Detection of Object Motion Regions in Aerial Image Pairs with a Multi-Layer Markovian Model

Csaba Benedek, Tamás Szirányi, *Senior Member, IEEE,* Zoltan Kato, *Senior Member, IEEE,* and Josiane Zerubia, *Fellow, IEEE*

Abstract

We propose a new Bayesian method for detecting the regions of object displacements in aerial image pairs. We use a robust but coarse 2-D image registration algorithm. Our main challenge is to eliminate the registration errors from the extracted change map. We introduce a three-layer Markov Random Field (L^3MRF) model which integrates information from two different features, and ensures connected homogenous regions in the segmented images. Validation is given on real aerial photos.

Index Terms

Aerial images, change detection, camera motion, MRF

DOCUMENT AVAILABILITY

For information please contact the authors bcsaba@sztaki.hu

A. *Evaluation versus different fusion models*

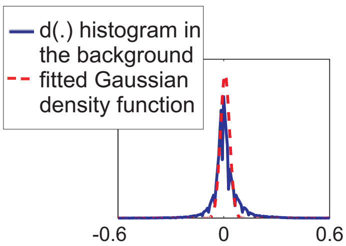
I. ACKNOWLEDGEMENT

This work was partially supported by the EU project MUSCLE (FP6-567752). The authors would like to thank the MUSCLE Shape Modeling E-Team for financial support and Xavier Descombes from INRIA for his kind remarks and advices.

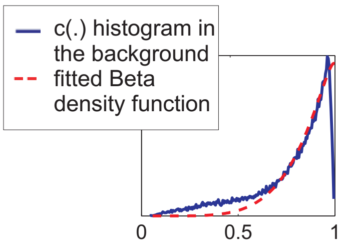
Cs. Benedek and T. Szirányi are with the Distributed Events Analysis Research Group, Computer and Automation Research Institute Budapest, Hungary. Z. Kato is with the Image Processing and Computer Graphics Dept., University of Szeged, Hungary, J. Zerubia is with the Ariana project-team (joint research group INRIA/CNRS/UNSA), Sophia Antipolis, France

Fig. 1. High resolution stereo image pair taken by the Hungarian Ministry of Defence Mapping Company© above Budapest with a few sec. time difference.

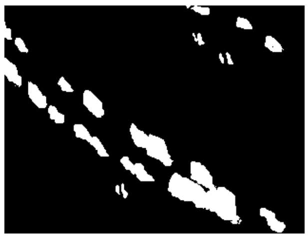
(a) First input image: X_1



(c) $d(.)$ feature statistics

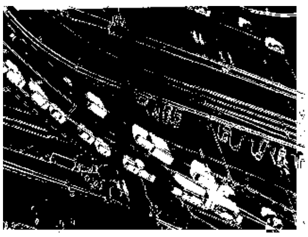


(e) $c(.)$ feature statistics



(g) Ground truth

(b) Registered second image: $\widetilde{X}_2$



(d) D image - segmented image based only on $d(.)$

(f) C image - segmented image based only on $c(.)$

(h) Result of the AND operation on D and C images

Fig. 2. Feature selection.

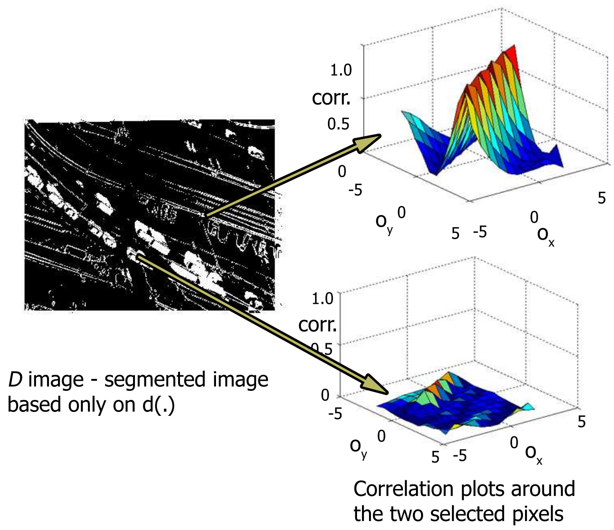


Fig. 3. Plot of the correlation values over the search window around two given pixels. The upper pixel corresponds to a parallax error in the background, while the lower pixel is part of a real object displacement.

