Novel Markovian Change Detection Models in Computer Vision

Csaba Benedek

Scientific adviser: Professor Tamás Szirányi, D. Sc

Pázmány Péter Catholic University
Faculty of Information Technology

Computer and Automation Research Institute
Hungarian Academy of Sciences

Public defense of *Ph.D.* dissertation, 30 June 2008

Content

Content

Introduction

- Image/video based computer systems - digital visual information streams
  - Video surveillance for police
  - Cartography and remote sensing - aerial image analysis
- Change detection - goals
  - Decreasing the number of interesting photos or video frames
  - Extracting object descriptors for higher level image processing modules

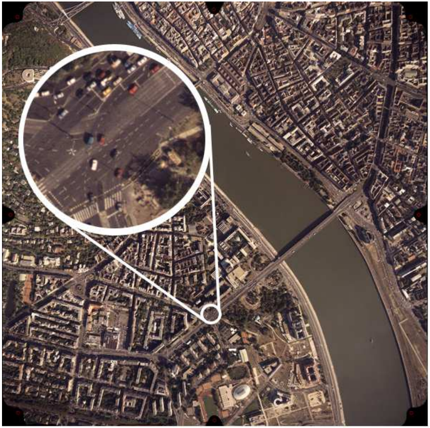
Video surveillance

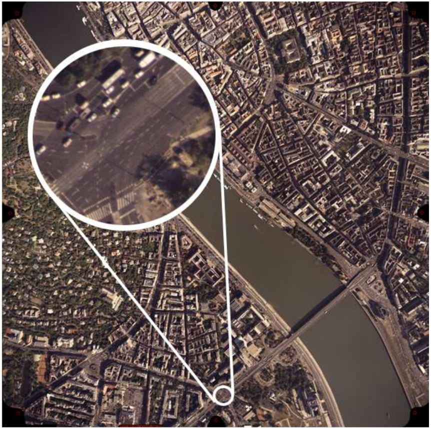
Remote sensing

Foreground and Shadow Detection in Video Sequences

Object Motion Detection in Image Pairs Taken by Moving Airborne Vehicles...

- Stereo reconstruction of static scenes

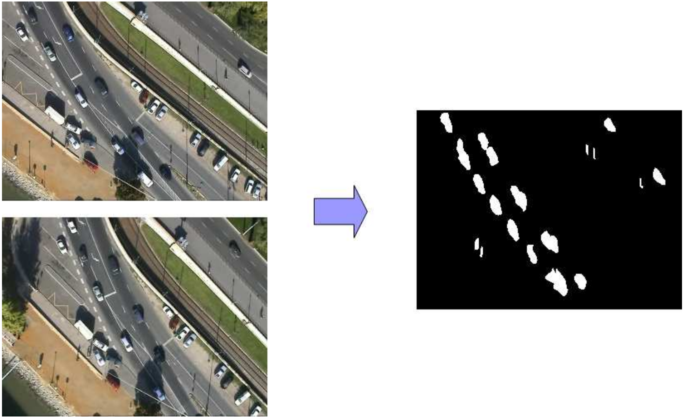




... and Processing Low Frame-Rate Aerial Videos

- Large and unpredictable camera motion
- Low frame-rate
- Frame differencing instead of video based techniques

... and Processing Low Frame-Rate Aerial Videos



Detecting Built-in Changes in Image Pairs Taken with Large Time Differences

- Comparing registered photos (Institute of Geoscience, Cartography and Remote Sensing)

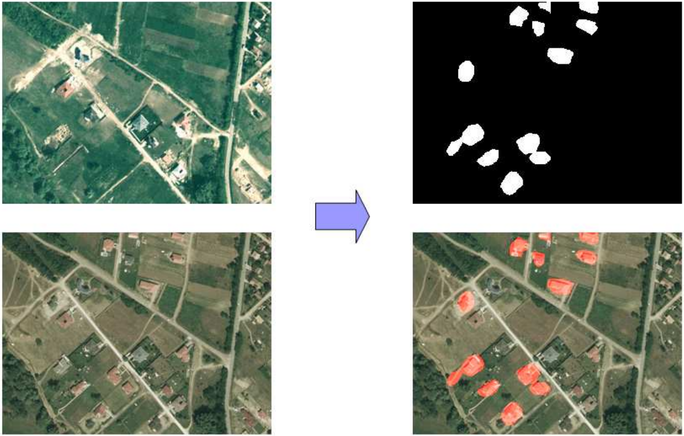
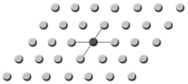


Image Segmentation with Markov Random Fields

- 2-D pixel lattice → graph: $S = \{s\}$
  - nodes: image points (s is a pixel)
  - edges: interactions → cliques

Lattice S

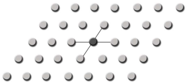


- Goal: generate a K-colored segmented image, with a task dependent label set $L = \{C_1, \ldots, C_K\}$
  - $\omega_s \in L$: label of pixel s which mark its segmentation class
  - Task 1: $K = 3$; C_1=foreground, C_2=background and C_3=shadow.
- Segmentation with Markov Random Fields (MRF):
  - f_s: local feature observed at pixel s (color, texture etc.)
    - Pixels' feature-values should fit the class models specified by their label
    - Classes are described by feature distributions or probability density functions e.g. $P(f_s|\omega_s = \text{background})$.
  - Segmented image is "smooth": we penalize, if two neighboring pixels have different labels

Image Segmentation with Markov Random Fields

- 2-D pixel lattice $\rightarrow$ graph: $S = \{s\}$
  - nodes: image points (s is a pixel)
  - edges: interactions $\rightarrow$ cliques

Lattice S

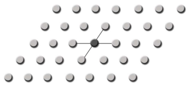


- Goal: generate a K-colored segmented image, with a task dependent label set $L = \{C_1, \ldots, C_K\}$
  - $\omega_s \in L$: label of pixel s which mark its segmentation class
  - Task 1: $K = 3$; C_1=foreground, C_2=background and C_3=shadow.
- Segmentation with Markov Random Fields (MRF):
  - f_s: local feature observed at pixel s (color, texture etc.)
    - Pixels' feature-values should fit the class models specified by their label
    - Classes are described by feature distributions or probability density functions e.g. $P(f_s|\omega_s = \text{background})$.
  - Segmented image is "smooth": we penalize, if two neighboring pixels have different labels

Image Segmentation with Markov Random Fields

- 2-D pixel lattice $\rightarrow$ graph: $S = \{s\}$
  - nodes: image points (s is a pixel)
  - edges: interactions $\rightarrow$ cliques

Lattice S



- Goal: generate a K-colored segmented image, with a task dependent label set $L = \{C_1, \ldots, C_K\}$
  - $\omega_s \in L$: label of pixel s which mark its segmentation class
  - Task 1: $K = 3$; C_1=foreground, C_2=background and C_3=shadow.
- Segmentation with Markov Random Fields (MRF):
  - f_s: local feature observed at pixel s (color, texture etc.)
    - Pixels' feature-values should fit the class models specified by their label
    - Classes are described by feature distributions or probability density functions e.g. $P(f_s|\omega_s = \text{background})$.
  - Segmented image is "smooth": we penalize, if two neighboring pixels have different labels

Image Segmentation with Markov Random Fields

- Global labeling: $\underline{\omega} = \{\omega_s | s \in S\}\}$
- Observation process: $\mathcal{F} = \{f_s | s \in S\}$
- MAP estimation of the optimal global labeling:

$$\widehat{\underline{\omega}} = \mathrm{argmax}_{\underline{\omega} \in \Omega} P(\underline{\omega}|\mathcal{F})$$

  where Ω denotes the set of all the possible global labelings.
- **(Hammersley-Clifford theorem)**: $P(\underline{\omega}|\mathcal{F})$ can be factorized into individual terms whose domains are the cliques of the graph.

$$P(\underline{\omega}|\mathcal{F}) \propto \underbrace{\prod_{s \in S} P(f_s|\omega_s)}_{P(\mathcal{F}|\underline{\omega})} \cdot \underbrace{\frac{1}{Z} \exp\left(- \sum_{C \in \mathcal{C}} V_C(\underline{\omega}) \right)}_{P(\underline{\omega})}$$

  - where C is an arbitrary clique and V_C is the potential of C.
  - MRF energy function: $-\log P(\underline{\omega}|\mathcal{F})$

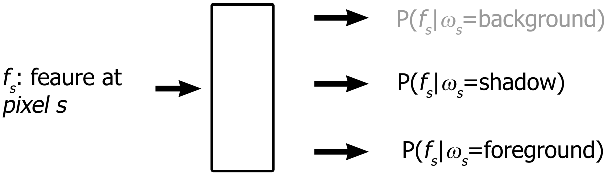
Content

Thesis Group 1: Bayesian Foreground and Shadow Detection in Video Scenes

- I have worked out a novel spatio-temporal probabilistic model based on MRF for foreground - background separation and cast shadow detection in video frames. I have experimentally shown that the proposed method outperforms the recently published models with the same goals and scene assumptions.

Bayesian Foreground and Shadow Detection in Video Scenes

- *Likelihood* model of pixel s:



- Field energy:

$$\sum_{s\in S} -\log P\left(f_s \mid \omega_s\right) + \sum_{r,s\in\mathcal{C}} \Theta\left(\omega_r, \omega_s\right)$$

$$\Theta(\omega_r, \omega_s) = \begin{cases} -\delta & \text{if } \omega_r = \omega_s \\ +\delta & \text{if } \omega_r \neq \omega_s \end{cases}$$

Thesis 1.1: Shadow Model

- I have proposed a novel statistical and adaptive color model for detecting cast shadows. I have shown that the procedure is more efficient than using previous approaches if the scene reflection properties are not ideally Lambertian.
  - Photometrical description of the measured color as a function of illumination $e(\lambda, s)$

$$g(s) = \int e(\lambda, s)\rho(\lambda, s)\nu(\lambda)d\lambda$$

  - Former similar models: simplifying assumptions
    - uniform illumination
    - purely Lambertian reflecting surfaces
    - decreased performance in complex scenarii
  - Proposed approach: global statistical characterization of pixel level, physical shadow descriptors

