

EFFICIENT APPROXIMATION OF THE MAHALANOBIS DISTANCE FOR TRACKING WITH THE KALMAN FILTER

Pinho, R. R.^{*}; Tavares, J. M. R. S.^{**} & Correia, M. V.^{***}

^{*} FEUP - Faculdade de Engenharia da Universidade do Porto, INEGI - Instituto de Engenharia, Mecânica e Gestão Industrial, LOME - Lab. de Óptica e Mecânica Experimental

^{**} FEUP (DEMEGI), INEGI, LOME

^{***} FEUP (DEEC), INEB – Instituto de Engenharia Biomédica, LSI – Lab. de Sinal e Imagem
^{*,**,***} Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal

E-Mail: rpinho@fe.up.pt; tavares@fe.up.pt; mcorreia@fe.up.pt

Abstract

In this paper, we address the problem of tracking feature points along image sequences efficiently. Thus, to estimate the undergoing movement we use an approach based on Kalman filtering, which performs the prediction and correction of the features' movement in every image frame. Measured data is incorporated by optimizing the global association set built on efficient approximations of the Mahalanobis distance (MD). We analyze the difference between the usage in the tracking results of the original MD formulation and its more efficient approximation, as well as the related computational costs. Experimental results which validate our approach are presented.

(Extended paper from the International Symposium CompIMAGE – Coimbra, Portugal, 20-21 October 2006.)

Key Words: Tracking, Data Association, Mahalanobis Distance, Kalman Filter, Computational Vision

1. INTRODUCTION

Tracking features in image sequences is a complex problem since several specific problems may appear. In particular, one has to face difficult and ambiguous situations generated by cluttered backgrounds, occlusions, large geometric deformations, illumination variation or noisy data. On the other hand, with better computational and imaging resources, and a lot of research work done in object's tracking in the last decades, results are expected to be more accurate but also obtained in quicker processes, as some of the numerous applications of movement tracking (surveillance, analysis of object's deformation, traffic monitoring [1, 2]) require results in almost real-time. In the case of human's movement analysis it can be used for medical diagnosis, physical therapy (for instance, to study gait disorders related to knee or hip injury or pain) or in sports (for example, to control cycles of motion in training schemes).

Although the computational performance has improved significantly, tracking systems that are able to capture and analyze the undergoing movement, for instance to track simultaneously multiple features or to obtain results in real-time, often use some kind of simplification to speedup the computational process. So, a compromise must be achieved between the accuracy of the motion tracking and the related computational cost.

The Pfinder [1] is a real-time system for tracking humans and interpreting their temporal behaviour from image sequences. It does not require any accurate initialization procedure to segments the person from the background image and do its tracking in real-time using a standard computer workstation. However, that system expects that only one person is presented in the images and that the scene is almost static.

Bayesian networks are used in [3] to perform tracking even in the presence of occlusions, group formation and splitting. The on-line object tracking is performed by gradually discarding the influence of past information on the current decisions, avoiding with that a combinatorial explosion and keeping the network complexity within reasonable bounds.

On the other hand, many current works use a probabilistic representation of the uncertainty and stochastic filters to fuse and validate sensor data, and to estimate parameters describing the environment. The data association problem is combined with filtering in [4] to track ground targets using ground moving target indicator (GMTI) reports obtained from an airborne sensor.

The matching between the estimated features and the observations detected after a sensing operation is performed using data association techniques. Data association algorithms may include a hypothesis-validation step, which may be based on the Mahalanobis distance (MD) and its validation through the χ^2 -distribution.

1.1 Related Work

The use of the Mahalanobis distance for associating data in tracking systems is usual since it gives good performance, but few approaches have tried to speedup the related computational procedure. A method is presented in [5] for tracking and identifying moving individuals from image sequences taken by a fixed field-of-view camera, and MD is used to distinguish pixels either from image foreground or background. To speedup the computation procedure only the diagonals of the model's covariance matrix are used in the MD calculation. This corresponds to the same simplification we propose here, but applied in different tracking approaches and its results are not directly compared to those obtained with the original MD formulation.

On the other hand, MD and its simplified formulations are extensively used in pattern recognition. As the computation time of the original MD formulation will reach $O(n^2)$ for n -dimensional feature's vectors, to reduce the computational cost several approximations of the MD have been proposed as, for example, the quasi-Mahalanobis distance [6], the modified Mahalanobis distance [7], and the modified quadratic discriminant function [8]. All these approaches were proposed to improve the recognition accuracy, but not to approximate the quadratic discriminant function.

An approximation of the MD is used in [9] in Chinese and Japanese character recognition. Based on two kinds of approximations of the MD, the proposed algorithm has two stages: feature vector division and dimensional reduction. The stage of feature division is based on the characteristic of the model's covariance matrix. The second stage is done by regarding the values of small eigenvalues as constants. When compared to the well-known dimensional reduction method, the Karhunen-Loeve (K-L) expansion, experimental results showed that the proposed algorithm not only reduce the computational cost, but also improve the recognition accuracy. However, this approach requires the computation of the involved eigenvalues.

1.2 Our Approach

In our previous work we used the Mahalanobis distance to evaluate the quality of data's correspondence and in the association of predictions with measurements, by optimizing the sum of all the involved MD [10]. By doing so, the best global correspondence set is always guaranteed.

To track the acquired movement we use a well-known statistical technique: the Kalman filter (KF). The drawbacks of this stochastic procedure are due to its relatively high restrictive assumptions [11]. But combining the KF with optimization techniques for data association we increment the filters' robustness to situations of occlusion and non-linear movement.

As mentioned above, the correspondence between each features prediction and new measurement data is set upon the MD minimization. The MD ensures that the correspondence is done according to the known behaviour of each feature. Its approximation to the χ^2 -distribution allows the choice of a significance level, from which features will be considered as unmatched. Therefore, even if the KF restrictions are not satisfied (a very often situation in many real tracking applications) the results obtained with our tracking system may be corrected with the use of optimization techniques and MD.

In this paper we propose an efficient approximation of the original MD formulation, in order to sort and match corresponding features for their tracking along image sequences. Experimental results which validate our approach are also presented.

1.3 Paper Overview

This paper is organized as follows. In the next section a brief introduction is made to the MD. Then, in section 3, we describe the used efficient approximation to the original MD formulation and analyze the undergoing error. In section 4, we overview our tracking system, which uses the efficient approximation of the MD to find features' correspondence along image sequences. Some experimental results are shown in section 5, both considering synthetic and real image sequences. In the last section of this paper some conclusions will be held.

2. THE MAHALANOBIS DISTANCE

The Mahalanobis Distance, also known as statistical distance, is a distance that for each of its components (the variables) takes their variability into account when determining its distance to the corresponding centre. So, components with high variability receive less weight than components with low variability. This is done by rescaling the components considered; that is, for two points $X_i = (x_{1i}, x_{2i}, \dots, x_{ni})$ and $Y_j = (y_{1j}, y_{2j}, \dots, y_{nj})$ the MD is given by:

$$d_M = \sqrt{(X_i - Y_j)^T C^{-1} (X_i - Y_j)}, \quad (1)$$

where $C_{(n \times n)}$ is a non-singular covariance matrix (therefore symmetric positive definite). Features with the same distance of X_i satisfy the equation of an ellipsoid centred on X_i , and those features of the ellipsoid defined by C have unitary MD value.

The MD is a standard approach used in data association for tracking features along image sequences. However, this step is one of the most time-consuming operations of the computational process. After an operation of data acquiring, are available M features' location estimations and N measurements. The problem is how to associate each measurement, X_i ($i = 1, \dots, N$), with the correspondent estimated feature, Y_j ($j = 1, \dots, M$).

This association procedure is higher time-consuming because it involves a matrix inversion, and the determination of matrix C and vector $v = X_i - Y_j$ that is subject to linearization. To save computational cost, some data-association techniques perform a validation test for each match hypothesis in order to work with only a reduced set of hypotheses. That validation procedure can be performed using a statistical test based on the MD and its approximation by the χ^2 -distribution:

$$v^T C^{-1} v \leq \chi^2, \quad (2)$$

where v is the vector between a predicted feature state and an acquired measurement. This test should theoretically be computed for $M \times N$ hypotheses.