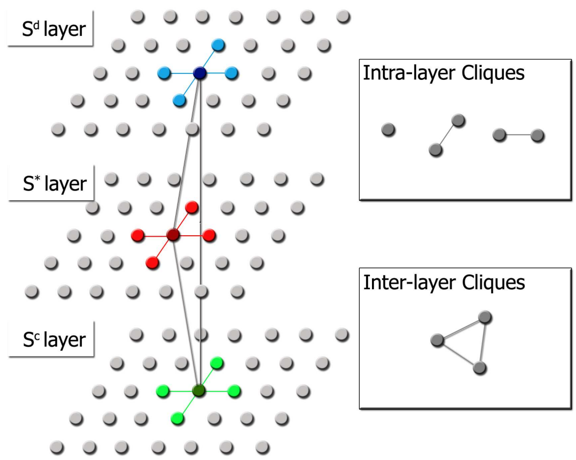


Fig. 4. Structure of the proposed three-layer MRF (L^3MRF) model

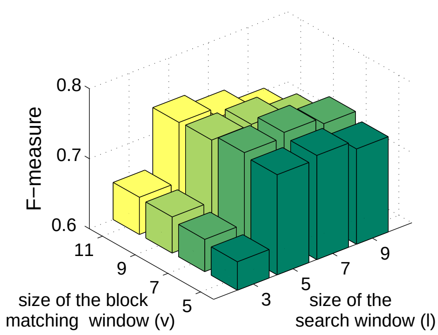


Fig. 5. Performance evaluation as a function of the block matching (v) and search window size (l) using training images from the 'balloon1' test set. Here, $v = 7$ and $l = 7$ proved to be optimal.

TABLE I

COMPARISON OF DIFFERENT RELATED METHODS AND THE PROPOSED MODEL. (NOTES FOR TEST METHODS: †IN FRAME-DIFFERENCING MODE ‡WITHOUT THE MULTIVIEW STRUCTURE CONSISTENCY CONSTRAINT)

Author(s)	Published paper(s)	Input of the method	Frame-rate of the image source	Compensated parallax	Expected object motions	Related test method
Reddy and Chatterji	TIP 1996	Image pair	no limit	none	arbitrary	Reddy
Irani and Anandan	TPAMI 1998	2 or 3 frames	no limit	no limit	arbitrary	Epipolar
Sawhney et al.	TPAMI 2000	3 frames	no limit	sparse, heavy	arbitrary	-
Pless et al.	TPAMI 2000	Sequence	video (≈ 25) fps	no limit	small	-
Kumar et al.	TIP 2006	Image pair	video fps	none	arbitrary	Affine
Farin and With	TCSVT 2006	Image pair†	no limit	dense/sparse, bounded	large	Farin †
Yin and Collins	CVPR 2007	Sequence	6fps	none	small	-
Yuan et al.	TPAMI 2007	Sequence	5fps	dense parallax	small	Epipolar †,‡
Jodoin et al.	TIP 2007	Image pair	video fps	bounded	small	KNNBF
Proposed method		Image pair	$0.3 - 1$ fps	dense/sparse, bounded	large	L^3MRF

