Change Detection in Optical Aerial Images by a Multi-Layer Conditional Mixed Markov Model

Csaba Benedek[1,2] Tamás Szirányi[1]

[1]Distributed Events Analysis Research Group
Computer and Automation Research Institute, Hungary

[2]ARIANA joint Project-team INRIA/CNRS/UNSA, Sophia Antipolis, France

Seminar at Florida State University, Department of Statistics,
Tallahassee, 16 December 2008

Content

Content

Introduction, research goals

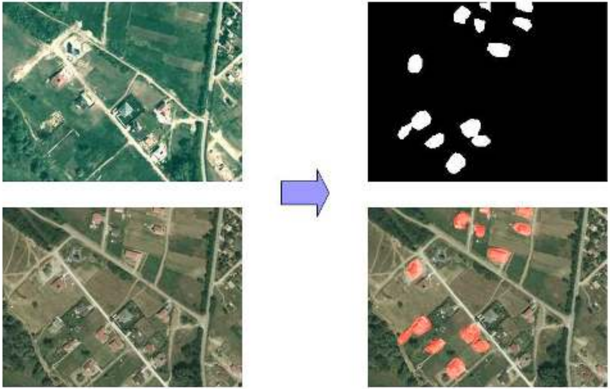
- Change detection in optical aerial image pairs
  - new built-up regions, building operations
  - planting of trees, fresh plough-land
  - groundwork before building-over etc
- Large (many years) time differences → different seasons, illumination conditions, vegetations etc.
- Input – preliminary registered orthophotos:

Image 1 (G_1)

Image 2 (G_2)

Task formulation

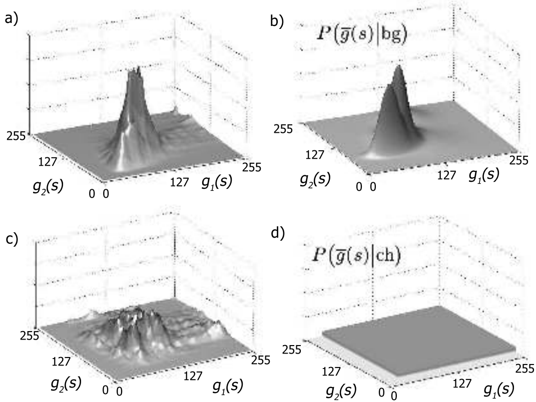
- Binary image segmentation problem:
  - Classifying each pixel s of the image lattice S as 'change' (below: white) or 'background' (i.e. unchanged, with black)



Content

Feature definition

- Global statistics of intensity co-occurrences
  - Feature vector of pixel s is pair of intensity values of s in the two images: $\overline{g}(s) = [g_1(s), g_2(s)]^T$, $g_1(s) \in G_1$, $g_2(s) \in G_2$
  - Global statistics in changed/background regions:



Feature density modeling

- Multi-Gaussian Intensity-based (MGI) change detection: 'change' class is modeled by a 2-D uniform pdf, while 'background' with a mixture of Gaussians in the $\overline{g}(s)$ feature space
  - Class 'background':

$$P\big(\overline{g}(s)\big|\mathrm{bg}\big) = \sum_{i=1}^{K} \kappa_i \cdot \eta\big(\overline{g}(s), \overline{\mu}_i, \Sigma_i\big)$$

    - using fixed K (e.g. $K = 5$) and EM parameter estimation
  - Class 'change':

$$P\big(\overline{g}(s)\big|\mathrm{ch}\big) = \begin{cases} \frac{1}{(b_1-a_1)\cdot(b_2-a_2)}, & \text{if } \overline{g}(s) \in \Gamma \\ 0 & \text{otherwise,} \end{cases}$$

    - where $\overline{g}(s) \in \Gamma$ iff $a_1 \leq g_1(s) \leq b_1$ and $a_2 \leq g_2(s) \leq b_2$

Validation of the Intensity Feature

- Result of the intensity based ML pixel classification

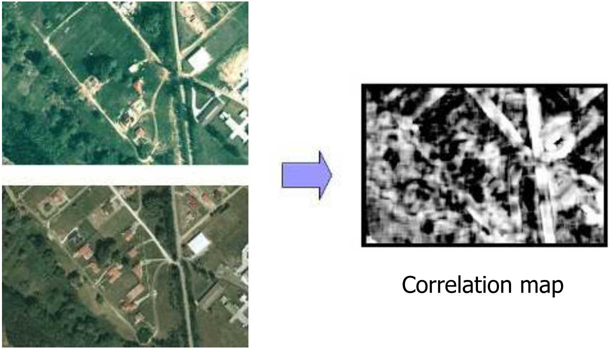
$$\phi_g(s) = \operatorname{argmax}_{\psi \in \{\text{ch},\text{bg}\}} P\left(\bar{g}(s) \middle| \psi\right)$$

Red: estimated changes

False alarms in **textured** image regions

Feature extraction 2

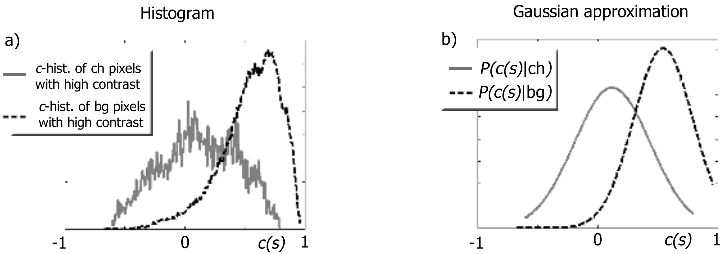
- Second feature: local block correlation
  - $c(s)$: normalized cross correlation between the $v \times v$ neighborhoods of pixel s in G_1 resp. G_2 images (used $v = 17$).



Correlation map

Feature extraction 2

- Feature statistics



$$P\big(c(s)\big|\mathrm{ch}\big) = \eta\big(c(s), \vartheta_{\mathrm{ch}}, \varsigma_{\mathrm{ch}}^2\big) = \frac{1}{\sqrt{2\pi\varsigma_{\mathrm{ch}}^2}} \exp\left(-\frac{(c(s) - \vartheta_{\mathrm{ch}})^2}{2\varsigma_{\mathrm{ch}}^2}\right)$$

$$P\big(c(s)\big|\mathrm{bg}\big) = \eta\big(c(s), \vartheta_{\mathrm{bg}}, \varsigma_{\mathrm{bg}}^2\big)$$

Feature extraction 2

- Result of the correlation based ML pixel classification

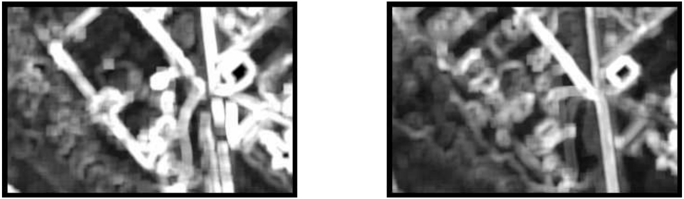
$$\phi_c(s) = \operatorname{argmax}_{\psi \in \{\mathrm{ch},\mathrm{bg}\}} P\big(c(s)\big|\psi\big)$$

Estimated changes

False alarms in **homogenous** image regions

Feature of feature selection

- Feature selection based on local contrast
  - $\nu_i(s)$, $i \in \{1,2\}$: variance of the gray levels over the $v \times v$ neighborhood of s in G_i
- Joint variance vector: $\overline{\nu}(s) = [\nu_1(s), \nu_2(s)]^T$
- Local variance (contrast) maps:



$\nu_1(.)$ $\nu_2(.)$

Feature integration

- Partitioning the pixels of the 'training' image pairs:

$$S_{\nu_1,\nu_2} = \{s \in S | \nu_1(s) \approx \nu_1,\ \nu_2(s) \approx \nu_2\}$$

- Reliability 'histogram' of the intensity map ϕ_g:

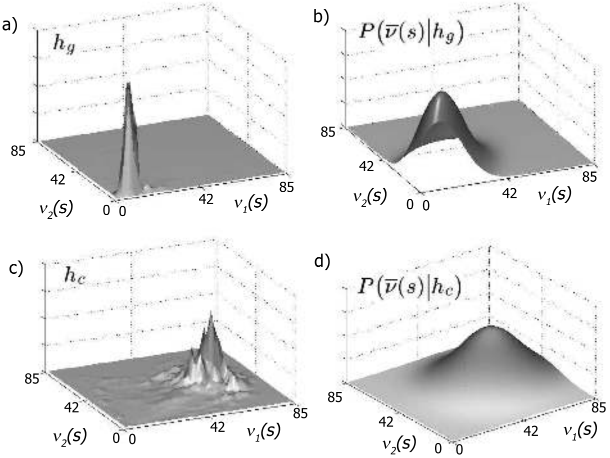
$$h_g[\nu_1, \nu_2] = \frac{\text{number of correctly classified pixels in } S_{\nu_1,\nu_2}}{\text{number of erroneously classified pixels in } S_{\nu_1,\nu_2}}$$

- Reliability 'histogram' of the correlation map ϕ_c:

$$h_c[\nu_1, \nu_2] = \frac{\text{number of correctly classified pixels in } S_{\nu_1,\nu_2}}{\text{number of erroneously classified pixels in } S_{\nu_1,\nu_2}}$$

Feature integration

- Reliability histograms h_g and h_c with 2-D Gaussian density approximations:



Feature integration

- Gaussian models for the reliability of the *g*/*c* features:

$$P\left(\overline{\nu}(s)\middle|h_g\right) = \eta\left(\overline{\nu}(s), \overline{\mu}_g, \overline{\overline{\Sigma}}_g\right)$$

$$P\left(\overline{\nu}(s)\middle|h_c\right) = \eta\left(\overline{\nu}(s), \overline{\mu}_c, \overline{\overline{\Sigma}}_c\right)$$

- Contrast-based feature selection-map (red where the correlation feature is estimated as more reliable):

$$\phi_\nu(s) = \text{argmax}_{\chi\in\{g,c\}} P\left(\overline{\nu}(s)\middle|h_\chi\right).$$

Feature integration

- Initial feature integration rule:
  - ϕ_*: final change mask

$$\phi_*(s) = \begin{cases} \phi_g(s) & \text{if } \phi_\nu(s) = g \\ \phi_c(s) & \text{if } \phi_\nu(s) = c \end{cases}$$

- Result of the pixel-by-pixel approach:

output $\phi_*(s)$ map

ground truth

- Observation: improved, but still noisy result

Towards a Robust Segmentation Approach

- Global labeling optimization over the image instead of pixel-by-pixel segmentation
  - pixel level feature descriptions
  - interaction constraints between neighbouring pixels
- Conventional Markov Random Field approaches must be extended:
  - multi layer model for considering the different label maps
  - particular role of the $\overline{\nu}(s)$ feature:
    - switching ON and OFF the $\overline{g}(s)$ respectively $c(s)$ features into the integration process
    - **data dependent** dynamic links are needed in the graph
    - application of **Mixed Markov models**

Content

Image Segmentation with Conventional MRFs

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

- Goal: generate a K-colored segmented image, with segmentation classes: $L = \{C_1, \ldots, C_K\}$
  - Here: $K = 2$; C_1=change and C_2=background.
- f_s: local feature observed at pixel s
- ω_s: label of pixel s which marks its segmentation class
- Segmentation with Markov Random Fields (MRF):
  - Pixels' feature-values must agree with the class models specified by their label:
    - Classes are characterized by 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 Conventional MRFs

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

- Goal: generate a K-colored segmented image, with segmentation classes: $L = \{C_1, \ldots, C_K\}$
  - Here: $K = 2$; C_1=change and C_2=background.
- f_s: local feature observed at pixel s
- ω_s: label of pixel s which marks its segmentation class
- Segmentation with Markov Random Fields (MRF):
  - Pixels' feature-values must agree with the class models specified by their label:
    - Classes are characterized by 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 Conventional MRFs

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

- Goal: generate a K-colored segmented image, with segmentation classes: $L = \{C_1, \ldots, C_K\}$
  - Here: $K = 2$; C_1=change and C_2=background.
- f_s: local feature observed at pixel s
- ω_s: label of pixel s which marks its segmentation class
- Segmentation with Markov Random Fields (MRF):
  - Pixels' feature-values must agree with the class models specified by their label:
    - Classes are characterized by 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 Conventional MRFs

- 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:

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

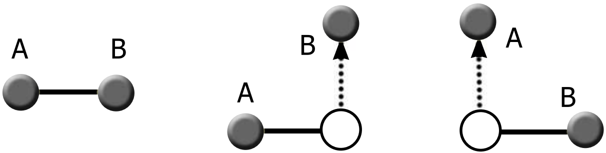
  where Ω denotes the set of all the possible global labelings.
- **(Hammersley-Clifford theorem)**: $P(\underline{\omega}|\mathcal{F})$ can be factorized into individual terms whose domains are the cliques of the graph.

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

  - where C is an arbitrary clique and V_C is the potential of C.

Step forward to Mixed Markov models

- In MRFs two nodes directly interact if and only if they are connected by a (static) edge
- In Mixed models the connections can also be data dependent
- Two types of nodes:
  - *regular nodes:* same role as nodes of MRF's
  - *address nodes:* their 'labels' are pointers to regular nodes
- Regular nodes A and B may interact iff they are connected by (i) a (static) edge OR (ii) a chain of a static edge and a dynamic address pointer



Three cases when A and B regular nodes may interact (address nodes are marked by white circles, edges by lines, pointers by dotted arrows)

Probability modeling in Mixed Markov models

- A priory probability of a global labeling:

$$P(\underline{\omega}) = \frac{1}{Z}\exp\left(-\sum_{C\in\mathcal{C}} V_C\left(\omega_C, \omega_C^{\mathcal{A}}\right)\right)$$

  - where C is a clique and ω_C is the set of labels inside C:

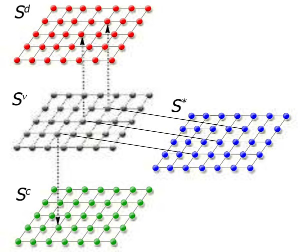
$$\omega_C = \{\omega(q) | q \in C\}$$

  - while $\omega_C^{\mathcal{A}}$ is the set of node labels pointed by the address nodes of clique C:

$$\omega_C^{\mathcal{A}} = \{\tilde{\omega}(a) \big| a \in \mathcal{A} \cap C, \omega(a) \neq \text{nil}\}$$

    - $\mathcal{A}$ is the set of address nodes and $\tilde{\omega}(a) = \omega(\omega(a))$ for $a \in \mathcal{A}$

4-layer Mixed Markov model for Change Detection

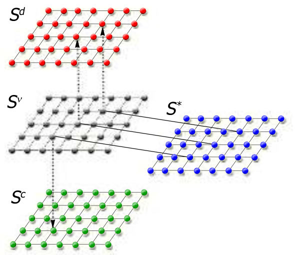


$s \longrightarrow \{s^d, s^c, s^\nu, s^*\}$

- Regular layers
  - S^g, S^c: change masks based on the $\overline{g}(s)$ resp. $c(s)$ features
  - S^*: combined layer – output change mask
- Address layer
  - S^ν: switch layer providing configurable, data-driven inter-layer connections

- Node labels: $\omega(s^i)$: $i \in \{d, c, \nu, *\}$, $s \in S$
- Cliques and clique potentials:
  - Singletons: data – label consistency
  - Intra-layer connections: smooth label maps V_{C_2}
  - Inter-layer interactions: label fusion V_{C_3}

4-layer Mixed Markov model for Change Detection



$s \rightarrow \{s^d, s^c, s^\nu, s^*\}$

- Regular layers
  - S^g, S^c: change masks based on the $\overline{g}(s)$ resp. $c(s)$ features
  - S^*: combined layer – output change mask
- Address layer
  - S^ν: switch layer providing configurable, data-driven inter-layer connections

- Node labels: $\omega(s^i)$: $i \in \{d, c, \nu, *\}$, $s \in S$
- Cliques and clique potentials:
  - Singletons: data – label consistency
  - Intra-layer connections: smooth label maps V_{C_2}
  - Inter-layer interactions: label fusion V_{C_3}

Singleton terms

- Assuming conditional independent observations, let be:

$$P(\mathcal{F}|\Omega) = \prod_{s\in S} P\big(\overline{g}(s)|\omega(s^g)\big) \cdot P\big(c(s)|\omega(s^c)\big) \cdot P\big(\overline{\nu}(s)|\omega(s^\nu)\big)$$

where we use previously defined densities for the S^g and S^c layers:

$$P\big(\overline{g}(s)|\omega(s^g) = \mathrm{bg}\big) = \sum_{i=1}^{K} \kappa_i \cdot \eta\big(\overline{g}(s), \overline{\mu}_i, \Sigma_i\big)$$

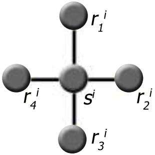
$$P\big(\overline{g}(s)|\omega(s^g) = \mathrm{ch}\big) = 1/[(b_1 - a_1) \cdot (b_2 - a_2)]$$

$$P\big(c(s)\big|\omega(s^c) = \psi\big) = \eta\big(c(s), \vartheta_\psi, \varsigma_\psi^2\big), \ \psi \in \{\mathrm{ch}, \mathrm{bg}\}$$

Singletons of S^ν will be later given.

Intra-layer Doubleton Potentials

- Doubleton cliques: smoothing priors of the segmentation within each layer.



- The potential of an intra-layer clique $C_2 = \{s^i, r^i\} \in \mathcal{C}_2$, $i \in \{g, c, *, \nu\}$:

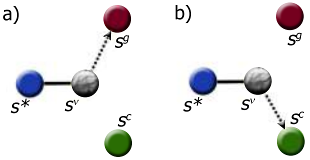
$$V_{C_2} = \begin{cases} -\delta^i & \text{if } \omega(s^i) = \omega(r^i) \\ +\delta^i & \text{if } \omega(s^i) \neq \omega(r^i) \end{cases}$$

for a constant $\delta^i > 0$.

Inter-layer interactions

- Inter-layer cliques: $\omega(s^*)$ should mostly be equal either to $\omega(s^g)$ or to $\omega(s^c)$, depending on the 'vote' of the $\nu(s)$ feature.

- Edge between s^* and s^ν
- Address node s^ν should point either to s^g or to s^c:

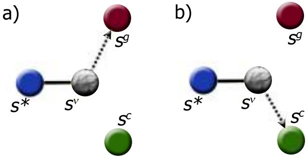


$$\forall s \in S: \ \omega(s^\nu) \in \{s^g, s^c\}$$

- The directions of the address pointers are influenced by the singletons of S^ν:

$$P\big(\overline{\nu}(s)|\omega(s^\nu) = s^\chi\big) = P\big(\overline{\nu}(s)|h_\chi\big), \ \ \chi \in \{g, c\}$$

Inter-layer interactions



- The potential function of the inter-layer clique $C_3 = \{s^*, s^\nu\}$:

$$V_{C_3}\Big(\omega(s^*), \tilde{\omega}(s^\nu)\Big) = \begin{cases} -\rho & \text{if } \omega(s^*) = \tilde{\omega}(s^\nu) \\ +\rho & \text{otherwise} \end{cases}$$

where $\rho > 0$, and $\tilde{\omega}(s^\nu) = \omega\big(\omega(s^\nu)\big)$.

Labeling optimization

- MAP estimation of the optimal global labeling $\underline{\hat{\omega}}$:

$$\widehat{\underline{\omega}} = \arg\min_{\underline{\omega}\in\Omega} \Bigg\{ \sum_{s\in S} -\log P\big(\overline{g}(s)|\omega(s^g)\big) +$$

$$+ \sum_{s\in S} -\log P\big(c(s)|\omega(s^c)\big) + \sum_{s\in S} -\log P\big(\overline{\nu}(s)|\omega(s^\nu)\big) +$$

$$+ \sum_{i;\{s,r\}\in\mathcal{C}_2} V_{C_2}\big(\omega(s^i), \omega(r^i)\big) + \sum_{s\in S} V_{C_3}\big(\omega(s^*), \tilde{\omega}(s^\nu)\big) \Bigg\}$$

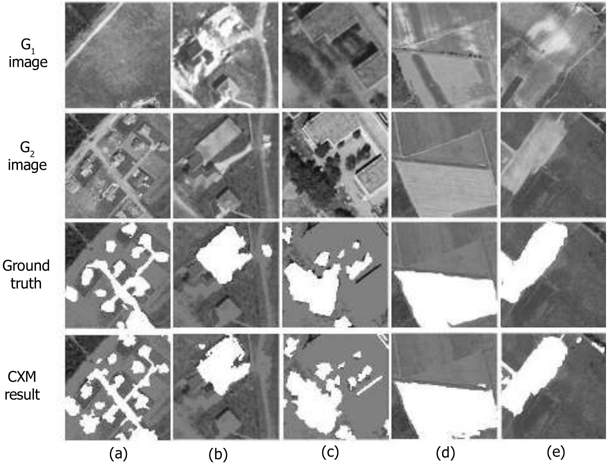
- Optimization by simulated annealing (Modified Metropolis algorithm)
- Output: labeling of the S^* layer.

Content

Test datasets and reference methods

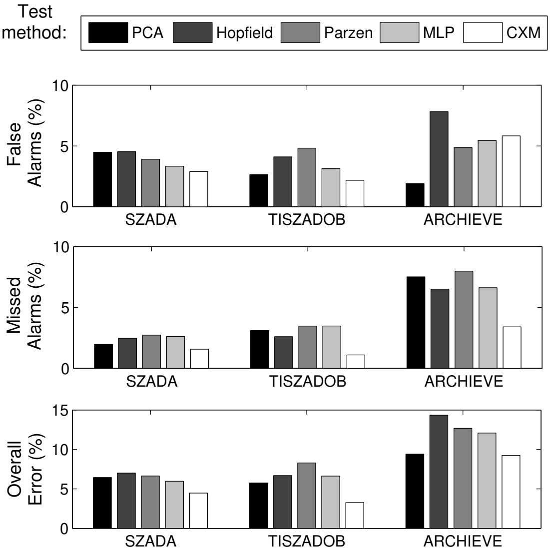
- Database: *three* sets of optical aerial image pairs provided by the Hungarian Institute of Geodesy Cartography & Remote Sensing (FÖMI) and Google Earth.
  - Data set SZADA: images by FÖMI from 2000 resp. 2005. *Seven* - also manually evaluated - photo pairs, covering in aggregate $9.5\mathrm{km}^2$ area at *1.5m/pixel* resolution.
  - Data set TISZADOB: *five* photo pairs from 2000 resp. 2007 ($6.8\mathrm{km}^2$) with similar size and quality parameters to SZADA.
  - Test pair ARCHIVE, an aerial image taken by FÖMI in 1984 and a corresponding Google Earth photo from around 2007.
- Manually generated ground truth masks
- Metrics: number of false and missed alarms
- 4 reference methods: PCA, Hopfield, MLP, Parzen

Ground truth generation

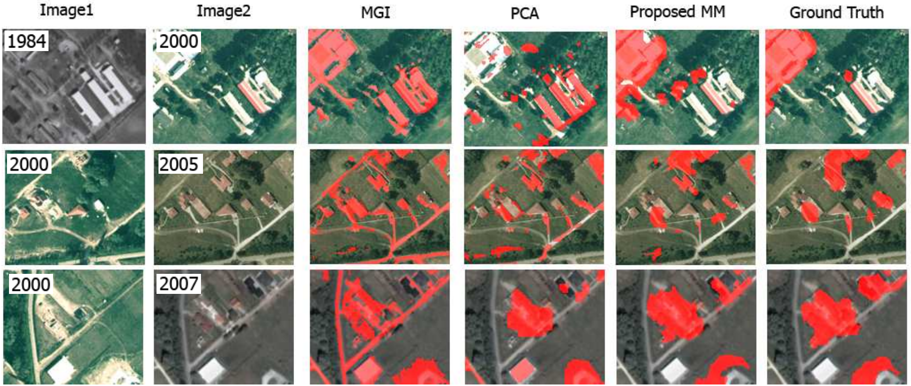


Change prototypes considered for ground truth generation (a) new built-up regions (b) building operations (c) planting of trees (d) fresh plough-land (e) groundwork before building over

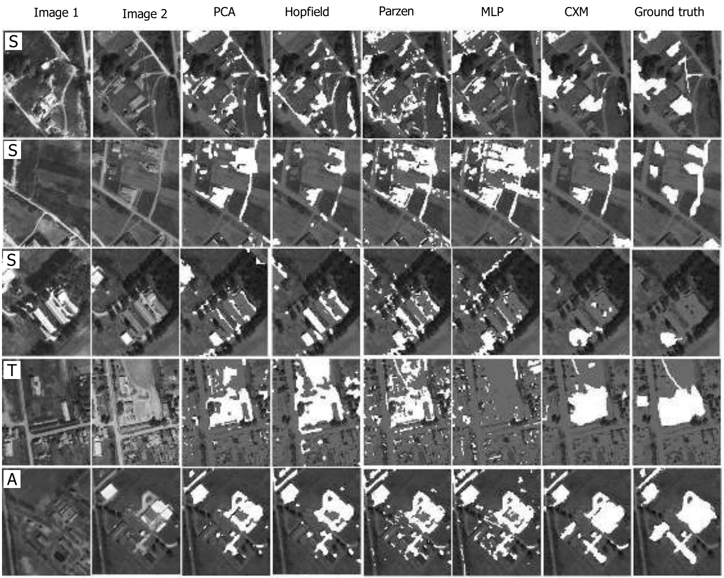
Quantitative comparison



Qualitative comparison



Qualitative comparison



References

- Cs. Benedek and T. Szirányi: "Change Detection in Optical Aerial Images by a Multi-Layer Conditional Mixed Markov Model", submitted to *IEEE Transacions of Geosciences and Remote Sensing*, 2008, before 2nd review round
- Cs. Benedek and T. Szirányi: "A Mixed Markov Model for Change Detection in Aerial Photos with Large Time Differences", *International Conference on Pattern Recognition (ICPR)*, Tampa, Florida, USA, December 8-11, 2008

Acknowledgement and contacts

- The authors would like to thank
  - Josiane Zerubia from INRIA for her kind advices regarding the proposed model
  - the MUSCLE Shape Modeling E-Team for financial support of this work
  - Prof. Anuj Srivastava for inviting me to the Florida State University
  - the Associated team Shapes (INRIA, FSU) for supporting my visit to FSU

- Contact me: Csaba Benedek
  - Url: `http://web.eee.sztaki.hu/~bcsaba/`
  - E-mail: `cbenedek@sophia.inria.fr`, `bcsaba@sztaki.hu`