3. AN EFFICIENT APPROXIMATION OF THE MAHALANOBIS DISTANCE

In our tracking system the acquired measurements are composed by their position coordinates in the image plane, and each of the tracked features has its own KF. This means that the matrix C and vector v are given by:

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{bmatrix}, \quad (3)$$

and

$$V = [v_1 \quad v_2]^T. \quad (4)$$

So, in this case, the original MD formulation is given by:

$$d_M = \frac{c_{22}v_1^2 - 2c_{12}v_1v_2 + c_{11}v_2^2}{c_{11}c_{22} - c_{12}^2}. \quad (5)$$

Rearranging the terms in the equation above, we can obtain:

$$d_M = \frac{v_1^2}{c_{11}} + \frac{v_2^2}{c_{22}} - \frac{2c_{12}v_1v_2}{c_{11}c_{22}} + \frac{c_{12}c_{22}v_1^2 + c_{12}c_{11}v_2^2 - 2c_{12}^3v_1v_2}{c_{11}c_{22}(c_{11}c_{22} - c_{12}^2)}. \quad (6)$$

The efficient approximation we propose for the MD consists on:

$$d_M = \frac{v_1^2}{c_{11}} + \frac{v_2^2}{c_{22}}, \quad (7)$$

which is thereby affected by the error:

$$\Delta \hat{d}_M = -\frac{2c_{12}v_1v_2}{c_{11}c_{22}} + \frac{c_{12}c_{22}v_1^2 + c_{12}c_{11}v_2^2 - 2c_{12}^3v_1v_2}{c_{11}c_{22}(c_{11}c_{22} - c_{12}^2)}. \quad (8)$$

The approximation of (7) uses only the diagonal elements of the covariance matrix, and so it may round up or down the Mahalanobis distance.

Although better approximations can be used for the MD, their computational cost can be questioned: the approximation that we propose is quite efficient as it only involves 5 arithmetic operations for each pair of features; instead of the 18 operations involved in the original MD formulation given by (5).

4. TRACKING WITH THE KALMAN FILTER AND MATCHING DATA USING THE MAHALANOBIS DISTANCE

The Kalman Filter is an optimal recursive Bayesian stochastic method. It provides *optimal* estimates that minimize the mean of squared error of the modelled process. In a Bayesian stochastic viewpoint, the filter propagates conditional probability density of the system state conditioned on the knowledge of the actual data acquired by the measuring devices used.

The equations for the KF fall into two steps: time update (or prediction) and measurement update (or correction). The time update equations are responsible for projecting forward (in time) the current state and error covariance to obtain the *a priori* estimates for the next time step. The measurement update equations deal with the feedback; thus, new measurements are incorporated into the *a priori* estimates to obtain improved *a posteriori* values [11].

The prediction step is based on the Chapman-Kolmogorov equation for a first order Markov process:

$$x_t^- = \Phi x_{t-1}^+, \quad (9)$$

where Φ relates the system state x_{t-1}^+ at the previous time step $t-1$ to the state x_t^- at the current step t . The superscripts $^+$ and $^-$ indicate if measurement data have been incorporated or not, respectively. The related uncertainty is given by:

$$P_t^- = \Phi P_{t-1}^+ \Phi^T + Q, \quad (10)$$

where P is the covariance matrix and Q models the process noise.

The correction equations, that update the predicted estimates upon the incorporation of new u_t measurements, are given by:

$$K_t = P_t^- H^T [H P_t^- H^T + R_t]^{-1}, \quad (11)$$

$$x_t^+ = x_t^- + K_t [u_t - H x_t^-], \quad (12)$$

$$P_t^+ = [I - K_t H] P_t^-, \quad (13)$$

where K is the filter gain that minimizes the *a posteriori* error covariance equation, H processes the coordinates transformation between the predicted space and the measurement space, R_t is the measurement noise involved, and I is the identity matrix [11].

One of the drawbacks of the KF is the restrictive assumption of Gaussian *posterior* density functions at every time step, as many tracking problems involve non-linear movement. To minimize such ambiguities, we evaluate all the $M \times N$ possible matches by estimating the MD, and optimize the global matching set.

The efficient MD can sort matches according to their MD value, and so the computational burden associated to the MD calculation in the matching process is overcome with a reduced error involved.

5. EXPERIMENTAL RESULTS

In each frame of the presented examples the tracked features correspond to the centre of mass of each segmented region of movement. Each predicted position is represented with a '+', with the corresponding uncertainty area centre in the predicted position and circumscribed by a solid ellipse, each measurement is the centre of the segmented region, and each corrected position is represented using an 'x'. The associations between predictions/measurements are represented by solid lines.