Thesis 1.1: Shadow Model

- I have proposed a novel statistical and adaptive color model for detecting cast shadows. I have shown that the procedure is more efficient than using previous approaches if the scene reflection properties are not ideally Lambertian.
  - Photometrical description of the measured color as a function of illumination $e(\lambda, s)$

$$g(s) = \int e(\lambda, s)\rho(\lambda, s)\nu(\lambda)d\lambda$$

  - Former similar models: simplifying assumptions
    - uniform illumination
    - purely Lambertian reflecting surfaces
    - decreased performance in complex scenarii
  - Proposed approach: global statistical characterization of pixel level, physical shadow descriptors

Foundations of the Shadow Model

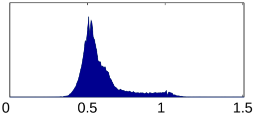
- Constant ratio model (noise sensitive):

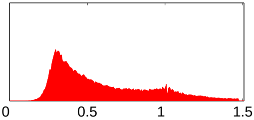
$$g_{\text{shadow}}(s) = A \cdot g_{\text{background}}(s)$$

- Proposed approach (1D visualization):

$$\psi(s) = g(s)/g_{\text{background}}(s)$$

  - spatiotemporal histograms of shadow- and foreground $\psi(s)$ values:

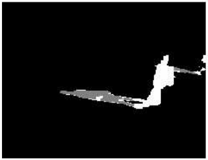




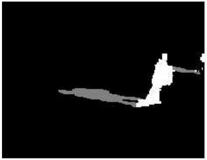
- Approximating the shadow domain: Gaussian density functions
- Color images: $\overline{\psi}(s)$ $3D$ vector, $3D$ Gaussian density

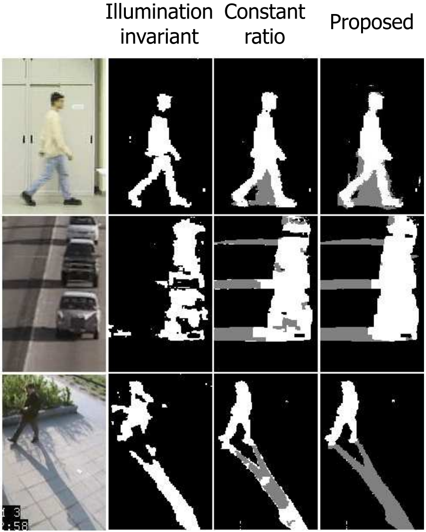
Experimental Validation of the Shadow Model

Constant ratio:



Proposed:





Thesis 1.2: Foreground Model

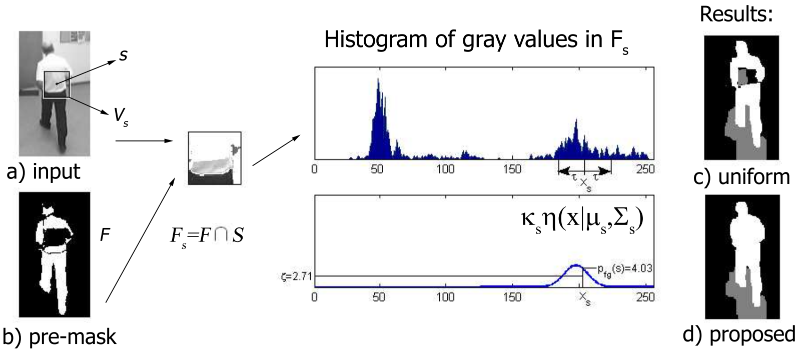
- A novel foreground description has been given based on spatial statistics of the nearby pixel values. I have shown that the introduced approach enhances the detection of background or shadow-colored object parts, even in low and/or unsteady frame rate videos.
  - Predicting the colors in the foreground
    - irrelevant temporal statistics
    - uniform color model - weak to detect fine differences
    - spatial statistics: in the neighborhood of a foreground pixel other foreground pixels are expected with similar color

Thesis 1.2: Foreground Model

- A novel foreground description has been given based on spatial statistics of the nearby pixel values. I have shown that the introduced approach enhances the detection of background or shadow-colored object parts, even in low and/or unsteady frame rate videos.
    - Predicting the colors in the foreground
        - irrelevant temporal statistics
        - uniform color model - weak to detect fine differences
        - spatial statistics: in the neighborhood of a foreground pixel other foreground pixels are expected with similar color

Introduction of the Foreground Model

- Estimation of the color statistics of the probably foreground pixels in each neighborhood

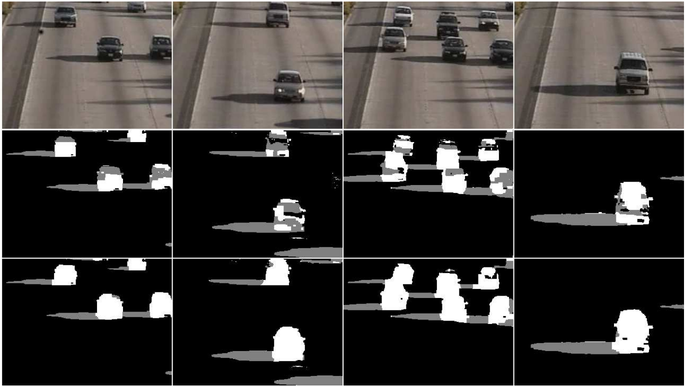


Introduction of the Foreground Model

- Example results:

