Short-term change detection for UAV video

Günter Saur and Wolfgang Krüger

Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB)
Department Video Exploitation Systems (VID)
Fraunhoferstrasse 1, 76131 Karlsruhe, Germany

ABSTRACT

In the last years, there has been an increased use of unmanned aerial vehicles (UAV) for video reconnaissance and surveillance. An important application in this context is change detection in UAV video data. Here we address short-term change detection, in which the time between observations ranges from several minutes to a few hours. We distinguish this task from video motion detection (shorter time scale) and from long-term change detection, based on time series of still images taken between several days, weeks, or even years. Examples for relevant changes we are looking for are recently parked or moved vehicles.

As a pre-requisite, a precise image-to-image registration is needed. Images are selected on the basis of the geo-coordinates of the sensor's footprint and with respect to a certain minimal overlap. The automatic image-based fine-registration adjusts the image pair to a common geometry by using a robust matching approach to handle outliers.

The change detection algorithm has to distinguish between relevant and non-relevant changes. Examples for non-relevant changes are stereo disparity at 3D structures of the scene, changed length of shadows, and compression or transmission artifacts. To detect changes in image pairs we analyzed image differencing, local image correlation, and a transformation-based approach (multivariate alteration detection). As input we used color and gradient magnitude images. To cope with local misalignment of image structures we extended the approaches by a local neighborhood search.

The algorithms are applied to several examples covering both urban and rural scenes. The local neighborhood search in combination with intensity and gradient magnitude differencing clearly improved the results. Extended image differencing performed better than both the correlation based approach and the multivariate alternation detection. The algorithms are adapted to be used in semi-automatic workflows for the ABUL video exploitation system of Fraunhofer IOSB, see Heinze et. al. 2010.[1] In a further step we plan to incorporate more information from the video sequences to the change detection input images, e.g., by image enhancement or by along-track stereo which are available in the ABUL system.

Keywords: change detection, video exploitation, UAV, cross-correlation, image differencing, short-term change

1. INTRODUCTION

In many application fields, there has been an increased use of unmanned aerial vehicles (UAV) during the last years. In particular for video reconnaissance and surveillance, UAVs have been proven to be a flexible and useful platform. An important application in this context is change detection in UAV video data. Here we address short-term change detection, in which the time between observations ranges from several minutes to a few hours. We distinguish this task from video motion detection (shorter time scale) and from long-term change detection based on time series of still images taken between several days, weeks, or even years. Examples for relevant changes we are looking for are recently parked or moved vehicles.

Further author information: Correspondence eMail {guenter.saur, wolfgang.krueger}@iosb.fraunhofer.de

Image and Signal Processing for Remote Sensing XVIII, edited by Lorenzo Bruzzone,
Proc. of SPIE Vol. 8537, 85370R · © 2012 SPIE · CCC code: 0277-786/12/$18
doi: 10.1117/12.975156

The challange using small UAVs lies in the instable flight behavior and using low weight cameras. Thus there is a need to stabilize and register the videos by image based methods since using only direct methods based on positional information by GPS and attitude and acceleration measurements by an inertial measurement unit (IMU) are not accurate enough. As a pre-requisite for change detection, a precise image-to-image registration is needed. Images are selected on the basis of the geo-coordinates of the sensor's footprint and with respect to a certain minimal overlap. The automatic image-based fine-registration adjusts the image pair to a common geometry by using a robust matching approach to cope with outliers.

The change detection algorithm has to distinguish between relevant and non-relevant changes. Examples for non-relevant changes are stereo disparity at 3D structures of the scene and compression or transmission artifacts. Other influences to be suppressed come from illumination changes. Even for short-term scene-revisiting the illumination may have changed by progressed sun elevation or changes of cloud covering. In combination with changed sensor positions there will be changes in the shading of the objects' surface and its' contour lines as well as changes of shape and size of the objects' shadows.

A systematic survey of change detection algorithms is given in Radke et al. 2005[2] and Lu et al. 2004.[3] The majority of the algorithms can be classified into the main categories algebra, transformation, and classification. In addition, there are specialized methods for remote sensing applications like land-cover and vegetation monitoring. The category algebra includes image differencing, image ratioing, regression analysis, and statistical hypothesis testing for pixel values. Transformation methods are often applied to multi- and hyperspectral images. The aim is to reduce the redundancy between the available images channels and to find a combination of channels which enhance the relevant changes. A well-known transformation method is principal component analysis. An image transformation specifically designed for change detection is multivariate alteration detection.[4] Classification-based methods need to find or have available application-specific class labels for image regions. Change information can be extracted by comparing the different label images, but the quality of change detection results depends strongly on the performance of the classifier, which in turn depends on quality and quantity of the available training data.

In the following section we present the pre-processing steps for the video data including image selection and registration. Data sets from three different scenes with both urban and rural characteristics are prepared and the observable changes are described. In the section dedicated to the algorithms we address the three approaches images differencing, cross correlation and multivariate alternation detection. We introduce extensions for handling local displacements. The result images are presented and the observed effects are discussed. Finally we draw some conclusions and the paper ends with an outlook to further work.

2. VIDEO EXAMPLES AND DATA PRE-PROCESSING

Places to be monitored by UAVs are, e.g., bridges, access points to enclosed areas, or traffic lanes. By repeated overflights, these places can be surveyed with respect to changes of, e.g., recently parked vehicles. As examples, three typical scenes were taken from an UAV mission over an extended area with both urban and rural characteristics. Fig. 1 left shows the flight path overlaid on a satellite image where the sensor's footprint is covered by a dense video image sequence. The data selection and pre-processing is performed by the following steps:

- For a certain place to be monitored, the covering images of the UAV video are selected by their spatial attributes deduced from the sensor's footprint. Each observation of a relevant place corresponds to a short video sequence of a near-nadir overflight. For the work presented here, we select single image frames from each overflight with respect to a certain minimal overlap, see Fig. 1 center.
- Starting from the geo-coordinates of the sensor's footprint, an automatic image-based fine-registration adjusts the image pairs to a common geometry by using a robust matching approach. Basis are the sensor parameters and the UAV's IMU information which give the relative sensor orientation between

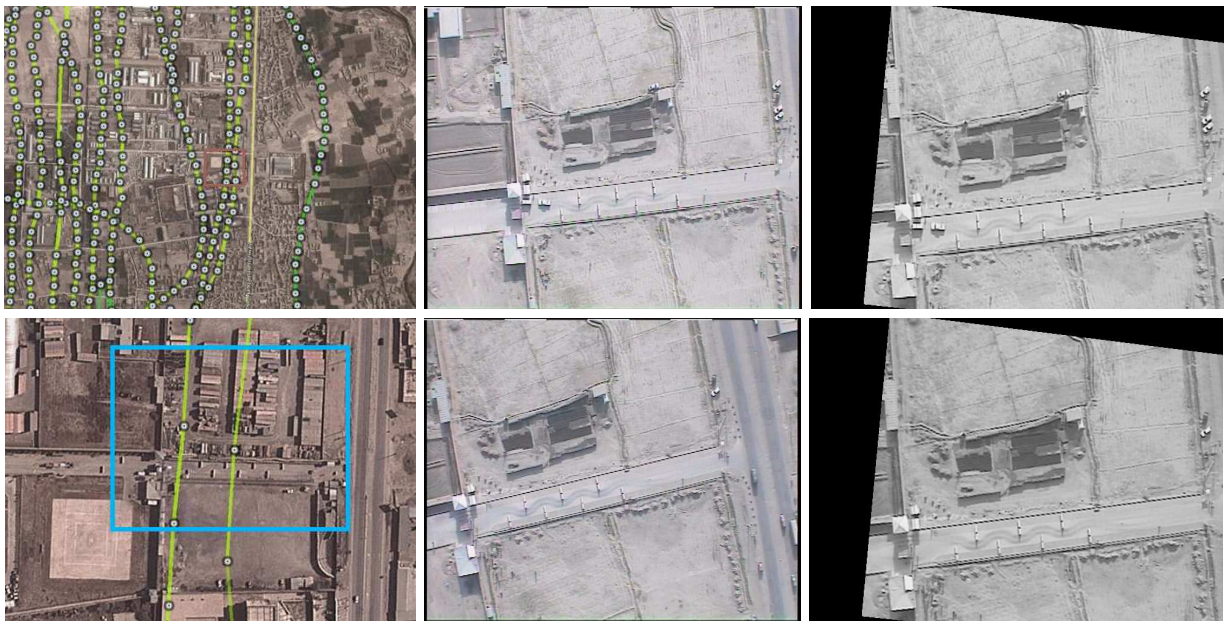

Figure 1. Frame selection for example scene *access point* from satellite image with overlaid flight path (left column), selected frames from two overflights (center), and registered images with non-overlapping areas masked out (right).

both image acquisitions. This relative orientation is used for image selection (to find images with a big overlap) and as pre-registration for an image-to-image matching algorithm. The result is a 2D homography transformation.

- One image is resampled onto the raster of the other. Some scene areas which are not covered by both images are labeled black, see Fig. 1 right. For evaluation purposes we clipped the images to get rid of the black corners, see Fig. 2 left.

To present the algorithms in the following section, we used three example scenes, see Fig. 2. In the example *access point* (Fig. 2 left), several cars have moved: on the road to the access point (left), in a parking area (right), and behind a building (on top). Several people have moved (to be recognized by their shadows), the length of shadows of the buildings has changed, and there are compression artifacts at the lines with high contrast along the road. In the example *road* (Fig. 2 center), the cars are relevant. In the example *bridge* (Fig. 2 right), the car on the road is relevant. Additionally, some people with their shadows are present in the images. In this image pair, a strong influence of 3D parallax comes from the building and the deep ditch the bridge is crossing.

3. ALGORITHMIC APPROACHES TO CHANGE DETECTION

Starting from commonly used approaches to change detection like image differencing and cross correlation, we introduce some extensions in order to cope with the main difficulties observed in video frames:

Local displacements: Object structures are displaced due to 3D parallax caused by different positions of the sensor at the time of image acquisition. The displacements often amount only to very few pixels and they are mixed with other effects like lens distortion or compression artifacts. Therefore the displacements are difficult to be used for stereo from motion algorithms. They are considered as noise and have to be suppressed by appropriate algorithm extensions.[5]

Figure 2. Registered and clipped color image pairs for change detection from example scenes *access point* in the left column, *road* in the center column, and *bridge* in the right column.

Illumination changes: Even if between the acquisitions of the two images only a few minutes have passed, the shading of the objects may have changed. Possible reasons are changes in cloud covering or slightly changed sensor positions which implies changes of the incidence angles of the observation ray at the objects's surfaces. An appropriate measure against these effects is using methods which are invariant to scale and offsets of image intensity like image differentation or selected texture features.[6]

In the following we propose three algorithmic approaches to change detection.

3.1 Differencing approach

Denoting by (f_{ij}) and (g_{ij}) the equal-sized gray scale images of the scene, the image (d_{ij}) of the absolute differences is derived by pixelwise differencing:

$$(d_{ij}) = (|f_{ij} - g_{ij}|)$$

To cope with local displacements, this approach is extended by a search in a neighborhood D_{ij} of each pixel (i, j) resulting in the extended difference image (e_{ij}) with

$$(e_{ij}) = (min_{(mn)\in D_{ij}}|f_{ij} - g_{mn}|).$$

In practice, as local neighborhood a small window of size 1×1 (no search) up to 11×11 is applied. Using absolute differences leads to dark areas where no changes occur and to bright areas otherwise. On the one hand, due to absolute values, we cannot distinguish between bright objects on dark background and dark objects on bright background. On the other hand, the images are handled symmetrically and, hence, it cannot be recognized in which one of the two images the objects appeared or disappeared. Color image are processed

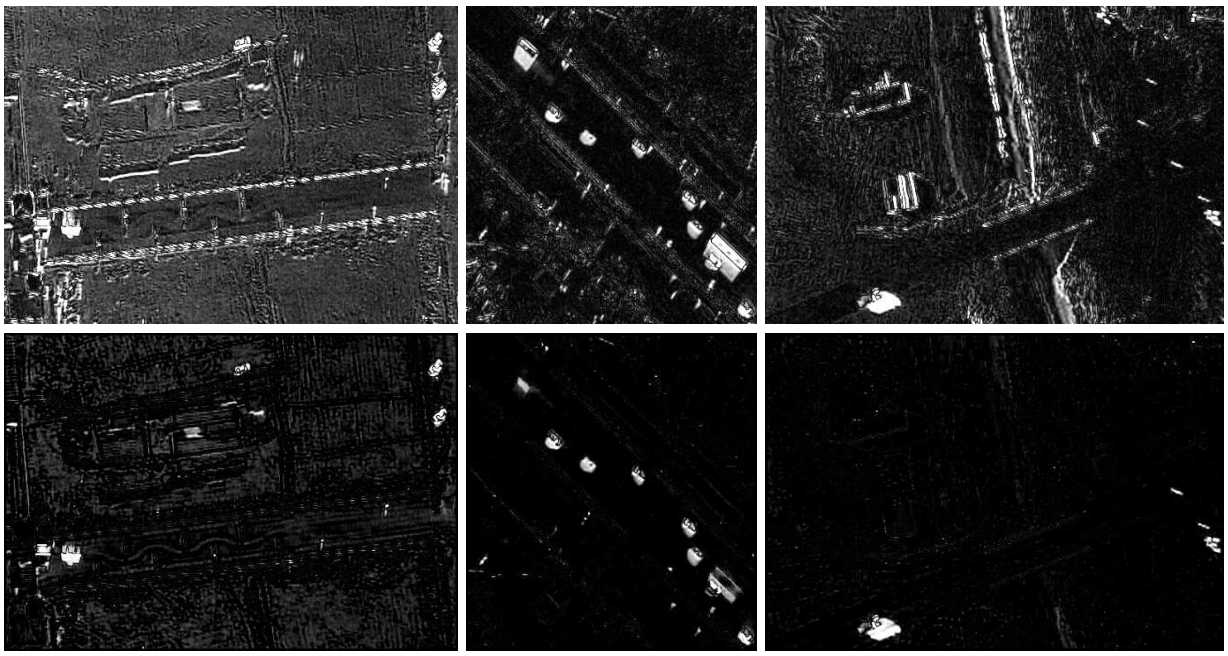

Figure 3. Absolute differences of gray values. In the second row a 7×7 search window is used.

after converting them into gray scale images using the arithmetic mean of the color channels. For visual inspection, the absolute differences are scaled up to get numbers in the range of the image gray values.

The results for our examples are shown in Fig. 3. Using no neighborhood search (first row), all the intensity disparities between the input images are still present in the difference images. When using a neighborhood of 7×7 (second row), smaller disparities vanish, e.g., the artifacts in the left image and the 3D effects in all images. The bright blobs at the places of changed vehicles get smaller, but they remain, while some smaller blobs indicating moved persons disappear. If we increase the size of the search window, even larger objects will vanish. In the left image (scene *access point*) the background is not all black since there is a shading difference between the original images.

An additional variation of image differencing consists in filtering the input images. Smoothing in combination with the search window has the effect of slightly decreasing the absolute differences. A means to overcome the effects of shading is image differentiation, e.g., using the gradient magnitude.

Fig. 4 shows results for absolute differences of the gradient magnitude. Using no search window (first row) has the same effect as before, all slight contrast disparities are still visible. For extended changed objects, only the contour line is present in the gradient magnitude image. When using a search window (second row), smaller change indications disappear, e.g., artifacts and 3D effects. As the contour blobs of the changed objects are getting smaller, they may get lost when using a larger search window. Unfortunately, the artifacts in the left image could not be suppressed completely when using the given parameters.

The combination of both gray value and gradient approach in Fig. 5 shows an interesting aspect. The results have been put to different color channels and we see that the change indications of both approaches are complementary. They supplement each other indicating diverse parts of the changed objects, i.e. the inner part (gray values) and the boundary part (gradient magnitude).

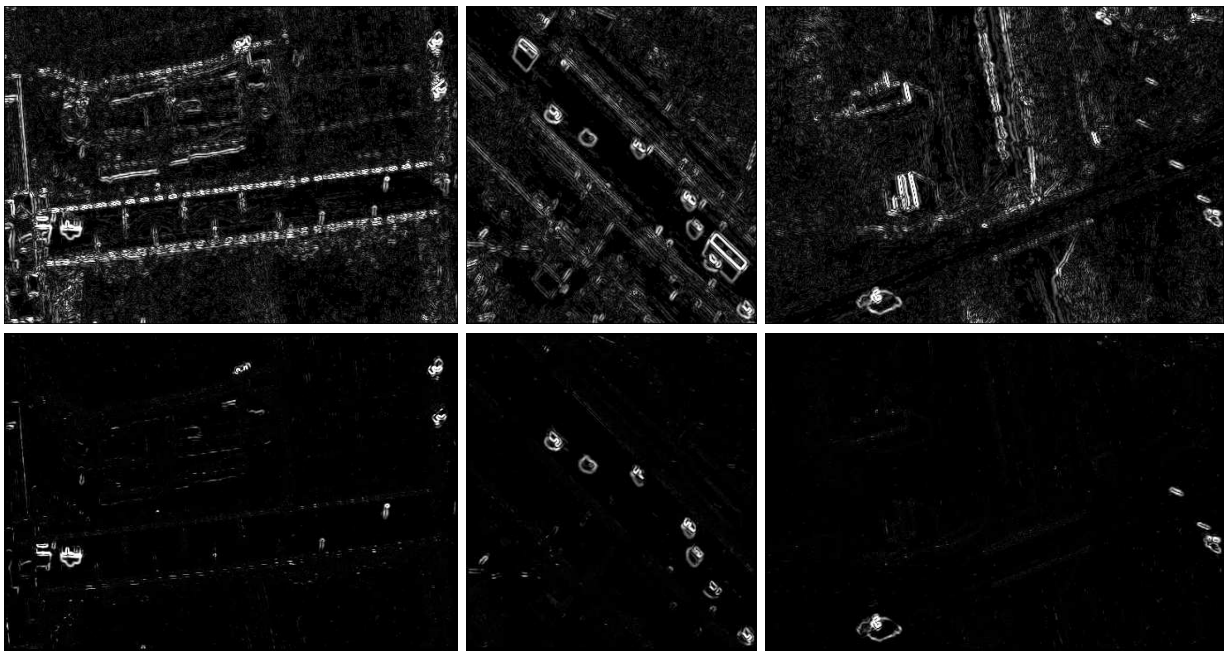

Figure 4. Absolute differences of gradient magnitudes, without (first row) and with 7×7 search window (second row).

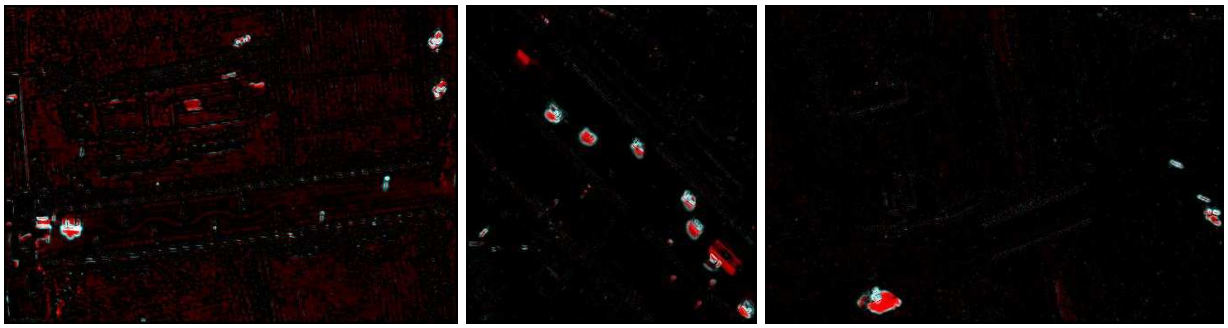

Figure 5. Composed difference images of gray values (red) and gradient magnitudes (cyan) using a 7×7 search window. Contributions from both components are mixed to white pixels.

3.2 Correlation approach

The normalized cross correlation (NCC) of two gray scale images (f_{ij}) and (g_{ij}) of same size is defined by their correlation coefficient in a neighborhood mask M_{ij}. This mask is moving over the images assigning to each center pixel (i, j) the correlation coefficient

$$(c_{ij}) = (\mathbf{cor}_{(kl)\in M_{ij}}(f_{kl}, g_{kl})) = \left(\frac{\sum_{(kl)\in M_{ij}} (f_{kl} - \overline{f}_{ij})(g_{kl} - \overline{g}_{ij})}{\sqrt{\sum_{(kl)\in M_{ij}} (f_{kl} - \overline{f}_{ij})^2 \sum_{(kl)\in M_{ij}} (g_{kl} - \overline{g}_{ij})^2}}\right)$$

denoting $\overline{f}_{ij} = (1/n) \sum_{(kl)\in M_{ij}} f_{kl}$, $\quad \overline{g}_{ij} = (1/n) \sum_{(kl)\in M_{ij}} g_{kl}$, $\quad$ and $n = \sum_{(kl)\in M_{ij}} 1$.

To cope with local displacements, this approach is also extended by a search in a neighborhood D_{ij} of each pixel (i, j) resulting in the extended correlation image

$$(k_{ij}) = (max_{(kl)\in D_{ij}} \mathbf{cor}_{(mn)\in M_{kl}}(f_{ij}, g_{mn})).$$

In fact, there are two local neighborhoods M_{kl} (cross correlation mask) and D_{ij} (search window) considered here and their effects and relations have to be analyzed. The correlation coefficients lie in the range $[-1, 1]$. Unchanged areas in the images have correlation coefficients near 1. From a practical point of view it is convenient to consider, e.g., the range $[0.75, 1]$ and scale it up to the gray value range $[0, 255]$ using inversion. The result are gray value images indicating changed areas by white blobs and unchanged areas by leaving them black. Using quadratic windows for M_{kl} and D_{ij} of size $(2m + 1)$ and $(2d + 1)$ respectively, the size of the image borders is $b = d + m$. Pixels inside this image border are not processed and are set to white.

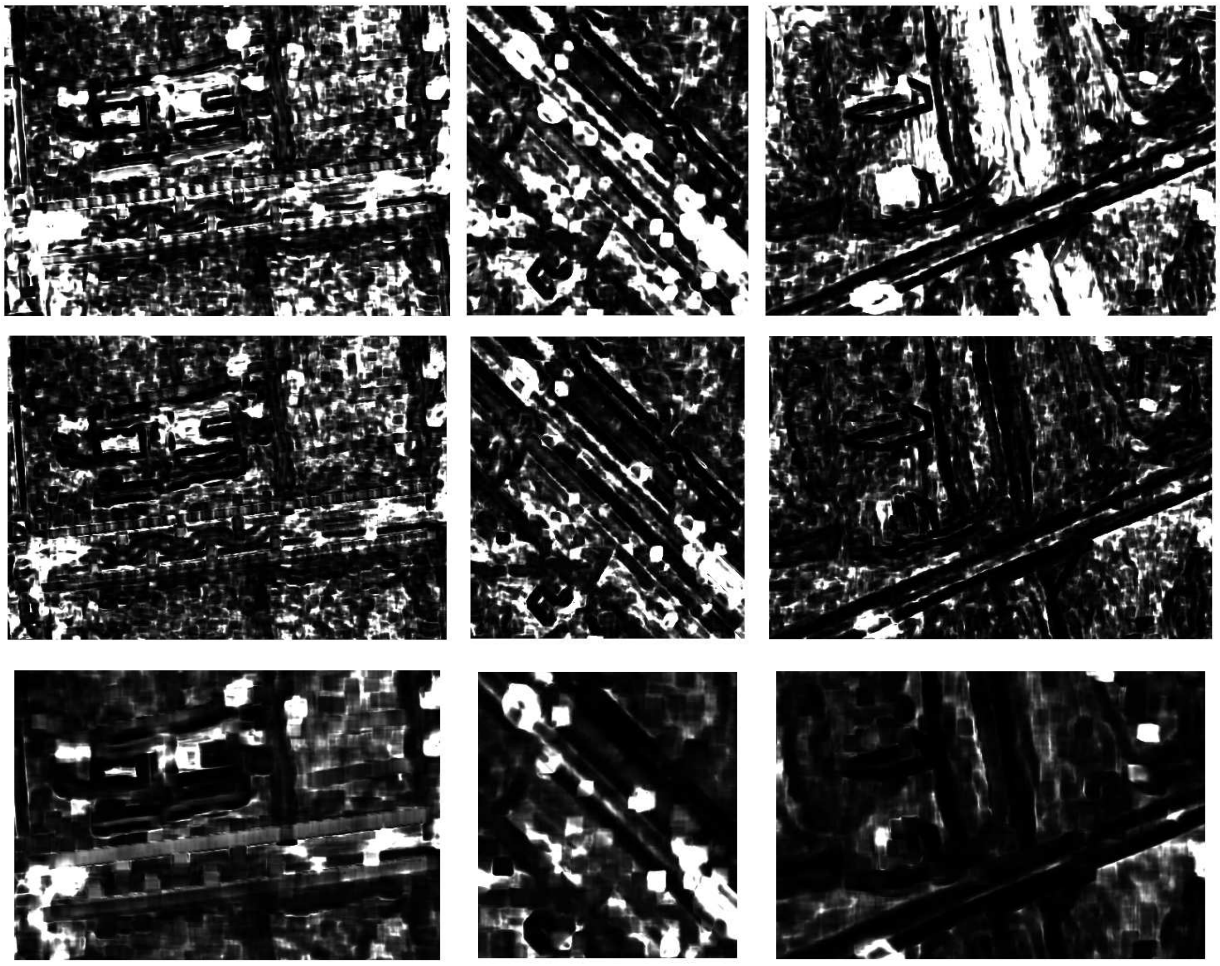

Figure 6. Correlation based approach using a 9×9 (first and second row) and a 17×17 (third row) mask. In the second row a 9×9 and in third row a 17×17 search window is applied. Note the white image borders of 4, 8 and 16 pixels.

Fig. 6 shows the results for our example scenes. In the first row, we used a correlation mask of 9×9 and applied no search window. It can be seen that many image details are left in the correlation image. Using smaller masks will cause even more details which are irrelevant for change detection. Applying a local search window (second row), some indications for irrelevant changes disappear, e.g., the 3D effects near the ditch in the right image. When using varying values d for the search window size, it can be observed, that a significant

influence of the search window will appear not until $d \geq m$, with m meaning the corresponding value for the correlation mask size.

In the third row of Fig. 6 even larger window sizes for search and correlation windows are used. It can be observed that many real changes are indicated by white blobs, but still the irrelevant changes have not been suppressed well. Additionally, there is an enormous loss of details since the resulting mask is very large. Furthermore, these big-sized windows lead to a significantly higher consumption of compution time compared to the approach using absolute differences.

3.3 Multivariate approach

In this section we present results of multivariate alteration detection (MAD) applied to our image data. MAD is a transformation-based change detection approach used in remote sensing.[4] The MAD algorithm is available* as a software extension for the image analysis system ENVI[4] and was used to compute the results in this section.

Given two N-channel images $\mathbf{F}$ and $\mathbf{G}$ of a scene, the MAD transformation yields N mutually uncorrelated difference images $D_1, \ldots, D_N$ with maximum variance.[4] The difference images are called MAD variates and are linear combinations of the input channels:

$$D_i = \mathbf{a_i}^\top \mathbf{F} - \mathbf{b_i}^\top \mathbf{G}, \quad (i = 1, \ldots, N)$$

The transformation coefficients $\mathbf{a_i}$ and $\mathbf{b_i}$ are computed by the MAD algorithm subject to the constraint that $\mathbf{a_i}^\top \mathbf{F}$ and $\mathbf{b_i}^\top \mathbf{G}$ have unit variance. The MAD variates are sorted in descending order of their variances, i.e. the first variate has the highest, and the last variate has the lowest variance in its pixel intensities. The algorithm also returns an image Z with chi-square values (sum of squares of the standardized MAD values) for each pixel:

$$Z = \sum_{i=1}^{N} \left(\frac{D_i}{\sigma_{D_i}} \right)^2$$

Assuming normally distributed and uncorrelated noise, the chi-square values for no-change pixels are expected to have a chi-square distribution with N degrees of freedom, i.e. sufficiently large chi-square values indicate change pixels. The results of the MAD algorithm are invariant with respect to linear transformations (scale and offset) of pixel intensities. Therefore, MAD analysis has the advantage of an implicit radiometric normalization.

The results of the MAD algorithm for our image data are shown in Fig. 7. Since our images are color RGB-images, we have three MAD variates ($N = 3$) and one chi-square image for each of the three input image pairs. We used the same image pairs as in the previous sections. Each column shows the results for one image pair. The rows show the MAD variates sorted in descending order of their variances (highest variance in first row). The last row contains the chi-square images. For visualization, the pixel intensities in each image are linearly stretched between the lower and upper one-percent quantile. It can be seen that the first two MAD variates contain the dominating large-scale changes between image which appear like blocking artifacts from JPEG compression. The small-scale changes at vehicles which are relevant for our application are concentrated in the MAD variate of highest order (third row) and resemble signed difference images. The chi-square images in the last row also reflect these changes. Our interpretation is that the change information from the first two MAD variates has less influence on the chi-square values because of the standardization with the larger variances of these variates.

However, large chi-square values appear not only at moved vehicles (wanted) but also at stationary infrastructure (unwanted). Examples are sampling and compression artifacts (e.g., road borders in images shown in left column) and registration errors at buildings or ditches caused by 3D parallax.

*http://mcanty.homepage.t-online.de/software.htm

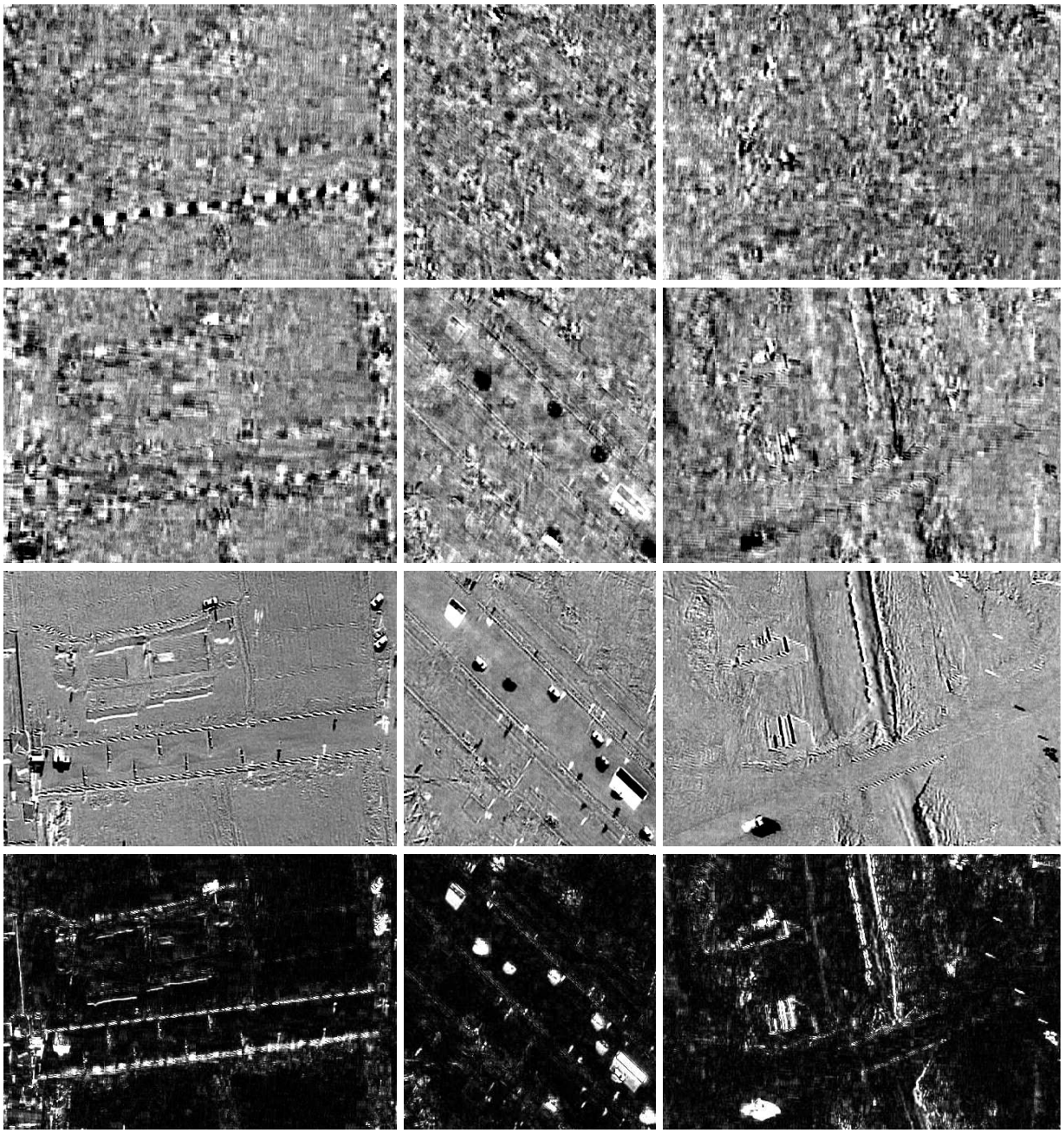

Figure 7. Results of MAD algorithm for our image data. Each column shows the results for one of three image pairs (cf. Fig. 2 for original input images). The first three rows show the MAD variates sorted in descending order of their variances (highest variance in first row). The last row contains the chi-square images.

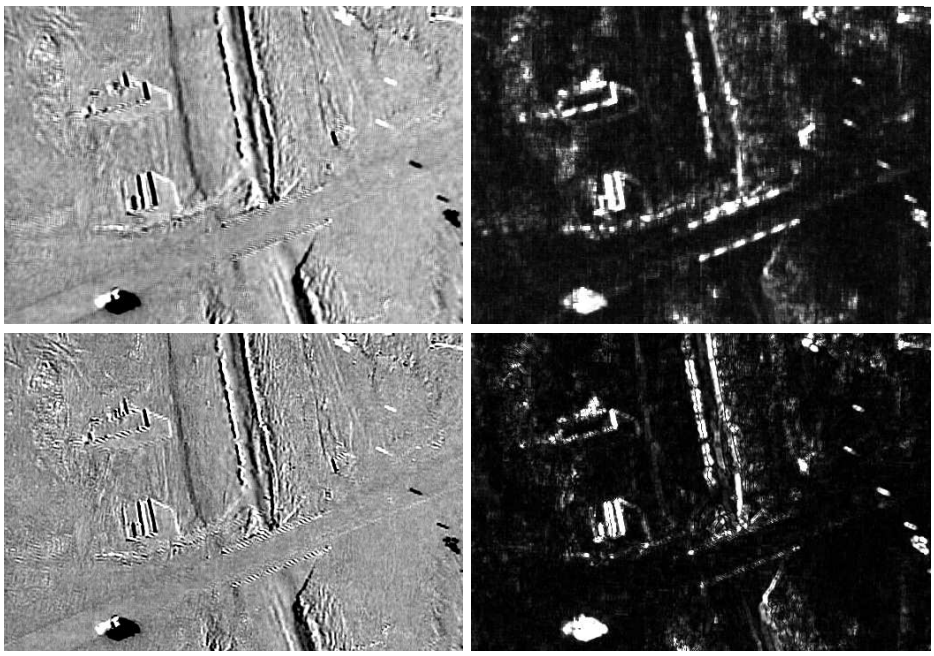

Figure 8. Results of MAD algorithm for 27-channel images with spatially shifted original channels (top row), and 4-channel images (gradient magnitude and the three color channels) at bottom.

In analogy to the extensions of the difference image approach in the previous sections, we tried to mitigate registration errors by adding spatially shifted versions of the original image channels to the MAD input. We tried shifts of ± 1 and ± 2 pixels in both image coordinates yielding a total of 27 and 75 channels for each input image, respectively. Results of the MAD algorithm for 27-channel images are shown in Fig. 8 (top row). The changes relevant for our application were still concentrated in the MAD variate of highest order (27). Adding shifted image channels resulted in blurred MAD variates and chi-square images, but did not reduce unwanted detections. The results for the 75-channel images and the two other image pairs were similar. Finally, we used the gradient magnitude image as an additional non-linear feature channel for the three RGB-channels. The results of the MAD algorithm for these 4-channel input images are also shown in Fig. 8 (last row). At the moment, we do not see how augmentation by additional channels would lead to enhanced robustness for our image data.

4. CONCLUSIONS AND FUTURE WORK

We proposed to extent image differencing for change detection by a local neighborhood search. This was applied to image intensities and to gradient magnitude. Our results show that false detections due to displacements from 3D parallax or image distortions could be widely suppressed. The approach based on cross correlation was extended also by a search in a local neighborhood. Its results for correlation are obviously worse than the previous results for image differencing. Additionally, the correlation needs computational extensive large mask sizes which lead to loss of object details in the images.

Image differencing of intensity images is not robust against shading variation of the images caused, e.g., by illumination changes in the scene. Using gradient magnitudes is invariant w.r.t. intensity offsets. Normalized cross correlation is invariant to offsets and to scaling.

The multivariate approach MAD is also invariant to intensity scaling and offsets. The small-scale changes which are relevant for our application are concentrated in the MAD variate of highest order (smallest variance).

The chi-square images resemble absolute intensity image differences. Since the multivariate approach is not robust against displacements, we tried to mitigate registration errors by adding spatially shifted versions of the original image channels to the MAD input. In contrary to the local neighborhood search used for image differencing, we did not observe any improvements.

Future work will be the generation of an explicit change mask. We think that a simple global threshold applied to the results of our algorithms will not be sufficient. The possibilities range from using an adaptive threshold over fusing several change result channels up to using a classifier which even can look back on the original images.

The algorithms are adapted to be used in semi-automatic workflows for the ABUL video exploitation system of Fraunhofer IOSB, see Heinze et. al. 2010[1] and Teutsch et al. 2011.[7] In a further step we plan to incorporate more information from the video sequences to the change detection input images, e.g., by image enhancement or along-track stereo, which are available in the ABUL system.

REFERENCES

[1] Heinze, N., Esswein, M., Krüger, W., and Saur, G., "Image exploitation algorithms for reconnaissance and surveillance with UAV," in [*Proceedings of SPIE Vol. 7668, Orlando, FL*], (Apr. 2010).

[2] Radke, R. J., Andra, S., Al-Kofahi, O., and Roysam, B., "Image change detection algorithms: A systematic survey," *IEEE Transactions on Image Processing* **14**, 294–307 (2005).

[3] Lu, D., Mausel, P., Brondzio, E., and Moran, E., "Change detection techniques," *International Journal of Remote Sensing* **25**(12), 2365–2401 (2004).

[4] Canty, M., [*Image Analysis, Classification and Change Detection in Remote Sensing, with Algorithms for ENVI/IDL*], Taylor & Francis, Second Edition: With Algorithms for ENVI/IDL ed. (2010).

[5] Bruzzone, L. and Cossu, R., "An adaptive approach to reducing registration noise effects in unsupervised change detection," *GeoRS* **41**, 2455–2465 (November 2003).

[6] Li, L. and Leung, M., "Integrating intensity and texture differences for robust change detection," *IEEE Transactions on image processing* **11**, 105–112 (February 2002).

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