For the first example, Fig. 1, consider a synthetic sequence of 9 frames, in which features are easily segmented using a simply threshold image operation. In the beginning of the image sequence only two blobs are visible. The circular blob will disappear definitively but the tracking approach keeps on trying to track it during the subsequent frames, although with gradually higher uncertainty (in frame (e), the uncertainty region surpasses the image border). (A detailed description of the model used in our tracking system to deal with missing features can be found in [12, 13].) In the second frame, a triangular blob appears, and in the third frame the square blob instantly disappears. In the fourth frame, the captured blobs overlap, and with the used image processing technique only one measurement is acquired and associated to a blob, but both features are continually correct tracked. From the seventh frame forwards the last one, 25 blobs are successfully tracked. The results presented in Fig. 1 were obtained by data association using the efficient MD but there are not any visual differences with those obtained with the original MD formulation; however, the computational cost associated to the efficient MD is obviously reduced, Fig. 2. Indeed, when few features are tracked the computational load of the original MD formulation is not significant, but as the number of tracked features increases the advantages of the efficient MD are more and more notable.

In Tables I and II are represented the results obtained with the efficient approximation of the MD and those obtained with the original MD formulation, respectively, to associate the tracked features and the measurements in frame (d) of Fig. 1. It can be noticed that the efficient approximation of the MD rounds up or down the complete value, but the differences are generally very small, Fig. 3.

Note that in frame (d) of Fig. 1, three features are being tracked and only two measurements are acquired, so to apply the optimization algorithm a fictitious variable was included and the cost of correspondence with it was defined as null [10].

Minimizing the overall correspondence costs evaluated in our tracking system, the same results are obtained using either the original MD formulation or its efficient approximation: the third tracked feature, Xe_3 is associated to the first measurement acquired, Xm_1 ; the second tracked feature and measurement are matched, Xe_2 and Xm_2 , respectively; and the first tracked feature Xe_1 is associated to the fictitious variable Xm_3 , which means that it is considered unmatched (Tables I and II).

What is more advantageous to our tracking system is that the efficient approximation of the MD can sort out the features to build the correct correspondences along the image sequence in an efficient and adequately manner.

