Robust Tracking of People in Crowds with Covariance Descriptors

Jürgen Metzler, Dieter Willersinn
Fraunhofer Institute for Information and Data Processing (IITB)
Fraunhoferstrasse 1, 76131 Karlsruhe, Germany

ABSTRACT

In order to control riots in crowds, it is helpful to get the ringleader under control. A great support to achieve this task is the capability to automatically track individual persons in a video sequence taken from a crowd. In this paper we address the robustness of such a tracking function.

We start from the results of a previous evaluation of tracking methods, where a so-called Covariance-Tracker was found to be most appropriate. This tracker uses covariance matrices as object descriptors, as proposed by Porikli et al. The set of all covariance matrices describes a Riemannian manifold that is used to compare and update the covariance descriptors during tracking.

We propose Covariance-Tracker adaptations to improve its performance. Furthermore, we summarize the performance evaluation results of the original method and compare these with the results of the adapted one. The result is a robust method for tracking people in crowds which can improve situational awareness.

Keywords: covariance descriptor, people tracking, riot control

1. INTRODUCTION

In these days political demonstrations belong to civil liberty and take place all over the world. Unfortunately peaceful demonstrations sometimes turn over to become violent. In these situations surveillance systems can assist security guards to re-establish order. We defined a system which supports the tracking and seizure of an offender in a crowd.

In this contribution we focus on the tracking of individual persons in crowds from elevated observation points or unmanned aerial vehicles. Observing crowds, especially tracking people in crowds from moving platforms, is a major challenge in many computer vision applications. In past years, quit a number of tracking methods have been developed. One consequence is that there are today many different suggestions for object representations. Of course the representations depend on the sensor system, for example color information is not available from grayscale or infrared cameras. For this reason a generic object representation is of interest that can take any object feature as an input. Porikli et al. introduced such an object representation, so-called covariance descriptors. The idea is to describe spatial and statistical properties as well as the correlations of an image region by means of a covariance matrix. As pointed out in [8] there are several advantages of using covariance descriptors:

- Low-dimensional descriptors
- Support of scale invariant features/properties
- Invariant to mean changes (e.g. invariant to identical shifting of color values)
- Insensitive to noise
- Simple replacing of features
- Efficient fusion of multiple features

The authors obtained excellent results in several tracking tasks by using covariance descriptors [8], and in [5] it was shown that these descriptors are also suitable for object detections tasks. In [2] we evaluated the performance of the Covariance-Tracker for tracking people in crowds and compared the results to other tracking methods. This evaluation confirmed the advantage of covariance descriptors. These convincing results and the generic-like object representation encouraged us to concentrate on Covariance-Tracker.

Visual Information Processing XVIII, edited by Zia-Ur Rahman, Stephen E. Reichenbach, Mark Allen Neifeld,
Proc. of SPIE Vol. 7341, 73410T · © 2009 SPIE · CCC code: 0277-786X/09/$18 · doi: 10.1117/12.820067

In the following section we first summarize the mathematical background of tracking with covariance descriptors. In Section 3, we describe the original Covariance-Tracker of Porikli et al. and propose several adaptations. We then present our data set used in this work as well as summarize the results obtained by the original and the adapted tracker in Section 4. Furthermore, we show the results of our extended Covariance-Tracker.

2. MATHEMATICAL BACKGROUND

A covariance matrix contains information about statistical dispersions and linear relationships of random variables. Let $X_1, \dots, X_n$ be quadratic integrable random variables, with expected values $E(X_1), \dots, E(X_n)$ and

$$Cov(X_i, X_j) = E[\left(X_i - E(X_i)\right)\left(X_j - E(X_j)\right)], \; i = 1 \dots n, \; j = 1 \dots n, \qquad (1)$$

the pairwise covariances, the covariance matrix $\sum$ is then given by

$$\begin{pmatrix} Cov(X_1, X_1) & \cdots & Cov(X_1, X_n) \\ \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & \cdots & Cov(X_n, X_n) \end{pmatrix}. \qquad (2)$$

The set of positive definite covariance matrices (positive definite symmetric matrices) describes a Riemannian manifold and is denoted by Sym_n^+. A Riemannian manifold or Riemannian space is a topological space that is only locally Euclidean: there is a tangent space at each element of the manifold (in our case at each covariance matrix). So Euclidean geometry is not appropriate to compare covariance matrices. In past years functions like the trace or determinant have been used to measure the similarity. However, these measures are not suitable [1] and so Pennec et al. [7] and Foerstner et al. [1] deduced invariant Riemannian metrics. These metrics are equivalent, thus it is sufficient to concentrate on Pennec's:

$$< y, z >_{\Sigma_1} = tr(\Sigma_1^{-\frac{1}{2}} y \Sigma_1^{-1} z \Sigma_1^{-\frac{1}{2}}), \qquad (3)$$

$\sum_1$ is a covariance matrix and y, z are elements of the tangent space at $\sum_1$. y and z are determined by a diffeomorphism which maps elements of the tangent spaces into the manifold of covariance matrices. Associated to the Riemannian metric (3) it is defined by the exponential map

$$exp_{\Sigma_1}(y) = \Sigma_1^{\frac{1}{2}} \exp(\Sigma_1^{-\frac{1}{2}} y \Sigma_1^{-\frac{1}{2}}) \Sigma_1^{\frac{1}{2}}. \qquad (4)$$

The exponential map is global in Sym_n^+ and thus there is an inverse mapping (logarithmic map) which is uniquely defined everywhere:

$$log_{\Sigma_1}(\Sigma_2) = \Sigma_1^{\frac{1}{2}} \log(\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}}) \Sigma_1^{\frac{1}{2}}. \qquad (5)$$

It maps points of the manifold into tangent spaces, and by substituting y and z from equation (1) by $log_{\Sigma_1}(\Sigma_1)$ and $log_{\Sigma_1}(\Sigma_2)$, respectively, we get the following equation for Pennec's metric:

$$< y, z >_{\Sigma_1} = tr(log^2(\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}})) = < log_{\Sigma_1}(\Sigma_1), log_{\Sigma_1}(\Sigma_2) >_{\Sigma_1}. \qquad (6)$$

Covariance distance

The distances between elements of one tangent space are defined by the Euclidean distance. Thus, it can be used to determine the distance between any two covariance matrices mapped into one tangent space, if it is a tangent space at the position of one of either matrices. Under this condition the distance between two covariance matrices $\sum_1$ and $\sum_2$ is given by

$$d(\Sigma_1, \Sigma_2) > = \sqrt{< log_{\Sigma_1}(\Sigma_1), log_{\Sigma_1}(\Sigma_2) >_{\Sigma_1}}. \qquad (7)$$

Empirical mean value

There are several definitions of the (empirical) mean value for a set of measures of the same positive definite symmetric matrix [6]. One applicable mean is the so-called Karcher or Fréchet mean which minimizes the sum of the squared distances between the matrices. According to [3], [4] and [6] this mean exists and is even unique in Sym_n^+, as this manifold has a non-positive curvature [9].

We are interested in an empirical mean value which can be determined by a gradient descent algorithm [6]. To this, matrices are mapped into the tangent space first, where then the Euclidean mean is calculated. Eventually, the mean value of the covariance matrices is given by mapping back the Euclidean mean. Let $\sum_1 \ldots \sum_N$ be a set of N measures of the positive definite symmetric matrix $\overline{\sum}^t$, then the new mean $\overline{\sum}^{t+1}$ of this set is given by

$$\overline{\sum}^{t+1} = \exp_{\overline{\sum}^t}\left(\frac{1}{N}\sum_{i=1}^{N}\log_{\overline{\sum}^t}(\sum_i)\right). \qquad (8)$$

An important point for this algorithm is to determine a good starting point. If there is no $\overline{\sum}^t$, an element of $\sum_1 \ldots \sum_N$ can be selected randomly as starting point here. In the case of the original Covariance-Tracker it is the initial covariance descriptor (details can be found in Section 3).

Additionally, the matrices can be weighted differently, for example by their distances to the mean value [8]. In this contribution we didn't weight the matrices, since it hadn't resulted in significant enhancements.

