



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

Атанас Филипов Николов

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

А В Т О Р Е Ф Е Р А Т

НА ДИСЕРТАЦИЯ

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 Димов

София, 2016 г.

Дисертацията е обсъдена и допусната до защита на разширено заседание на секция „Математически методи за обработка на сензорна информация“ на ИИКТ-БАН, състояло се на 06.01.2016 г.

Дисертацията съдържа 156 стр., в които 70 фигури, 17 таблици и 10 стр. литература, включваща 237 заглавия.

Защитата на дисертацията ще се състои на2016 г. от часа в зала на блок на ИИКТ-БАН на открито заседание на научно жури в състав:

1.
2.
3.
4.
5.

Материалите за защитата са на разположение на интересуващите се в стая, бл. на ИИКТ-БАН, ул. „Акад. Г. Бончев“.

Автор: *Атанас Филипов Николов*

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

Обща характеристика на дисертацията

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 личността¹ по формата на ушите. Тези две, донякъде различни области на изследователски интерес, са съвместени в една обобщаваща детайлизирана схема (виж Фиг. 2).

Свързващото звено между видеостабилизацията и 3D разпознаването чрез 3D модели за БД с прецеденти е в 3D реконструкцията на обекти от обхождащ видеоклип. 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. Също така, стабилизацията на заснемания обект е пряко свързана и с неговото ефективно сегментиране от фона (виж Фиг. 2, 2D Сегментация). 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.

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.

Методите за видеостабилизация целят подобряване качествата на нестабилно заснето видео (любителско и/или професионално) в точка или по гладка траектория (виж Глава 2, §2.1 и §2.2). 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 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 РС 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 се разчита на съвместното използване на вградените в хардуера им инерциални сензори (жироскопи и/или акселерометри) заедно с прецизни и бързи софтуерни подходи.

¹ Тук взаимнозамеяемо използваме философското понятие „личност“ с „индивид“ при разпознаването.

В тази връзка ние разработихме един бърз и ефективен софтуерен метод за 2D видео стабилизация, 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). В близка перспектива предвиждаме негова 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 хора [3], свидетелство за което са и множеството дисертационни резултати [4, 5], книги [6, 3] и обзорни статии в областта [7, 8, 9]. 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 [8]. 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 [10]. Lei et al. показват, че ушите могат да се използват и за класификация по пол [11]. 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 окончателно установено, че ушите на всички хора са уникални. Освен това, структурата на ушите не се променя значително с възрастта [12]. 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 [13], и могат да се използват в биометричната идентификация и верификация на индивиди. Въпреки че повече от 99% от биометричните системи, достъпни на пазара, са базирани на пръстови отпечатъци, лицево или ирисово разпознаване [14], използването на ушите има и своите предимства спрямо тях [9], 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 се оформи в дискусия на BIOMET'2014, като развитието и реализацията на тази идея стана възможно благодарение на модерния 3D лазерен скенер, закупен по проекта AComIn на ИИКТ-БАН (<http://www.iict.bas.bg/acomin/>).

Цели и задачи на дисертацията

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

1. Разработване на методи за ефективна стабилизация на видеоклипове, както по изградена 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 precedenti.

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 съседи, и др.;
- *Компютърно зрение*: 3D реконструкция на обекти, стереозрение, фотометрия.

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“.

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 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.

Съдържание на дисертацията

Настоящата дисертация се състои от списък на фигурите, списък на таблиците, речник на термините, увод, 3 глави, заключение, списък на цитираната литература и 4 приложения, общо 156 страници. Основното съдържание е поместено на 111 страници, а изложението е придружено от 70 фигури и 17 таблици. Списъкът на цитираната литература включва 237 заглавия.

Глава 1. Обзор на основните резултати в областта

Глава 1 съдържа обзор на основните резултати по направленията видеостабилизация и биометрично разпознаване на личността (индивида) по формата на ушите.

1.1 Видеостабилизация

Широкото използване на софтуерните средства за видеостабилизация за постигане на плавност на движението в неспринудено заснето видео набира все по-голяма популярност през последните години. Това се обуславя от една страна на непрактичното използване на скъпо и сложно стабилизиращо оборудване за любителски цели (стативи, камери, движещи се върху релси, операторски колички, преносими стабилизиращи механизми, закрепени към тялото и др.), както и на възможността за значително прецизиране на хардуерните алгоритми за работа в реално време, възползвайки се от вградените инерциални сензори (жироскопи и/или акселерометри) в преносимите устройства.

1.1.1 Софтуерни методи за видеостабилизация

В литературата са известни множество софтуерни методи за видеостабилизация по гладка траектория, които може да се класифицират като 2D, 3D, както и доста перспективните напоследък 2.5D.

А. 2D методи за стабилизация

Двумерните техники за видеостабилизация изчисляват 2D модел на движението (като твърдотелна сцена, афинен, перспективен, виж **Приложение 1** от дисертацията) 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, 15, 16].

В случаите на приблизително равнинни сцени (малка девиация в дълбочината на обектите) или когато движението на камерата е чисто ротационно без значение от дълбочината в сцената, 2D стабилизиращите методи са достатъчно прецизни, работят върху целия кадър, изискват следене на малък брой характерни точки (траектории във времето) и консумират минимална изчислителна мощност.

Б. 3D методи за стабилизация

Тримерният подход за видеостабилизация е описан за първи път от Buchler et al. [17]. 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 [18].

След като през 2009 г. Liu et al. [19] реализират за първи път 3D видеостабилизация върху динамични сцени, въпреки високото качество на плавността 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 в следните ситуации: (i) липса на паралакс; (ii) промяна на фокусното разстояние; (iii) собствена видеостабилизация на камерата; (iv) “rolling shutter” ефектът. Производителността на SfM също е проблем, понеже изисква глобална нелинейна оптимизация, както и изискването за произволен достъп до цялото видео.

В. 2.5D методи за стабилизация

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, 20, 21, 22, 23].

1.1.2 Исторически преглед на върховите постижения при софтуерната видеостабилизация

В този раздел разглеждаме развитието във времето² на върховите постижения на различните методи за видеостабилизация (2D, 3D и 2.5D), всеки един от тях, претендира за постигане на *“state-of-the-art”* резултат в сравнение с предходните методи.

1.1.3 Хардуерни методи за видеостабилизация, използващи инерциални сензори

Класически примери за 3D стабилизация чрез инерциални сензори са [24, 25, 26]. 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. 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

Тук разглеждаме 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. Този опростен подход за 2D видеостабилизация (но само по трансляция) работи със скорост над 100 fps за видео 1280×960 и е описан в White Paper на Texas Instruments (TI) [27, 28]. Методът е интересен за нас, защото използва BSC чипа [27], от състава на DaVinci процесора (TMS320DM365) на TI [28], и което бе изрично изискване на фирмата заявител на първата наша разработка по темата. BSC позволява ефективно изчисление на вертикалните и хоризонталните хистограми (акумулиращи проекции) на дадено изображение. Поточно, 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 SAD хистограми – по 2 за всяко подизображение. Ние заимстваме този BSC/SAD подход, който позволява ефективно оценяване на оптималния вектор на отместването, без да изисква специални ускорители, например FFT, като при нашите експерименти емулираме BSC и модифицираме SAD (виж Глава 2, §2.1.1).

1.2 Биометрика

Биометриката е науката за *автентикация* (биометричната автентикация е синоним на биометрично разпознаване, означаващо идентификация или верификация [29]) на индивида въз основа на неговите физически характеристики (лице³, ирис, пръстов отпечатък, геометрия на ръката, уши, глас и др.) или поведенчески характеристики (походка, ръкописен стил, подпис, динамика на писането чрез клавиатура и др.) [30].

В тази част от дисертацията представяме подробна справка и сравнение на предимствата и недостатъците на известните биометрични модалности, според литературните източници. Обоснована е актуалността на разпознаването на хора по формата на ушите,

² В периода 2009-2013 г. 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.

³ Под „лице“ ще разбираме „физиономия“, а не личност, които ще използваме като взаимнозаменяеми.

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 2D и 3D методите с най-високите резултати по детекцията и разпознаването на ушите.

1.2.1 Традиционни методи за автентикация

Традиционно, автентикацията на индивида (личността) се организира 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 отгатнато [31].

Появата на биометриката цели да адресира съпътстващите проблеми на традиционните методи за сигурност. Нейната популярност се дължи на по-голямата сигурност [32], 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 видеонаблюдение [33].

1.2.2 Класификация на биометричните технологии

Една възможна класификация на биометричните технологии може да бъде направена в зависимост от *размерността на входния сигнал*: 1D, 2D или 3D [34]. Повечето от техниките за биометрично разпознаване са базирани на 1D сигнали – глас [35] и on-line подпис [36], или 2D изображения – off-line подпис [37], пръстови отпечатъци [37, 38], лица [39, 40], ирис [41, 42], отпечатък на дланта [43], геометрия на ръката [44], уши [7, 8, 9]. Поради това, че 2D представяне на биометричните модалности лесно може да бъде фалшифицирано, се наблюдава засилен интерес към изследването на нови методи за разпознаване по 3D биометрични характеристики, като: 3D лица [45, 46]; 3D пръстови отпечатъци [47]; 3D отпечатък на дланта [48]; 3D уши [49, 50, 51, 4, 5, 6]. Тези 3D данни 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

Тук се разглеждат в хронологичен ред възникването на базовите биометрични модалности и технологии за разпознаване на индивиди, на които текущо се базират най-съвременните биометрични системи на пазара. Това са *пръстовите отпечатъци* (началото на 20-ти век); *ириса* (1936 г.); *физиономии* (60-те години на 20-ти век); *подписът* (1965 г.); *геометрията на ръката* (1974 г.); *четец на пръстови отпечатъци* (1975 г.); система за *разпознаване на диктори* (1976 г.); *автоматичното разпознаване на лица в реално време* (1991 г.); първата автоматизирана система за разпознаване на *отпечатъци от дланта и пръстите* (1994 г.); първата комерсиална система за *разпознаване по ириса* (1995 г.); ФБР създава база данни с *ДНК* информация (1998 г.); първата комерсиална система за *разпознаване по кръвоносната система на човека* (2000 г.).

1.2.4 Изисквания към биометричните модалности и системи

Според Jain et al. [13], за да бъдат използвани в биометрията, човешките физиологични или поведенчески характеристики трябва да бъдат *универсални, уникални, постоянни и лесни за регистрация (измерване)*. 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 свързаност между множество електронни устройства. Също, от анонимни данни за местоположението чрез проследяване на координатите на мобилния телефон, всеки може да разбере „кой си ти“. Шаблонът на движението, нанесен върху карта, създава уникален отпечатък за всеки човек.

1.2.6 Моно- и мултибиометрични системи

Повечето съвременни биометрични системи типично използват единствена биометрична модалност за установяване на идентичността на индивида, което води до несигурност в системата и намалена степен на разпознаваемост [52]. За ефективното преодоляване на недостатъците на едномодалните системи особен интерес представляват мултимодалните биометрични системи. 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 монобиометрична система [53].

1.2.7 Сравнителен анализ на биометричните модалности

Повече от 99% от биометричните системи, налични на пазара, са базирани на пръстови отпечатъци, лицево или ирисово разпознаване [14], подходящи за работа както в режим на верификация (1:1 сравнения), така и в режим на идентификация (1:N сравнения), виж Фиг. 11.



Фиг. 11. Най-популярните биометрични модалности за автентикация на индивиди:
а) пръстови отпечатъци; б) физиономии; в) ирис. *Заимствано от Jain*⁴.

⁴ http://biometrics.cse.msu.edu/Presentations/AnilJain_Biometrics_TechnologyforHumanRecognitionNCVPRIPG13.pdf

Пръстовите отпечатащи и разпознаването по лица едва ли някога ще бъдат заместени от други биометрични модалности поради наличието на огромни правителствени бази данни в отделните държави.

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

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.

А биометрични модалности като разпознаване по походката [54], динамика на писане на клавиатура [55], кръвоносни съдове на очни дъна, кардиограма [56], околоочният контур [57], *ушите*, белези и татуировки, стави на пръстите [58], биоелектричната активност на мозъка, са все още на ниво лабораторни експерименти (Фиг. 15).



Фиг. 15. Биометрични модалности в лабораторен стадии на разработка. Заимствано от Jain⁵.

За пълнота ще споменем и други биометрични модалности като ретина, термограми на лицето и ръката, зъби, 3D отпечатък на дланта, отпечатък от ухото, отпечатащи от ходилата [59], отпечатък от подметката, миризма, движение на устните, стъпки, форма на носа [60], нокти на пръстите [61], предни синуси [62], генериран звук от вътрешното ухо [63].

В *Приложение 2* от дисертацията 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 приложения⁶.

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 [64]: *комерсиални*; *държавно-административни*; *приложения в криминалистиката и съдебната експертиза*, които могат да се разложат на по-детайлна таксономия от 12 категории за обхващане на пълния спектър на биометричните приложения [65].

⁵ http://biometrics.cse.msu.edu/Presentations/AnilJain_Biometrics_TechnologyforHumanRecognitionNCVPRIPG13.pdf

⁶ 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).

1.2.9 Разпознаване на индивиди по формата на ушите

За първи път идентификация на индивида по морфологията на ухото се споменава през 1890 г. във Франция от Bertillon [66]. Въпреки, че научните изследвания за използването на човешкото ухо за целите на разпознаването в биометриката се намират в начален стадий на разработка, ушите са обект на особен интерес, защото се смятат за надеждна отличителна характеристика за биометричното разпознаване на личността [3].

Известно е, че формата на ухото еволюира по време на ембрионалния растеж от шест развиващи се базови възела. Това обуславя и неговата структура, която не е напълно произволна. Затова, детайлната структура на човешкото ухо се разглежда като универсална и уникална, въпреки че все още не е окончателно установено, че ушите на всички хора са уникални. Освен това, структурата на ушите не се променя значително с възрастта [12]. Ушите са лесни за (отдалечено) заснемане по ненаатрапчив начин. Горните особености водят до извода, че ушите удовлетворяват изискванията, формулирани от Jain [13], за да бъдат използвани в биометричната идентификация и верификация на личността (виж §1.2.4).

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 отпечатаци, ирис и лице [9]:

- 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);
- За разлика от пръстовите отпечатаци 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 отпечатаци;
- 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.

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

- Увеличената чувствителност към припокривания (от коса, шапки и др.) поради относително по-малкия размер на ухото, съчетано с ниска разделителна способност на изображенията на уши, както и вариацията в ориентацията и осветлението водят до намаляване на производителността на 2D методите за разпознаване.

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. Разпознаването на уши разчита на сегментираните изображения на уши за извличане на признаци от тях и след това прилагане на определена стратегия за сравнението им.

Двумерни (2D) подходи за разпознаване на уши

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 [9].

Тримерни (3D) подходи за разпознаване на уши

Биометричните подходи за разпознаване на уши разчитат на морфологичните свойства на ухото, защото 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 вид и отличителните черти на ухото. 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 повишаване на точността на разпознаването по ушите.

Измежду най-значимите разработки в областта на тримерното разпознаване на уши са тези на Chen и Bhanu [50, 67] и Yan и Bowyer [51, 68].

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

В тази част са разгледани публично известните до момента БД от 2D изображения и 3D модели на човешки уши. В *Приложение 3* (от дисертацията) 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 съдържанието и предназначението им [7, 8, 9, 69]. В *Приложение 4* (от дисертацията) са показани примерни изображения на разгледаните БД от Приложение 3.

Глава 2. Методи за бърза и качествена 2D видеостабилизация

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 рекламни видеоклипове и/или произволни части от тях в базов TV видеопоток (§2.3).

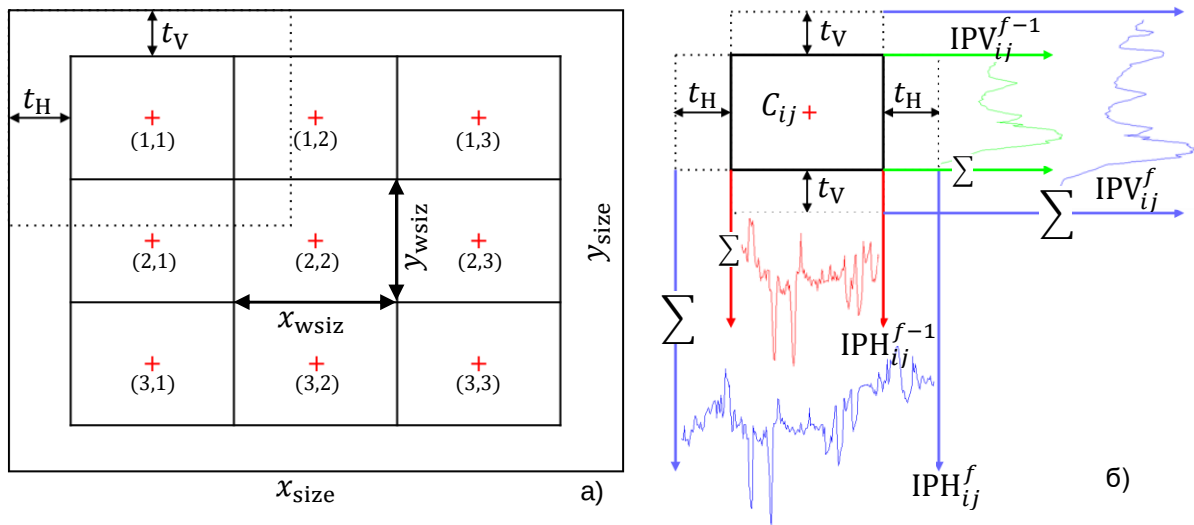
2.1 Видеостабилизация по гладка траектория

В настоящото изследване ние предлагаме нов, бърз и ефективен, софтуерен метод за 2D видеостабилизация по гладка траектория, при което се цели изглаждане на високочестотната нестабилност на камерата със запазване на оригинално замислената от оператора траектория на движение. Методът, подобно [27, 28], е предназначен за работа в реално време върху портативни устройства и е адаптиран (без ограничение на общността) към BSC чипа на TI за търсене на „векторите на движение“ между двойка кадри от видеото чрез корелации по SAD метода между съответните 1D акумулиращи проекции по хоризонтали и вертикали.

2.1.1 Определяне на векторите на движение

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.

Тук разглеждаме интегралното определяне на векторите на движение⁷, подобно на [27, 70, 71]. Целта е (желаната, търсената) съвместимост 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 [27, 28], което определя броя 9 на търсените вектори на движението $\vec{t}_{ij}$ ($i, j = 1, 2, 3$) между съответните 9 центъра на подизображенията в текущия и предходния кадър (виж Фиг. 18а).



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.

Ние също използваме идеята за оценка на 9-те вектора на движението чрез SAD подхода [27], 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 области с център C_{ij} за текущия кадър (f) и предходния кадър ($f - 1$), виж Фиг. 18б.
- Съответните вектори на движението $\vec{t}_{ij} = (t_{xij}^{(f)}, t_{yij}^{(f)})$ се получават като min по SAD:

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

и аналогично за $t_{yij}^{(f)}$ чрез вертикалните акумулиращи проекции IPV_{ij}^f .

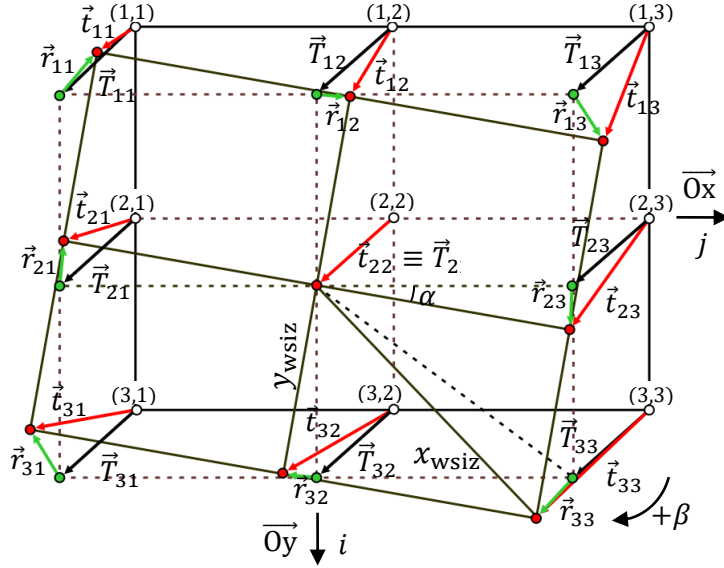
Нашето подобрение на SAD подхода се състои в това, че вместо класическите проекции IPH_{ij}^f и IPV_{ij}^f , ние провеждаме горното изчисление върху нормализираните по средна стойност проекции $\bar{IPH}_{ij}^f$ и $\bar{IPV}_{ij}^f$, при това – средната стойност е „плаваща“:

⁷ Въпреки че във видеофайловете с трег 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 с цел стабилизация в момента на заснемане.

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

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 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 $AX = B$, чието решение X по метода на най-малките квадрати е следното:

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

Така, за финалното решение, получаваме:

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

където:

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

$k = y_{wsiz} / x_{wsiz}$, 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), след елиминиране на „големите отклонения“ според Гаусовото разпределение на грешките, получени при първия LSM пас.

По този начин извеждаме 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” области, като с увеличаване на броя на търсените векторите на движение постигаме още по-прецизно решение.

2.1.3 Мултипликация на базовия векторен модел за 3×3 → 9×9 области

Тук е предложена мултипликация по аналогия на базовия векторен модел (виж Фиг. 21), 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”.

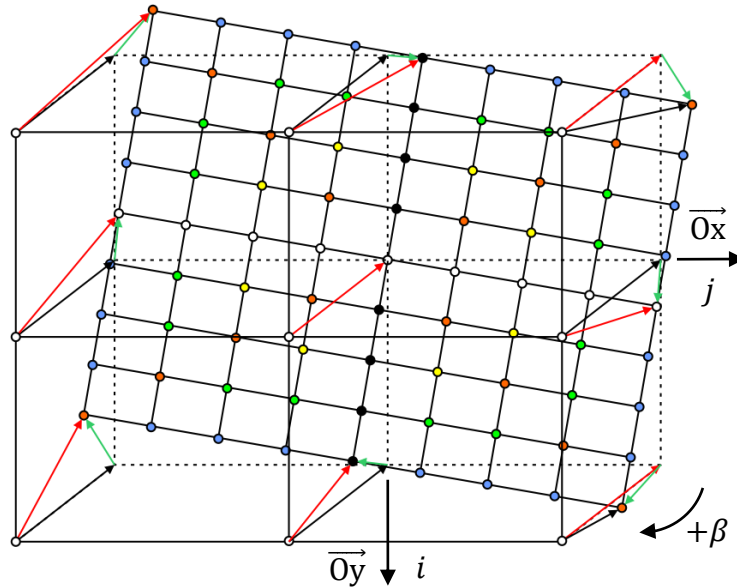
По аналогия на векторния модел с “3×3” разбиване съставяме система от 162 по-компонентни уравнения (по $\vec{Ox}$ и $\vec{Oy}$), в която неизвестните параметри (T_x , T_y , r_x , r_y) се намират отново по матричния LSM метод.

За финалното решение получаваме:

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

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 определянето на коефициентите a, b, c, d виж съответния раздел от дисертацията.

Тук отново (както при метода “3×3”) може да се приложи LSM методът втори път с цел подобряване на точността на решението.



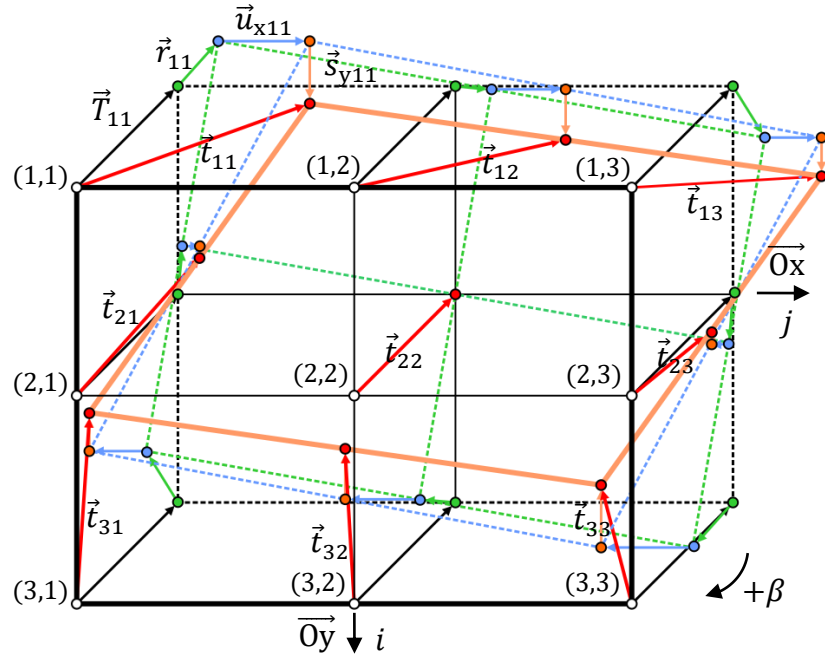
Фиг. 21. Мултипликация на базовия векторен модел за разбиване на кадрите на “9×9” области, отчитащ трансляция и ротация между два последователни кадъра във видеото.

2.1.4 “Rolling shutter” разширение на базовия векторен модел

В тази част е предложено 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.

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. По тази причина, освен към обичайните трансляция и ротация, в резултат от потрепването на камерата се наслагват още две съставки (накланяне и мащабиране), които трябва да се компенсират за осигуряване на оригиналната твърдост 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 и Фиг. 29).



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

В Таблица 12 е представена базовата система от уравнения за разширения векторен модел за случая с “3×3” разбиване на кадрите, т.е. за 9 векторни или 18 компонентни уравнения.

Таблица 12. Система от 18 покомпонентни уравнения за “3×3” разширения векторен модел.

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

Методът е успешно експериментиран и за случая с “9×9” разбиване на кадрите, без извеждане на прякото решение за търсените параметри. Експериментите са проведени чрез функцията в Матлаб за решаване на линейни системи по LSM метода, където са необходими единствено коефициентите пред съответните базови неизвестни параметри за трансляция и ротация $\vec{T}_{ij}$, $\vec{r}_{ij}$, както и коефициентите за допълнителните параметри на накланяне и мащабиране $\vec{u}_{xij}$, $\vec{s}_{yij}$.

2.1.5 Експериментални резултати

За количествена оценка на качеството на предложените алгоритми за видеостабилизация по гладка траектория ние използваме мярката *Interframe Transformation Fidelity* (ITF), предложена в [72] и приложена също в [70], която модифицираме за граничните случаи:

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

където PSNR 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. Но ние не използваме класическата PSNR оценка, а нейна модификация, която разглежда подобие между съответните пиксели от съседните кадри в рамките на предварително дефиниран праг, чиято цел е компенсиране на малки отмествания между стабилизираните кадри в рамките на един тематично свързан видеоклип (шот).

Експериментите са проведени върху 5 видеоклипа (от различни шотове), с размер 640×480 и с различна степен на трептене, оценена по скалата от 1 (леко) до 5 (силно).



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

Съответните ITF стойности за качеството на експериментите върху тестовите видеоклипове 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	original ⁸	3x3/noAVG/I ⁹	3x3/AVG/I ¹⁰	3x3/noAVG/II ¹¹	3x3/AVG/II ¹²	9x9/AVG/II ¹³	FAST ¹⁴
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.

- ⁸ (original): → в тази колона са дадени ITF стойностите на оригиналните (нестабилизирани) клипове;
- ⁹ (3×3/noAVG/I): → използва 3×3 схема на разделяне, без IP нормализации и прилагане на LSM само веднъж;
- ¹⁰ (3×3/AVG/I): → същото като горното, но с нормализация на акумулираните проекции (IP);
- ¹¹ (3×3/noAVG/II): → използва 3×3 схема на разделяне, без IP нормализации, но прилагане на LSM два пъти;
- ¹² (3×3/AVG/II): → същото като горното, но с нормализация на акумулираните проекции (IP);
- ¹³ (9×9/AVG/II): → използва 9×9 схема на разделяне, с IP нормализации и втори пас на LSM (LSM-II);
- ¹⁴ (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 1**). 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.:** В Матлаб реализацията на изследваните алгоритми, в таблицата не са включени времената за външните за алгоритмите операции, като: (i) *време за четене* ≈ 10 ms/frame; (ii) *обратна трансформация* $\approx 20^{15}$ ms/frame; (iii) *запис на стабилизираните 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.

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. Методът е базиран на прецизен векторен модел, позволяващ нарастващи по сложност интерпретации на движението – от твърдотелен (CCD) модел на трансформация между кадрите до “rolling shutter” (CMOS) модел, характерен за болшинството портативни устройства, като мобилни (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. Проведените експерименти върху тестови видео клипове са обещаващи за бъдещото развитие на метода в насока за неговата евентуална комбинация с (известните) методи за видеостабилизация, използващи инерциални сензори, за постигане на една задоволително прецизна видеостабилизация.

2.2 Видеостабилизация за индустриални високоскоростни камери

В този раздел от дисертацията се разглежда проблемът за видеостабилизацията в „точка“, където се цели стабилизация на всички видео кадри спрямо един избран референтен кадър за произвеждане на ново видео, 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, заснети от високоскоростни камери. За тази цел е предложен комбиниран метод “3×3OF9×9”, съчетаващ базовия “N×N” метод, с известния метод на „оптичния поток“ (OF), който при запазване на високата точност неутрализира ефекта „пълзене на сцената“, характерен и за други известни разработки по стабилизация в „точка“.

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-6¹⁶, закупена по проекта AComIn на ИИКТ-БАН. Необходимостта от видеостабилизация при този вид камери възниква, когато е невъзможно да се избегне динамиката на средата, т.е. камерата не може да бъде статично фиксирана. В този случай трептенето във видеото е причинено от нестабилности на фундамента на 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

¹⁵ При извършването на обратното възстановяване от оригиналните към стабилизираните кадри, последователността на трансформации е в обратен ред на изчислените, за да се избегне антикомутативността.

¹⁶ <http://www.nacinc.com/products/memrecam-high-speed-digital-cameras/hx-6/>

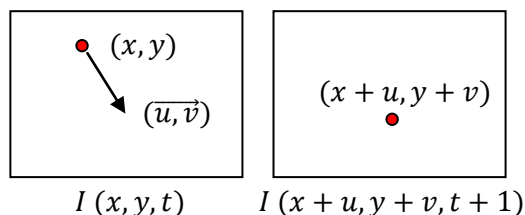
сътресения в близост до камерата); (3) снимане на ръка от превозно средство; (4) снимане от ръка „на момента”, в „полеви условия” и др.

Поради спецификата на камерата – скорости на снимане от 50 до 370330 fps, видеоостабилизацията се провежда в off-line режим. Тук, 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 „пълзене 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*).

2.2.2 Методът на Оптичния поток

В този раздел се разглежда същността на метода на оптичния поток [74, 75], 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 в 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 (Фиг. 31).



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

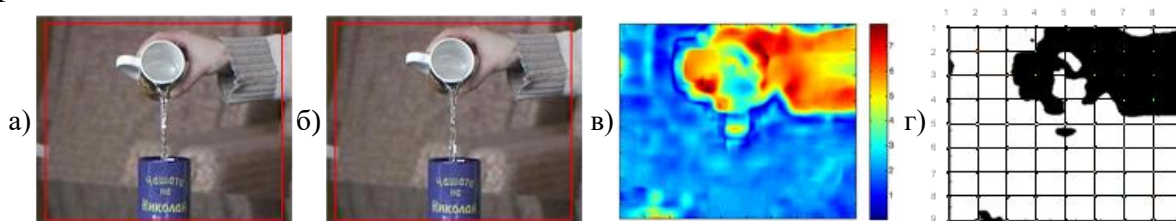
2.2.3 Комбиниран подход “3×3OF9×9” за видеоостабилизация в точка

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. За преодоляване на недостатъците на трите метода поотделно, разглеждайки ги независимо, **е предложен оригинален комбиниран подход**, наречен “3×3OF9×9”, за видеоостабилизация в „точка“. Той се състои от два последователни етапа на стабилизация. На първия етап се използва “3×3” методът (виж §2.1.2) за бърза груба стабилизация, след което се прилага “9×9” методът (виж §2.1.3) върху 9×9 центъра на съответните области от векторното поле на „оптичния поток“.

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. 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.

На втория етап, чрез бинаризиране на картата от големините на векторите на оптичния поток, се изключват определени вектори на движението, участващи в схемата на

“9×9” метода, и по този начин се подобрява точността на стабилизацията на статичния фон.



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 [76], приложена към “9×9” векторите на движение.

2.2.4 Експериментални резултати

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

1. Видеоклип, свален от интернет страницата на производителя на камерата NAC Memrecam НХ-6, съдържащ запис на контролиран взрив (виж Фиг. 33), където е налице огромен динамичен обект (облаци с пушек от взрива). Видеото е с разделителна способност 1280×720 пиксела и служебна скорост от 25 кадъра в секунда.



Фиг. 33. Стабилизация на видеоклип на контролиран взрив: а) първи кадър; б) последен кадър след стабилизация в точка; в) последен кадър след стабилизацията по траектория.

2. Видеоклипове, заснети от ръка с нарочно (симулирано) трептенене, 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 fps.

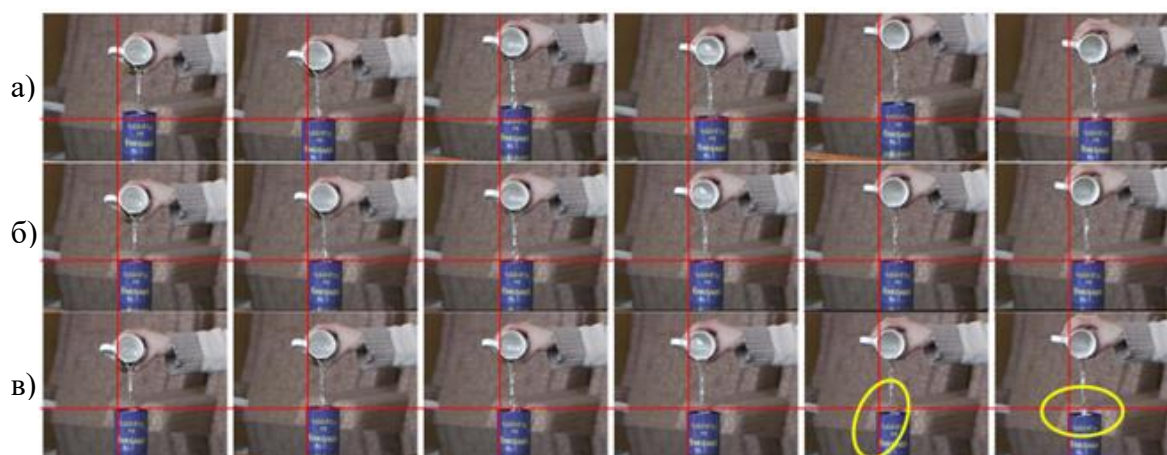
В Таблица 15 са дадени максималните грешки по Ох и Оу посоките за съответните скорости на видеоклиповете с преливане на вода за двата етапа на стабилизацията.

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

Резолуция 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 %

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 явно е указана опцията за стабилизация в статична камера (фон).



Фиг. 36. Нашият резултат за стабилизация в статичен фон в сравнение с този от ААЕ: 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 нашия базов “N×N” векторен модел с известния метод на „оптичния поток“, с детекция и отстраняване на подвижните обекти в сцената за неутрализиране на ефекта „пълзене на сцената“. Методът се отличава с достатъчно бързодействие и прецизност на стабилизацията. Той е предназначен за видео експерименти (напр. 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 „пълзене 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 поток

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. Текущо, задачата за следенето на излъчването на реклами според условията на договора с дадената телевизия 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.

2.3.1 Формулировка на задачата

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



Фиг. 37. TV видео, съдържащо множество рекламни блокове, в рамките на новинарска емисия.

Вход:

1. Множество от 30 мин. записи в MPEG4 формат (25 fps, с размер 448x336 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.

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 $\sim$ 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“ чрез дължините им в брой кадри. Този алгоритъм има предимство в случая, когато броят на рекламните шаблони е голям. Същността на алгоритъма се състои в представянето на *видеозаписите като редици 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) после-

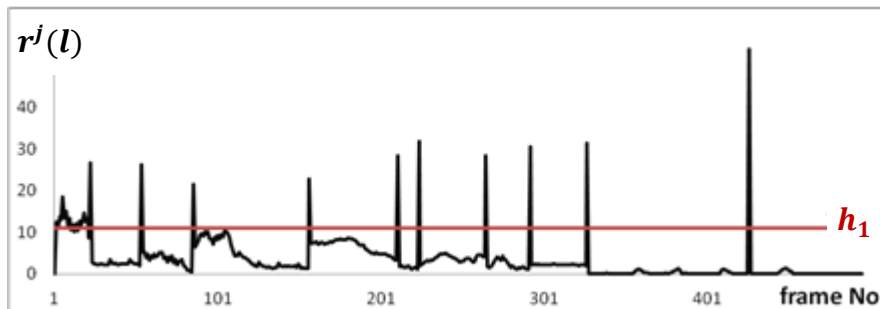
дователни кадъра, като указание за следваща сцена. По този начин, първоначално можем да сравняваме сцените по дължина (продължителност в брой кадри) и само за „подозрителните“ сцени (т.е. с близки дължини) извършваме покадрови сравнения, от типа „едно-към-много“. Предимството на тази стратегия е в много по-малкия брой покадрови сравнения, отколкото при предходните сценарии.

2.3.3 Описание на предложения алгоритъм по третия сценарий

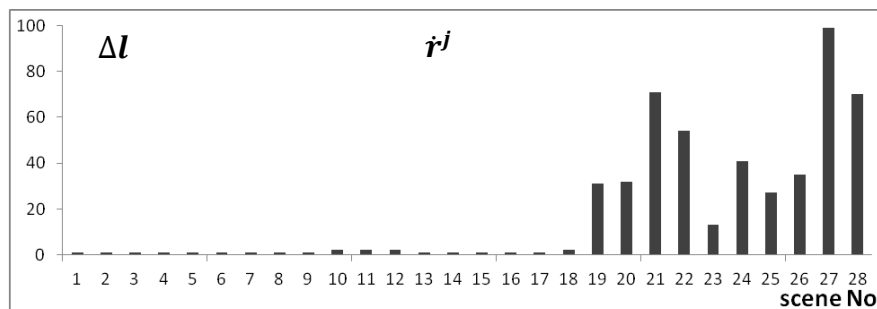
Понеже се стремим към работа в реално време (или по-бързо), за детекцията на сцени ние предлагаме един много прост и бърз подход, основан на разликата между два последователни кадъра в рамките на предварително избран праг. Този подход осигурява достатъчно надеждна детекция на цели реклами и/или произволни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

На Фиг. 41 и 42 са показани резултати от експеримент за детекция на сцени и представяне на рекламен видеоклип (~20 сек.) като целочислена редица. При първите 20 кадъра се наблюдава голямо струпване на сцени, което се дължи на факта, че началото на рекламната включва едновременно движение на обекти и на фона, което води до рязка промяна между два последователни кадъра. Тази промяна е указание за следваща сцена според прага h_1 , приложен към диференциалното представяне $r^j(l)$ на даден рекламен клип, където $l = 1, \dots, (N_j - 1)$ е последователният номер на кадър от j -тата реклама.

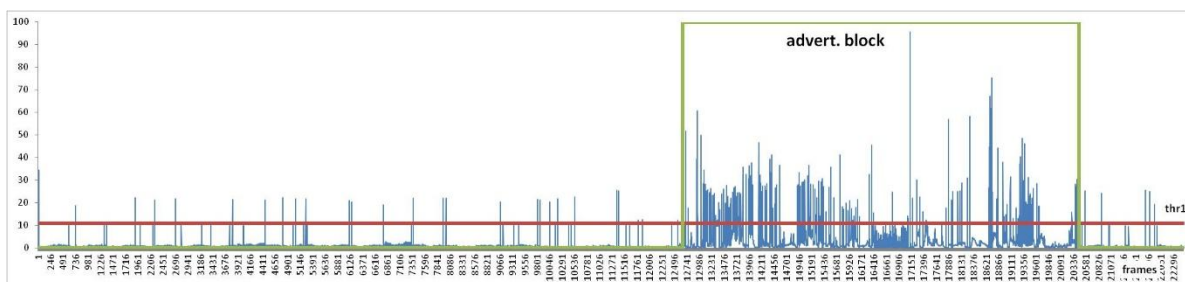


Фиг. 41. Детекция на сцени в рекламен видеоклип (~20 сек.).



Фиг. 42. Целочислена редица $\dot{r}^j = (\Delta l_1^j, \Delta l_2^j, \dots, \Delta l_{s_j}^j)$ 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.

2.3.4 Заключение и идеи за бъдеща работа

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. Същността на алгоритъма се състои в разделяне на видеоклиповете на последователност от сцени 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 извършваме покадрови сравнения, от типа „едно-към-много“. Този алгоритъм изисква много по-малък брой 2D покадрови сравнения (спрямо другите разгледаните методи), което е съществено предимство в случаите на значителен брой реклами за търсене в TV потока.

Експериментиран е прост подход за детекция на сцени, базиран на диференциалното представяне на видеоклиповете чрез разликата между два последователни кадъра в рамките на дефиниран, предварително зададен „праг за подобие между кадрите“ (PST). Целта е да се намери такъв PST праг, който 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

В тази глава е описана разработената многомоделна БД (3DEarDB), съдържаща различни моделни представяния на десните уши на ~100 индивиди. Разгледан 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. Очертани са насоките за бъдещата работа по дисертацията във връзка с въпроса за 3D реконструкцията на уши чрез методите на стереозрение или фотометричното стерео.

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

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 AComIn. 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 VIUscan на ИИКТ, който настроихме за 1 mm резолюция на сканиране (разстояние между съседните 3D точки), при точност от 0.05 mm за всяка 3D точка.

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.

3.1.1 Устройства за събиране на данни

VIUscan 3D лазерен скенер¹⁷

Този портативен скенер на Creaform беше закупен чрез проекта AComIn за „Умната лаборатория“ на ИИКТ-БАН, в края на 2013 г. Той може да възпроизвежда 3D фасетъчен (триангулачен) модел на сканирания обект, както и неговата текстура, и цветове. Изключването на цветността прави процедурата по сканиране по-бърза, до 10 мин. за ухо.

Kinect Xbox One сензор¹⁸

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 Camera¹⁹

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 [77], е представен на Фиг. 46б. 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.

¹⁷ http://www.creaform3d.com/sites/default/files/assets/brochures/files/handyscan/Handyscan3D_Brochure_EN_HQ_22052012.pdf

¹⁸ https://en.wikipedia.org/wiki/Kinect_for_Xbox_One#Specifications

¹⁹ <http://www.olympus-global.com/en/news/2011b/nr111110sh21e.jsp>



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

3.1.3 Обработка на суровите 3D модели

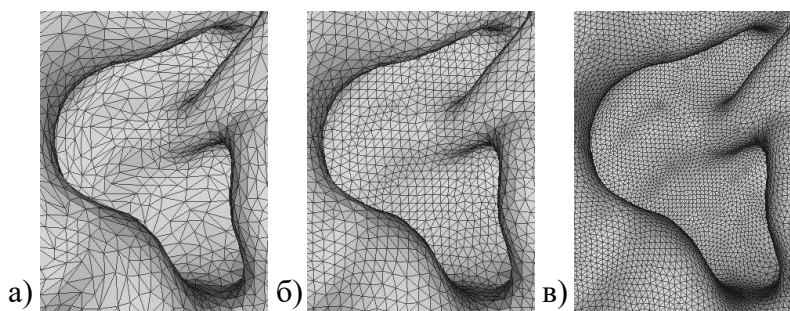
За създаване на завършен и подходящо изгладен 3D модел за всяко ухо провеждаме следната пост-обработката в 6 стъпки:

1. Груба сегментация от заобикалящия фон (чрез VElements)
2. Запълване на случайни празнини/дупки (чрез VElements)
3. Фина обработка на 3D фасетките на ушите (чрез MeshLab)
4. Допълнително изглаждане на модела (чрез MeshLab)

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).

5. Огрубяване („decimation“) и изфинване („subdivision“) (чрез MeshLab)

Тези манипулации на гъстотата на мрежа на 3D моделите (Фиг. 48) е необходима при генерирането на тестови данни за нашия EGI метод за класификация [31].



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

6. Геометрична нормализация (в MATLAB)

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

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.

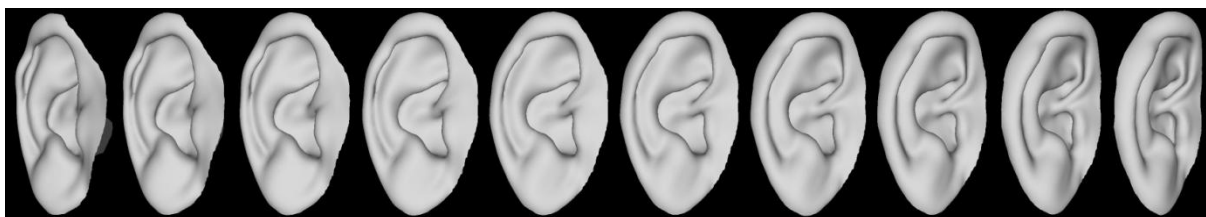
3.1.5 Full HD обхождащи видеоклипове на уши

За болшинството от ушите са направени видеоклипове с разделителна способност 1920×1080, равномерно заснети по азимут от -80° до +80° чрез 3 хоризонтални обхож-

дания (с различна деклинация – горе, център и долу) спрямо центъра на фронталния изглед. 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.

3.1.6 2D интензитетни проекции

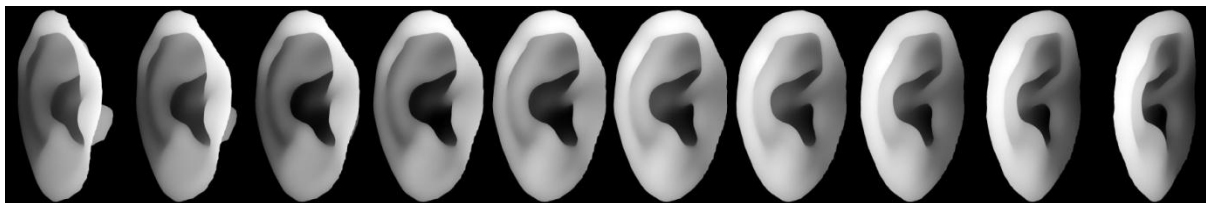
2D интензитетните проекции се генерират в MeshLab според избрана схема за 3D ротация – в 100 посоки, равномерно разпределени спрямо центъра на тежестта на ухото, т.е. по 10 деклинации и 10 азимута, равномерно избрани в интервала $(-45^\circ, +45^\circ)$. Този вид 3D представяне на ушите, наречен „Многоизгледно 3D моделиране“, беше разработено за наши експерименти по 3D разпознаване на уши чрез множество от техни 2D проекции [78]. Илюстрация на 10 двумерни (2D) проекции на дадено ухо, произведени от негов тримерен (3D) модел, са показани на Фиг. 52.



Фиг. 52. 2D изображения от даден ред на ротационната схема за 3D моделиране.

3.1.7 2D дълбочинни изображения

При генерирането на 2D дълбочинни изображения от 3D модел на ухо, 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)

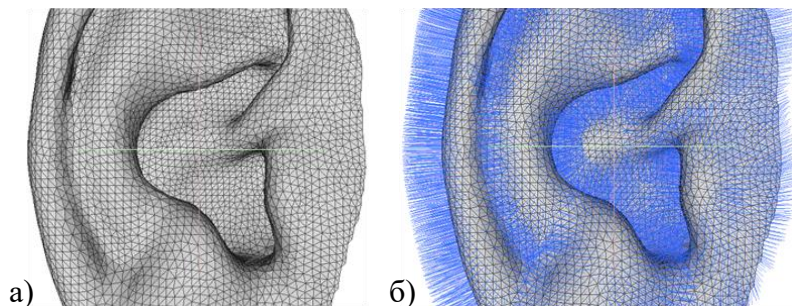
Разширеното Гаусово изображение (EGI) първоначално е предложено от В.К.Р. Horn [79], през 1984 г., за недвусмислено представяне на изпъкнала 3D повърхност във вид на хистограма на ориентациите върху сфера.

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

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

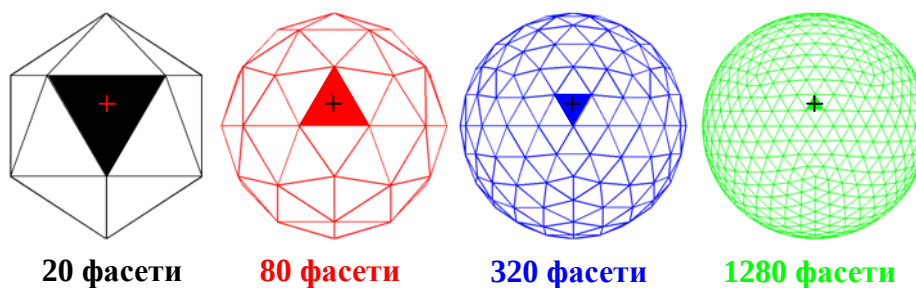
където N е броят на фасетките (триъгълниците) на 3D модела, а всяка фасетка T_i се представя със своите три върха:

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

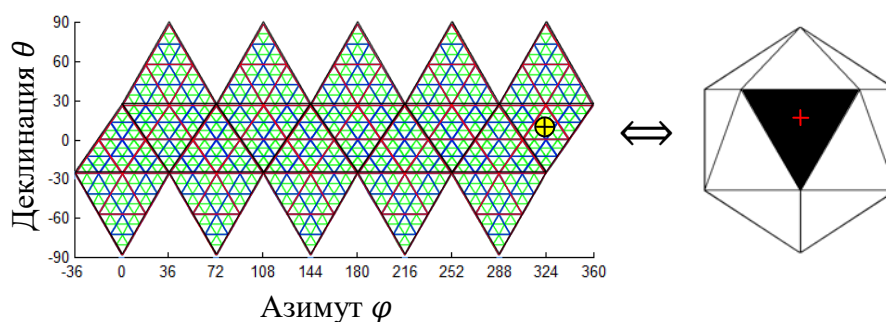


Фиг. 54. а) 3D мрежест (триангулачен) модел; б) нормални вектори към фасетките на модела.

Най-общо казано, EGI представянето на една 3D повърхнина представлява хистограма от всички нейни ориентации върху единична (Гаусова) сфера. На практика, Гаусовата сфера също се представя в дискретен вид, започвайки от икосаедър (правилен многостен с 20 триъгълни страни) на най-ниското ниво на дискретизация и достигайки до по-прецизна апроксимация на сферата на следващите нива, като всяка триъгълна страна от икосаедъра се разбива рекурсивно 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 $m = 4^n 20, n = 0, 1, \dots$.



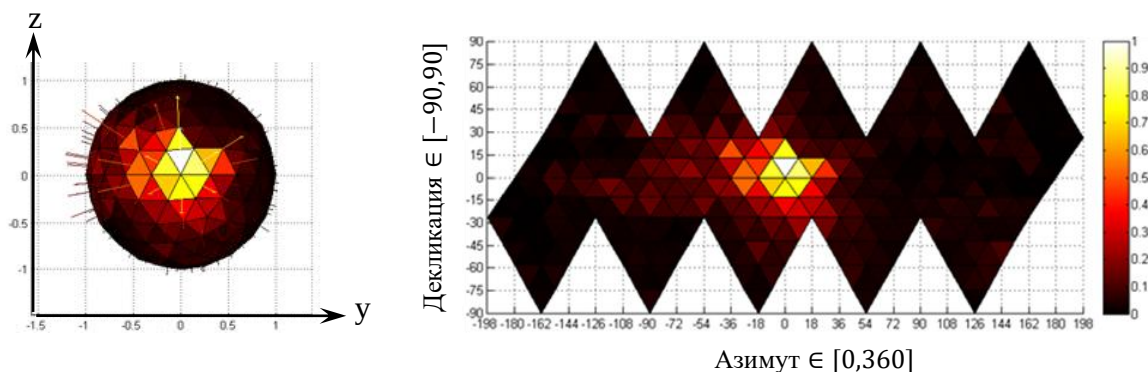
Фиг. 55. 3D рекурсивно разбиване, базирано на икосаедър (20-тостен) за апроксимиране на сфера чрез 20, 80, 320 и 1280 фасетки.



Фиг. 56. 2D представяне (по азимут и деклинация) на дадена апроксимация на сфера чрез съответното разрязване и разгъване на базовия икосаедър.

Така, според ориентацията на нормалния вектор към всяка фасетка на входната 3D повърхнина (Фиг. 54б), площта на фасетката се акумулира към съответния „бин“ от дискретната Гаусова сфера (Фиг. 57). За бина като индекс е достатъчно да се намери максимумът от всички скалярни произведения на нормалите на входния обект и бин-нормалите на сферата. 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.

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.



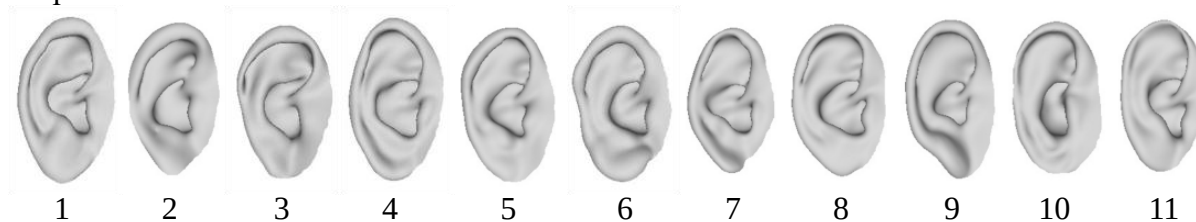
Фиг. 57. 3D и 2D EGI хистограми за някаква входна 3D повърхнина – във всеки EGI бин се акумулира площта на съответните триъгълните фасетки от входа.

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

Тук е описан процесът на построяване на икосаедър и неговото 2D представяне. Изчисляват се координатите (x, y, z) за всеки връх на икосаедъра, след което се преобразуват в 2D координати според ъглите на азимута и деклинацията до съответните 3D върхове.

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

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 всеки 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 разпознаемостта, разбира се до определени граници на зашумяването, което прилагаме. Нашето обяснение за този позитивен експеримент е, че ухото е по-скоро плосък 3D обект.



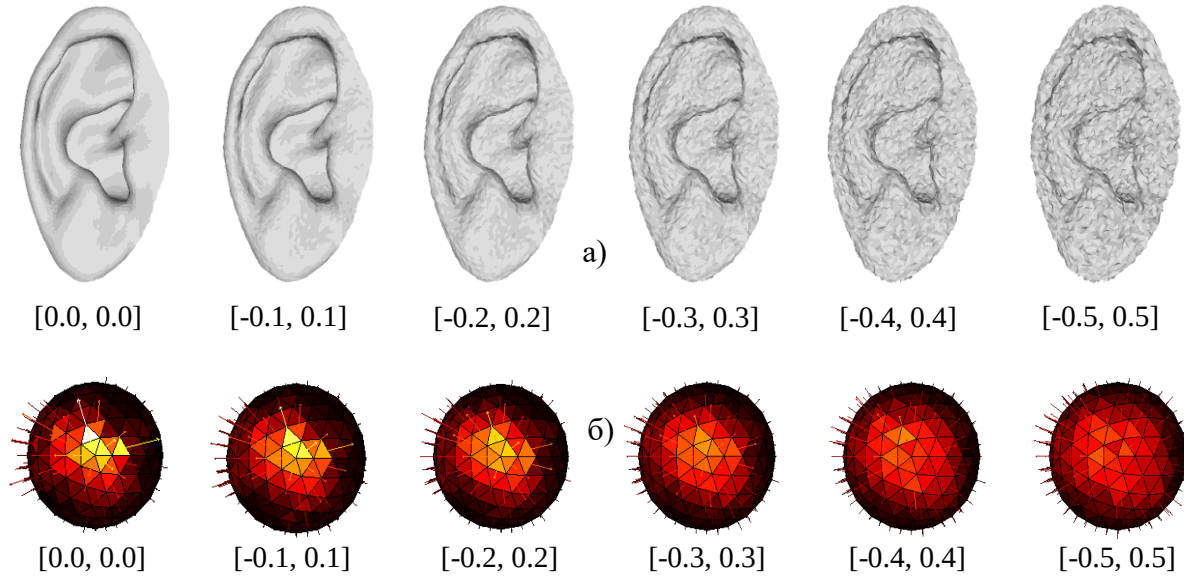
Фиг. 59. Първоначалната БД от 11 3D модела на уши, създадена през май 2014 г.

Текущата цел е да потвърдим този позитивен експеримент върху цялата налична БД, като само за простота (целочисленост на резултатите в %, Табл. 16) разглеждаме точно 100 от 3D моделите в 3DEarDB.

Така, ние използваме разширеното Гаусово изображение, като 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.

Нашият доказателствен експеримент се състои в следното:

Първо: За всяко ухо в БД се построява съответното му EGI представяне (виж §3.2). 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 въведен 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 дадено 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.

Второ: 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:

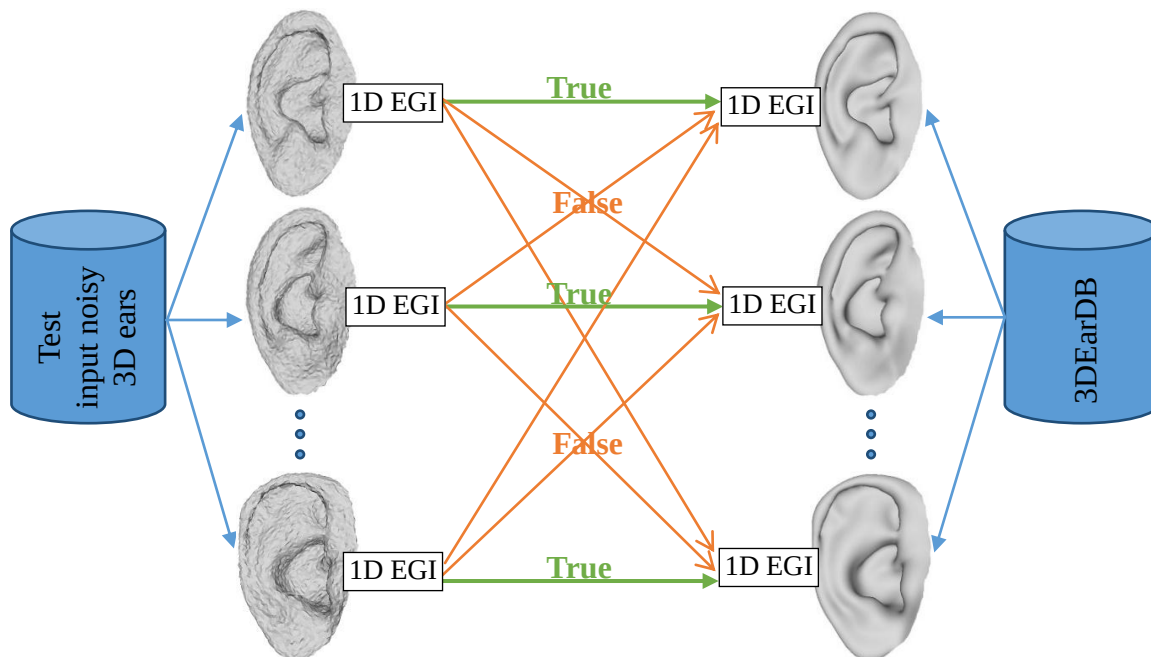
$$E_2 = \sqrt[2]{\sum_{i=1}^m (M_i - S_i)^2}; \quad (25a)$$

- Bray-Curtis разстоянието [80]:

$$E_{BC} = \frac{\sum_{i=1}^m |M_i - S_i|}{\sum_{i=1}^m (M_i + S_i)}, \quad 0 \leq E_{BC} \leq 1, \quad (25b)$$

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

Трето: 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).

Очевидно, при някакво достатъчно високо ниво на входния шум, БД ще започне да греша. Целта на предложението тук (от типа „всеки-към-всеки“) за тестване на нашата 3DEarDB е да се оцени максимално допустимата стойност на входния шум Δ , 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.

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 (25a) и (25б), за 4 последователни нива на EGI дискретизация (20, 80, 320 и 1280 фасетки) и за 3 различни резолюции на 3D моделите: 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). Същевременно, вариациите за Δ организираме по принципа „увеличаваме Δ до първо срещане на грешно разпознаване“, плюс няколко, примерно 2, опита наоколо.

Така, 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 (Глава 3, §3.1), ние разглеждаме 3DEarDB като „памет“ на вече обучен (100 класа, всеки 3D модел е център на клас) класификатор от типа „най-близкия съсед“, при дадена дефиниция за разстояние, например (25a), (25б) или друго [42]. А използваната от нас (единствена и достатъчна в нашия случай) характеристика TRR (True Recognition Rate) за оценка на разпознаемостта може да се изрази с популярната на-

последък характеристика “accuracy”²⁰:

$$TRR = \frac{\text{брой правилно разпознати изходи (позитивни и негативни)}}{\text{брой експерименти}} = \text{accuracy.} \quad (26)$$

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

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] ²¹		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
	фасети E_{BC}	100	100	92	100	100	89	100	99	80
80	фасети E_2	100	93	17	100	92	16	100	88	40
	фасети E_{BC}	100	100	67	100	100	70	100	100	80
320	фасети E_2	100	84	25	100	83	27	100	80	44
	фасети E_{BC}	100	100	80	100	100	66	100	100	80
1280	фасети E_2	100	63	13	100	52	16	100	56	30
	фасети E_{BC}	100	100	62	100	95	48	100	96	78

3.3.1 Анализ на експерименталните резултати

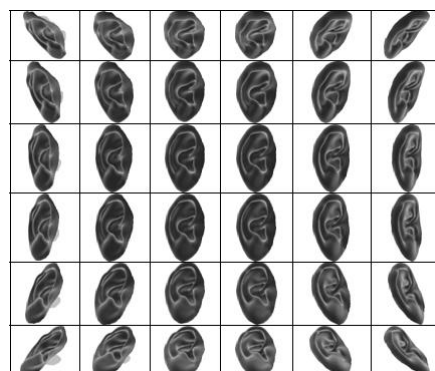
- ❖ Проведените експерименти върху текущата 3DEarDB (от 100 3D модела на уши) потвърждават възможността за използване на EGI представянето за недвусмислена идентификация на уши, независимо че тяхната повърхнина е съставена от комбинация между изпъкнали и вдлъбнати области. Това се потвърждава от резултатите 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.
- ❖ 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.

²⁰ https://en.wikipedia.org/wiki/Confusion_matrix

²¹ 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 елипсoid: ширина (32.3 mm по O_x), височина (50.3 mm по O_y), дълбочина (13.2 mm по O_z).

3.4 Разпознаване на уши чрез подхода „Многоизгледно 3D моделиране“²²

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 [81]. Разработката, авторско дело на доц. д-р Димо Димов, е тествана в неговата експериментална софтуерна среда – 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.



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

Връзката на системата EFCIOR с изследванията по дисертацията (виж Фиг. 2) 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 личността по формата на ушите. Проведените експерименти с малка БД (с проекциите на 3D модели на уши на 10 различни личности) [78], 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-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. Сампери, Сицилия, Италия. Там бяха демонстрирани възможностите на вариационните методи за 3D реконструкция: (i) от стерео изображения [82]; (ii) от много изгледи [83]; (iii) на динамични сцени [84]; (iv) на плътни сцени в реално време (24 fps за резолюция 480×360 пиксела) [85]; (v) на сцени в голям мащаб [86]; (vi) на сцени в реално време, заснети от движещ се автомобил [87]. Бе спомената и системата „CoryMe3D“ за 3D моделиране на обекти в реално време, работеща на PC с помощта на Kinect дълбочинен сензор [88], където сканирането на обекта отнема около 10 сек. (за 300 кадъра), докато пълната обработка и извеждането на резултата типично отнемат около 1 мин. на PC от среден клас. Беше демонстрирано и моделиране на градове в реално време (30 fps) от видео клип, чрез GPU имплементация на PC [89]. Представена е и разработка за превръщането на мобилните телефони в 3D скенери [90], с приложение към заснемане на културно наследство, идеята на 3D селфита и др.

Поради недостижимия бум в областта на плътната 3D реконструкция на обекти и сцени, работеща в реално време на PC и дори на мобилни телефони, разработваният метод, базиран на стереозрение, за определяне на дълбочина до определени точки от обекти [91], 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). В тази връзка, особен интерес представлява мобилната система [92] за 3D реконструкция на лица от видеоклип с конкретно приложение в биометриката (Фиг. 68).

²² EFCIOR методът не е разработен в рамките на текущата дисертация, но се разглежда във връзка с 3DEarDB.



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.

Друг евентуален подход за разпознаване на уши в контекста на EGI подхода са методите, базирани на фотометрично стерео [93, 94, 95], 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*²³, 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 повече от 100 личности (главно колеги от БАН), 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 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 нови такива, базирани на 3D информацията от триангулационните модели.

Резултатите от проведените експерименти върху собствената БД от 100 3D модела на уши (3DEarDB) потвърждават възможността за използване на конвенционалния EGI дескриптор за недвусмислена идентификация на уши (независимо че тяхната повърхнина не е изцяло изпъкнала), което можем да обясним с факта, че ушите са по-скоро плоски 3D обекти. 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.

²³ <http://www.trimensional.com/>

Заклучение

Резюме на постигнатото и идеи за бъдеща работа

Предложен е нов ефективен и робастен софтуерен метод за 2D видеостабилизация 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, вградени в съответното портативно устройство, на които се разчита за една ефективна хардуерна видеостабилизация.

Разработен е комбиниран метод $3 \times 3 \text{OF}9 \times 9$, предназначен за видеостабилизация 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 „пъзене 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 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 извършваме покадрови сравнения, от типа „едно-към-много“. Този алгоритъм изисква много по-малък брой 2D покадрови сравнения (спрямо другите разгледаните методи), което е съществено предимство в случаите на значителен брой реклами за търсене в TV потока. Разработена е многомоделната 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 резо-

люция 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.

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 значимо 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 " $3 \times 3 \text{OF9} \times 9$ ", 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).

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] Bhanu, Bir., Chen, Hui.: Human Ear Recognition by Computer. Springer, 2008.
- [4] 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.
- [5] 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.
- [6] Prakash, S., Gupta, P.: Ear Biometrics in 2D and 3D. Springer, 2015.
- [7] 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).
- [8] 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).
- [9] 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).
- [10] 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).
- [11] 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).
- [12] 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).
- [13] 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).
- [14] 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>
- [15] Morimoto, C., Chellappa, R.: Evaluation of Image Stabilization Algorithms. In DARPA Image Understanding Workshop DARPA97, pp. 295-302 (1997).
- [16] 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).
- [17] 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).
- [18] Hartley, R. I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, 2000.
- [19] 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).
- [20] Goldstein, A., Fattal, R.: Video Stabilization using Epipolar Geometry. ACM Transactions on Graphics, Vol. 31, Issue 5, pp. 126:1-126:10 (2012).
- [21] 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).
- [22] 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).
- [23] 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).

- [24] 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).
- [25] 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.
- [26] 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.
- [27] Fitzgerald, J. A., Translational Motion Stabilization for Embedded Video Applications, White Paper SPRAAX8, Texas Instruments, 2008.
- [28] TMS320DM365 Video Stabilization Algorithm Datasheet, Texas Instruments, January 2010.
- [29] 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).
- [30] Ross, A., Jain, A. K.: Biometrics, Overview. In: Li, S. Z., Jain, A. K. (Eds.) *Encyclopedia of Biometrics*, pp. 168-172. Springer, Heidelberg (2009).
- [31] 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).
- [32] 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).
- [33] 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).
- [34] Zhang, D., Lu, G.: *3D Biometrics - Systems and Applications*. Springer, 2013.
- [35] Barbu, T.: Comparing Various Voice Recognition Techniques. *Proc. of the 5th Conf. on Speech Technology and Human-Computer Dialogue*, pp. 1-6 (2009).
- [36] 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).
- [37] Maltoni, D., Maio, D., Jain, A. K., Prabhakar, S.: *Handbook of Fingerprint Recognition*. 2nd Edition. Springer, 2009.
- [38] Yoon, S.: *Fingerprint Recognition: Models and Applications*, Ph.D. Thesis, 2014.
- [39] Li, S. Z., Jain, A. K.: *Handbook of Face Recognition*. 2nd Edition. Springer, 2011.
- [40] Rowden, L. B., Jain, A. K.: A Longitudinal Study of Automatic Face Recognition. *ICB*, Phuket, Thailand, May 19-22, 2015.
- [41] 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).
- [42] Rathgeb, Ch., Uhl, A., Wild, P.: *Iris Biometrics - from Segmentation to Template Security*. Springer, 2013.
- [43] 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).
- [44] 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).
- [45] 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).
- [46] 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).
- [47] 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).
- [48] 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).

- [49] 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).
- [50] 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).
- [51] 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).
- [52] Ross, A.: An Introduction to Multibiometrics. Proc. of the 15th European Signal Processing Conference (EUSIPCO), Poznan, Poland (2007).
- [53] 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).
- [54] 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).
- [55] Bender, S., Postley, H.: Key sequence rhythm recognition system and method. US Patent No. 7206938, 2007.
- [56] 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).
- [57] Park, U., Ross, A., Jain, A. K.: Periocular Biometrics in the Visible Spectrum: A Feasibility Study. Proc BTAS, Washington DC, 2009.
- [58] Kumar, A., Ravikanth, C.: Personal Authentication using Finger Knuckle Surface. IEEE Trans. on Inf. Forensics & Security, Vol. 4, No. 1, pp. 98-110 (2009).
- [59] 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).
- [60] 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).
- [61] 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.
- [62] 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).
- [63] 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).
- [64] Jain, A. K., Ross, A., Nandakumar, K.: Introduction to Biometrics. Springer, 2011.
- [65] Day, D.: Biometric Applications, Overview. In: Li, S. Z., Jain, A. K. (Eds.) Encyclopedia of Biometrics, 2nd Edition, pp. 169-174. Springer, Heidelberg (2015).
- [66] Bertillon, A.: La Photographie Judiciaire: Avec Un Appendice Sur La Classification Et L'Identification Anthropométriques. Gauthier-Villars, Paris, 1890.
- [67] Chen, H., Bhanu, B.: Contour Matching for 3D Ear Recognition. Proc. of the 7th IEEE Workshop on Applications of Computer Vision (WACV/MOTION), 2005.
- [68] Yan, P., Bowyer, K. W.: ICP-Based Approaches for 3D Ear Recognition. Proc. of SPIE, Vol. 5779, pp. 282-291 (2005).
- [69] 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).
- [70] 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).
- [71] 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.
- [72] 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).
- [73] 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).
- [74] 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).
- [75] Bouguet, J.: Pyramidal Implementation of the Lucas-Kanade Feature Tracker: Description of The Algorithm. Technical report, OpenCV Document, Intel Microprocessor Res. Labs, 2000.

- [76] 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).
- [77] 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
- [78] 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).
- [79] Horn, B.K.P.: Extended Gaussian images. Proc. of the IEEE, Vol. 72, pp. 1671-1686 (1984).
- [80] Bray, J. R., Curtis, J. T.: An Ordination of Upland Forest Communities of Southern Wisconsin. Ecological Monographs, Vol. 27, pp. 325-349 (1957).
- [81] 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).
- [82] 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.
- [83] Kolev, K., Klodt, M., Brox, T., Cremers, D.: Continuous Global Optimization in Multiview 3D Reconstruction. In Int. Journal of Computer Vision, Vol. 84 (2009).
- [84] Oswald, M. R., Stühmer, J., Cremers, D.: Generalized Connectivity Constraints for Spatio-temporal 3D Reconstruction. In European Conference on Computer Vision (ECCV), 2014.
- [85] 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
- [86] Engel, J., Schöps, T., Cremers, D.: LSD-SLAM: Large-Scale Direct Monocular SLAM. In European Conference on Computer Vision (ECCV), 2014.
- [87] 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.
- [88] Sturm, J., Bylow, E., Kahl, F., Cremers, D.: CopyMe3D: Scanning and Printing Persons in 3D. In German Conference on Pattern Recognition (GCPR), 2013.
- [89] 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
- [90] 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).
- [91] 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).
- [92] 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).
- [93] Woodham, R. J.: Photometric Method for Determining Surface Orientation from Multiple Images. Opt. Eng., Vol. 19, Issue 1, 191139 (1980).
- [94] Shi, B.: Photometric Stereo for General Reflectance and Lighting. A doctoral dissertation, the University of Tokyo, Supervisor: Katsushi Ikeuchi, June, 2013.
- [95] 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.