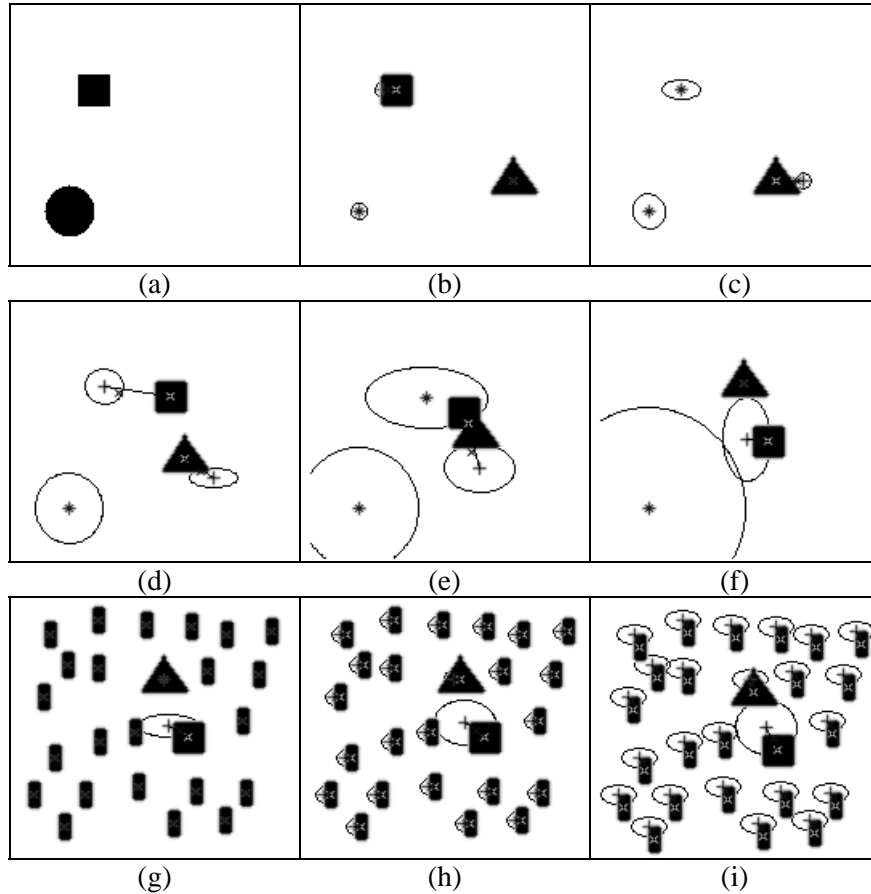


Figure 1: Tracking blobs in a 9 frames image sequence: (a) - original 1st frame; (b) to (i) - KF results: search areas defined by solid ellipses, the predicted position for each marker is represented by a '+', and the corrected position is indicated using an 'x'.

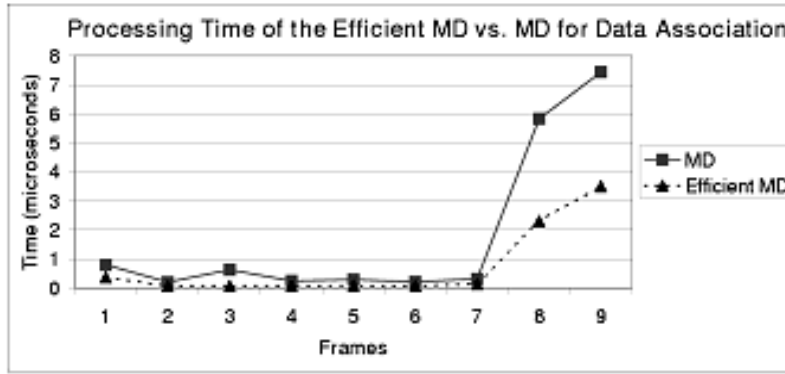


Figure 2: Time consumed in data association using the efficient MD or the original MD formulation on a computer with a Mobile AMD Athlon (tm) 4 at 1.20 GHz and 256 MB of RAM.

Table I and Table II: Associating tracked features and measurements in frame (d) of Fig. 1.

Values of the efficient approximation of the MD.

T. I	Measurements		
Tracked Features	Xm_1	Xm_2	Xm_3
Xe_1	5.635	7.248	0.0
Xe_2	3.910	1.678	0.0
Xe_3	0.406	2.570	0.0

Values of the original MD formulation.

T. II	Measurements		
Tracked Features	Xm_1	Xm_2	Xm_3
Xe_1	5.930	7.816	0.0
Xe_2	3.652	1.657	0.0
Xe_3	0.382	2.416	0.0

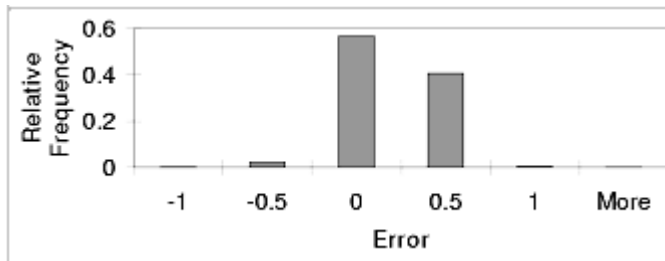


Figure 3: Relative frequency of errors in the efficient approximation of MD.

For the next experimental example, consider the sequence of real images presented in [14] for the evaluation of surveillance systems' performance. In this example, a person walks into the room and stops at the centre; meanwhile, two other persons enter, walk around the room and exit; then, the remaining person walks out of the room.

One usual difficulty related with persons' tracking using low level image processing techniques is the detection of 2 regions of movement, one of them corresponding to the peoples' head and upper body, and the other to their legs and feet. In order to test the proposed data association algorithm, in the images presented we did not try to remove all the existing noisy features. To detect the regions of motion we used image background subtraction and subsequent image frame subtraction with erode and dilation image operators to partially remove the noise presented.

We tracked the moving features using both the original MD formulation as well as its efficient approximation and, generally, the same results are obtained, Fig. 4. But in some uncommon cases the data association may differ, as in Fig. 5, but even in those cases the tracking process is correctly recovered in the next frames.

