Creaform, Inc., Handyscan 3D Laser Scanners, <http://www.creaform3d.com>
- [178] Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 519-528 (2006).
- [179] Georgopoulos, A., Ioannidis, Ch., Valanis, A.: Assessing the Performance of A Structured Light Scanner. *Proc. of the ISPRS Commission V Mid-Term Symposium on Close Range Image Measurement Techniques*, Vol. XXXVIII, Newcastle upon Tyne, UK (2010).
- [180] Li, H., Adams, B., Guibas, L. J., Pauly, M.: Robust Single-View Geometry and Motion Reconstruction. *Proc. of ACM SIGGRAPH*, Vol. 28, Issue 5, Article No. 175, 10 pages (2009).
- [181] Chenglei, W., Wilburn, B., Matsushita, Y., Theobalt, C.: High-Quality Shape from Multi-View Stereo and Shading under General Illumination. *Proc. of the 2011 IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 969-976 (2011).
- [182] Nießner, M., Zollhöfer, M., Izadi, S., Stamminger, M.: Real-Time 3D Reconstruction at Scale using Voxel Hashing. *Journal of ACM Trans. on Graphics*, Vol. 32, Issue 6, Article No. 169, 11 pages (2013).
- [183] Pradeep, V., Rhemann, C., Izadi, S., Zach, C., Bleyer, M., Bathiche, S.: MonoFusion: Real-Time 3D Reconstruction of Small Scenes with a Single Web Camera. *IEEE Int. Symposium on Mixed and Augmented Reality*, pp. 83-88 (2013).
- [184] Zollhöfer, M., Nießner, M., et al.: Real-Time Non-Rigid Reconstruction using an RGB-D Camera. *Journal of ACM Trans. on Graphics*, Vol. 33, Issue 4, Article No. 156, 12 pages (2014).
- [185] Pock, T., Schoenemann, T., Graber, G., Bischof, H., Cremers, D.: A Convex Formulation of Continuous Multi-Label Problems. In *European Conference on Computer Vision (ECCV)*, 2008.
- [186] Kolev, K., Klodt, M., Brox, T., Cremers, D.: Continuous Global Optimization in Multiview 3D Reconstruction. In *Int. Journal of Computer Vision*, Vol. 84 (2009).
- [187] Oswald, M. R., Stühmer, J., Cremers, D.: Generalized Connectivity Constraints for Spatio-temporal 3D Reconstruction. In *European Conference on Computer Vision (ECCV)*, 2014.
- [188] Stühmer, J., Gumhold, S., Cremers, D.: Real-Time Dense Geometry from a Handheld Camera, http://link.springer.com/chapter/10.1007%2F978-3-642-15986-2_2
- [189] Engel, J., Schöps, T., Cremers, D.: LSD-SLAM: Large-Scale Direct Monocular SLAM. In *European Conference on Computer Vision (ECCV)*, 2014.
- [190] Usenko, V., Engel, J., Stueckler, J., Cremers, D.: Reconstructing Street-Scenes in Real-Time From a Driving Car. In *Proc. of the Int. Conference on 3D Vision (3DV)*, 2015.
- [191] Sturm, J., Bylow, E., Kahl, F., Cremers, D.: CopyMe3D: Scanning and Printing Persons in 3D. In *German Conference on Pattern Recognition (GCPR)*, 2013.
- [192] Pollefeys, M., Gool, L. V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual Modeling with a Hand-Held Camera. *Int. J. of Comp. Vision*, Vol. 59, Issue 3, pp. 207-232 (2004).
- [193] Gallup, D., Pollefeys, M., Frahm, J.-M.: 3D Reconstruction Using an n-Layer Heightmap, http://link.springer.com/chapter/10.1007%2F978-3-642-15986-2_1

- [194] Tanskanen, P., Kolev, K., Meier, L., Camposeco, F., Saurer, O., Pollefeys, M.: Live Metric 3D Reconstruction on Mobile Phones. Proc. of the IEEE Int. Conf. on Computer Vision (ICCV '13), pp. 65-72 (2013).
- [195] Haene, C., Sattler, T., Pollefeys, M.: Obstacle Detection for Self-Driving Cars Using Only Monocular Cameras And Wheel Odometry. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (2015).
- [196] Fraundorfer, F., Lionel, H., Honegger, D., Lee, G. H., Meier, L., Tanskanen, P., Pollefeys, M.: Vision-Based Autonomous Mapping and Exploration Using a Quadrotor MAV. Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), 2012.
- [197] Raghavendra, R., Raja, K. B., Pflug, A., Yang, B., Busch, C.: 3D Face Reconstruction and Multimodal Person Identification from Video Captured using Smartphone Camera. Proc. of IEEE Int. Conf. on Technologies for Homeland Security, pp. 552-557 (2013).
- [198] Nikolova, I., Nikolov, A., Zapryanov, G.: Depth Estimation Using Shifted Digital Still Camera. Proc. of the 12th Int. Conf. on Computer Systems and Technologies (CompSys-Tech'11), June 16-17, 2011, Vienna, Austria, ACM ICPS Vol. 578, pp. 234-240 (2011).
- [199] Woodham, R. J.: Photometric Method for Determining Surface Orientation from Multiple Images. Opt. Eng., Vol. 19, Issue 1, 191139 (1980).
- [200] Shi, B.: Photometric Stereo for General Reflectance and Lighting. A doctoral dissertation, the University of Tokyo, Supervisor: Katsushi Ikeuchi, June, 2013.
- [201] Xie, W., Zhang, Y., Wang, C.C.L. Chung, R.C.-K.: Surface-from-Gradients: An Approach Based on Discrete Geometry Processing. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2014.
- [202] Szeliski, R.: Computer Vision: Algorithms and Applications. Springer, New York, 2011.
- [203] Flynn, P. J., Bowyer, K. W., Phillips, P. J.: Assessment of Time Dependency in Face Recognition: An Initial Study. Audio and Video-Based Biometric Person Authentication, pp. 44-51 (2003).
- [204] Chang, K., Bowyer, K. W., Sarkar, S., Victor, B.: Comparison and Combination of Ear and Face Images in Appearance Based Biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 9, pp. 1160-1165 (2003).
- [205] Yan, P., Bowyer, K. W.: Empirical Evaluation of Advanced Ear Biometrics. IEEE Computer Society Workshop on Empirical Evaluation Methods in Computer Vision, 41 pages (2005).
- [206] Zhou, J., Cadavid, S., Abdel Mottaleb, M.: Histograms of Categorized Shapes for 3D Ear Detection. Proc. of the IEEE Conf. on Biometrics: Theory, Applications, and Systems (2010).
- [207] Yan, P., Bowyer, K. W.: An Automatic Ear Recognition System. Proc. of 3rd Int. Symposium on 3D Data Processing, Visualization and Transmission (3DPVT'2006), pp. 326-333 (2006).
- [208] Faltemier, T., Bowyer, K. W., Flynn, P. J.: Rotated Profile Signatures for Robust 3D Feature Detection. Proc. of Automatic Face & Gesture Recognition (FG '08), Amsterdam, 2008, <http://dx.doi.org/10.1109/AFGR.2008.4813413>
- [209] IITK: Biometric Research Lab (Indian Institute of Technology in Kanpur), <http://www.cse.iitk.ac.in/users/biometrics/>
- [210] USTB: Ear Recognition Laboratory (University of science and technology, Beijing, USTB database), <http://www1.ustb.edu.cn/resb/en/index.htm>
- [211] Introduction to USTB ear image databases I, II, III, http://www1.ustb.edu.cn/resb/en/doc/Imagedb_123_intro_en.pdf
- [212] Attarchi, S., Faez, K., Rafiei, A.: A New Segmentation Approach for Ear Recognition. Proc. of Int. Conf. on Advanced Concepts for Intelligent Vision Systems, pp. 1030-1037 (2008).
- [213] Wang, Z.-Q., Yan, X.-D.: Multi-Scale Feature Extraction Algorithm of Ear Image. Proc. of Int. Conf on Electric Information and Control Engineering (ICEICE), p. 528 (2011).
- [214] Introduction to USTB ear image databases IV, http://www1.ustb.edu.cn/resb/en/doc/Imagedb_4_intro_en.pdf
- [215] Zhang, Z. Liu, H.: Multi-View Ear Recognition Based on B-Spline Pose Manifold Construction. Proc. of the 7th IEEE World Congress on Intelligent Control and Automation, 2008.
- [216] WPUT-DB (West Pomeranian University of Technology - DataBase), <http://ksm.wi.zut.edu.pl/wputedb/>