Empirical covariance matrix

Let $\sum_1 \ldots \sum_N$ be a set of N positive definite symmetric matrices with the empirical mean value $\overline{\sum}$. According to [6] the empirical covariance matrix which is a generalization of the usual definition, is given by

$$Cov_{\overline{\sum}} = \frac{1}{N-1}\sum_{i=1}^{N} Vec_{\overline{\sum}}(\overrightarrow{\overline{\sum}\sum_i})\, Vec_{\overline{\sum}}(\overrightarrow{\overline{\sum}\sum_i})^T. \quad (9)$$

$Vec_{\overline{\sum}}$ is an isomorphism between the tangent space at $\overline{\sum}$ and $\mathbb{R}^{n(n+1)/2}$:

$$Vec_{\overline{\sum}}(\overrightarrow{\overline{\sum}\Lambda}) = Vec_{Id}(log(\overline{\sum}^{-\frac{1}{2}}\Lambda\overline{\sum}^{-\frac{1}{2}^{\mathrm{T}}})), \qquad (10)$$

where $Vec_{Id}(W) = \left(w_{1,1}, \sqrt{2}w_{1,2}, w_{2,2}, \sqrt{2}w_{1,3}, \sqrt{2}w_{2,3}, w_{3,3}, \ldots \sqrt{2}w_{1,n}, \ldots \sqrt{2}w_{(n-1),n}, w_{n,n}\right)^T$.

Mahalanobis distance and χ^2 distribution

Now we have definitions for the empirical mean and the empirical covariance matrix. Moreover, one may also define a Mahalanobis distance $\mu_{(\overline{\sum}, Cov_{\overline{\sum}})}$ in Sym_n^+ [6] which is well defined for any distribution of a random point $\sum \sim(\overline{\sum}, Cov_{\overline{\sum}})$:

$$\mu_{(\overline{\sum}, Cov_{\overline{\sum}})} = \sqrt{Vec_{\overline{\sum}}(\overrightarrow{\overline{\sum}\hat{\sum}})^T\, Cov_{\overline{\sum}}^{-1}\, Vec_{\overline{\sum}}(\overrightarrow{\overline{\sum}\hat{\sum}})},$$

where $Vec_{\overline{\sum}}$ is the same isomorphism as defined by (10). The Mahalanobis distance measures the distance between an observation $\hat{\sum}$ and the mean $\overline{\sum}$ according to $Cov_{\overline{\sum}}^{-1}$.

Furthermore, Pennec generalized the normal distribution for complete Riemannian manifolds such as Sym_n^+ by looking for the probability density function that minimizes the entropy with a constrained mean and covariance. Assuming the random point $\sum \sim(\overline{\sum}, Cov_{\overline{\sum}})$ is normal and $\hat{\sum}$ an observation, then the Mahalanobis distance should be χ^2 distributed if the observation is correct. Then a χ^2 test can be used for determining outliers (details about the normal and the χ^2 distribution as well as the χ^2 test can be found in [6]).

3. COVARIANCE-TRACKER

The main idea of the Covariance-Tracker of Porikli et al. is to describe image regions by covariance matrices [8]. For each region of interest (ROI) a number of features such as position, color and gradients of pixels is measured and coded in a covariance matrix. To this, a d-dimensional feature vector at each pixel inside the ROI is constructed first. As features we use the x- and y-coordinates of a pixel, RGB color values and gradient information.

Let $\{f_i\}_{i=1\ldots n}$ be a set of features vectors of a W-width and H-height rectangular ROI R_1 and

$$f_i = (x, y, R(x,y), G(x,y), B(x,y), I_x(x,y), I_y(x,y), I_{xx}(x,y), I_{yy}(x,y))^T$$

a feature vector at the pixel with the coordinates (x, y). The covariance matrix for R_1 is then given by

$$Cov_{R_1} = \frac{1}{WH}\sum_{i=1}^{WH}(f_i - \mu_R)(f_i - \mu_R)^T,$$

where μ_R is the mean-vector of $\{f_i\}_{i=1\ldots n}$.

Cov_{R_1}, also referred to as covariance descriptor, contains information about spatial and statistical properties of the object as well as linear correlations between these properties.

After the detection of the region of interest the covariance descriptor is computed and stored in an object model. In the next step the model is searched in a target image. To this, the original algorithm determines the descriptor that has the minimum covariance distance to the object model, and assigns its region center as the estimated object position. In order to adapt to illuminations or object variations the object model is additionally updated by calculating the mean of the last n found descriptors. The number n is, for example, depending on the frame rate or variation speed of the target object. Fig. 1 gives an overview of the Covariance-Tracker (details can be found in [8]).

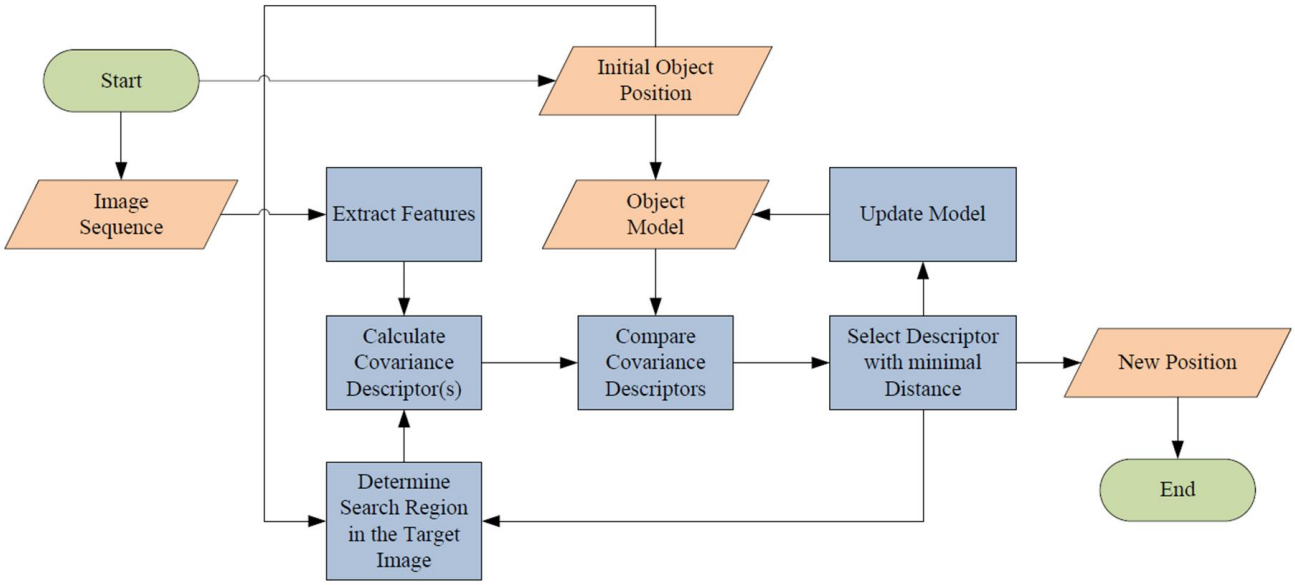

Fig. 1. Covariance-Tracker

The original Covariance-Tracker estimates the new object position by finding the *best* covariance descriptor (covariance descriptor that has the minimum covariance distance to the object model). Thereby, through noise or similar regions beside the object, it can happen that the best covariance descriptor does not contain the object anymore. Thus, the object model would be updated using a false descriptor and the tracker could lose the object. In order to avoid this, we additionally consider the neighbor regions in the object localization. Furthermore, we replaced the covariance distance by the Mahalanobis distance.

Generally, the covariance distances around the position of the best covariance descriptor weakly increase in direction away from this one, and moreover, they increase for different regions differently. The idea is now, to confirm respectively disconfirm the new object position by considering these observations. Therefore we introduce several conditions that the tracker has to fulfill.

Let $\mathcal{R}$ be the region around the position of the best covariance descriptor (descriptor that has the minimum Mahalanobis distance to the object model) with the radius r and let us define descriptors inside $\mathcal{R}$ that have a small Mahalanobis distance to the object model as *object-like-models*. Furthermore, let us assume that the covariance descriptors inside $\mathcal{R}$ are normal distributed around the best descriptor. Then the conditions are defined as follows:

1. The Mahalanobis distances inside $\mathcal{R}$ have to increase in direction away from the best covariance descriptor.
2. The number of *object-like-models* has to be higher than a threshold.
3. The number of *object-like-models* inside $\mathcal{R}$ may not be strongly changed during the tracking.
4. The variance of the descriptors inside $\mathcal{R}$ has to be small.

There are several ways to apply these conditions. In this work we update neither the object position nor the object model, if one is not fulfilled.

4. RESULTS

In performance evaluation we used data which we collected during a Crowd and Riot Control (CRC) training. For the acquisition of the supervised areas we installed cameras on elevated observation points. The cameras were installed 25 meter high on a crane platform and recorded data from nadir view (see Fig. 2). The CRC scenarios consisted of three different levels of escalation. First the crowd started with a peaceful demonstration. Later there were violent protests, and third, escalations of the riot where offenders bumped into the chain of guards.

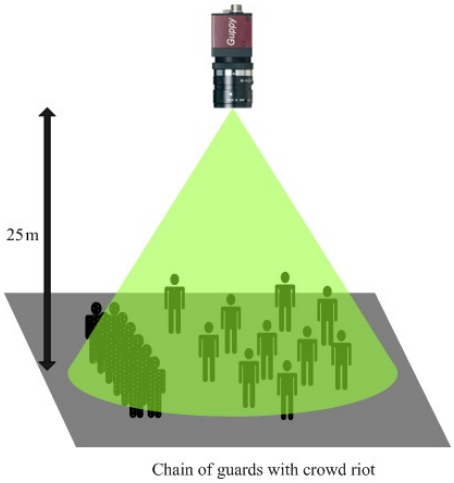

Fig. 2. Technical data acquisition details

In data acquisition we used a single-chip CMOS camera with a 20Hz sampling rate and a resolution of 752 x 480 pixels. The resolution of an individual is about 16 x 16 pixels. Each pixel of the image sensor chip captures only 8 bit: one of the three basic colors red, green or blue. The 24 bit color information for each pixel is computed in an embedded software module.

As a result of the nadir view there are only few object occlusions which simplify the tracking task. On the other hand, pictures from unmanned aerial vehicles contain a crowd consisting of small individuals which is a challenge for tracking methods (compare Fig. 3). Additionally, in our real data sets we have got to deal with variations of contrast of the individuals to the background. There are individuals with weak (see Fig. 4 Circle 1) and significant (see Fig. 4 Circle 2)

contrast to the background. Another challenge is the similarity of objects (see Fig. 4 Circle 3). As a consequence of fast changing lightning conditions there are local and global variations of luminance linked with strong shadowing effects (see right image in Fig. 3).

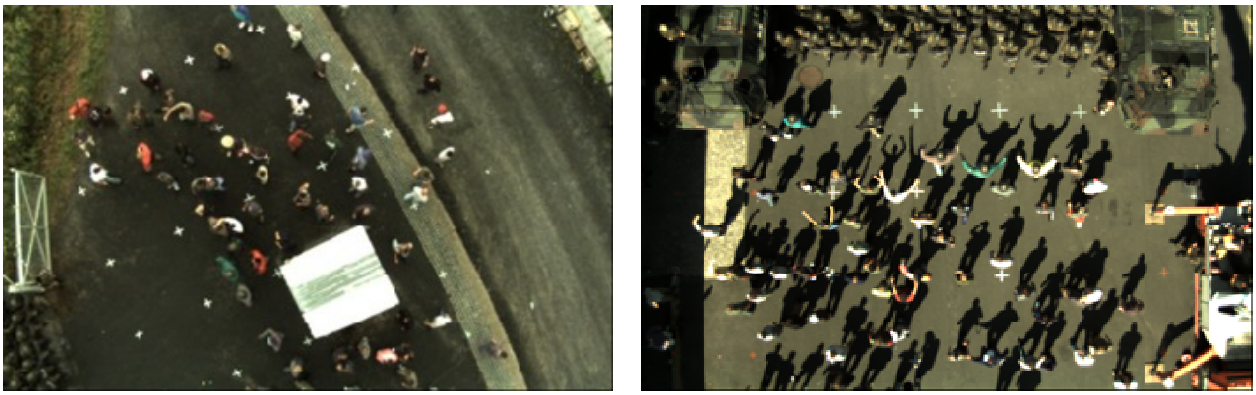

Fig. 3. Snapshots from the CRC data set

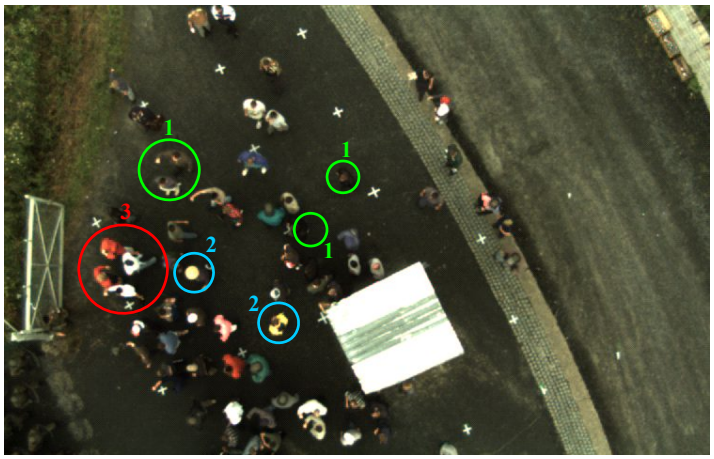

Fig. 4. Data set challenges

In [2] we evaluated the performance of three different tracking methods that are shown as suitable for tracking individual people in crowds from elevated observation points or unmanned aerial vehicles. The results are summarized in Table 1.

Metrics	KLT	Covariance	ColHist
Total ground truth	1000	1000	1000
False Negative Rate (1-f_n)	7,44%	3,74%	46,27%
True Positive Rate (f_n)	92,56%	96,26%	53,73%
OTE	11,0	9,8	31,0

Table 1. Performance evaluation overview. The evaluated tracking methods were a color based Kanade-Lucas Tracker (KLT), a Color-Histogram-Tracker (ColHist) and the original Covariance-Tracker (Covariance). The (original) Covariance-Tracker obtained the best True Positive Rate/Positive Prediction respectively False Negative Rate/False Alarm Rate and Object Tracking Error.

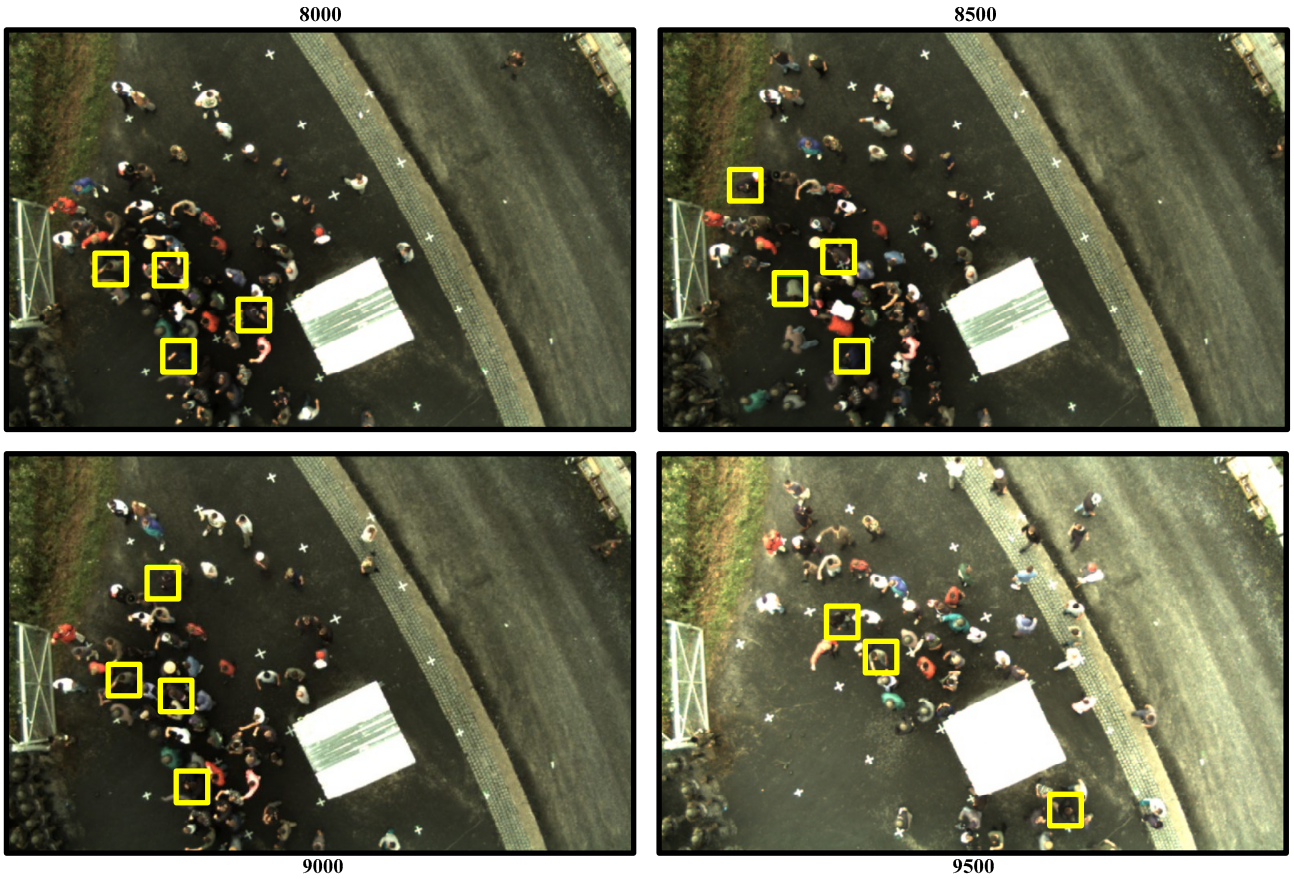

Fig. 5. Snapshots of the evaluated image sequence. Yellow squares show examples of challenging tracking tasks.

Here we concentrate on the Covariance-Tracker. We evaluated the performance of the original and our adapted method on **difficult** to track individuals (see Fig. 5). To this, we used the following metrics:

- False Negative Rate f_n: The relation between the numbers of not detected objects (N_{FN}) and the number of ground truth objects (N_{GT}):

$$f_n = \frac{N_{FN}}{N_{GT}}.$$

- The Object Tracking Error (OTE): The Euclidean Distance of the ground truth center of gravity and the hypothesis center of gravity for one ground truth object of an image:

$$OTE = \frac{1}{N_{rg}} \sum_{i \in g(t_i) \wedge r(t_i)} \sqrt{\left(x_i^g - x_i^r\right)^2 + \left(y_i^g - y_i^r\right)^2},$$

whereas N_{rg} is the number of images containing both ground truths and tracking results. x_i^g and y_i^g define the x- and y-coordinate of the ground truth center of gravity in the i^{th} image respectively x_i^r and y_i^r are the coordinates of the hypothesis center of gravity in the i^{th} image.

We obtained slight better results by replacing the covariance distance by the Mahalanobis distance. Furthermore, consideration of statistical conditions could further increase the performance. The results of the original and the adapted Covariance-Tracker (Mahalanobis distance + statistical conditions) are summarized in Table 2.

Metrics	Covariance	Adapted Covariance
Total ground truth	2000	2000
False Negative Rate (f_n)	44,21 %	25,09 %
True Positive Rate (1-f_n)	55,79 %	74,91 %
OTE	76,31	14,11

Table 2. Covariance-Tracker evaluation

5. SUMMARY

We proposed adaptations for the Covariance-Tracker of Porikli et al. and evaluated the original and adapted one on a riot control image sequence observed from an elevated observation point. The performance evaluation confirms the usefulness of tracking people with covariance descriptors in such scenarios. Furthermore, we showed that the consideration of statistical measures for covariance descriptors can further improve the performance. The result is a robust method for tracking people in crowds which can improve situational awareness.

ACKNOWLEDGEMENTS

The research reported in this contribution was funded by the Technical Center for Information Technology and Electronics (WTD 81) of the German Federal Office for Armament and Procurement (BWB).

REFERENCES

[1] W. Förstner and B. Moonen, “A Metric for Covariance Matrices”, Technical report, Department of Geodesy and Geoinformatics, Stuttgart University, 1999.

[2] Y. Hübner, J. Metzler, B. Dürr, U. Jäger and D. Willersinn, “Assessment and Optimization of Methods for Tracking People in Riot Control Scenarios”, Proc. of SPIE Security+Defence Europe, Cardiff, 2008.

[3] H. Karcher, “Riemannian center of mass and mollifier smoothing”, Comm. Pure Appl. Math., 30:509-541, 1977.

[4] W.S. Kendall, “Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence”, Proc. London Math. Soc., 61(2):371-406, 1990.

[5] S. Paisitkriangkra, C. Shen and J. T. Zhang, "An Experimental Evaluation of Local Features for Pedestrian Classification", Proc. of IEEE Conf. on Digital Image Computing Techniques and Applications, pp. 53-60, 2008.
[6] X. Pennec, "Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements", Journal of Mathematical Imaging and Vision (JMIV), 25:127-154, 2006.
[7] X. Pennec, P. Fillard and N. Ayache, "A Riemannian Framework for Tensor Computing", International Journal of Computer Vision (IJCV), 66:41–66, 2006.
[8] F. Porikli, O. Tuzel and P. Meer, "Covariance Tracking using Model Update Based on Means on Riemannian Manifolds", Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, New York, 2005.
[9] L. Skovgaard, "A Riemannian geometry of the multivariate normal model", Scand. J. Statistics, 11:211-223, 1984.