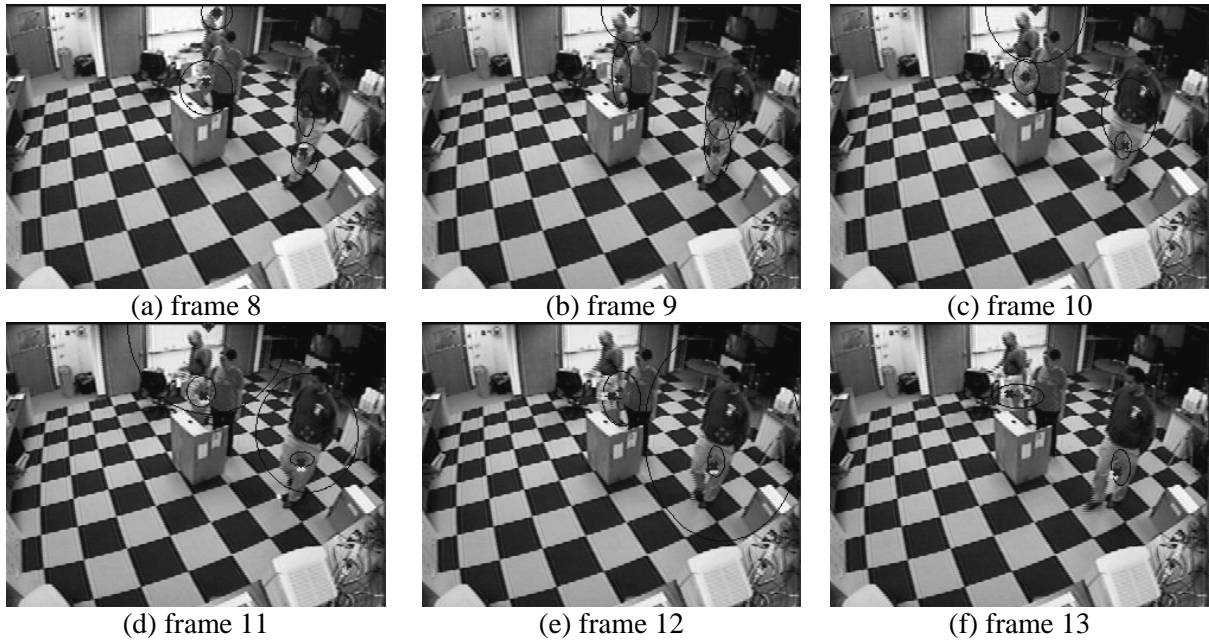


Figure 4: Tracking persons in a surveillance image sequence: results obtained with data association using the efficient approximation of the MD.

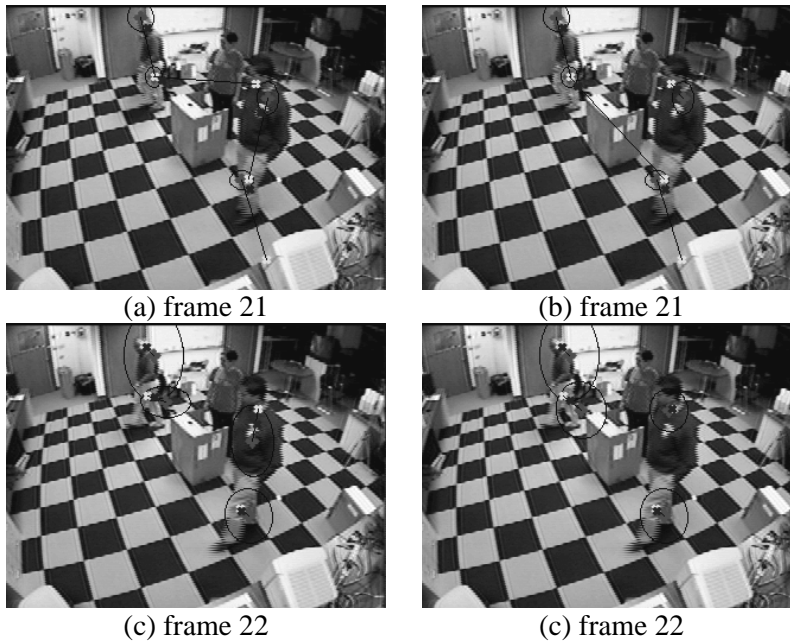


Figure 5: Results obtained with data association using the efficient approximation of the MD (left column) and the original MD formulation (right column).

6. CONCLUSION

To track features in image sequences we used in this paper a Kalman filter, which performs the prediction and correction of the features movement in every image frame. With our approach the association in each frame between each prediction, obtained by the KF, and the correspondent acquired measurement is established by optimizing the global matching set based on the MD. In order to reduce the computational cost of associating data in the KF for tracking along image sequences we proposed in this paper an efficient approximation of the original MD formulation. We have exemplified that this approach is especially advantageous

when a large number of features are tracked, and have shown that the resulting correspondences are generally very similar to those obtained with the original MD formulation.

7. ACKNOWLEDGEMENTS

This work was partially done in the scope of project “*Segmentation, Tracking and Motion Analysis of Deformable (2D/3D) Objects using Physical Principles*”, with reference POSC/EEA-SRI/55386/2004, financially supported by FCT - *Fundação para a Ciência e a Tecnologia* from Portugal. The first author would like to thank the support of the PhD grant, with reference SFRH/BD/12834/2003, from FCT as well.

REFERENCES

- [1] Wren, C.; Azarbayejani, A.; Darrell, T.; Pentland, A. (1997). Pfinder: Real-Time Tracking of the Human Body, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, 780-785
- [2] Cucchiara, R.; Grana, C.; Piccardi, M.; Prati, A. (2000). Statistic and Knowledge-based Moving Object Detection in Traffic Scenes, *Proceedings of the 3rd IEEE Conference on Intelligent Transportation Systems (ITSC2000)*, Dearborn, Indiana, USA
- [3] Jorge, P.; Abrantes, A.; Marques, J. (2004). On-line Tracking Groups of Pedestrians with Bayesian Networks, *Workshop PETS*, Prague, Czech Republic
- [4] Lin, L.; Bar-Shalom, Y. (2004). New Assignment-Based Data Association for Tracking Move-Stop-Move Targets, *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 40, No. 2, 714-725
- [5] O'Malley, P.; Nechyba, M.; Arroyo, A. (2002). Human Activity Tracking for Wide-Area Surveillance, *2002 Florida Conference on Recent Advances in Robotics*, Miami, USA
- [6] Kurita, M.; Tsuruoka, S.; Yokoi, S.; Miyake, Y. (1983). *Handprinted "Kanji" and "Hiragana" character recognition using weighting direction index histograms and quasi-Mahalanobis distance*, IEICE Technical Report PRL82-79
- [7] Kato, N.; Abe, M.; Nemoto, Y. (1996). A Handwritten Character Recognition System Using Modified Mahalanobis Distance, *Transactions Institute of Electronics, Information and Communication Engineers (IEICE)*, Vol. J79-D, No. 1, 45-52
- [8] Kimura, F.; Shridhar, M. (1991). Handwritten numerical recognition based on multiple algorithms, *Pattern Recognition*, Vol. 24, No. 10, 969-983
- [9] Sun, F.; Omachi, S.; Kato, N.; Aso, H.; Kono, S.; Takagi, T. (2000). Two-Stage Computational Cost Reduction Algorithm Based on Mahalanobis Distance Approximations, *International Conference of Pattern Recognition*, Barcelona, Spain
- [10] Pinho, R.; Tavares, J.; Correia, M. (2005). Human Movement Tracking and Analysis with Kalman Filtering and Global Optimization Techniques, *II International Conference On Computational Bioengineering*, Lisbon, Portugal
- [11] Welch, G.; Bishop, G. (1995). *An Introduction to Kalman Filter*, University of North Carolina at Chapel Hill, Technical Report, TR 95-041
- [12] Pinho, R.; Tavares, J.; Correia, M. (2005). A Movement Tracking Management Model with Kalman Filtering, Global Optimization Techniques and Mahalanobis Distance, *ICCMSE 2005 - International Conf. on Computational Methods in Science and Engineering*, Loutraki, Greece
- [13] Pinho, R.; Tavares, J.; Correia, M. (2007). An Improved Management Model for Tracking Missing Features in Computer Vision Long Image Sequences, *WSEAS Transactions on Information Science and Applications*, Vol. 4, No. 1, 196-203
- [14] Brown, L.; Senior, A. W.; Tian, Y.; Connell, J.; Hampapur, A.; Shu, C.; Merkl, H.; Lu, M. (2005). Performance Evaluation of Surveillance Systems Under Varying Conditions, *IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance*, Colorado, USA