- [217] Frejlichowski, D., Tyszkiewicz, N.: The West Pomeranian University of Technology Ear Database a Tool for Testing Biometric Algorithms. In: Campilho, A., Kamel, M. (Eds.) Image Analysis and Recognition. LNCS, Vol. 6112, pp. 227-234. Springer Berlin / Heidelberg (2010).
- [218] IIT Delhi. Biometrics Research Laboratory at IIT Delhi, http://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Ear.htm
- [219] Kumar, A., Wu, C.: Automated Human Identification Using Ear Imaging. Int. Conf. on Pattern Recognition, Vol. 45, Issue 3, pp. 956-968 (2012).
- [220] ScFace (Surveillance Cameras Face Database) Technical University of Zagreb, <http://www.scface.org/>
- [221] UMIST (the Sheffield Face Database) Image Engineering Laboratory, University of Sheffield, <http://www.sheffield.ac.uk/eee/research/iel/research/face#>
- [222] Abaza, A., Hebert, C., Harrison, M.: Fast Learning Ear Detection for Real-Time Surveillance. In Fourth IEEE Int. Conf. on Biometrics: Theory Applications and Systems, pp. 1-6 (2010).
- [223] Nizami, H. A., Adkins-Hill, J. P., Zhang, Y., Sullins, J. R., McCullough, C., Canavan, S., Yin, L.: A Biometric Database with Rotating Head Videos and Hand- Drawn Face Sketches. Proc. of the 3rd IEEE Int. Conference on Biometrics: Theory, Applications and Systems (BTAS'09), pp. 38-43, Piscataway, NJ, USA (2009).
- [224] NCKU (The National Cheng Kung University in Taiwan), http://robotics.csie.ncku.edu.tw/Databases/FaceDetect_PoseEstimate.htm
- [225] UBEAR (University of Beira Interior, Department of Computer Science, Portugal), <http://ubear.di.ubi.pt/ubear.html>
- [226] Fahmy, G., Elsherbeen, A., Mandala, S., Abdelmottaleb, M., Ammar, H.: The Effect of Lighting Direction / Condition on the Performance of Face Recognition Algorithms. Proc. of the SPIE Conf. on Human Identification, 2006.
- [227] Hajsaid, E., Abaza, A., Ammar, H.: Ear Segmentation in Color Facial Images using Mathematical Morphology. Proc. of the 6th IEEE Biometric Consortium Conference (BCC), 2008.
- [228] XM2VTS (Extended Multi Modal Verification for Teleservices and Security applications), <http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/>
- [229] Ibrahim, M.I.S., Nixon, M.S., Mahmoodi, S.: Shaped Wavelets for Curvilinear Structures for Ear Biometrics. Proc. of 6th Int. Conf. on Advances in Visual Computing (ISVC'10), Vol. Part I, pp. 499-508 (2010).
- [230] NIST (Mugshot Identification Special Database 18), <http://www.nist.gov/srd/nistsd18.cfm>
- [231] Watson, C.I.: NIST Special Database 18 - Mugshot Identification Database. National Institute of Standards and Technology, Visual Image Processing Group, December, 1994.
- [232] FERET (The Facial Recognition Technology), <http://www.itl.nist.gov/iad/humanid/feret/>
- [233] CAS-PEAL face database (Chinese Academy of Sciences - Poses, Expressions, Accessories, and Lighting), <http://www.jdl.ac.cn/peal/home.htm>
- [234] Carreira-Perpiñán database, <http://faculty.ucmerced.edu/mcarreira-perpinan/papers/msc-thesis.html>
- [235] Carreira-Perpiñán, M. Á: Compression neural Networks for Feature Extraction: Application to Human Recognition from Ear Images (in Spanish). MSc Thesis, Faculty of Informatics, Technical University of Madrid, Spain, 1995.
- [236] Nosrati, M., Faez, K., Faradji, F.: Using 2D Wavelet and Principal Component Analysis for Personal Identification based on 2D Ear Structure. Proc. of the IEEE Int. Conf. on Intelligent and Advanced Systems, 2007.
- [237] AMI Ear database, http://www.ctim.es/research_works/ami_ear_database/

Приложения

Приложение 1. Йерархично представяне на линейните 2D трансформации.⁴³

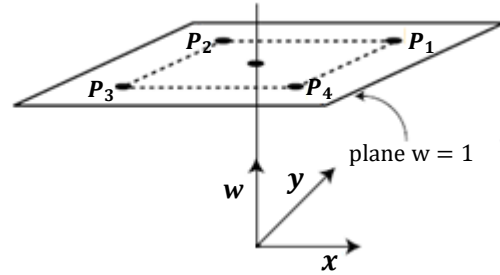
$$\text{Нека } X = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \equiv w \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} u \\ v \\ w \end{pmatrix}$$

е хомогенното представяне на дадена точка (x, y) , където $x = u/w$, $y = v/w$ дефинира обратното преобразуване от хомогенни в евклидови координати, $w \neq 0$.