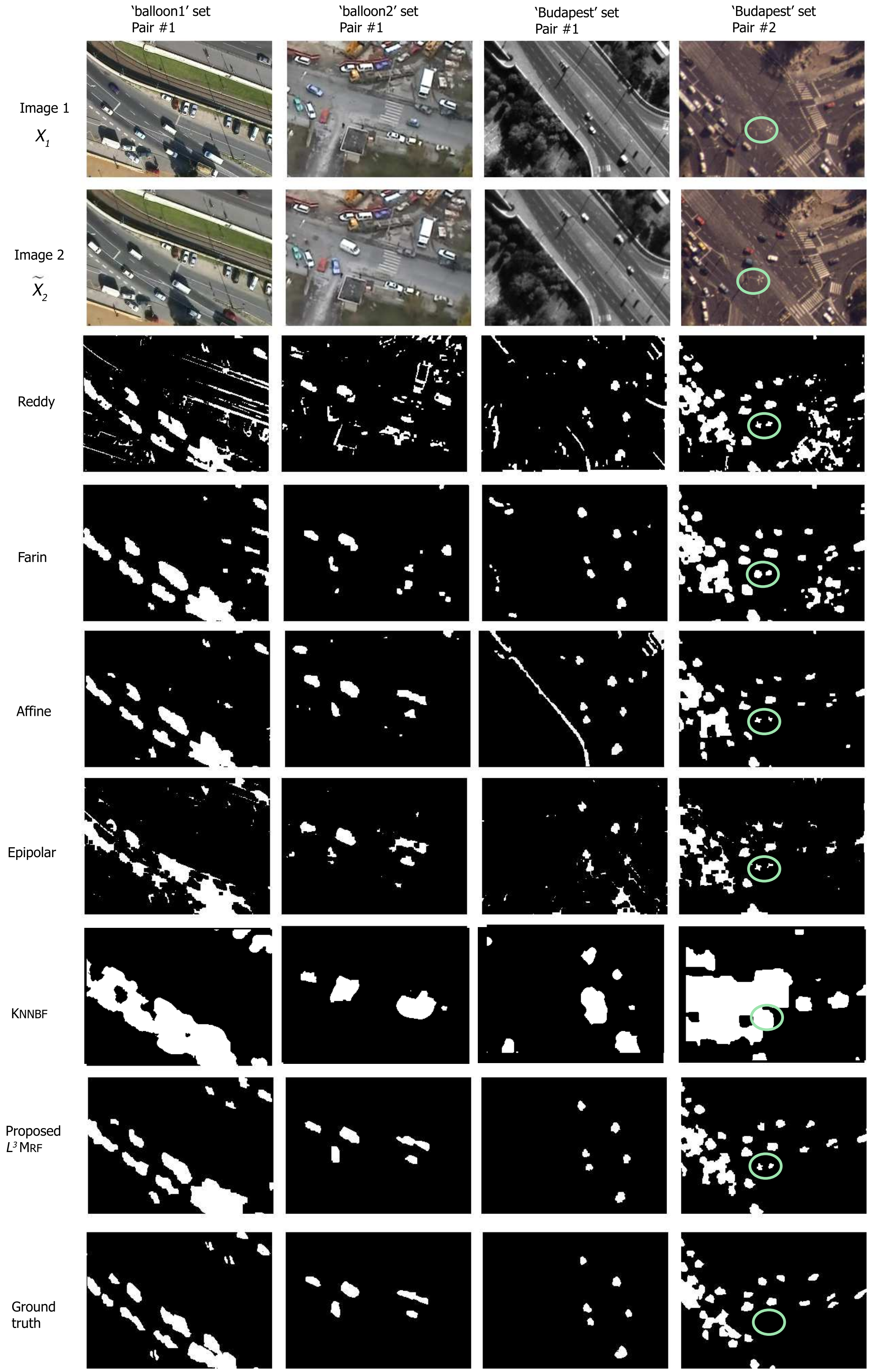


Fig. 6. Comparative segmentations: four selected test image pairs, segmentation results with different methods and ground truth. In the right column, the ellipses demonstrate a limitation: a high standing lamp is detected as a false moving object by all methods.

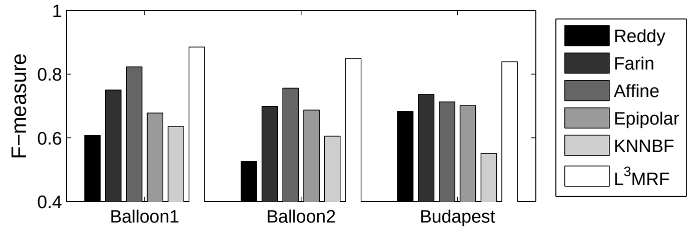


Fig. 7. Numerical comparison of the proposed model (L^3MRF) to five reference methods, using three test sets: 'balloon1' (52 image pairs), 'balloon2' (22) and 'Budapest' (9).

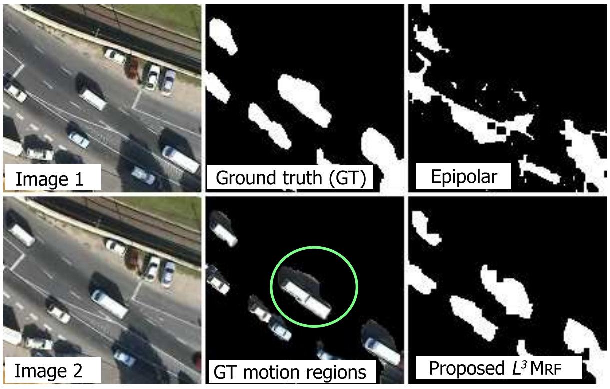


Fig. 8. Segmentation example with the *Epipolar* method and the proposed L^3MRF model. Circle in the middle marks a motion region which erroneously disappears using the *Epipolar* approach.

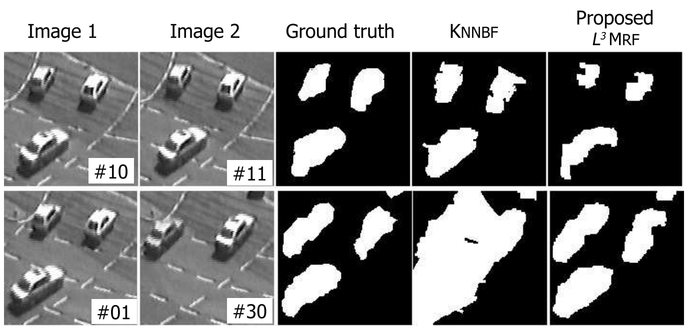


Fig. 9. Comparison of the proposed L^3MRF model to the KNNBF method, using image pairs from the KARLSRUHE sequence (# denotes the frame number). In consecutive frames of the video (above) KNNBF produces better results, however, our L^3MRF model significantly dominates if (below) we chose two frames with 1 second time difference

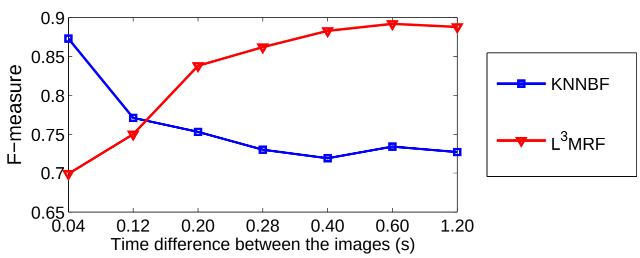


Fig. 10. Comparing KNNBF to L^3MRF. Quantitative segmentation results (F-measure) of different frame pairs from the KARLSRUHE test sequence, as a function of the time difference between the images. The proposed method dominates if the images are taken with larger elapsed time, which results in large object displacements.

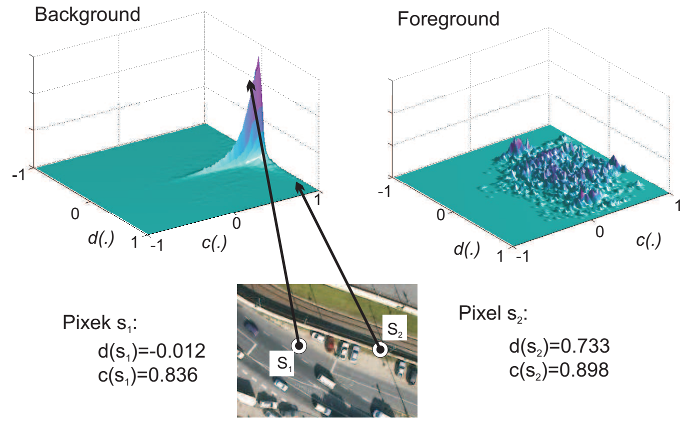
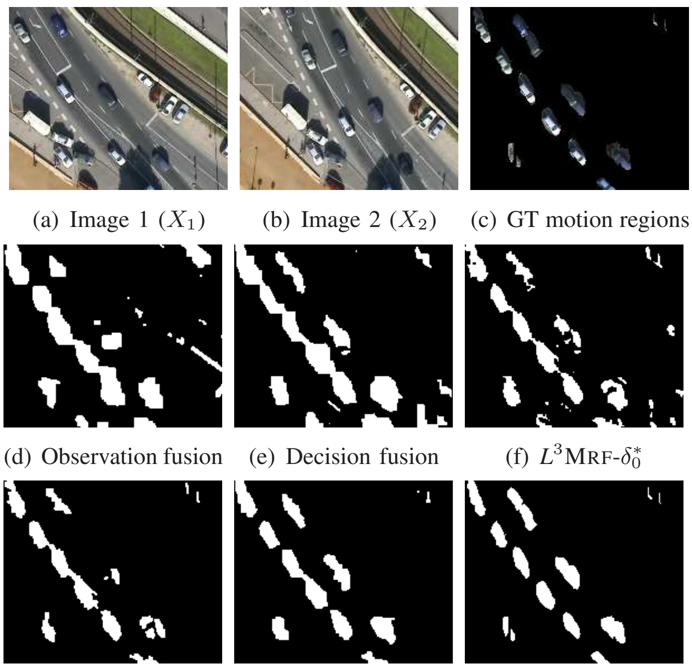


Fig. 11. Limitations of the observation fusion approach with the proposed feature selection. Above: 2-D joint histogram of the $\overline{f}(s) = [d(s), c(s)]$ vectors obtained in the background and in the foreground training regions. Below: two selected *background* pixels and backprojection of the corresponding feature vectors to the background histogram.



(a) Image 1 (X_1) (b) Image 2 (X_2) (c) GT motion regions

(d) Observation fusion (e) Decision fusion (f) L^3MRF-δ_0^*

(g) L^3MRF-$\delta_0^{d,c}$ (h) Proposed L^3MRF (i) Binary ground truth

Fig. 12. Evaluation of the proposed L^3MRF model versus different fusion approaches. Methods are described in Section -A.

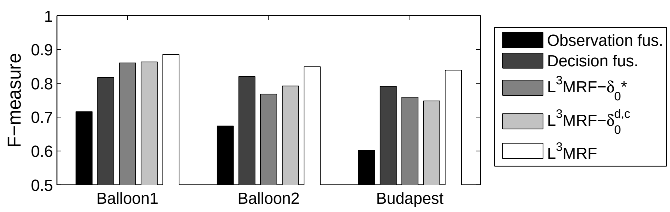


Fig. 13. Numerical comparison of the proposed model (L^3MRF) to different information fusion techniques with the same feature selection.