Uniform foreground model

Proposed foreground model



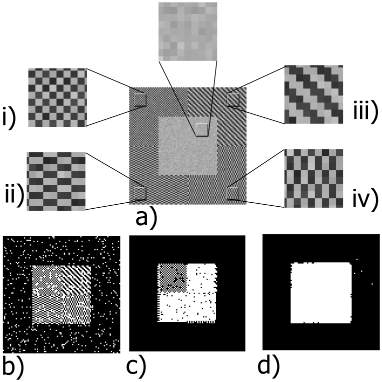
Thesis 1.3: Microstructure Model

- I have given a probabilistic model of the microstructural responses in the background and in the shadow. Thereafter, I have completed the MRF segmentation model with microstructure analysis. The proposed adaptive kernel selection strategy considers the local background properties. I have shown via synthetic and real-world examples, that the improved framework outperforms the purely color based model, and methods using a single kernel.
  - goal: considering textural differences in the separation
  - texture distribution parameters can be analytically estimated

Effects of the Microstructure Model

Synthesised Example

- Input image - Fig. a)
  - homogenous but noisy foreground (light rectangle in the middle)
  - inhomogeneously textured background
- Results:
  - b) only intensity based separation
  - c) intensity + edge based model
  - d) proposed adaptive kernel selection



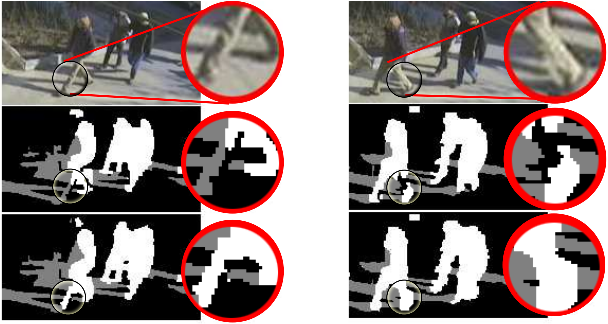
Effects of the Microstructure Model

Real Image Examples

- Improvements in regions of finely textured details

Without texture analysis

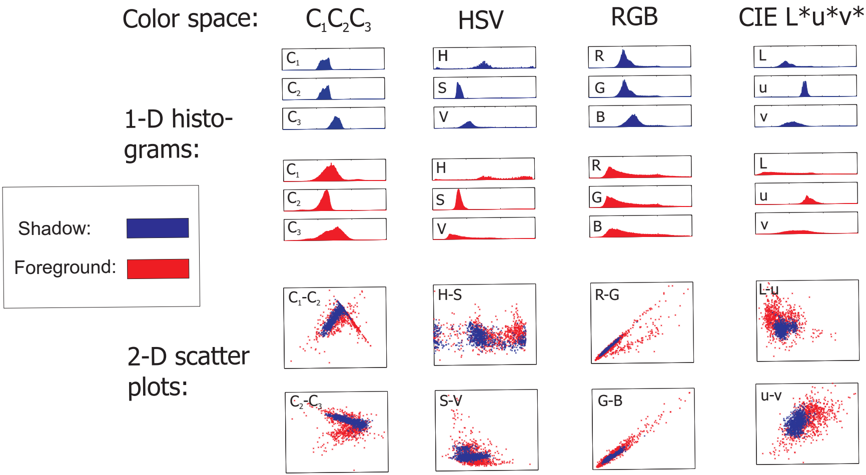
With texture analysis



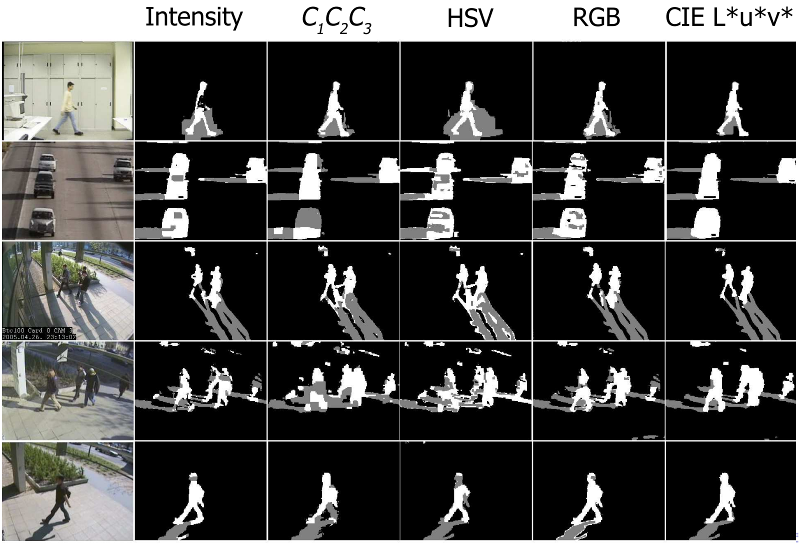
Thesis 1.4: Color Space Selection for Shadow Detection

- I have experimentally shown that among the widespread color spaces, the CIE L*u*v* model is the best for cast shadow detection, both using an elliptical separation in the space of the pixel-level descriptors and regarding a color space independent extension of the proposed MRF-segmentation model.

Shadow Descriptors in Different Color Spaces

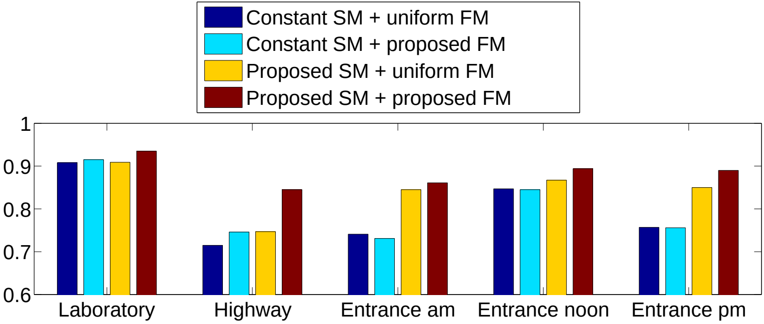


Comparing Color Spaces in the Proposed MRF Framework



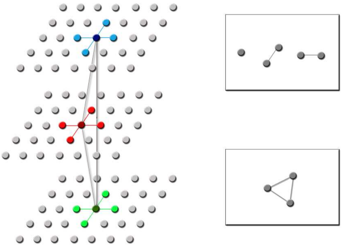
Quantitative Evaluation of the Results from Thesis Group 1.

- 2 benchmark sequences and 3 real surveillance videos, in aggregate 861 evaluated frames
- Metric: *F*-measure (harmonic mean of recall and precision of foreground detection)
- Notation: SM = shadow model, FM = foreground model



Thesis Group 2: Three-Layer Markovian Models

- I have developed novel three-layer MRF models for object motion detection in unregistered aerial image pairs and built-in change detection in aerial photos captured with several years time difference. I have experimentally validated the proposed models.



Content

Thesis 2.1

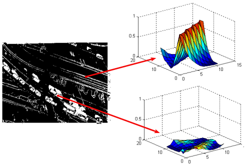
- I have developed a novel statistical model for object motion detection in image pairs captured by moving airborne vehicles. I have experimentally shown that the proposed approach outperforms previous models which use purely linear image registration techniques or local parallax removal.

Feature Extraction

- 1. feature: gray level difference

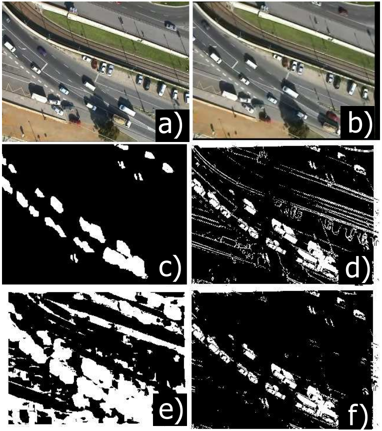
$$d(s) = \tilde{x}_2(s) - x_1(s)$$

- 2. feature: local correlation peak value $c(s)$



- Pixel s belongs to background, if:

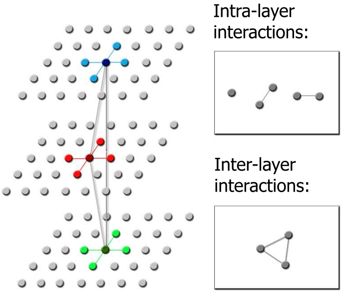
$$|d(s)| < T_1 \ \text{OR} \ c(s) > T_2$$



- Spatial smoothing is necessary - composite Markovian model due to feature integration

3-Layer Markov Random Field Model

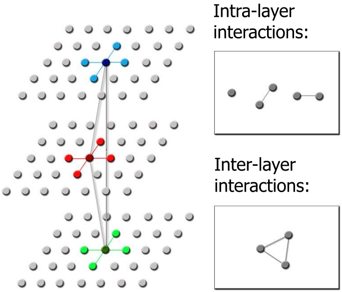
- Classes: object motion (i.e. foreground), background
- Layers:
  - 1. observation layer: MRF based on intensity difference
  - Segmentation layer: final result by feature integration
  - 2. observation layer: MRF based on the correlation peak feature



- Singletons
  - Labels in the observation layers should be consistent with local features $d(s)$ resp. $c(s)$
- Intra layer connections
  - Smooth segmentation in each layer
- Inter layer interactions
  - Syncronizing the segmentations by label fusion.

3-Layer Markov Random Field Model

- Classes: object motion (i.e. foreground), background
- Layers:
  - 1. observation layer: MRF based on intensity difference
  - Segmentation layer: final result by feature integration
  - 2. observation layer: MRF based on the correlation peak feature

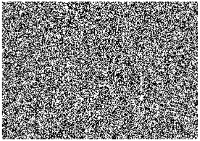


- Singletons
  - Labels in the observation layers should be consistent with local features $d(s)$ resp. $c(s)$
- Intra layer connections
  - Smooth segmentation in each layer
- Inter layer interactions
  - Syncronizing the segmentations by label fusion.

Field Energy Optimization

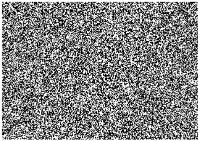
$$\widehat{\underline{\omega}} = \operatorname{argmin}_{\underline{\omega}\in\Omega}\Big\{ -\sum_{s\in S}\log P(d(s)|\omega(s^d)) - \sum_{s\in S}\log P(c(s)|\omega(s^c)) +$$

$$+ \sum_{\{r^i,s^i\}\in\mathcal{C}_2} \beta\cdot\delta\big(\omega(r^i),\omega(s^i)\big) + \sum_{s\in S} I\big(\omega(s^d),\omega(s^c),\omega(s^*)\big)\Big\}$$

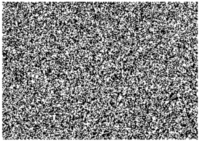


d(s) layer

ground truth



c(s) layer



result layer

Test Datasets and Reference Methods

- Database: 83 image pairs from 3 test sets
- Comparison to manual segmentation
- Metric: F-measure
- Reference methods:
  - FFT similarity matching
  - Method of Farin and With , ICIP 2005[1]
  - Supervised affine matching

[1] D. Farin and P. With, "Misregistration Errors in Change Detection Algorithms and How to Avoid Them," in *Proc. International Conference on Image Processing (ICIP)*, vol. 2, pp. 438-441, Genoa, Italy, Sept. 2005.

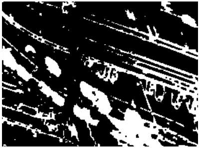
Results

First image

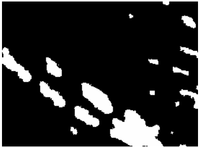
Second image

Ground truth

FFT similarity



Farin's method



Supervised affine

3-layer MRF

Results

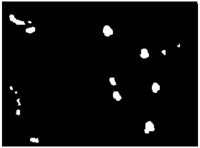
First image

Second image

Ground truth

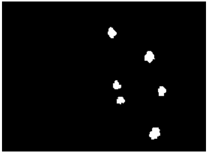
FFT similarity

Farin's method



Supervised affine

3-layer MRF



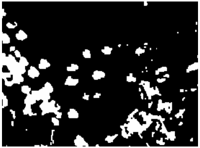
Results

First image

Second image

Ground truth

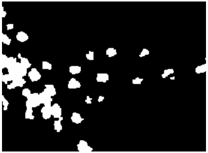
FFT similarity



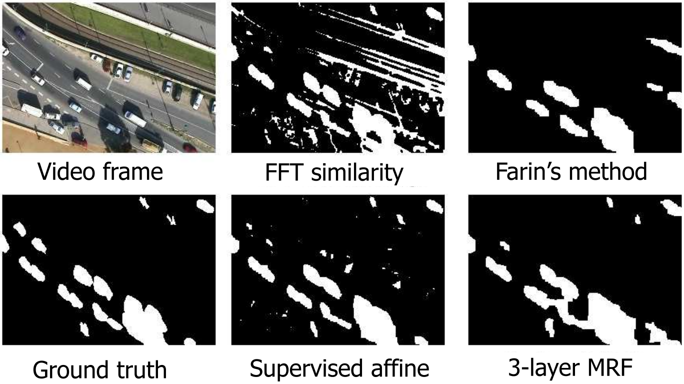
Farin’s method

Supervised affine

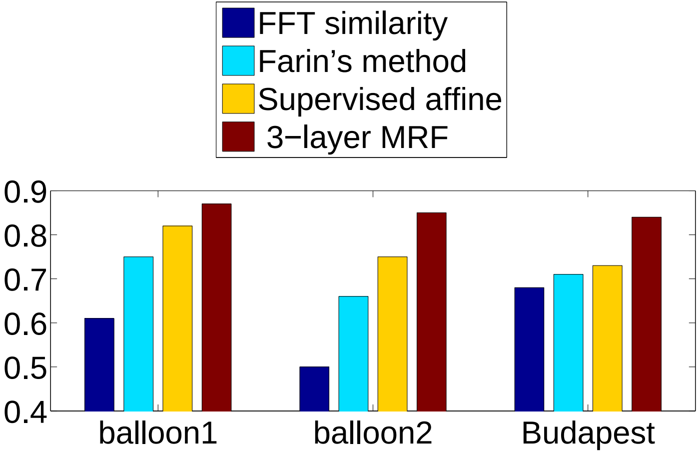
3-layer MRF



Results in ‘Balloon 1’ Test Set



Quantitative Results (*F*-measure)



Content

Thesis 2.2

- I have developed a Markovian framework for structural change detection in aerial photos captured with significant time difference. I have shown through an application on built-in change detection that connecting the segmentations of the different images via pixel-level links results in an efficient region based change detection method, which is robust against the noise and incompleteness of the class descriptors.

Structural Change Detection in Aerial Images Captured with Large Time Differences

- Preliminary registered aerial photos
- 5-20 years difference
- Pixel-level comparison is irrelevant
- Region based change detection
  - Segmenting the images with the same clusters: built-in and natural areas
  - Detecting regions with changed clusters

Feature Selection for Built-in Change Detection

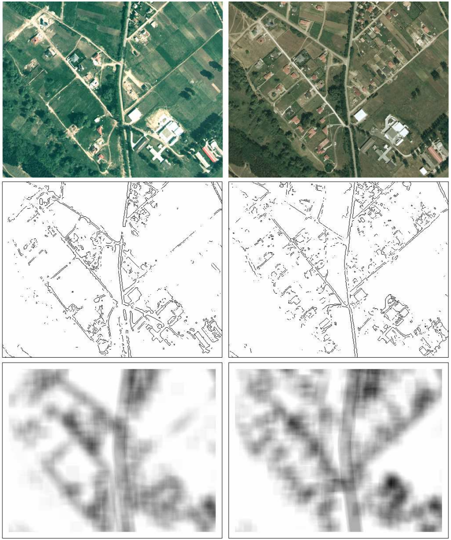
- “Edge-density” textural descriptor
- Edge map:

$$E = \{E(s) | s \in S\}$$

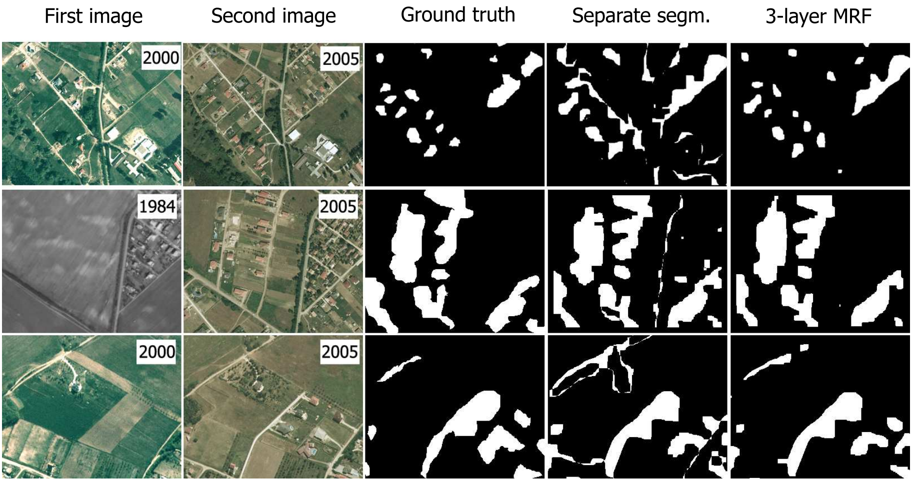
- Edge density map:

$$T(s) = \frac{1}{(2W+1)^2} \sum_{|r-s| \leq W} E(r)$$

- In built-in areas the edge density is high
- Region borders are ambiguous



Built-in Change Detection: Results



Content

Journal Publications

- Cs. Benedek and T. Szirányi: “Bayesian Foreground and Shadow Detection in Uncertain Frame Rate Surveillance Videos”, *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 608 - 621, April 2008. (IF: 2.715)
- Cs. Benedek and T. Szirányi: “Study on Color Space Selection for Detecting Cast Shadows in Video Surveillance,” *International Journal of Imaging Systems and Technology, Special Issue on Applied Color Image Processing*, vol. 17, no. 3, pp. 190-201, Wiley, 2007 (IF: 0.983)

International Conference Publications 2/1

- Cs. Benedek , T. Szirányi, Z. Kato and J. Zerubia:“A Multi-Layer MRF Model for Object-Motion Detection in Unregistered Airborne Image-Pairs,” in Proc. *IEEE International Conference on Image Processing (ICIP)*, 2007
- Cs. Benedek and T. Szirányi: “Markovian Framework for Foreground-Background-Shadow Segmentation of Real World Video Scenes”, *Asian Conference on Computer Vision (ACCV), Lecture Notes in Computer Science, Springer*, 2006
- Cs. Benedek and T. Szirányi: “Color Models of Shadow Detection in Video Scenes”, in Proc. *International Conference on Computer Vision Theory and Applications (VISAPP)*, 2007
- Cs. Benedek and T. Szirányi: “Markovian Framework for Structural Change Detection with Application on Detecting Built-in Changes in Airborne Images,” in Proc. *IASTED International Conference on Signal Processing, Pattern Recognition and Applications (SPPRA)*, 2007

International Conference Publications 2/2

- D. Szolgay, Cs. Benedek and T. Szirányi: "Fast Template Matching for Measuring Visit Frequencies of Dynamic Web Advertisements", *International Conference on Computer Vision Theory and Applications (VISAPP)*, 2008
- Z. Szlávik, L. Havasi, Cs. Benedek and T. Szirányi: "Motion-based Flexible Camera Registration", *IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, 2005
- Z. Szlávik, T. Szirányi, L. Havasi and Cs. Benedek: "Optimizing of Searching Co-Motion Point-Pairs for Statistical Camera Calibration", *IEEE International Conference on Image Processing (ICIP)*, 2005
- Z. Szlávik, T. Szirányi, L. Havasi and Cs. Benedek: "Random Motion for Camera Calibration", *European Signal Processing Conference (EUSIPCO)*, 2005
- L. Havasi, Z. Szlávik, Cs. Benedek and T. Szirányi, "Learning human motion patterns from symmetries", *ICML Workshop on Machine Learning for Multimedia*, 2005
- L. Havasi, Cs. Benedek, Z. Szlávik and T. Szirányi: "Extracting Structural Fragments from Images Showing Overlapping Pedestrians", *IASTED International Conference on Visualization, Imaging, and Image Processing*, 2004