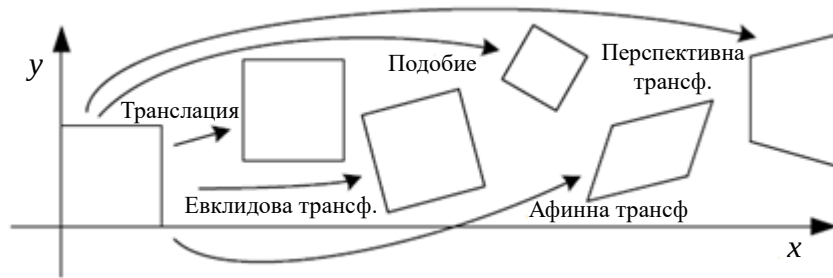
Нека (x', y') е образ на дадената точка (x, y) след някаква трансформация и

$$X' = \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} \equiv w' \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} u' \\ v' \\ w' \end{pmatrix}$$

е нейното хомогенно представяне на (x', y') , където $x' = u'/w'$, $y' = v'/w'$ е съответното обратное преобразуване от хомогенни към евклидови координати, $w' \neq 0$.



Фиг. 69. Хомогенно представяне на 2D точки върху равнината $w = 1$.



Фиг. 70. Йерархично множеството от линейни 2D трансформации. *Заимствано от [202].*

1. Транслация (Translation)

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$\text{или } X' = H_T X = \begin{bmatrix} I & t \\ 0^T & 1 \end{bmatrix} X, \text{ където } t = \begin{bmatrix} t_x \\ t_y \end{bmatrix}, I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{aligned} x' &= x + t_x \\ y' &= y + t_y \end{aligned}$$

⁴³ За повече информация виж книгите на Hartley [9] и Szeliski [202].

2. Евклидова трансформация (Euclidean transformation)

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & t_x \\ \sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$\text{или } X' = H_E X = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} X, \text{ където } R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$$\begin{aligned} x' &= x(\cos\theta) - y(\sin\theta) + t_x \\ y' &= x(\sin\theta) + y(\cos\theta) + t_y \end{aligned}$$

Евклидовата трансформация моделира движението на “rigid” обекти. Тя представлява композиция от ротация и трансляция. Двумерната Евклидова трансформация има 3 степени на свобода (DoF) – една за ротацията и две за трансляцията. Затова, съответните три параметъра могат да се определят по две дадени точки и техните образи, като дължините на съответните отсечки трябва да са еднакви.

Тази трансформация запазва дължините на отсечките, ъглите между тях и площта на обектите (т.е. линейните, квадратичните и ъгловите съотношения в 2D обектите).

3. Трансформация на подобие (Similarity)

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} s \cdot \cos\theta & -s \cdot \sin\theta & t_x \\ s \cdot \sin\theta & s \cdot \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$\text{или } X' = H_S X = \begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix} X, \text{ където } s \text{ е коефициент на мащабиране}$$

$$\begin{aligned} x' &= x(s \cdot \cos\theta) - y(s \cdot \sin\theta) + t_x \\ y' &= x(s \cdot \sin\theta) + y(s \cdot \cos\theta) + t_y \end{aligned}$$

Двумерната трансформация на подобие има 4 степени на свобода – същите три, като на евклидовата трансформация + още една, поради изотропното мащабиране. Тази трансформация може да се изчисли чрез две съответстващи си точки (и съответните им образи).

Трансформацията на подобие запазва ъглите, както и успоредните направления в обектите. Дължините между две точки не се запазват, но тяхното съотношение – да.

4. Афинна трансформация (Affine)

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} a.s.\cos\theta & -b.s.\sin\theta & t_x \\ a.s.\sin\theta & b.s.\cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$\text{или } X' = H_A X = \begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix} X,$$

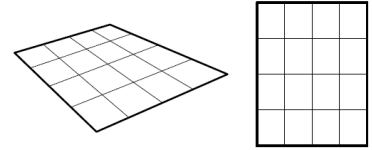
където a и b са различни мащабиращи коефициенти

$$\begin{aligned} x' &= x(a.s.\cos\theta) - y(b.s.\sin\theta) + t_x = xa_{11} + ya_{12} + t_x \\ y' &= x(a.s.\sin\theta) + y(b.s.\cos\theta) + t_y = xa_{21} + ya_{22} + t_y \end{aligned}$$

Единствената разлика на тази трансформация с трансформацията на подобие е неизотропното мащабиране. Двумерната афинна трансформация има 6 степени на свобода, като тези параметри могат да се определят по *три* точки (и техните образи).

Тук, паралелните линии се запазват, но съотношенията на дължините, както и на ъглите – не.

5. Проективна трансформация (Projective $\equiv$ Homography)



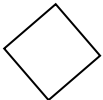
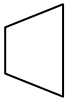
$$\begin{pmatrix} u' \\ v' \\ w' \end{pmatrix} = \begin{bmatrix} a.s.\cos\theta & -b.s.\sin\theta & t_x \\ a.s.\sin\theta & b.s.\cos\theta & t_y \\ p_1 & p_2 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$\text{или } X' = H_P X = \begin{bmatrix} A & t \\ p^T & 1 \end{bmatrix} X, \text{ където } p = (p_1, p_2)^T$$

$$\begin{aligned} u' &= xh_{11} + yh_{12} + h_{13} \\ v' &= xh_{21} + yh_{22} + h_{23} \\ w' &= xh_{31} + yh_{32} + 1 \end{aligned} \Rightarrow x' = \frac{xh_{11} + yh_{12} + h_{13}}{xh_{31} + yh_{32} + 1}, \quad y' = \frac{xh_{21} + yh_{22} + h_{23}}{xh_{31} + yh_{32} + 1}$$

Проективната трансформация между две равнини може да се изчисли чрез *четири* съответстващи си точки, които трябва да не лежат на една права. Тази трансформация запазва правите линии след трансформацията.

Таблица 17. Йерархия на линейните 2D трансформации.

Трансформация	Матрица	# DoF (# параметри)	Запазва	Означение
Транслация	$\begin{bmatrix} I & t \\ 0^T & 1 \end{bmatrix}$	2	ориентацията дължините ъглите успоредност прави линии	
Евклидова	$\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$	3	дължините ъглите успоредност прави линии	
Подобие	$\begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix}$	4	ъглите успоредност прави линии	
Афинна	$\begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix}$	6	успоредност прави линии	
Проективна	$\begin{bmatrix} A & t \\ p^T & 1 \end{bmatrix}$	8	прави линии	

Всяка трансформация от по-нисък ред (степен на свобода) запазва свойствата на тези с по-висок ред. Например, подобие запазва ъглите, успоредността и правите линии.

Приложение 2. Сравнителен анализ на биометричните признаци – предимства, недостатъци, приложения.

Вид	Биометрична модалност	Описание	Предимства	Недостатъци	Точност	Приложения
ФИЗИОЛОГИЧНА БИОМЕТРИКА	ДНК анализ* (DNA)	ДНК се съдържа в ядрото на всяка клетка от човешкото тяло; може да установи биологичната идентичност на жив човек или на негови останки, както и да разкрие роднински връзки.	Високо устойчива биометрична характеристика; ДНК структурата на всеки човек е уникална (с изключение на еднородните близнаци).	Резултатите не са незабавни и не са податливи на пълна автоматизация; цената на анализа е висока; има проблеми със замърсяване, преноса и устойчивостта; натрапчиво за индивида.	100%	В съдебната експертиза, криминални разследвания и в цялата правосъдна система.
	Пръстови отпечатъци (Fingerprint)	Пръстовите отпечатъци се състоят от текстурни шарки, съставени от била и долини, които се формират по време на първите седем месеца от развитието на плода.	Уникалност и неизменчивост във времето; висока точност на разпознаването; възможност за работа в реално време.	N/A	Very High → High	В съдебната експертиза (анализ на криминални сцени); електронни документи, паспорти, визи, лични карти; гранична сигурност; АТМ автентикация; контрол на физическия достъп (ключалки с пръстови четци); контрол на логическия достъп (компютри); системи за управление на работното време и посещаемостта; лесно вграждане на четци в лаптопи, клавиатури, мишки, умни телефони
		2D	Получаване на 2D изображение на върха на пръста чрез оптични, чувствителни на допир, използващи ултразвук, или термални четци.	Много добро съотношение производителност/цена на пръстовите четци; налични огромни правителствени БД; лесно използване и широка обществена приемственост; не са толкова натрапчиви като ириса.	High	
		3D	Използване на информацията за 3D повърхнината на пръста, заснета / сканирана чрез 3D сензор.	Основното предимство на използването на 3D геометрия на пръста е възможността за безконтактна работа.	Very High	

ФИЗИОЛОГИЧНА БИОМЕТРИКА		Вени на пръста (Finger vein)	Вените на пръстите се намират под кожата и са видими чрез сензори, активни в близката инфрачервена област.	Един от най-точните и сигурни биометрични характеристики; изключително труден за фалшифициране, кражба или дублиране); висока използваемост и потребителска приемственост.	N/A	Very High	Банкови транзакции (80% от японските финансови институции); система за достъп до вратите на автомобил (Hitachi Engineering Co); логически достъп до РС; терминал за управление на работното време и посещаемостта (Amano Corporation, Japan).
	Очи (Ирис и Ретина)	Ирис	Ирисът е цветният пръстен около зеницата, чиито текстурни характеристики се използват за разпознаване на индивида.	Много висока точност и бързодействие на сравнението; висока универсалност, отличимост и устойчивост във времето; малка вероятност за фалшифициране; лесно достъпни по безконтактен начин.	Натрапчиво за потребителя.	Very High	В системи за идентификация от голям мащаб (граничен контрол в Обединените Арабски емирства) и националната сигурност.
		Ретина	Това е слой от кръвоносни съдове, разположени в задната част на очната ябълка, които са видими в близката инфрачервена област.	Имат много богата и уникална структура.	Необходимо е много дълго време за обработка и формиране на изображения в сравнение с другите биометрични технологии.		N/A
	Физиономии / Лица (Face)		Лицевото разпознаване е процедурата по разпознаване на личността по техните лицеви отличителни черти.	Разпознаването може да бъде извършено от разстояние, без съдействието или знанието на потребителя; широко възприето и ненаатрапчиво в сравнение с признаци като пръстови отпечатъци, ирис, ретина, геометрия на ръката.	Лицето се променя чувствително с годините; също промените може да се дължат на грим и прическа, емоции, очила, брада и мустаци; не може да осигури толкова голяма точност както пръстовите отпечатъци, ирисът или ДНК.	High → Medium	При взаимодействието човек-компютър; достъп до защитени области; верификация на хора на летища.

ФИЗИОЛОГИЧНА БИОМЕТРИКА	Физиономии / Лица (Face)	2D	Тук се използват 2D изображения, заснети чрез фото / видео камера.	Наличие на огромни БД от физиономии; лесно вграждане в ежедневните устройства (лаптопи, умни телефони, таблети) и системите за видеонаблюдение на обществени места – метростанции, гари, летища.	Проблеми с вариациите по ориентация на лицето, гледна точка на камерата, разделителна способност на изображенията, осветление, лицеви изражения, дългосрочни промени в косата и кожата; ниска устойчивост в неконтролирана среда; 2D алгоритмите могат да бъдат заблудени чрез снимка на лицето.	Medium	N/A
		3D	Тук се използват данните от 3D сензори, които възприемат 3D геометричната информация на физиономията	3D данните са инвариантни по ориентация и осветление; драстично намаляване на цената на 3D скенерите.	Генериране на значително повече информация от една 2D камера, което води до по-големи изисквания към времето за обработка и съхранение; липса на големи БД от 3D лица.	High	N/A
		Near-Infrared	Заснемането се извършва в близкия инфрачервен спектър 0.7 – 0.9 μm .	Преодолява проблемите от вариацията на осветлението; постига значително по-висока производителност от подхода във видимия спектър.	Не е широко възприет, поради високата цена на инфрачервените камери; проблеми при визуални особености като белези и косми.	High	N/A
		Thermal	Температурните промени по повърхността на лицето произвеждат топлинни шарки (термограми), които могат да бъдат заснети като изображения чрез термокамери в диапазона 2.4–14 μm	Термо изображенията са почти инвариантни на промени в заобикалящото осветление и могат да се получат при пълна тъмнина; улеснена детекцията на физиономии, при разлика в температурата на фона.	Висока цена на термокамерите; ниска разделителна способност на изображенията; липса на общодостъпни БД; чувствителност към топлинните шарки на лицето; непрозрачни за стъкло.	Medium	N/A

ФИЗИОЛОГИЧНА БИОМЕТРИКА	Физиономии / Лица (Face)	Video-based	Комбинираща временните характеристики на движението на физиономията с вариациите в гледната точка чрез множество видео кадри.	Възможност за построяване на 3D модел (преодоляване на вариациите в ориентацията и осветлението) или изображение със суперрезолюция; приложимост при сценарий на следене или сегментиране на физиономии.	Повечето съществуващи видео базирани подходи за разпознаване по лица са тествани върху независимо заснети много малки БД.	High	N/A
		Appearance-based	Базиран е на различни статистически подходи, работещи директно със стойностите на пикселите в лицевото изображение.	Превъзходство над геометричния метод за една и съща БД; водещите методи за разпознаване по лица са appearance базирани.	N/A	High	N/A
		Geometry-based	Използва местоположението на отличителните лицеви особености като очи, уста, нос и брадичка и измерва тяхната относителна позиция, ширина и др.	Появата на нови техники за детекция на лицевите особености може да възвърне интереса към този метод.	Липса на робастни детектори на лицевите особености.	Medium	N/A
	Ръка (Hand)	Геометрия на ръката (Hand geometry)	Основава се на измерване на формата на контура на ръката (дължина, ширина, дебелина, обиколка, площ на пръстите, форма на дланта).	Високо ниво на обществено одобрение; ниска изчислителна сложност.	Висока цена и ниска точност (само за верификация, 1:1 сравнения); уникалността на особеностите на ръката не е гарантирана.	Low	Контрол на достъпа с ниско ниво на сигурност; системи за управление на работното време и посещаемостта на работното място.
		Отпечатък на дланта (Palmprint) 2D/ 3D	Състои се от текстурните шарки на кожата на дланта.	Много по-голяма площ, отколкото на пръста; богата на уникални и постоянни особености; висока степен на одобрение.	Липса на обширни изследвания върху 3D отпечатъка на дланта; 2D методите са чувствителни към замърсявания и ~ в осветление.	High	В съдебната експертиза (криминални сцени); физически контрол на достъпа; следене на посещаемостта на работното място; за лична верификация.

ФИЗИОЛОГИЧНА БИОМЕТРИКА		Вени на ръката (Hand vein)	Измерване на различни шарки на кръвоносната система на гърба на ръката, видими чрез заснемане в инфрачервената област.	Висока точност; уникални; стабилни при възрастни (20-50 години); трудни за фалшифициране (кражба или дублиране); безконтактна регистрация, което осигурява изключително удобство и условия за използване в санитарни условия.	Основен проблем със заснемането в близката инфрачервена област е влиянието на видими особености като белези по кожата и косми.	Very High	N/A
		Вени на дланта (Palm vein)	Измерване на различни шарки на кръвоносната система на дланта, видими в близката инфрачервена област.		Заснемането в далечната инфрачервена област среща трудности при вените на дланта, поради ниската термална енергия, която излъчват.		Банки в Япония; системи за сигурност на входа; системи за автентикация на PC; здравеопазване.
	Уши (Ears)		Разпознаването по уши се отнася до автоматичната идентификация на индивиди на база на анатомичните особености на ушите, заснети като 2D/3D изображения.	Имат много богата морфологична структура; не се променят значително по време на човешкия живот; не се променят от емоции; обикновено са видими, поради необходимостта от добра чувасемост; висока обществена степен на одобрение; по-малко стресиращи от пръстовите отпечатьци – безконтактни (без проблеми с хигиената); лесно отдалечено заснемане.	Увеличена чувствителност (поради малката площ на ушите) към закривания (коса, шапки); ниска разделителна способност на изображенията при отдалечено заснемане; най-добрите резултати в литературата са тестовани върху контролирани БД, съдържащи профил на лице, без закривания и на относително близко разстояние.	High → Medium	Човешкото ухо е подходящ източник на данни за пасивна идентификация на хора; разпознаването по уши може да се използва в приложения за ниска и висока сигурност, както и в комбинация с други биометрични признаци като лице.
		2D	Тук се използват 2D изображения от фото / видео камера.	Изображенията на уши са по-сигурни от тези на лица, защото асоциацията на снимка на ухо с принадлежността към даден човек е много трудна.	Производителността на 2D методите, значително се влияе от вариацията в ориентацията и осветлението; възможност от закриване от коса, шал или обеци.	Medium	

ФИЗИОЛОГИЧНА БИОМЕТРИКА	Уши (Ears)	3D	Тук се използват данните от 3D сензори, съдържащи информация за формата на повърхнината, свързана с анатомичната структура на ухото.	Тримерните изображения са относително нечувствителни към осветление; точността на тримерните подходи за разпознаване по уши е значително по-висока от тази на двумерните.	Това е относително нова област на научни изследвания в биометриката.	High	С намаляването на цената и размерите на 3D скенерите и увеличаване на тяхната производителност, тримерното разпознаване на уши би могло да бъде много полезно в биометричните приложения в близко бъдеще.
		Отпечатъци от уши (Earprints)	Отпечатъците от уши са налични върху повърхността на допир до ухото.	Изследвания за приложената сила на натиск по време на слушане показват, че вътрекласовата вариация е относително малка в сравнение с междукласовата.	Високо субективен процес на ръчен анализ; липса на обширно емпирично изследване за уникалността на отпечатъците от уши; научните изследвания в областта са в ранен стадий – само около 60 публикувани статии.	Medium	В съдебната експертиза - при анализ на криминални сцени с наличие на отпечатъци от уши.
	Разпознаване на диктори (Speaker recognition)		Това е задачата за разпознаване на хора по техния глас; могат да се дефинират четири основни подзадачи: идентификация, верификация, сегментация и следене на речта на диктора.	Директно приложение в автоматизираните телефонни услуги; ненаатрапчиво за потребителите и висока приемственост; не е необходим физически контакт; ниска цена на устройствата за запис.	Не толкова висока точност в сравнение с другите биометрики, поради вътрекласовата вариация на говорния сигнал (поради емоционално и здравно състояние, възраст, роднинска връзка); голяма чувствителност към фоновия шум; податливост на атаки чрез пускане на запис на говора.	Medium	Приложения в много различни области като автентикация на телефонните транзакции, контрол на достъпа, съдебната експертиза (чрез запис на гласа на нарушителя), управление на речта (кой и кога говори); основно се използва в режим на верификация (1:1).
	Зъби (Dental Biometric)		Тази биометрика използва данните за структурата на зъбите за автоматично идентифициране на човешки останки.	Значимостта на деналната идентификация стана очевидна след множеството терористични атаки и природни бедствия.	N/A	Very High	Основно приложение за идентификация на жертви от големи бедствия.

	Миризма (Odor)	Човешката миризма може да се използва като отличителна особеност измежду индивидите чрез обучени кучета за проследяване на хора или идентификация на обекти, носени от определен човек.	За разлика от много други биометрики, човешката миризма може да бъде уловена по следа от допир и може да се използва по неинвазивен начин.		Medium	Приложение като съдебен инструмент за идентификация на хора в множество европейски държави.
	Отпечатъци от ходилата (Barefoot)	Отпечатъка от босите ходила се формира от натиска на тежестта на тялото и се използват за търсене на връзка на заподозрени с криминалната сцена.	Много висока разграничителна сила, когато качеството на отпечатъка е добро; може да се използва като мощно доказателство в съда.	Сравнението на отпечатъците от ходилата все още не е рутинна процедура, възприета при криминално разследване.	High	Тяхното използване в криминалните разследвания предхожда използването на пръстовите отпечатъци.
ПОВЕДЕНЧЕСКА БИОМЕТРИКА	Подпис (Signature)	Разпознаването по подпис е исторически установен метод за автентикация на индивиди или документи	Широко обществено възприетие; фалшифицирането на динамиката на онлайн подпис е значително по-трудно отколкото неговата форма; вътрекласовата вариация в подписа във времето се смята за по-ниска в сравнение с много физиологични биометрики.	Използването на off-line подписи като изключително средство за разпознаване е ограничено, понеже изображението на подписа може да бъде лесно копирано.	High (on-line) Medium (off-line)	В исторически план, подписите и фалшифицирането на ръкопис винаги са били от интерес за експертите в криминологията; има приложения в банковите трансакции, споразумения и официални документи.
	Походка (Gait)	Походката се дефинира като стил или начин на вървене; изучаването на походката на човека стана актуално през последните години.	Не се изисква активното съдействие на потрeбителя; има голям диапазон на работа (няколко десетки метра от камерата); ненапращив метод.	Отличителните характеристики на походката не са толкова силни както при традиционните модалности – пръстови отпечатъци и ирис.	Low	Фундаментално предизвикателство тук е разработването на робастни алгоритми, инвариантни на гледната точка, облеклото, повърхността за ходене, вида на обувките, носените предмети; от особен инте-

ПОВЕДЕНЧЕСКА БИОМЕТРИКА						рес е обединяването на походката с лицевото разпознаване за повишаване на точността.
	Отпечатъци от подметката (Footwear)	Това е процесът на идентифициране по оставения отпечатък от подметката на обувката.	Има дълга история, като важно средство за разпознаване в криминалистиката; източник е на потенциално доказателство при анализа на криминални сцени.	Наличие на големи промени от изтъркване на подметката във времето; липса на стандарти за цифров запис на отпечатъците и липса на свободно достъпна БД.	Low	Може да се използва от полицията и други разследващи институции за идентификация на криминално заподозрени.
	Динамика на писането чрез клавиатура (Keystroke Recognition)	Разпознаването по динамиката на писането чрез клавиатура се основава на уникалния начин, по който хората пишат на различни електронни устройства.	Има дълга история (комуникация чрез Морзовата азбука); има потенциала за подобряване на традиционната автентикация чрез потребителско име и парола; ниска цена на разработката.	N/A	Low	Приложение за подобряване на надеждността на текущите съществуващи комерсиални и правителствени приложения.
	Движение на устните (Lip Movement)	Разпознаване на говорещия по динамиката на характерни точки, извлечени от областта на устата.	Използва се като добавка към акустичната информация с цел подобряване на сигурността на системите за гласово разпознаване; лесни за използване и ниска цена.	Системите, базирани само на звук или само на статични изображения, са податливи на атаки чрез притежаването от фалшификатора на запис на речта на потребителя или неговата фотография.	Medium	Приложения при проверка за допускане до спортни събития; достъп по РС; достъп до сгради.
	Стъпки (Footstep)	Относително нова биометрична характеристика, базирана изцяло на сигнала от стъпките на човек върху сензорно чувствителна област.	Сигналите могат да бъдат записани с минимално потребителско съдействие; устойчивост на шум в заобикалящата среда.	Наличие на висока степен на вътрекласова вариация – различни обувки, пренасяне на тежък багаж, различна скорост на вървене.	Low	В медицината (анализ на отклонения в походката); умни домове (определяне на местоположението на човек в стаята); мултимедийни приложения (чрез аудио инсталация в пода).

СОФТ БИОМЕТРИКА	Анатомична	Включва анатомични (пол, етнос, възраст, цвят на очите, височина, тегло, белези) или поведенчески характеристики (акцент на гласа, татуировки, белези), които осигуряват допълнителна, но не и достатъчна информация за точно определяне на идентичността.	Софт биометриката допълва персоналната информация от традиционните биометрични идентификатори като пръстови отпечатыци, лице, ирис, глас, като по този начин може да доведе по подобряване на точността на разпознаване на основните биометрични системи.	Много трудно извличане на софт биометричните признаци със 100% точност; обединяващите схеми за включването на тази допълнителна информация не са изследвани достатъчно задълбочено.	Medium	Софт биометриката е един алтернативен начин на мулти-биометриката за увеличаване на прецизността на разпознаващите системи.
	Поведенческа					

Приложение 3. Публични БД от 2D изображения и 3D модели на уши - брой участници и общ брой изображения, максимални резултати за разпознаване и/или сегментация, описание, предназначение.

Име на БД		Субекти	Общо	Достъп, [референции]	Детекция Разпознаване		Описание
UND [160]	Collection E	114	464	Лиценз, [203, 204]	96.34% [66]	96.75% [66]	За всеки потребител са заснети между 3 и 9 профилни 2D изображения на уши с различна ориентация и осветеност. Създадена през 2002.
	Collection F	302	942	Лиценз, [205]	2D	91% [113]	Съдържа 3D (дълбочинни изображения) и съответните им 2D профилни изображения, заснети през 2003 и 2004.
					3D	100% в/у 142 изоб. [206]	
	Collection G	235	738	Лиценз, [207]			Съдържа 3D (дълбочинни изображения) и съответните им 2D профилни изображения, заснети през 2003 и 2005.
	Collection J2	415	1800	Лиценз, [125]	2D	96.63% [138]	Съдържа 3D (дълбочинни изображения) и съответните им 2D профилни изображения, заснети през 2003 и 2005. За заснемането на дълбочинни изображения и съответните им цветни изображения (с размер 480×640) е използван скенерът Minolta Vivid 910.
					3D	99.38% [139]	
IIT Kanpur [209]	Collection NDOff-2007	387	6940	Лиценз, [208]			Съдържа 3D (дълбочинни изображения) и съответните им 2D изображения на лица. Заснети са т.нар. ротации „yaw“ и „pitch“ на главите на участниците.
	Subset I	190	801	Лиценз, [108]	99.25% [66]	97.35% [66]	Съдържа между 2 и 10 фронтални изображения на уши за субект.
	Subset II	89	801	Лиценз, [108]	98.5% [66]	98.79% [66]	За всеки участник са заснети 9 изображения: три ориентации (поглед направо, ротации на +/-20° в равнината на ухото) и три различни мащаба за всяка.
	Subset III	107	1070	Лиценз, [108]	95.61% [66]		Изображенията са заснети чрез движеща се камера в кръг около субекта на -40°, -20°, 0°, +20°, +40°. За всеки ъгъл са направени по 2 изображения, т.е. по 10 изображения на човек.

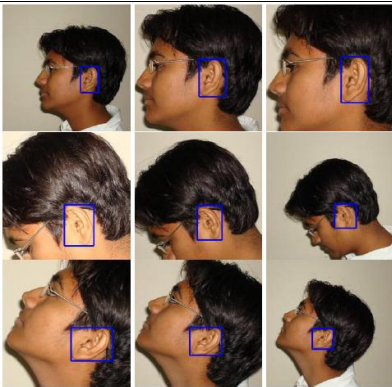
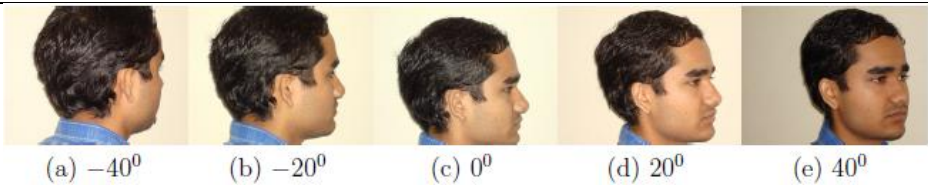
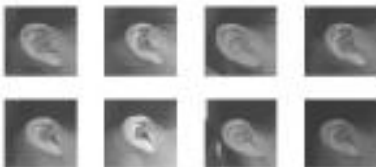
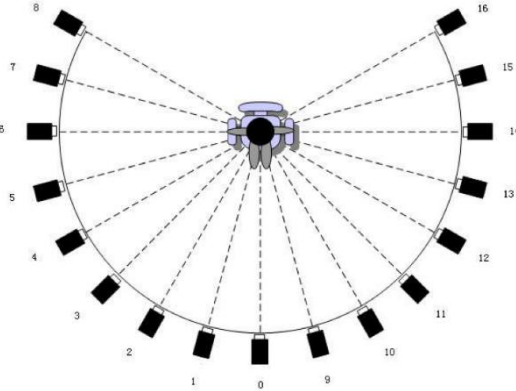
USTB [210]	Dataset 1	60	180	Лиценз, [211]		100% [118]	Съдържа по 3 изображения на десни уши за всеки субект: фронтално; с малък ъгъл на ротация; с различна осветеност. Дата на събиране: юли-август 2002.
	Dataset 2	77	308	Лиценз, [211]	98.05% [212]	100% [213]	Съдържа по 4 десни изображения на уши за човек: 1 фронтално (0°); 2 изображения на +30°, -30°; 1 с промяна в осветеността. Резолюцията е 300×400. Дата на събиране: ноември 2003 - януари 2004.
	Dataset 3	79	1600	Лиценз, [211]		95.1% [117]	Съдържа ротации на главата на участниците на 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60 градуса надясно и от 0 до 45 градуса наляво. Включва изображения на уши с малко и нормално закриване от коса. Резолюцията е 768×576. Дата на събиране: ноември-декември 2004.
	Dataset 4	500	25500	Лиценз, [214]		97.7% [215]	Съдържа изображения на уши и лица, заснети със 17 CCD камери, разположени в кръг през 15 градуса около участниците, които гледат нагоре, надолу, напред, наляво и надясно.
UCR [64]		155	902	Частна, [64]	99.3% [64]	96.77% [65]	Съдържа 2D и 3D изображения, като за всеки субект са направени поне 4 екземпляра. 12 субекта са със частично закрити уши, а 6 носят обеци. Използван е скенерът Minolta Vivid 300 с резолюция от 200×200 пиксела за регистрацията на 3D и 2D изображенията.
WPUT [216]		501	2071	Свободен, [217]			За всеки субект за заснети между 4 и 8 изображения с различна осветеност. Някои от ушите са закрити от шапки, коса, обеци, слухови устройства.
IIT Delhi [218]		121	471	Свободен, [219]			Съдържа най-малко по 3 изображения за субект (14-58 години), заснети в закрито помещение. Резолюцията е 272×204 пиксела. В добавка, тази БД съдържа автоматично нормализирани и изрязани 754 изображения на уши от 212 потребители. Дата на събиране: октомври 2006 - юни 2007.
ScFace [220]		130	4160	Лиценз, [143]			Съдържа фотографии с висока разделителна способност на ляв и десен профил на лицата (20-75 г.). БД е предназначена основно за тестване на алгоритми за разпознаване на физиономии в неконтролирана обстановка на закрито, използвайки обзорни видео камери с различно качество на изображенията.

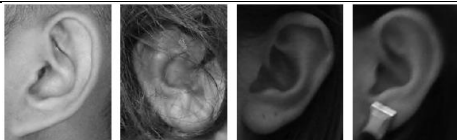
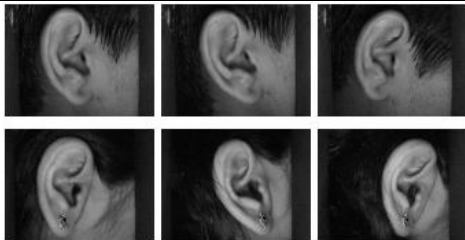
UMIST [221]	20	564 видео- клипа	Свободен, [140]	95% в/у 940 изобр. [222]		Съдържа субекти от различна раса и пол. Всеки индивид е заснет, докато завърта главата си от профил до фронтален изглед. В литературата UMIST БД се използва само за детекция на уши.
YSU [223]	259	2590	Свободен, [223]			За всеки субект са извлечени по 10 изображения от видео поток. Те съдържат ротации от 0° до 90° (от десен профил към фронтален изглед).
NCKU [224]	90	3330	Свободен			Съдържа 37 изображения за субект, заснети от ляв (-90°) до десен профил (+90°) през стъпка от 5° при еднаква осветеност. Някои уши са частично или напълно закрити от коса. Резолуцията е 640×480 пиксела.
UBEAR [225]	126	4430	Лиценз, [142]			Съдържа изображения от видео поток на субекти, гледащи напред, надолу и нагоре. Видеоето е заснето в реалистична обстановка (вариация на осветеността, наличие на коса, обеци), което ги прави подходящи за по-усъвършенстваните методи за разпознаване.
WVU [226]	402	460 видео- клипа	Частна, [226]	90% в/у 3750 изобр. [227]	95% [128]	Съдържа видео последователност, която започва от левия (0°) до десния профил (180°) на субекта. Ушите са частично или напълно закрити – обеци, очила.
XM2VTS [228]	295	252 изобр.	Платен, [144]	100% в/у 252 изобр. [113, 229]	99.2% [98]	Съдържа високо качествени цветни изображения, 32 KHz 16-bit звукови файлове, видеоклипове (ротация на главата) и 3D модели.
NIST [230]	1573	3248	Платен, [231]			БД съдържа сегментирани 8-bit интензитетни изображения на престъпници. 1333 субекта са заснети в профил и фронтален изглед.
FERET [232]	1199	14126	Свободен / Платен, [145, 146]			Голяма БД от изображения на физиономии, заснети в полу-контролирана заобикаляща среда. Някои изображения съдържат ляв и десен профил на лицето и са подходящи за 2D разпознаване по уши. Дата на събиране: между 1993 и 1997 г.
CAS-PEAL [233]	1040	99594	Свободен, [141]			Изображенията са заснети чрез 9 видео камери, разположени равномерно в полукръг, като всеки субект е помолен да гледа надолу и нагоре. БД също включва лицеви изражения, аксесоари (очила, шапки) и 15 посоки на осветление. В литературата CAS-PEAL се използва само за детекция на уши.

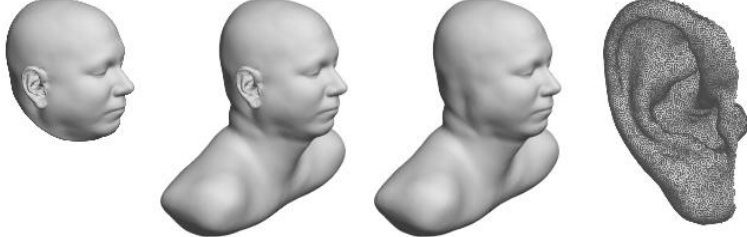
Carreira-Perpiñan [234]	17	85	Свободен, [235]	97.05% [212]	95.05% [236]	Съдържа изрязани и ротирани по 5 изображения на ухо за човек, заснети при еднаква осветеност със съотношение 1.6 на височината към ширината.
AMI [237]	100	700	Свободен	няма данни		Съдържа по 7 изображения (6 на дясното ухо и 1 на лявото) на индивид (19-65 г.), заснети в закрыта среда при еднаква осветеност. Субектите гледат нагоре, надолу, наляво, надясно.
OpenHear [136]	20	20	Свободен	няма данни		Съдържа 3D модели на главата с ушите на участниците, създадени чрез стерео-методи за изчисляване на дълбочината, както и чрез софтуерна реконструкция. Целта на БД е акустична симулация при дизайна на слухови устройства.
SYMARE [137]	61	61	Свободен	няма данни		БД включва 3D модели с висока резолюция на горната част от торса, главата и ушите на 48 мъже и 13 жени. 3D информацията е пресъздадена чрез MRI скенер. Целта на БД е подобряване на индивидуализацията на 3D аудио за бъдещи персонални аудио устройства.

Приложение 4. Извадка на изображения от разгледаните БД в Приложение 3.

База от данни		Извадка от примерни изображения			
UND [160]	Collection E	 Цветни 2D изображения			
	Collection F	 Цветни 2D / 3D дълбочинни изображения			
	Collection G	 Цветно 2D изображение			
	Collection J2	 цветни 2D изображения 3D профили на дълбочинни изображения с вариация в ориентацията и мащаба			
	Collection NDOff-2007	N/A			
ИТТ Kanpur [209]	Subset I	 Фронтален изглед на изображения, съдържащи уши			

USTB [210]	Subset II	 Профил на субект гледащ на право, с ротация на of $\pm 20^\circ$ и при 3 различни мащаба
	Subset III	 Изображения от ротация на субекта на -40°, -20°, 0°, $+20^\circ$, $+40^\circ$ градуса
	Dataset 1	 Нормални изображения на уши, тривиална ротация и различна осветеност
	Dataset 2	 1 профилно изображение (0°); 2 - с ротация на $+30^\circ$; 1 - с вариация в осветеността
	Dataset 3	 изображения с ротация от 0 до 45 градуса
	Dataset 4	 17 CCD камери са разположени около субекта през 15° градуса

UCR [64]	 Дълбочинни 3D изображения и техните съответни 2D изображения																												
WPUT [216]	 Изображения с различна осветеност; наличие на обеци и закриване от коса																												
IIT Delhi [218]	 3 изображения на уши за субект с различна ориентация, заснети в закрыта среда																												
ScFace [220]	N/A																												
UMIST [221]	 Субект с ротация от профил към фронтален изглед																												
YSU [223]	<table><tr><th>Angle</th><th colspan="6">Reference Images</th></tr><tr><td>0°</td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>30°</td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>90°</td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table> По 10 изображения за субект са извлечени от видео клип (с ротация от 0° до 90°)	Angle	Reference Images						0°							30°							90°						
Angle	Reference Images																												
0°																													
30°																													
90°																													
NCKU [224]	 37 изображения за субект са заснети от ляв (-90°) до десен (+90°) профил през стъпка от 5° при еднаква осветеност																												
UBEAR [225]	 (a) Good quality image (b) Over illumination (c) Blurred image (d) Earrings occlusion (e) Strong hair occlusion (f) Weak hair occlusion																												
WVU [226]	N/A																												

XM2VTS [228]	 Висококачествени цветни изображения на фронтален, ляв и десен профил
NIST [230]	 Профилен изглед на изображения на престъпници
FERET [232]	 Примерни изображения от профилната БД с физиономии
CAS-PEAL [233]	 Едновременно заснемане на изображения чрез масив от видеокамери
Carreira-Perpiñan [234]	 Изрязани и ротиращи изображения с еднаква осветеност
AMI [237]	 BACK FRONT UP DOWN LEFT RIGHT ZOOM Изображения на уши, заснети от различна посока
OpenHear [136]	 Примерна извадка на 3D физиономии и уши
SYMARE [137]	 За всеки субект от БД са налични 4 вида 3D модели – на глава, торс, или уши

