



Секция

„Обработка на сигнали и разпознаване на образи“

Семинар за отчисляване от докторантура, ИИКТ-БАН, 12.12.2014

Видеостабилизация и 3D разпознаване в реално време



Докторант: Атанас Филипov Николов

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



Съдържание

1. Видеостабилизация в реално време

- 1.1. Определяне векторите на движение
- 1.2. Векторен модел за намиране параметрите на движението
- 1.3. Експериментални резултати

2. Представяне на 3D обекти чрез EGI и разпознаването им

- 2.1. Extended Gaussian Image (EGI) и неговият 2D аналог
- 2.2. Разпознаване на уши
- 2.3. Експериментални резултати

3. Тримерна реконструкция на облак от точки

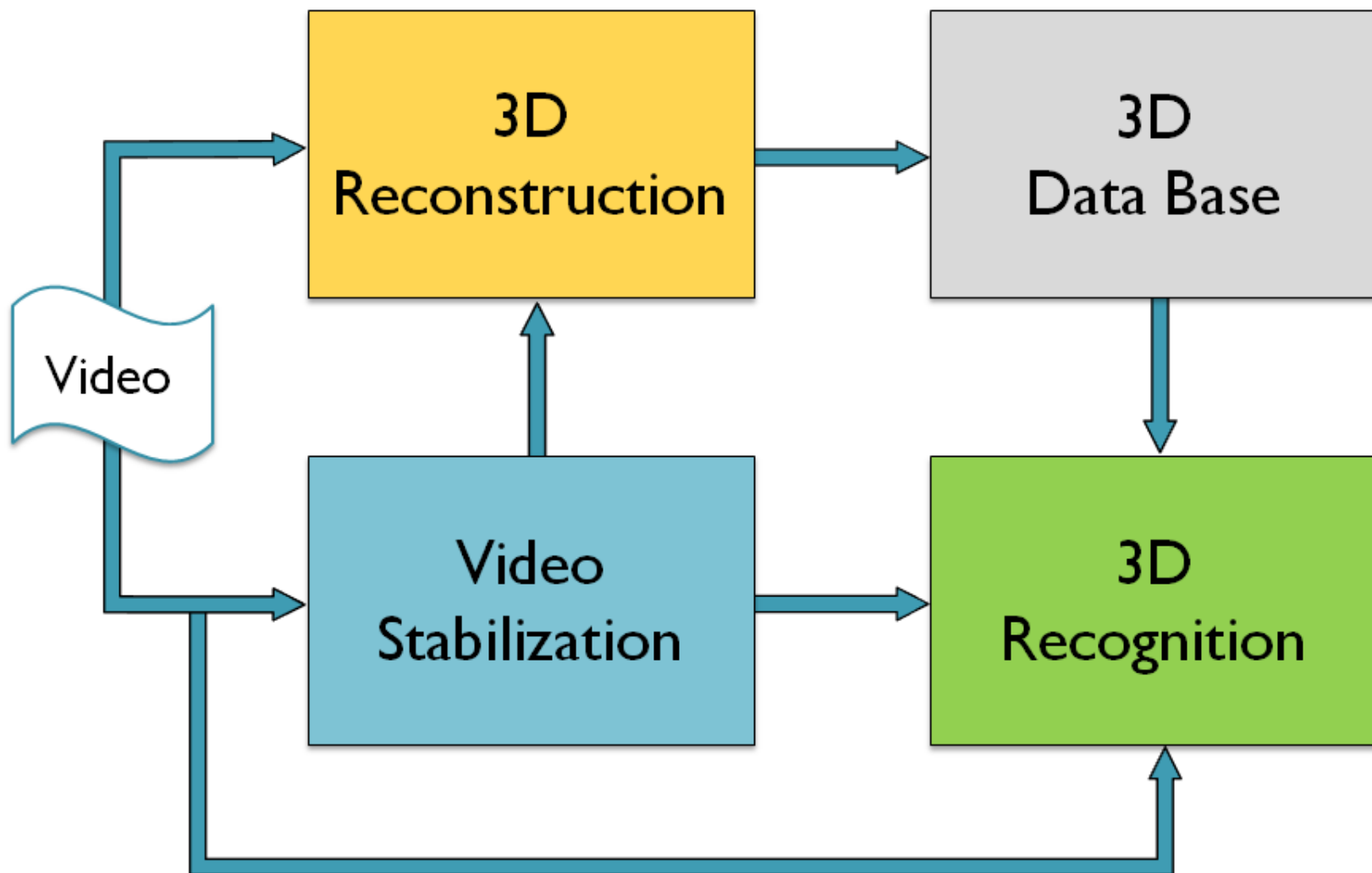
- 3.1. Канонична система за стереозрение
- 3.2. Проведени експерименти

4. Структура на дисертацията – схематично представяне

5. Текущи и финални изследвания

6. Насоки за бъдеща работа (след защитата)

Взаимовръзка на изследванията по дисертацията



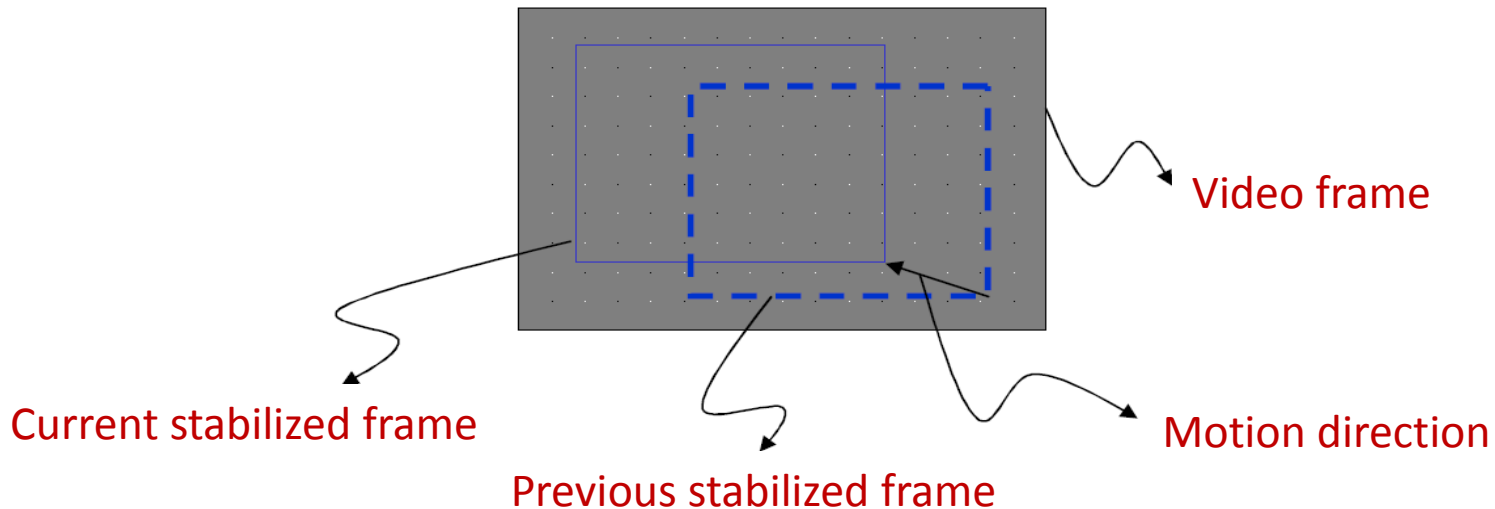
1. Видеостабилизация в реално време

- Видео стабилизацията цели да създаде **изгладен вариант на нестабилно заснето видео** (с преносимо устройство като мобилен телефон, видеокамера и т.н.), при което се проявяват всички недостатъци на нестационарното филмиране.
- Най-общо, методите за 2D видео стабилизация се основават на оценяване на параметрите на **оптимална линейна трансформация** на текущия кадър спрямо референтен кадър (напр. предходния) от видео последователността.
- Фактическата стабилизация се реализира чрез **изглаждане във времето на параметрите на движението**, след което се прилага трансформация (обратна на изчислената), която преобразува (транслира, ротира, мащабира, и т.н.) кадрите от оригиналното движение в кадри на изгладения модел на движението.
- Оценката на реалното движение на камерата, т.е. последователността от трансформации на текущия кадър към текущата референтна позиция, се основава на изчисляване на така наречените **„вектори на движението“**, които се дефинират интегрално – за съответстващи области в кадрите или локално – за съответстващи характерни точки в тях.

1. Видеостабилизация в реално време

- Един бърз (над 100 fps за видео 1280x960) и опростен подход за 2D видеостабилизация (само по трансляция) е описан във White Paper of Texas Instruments: Motion Stabilization for Embedded Video Application (Fitzgerald'08).
- Методът използва BSC чипа на TI, който позволява ефективно изчисляване на вертикални и хоризонтални акумулиращи проекции (“хистограми”) на дадено изображение.
- BSC може равномерно да разбие входното изображение на няколко части - примерно на 9 подизображения (3 по вертикал и 3 по хоризонтал), за които да изчисли съответните 18 “хистограми” – по 2 за всяко подизображение.

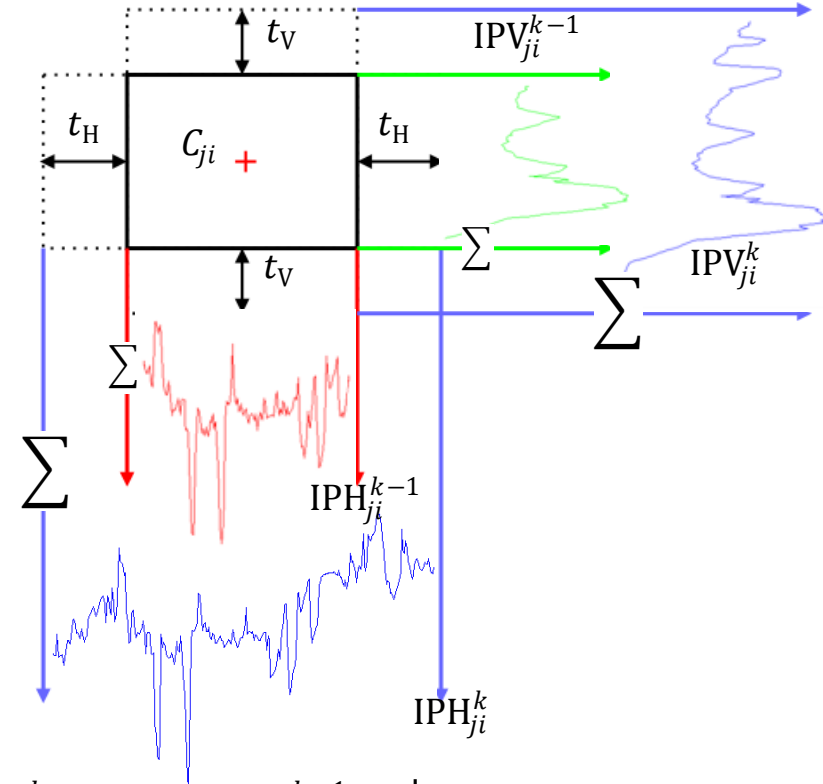
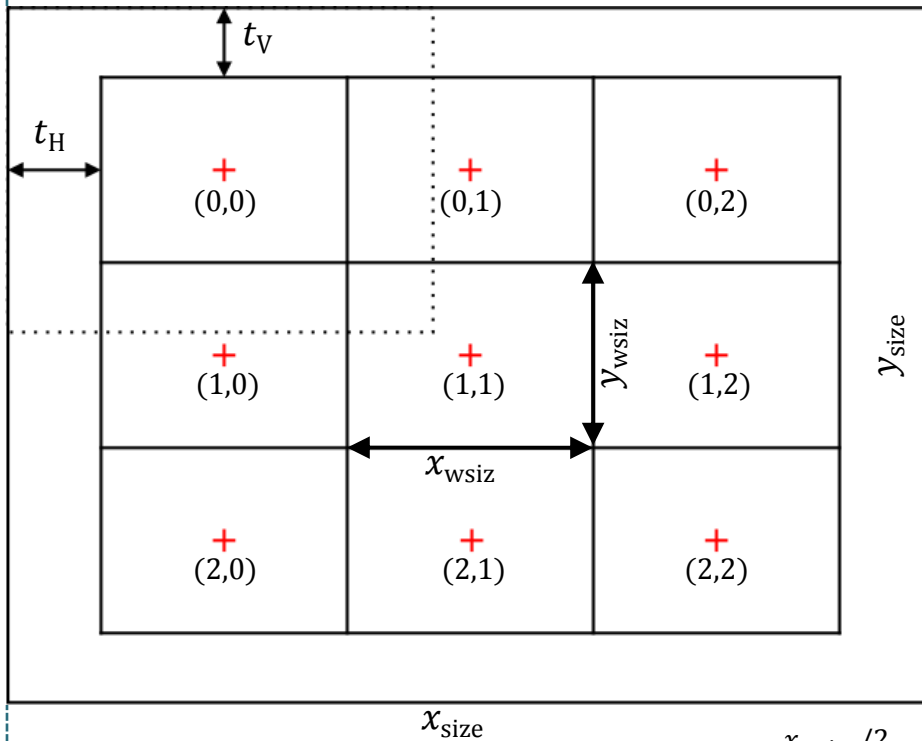
TMS320DM365 Video Stabilization Algorithm Datasheet (January 2010)



1. Видеостабилизация в реално време [1]

1.1. Определяне векторите на движение

Всеки кадър се разделя на 3×3 области, което определя 9 на брой търсени вектори на движението \vec{t}_{ji} ($j, i = 0, 1, 2$) между съответните 9 центъра на подобластите в текущия и предходния кадър. Търси се минимум на Sum of Absolute Diff. (SAD).

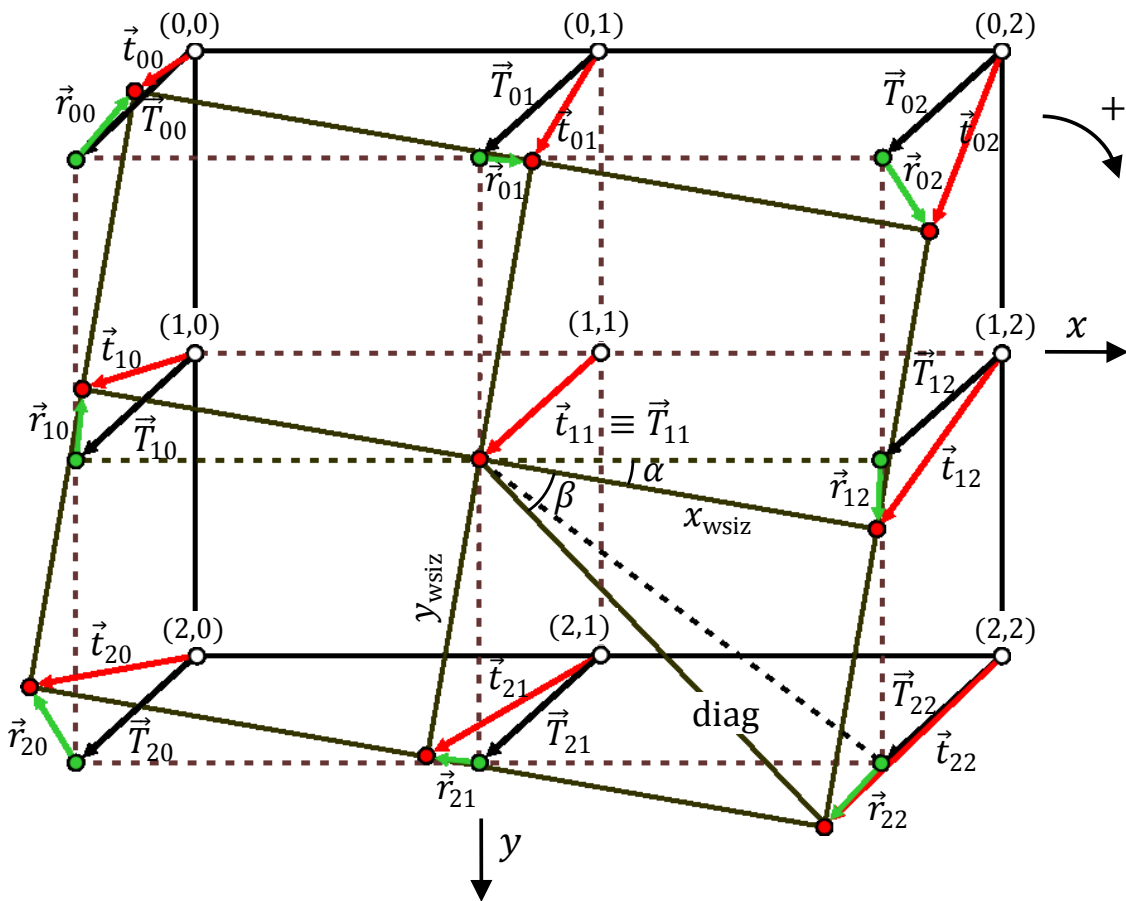


$$t_{xji}^{(k)} = \underset{(\tau)}{\operatorname{argmin}} \{ \operatorname{SADH}_{ji}^k(\tau) \}; \operatorname{SADH}_{ji}^k(\tau) = \sum_{x=-x_{wsiz}/2}^{x_{wsiz}/2} |\operatorname{IPH}_{ji}^k(x + \tau) - \operatorname{IPH}_{ji}^{k-1}(x)|, \quad -t_H < \tau < t_H$$

$$t_{yji}^{(k)} = \underset{(\tau)}{\operatorname{argmin}} \{ \operatorname{SADV}_{ji}^k(\tau) \}; \operatorname{SADV}_{ji}^k(\tau) = \sum_{y=-y_{wsiz}/2}^{y_{wsiz}/2} |\operatorname{IPV}_{ji}^k(y + \tau) - \operatorname{IPV}_{ji}^{k-1}(y)|, \quad -t_V < \tau < t_V$$

1. Видеостабилизация в реално време [1]

1.2. Намиране параметрите на движението – векторен модел



Всеки оценен по SAD вектор на движение \vec{t}_{ji} между два кадъра се разлага на сума от два вектора: вектор на трансляция \vec{T}_{ji} и вектор на ротация \vec{r}_{ji} , т.е. $\vec{t}_{ji} = \vec{T}_{ji} + \vec{r}_{ji}$

По този начин съставяме система от 18 покомпонентни уравнения, в която имаме 4 неизвестни параметъра (T_x, T_y, r_x, r_y), които намиране по Least Squares метода.

Прякото решение за търсените параметри е следното:

$$T_x = \frac{1}{9} \sum_{j,i} t_{xji}, \quad T_y = \frac{1}{9} \sum_{j,i} t_{yji}$$

$$r_x = \frac{(t_{x12} - t_{x10} + t_{x22} - t_{x00} + t_{x02} - t_{x20}) + k(t_{y21} - t_{y01} + t_{y22} - t_{y00} - t_{y02} + t_{y20})}{6 + 6k^2}, \quad k = y_{wsiz} / x_{wsiz}$$

$$r_y = \frac{(t_{y12} - t_{y10} + t_{y22} - t_{y00} + t_{y02} - t_{y20}) - k(t_{x21} - t_{x01} + t_{x22} - t_{x00} - t_{x02} + t_{x20})}{6 + 6k^2}, \quad \alpha = 2 \arctg(r_x / r_y)$$

1. Видеостабилизация в реально време [1]

1.2. Намиране параметрите на движението (транслация + ротация)

Centre (j,i)	LSM on \overrightarrow{Ox}	$\frac{\partial \varepsilon_{xji}^2}{2\partial r_x}$	$\frac{\partial \varepsilon_{xji}^2}{2\partial r_y}$	LSM on \overrightarrow{Oy}	$\frac{\partial \varepsilon_{yji}^2}{2\partial r_x}$	$\frac{\partial \varepsilon_{yji}^2}{2\partial r_y}$
(1,1)	$T_x - t_{x11} = \varepsilon_{x11}$	0	0	$T_y - t_{y11} = \varepsilon_{y11}$	0	0
(1,2)	$T_x + r_x - t_{x12} = \varepsilon_{x12}$	ε_{x12}	0	$T_y + r_y - t_{y12} = \varepsilon_{y12}$	0	ε_{y12}
(1,0)	$T_x - r_x - t_{x10} = \varepsilon_{x10}$	$-\varepsilon_{x10}$	0	$T_y - r_y - t_{y10} = \varepsilon_{y10}$	0	$-\varepsilon_{y10}$
(2,1)	$T_x - kr_y - t_{x21} = \varepsilon_{x21}$	0	$-k\varepsilon_{x21}$	$T_y + kr_x - t_{y21} = \varepsilon_{y21}$	$k\varepsilon_{y21}$	0
(0,1)	$T_x + kr_y - t_{x01} = \varepsilon_{x01}$	0	$k\varepsilon_{x01}$	$T_y - kr_x - t_{y01} = \varepsilon_{y01}$	$-k\varepsilon_{y01}$	0
(2,2)	$T_x + r_x - kr_y - t_{x22} = \varepsilon_{x22}$	ε_{x22}	$-k\varepsilon_{x22}$	$T_y + kr_x + r_y - t_{y22} = \varepsilon_{y22}$	$k\varepsilon_{y22}$	ε_{y22}
(0,0)	$T_x - r_x + kr_y - t_{x00} = \varepsilon_{x00}$	$-\varepsilon_{x00}$	$k\varepsilon_{x00}$	$T_y - kr_x - r_y - t_{y00} = \varepsilon_{y00}$	$-k\varepsilon_{y00}$	$-\varepsilon_{y00}$
(0,2)	$T_x + r_x + kr_y - t_{x02} = \varepsilon_{x02}$	ε_{x02}	$k\varepsilon_{x02}$	$T_y - kr_x + r_y - t_{y02} = \varepsilon_{y02}$	$-k\varepsilon_{y02}$	ε_{y02}
(2,0)	$T_x - r_x - kr_y - t_{x20} = \varepsilon_{x20}$	$-\varepsilon_{x20}$	$-k\varepsilon_{x20}$	$T_y + kr_x - r_y - t_{y20} = \varepsilon_{y20}$	$k\varepsilon_{y20}$	$-\varepsilon_{y20}$

$$T_x = \frac{1}{9} \sum_{j,i} t_{xji}, \quad r_x = \frac{(t_{x12} - t_{x10} + t_{x22} - t_{x00} + t_{x02} - t_{x20}) + k(t_{y21} - t_{y01} + t_{y22} - t_{y00} - t_{y02} + t_{y20})}{6 + 6k^2}$$

$$T_y = \frac{1}{9} \sum_{j,i} t_{yji}, \quad r_y = \frac{(t_{y12} - t_{y10} + t_{y22} - t_{y00} + t_{y02} - t_{y20}) - k(t_{x21} - t_{x01} + t_{x22} - t_{x00} - t_{x02} + t_{x20})}{6 + 6k^2}$$

$$\alpha = 2 \arctg(r_x / r_y), \quad k = y_{\text{wsiz}} / x_{\text{wsiz}}$$

1. Видеоостабилизация в реално време [1]

1.3. Експериментални резултати

$$ITF = \frac{1}{N-1} \sum_{k=1}^{N-1} 10 \left| \log_{10} \left(\frac{1}{x_{size} y_{size}} \left(1 + \sum_{y=0}^{y_{size}-1} \sum_{x=0}^{x_{size}-1} \text{diff}(I_k(x, y), I_{k-1}(x, y)) \right) \right) \right|$$

където $I_k(x, y)$ е стойността за пиксел (x, y) в кадър (k) , а $\text{diff}(a, b) = \begin{cases} 1, & \text{if } |a - b| > t \\ 0, & \text{otherwise} \end{cases}$
 $0 \leq a, b \leq 255$, където t е праг за компенсиране на отместването м/у стабилизираните кадри.



Size: 640x480

ITF оценки за 5 видеоклипа: оригиналните и техните стабилизирани по 6-те подхода

метод \ видео	original	3x3/noAVG/I	3x3/AVG/I	3x3/noAVG/II	3x3/AVG/II	9x9/AVG/II	FAST
1	11.85	16.57	17.45	17.68	17.99	17.99	18.09
2	11.09	14.36	16.06	16.01	16.43	16.29	16.37
3	10.01	14.02	14.79	14.84	15.30	15.51	15.70
4	9.61	13.05	13.66	13.85	14.03	14.31	14.32
5	9.40	12.58	13.07	13.17	13.29	13.32	13.40
AVM [ms]	-	2	2	2	2	15	142
Общо [ms]	33	2+26	2+26	2+26	2+26	15+26	142+26
Кадри [fps]	30	35.7	35.7	35.7	35.7	24.4	5.95

Допълнителни времена в Matlab: четене $\approx 10ms$; запис $\approx 10ms$; обратна трансформация $\approx 6ms$

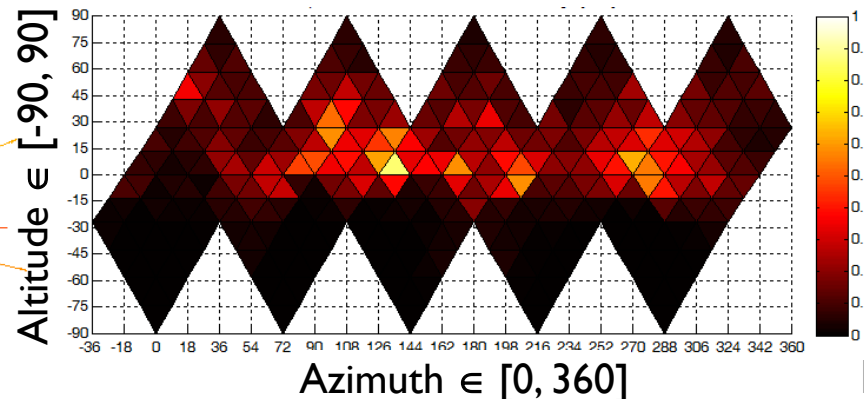
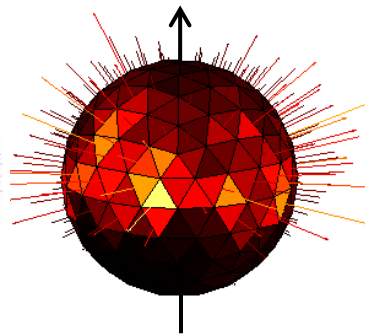
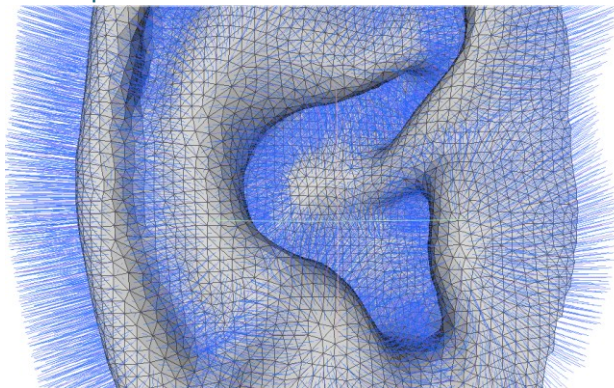
2. Представяне на 3D обекти чрез EGI [2]

2.1. Extended Gaussian Image и неговият 2D аналог

- В резултат от тримерното сканиране на един обект се получава апроксимиращата триангулационна мрежа $T = \{T_1, \dots, T_N\}$, $T_i \subset R^3$, където N е броят на триъгълниците от T , зададени чрез техните върхове: $T_i = \{V_{A_i}, V_{B_i}, V_{C_i}\}$
- Нормалният вектор n_i и площта S_i на триъгълника T_i , както и общата площ S на апроксимиращата мрежа T , са както следва:

$$\hat{n}_i = (V_{B_i} - V_{A_i}) \times (V_{C_i} - V_{A_i}), \quad S_i = \frac{1}{2} |(V_{B_i} - V_{A_i}) \times (V_{C_i} - V_{A_i})|, \quad S = \sum_{i=1}^N S_i$$

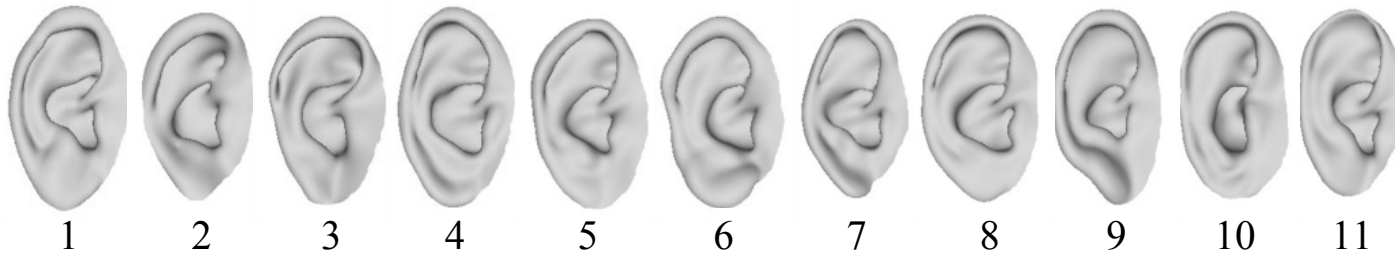
- EGI на един 3D обект представлява хистограма върху единична сфера, апроксимирана чрез правилен многостен, съдържаща разпределението на единиците площи на обекта по отношение на тяхната ориентация.
- Теглото на всеки bin с дадена ориентация от хистограмата е равно на сумата от единиците площи от обекта със същата ориентация.



2. Представяне на 3D обекти чрез EGI [2]

2.2. Разпознаване на 3D обекти (уши)

- По случай международния семинар по Биометрика, проведен на 23-24.06.2014г. в дома на учения на БАН, беше създадена малка **БД от 3D модели на уши** на 11 човека. За всеки индивид, БД съдържа 6 3D уши (където 5 са умишлено зашумени за целите на извършените изследвания), което прави общо 66 3D модела на уши.
- Всички 3D уши са сканирани чрез VIUscan 3D скенера на ИИКТ, който позволява геометрична разрешаваща способност от **0.1mm** и точност от **50μm**.



Изчисляване на подобие между EGI хистограмите

1) Разстояние по Minkowski:

$$E_M = \sqrt[p]{\sum_{i=1}^m |M_i - I_i|^p}$$

Manhattan ($p=1$),

Euclidean ($p=2$)

2) Разстояние по Bray-Curtis:

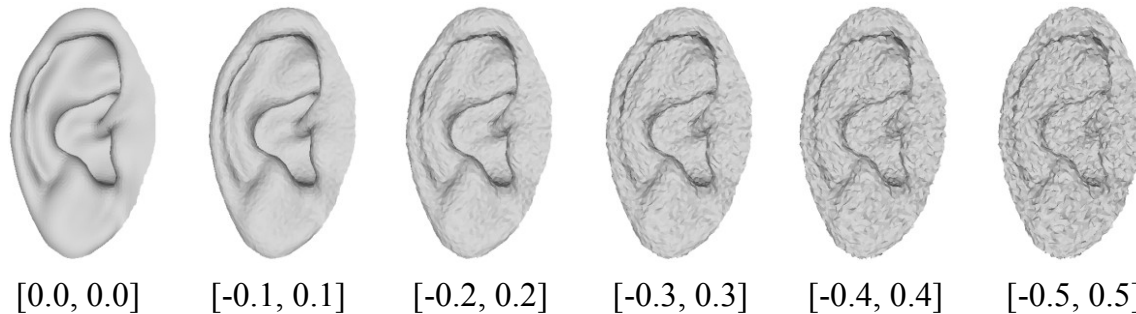
$$E_{BC} = \frac{\sum_{i=1}^m |M_i - I_i|}{\sum_{i=1}^m |M_i + I_i|}, \quad 0 \leq E_{BC} \leq 1$$

M_i и I_i са хистограмите на модела и входния обект.

2. Представяне на 3D обекти чрез EGI [2]

2.3. Експериментални резултати

- Целта на преведените експерименти е да се определи най-подходящото разстояние, устойчиво на шум, за оценка на подобие между EGI хистограмите за целите на разпознаването.
- Предназначението на шума е да се симулира различна точност на 3D сканиращото устройство. За тази цел е генериран равномерен шум в определен диапазон, който е прибавен към всяка от 3D координатите (x, y, z) на сканирания обект.



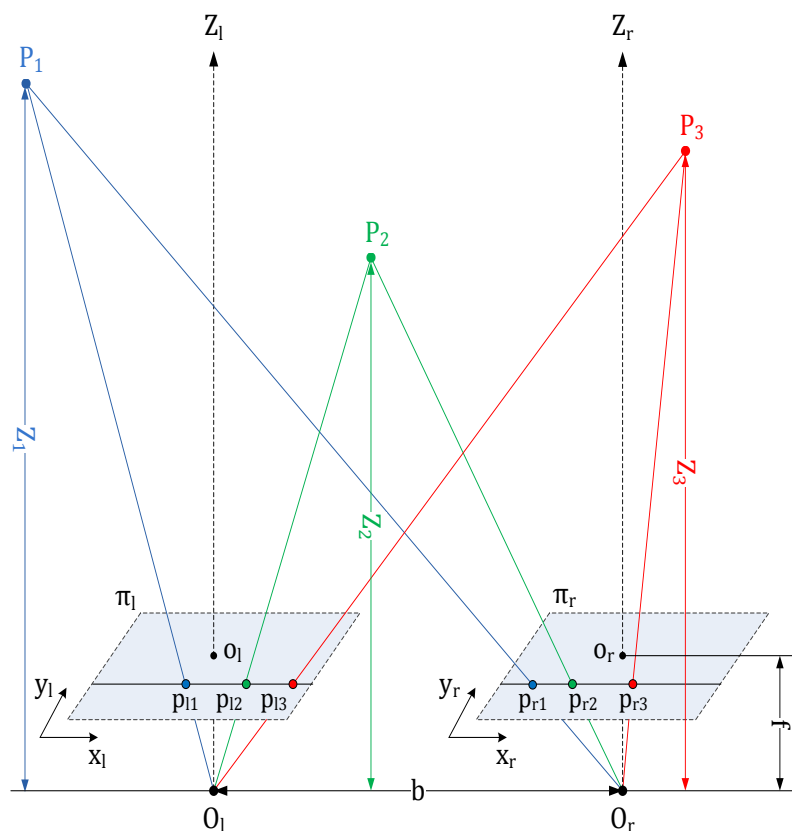
За разпознаването е приложен методът на най-близкия съсед, който разпознава по най-късото разстояние между EGI хистограмата на входното зашумено ухо, сравнена с всички модели.

Distance \ Noise	± 0.1		± 0.2		± 0.3		± 0.4		± 0.5	
	$E_M(p = 1)$	100%	0.415	100%	0.655	100%	0.828	72.7%	0.939	36.4%
$E_M(p = 2)$	100%	0.427	100%	0.705	63.6%	0.917	45.5%	1.031	36.4%	1.101
E_{BC}	100%	0.418	100%	0.657	100%	0.830	90.9%	0.928	54.5%	0.987

True Recognition Rate (TRR) и съответната осреднена достоверност на разпознаването (the smaller, the better), базирани на изследваните разстояния и нива на шум. 12

3. Тримерна реконструкция на облак от точки

3.1. Канонична система за стереозрение



Основни параметри:

- **Базово разстояние** b [cm] между оптичните центрове O_l и O_r ;
- **Фокусно разстояние** f [mm] - от оптичните центрове до равнините на сензорите;
- **Физическия размер** P_{hsiz} [μm] на един пиксел от сензорите на камерата;
- **Координатите** x_{li} и x_{ri} [pixels] на съответстващите си точки в лявото и дясното изображение, проекции на една и съща 3D точка P_i ;
- **Разстояние** Z_i [m] между линията, свързваща оптичните центрове на камерите и 3D точката P_i от сцената.

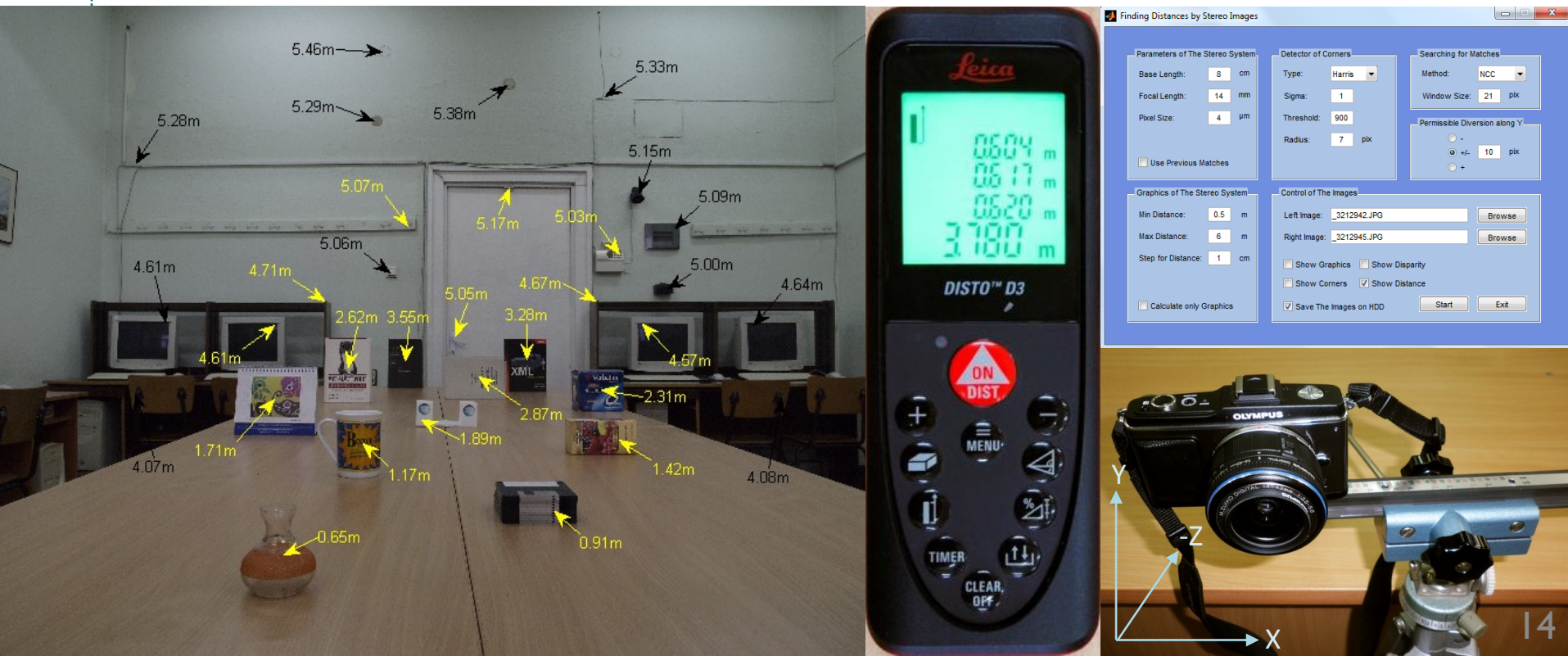
$$Z_i = \frac{bf}{(x_{li} - x_{ri}) * p_{hsiz}}$$

Работна идея: Тази канонична стереодвойка може да се приложи към всеки два припокриващи се кадъра от видеоклип. Ако клипът съдържа трансляция (евентуално и ротация), но само във фокалната равнина, т.е. подходящ е за нашата 2D видеостабилизация, то можем да изчислим базата b за двойката кадри, и т.н. за всяка следваща двойка. Оттук - идеята за 3D разредена реконструкция на обект в сцената.

3. Тримерна реконструкция на облак от точки [3]

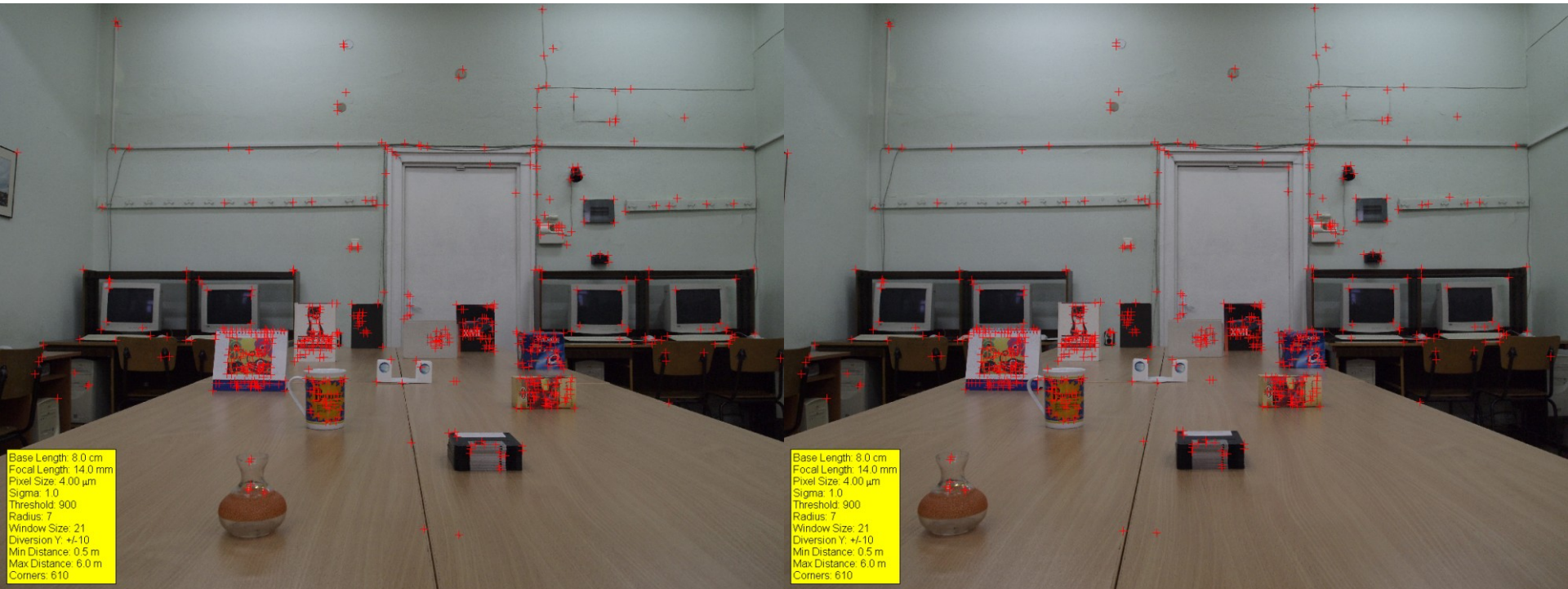
3.2. Проведени експерименти – постановка на задачата

- Всички двойки изображения са заснети с фотоапарат Olympus E-P2 чрез хоризонталното му транслиране на определено разстояние (6, 7, 8, 9 и 10 cm) от дадена референтна позиция.
- Измерват се разстоянията до различни обекти от сцената с помощта на лазерен далекомер Leica Disto™D3 с точност ± 1 mm. Целта е сравнение на измерените с изчислените разстояния до определени обекти и оценка на грешката.



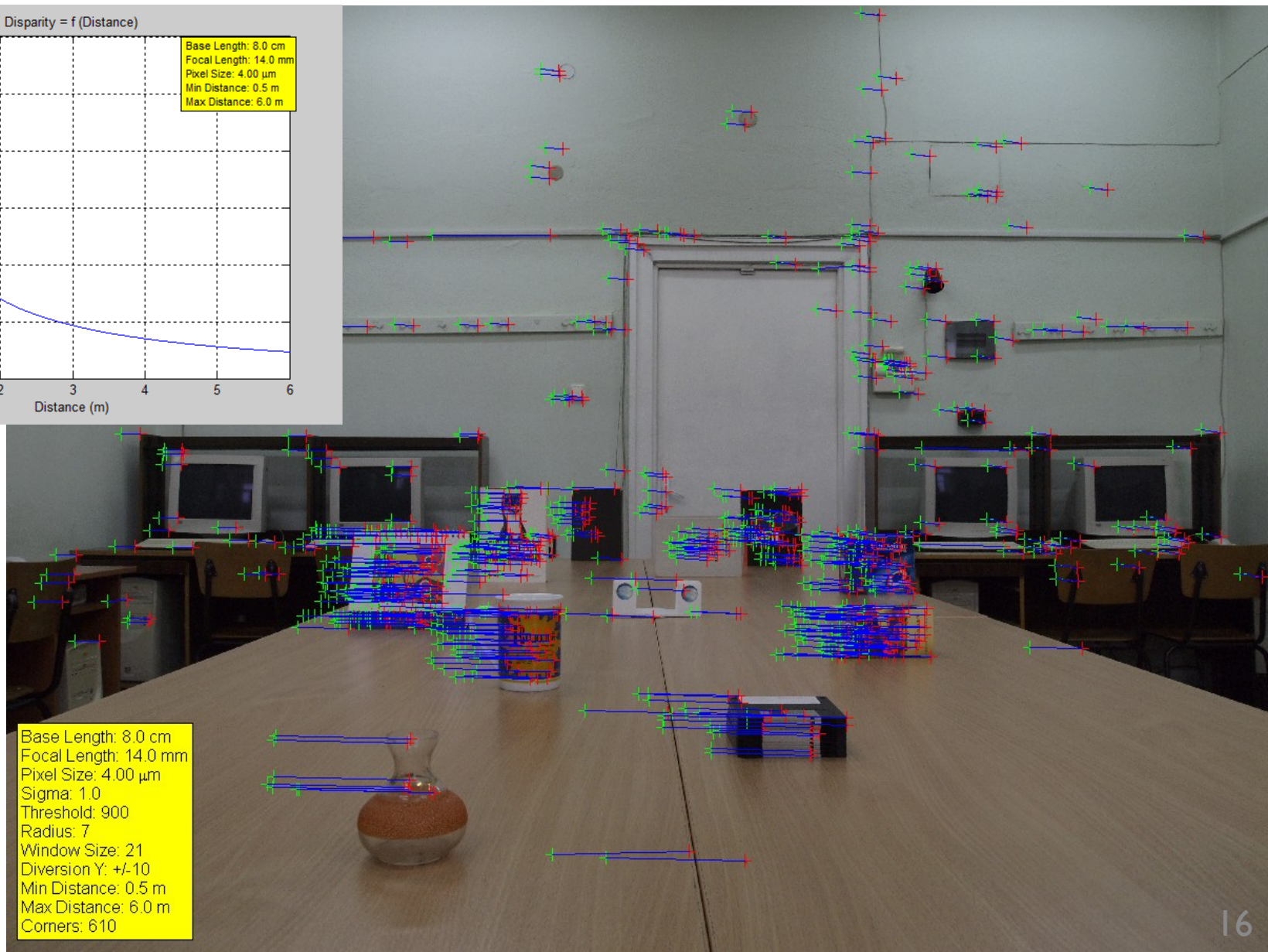
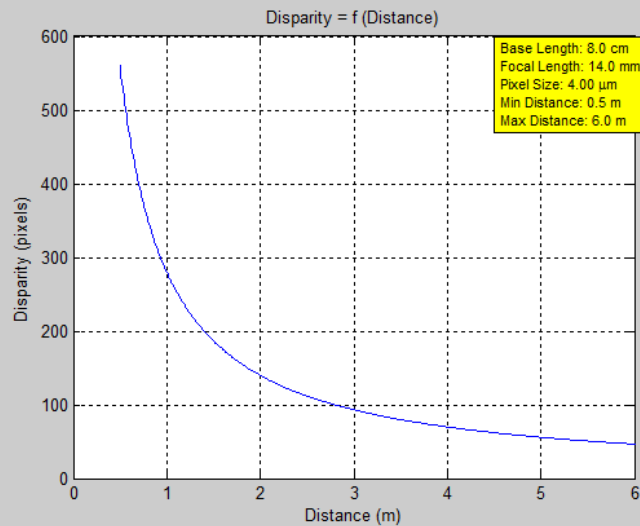
3. Тримерна реконструкция на облак от точки [3]

3.2. Проведени експерименти – намиране на характерни точки по Харис



3. Тримерна реконструкция на облак от точки [3]

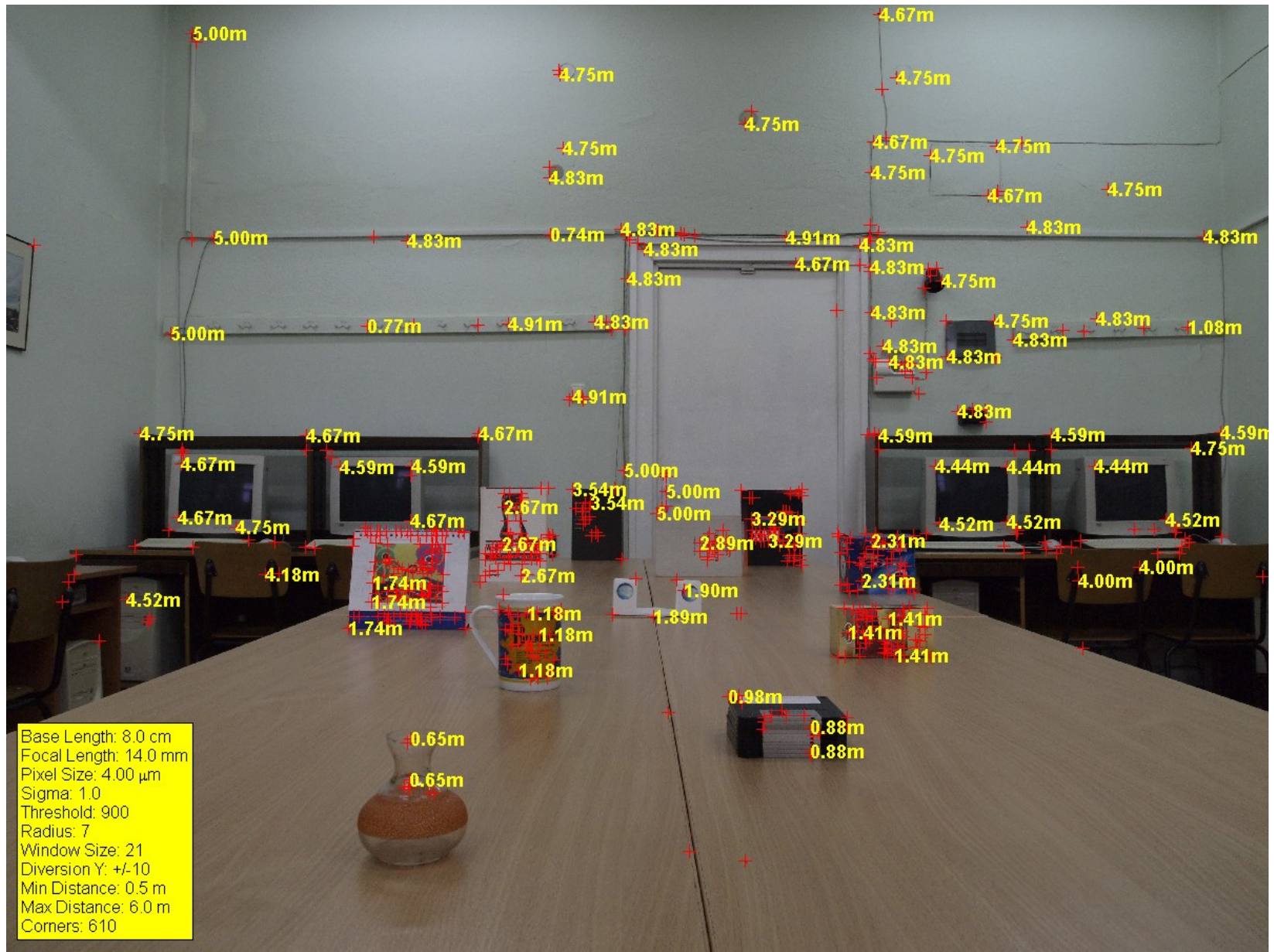
3.2. Проведени експерименти – намиране на съответни точки



Base Length: 8.0 cm
Focal Length: 14.0 mm
Pixel Size: 4.00 μm
Sigma: 1.0
Threshold: 900
Radius: 7
Window Size: 21
Divergence Y: +/-10
Min Distance: 0.5 m
Max Distance: 6.0 m
Corners: 610

3. Тримерна реконструкция на облак от точки [3]

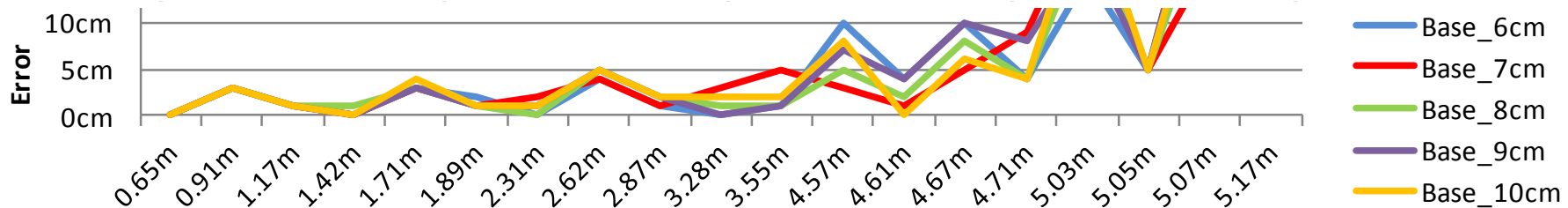
3.2. Проведени експерименти – изчислени разстояния



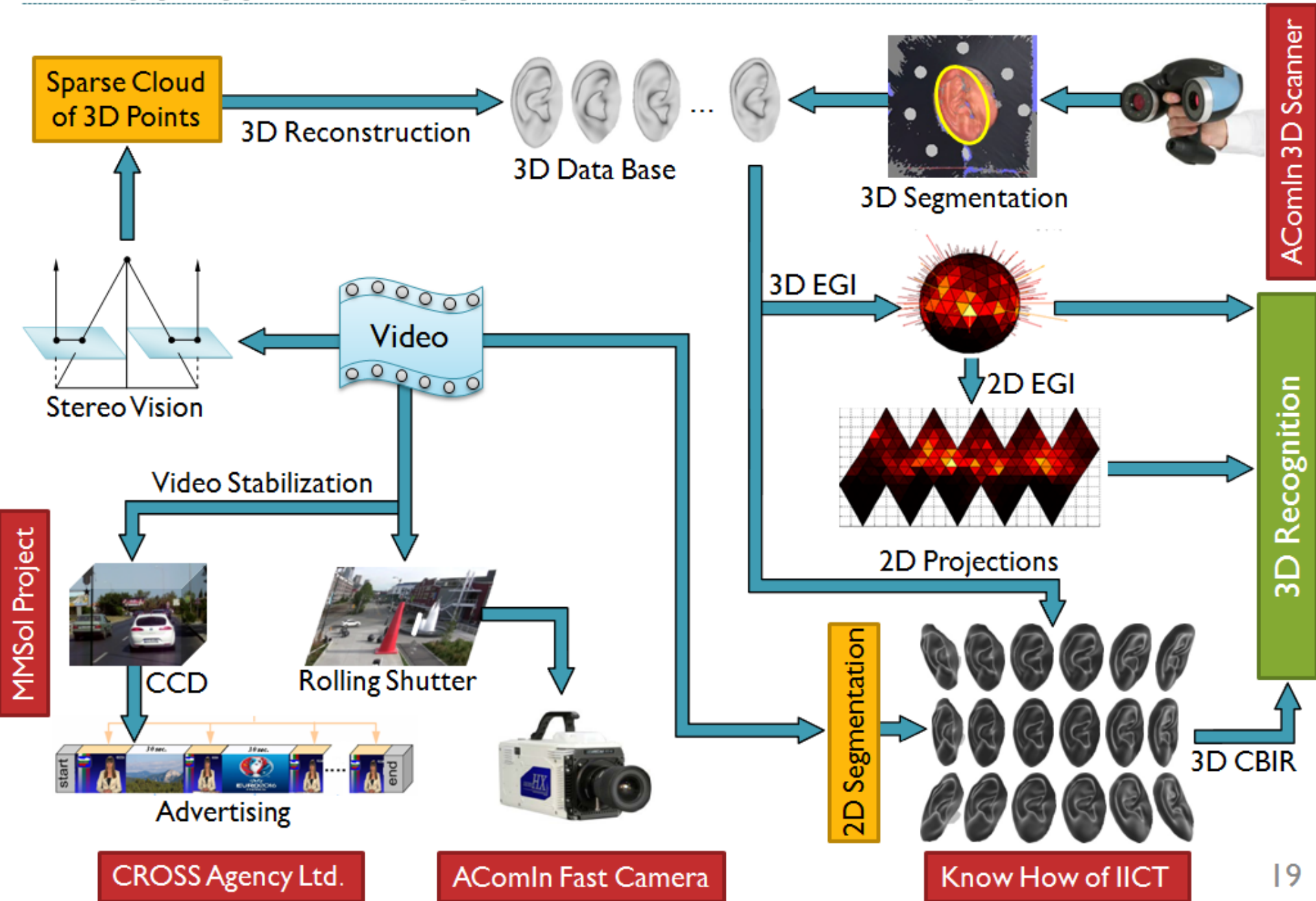
3. Тримерна реконструкция на облак от точки [3]

3.2. Проведени експерименти – оценка на грешката

Real Distance	Base length between cameras with focal length 14 mm									
	6cm		7cm		8cm		9cm		10cm	
	Est.	Error	Est.	Error	Est.	Error	Est.	Error	Est.	Error
0.65m	0.65m	0cm	0.65m	0cm	0.65m	0cm	0.65m	0cm	0.65m	0cm
0.91m	0.88m	3cm	0.88m	3cm	0.88m	3cm	0.88m	3cm	0.88m	3cm
1.17m	1.18m	1cm	1.18m	1cm	1.18m	1cm	1.18m	1cm	1.18m	1cm
1.42m	1.42m	0cm	1.42m	0cm	1.41m	1cm	1.42m	0cm	1.42m	0cm
1.71m	1.74m	3cm	1.74m	3cm	1.74m	3cm	1.74m	3cm	1.75m	4cm
1.89m	1.91m	2cm	1.90m	1cm	1.90m	1cm	1.90m	1cm	1.90m	1cm
2.31m	2.31m	0cm	2.33m	2cm	2.31m	0cm	2.32m	1cm	2.30m	1cm
2.62m	2.66m	4cm	2.66m	4cm	2.67m	5cm	2.67m	5cm	2.67m	5cm
2.87m	2.88m	1cm	2.88m	1cm	2.89m	2cm	2.89m	2cm	2.89m	2cm
3.28m	3.28m	0cm	3.31m	3cm	3.29m	1cm	3.28m	0cm	3.30m	2cm
3.55m	3.56m	1cm	3.60m	5cm	3.54m	1cm	3.54m	1cm	3.57m	2cm
4.57m	4.47m	10cm	4.54m	3cm	4.52m	5cm	4.50m	7cm	4.49m	8cm
4.61m	4.57m	4cm	4.62m	1cm	4.59m	2cm	4.57m	4cm	4.61m	0cm
4.67m	4.57m	10cm	4.62m	5cm	4.59m	8cm	4.57m	10cm	4.61m	6cm
4.71m	4.67m	4cm	4.62m	9cm	4.67m	4cm	4.63m	8cm	4.67m	4cm
5.03m	4.88m	15cm	4.80m	23cm	4.83m	20cm	4.85m	18cm	4.79m	24cm
5.05m	5.00m	5cm	5.00m	5cm	5.00m	5cm	5.00m	5cm	5.00m	5cm
5.07m	4.77m	30cm	4.90m	17cm	4.83m	24cm	4.77m	30cm	4.79m	28cm
5.17m	4.77m	40cm	4.80m	37cm	4.83m	34cm	4.77m	40cm	4.79m	38cm



4. Структура на дисертацията – схематично представяне



5. Текущи и финални изследвания

- Актуализиране на направения обзор по видеостабилизация, отразявайки върховите постижения в областта през 2014 + обзор по 3D разпознаване.
- Разписване и получаване на прякото решение (не чрез матричен LSM) на търсените параметри (транслация и ротация) за метода с 9x9 разбиване на кадрите, подобно на разработения 3x3 вариант.
- Съставяне на аналогичен векторен модел за намиране и на други параметри на движението като мащабиране и накланяне, компенсирани Rolling Shutter ефекта, наблюдаван при камерите със CMOS сензори.
- Изследване на различни подходи за отстраняване на грешните *вектори на движение*: 2nd pass of Least Squares Method (LSM); RANSAC; Histogram of Oriented Gradients (HoG); Dynamic Time Warping (DTW); Floating Average.
- Адаптиране на известни или прилагане на нови методи за 2D разпознаване към 2D EGI хистограмата на даден 3D обект.

6. Насоки за бъдеща работа (след защитата)

- Обединяване на резултатите от софтуерната видеостабилизация с информацията от инерциалните сензори (жироскопи, акселерометри) на "GSM".
- Плътна (неразредена) 3D реконструкция на обекти от видеопоток.
- Сегментиране на 2D/3D обекти, запълване на дупки в 3D структурата им.

Списък с публикации

1. 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, ISBN: 978-1-4503-2753-4, pp. 124-133.
2. 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 International Workshop on Biometrics, BIOMET'2014, June 23-24, 2014, Sofia, Bulgaria, LNCS, Springer, 15 pages (to appear).
3. Nikolova, I., A. Nikolov, A., and G. Zapryanov: *Depth estimation using shifted digital still camera*, In: Proceedings of the 12th Int. Conf. on CompSysTech'11, June 16-17, 2011, Vienna, Austria, ACM ICPS, Vol. 578, ISBN: 978-1-4503-0917-2, pp. 234-240.
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, ISBN: 978-954-9526-87-5, pp. 35-47.
5. Николов, А.: *Методи за откриване на характерни точки и тяхното съответствие в стереодвойки изображения*, В: Сборник трудове „Дни на науката 2011”, XV Майски четения, 27 Май, 2011, В.Търново, Том 2, ISSN: 1314-2283, pp. 259-276.
6. Николов, А., В. Иванова: *Статистически подход за анализ и графично представяне на данни от анкета по проблеми на докторантското обучение*, В: Сборник научни трудове на Русенския университет, Научна конференция РУ&СУ'11, 28-29 Октомври, 2011, Русе, Том 50, серия 6.1, ISSN: 1311-3321, pp. 28-33).

Участие в проекти / договори

1. Математически алгоритми за видео стабилизация и филтриране на шум от сензори за движение. Договор за изследвания и разработка (№ 210214) между “ММ Солушънс” АД и ИИКТ от 07.05.2012г. до 02.10.2012г.
2. Подпроект “Стабилизация на видео от високоскоростна камера”, по договор ДО1-192 на МОН “Съфинансиране на проект АКОМИН 316087 по 7РП на ЕК”.
3. „Изграждане и развитие на млади висококвалифицирани изследователи за ефективно прилагане на биомедицинските изследвания за подобряване качеството на живот”, по Оперативна програма “Развитие на човешките ресурси” (Европейски Социален Фонд 2007-2013) на МОН, договор BG051PO001-3.3.06-0048/04.10.2012 с Институт по експериментална морфология, патология и антропология с музей – БАН (координатор).
4. “Подкрепа на творческото развитие на докторанти, пост-докторанти и млади учени в областта на компютърните науки”, по Оперативна програма “Развитие на човешките ресурси” (Европейски Социален Фонд 2007-2013) на МОН, договор BG051PO001-3.3.04/13 с Великотърновски университет “Св.св. Кирил и Методий” (координатор).



Секция

„Обработка на сигнали и разпознаване на образи“

Семинар за отчисляване от докторантура, ИИКТ-БАН, 12.12.2014

**Благодаря
за вашето внимание !**