Spatio-Temporal Fusion of Object Segmentation Approaches for Moving Distant Targets

Michael Teutsch and Wolfgang Krüger
Fraunhofer IOSB
Fraunhoferstr. 1, 76131 Karlsruhe, Germany
Email: {michael.teutsch, wolfgang.krueger}@iosb.fraunhofer.de

***Abstract*—Object segmentation can be an important step between object detection and tracking. Especially in image processing applications, where object detection is difficult due to high object distance, camera motion, or noise, the detection result might not be precise enough to robustly initialize tracks and perform multi-target tracking. In this paper we present the detection and segmentation of moving objects in image sequences coming from a small Unmanned Aerial Vehicle (UAV). Based on the detection and tracking of local image features, camera motion is compensated and independent motion created by moving vehicles and people on the ground is found. By clustering the independent motion vectors initial object hypotheses are generated which may be affected by over- and under-segmentation. For improvement, several object segmentation approaches are introduced and tested. Best results are achieved with a spatio-temporal fusion of some approaches. Both spatial and temporal information is provided by the local image features. The object segmentation approaches and the fusion methods are evaluated for their completeness and precision.**

Keywords: moving target detection, independent motion, segmentation, surveillance, reconnaissance, UAV video.

I. Introduction

Automatic camera-based surveillance including detection and tracking of moving objects from Unmanned Aerial Vehicles (UAVs) can be a powerful support for many application areas of public security and defense. However, it is also challenging due to camera motion, varying environmental conditions, high object distance, changing object background, object shading, or noise. Object detection and tracking are well-discussed problems, but while facing these difficult challenges the introduction of object segmentation can be helpful to improve the detection result and, thus, enable better tracking performance.

Our image data is coming from a small UAV flying at an altitude of about 400 m. The camera is directed perpendicularly to the ground with a Ground Sampling Distance (GSD) of about 0.345 m/pixel. A standard car covers about 15×5 pixels in the image. We present the detection and segmentation of moving objects to allow reliable and robust multi-target tracking. The whole processing chain with an emphasis to tracking is described in more detail in [1], whereas we focus in this paper on object segmentation and the implementation of several approaches which are spatially fused for higher robustness compared to the single approaches. In addition, temporal fusion is introduced by the registration and stacking of consecutive images relative to the same moving object. This way, the object stays in the same position in the image stack and the background is moving leading to a blurred background improving the segmentation performance.

Related Work

The related work will focus on articles in which segmentation of distant objects is applied and discussed. Approaches developed for both visual-optical and infrared images are considered, since many segmentation algorithms are suitable for different kinds of image data.

Moving vehicles are segmented by Zhang et al. [2] in gray-value images. A fusion is applied of regions provided by Mean Shift algorithm and moving pixels detected by block-based motion estimation and refined by spatio-temporal continuity constraints. Segmentation and tracking of moving objects in UAV video sequences is the aim of Zhang [3]. Since object size can vary strongly in the image, two segmentation approaches are implemented: statistical snake for bigger objects as well as standard change detection after registration of two consecutive images and optional Sobel edge detection with morphological dilation for small objects. Li et al. [4] are first locating targets in the temporal domain by image registration and then using Fuzzy C-means clustering for precise segmentation in infrared images. Oreifej et al. [5] segment detected ROIs for body blobs aiming to identify humans in aerial images. Pixels are iteratively assigned to be fore- or background with kernel density estimation. Ibrahim et al. [6] use Gaussian Mixture learning to determine a motion ROI and segment the moving object using Graph-Cuts. Yao et al. [7] extract object blobs from the difference images of UAV videos with morphological operations. Finally, Qi et al. [8] segment man-made objects in UAV images by advanced thresholding using Bézier histograms. The idea is that histogram enhancement provides similar results as image enhancement, but faster.

The remainder of the paper is organized as follows: in Section II the detection and clustering of independent motion is introduced. Object segmentation including spatio-temporal fusion is described in Section III. A quantitative evaluation based on real data with respect to segmentation completeness and precision is given in Section IV. Finally, the conclusions are presented in Section V.

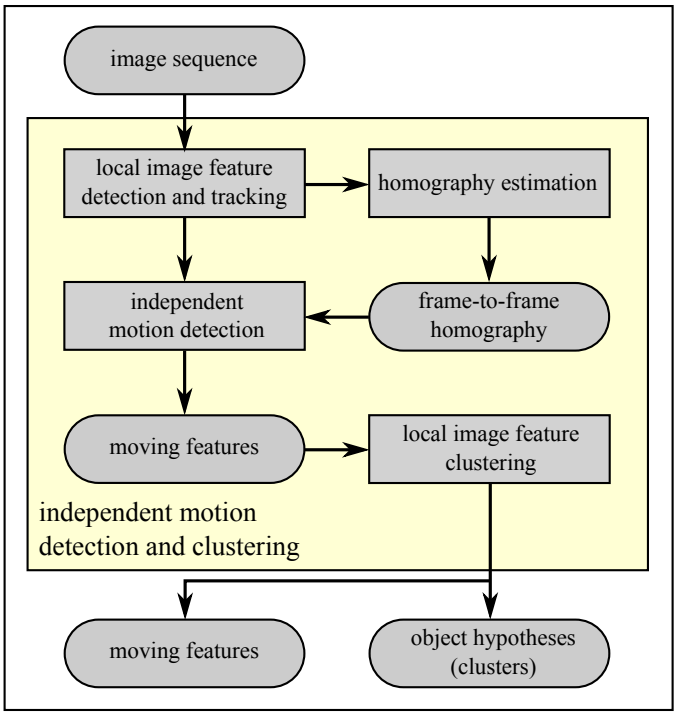

Figure 1. The concept of independent motion detection and clustering.

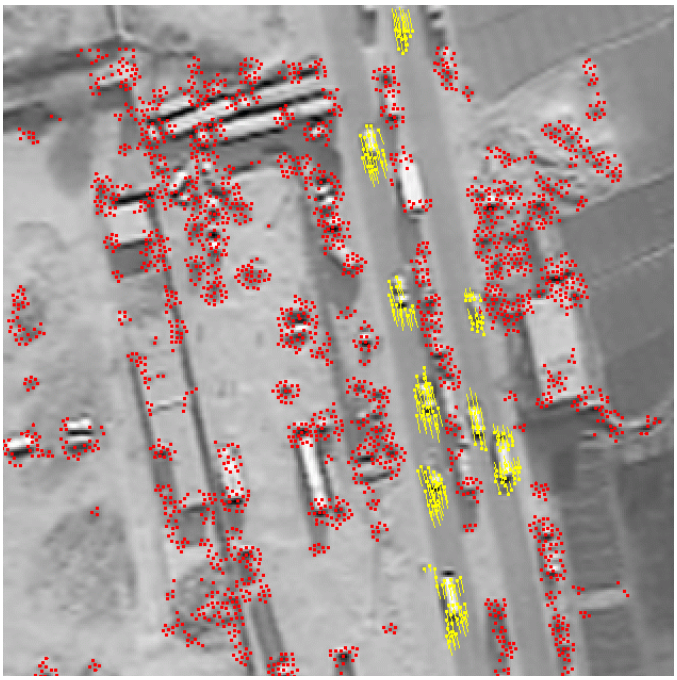

Figure 2. Stationary (red) and moving (yellow) detected and tracked local image features.

II. Independent Motion Detection and Clustering

Independent motion detection is based on local image feature tracking, image registration, and motion clustering as seen in Fig. 1. The first step is to detect and track corner features [9] in the processed image sequence. Feature correspondences between subsequent image frames are used to estimate homographies as global image transformations [10] for frame-to-frame alignment. These homographies are needed to eliminate the dominating image motion caused by the moving camera platform. After motion compensation, image velocities relative to the static background of the monitored scene are estimated for each feature track. Independent motion is then detected at features with significant relative velocities and discriminates features at moving objects from features at the static background [11].

The final step of independent motion detection is to cluster the local image features classified as moving relative to the static background. We use a single-linkage clustering based on the position and velocity estimates of the moving features. The idea is to start with finding isolated motion clusters and to handle more complicated cases in the following object segmentation steps. The selection of distance thresholds is based on the known ground sampling distance and the expected size of vehicles.

Our approach does not require camera calibration and is largely independent of object appearance. Using homographies instead of plane+parallax decompositions with multiview geometric constraints [12], [13] is adequate since our image data is mainly from UAVs operating at higher altitudes. We do not use motion compensated difference images [14] because we achieved much more reliable results compared to our previous work [15] which was based on difference images.

Fig. 2 shows typical results for our independent motion detection. In total, there are 6934 feature tracks. From these tracks 283 are correctly classified as coming from moving objects. Shown are the estimated relative velocities (scaled by factor 5) for all tracks. Note the large number of feature tracks at parked vehicles which are correctly classified as part of the static background. Fig. 3 shows the distribution for the magnitude of the relative velocities (pixels per frame) estimated for the 283 moving features. It can be seen that our system is able to reliably estimate and classify sub-pixel relative motion.

Fig. 7 shows the clustering result for the moving features from Fig. 2. In total, 14 motion clusters are found with 11 clusters corresponding to the successful segmentation of a vehicle or pedestrian/bicyclist. Two clusters at a long white vehicle display an over-segmentation, since there is an isolated motion cluster at the front and at the rear of the vehicle. The remaining cluster shows an under-segmentation, where two vehicles which are too close and have similiar relative velocities are grouped together.

Our experience is that especially in crowded scenes such cases of motion-based over-segmentation or under-segmentation cannot be avoided at the same time and additional appearance-based image features should be exploited for a more precise object segmentation.

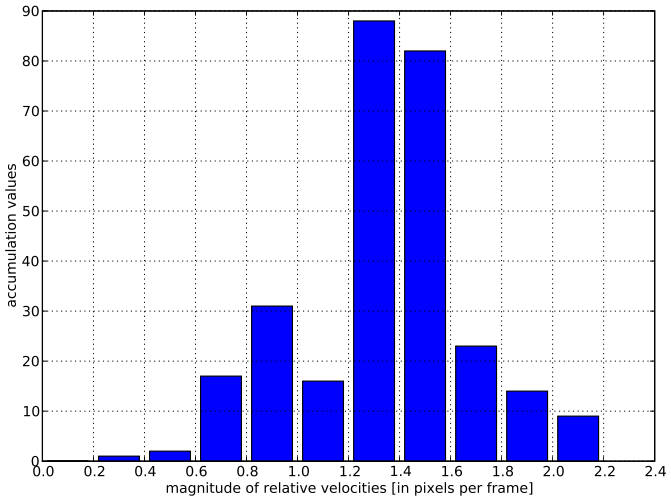

Figure 3. The velocities of the moving features (yellow) in Fig. 2.

III. Object Segmentation and Spatio-Temporal Fusion

The fundament for object segmentation is the set of independent motion clusters. Over- and under-segmentation is likely to occur especially in cases where motion direction and velocity of objects close to each other is very similar. Hence, advanced object segmentation is needed to provide robust and reliable object hypotheses to subsequent multi-target tracking. In this section, the general concept, the different implemented approaches, and the fusion methods are presented.

A. Concept

The concept of object segmentation is shown in Fig. 4. As input data the object hypotheses (clusters) and the moving features provided by independent motion detection and clustering are used. For each local feature image stacking is performed as temporal fusion. The result is more object hypotheses with smoothed background due to image stacking and taking the pixelwise mean gray-value. This set of hypotheses is processed by five different approaches for object boundary or blob detection generating an accumulation image. With adaptive thresholding and connected-component labeling, blobs are extracted and finally stored as bounding boxes. Spatial fusion is applied during accumulation image calculation and segment clustering. The result of the second spatial fusion is a set of object hypotheses represented by segments.

B. Accumulator Image Calculation

Based on the cluster boundaries a Region Of Interest (ROI) is generated for each cluster bigger than the original cluster to be able to handle over- and partial segmentation. The first part of object segmentation is the calculation of an accumulator image. This image has the same size as the ROI. In a pre-processing step, motion direction and GSD are used to generate normalized ROI images as presented in former work [11]. With normalized motion direction, paraxial bounding boxes can be used, and with normalized GSD the sensibility and dependency of the approaches towards their parameters is decreased. Furthermore, object size estimation in meters is possible.

We have chosen approaches generating accumulator images for better and easier fusion. Values such as gradient magnitudes or relative connectivity numbers are accumulated and object boundaries or blobs appear as bright areas in the accumulator. In the following subsections the five implemented approaches for accumulator image calculation are presented. The analysis of these images will be described in Subsection III-C.

1) Morphological Gradient: Morphological image processing [16] is versatilely applicable. Widely used are morphological erosion, dilation, opening and closing, where a binary or gray-value image is manipulated using a structuring element of arbitrary size and shape. Erosion e and dilation d are needed to calculate gradient magnitudes G. For gray-value image I, gray-value structuring element s, and pixel position (x, y) the definition is:

$$I \ominus s = e(x,y) = \min_{i,j}\{I(x+i, y+j) - s(i,j)\}, \quad (1)$$

$$I \oplus s = d(x,y) = \max_{i,j}\{I(x-i, y-j) + s(i,j)\}. \quad (2)$$

where (i, j) are the pixel positions of s. Morphological gradient calculation as proposed by Lee et al. [17] is performed using the following equation:

$$G = (I \oplus s) - (I \ominus s)\,. \quad (3)$$

For gradient calculation all values of the structuring element s are 0, as only the size of s is relevant for G. Morphological operations are non-linear and, hence, have different properties compared to linear gradient calculation with Sobel or Prewitt filters. We use quadratic structuring elements s_i of different size to perform multi-scale gradient calculation [18] enhancing the robustness of object segmentation. In our application this leads to the following equation:

$$G = \sum_{i=1}^{n} ((I \oplus s_i) - (I \ominus s_i))\,. \quad (4)$$

2) LBP Gradient: Local Binary Patterns (LBPs) are an unique encoding of a local pixel neighborhood and well-proven in the application of texture classification [19]. A special subset of LBPs are rotation-invariant, uniform LBPs, which are called $LBP^{riu2}_{P,R}$ and can be interpreted as *texture primitives* such as edge, corner, or spot [20]. P denotes the number of considered neighbors and R the radius, as LBPs are organized circularly. We assume that all LBPs not being texture primitives and all spots are affected by noise, since spots might be the result of salt-and-pepper noise. All pixel positions, where such kinds of LBPs occur, are not considered for gradient calculation to suppress noise [21]. For all other pixel positions (x, y) we calculate the local variance between the neighbors as proposed by Ojala et al. [19]:

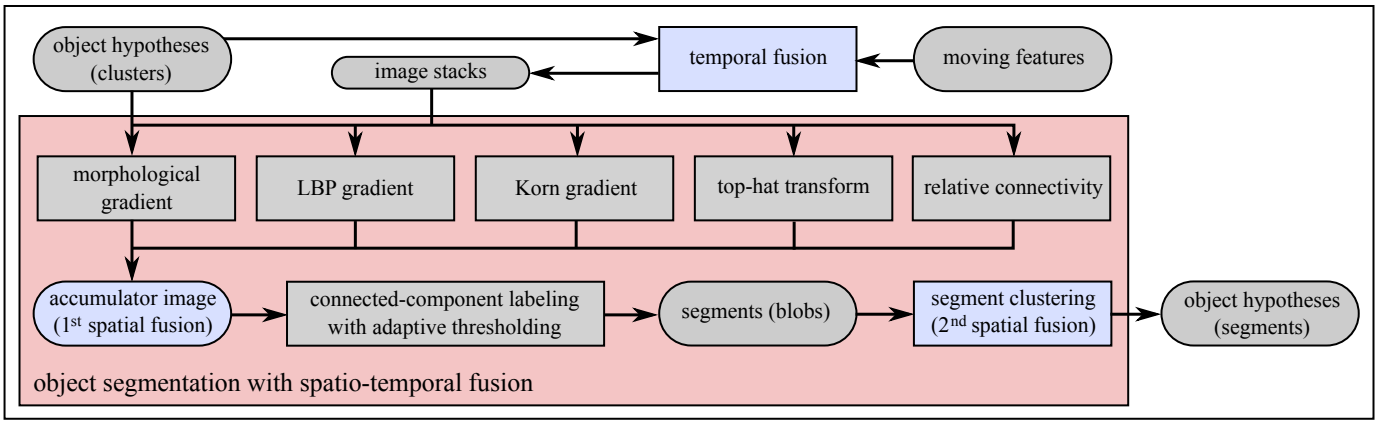

Figure 4. The concept of object segmentation.

$$VAR_{P,R} = \frac{1}{P}\sum_{p=0}^{P-1}(g_p - \mu)^2, \quad \text{where } \mu = \frac{1}{P}\sum_{p=0}^{P-1} g_p. \quad (5)$$

g_p denotes the gray-value of neighbor p. For higher robustness multi-scale LBPs [19] are used with varied radius r. Furthermore, the standard deviation produced better results especially in cases, where two objects with different brightness were in the same cluster. This leads to the following equation for the calculation of gradient magnitude G:

$$G(x,y) = \sum_{r=R_1}^{R_n} \sqrt{\mathrm{VAR}_{P,r}(x,y)}. \quad (6)$$

LBP gradient calculation is also non-linear. The main difference to the morphological approach, except of the filtering strategy, is the preservation of edge sharpness.

3) Korn Gradient: In contrast to morphological or LBP gradients, edge detection as proposed by Korn [22] is a linear approach. It is pretty similar to the well-known Canny edge detector [23] but using a modified filter matrix which is normalized.

4) Top-hat Transform: The top-hat transform [16] is another morphological operation. It is applied to extract bright or dark peaks from images. Therefore, morphological opening O and closing C are used:

$$O = I \circ s = (I \ominus s) \oplus s, \quad (7)$$
$$C = I \bullet s = (I \oplus s) \ominus s. \quad (8)$$

Depending on size and shape of the structuring element s morphological opening chops bright peaks (*hats*) smaller than s in images and morphological closing dark peaks, respectively. By subtracting the original image and the morphologically opened/closed one, only the peaks are kept as result. This is given by the following formalization:

$$T_w = I - (I \circ s), \quad (9)$$
$$T_b = (I \bullet s) - I, \quad (10)$$

where T_w is called white top-hat transform and T_b black top-hat transform, respectively.

Since objects in our UAV video data appear as bright and dark peaks both top-hat transforms are combined in one accumulator image.

5) Relative Connectivity: Relative connectivity was introduced by Yip [24] as a measure for a set of collinear points to belong together. Originally, it was used as an approach for line detection, namely Line Pattern Hough Transform (LPHT), since the regular Hough transform does not provide start and end points. But this algorithm is also applicable for object segmentation as we presented in former work [25]. Especially scattered, fragmentary, or noise affected object appearances can be successfully segmented.

The ROI image is scanned pixelwise in vertical and horizontal direction for pixels potentially belonging to an object. If pixel value threshold t is exceeded at image position (x, y), the pixel is considered. This leads to the binary belonging function B:

$$B(x,y) = \begin{cases} 1, & \text{if } I(x,y) \geq t \\ 0, & \text{if } I(x,y) < t. \end{cases} \quad (11)$$

For a set of collinear and equidistant points $P_i(x_i, y_i)$ with $i \in \{1, \ldots, n\}$ and $\forall i$: $B(x_i, y_i) = 1$ the relative displacement $(\Delta x, \Delta y)$ is given by

$$\Delta x = x_2 - x_1, \quad (12)$$
$$\Delta y = y_2 - y_1. \quad (13)$$

For being a start point, P_1 has to satisfy the constraint

$$B(x_1, y_1) = 1 \qquad \text{and} \qquad B(x_1 - \Delta x, y_1 - \Delta y) = 0. \quad (14)$$

The definition of P_n being an end point is done analogously

$$B(x_n, y_n) = 1 \qquad \text{and} \qquad B(x_n + \Delta x, y_n + \Delta y) = 0. \quad (15)$$

n is the connectivity number since

$$B(x_1 + i \cdot \Delta x, y_1 + i \cdot \Delta y) = 1 \qquad \forall i \in \{0, \ldots, n-1\}. \quad (16)$$

n is entered to the accumulator at start and end point position (x_1, y_1) and (x_n, y_n) to store this found relative connectivity.

Robustness is achieved by varying the relative displacement Δ.

Unfortunately, this approach is rather sensitive to the value of threshold t. Furthermore, it is not easily invertible, so dark objects are likely to get lost.

C. Accumulator Image Analysis

After accumulator image calculation high accumulation values (bright pixels) represent object boundaries or interior. To extract object blobs we apply connected-component labeling with adaptive gray-value threshold. The threshold is determined based on the assumption that a certain number of pixels in each cluster belongs to objects. By generating a gray-value histogram and using a fixed quantile, a certain percentage of the brightest pixels is accepted for connected-component labeling. The labeling process creates object blobs which might be scattered. Thus, morphological closing is applied to fill potential holes in these blobs. For each blob the minimal surrounding bounding box with same orientation as the related cluster represents the object hypothesis.

Example images for the whole object segmentation process are shown in Fig. 5. An under-segmented cluster with two merged objects is given and analyzed. The five segmentation approaches generate different accumulator images. The three gradient algorithms produce similar but not identical gradient magnitudes. The connected components have holes and gaps after labeling, but morphological closing generates good object segments (blobs). Three out of five approaches produce a correct result while the other two cannot handle the under-segmentation in this example. However, for the LBP approach it was only a matter of one pixel making the blobs growing together.

D. Spatio-Temporal Fusion

The spatial and temporal fusion at three points in the processing chain, as seen in Fig. 4, aims to increase the performance and robustness of object segmentation. Different approaches may work well alone, but they might work better in combination.

1) Temporal Fusion: For temporal fusion we take advantage of local feature tracking which provides the necessary temporal information. Sub-pixel motion is detected precisely and is the fundament for this fusion. For each moving feature an image stack is constructed. Around the feature position a ROI of fixed size is taken at each time step. If feature tracking is successful, the feature will be at the same position of the related object. This is the center of all stacked images. Since feature orientation can be unstable, we additionally consider the related cluster. The cluster orientation is stable, as all local features inside the cluster are used to estimate it. And this information can be used to stabilize the orientation of each feature before the current ROI image is stacked.

The idea of image stacking is to generate a pixelwise mean image for each local feature. Since the considered object is always at the same position during stacking, it will stay there in the mean image, too. The background may contain confusion objects such as strongly textured buildings close to the street or slower vehicles which can disturb object segmentation. As the background is moving along the whole stack, taking the pixelwise mean will generate a blurring effect as seen in Fig. 6. The image center is the position of the moving local feature coming from the central vehicle. In the lower row the stationary vehicle right next to the central vehicle is clearly visible in the first frame, but during the stacking process it gets more and more blurred. Even vehicles moving in the same direction as the central vehicle start blurring after some time as seen in both rows. Object segmentation is improved this way, as we will show later in the experiments.

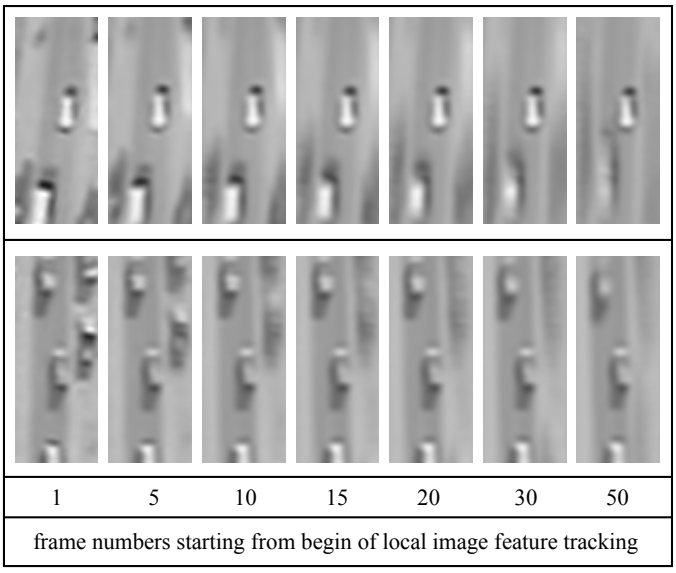

Figure 6. Temporal fusion (image stacking) for two local features, which are tracked along 50 frames. The central vehicle is moving and stacked.

The temporal fusion is used for object segmentation by constructing the image stack for each feature, finding the overlap between the feature-related and the cluster-related image ROI, and extracting the object hypotheses (segments) in this overlap area. Hence, there will be several bounding boxes for each object, which will be fused also in the second spatial fusion step. Only features with a stack higher than ten images are considered to ensure reliability and good quality of the image stack.

2) First Spatial Fusion: All approaches for accumulator image calculation generate similar but not identical results, so they can be combined in an accumulator fusion image. Therefore, two steps are needed: the normalization and the fusion strategy. Normalization of each accumulator aims to prevent one approach dominating the whole fused segmentation process by generally producing higher accumulation values than the other approaches. Thus, the value range is mapped to a fixed interval, for example $[0; 255]$. Standard methods for fusion strategies are pixelwise addition or multiplication. During some experiments we tested both fusion strategies together with several combinations of the five approaches.

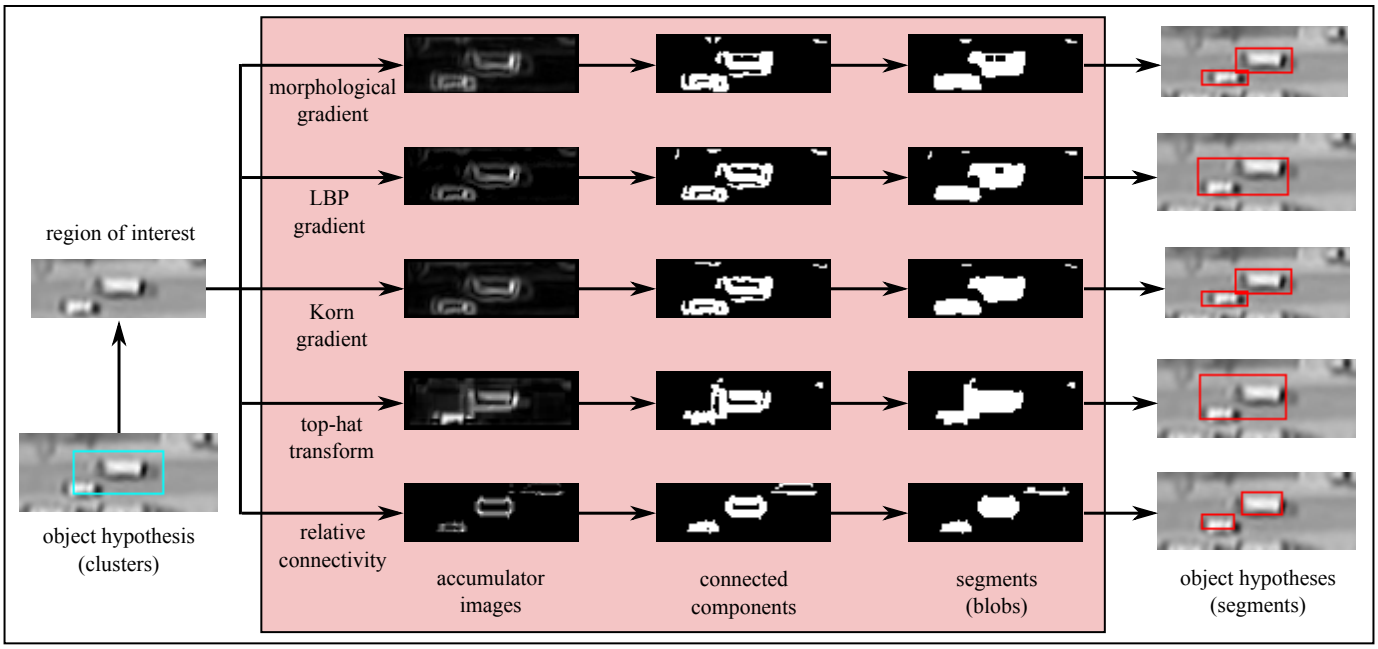

Figure 5. Object segmentation examples for each segmentation approach.

In summary, there was no big difference between the two strategies, but addition produced slightly better results than multiplication. The evaluation of the approach fusion showed that top-hat transform did not improve the results. All other approach combinations performed pretty similar and mostly better than the single approaches. Finally, the fusion of all approaches without top-hat transform produced the best results with our image data.

3) Second Spatial Fusion: The second spatial fusion is only needed if temporal fusion is applied. Without temporal information each object hypothesis is analyzed only once, as it is only covered by one cluster initially. When considering the local features with their image stacks, there will be many bounding boxes per object in the end. Most of them will be nearly at the same position in the image, but some of them can be incorrect. All bounding boxes strongly overlapping each other are clustered. For each bounding box cluster the median values for center point, width, and length are taken as final object hypothesis (segment). This way, outliers are eliminated.

IV. Experimental Results

In our experiments we evaluate both segmentation completeness and precision. Completeness is the rate of correctly found objects against the over- and under-segmented as well as the missed objects. An image sequence of 100 frames has been labeled manually for all appearing moving objects. The image size is 687×547 pixels and the mean object size in the image is about 15×5 pixels. Since it is difficult to find all moving objects, we only rely on the detected independent motion. Therefore, we assume that all objects which are not found by independent motion detection and clustering are moving too slow or not at all. This is reasonable, since no ground-truth is available and we mainly want to evaluate the relative performance between the approaches. In summary, 1535 objects have been labeled and Table I shows the segmentation completeness. The independent motion clustering is prone to under-segmentation, where several objects are merged together in one cluster. This is the reason, why advanced object segmentation is needed. The three gradient-based approaches perform very similar, but with spatial and especially spatio-temporal fusion much better results are achieved compared to each single approach. Objects are missed due to weak contrast to the background.

Table I
Evaluation of object segmentation completeness: correct, incorrect, and miss rates.

method	correct	under-seg	over-seg	miss
spatio-temporal fusion	0.85	0.04	0.04	0.07
spatial fusion	0.81	0.07	0.04	0.08
morphological gradient	0.78	0.09	0.03	0.10
LBP gradient	0.79	0.10	0.02	0.09
Korn gradient	0.77	0.08	0.05	0.10
top-hat transform	0.72	0.18	0.01	0.09
relative connectivity	0.62	0.04	0.10	0.24
independent motion clusters	0.56	0.35	0.09	0

For the evaluation of segmentation precision 15 objects have been labeled along 100 frames for position (center point x and y) and size (width w and length l). Table II shows the mean errors in pixels of the four object attributes for all objects along all frames. Position y and length l produce much higher errors than x and w, as the objects are moving vertically on the one hand and are affected by shadows in vertical direction on the other. The shadow of moving objects is moving as well,

Table II
EVALUATION OF OBJECT SEGMENTATION PRECISION: MEAN ERRORS IN PIXELS FOR POSITION (x, y) AND SIZE (w, l).

method	$\overline{e}_x$	$\overline{e}_y$	$\overline{e}_w$	$\overline{e}_l$
spatio-temporal fusion	1.38	1.82	3.72	7.70
spatial fusion	1.27	2.69	5.62	9.89
morphological gradient	1.29	2.94	5.45	10.04
LBP gradient	1.44	2.58	5.78	10.53
Korn gradient	1.33	2.99	5.82	10.42
top-hat transform	2.23	4.89	5.57	15.79
relative connectivity	1.22	3.65	4.63	10.32
independent motion clusters	2.32	5.37	8.68	26.69

so it is detected and clustered. Strong shadows even cannot be handled by object segmentation, since their appearance is similar to the one of dark vehicles. Modeling or learning object shading could help but is not considered in this work.

The same conclusions can be made as for object completeness evaluation: the three gradient-based approaches are working well but are clearly outperformed by spatial and spatio-temporal fusion. Relative connectivity has good precision, too, but already has shown a high object miss rate during completeness evaluation. The high errors of independent motion clustering are the result of under-segmentation. As the GSD is available, the mean errors can be converted from pixels to meters. For a GSD of 0.345 m/pixel the mean position error for spatio-temporal fusion is 0.47 m in x- and 0.63 m in y-direction as well as 1.28 m for w and 2.66 m for l. An example image visualizing the results of independent motion detection and custering as well as object segmentation including spatio-temporal fusion is given in Fig. 7.

V. CONCLUSION

In this paper we have presented a processing chain for precise segmentation of moving distant objects in video data coming from an UAV. Independent motion is detected and tracked by image registration using local image features. The clustering of independent motion provides object hypotheses which are prone to over- or under-segmentation. For the initialization of tracks and multi-target tracking in general this can be a problem. Thus, the focus of this work lies on advanced object segmentation using five different algorithms to extract image features which are suitable for object blob detection and bounding box generation. The resulting object hypotheses are better but significantly improved by the introduction of spatio-temporal fusion. Spatial fusion of the five segmentation algorithms improves the robustness. Furthermore, the local features are used for temporal fusion (image stacking) to blur and mask out the background of a moving object. On our UAV image data we achieved 85 % correctly segmented objects with a mean error of about 0.5 m for object position and about 2 m for object size. Moving object shading is the main problem since the moving shadows are detected, too, and have a similar appearance as dark objects.

REFERENCES

[1] M. Teutsch and W. Krüger, "Detection, Segmentation, and Tracking of Moving Objects in UAV Videos," in *Proceedings of the IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, Beijing, China, Sep. 2012, accepted.

[2] S. Zhang, H. Chen, Z. Chi, and P. Shi, "An algorithm for segmenting moving vehicles," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Hong Kong, China, April 6-10 2003.

[3] S. Zhang, "Object Tracking in Unmanned Aerial Vehicle (UAV) Videos Using a Combined Approach," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, PA, USA, March 18-23 2005.

[4] J. Li, D. Shi, and W. Yang, "Infared Target Extraction in FLIR Imagery based on Spatio Temporal Using Fuzzy Clustering," in *Proceedings of the Second International Symposium on Intelligent Information Technology Application (IITA)*, Shanghai, China, December 21-22 2008.

[5] O. Oreifej, R. Mehran, and M. Shah, "Human identity recognition in aerial images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, USA, June 13-18 2010.

[6] A. W. N. Ibrahim, P. W. Ching, G. Seet, M. Lau, and W. Czajewski, "Moving Objects Detection and Tracking Framework for UAV-based Surveillance," in *Proceedings of the Fourth Pacific-Rim Symposium on Image and Video Technology*, Singapore, November 14-17 2010.

[7] F. Yao, A. Sekmen, and M. J. Malkani, "Multiple moving target detection, tracking, and recognition from a moving observer," in *Proc. of the IEEE Intern. Conf. on Information and Automation (ICIA)*, Hunan, China, June 20-23 2008.

[8] H. Qi, W. E. Snyder, and D. Marchette, "Efficient approach to segment man-made targets from unmanned aerial vehicle imagery," *Optical Engineering*, vol. 39, no. 5, pp. 1267–1274, 2000.

[9] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, June 21–23 1994, pp. 593–600.

[10] R. Hartley and A. Zisserman, *Multiple-View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, March 2004.

[11] M. Teutsch, W. Krüger, and N. Heinze, "Detection and classification of moving objects from UAVs with optical sensors," in *Proceedings of SPIE Vol. 8050*, Orlando, FL, USA, Apr. 2011.

[12] M. Irani and P. Anandan, "A unified approach to moving object detection in 2D and 3D scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 6, pp. 577–589, Jun. 1998.

[13] C. Yuan, G. Medioni, J. Kang, and I. Cohen, "Detecting motion regions in the presence of a strong parallax from a moving camera by multiview geometric constraints," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1627–1641, 2007.

[14] R. Kumar, H. Sawhney, S. Samarasekera, S. Hsu, H. Tao, Y. Guo, K. Hanna, A. Pope, R. Wildes, D. Hirvonen, M. Hansen, and P. Burt, "Aerial video surveillance and exploitation," *Proc. of the IEEE*, vol. 89, no. 10, pp. 1518–1539, Oct. 2001.

[15] N. Heinze, M. Esswein, W. Krüger, and G. Saur, "Automatic image exploitation system for small UAVs," in *Proceedings of SPIE Vol. 6946, Airborne intelligence, surveillance, reconnaissance (ISR) systems and applications V*, Orlando, FL, USA, March 2008.

[16] E. R. Dougherty, *An Introduction to Morphological Image Processing*. SPIE - International Society for Optical Engineering, Jan. 1992.

[17] J. S. J. Lee, R. M. Haralick, and L. G. Shapiro, "Morphologic edge detection," *IEEE Journal of Robotics and Automation*, vol. 3, no. 2, pp. 142–156, April 1987.

[18] B. Chanda, M. K. Kundu, and Y. V. Padmaja, "A Multi-Scale Morphologic Edge Detector," *Pattern Recognition*, vol. 31, no. 10, pp. 1469–1478, October 1998.

[19] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, July 2002.

[20] T. Mäenpää, *The Local Binary Pattern Approach to Texture Analysis - Extensions and Applications*. Dissertation, University of Oulu, Finland, Aug. 2003.

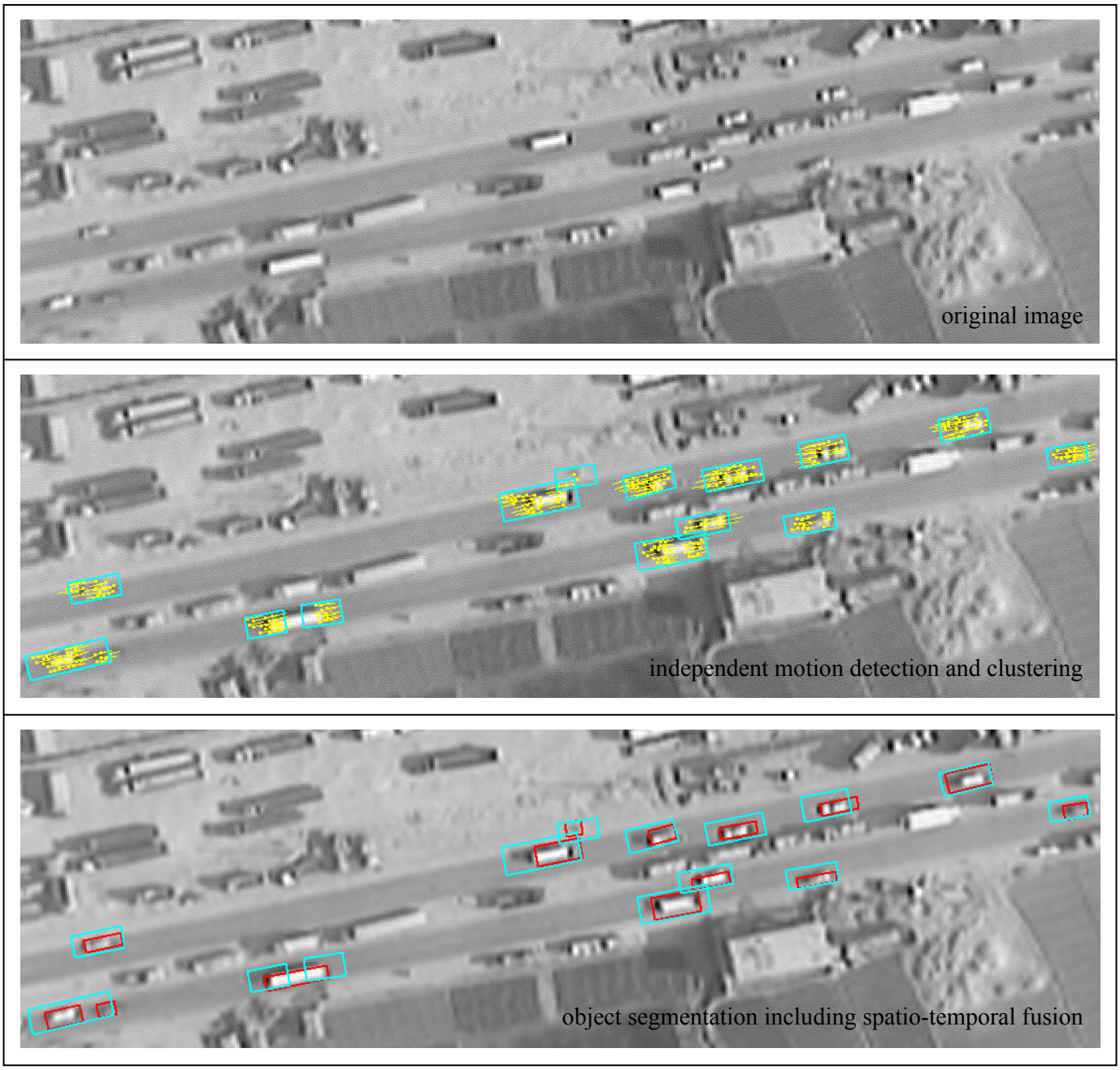

Figure 7. Examples for independent motion detection (yellow) and clustering (cyan) as well as object segmentation including spatio-temporal fusion (red).

[21] M. Teutsch and G. Saur, "Segmentation and Classification of Man-made Maritime Objects in TerraSAR-X Images," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Vancouver, Canada*, Jul. 2011.

[22] A. Korn, "Toward a Symbolic Representation of Intensity Changes in Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 5, pp. 610–625, September 1988.

[23] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[24] R. K. K. Yip, "Line Patterns Hough Transform for Line Segment Detection," in *Proceedings of the 1994 IEEE Region 10's Ninth Annual International Conference (TENCON'94)*, Aug. 1994, pp. 319–323.

[25] M. Teutsch and T. Schamm, "Fast Line and Object Segmentation in Noisy and Cluttered Environments Using Relative Connectivity," in *International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, Las Vegas, NV, USA, Jul. 2011.