Ph.D. Incoming Erasmus Student Seminar
at Computer Vision & Multimedia Lab,
University of Pavia, 27/03/2013

Depth Estimation using Single Digital Still Camera & 2D Real-Time Video Stabilization

Atanas Nikolov, Ph.D. student,

at Bulgarian Academy of Sciences,
Institute of Information and Communication Technologies,
Signal Processing and Pattern Recognition Department,
Web page: http://www.iict.bas.bg/SPIR/index-en.html
E-mail: a.nikolov@iinf.bas.bg

Education and Culture DG
Lifelong Learning Programme

Content

PART 1

Depth Estimation using Single Digital Still Camera

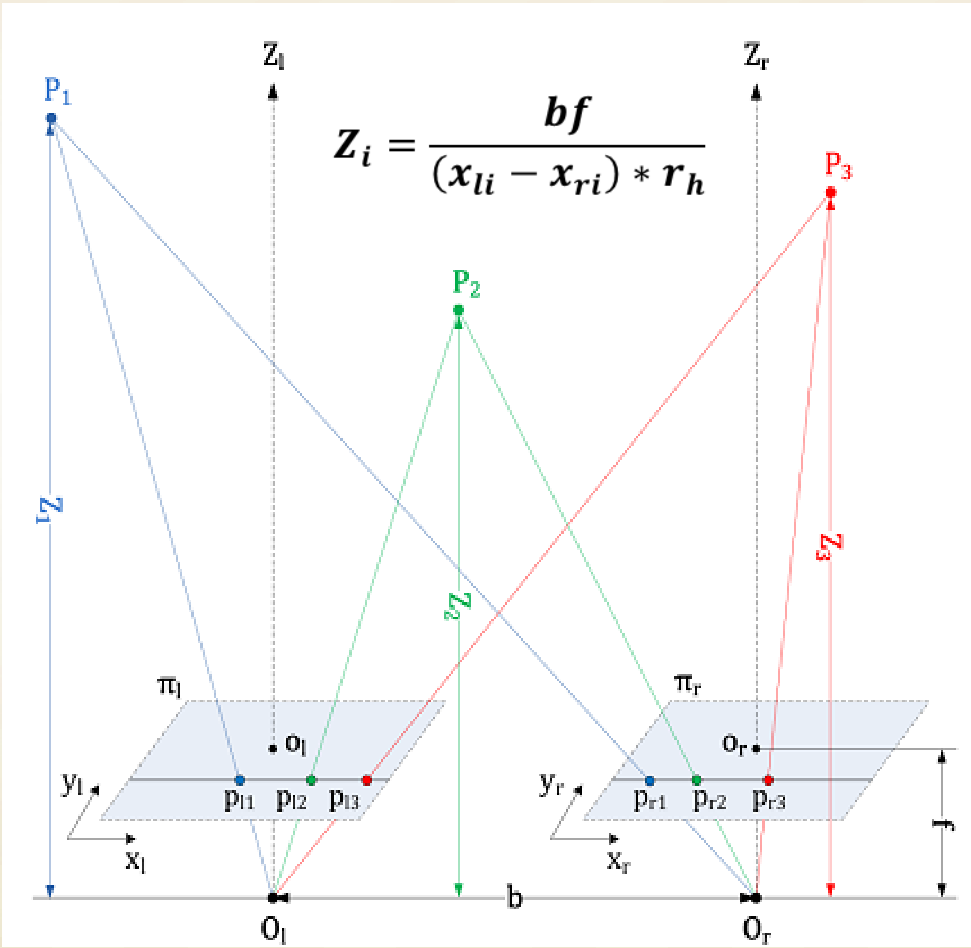
PART 2

2D Real-Time Video Stabilization

Depth Estimation using Single Digital Still Camera

- The present work's objective is to investigate the possibilities of a simple method for acquiring the depth using the principles of a canonical stereo vision system.
- The aim is to prove by physical experiments, using conventional digital still camera, when a real stereovision system is not available, is possible to effectively determine the depth to particular object points in a given static scene.
- The main requirement is that the camera should have precise horizontal movement, high resolution and possibilities for adjusting the parameters of its optical system.

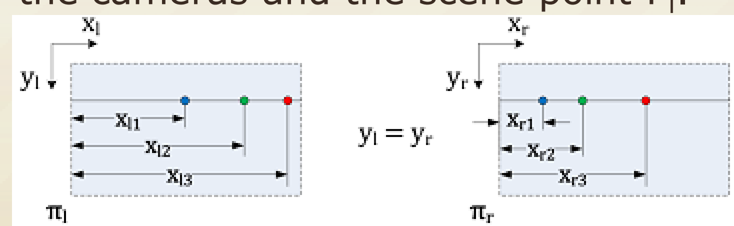
Geometrical Model of a Canonical Stereovision Configuration



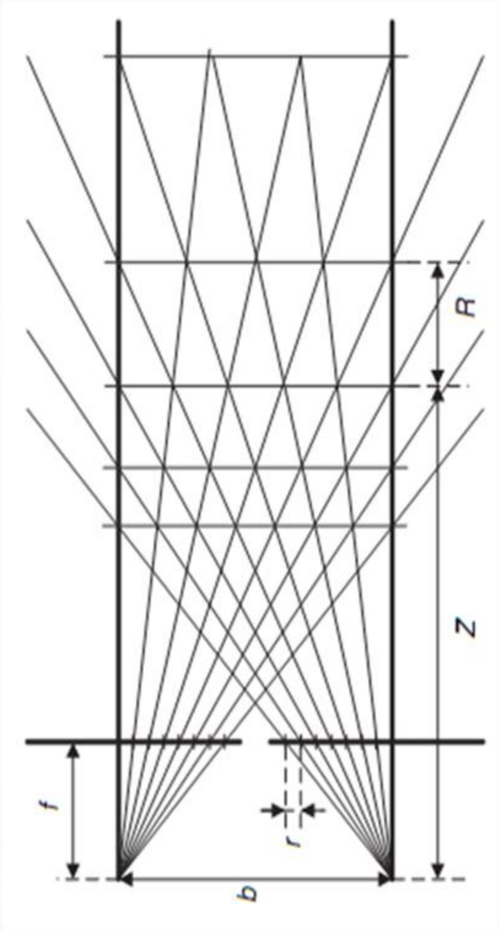
Geometrical model of a canonical stereo configuration

Parameters of a CSC:

(1) O_l and O_r - optical centers of the cameras; ***(2) Z_l and Z_r*** - parallel principal axes that pass through points O_l and O_r which are perpendicular to the image planes π_l and π_r; ***(3) base length, b*** - the distance between the points O_l and O_r; ***(4) focal length, f*** - the distance from the image planes to the central points O_l and O_r; ***(5) p_{li} and p_{ri}*** - the projections of an arbitrary point P_i within the scene on the image planes π_l and π_r; ***(6) x_{li} and x_{ri}*** - distances, measured from the top left corner of the images to the corresponding projections of a given point P_i (p_{li} and p_{ri}); ***(7) Z_i*** - the distance between the line connecting the optical centers of the cameras and the scene point P_i.



Depth Resolution of a Canonical Stereovision System



The phenomenon of diminishing accuracy of depth measurement with increasing distance from the camera planes is a geometrical limitation since it depends exclusively on geometrical parameters of a stereovision system.

$$R \approx \frac{rZ^2}{fb}$$

For most image acquisition systems, the values of ***r***, ***b*** and ***f*** are constant, at least for a single acquisition. This means that there is such a value ***Z*** for which it is not possible to measure the depth of the observed scene due to geometrical limitations of a stereovison setup.

Phenomenon of a limited accuracy of depth measurement with increasing distance from the camera

Depth Estimation Algorithm

The main problem in determining the distance to objects in a scene from a pair of stereo images, obtained by a canonical stereo configuration, is to find pairs of corresponding image points, which represent projection of one point from a 3D scene.

Step 1: Features detection in the stereo images by cornerness measure, proposed by Alison Noble, on the basis of the Harris corner detector.

Step 2: Searching for matches by the method *Sum of Absolute Differences* between the previously found feature points (corners).

Step 3: Depth estimation using the relationship between disparity, base length and focal length, obtained on the basis of the geometrical model of a canonical stereo configuration.

Corners Detection

Harris cornerness measure

$$C(x,y) = \det(M) - k\big(trace(M)\big)^2$$

where $k = 0.04 \div 0.06$

A. Noble cornerness measure

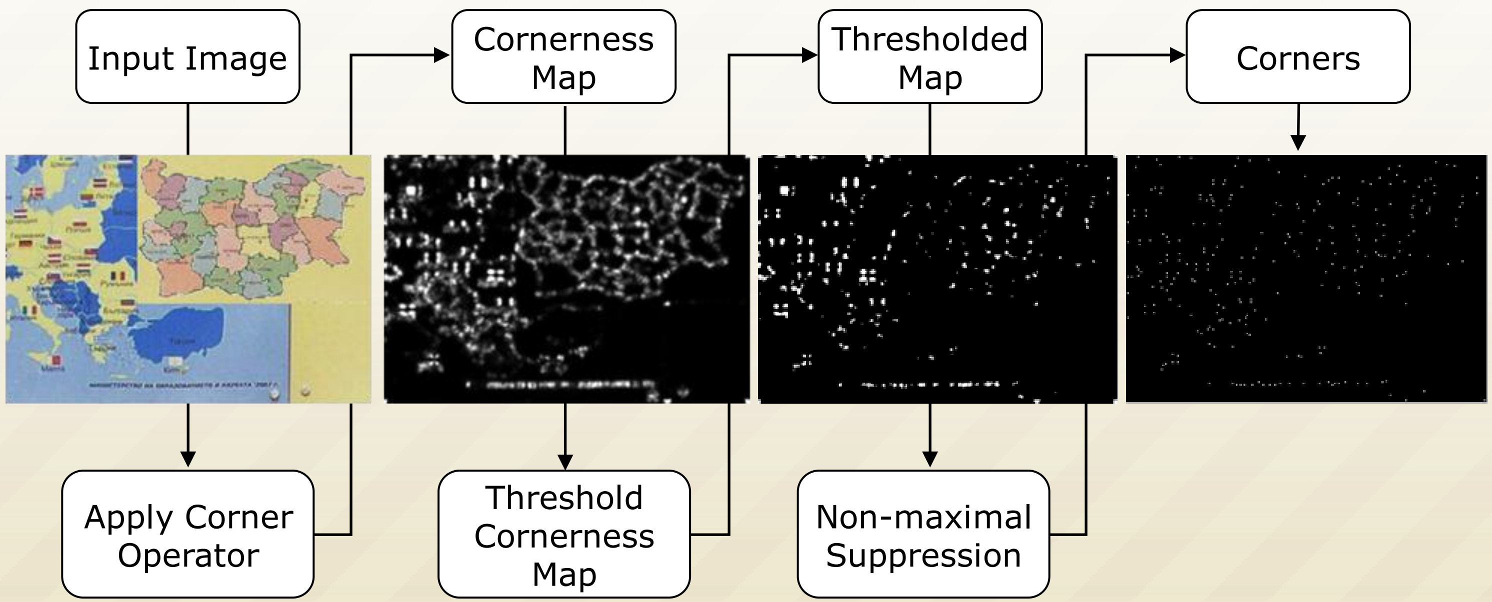
$$C(x,y) = \det(M)\,/\,trace(M)$$

$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix}$, $\det(M) = AB - C^2$, $trace(M) = A + B$, $A = I_x^2 \otimes w$, $B = I_y^2 \otimes w$, $C = I_x I_y \otimes w$

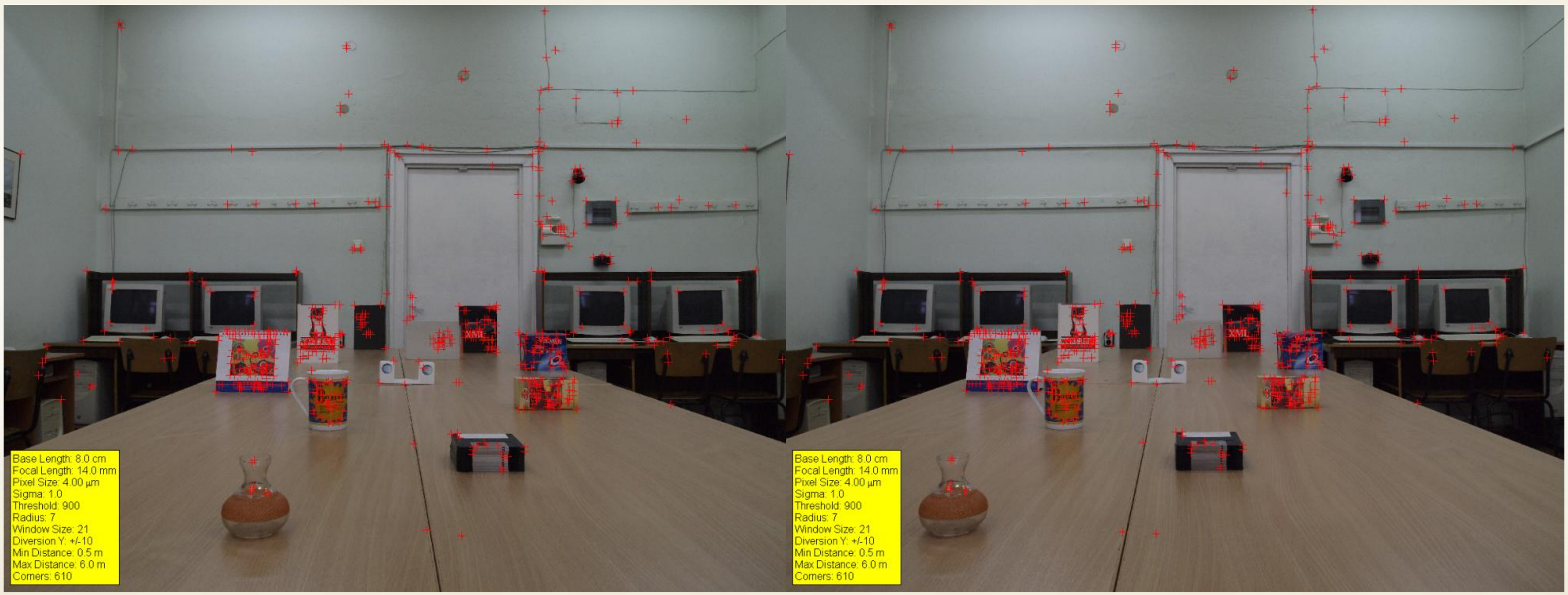
$\otimes$ *is the convolution opertor,*

w is the Gaussian window

$I_x = I \otimes (-1,0,1) \approx \dfrac{dI}{dx}$ *and* $I_y = I \otimes (-1,0,1)^T \approx \dfrac{dI}{dy}$

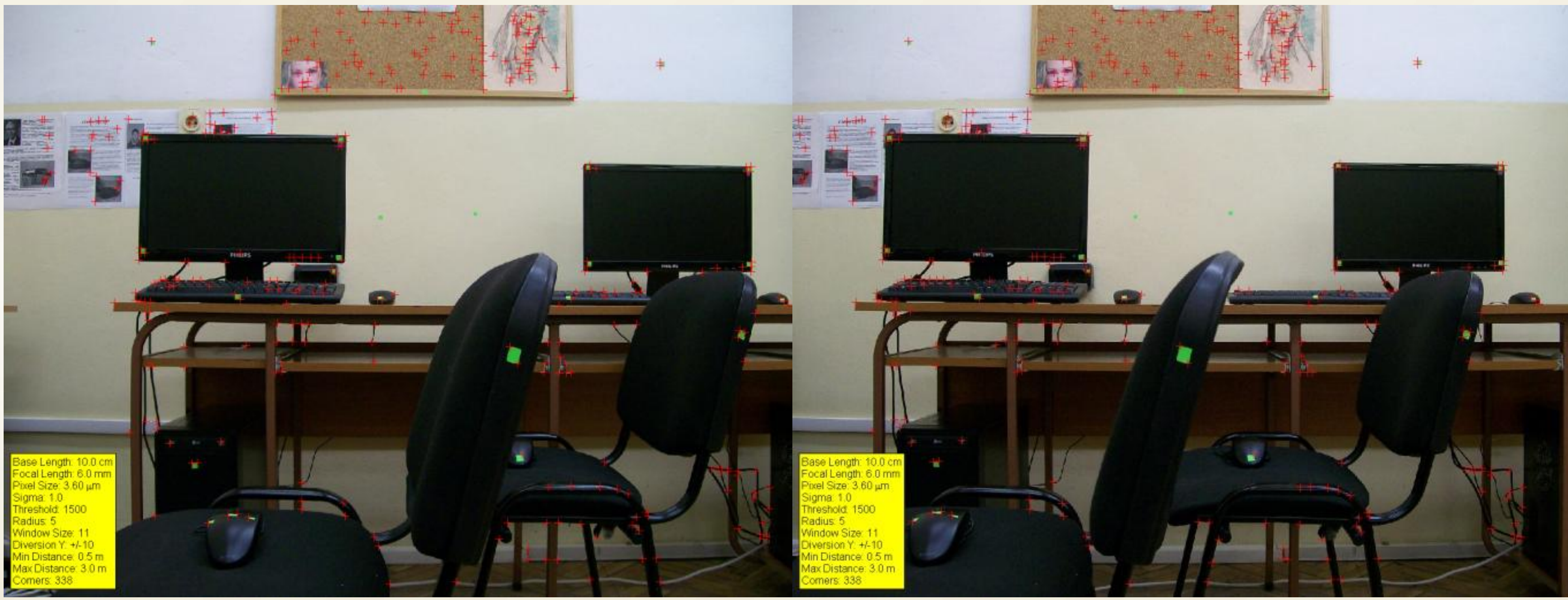


An Example of Found Corners in a Pair of Stereo Images



(Olympus)

An Example of Found Corners in a Pair of Stereo Images



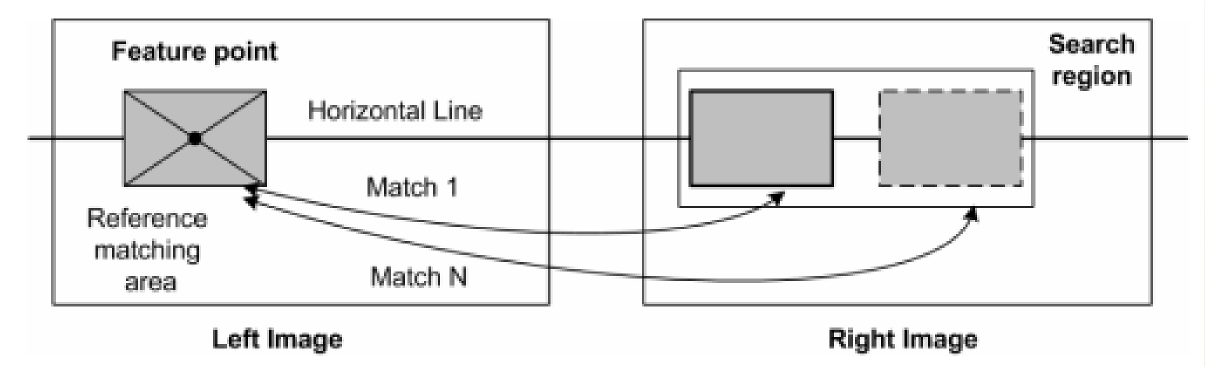
(Kodak)

Searching for Matches

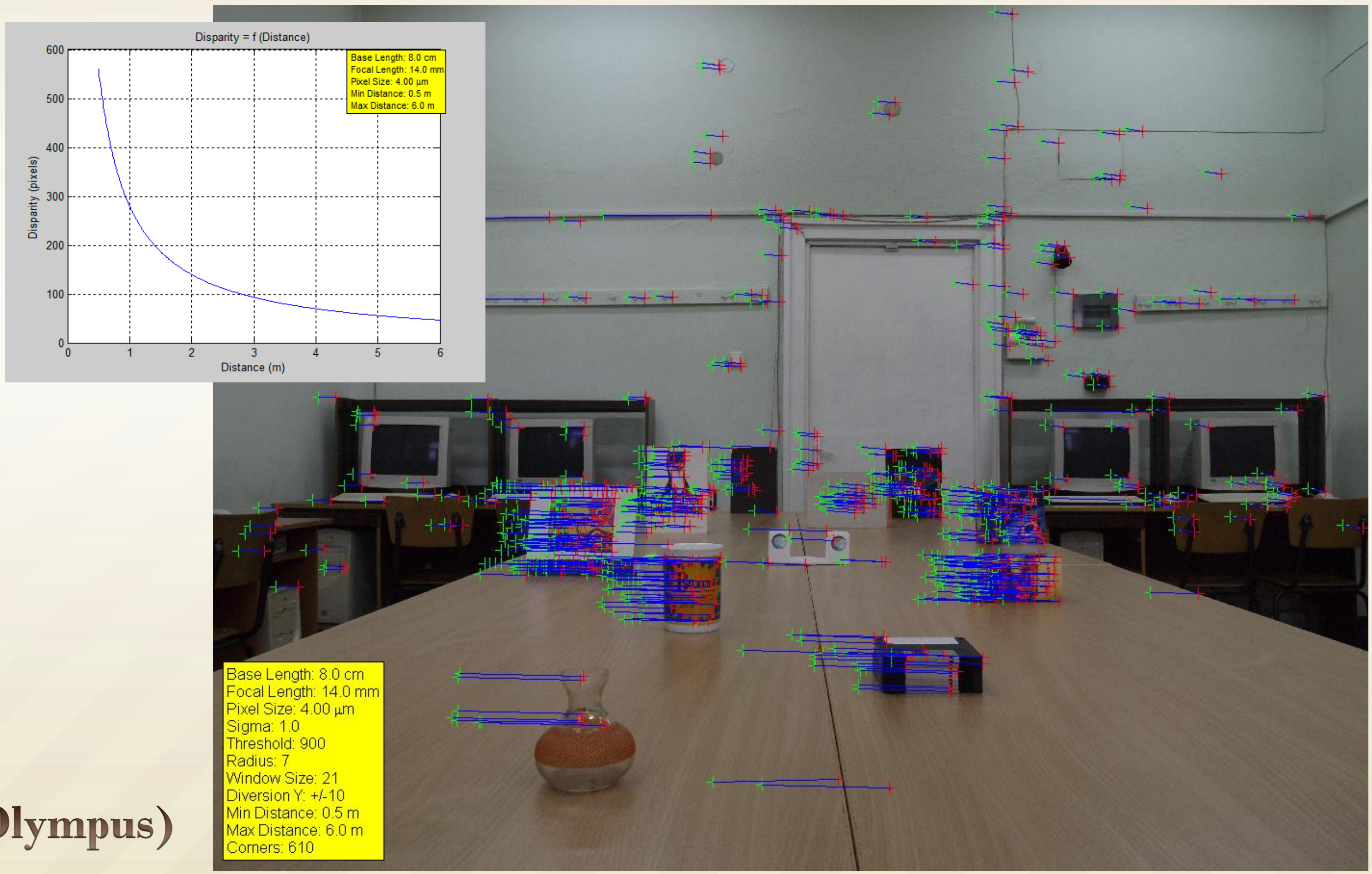
The metric used to determine which features in the stereo pair correspond to each other is based on *Sum of Absolute Differences*:

$$SAD = \sum_{(i,j)\in U} \left| I_1(x+i, y+j) - I_2\left(x+d_x+i, y+d_y+j\right) \right|$$

where: $\boldsymbol{I_1}$ and $\boldsymbol{I_2}$ are two image regions being compared. The region $\boldsymbol{I_1}$ is built around a reference point $\boldsymbol{(x , y)}$, and the region $\boldsymbol{I_2}$ - around point $\boldsymbol{(x+d_x , y+d_y)}$, where with $\boldsymbol{d_x}$ and $\boldsymbol{d_y}$ are denoted the relative horizontal and vertical displacements of the two image blocks being compared. The matching regions are defined by a set $\boldsymbol{U}$ of offset values, measured from their reference points, i.e. $\boldsymbol{(x , y)}$ and $\boldsymbol{(x+d_x , y+d_y)}$, respectively.

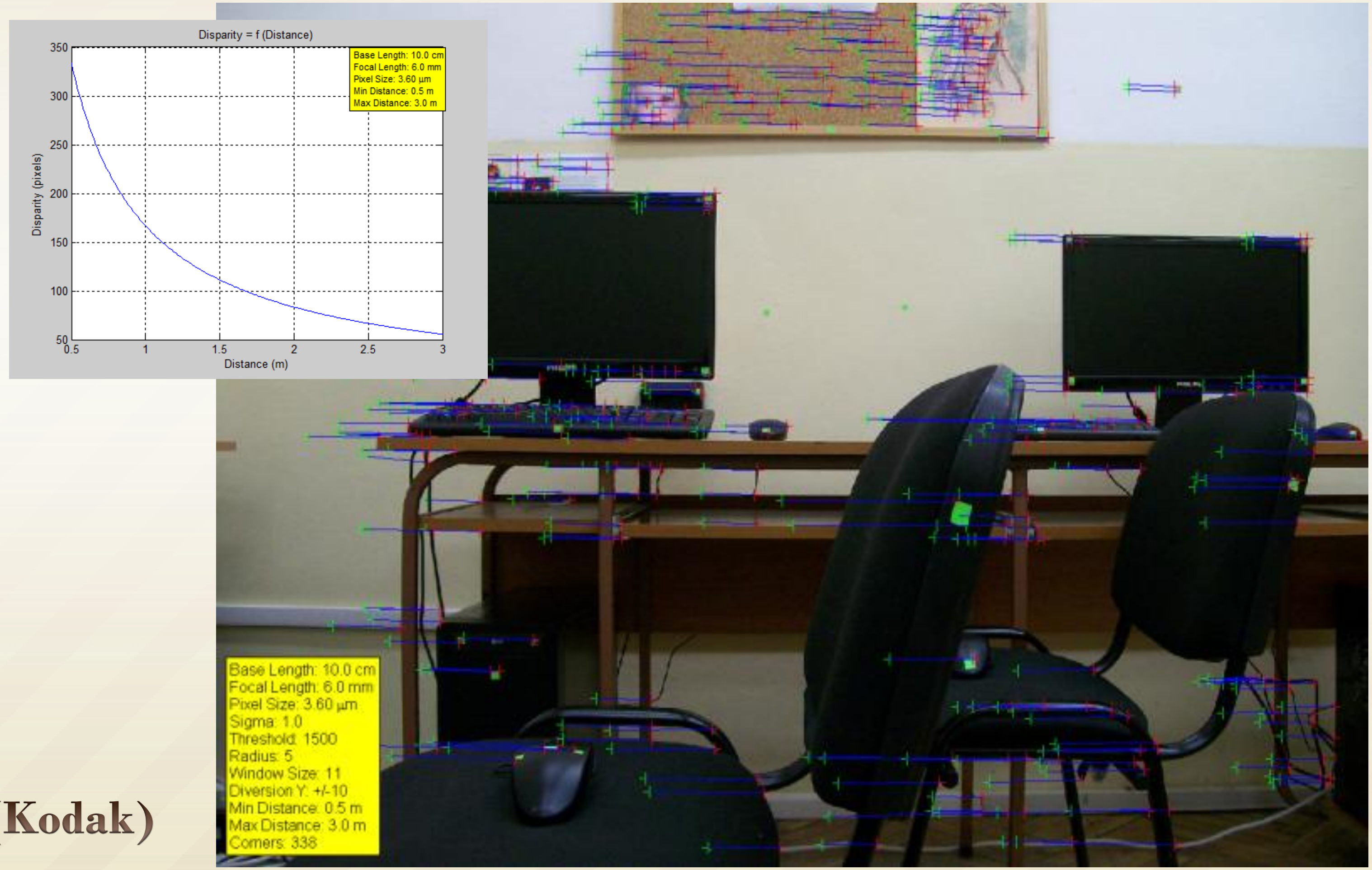


An Example of Found Correspondences Between Previously Detected Corners



(Olympus)

An Example of Found Correspondences Between Previously Detected Corners



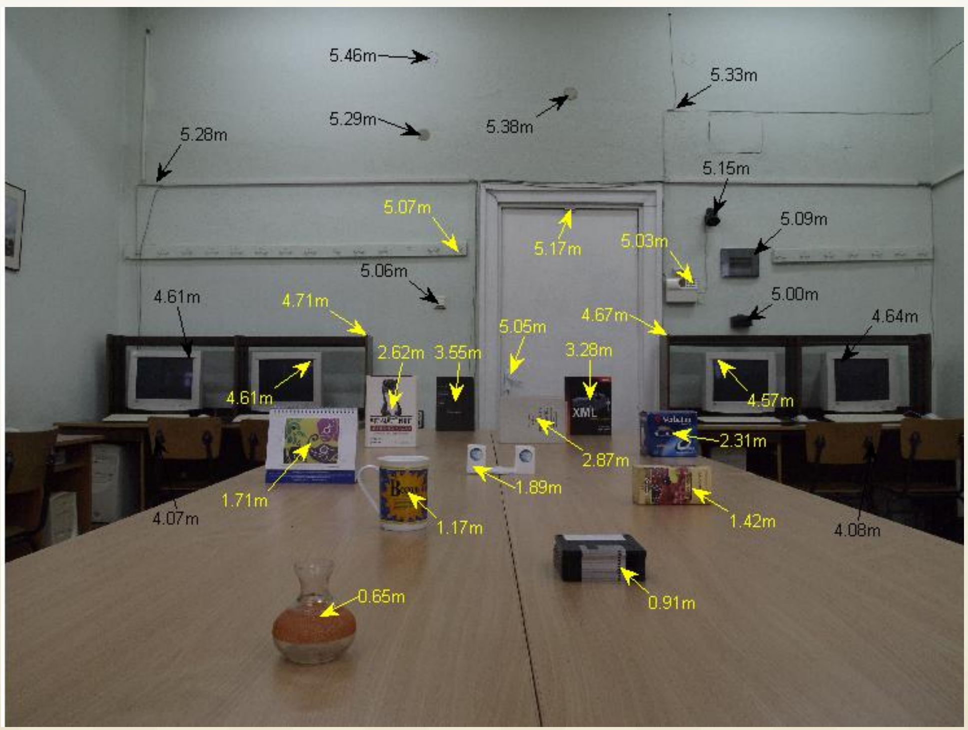
(Kodak)

Experimental Results

Our experimental work has two goals:

1) to verify the applicability of the mathematical model, using a non-real stereovision system;

2) to test the accuracy of the estimated distances in a real static scene.



Real Distances to selected objects (Olympus)

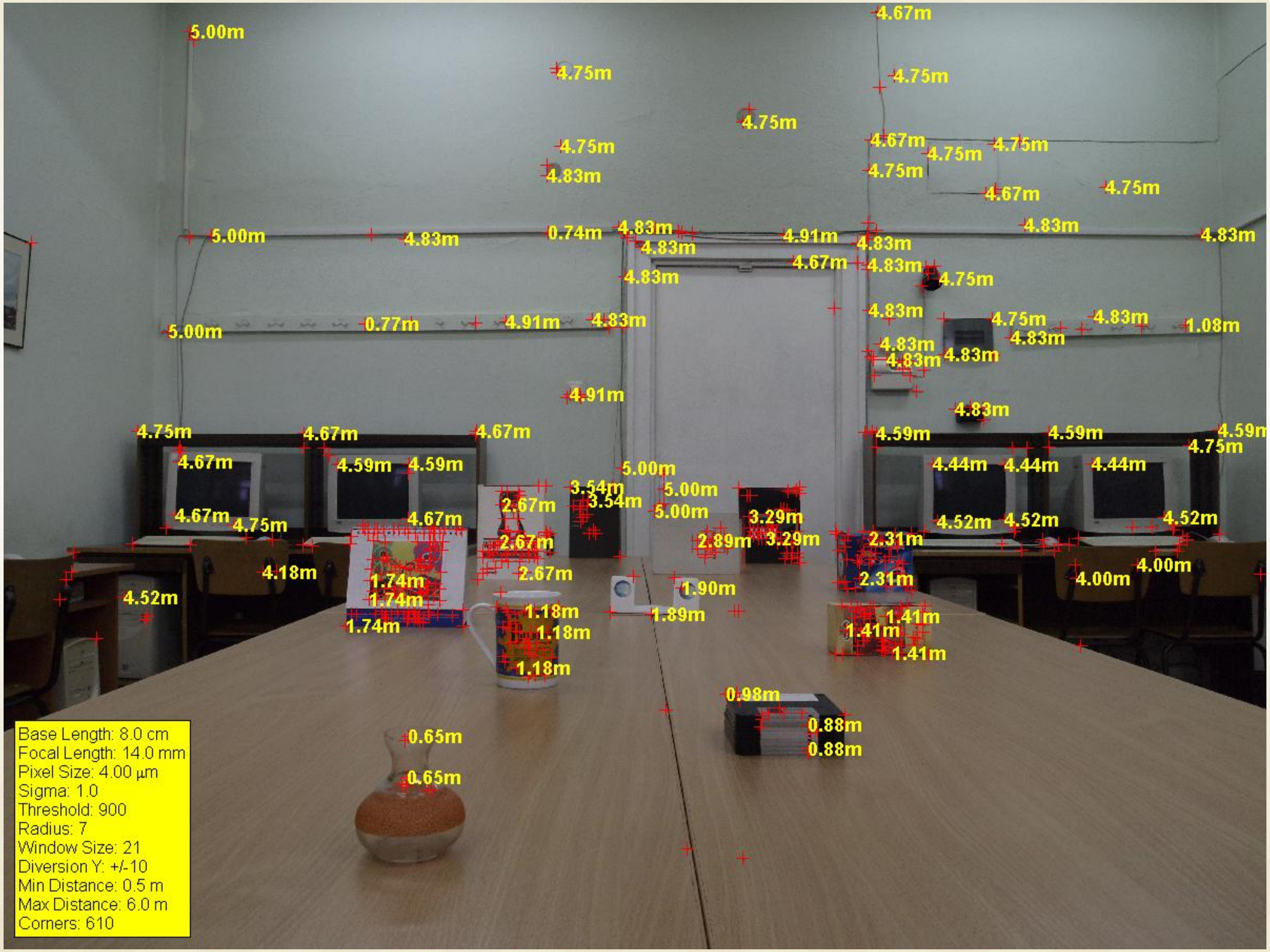
Experimental Platform

Laser distancemeter

Experimental Results

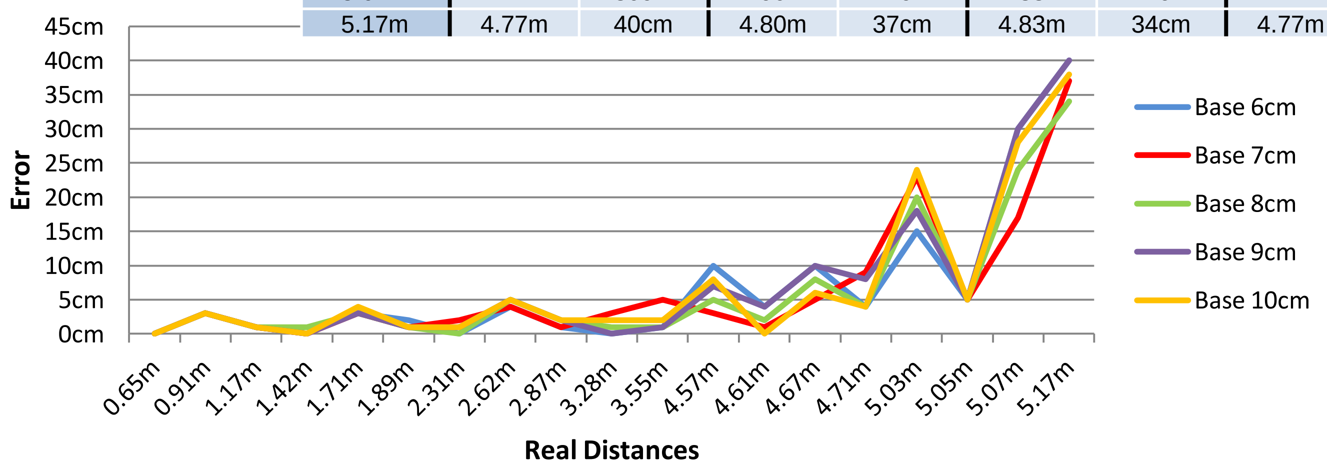
Real Distances to selected objects (Kodak) – unprofessional, but effective method ☺

Depth Estimation to Definite Points (Olympus)

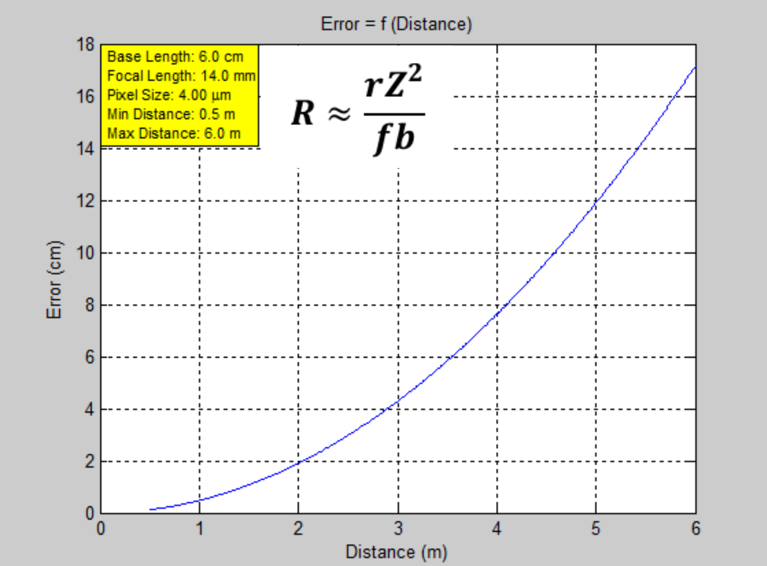


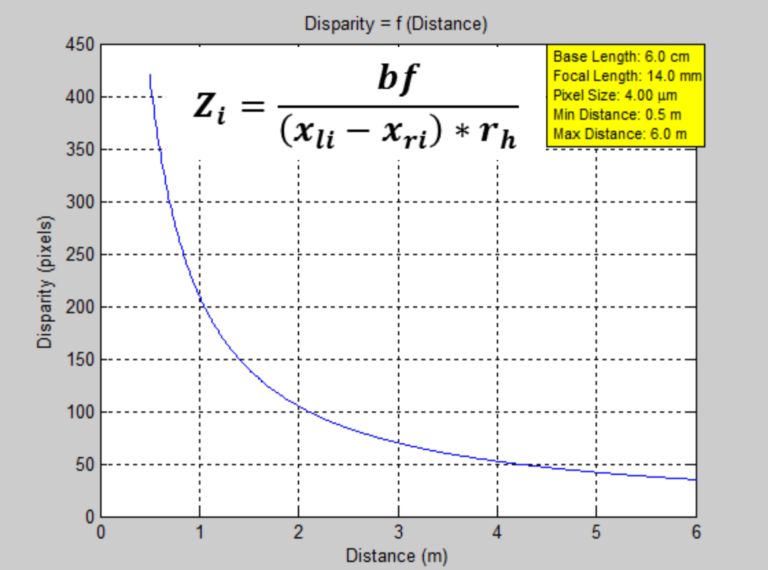
Experimental Results (Olympus)

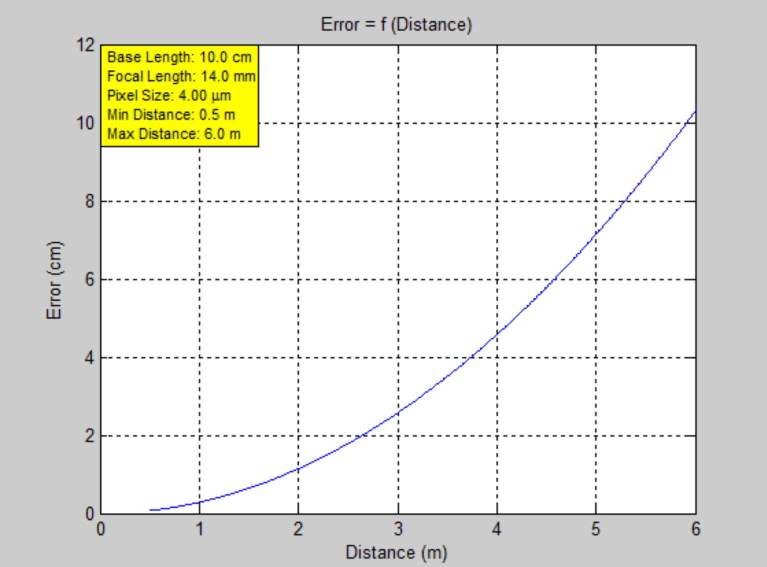
Real Distance	Base length between cameras with focal length 14 mm									
	6cm		7cm		8cm		9cm		10cm	
	Estimate	Error	Estimate	Error	Estimate	Error	Estimate	Error	Estimate	Error
0.65m	0.65m	0cm	0.65m	0cm	0.65m	0cm	0.65m	0cm	0.65m	0cm
0.91m	0.88m	3cm	0.88m	3cm	0.88m	3cm	0.88m	3cm	0.88m	3cm
1.17m	1.18m	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

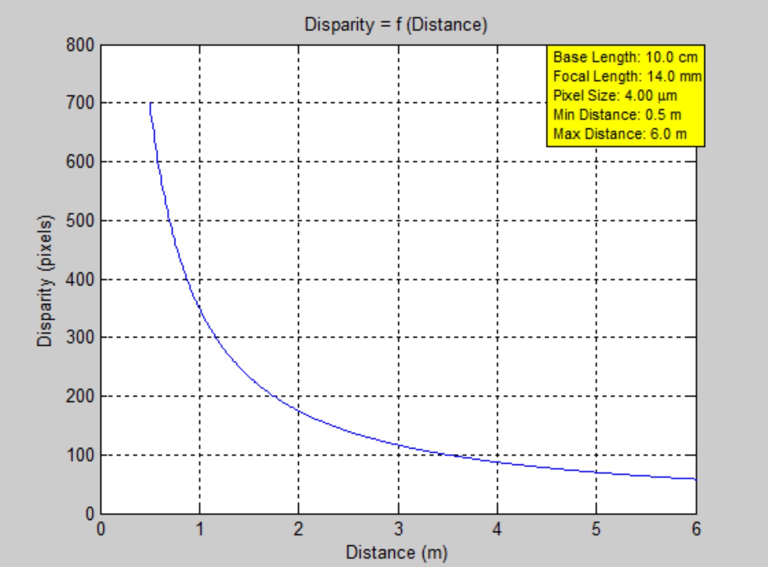


Depth Resolution Error (Olympus)

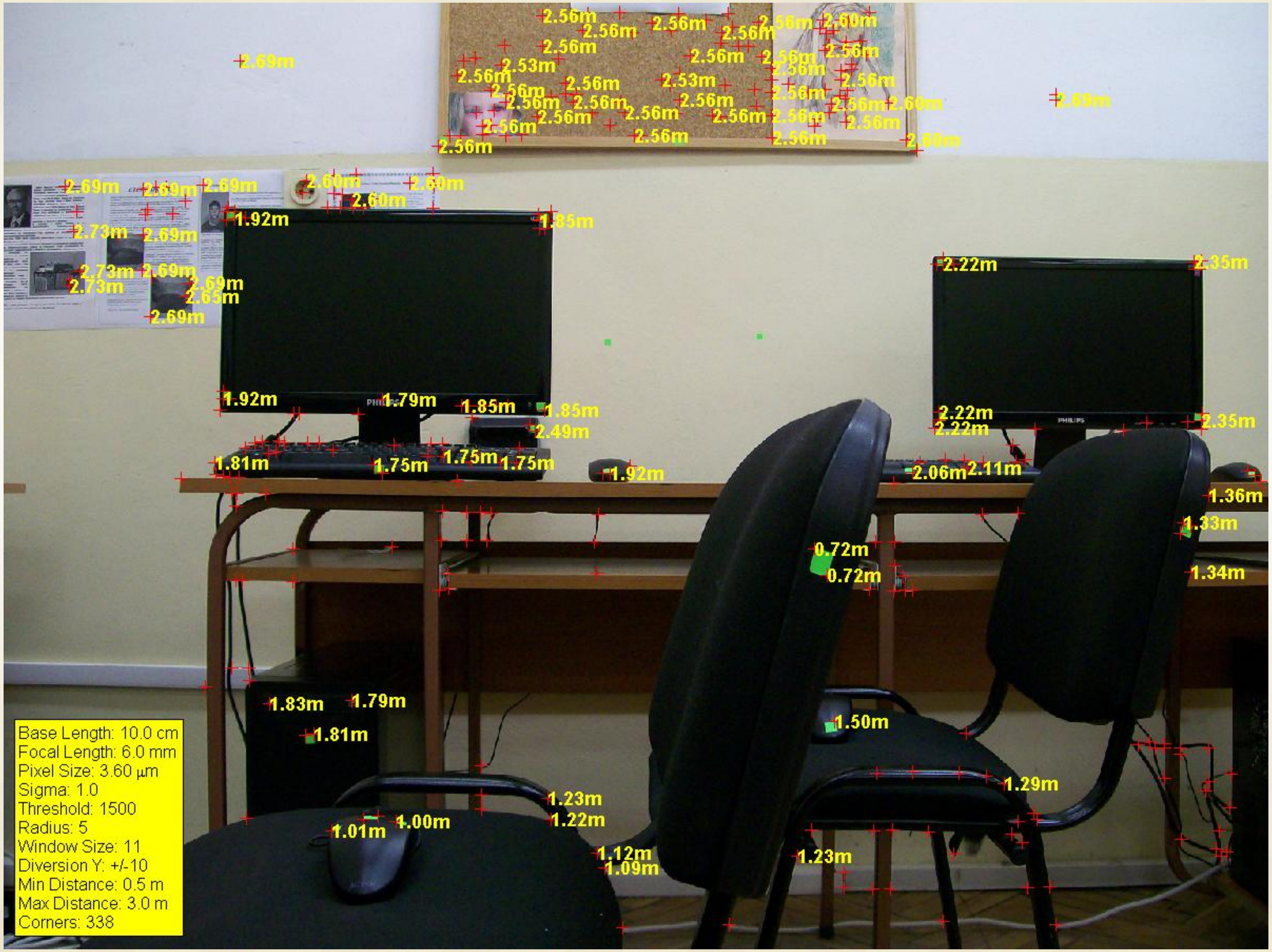






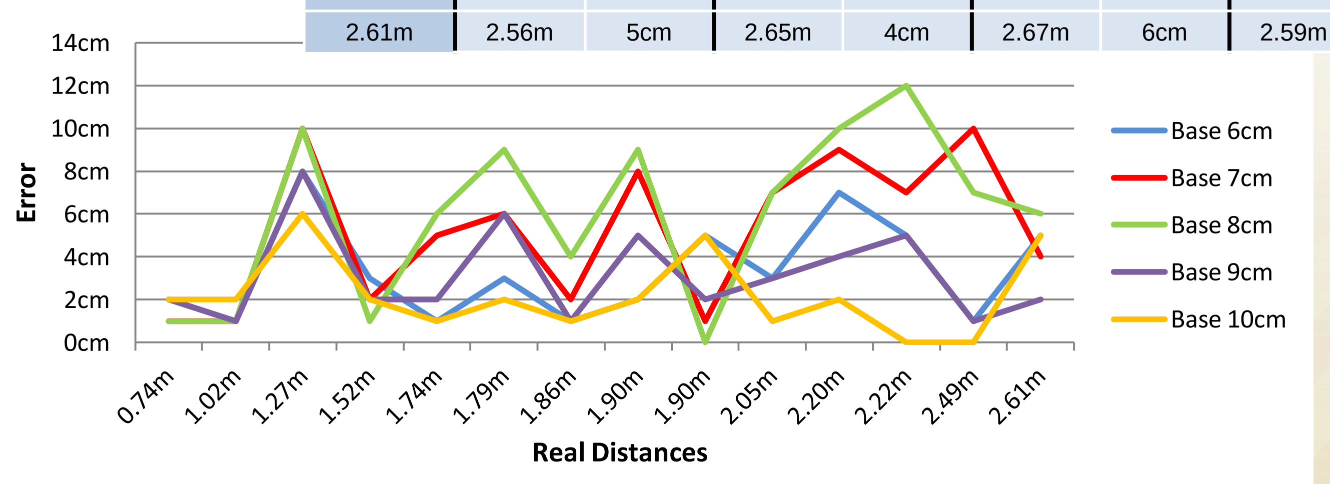


Depth Estimation to Definite Points (Kodak)

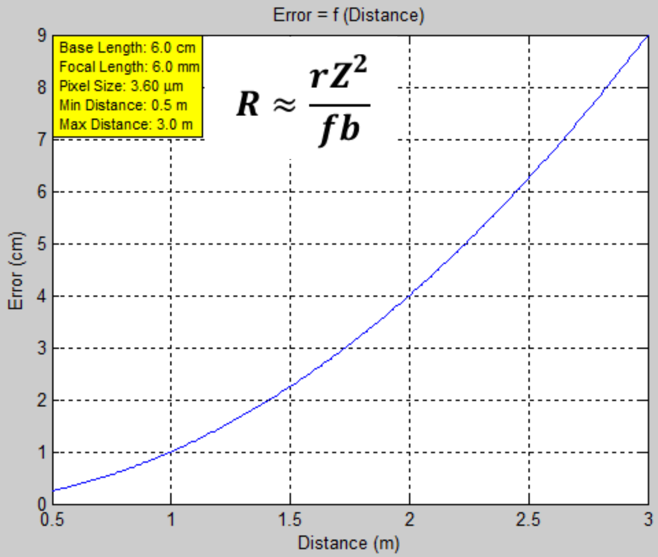


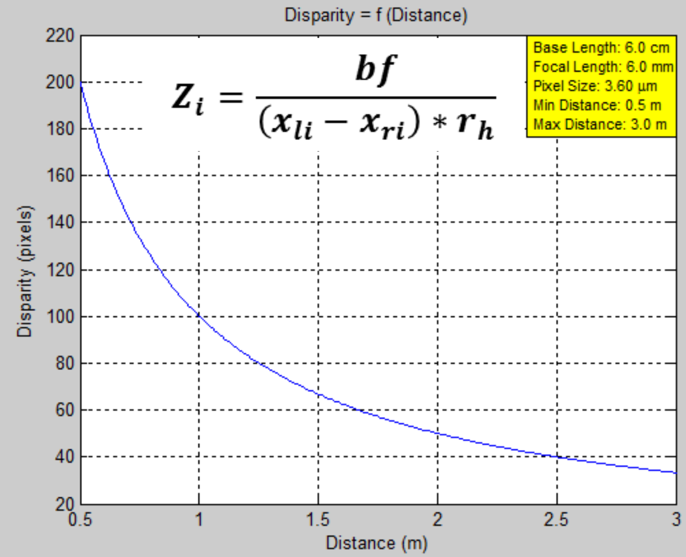
Experimental Results (Kodak)

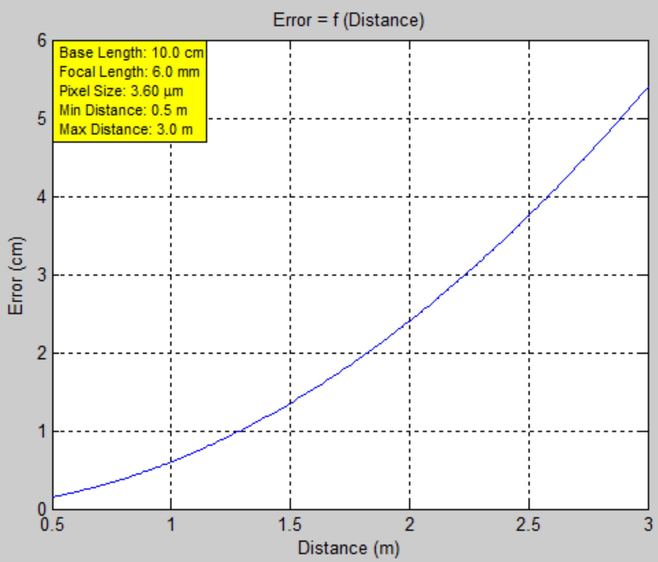
Real Distance	Base length between cameras with focal length 6mm									
	6cm		7cm		8cm		9cm		10cm	
	Estimate	Error	Estimate	Error	Estimate	Error	Estimate	Error	Estimate	Error
0.74m	0.72m	2cm	0.73m	1cm	0.73m	1cm	0.72m	2cm	0.72m	2cm
1.02m	1.01m	1cm	1.01m	1cm	1.03m	1cm	1.01m	1cm	1.00m	2cm
1.27m	1.35m	8cm	1.37m	10cm	1.37m	10cm	1.35m	8cm	1.33m	6cm
1.52m	1.49m	3cm	1.54m	2cm	1.53m	1cm	1.50m	2cm	1.50m	2cm
1.74m	1.75m	1cm	1.79m	5cm	1.80m	6cm	1.76m	2cm	1.75m	1cm
1.79m	1.82m	3cm	1.85m	6cm	1.88m	9cm	1.85m	6cm	1.81m	2cm
1.86m	1.85m	1cm	1.88m	2cm	1.90m	4cm	1.85m	1cm	1.85m	1cm
1.90m	1.92m	2cm	1.98m	8cm	1.99m	9cm	1.95m	5cm	1.92m	2cm
1.90m*	1.85m	5cm	1.91m	1cm	1.90m	0cm	1.88m	2cm	1.85m	5cm
2.05m	2.08m	3cm	2.12m	7cm	2.12m	7cm	2.08m	3cm	2.06m	1cm
2.20m	2.27m	7cm	2.29m	9cm	2.30m	10cm	2.24m	4cm	2.22m	2cm
2.22m	2.27m	5cm	2.29m	7cm	2.34m	12cm	2.27m	5cm	2.22m	0cm
2.49m	2.50m	1cm	2.59m	10cm	2.56m	7cm	2.50m	1cm	2.49m	0cm
2.61m	2.56m	5cm	2.65m	4cm	2.67m	6cm	2.59m	2cm	2.56m	5cm

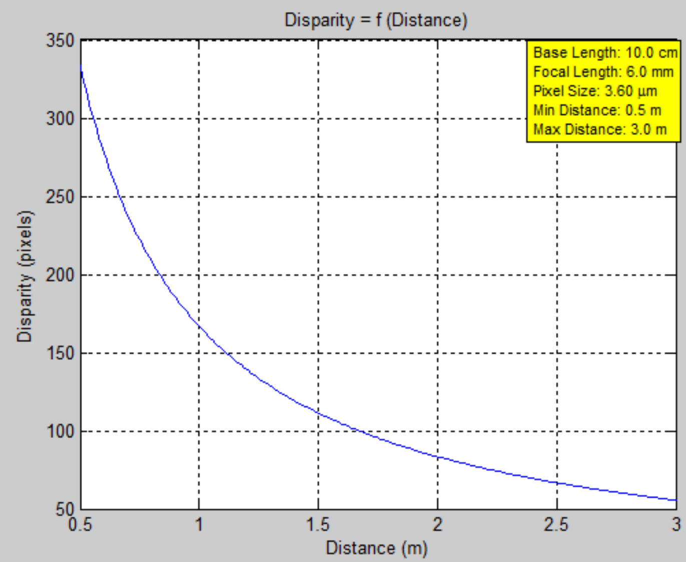


Depth Resolution Error (Kodak)









Conclusion and Future Work

- The accuracy of the investigated method for determining distance to given objects in a static scene by precise horizontal translation of one camera can be viewed from physical and algorithmic point:
  - For the physical accuracy improvement a subjective adjustments (like choice of optimal base and focal length, knowledge for optical distortions) need to be applied, in order to calibrate the stereovision system, according to the working distance range.
  - The algorithmic accuracy aspect depends on the software methods for determining the feature points (corners) and their correspondence.

- The future development of the investigated method can be focused on:
  - ✓ As much as possible points from the stereoimages to be viewed as a characteristic and their correspondence to be found in the other image;
  - ✓ Determining the 3D coordinates of a random characteristic point;
  - ✓ The distance determining algorithm to be improved to work with arbitrary translation and rotation between the cameras (or maybe a single camera?);
  - ✓ 3D Object Recognition;
  - ✓ 3D Video Stabilization.

THANK YOU FOR YOUR QUESTIONS

(before PART 2)

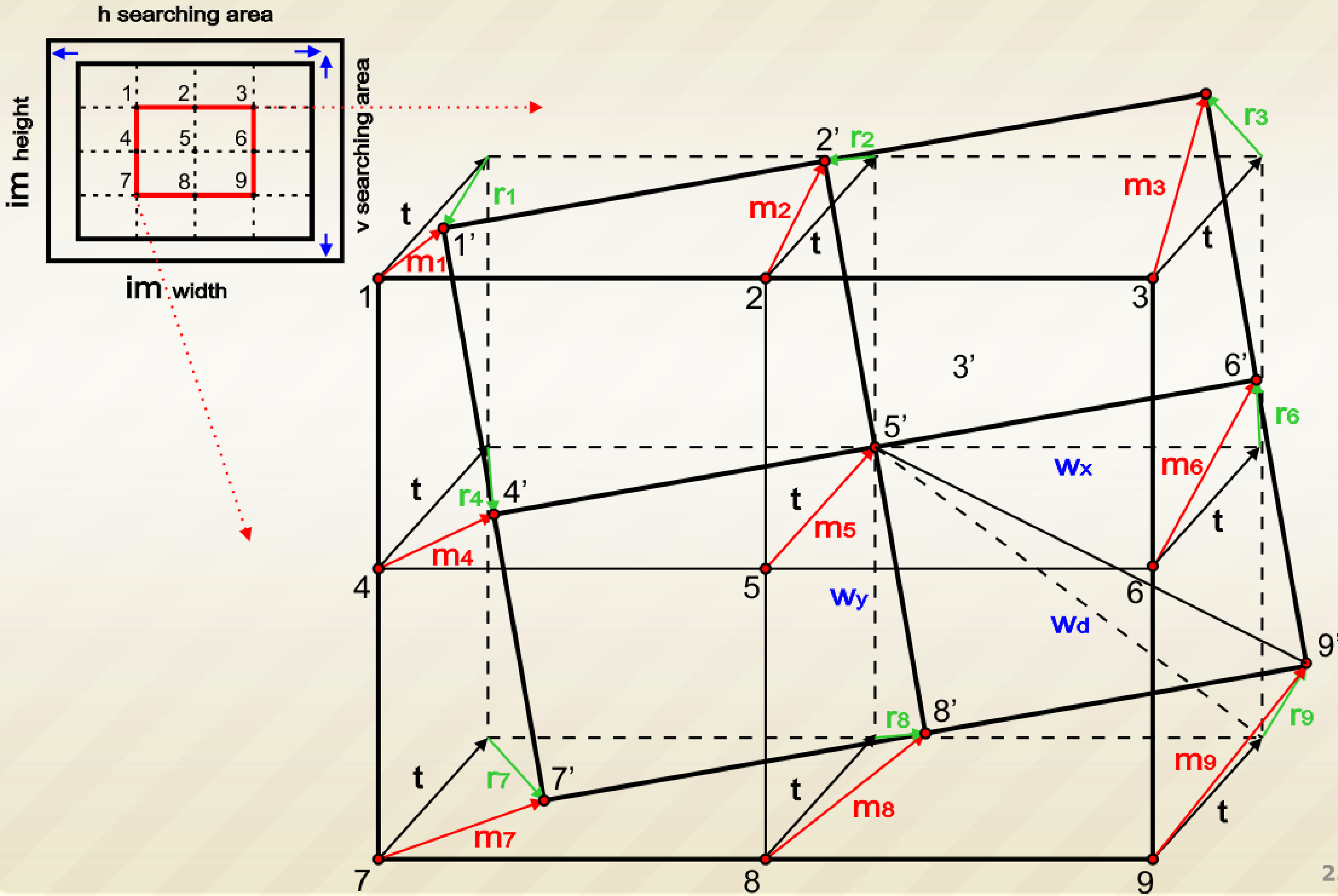
2D Real-Time Video Stabilization - Introduction

- Video stabilization seeks to create a stable version of casually shot video (usually filmed on a handheld device, such as a mobile phone or a portable camcorder) which is typically shaky and undirected.
- 2D video stabilization techniques work by estimating a 2D motion model between consecutive frames and applying per-frame warps between the original and filtered motion models.
- In case of approximately planer scenes (with an arbitrary camera movement) or cases where the camera shake is strictly rotational (within an arbitrary scene), unwanted jitters can be effectively reduced based on two-dimensional reasoning of the video.
- Assuming the scene geometry and camera motion do fall into these categories, such 2D stabilization methods are robust, operate on the entire frame, require a small number of tracked points and consume minimal computing efforts.
- This type of stabilization became very common in still and video cameras where it is implemented via mechanical means, either in the lens or the camera sensor.

An Algorithm for 2D Video Stabilization

1) Motion Vectors Model;
2) Estimating Horizontal and Vertical motion vectors for each block between Next and Previous frame by the method of *SAD*;
3) 1st Solving a Linear System of 18 equations with 4 unknowns by the *Least Square Method*;
4) Eliminating equations with values outside 1.5*standard deviation (const*sigma);
5) 2nd Solving a Linear System with less or equal number of equations with 4 unknowns by the *Least Square Method*;
6) Determining the interframe translational and rotational vectors based on the 2nd LSM;
7) Smoothing Global H/V/R Motion.

Motion Vectors Model

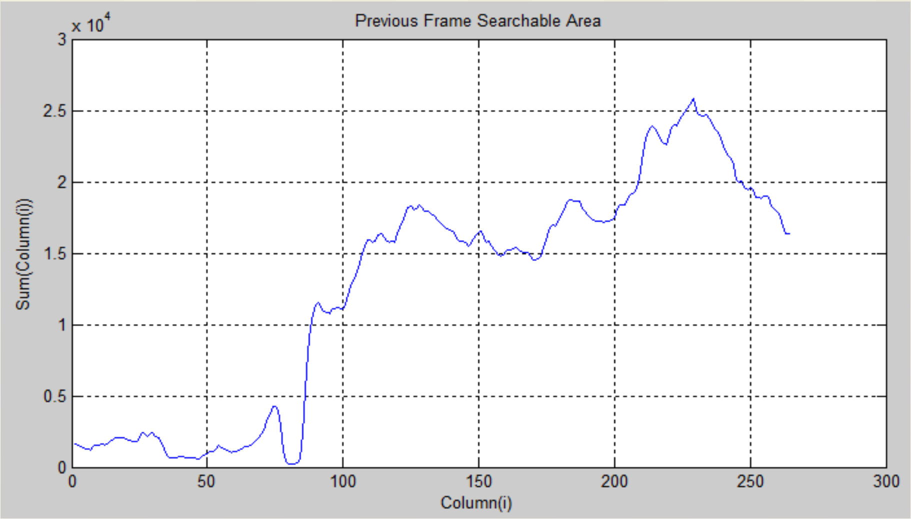


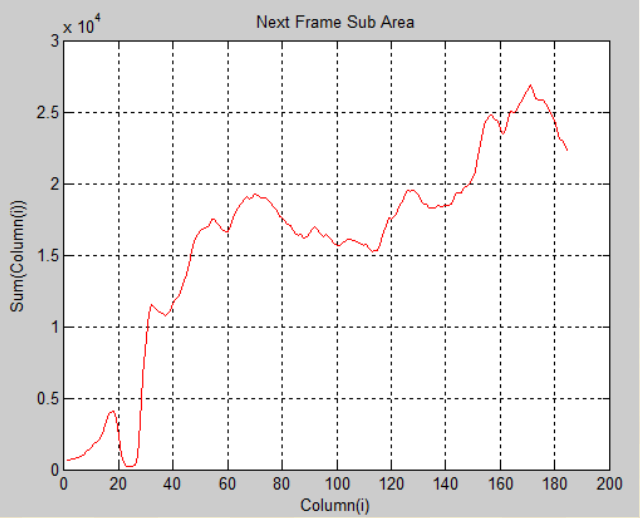
Example of Given Previous Frame

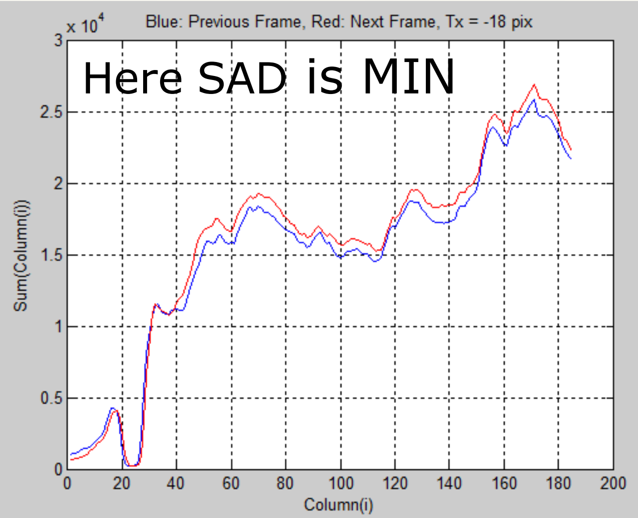
Example of Given Next Frame

Example of SAD between 1st areas

Horizontal Searching

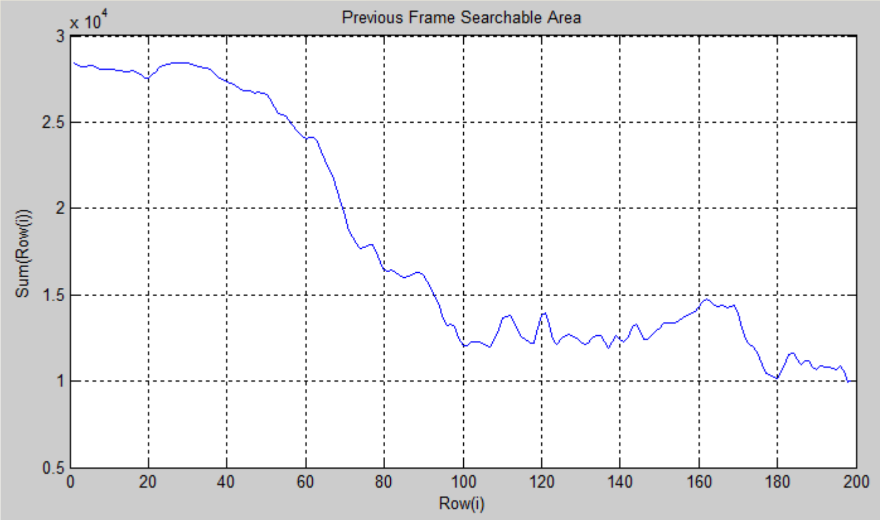


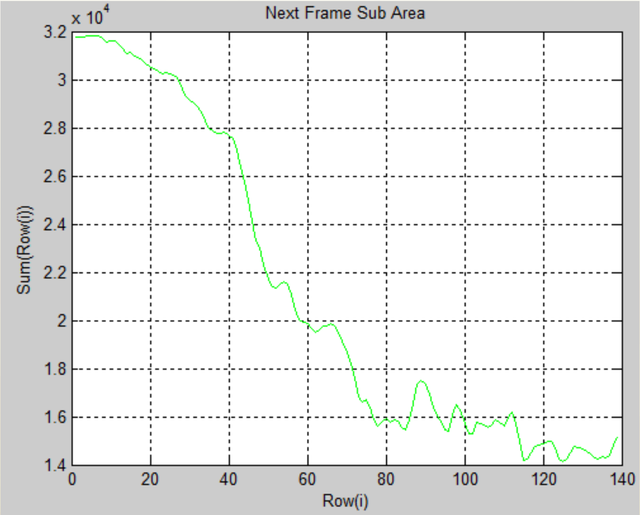


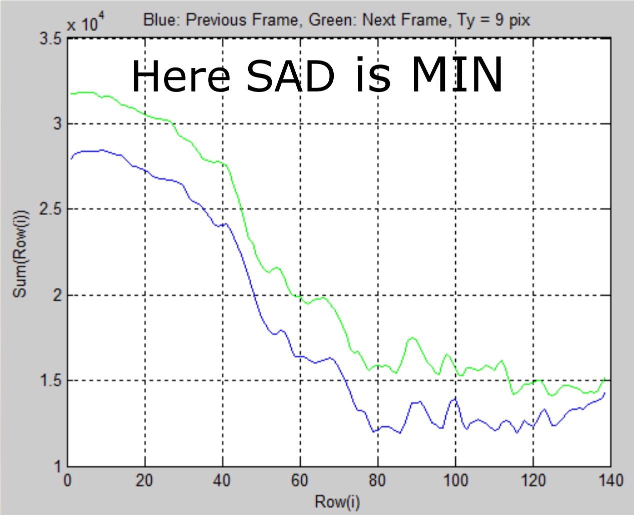


Example of SAD between 1st areas

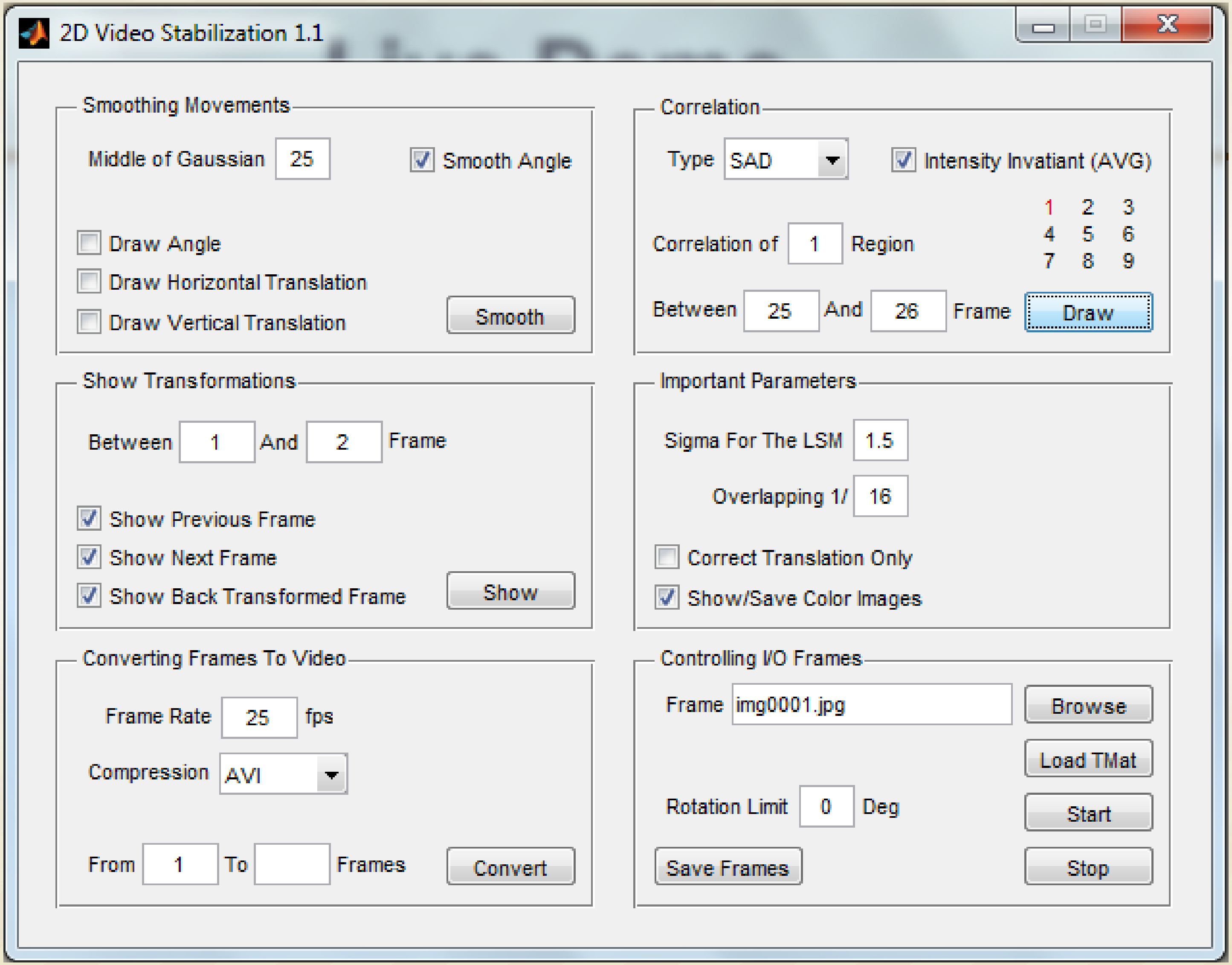
Vertical Searching







Live Demo



2D Video Stabilization 1.1
Smoothing Movements
Middle of Gaussian
25
Smooth Angle
Draw Angle
Draw Horizontal Translation
Draw Vertical Translation
Smooth
Correlation
Type
SAD
Intensity Invatiant (AVG)
1 2 3
4 5 6
7 8 9
Correlation of
1
Region
Between
25
And
26
Frame
Draw
Show Transformations
Between
1
And
2
Frame
Show Previous Frame
Show Next Frame
Show Back Transformed Frame
Show
Important Parameters
Sigma For The LSM
1.5
Overlapping 1/
16
Correct Translation Only
Show/Save Color Images
Converting Frames To Video
Frame Rate
25
fps
Compression
AVI
From
1
To
Frames
Convert
Controlling I/O Frames
Frame
img0001.jpg
Browse
Load TMat
Rotation Limit
0
Deg
Start
Save Frames
Stop

Conclusion

- Limitations of the 2D Video Stabilization:
  - In cases where the scenes contain objects at arbitrary depths, a full-frame warp cannot model the parallax that is induced by a translational shift in viewpoint and this level of scene modelling is insufficient for video stabilization.
  - The second limitation of 2D motion models is that there is no knowledge of the 3D trajectory of the input camera, making it impossible to simulate an idealized camera path similar to what can be found in professional tracking shots.

- The 3D Video Stabilization could overcome these limitations:
  - Using structure-from-motion technique (where a sparse set of feature points along the video are tracked and their correspondences are used) to recover the 3D camera pose and the 3D location of every feature point.
  - Estimating 2D feature trajectories from the input video (using the standard KLT approach), smooth them and synthesize new frames by video warping.
  - Other techniques use: Epipolar Geometry, L1 optimal camera path, etc.

THANK YOU