Acknowledgement

- Scientific Advisor: Tamás Szirányi
- Supporting Institutes:
  - Pázmány Péter Catholic University
  - MTA-SZTAKI
- Tamás Roska, leader of the Doctoral School
- Colleagues from the EEE SZTAKI Research Group
  - Zoltán Szlávik, László Havasi, István Petrás, Levente Kovács
- Colleagues at the doctoral school
- Collaborators
  - Josiane Zerubia (INRIA France), Zoltan Kato (University of Szeged)
- Tibor Vámos, leader of my former laboratory at SZTAKI
- Reviewers of the Dissertation
- All colleagues at SZTAKI, PPKE and INRIA institutes
- My Lívi, family & friends

Thank you for your attention

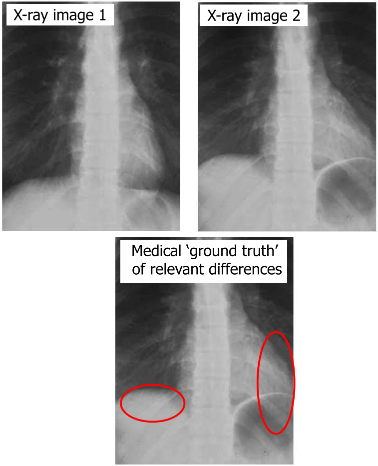
Content

Kernel Size Used for Texture Analysis

- Question: Can be the quality of segmentation enhanced by using 5×5 or larger kernels to compute the microstructural responses instead of 3×3 kernels?
- Effects of using larger kernels:
  - $\oplus$ improved detection of the internal parts of object/background regions
  - $\ominus$ increased artifacts appear near to the class-boundaries
- Optimal kernel size depends on:
  - Image resolution
  - Size of objects
- In 320×240 video frames 3×3 proved to be a good comprise.

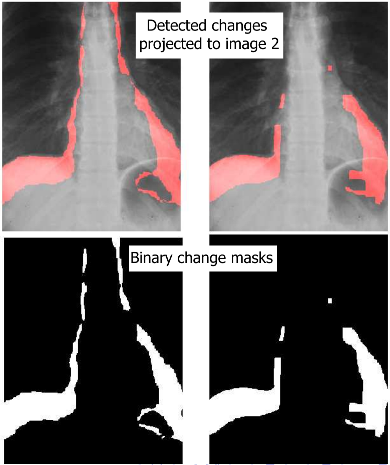
Segmenting X-ray Images with the Proposed Markovian Structure

Input images and `ground truth'



Change detection results with

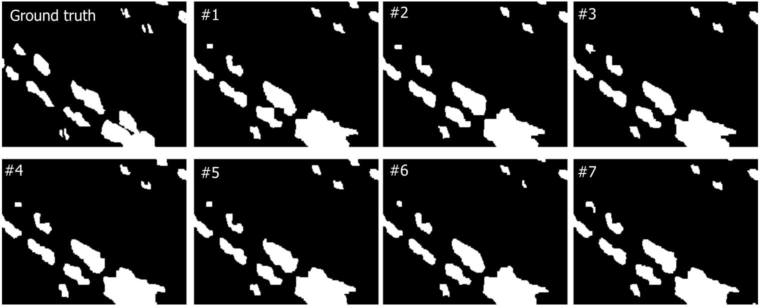
a) separate segm. b) three-layer MRF



Experimental Validation of the Stochastic Optimizer

Repeatability of the experiments for the 3-layer MRF model

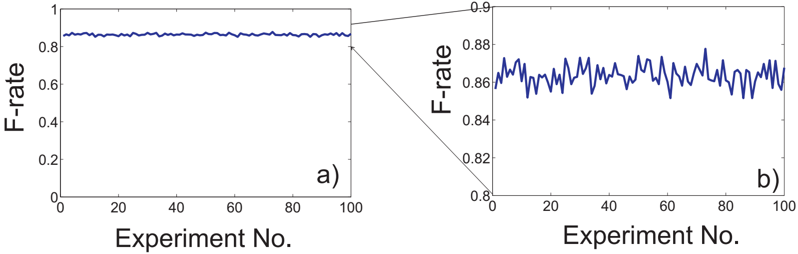
- Qualitative results of 7 different pseudo-stochastic optimization experiments with the *same* image pair, *same* parameters and relaxation settings, but *different* seeds for the RANDOMIZE calls



Experimental Validation of the Stochastic Optimizer

Repeatability of the experiments for the 3-layer MRF model

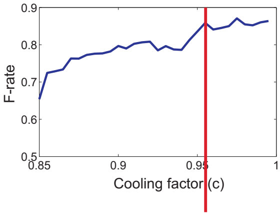
- Quantitative results of 100 different pseudo-stochastic optimization experiments with the *same* image pair, *same* parameters and relaxation settings, but *different* seeds for the RANDOMIZE calls
  - measured mean value of F-rates: 0.8635
  - measured standard deviation: 0.0057



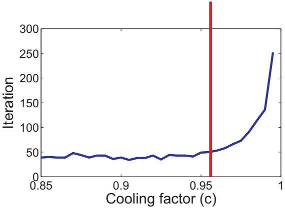
Experimental Validation of the Stochastic Optimizer

Effects of changing the cooling factor and the iteration number

Performance as a function of the cooling factor (c), obtained at convergence:



Number of MMD iterations till convergence as a function of the cooling factor (c):



Study on the convergence speed obtained at c=0.96: