Video Geometry without Shape

Tamás Szirányi

MTA SZTAKI, Budapest

SSIP 2007, Szeged

Retrieving geometrical information in videos without any a priori information about the image structure or possible shapes:

- Registration of different views, or mirror, or shadows through co-motion statistics
- Focus-map through Bayesian iterations and a new error metric.

SSIP 2007, Szeged 2

Structure from conditional probabilities

Searching for statistical interaction among image points. This statistical information is given by

- Conditional/Concurrent Motion changes: camera registration, vanishing point of mirror, shadow, horizon
- Spatial coherence: relative focus depth; here conditional probabilities are given by the light distribution via the imaging system

- Lucy – Richardson Bayesian iteration schema is used three times here:
  - Co-motion statistics for common points of two cameras
  - Shadow modelling
  - Focus depth through blind deconvolution

SSIP 2007, Szeged 3

Co-motion: correlated motion in mirror and in stereo image pair

SSIP 2007, Szeged 4

The TASK for Stereo Wide Baseline Video Registration

- Given two or more views.
- Track objects across different views.

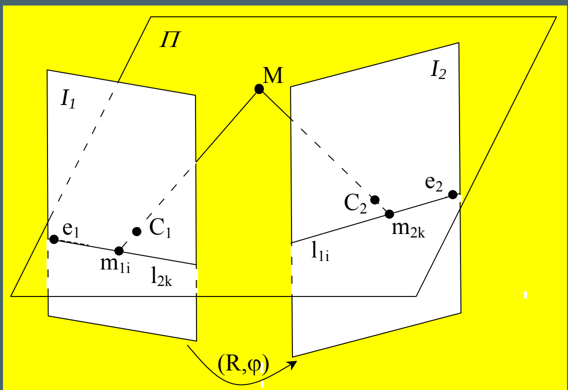
SSIP 2007, Szeged 5

General scheme

1. Background modeling.
2. Detection of features.
3. Extraction of point-correspondences – extraction of candidates, rejection of outliers.
4. Alignment of the cameras' views.

SSIP 2007, Szeged 6

The epipolar geometry for two cameras



Concurrently moving points for matching
Co-motion statistics

Motion statistics for a given pixel:

Detected objects | Motion map | Collected statistics

Co-motion statistics – Collecting statistics of co-motions simultaneously from two cameras/videos.

Motion, global motion statistics, co-motion statistics

$$m(t,\vec{x}) = \begin{cases} 1, & \text{where motion is detected} \\ 0, & \text{otherwise} \end{cases}$$

$$P_g(\vec{x}) = \frac{\sum_t m(t,\vec{x})}{\Delta t}$$

$$f(\vec{u},\vec{x}) = P_{co}(\vec{u}\,|\,\vec{x}) = \frac{1}{P_g(\vec{x})} \frac{\sum_t m(t,\vec{x})\, m(t,\vec{u})}{\Delta t}$$

Extracting candidate point pairs:
Co-motion detection in local and remote views

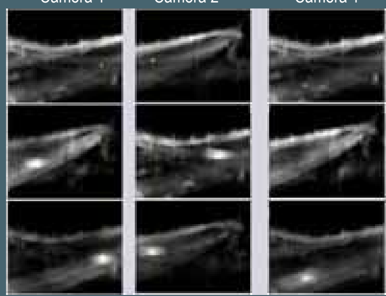
Inliers | Outlier

Camera 1 | Camera 2 | Camera 1

Overall motion statistics for Camera 1

Remote motion Statistics for the other camera considering a point

Local motion co-motion statistics for an image point



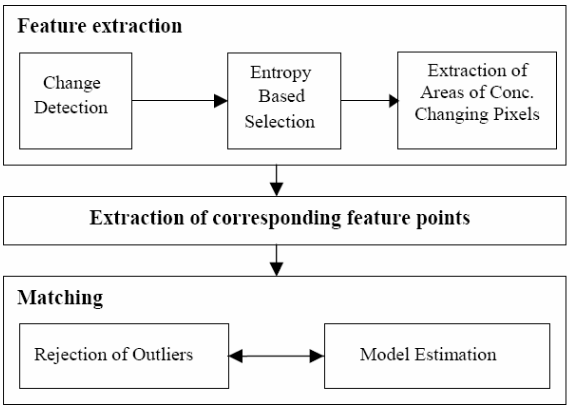
Local maximums on the created statistical images

Resolution is 80×60

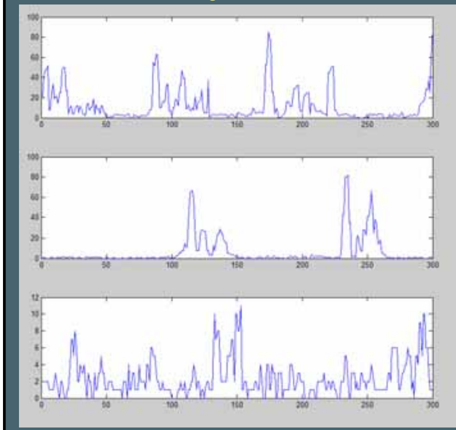
Composite view after optimized alignmenr by using RANSAC

Exhaustive search method needs huge amounts of memory for storing co-motion statistics for each pixel

Probabilistic model without shapes or tracks



Change histories

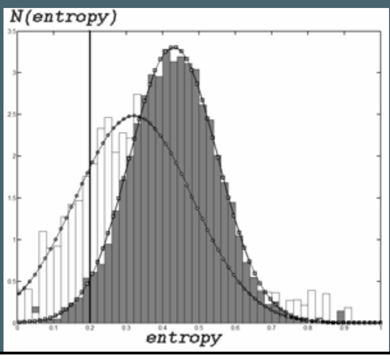


Continuous motion

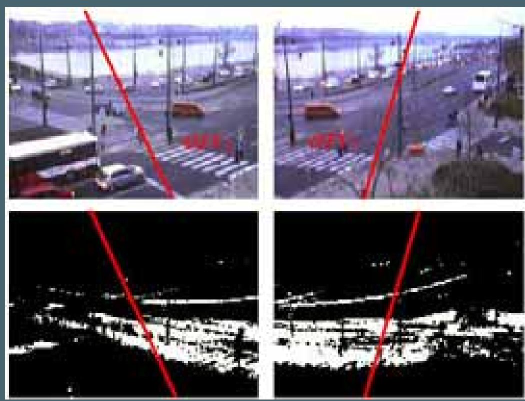
Motion with significant sudden changes

Noise flickerings (video noise)

- The distribution of entropy for noise flickerings (*boxes*) and deterministic motions (*circles*) in real life videos.
- The line shows a selected threshold at which the proportion of potential outliers is 15%.
- *Estimated Gaussian distributions*



Samples from input videos and Results of the *entropy* based preselection of feature points



Parameters of entropy distribution for different test videos. Last column shows the proportion of noise flickerings among pixels of detected changes if the threshold value is 0.2.

	Real motions		Noise flickerings		Proportion of noise flickerings
	exp. value	variance	exp. value	variance	
Video 1	0.23	0.14	0.43	0.14	12%
Video 2	0.26	0.11	0.41	0.13	15%
Video 3	0.31	0.05	0.47	0.1	19%
Video 4	0.28	0.16	0.48	0.16	13%

Probabilistic interaction among points of different views

for motion / no-motion functions

$$P(m_{1i} \mid m_{2k}) = \frac{1}{\sum_{t=1}^{T} \underline{b}_{2k}(t)} \sum_{t=1}^{T} \underline{b}_{1i}(t)\underline{b}_{2k}(t)$$

$$P(m_{1i} \mid m_{2k}) = \frac{P(m_{2k} \mid m_{1i})P(m_{1i})}{\sum_{j} P(m_{2k} \mid m_{1j})P(m_{1j})}$$

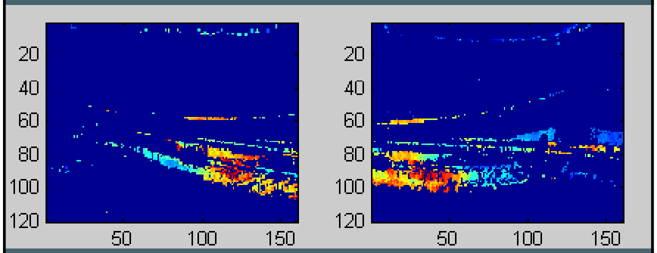
Ergodic regular Markov chain has a unique stationary distribution

$$\left(\underline{p}_1 \quad \underline{p}_2\right) = \left(\underline{p}_1 \quad \underline{p}_2\right)\underline{\underline{\Pi}}$$

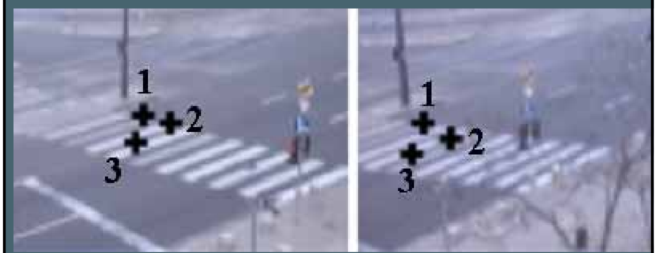
$$P(m_{1i})_{r+1} = P(m_{1i})_r \sum_k \frac{P(m_{2k} \mid m_{1i}) P(m_{2k})_r}{\sum_j P(m_{2k} \mid m_{1j}) P(m_{1j})_r}$$

$$P(m_{2k})_{r+1} = P(m_{2k})_r \sum_i \frac{P(m_{1i} \mid m_{2k}) P(m_{1i})_r}{\sum_j P(m_{1i} \mid m_{2j}) P(m_{2j})_r}$$

Bayesian iterations of Ergodic regular Markov chain with a unique stationary distribution



Sample point pairs obtained by Bayesian iterations. The nearly corresponding points are numbered with the same number.

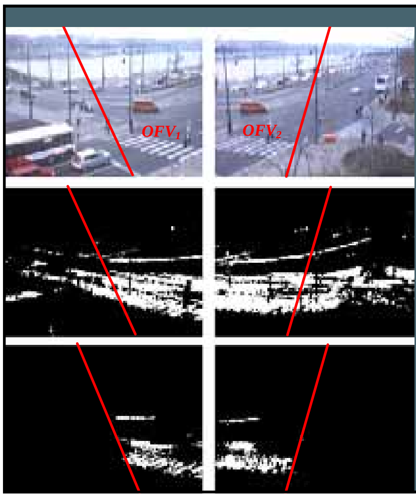


Short iteration length for the exceeding point sets

After four of the double iteration steps

the algorithm is stopped and those feature-points are selected for which

$P(m_{1i})$ and $P(m_{2k})$ are greater than $1/N_1$ and $1/N_2$.



Images show the resulting point sets of the feature extraction.

- Samples from input videos.
- Results of the *entropy* based preselection of feature points.
- Result of Bayesian estimation of OFV.

The change of the relative impact (in %) of feature points within the estimated OFV areas relative to the whole image before and after Bayesian iteration.

	Impact of feature points from OFV (%)	
	Before (17)	After (18)
video1	51	81
video2	59	78
video3	37	70
video4	32	75

The proportion of estimated OFV points to the real OFV points

after correlation based selection and

Bayesian iteration.

	Correlation	Bayesian
video1	71%	98%
video2	68%	99%
video3	73%	93%
video4	72%	90%

Reduction of ROI after feature extraction steps

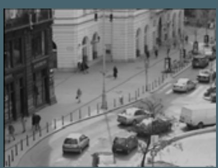
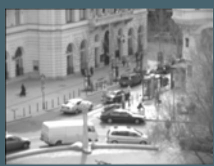
Feature extraction step	Size of ROI in pixels	
	I_1	I_2
Input	19200	19200
Entropy based preselection	2311	3253
Bayesian iteration	636	671

Co-motion statistics

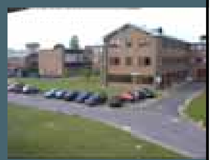
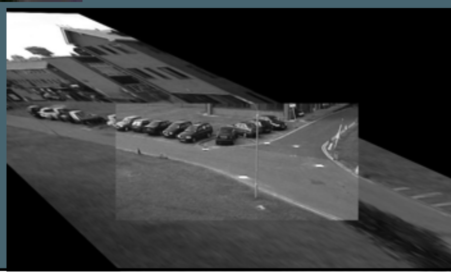
GELLÉRT video

- Same cameras
- Same zoom
- 160×120 resolution, 10 fps

Grounplane fitting of two views

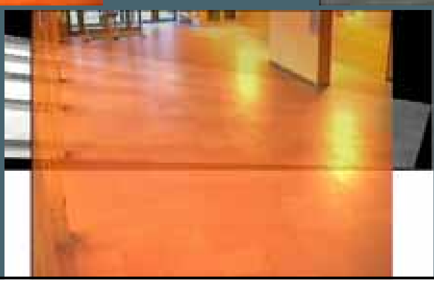




Final alignment of two views' ground planes for PETS2001

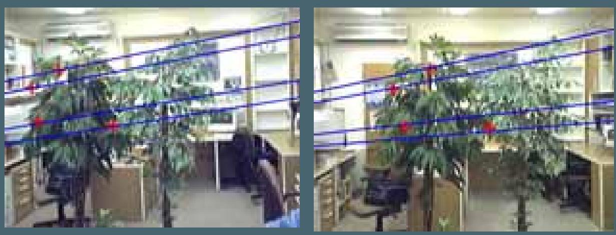
Indoor ground planes: Alignment of two views' detecting concurrently moving shadows





Random motion for registration


Sample point-pairs with corresponding epipolar lines obtained for random motion videos



Numerical results of model fitting for different experiments

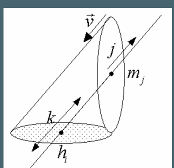
Experiment	Average Error	Min Error	Model
Gellert	5.40	0.16	H
Ferenciek	6.54	0.44	H
PETS2001	4.18	0.88	H
Indoor	9.00	0.63	H
LAB	6.15	0.16	F
LAB	21.22	12.30	H

Shadow detection with an iteration scheme

Based on this formula the following iteration scheme can be written (for shadow pixels):

$$P_{k+1}(h_i) = P_k(h_i) \sum_j \frac{P(m_j|h_i)\,P(m_j)}{\sum_k P(m_j|h_k)\,P(h_k)}, \quad i, j, k \in S$$

The key issue in the formula is the determination of P(m|h) conditional probability. According to the geometrical model the computation can be summarized as follows:



There is uniform initialization value along the line. The j and k indices demonstrate the cycles only.

Results of iterative shadow process

Advantages:

- well defined formulas
- flexible for further parameters
- can handle geometrical information

Disadvantages:

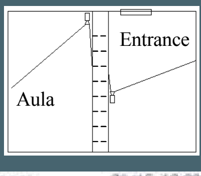
- heavy computation time
- there are problematic situations
- on-line estimation of light direction is needed

- It is a *complementary* method with others.

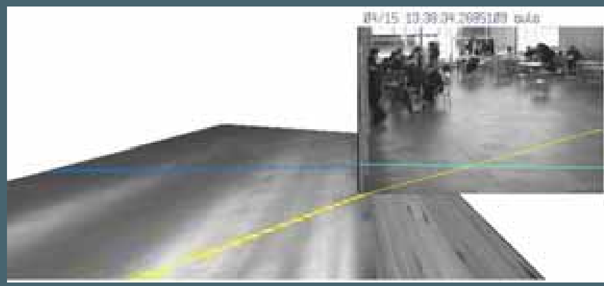
Indoor sample video

Detection of human walking

Non-overlapping views – detecting walkers' leg



Alignment non-overlapping views



More about Co-motion registration for two views

- Z. Szlávik, T. Szirányi, L. Havasi:

”Stochastic view registration of overlapping cameras based on arbitrary motion”, **IEEE Tr. Image Processing**, March, 2007

- Z. Szlávik, T. Szirányi, L. Havasi:

”Video camera registration using accumulated co-motion maps”, **ISPRS J Photogrammetry and Remote Sensing**, January, 2007

- L. Havasi , Z. Szlávik, T. Szirányi:

”Detection of Gait Characteristics for Scene Registration in Video Surveillance System”, **IEEE Tr. Image Processing**, February, 2007

Vanishing Point in case of Mirror: Original object and its reflection are presented

Havasi, L., Szirányi, T.:
Estimation of Vanishing Point in Camera-Mirror Scenes Using Video, **Optics Letters** (2006)

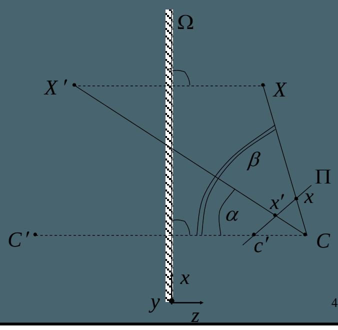
Use of Co-motion statistics for vanishing point estimation in camera-mirror scenes and in case of cast shadow

Vanishing point describes a skew symmetric fundamental matrix between two views:

A reflective surface (e.g. a mirror), denoted by Ω, which lies in the x-y plane (right-handed system).

C denotes the camera center, and the image plane is denoted by Π

(3-D points are mapped to this plane via central projection).



Fundamental constraint

- *The fundamental matrix corresponds to the original image and the virtual image in a camera-mirror scene. Consequently, F has 2 degrees of freedom and is identified with the VP.*

$$\tilde{\mathbf{x}}_1^T F \tilde{\mathbf{x}}_2 = \tilde{\mathbf{x}}_1^T \left(\tilde{\mathbf{c}}(F) \times \tilde{\mathbf{x}}_2 \right) = \left\langle \tilde{\mathbf{x}}_1, \tilde{\mathbf{c}}(F) \times \tilde{\mathbf{x}}_2 \right\rangle = 0$$

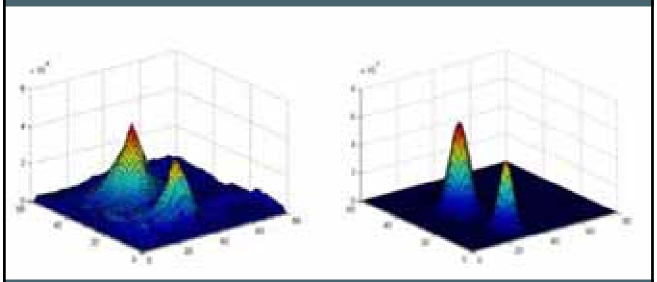
$$F = \begin{bmatrix} 0 & -1 & c_2' \\ 1 & 0 & -c_1' \\ -c_2' & c_1' & 0 \end{bmatrix}$$

Co-motion statistics in camera-mirror scenes

$$P_{co}(\vec{u}\,|\,\vec{x}) \approx w_1^{\vec{x}} P_{near}(\vec{u}\,|\,\vec{x}) + w_2^{\vec{x}} P_{coll}(\vec{u}\,|\,\vec{x})$$
$$= \sum_{i=1}^{2} w_i^{\vec{x}} \mathrm{N}(\vec{u}, \vec{\mu}_i^{\vec{x}}, \Sigma_i^{\vec{x}}), \text{ where } w_1^{\vec{x}} + w_2^{\vec{x}} = 1$$
$$\text{and } \left\|\vec{\mu}_1^{\vec{x}} - \vec{x}\right\| < \left\|\vec{\mu}_2^{\vec{x}} - \vec{x}\right\|$$

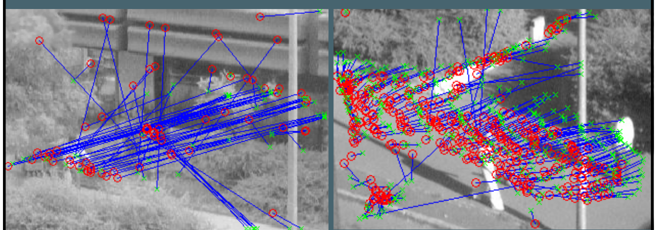
In case of a visible reflective surface two peaks are probable, thus the pdf is modeled with a simple Gaussian mixture model with two components:

Co-motion statistics and its Gaussian mixture model estimation



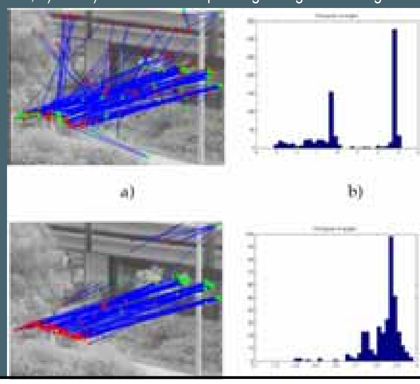
Corresponding point pairs

Depending on the scene configuration, not every moving point will have a visible reflection.



Rejection of outliers for the “Shop” sequence.

Only the directions corresponding to the main peak (mode) of the histogram (determined from the line directions) will be used for later computations. a) before rejection, c) after rejection; b) and d) show the corresponding histograms of angles.

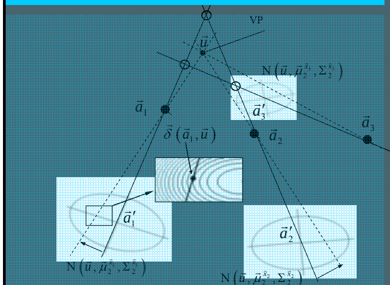


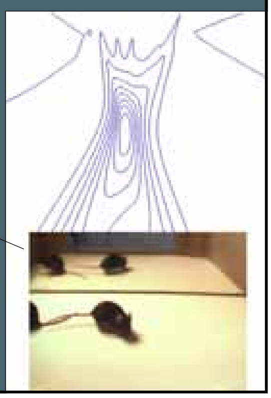
The determination of VP is carried out by using an objective function

Goodness-of-fit function represented using a contour graph; the VP is marked with point

$$\vec{\delta}(\mathbf{x},\mathbf{u}) = \arg\max_{\mathbf{v}\in S_2} P_{coll}(\mathbf{v}|\mathbf{x}) \text{ and } \left\langle \tilde{\boldsymbol{\mu}}_{near}^{\mathbf{x}} \times \tilde{\mathbf{u}}, \tilde{\mathbf{v}} \right\rangle = 0$$

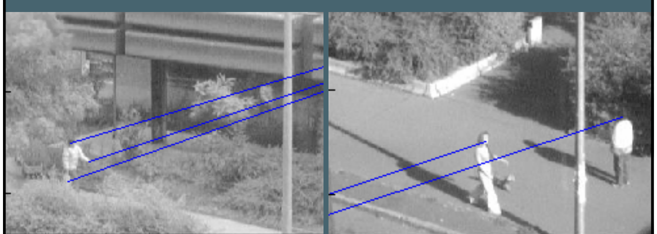
$$VP = \mathbf{c}' = \arg\max_{\mathbf{u}} \sum_{\mathbf{x}\in S_2} P_g(\mathbf{x}) P_{coll}\left(\vec{\delta}(\mathbf{x},\mathbf{u})|\mathbf{x}\right)$$





Experimental results on VP estimation

The results demonstrate the collinearity constraint



Co-motion summary

We have shown that cameras can be registered in several rather miserable conditions, based on:

- Unpredictable motion without structured background or defined object shapes or
- Shadows of undefined structures in front of flickering background.
- Detection of Vanishing Pont from arbitrary motion in case of mirror or shadow

It can joint to detection and search for video events of multicamera systems

Relative Focus Area Extraction by Blind Deconvolution for Defining Regions of Interest

Focus estimation with deconvolution

- Blind deconvolution:
  - Given observation g, give an estimation of the original image f and the blurring function (PSF) h : $g = f * h$
  - Starting from Richardson's original formula based on Bayesians:

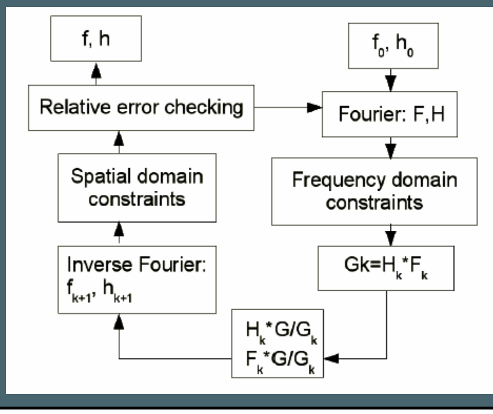
$$P(f_i|g) = [P(g|f)P(f_i)] \,/\, \sum_j [P(g|f_j)P(f_j)]$$

$$P(f|g) = P(fg)/P(g)$$

$$P(f_i) = \sum_l P(f_i g_l) = \sum_l P(f_i|g_l)P(g_l) = \sum_l \frac{P(g_l|f_i)P(f_i)P(g_l)}{\sum_j P(g_l|f_j)P(f_j)}$$

$$P_{k+1}(f_i) = P_k(f_i) \sum_l \frac{P(g_l|f_i)P(g_l)}{\sum_j P(g_l|f_j)P_k(f_j)}$$

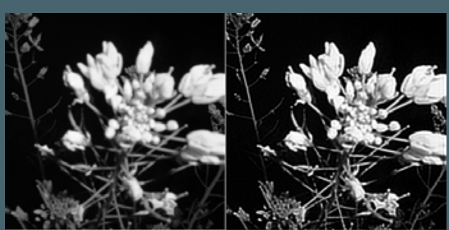
Focus Area Extraction by Blind Deconvolution for Defining Regions of Interest



Blind deconvolution with Lucy - Richardson double irerations

$$f_{i,k+1} = f_{i,k} \sum_l \frac{h_{i,l} g_l}{\sum_j h_{j,l} f_{j,k}} = f_{i,k} \sum_l h_{i,l} \frac{g_l}{\sum_j f_{j,k} h_{j,l}}$$

$$f_{k+1} = f_k \left(h_k \otimes \frac{g}{f_k \otimes h_k} \right)$$



The double iteration

- We create a localized double iteration scheme for locally varying f and PSF estimation (r – location vector):

$$\begin{cases} f_{k+1}(r) = f_k(r) \left[h_k(r) * \frac{g}{g_k}(r) \right] \\ h_{k+1}(r) = \frac{h_k(r)}{\gamma} \left[f_k(r) * \frac{g}{g_k}(r) \right] \end{cases}$$

Automatic rel. focus map extraction

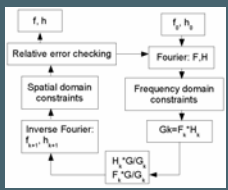
- From the classic Richardson iterative blind deconvolution formula

$$g(x) = h(x) * f(x)$$

$$P_{k+1}(f_i) = P_k(f_i) \sum_i \frac{P(g_i|f_i)P(g_i)}{\sum_j P(g_i|f_{k,j})P(f_{k,j})}$$

$$f_{k+1} = f_k \left[h_k * \frac{g}{f_k * h_k} \right] = f_k \left[h_k * \frac{g}{g_k} \right]$$

g – observed
h – PSF
f – unknown original



- *f* and the PSF vary locally according to the amount of blur (distortion) present on the image locally
- Stop the double iteration at a finite step (here #5) and check the error between the measured and the estimated blurred image blocks: $\| g - g_k \|$
- Is MSE usable for comparison the BD residual errors of different blocks?

Constraints and ill-posedness

- In the local deconvolution we consider only a few constraints
  - symmetricity,
  - non-negativity,
  - zero phase.
  - and nothing about the image content regularization (e.g. edges). Localized deconvolution runs on small blocks, range of the PSF. Thus the ill-posed iteration process tends to be noisy.
- For the classification we stop at a low iteration count and we need a stable error measure which gives different values for differently focused areas, and which is not much affected by the process's noisy nature.

ADE : angle deviation error

Orthogonality criterion: signal and noise are independent

$$\left| \arc\sin \frac{\langle g, g - g_k \rangle}{|g| \, |g - g_k|} \right|$$

In case of

$g - g_k = [\,+1, -1, -1, +1, -1, +1 \ldots\ldots -1, +1\,]$

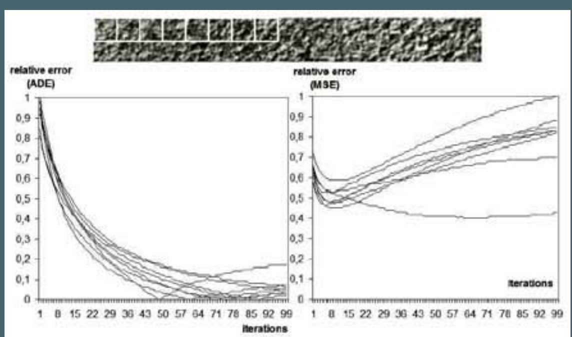
$g = [10, 10, 10, 10, 10, 10 \ldots\ldots 10, 10]$

$\| g - g_k \|$ is high, while

$< g, g - g_k > \;\longrightarrow$ zero

Error curves for 8 neighboring blocks (each curve stands for one block) on a blurred texture sample (top) for the same blur with **ADE** (left), and **MSE** (right) .

Ideally, curves of the same measure should remain close to each other.



The error function

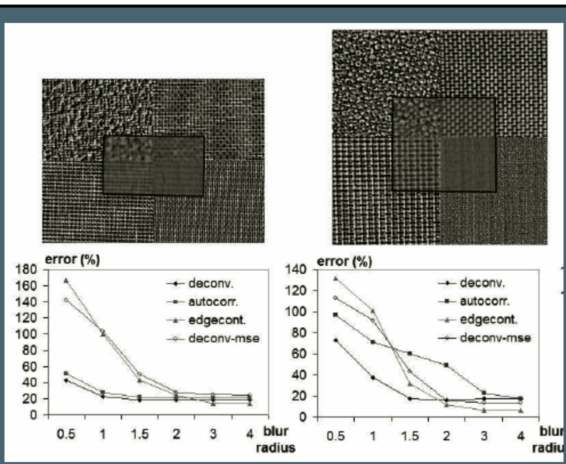
- Localised blind deconvolution for focus map estimation:
  - run local deconvolution with a low iteration count
  - calculate local residual errors, with contrast weighting

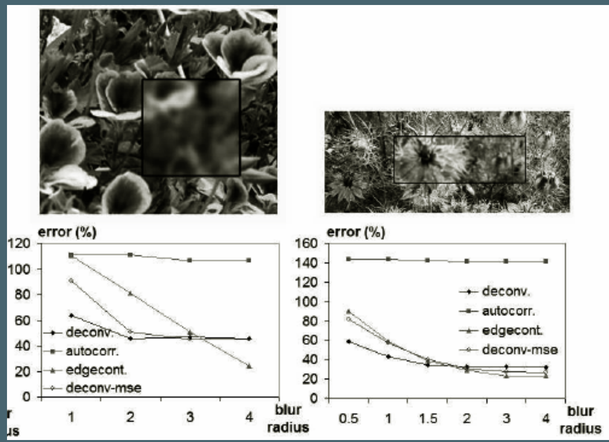
$$E_r(g, g_k) = arc\,sin \frac{< g - g_k, g >}{|g - g_k| \cdot |g|} \cdot \frac{C_r(g_r)}{max_r \{C_r(g_r)\}}$$

$$C_r(g_r) = \frac{g_{max\{x \in T_r\}} - g_{min\{x \in T_r\}}}{g_{max\{x \in T_r\}} + g_{min\{x \in T_r\}}}$$

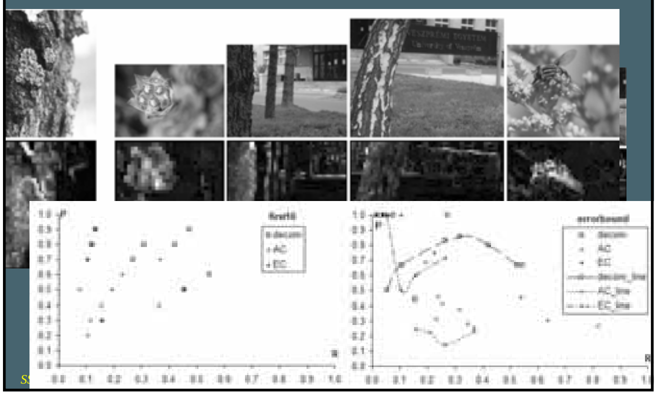
  - use the local residuals for relative classification of areas

$$F(r) = \frac{c \cdot (E_r(g, g_k) - min\{E_r(\cdot, \cdot)\})}{max\{E_r(\cdot, \cdot)\} - min\{E_r(\cdot, \cdot)\}}$$





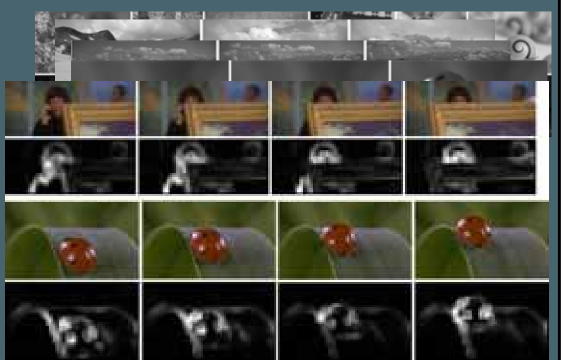
Find images with similar relative focused objects:

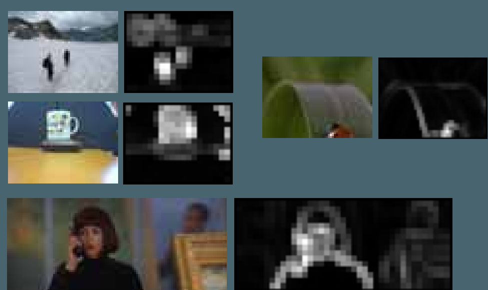


L. Kovács, T. Szirányi:

- “Relative Focus Map Estimation Using Blind Deconvolution”, Optics Letters, 2005
- “Image Indexing by Focus Map” Lecture Notes in Computer Science, 2005
- Focus Area Extraction by Blind Deconvolution for Defining Regions of Interest IEEE T Pattern Anal. Mach. Int. (June 2007)

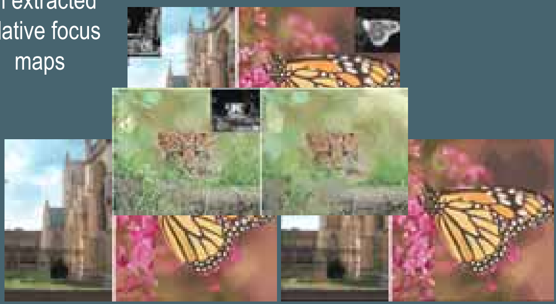
Image / Video indexing matching sample image or semantic description





Application in Non Photorealistic Rendering

Painting detail variation based on extracted relative focus maps



A demostration on

A. Licsár, T. Szirányi: User-adaptive hand gesture recognition system with interactive training" , *Image and Vision Computing*, 2005 and *ACM MM WS,* 2006 Oct.

Tillarom: an AJAX Based Folk Song Search and Retrieval System with Gesture Interface Based on Kodály Hand Signs

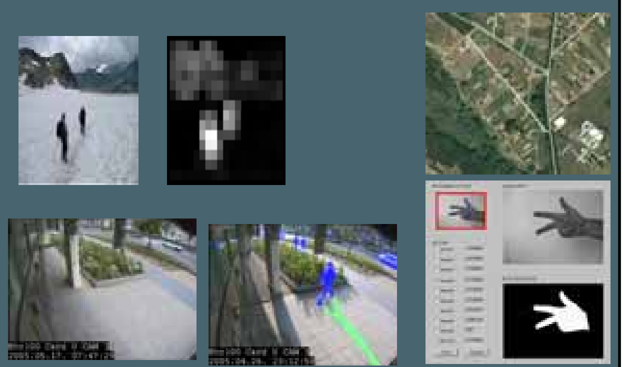
Thank you for your attention!

Distributed Events Analysis (DEVA) Research Group

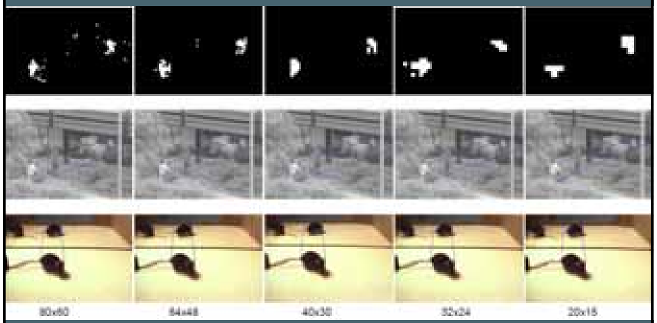
sztaki.hu/department/EEE/

Distributed Events Analysis (DEVA) Research Group

http://www.sztaki.hu/~sziranyi/DOC/eee/

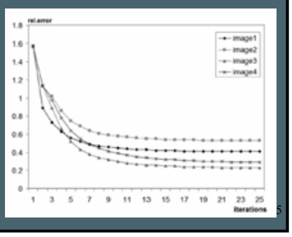


Results for different scales

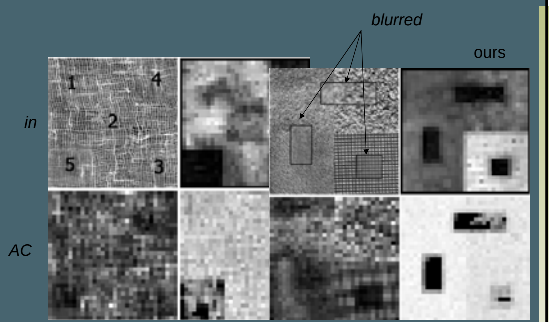


Sequence	#Point	Vanishing point		
		Initial [3]	Optimized	True
Shop	790	104,165	4881, -1272 (14.6°)	4500,-1300 (16.1°)
Shadow	3509	-13, 53	-1918,680 (199.5°)	-2110,850 (201.9°)

- Constraints:
  - In the space domain:
    - size of PSF (region of support constraint): spurious elements of h_k outside the initial region of support will be zeroed during the iteration;
    - pixel amplitude bound $0<=f_k<=255$;
    - non-negativity for the PSF (h_k) and the image (f_k).
  - In the frequency domain:
    - zero phase of h_k, so as not to induce phase distortion when filtering the image f_k



- Rel. focus map extraction - comparison on textures



Apps: feature extraction

Find images with similar relative focused objects:

Follow the path of relative focus changes: