

Tracking Features in Image Sequences with Kalman Filtering, Global Optimization, Mahalanobis Distance and a Management Model

Raquel R. Pinho¹ and João Manuel R. S. Tavares¹

Abstract: This work addresses the problem of tracking feature points along image sequences. In order to analyze the undergoing movement, an approach based on the Kalman filtering technique has been used, which basically carries out the estimation and correction of the features' movement in every image frame. So as to integrate the measurements obtained from each image into the Kalman filter, a data optimization process has been adopted to achieve the best global correspondence set. The proposed criterion minimizes the cost of global matching, which is based on the Mahalanobis distance. A management model is employed to manage the features being tracked. This model adequately deals with problems related to the occlusion of the tracked features, the appearance of new features, as well as optimizing the computational resources used. Experimental results obtained through the use of the proposed tracking framework are presented.

Keywords: Stochastic Filter, Data Association, Motion Correspondence, Optimization, Mahalanobis Distance, Image Analysis, Tracking

1 Introduction

Object tracking based on image processing and analysis techniques is a complex issue that has evolved considerably over the past decade. Movement analysis using video systems for motion acquisition and interactive modeling can assist one in terms of the analysis, diagnosis and assessment of movements through the use of tools that are exceptionally useful in a number of application areas such as in [Han, Feng and Owen (2007)] a study dealing with the transportation of irregular particles in turbulent flows which can be accomplished by tracking the particles along the pipeline structure; additionally, it could be applied to a study of the interspinous

¹ Instituto de Engenharia Mecânica e Gestão Industrial (INEGI), Lab. de Óptica e Mecânica Experimental (LOME) / Faculdade de Engenharia da Universidade do Porto (FEUP), Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal, Email: {rpinho, tavares}@fe.up.pt

process spacer device placed at L2-L3 if a tracking framework with orientation is employed [Zhang, Cheng, Oh, Spehar and Burgess (2008)]. There are numerous examples of movement tracking applications, such as: surveillance, deformation analysis, gait analysis, traffic control or even medical diagnosis, [Azarbayejani, Wren and Pentland (1996); Chen, Huang and Arrott (1998); Cucchiara, Grana, Piccardi and Prati (2000); Feldman and Balch (2003); Fish and Nielsen (1993); Zhou and Hu (2008)]. For instance, the analysis of human movement can be employed in medical diagnosis procedures, physical therapy or sports, to improve the study of gait disorders related to knee or hip pain, or even to help the control of motion cycles in rehabilitation or training processes, [Aggarwal and Cai (1999); Deutscher and Reid (2005); Veeraraghavan, Roy-Chowdhury and Chellappa (2005); Vieth (2007); Wang and Singh (2003); Zhou and Hu (2008)].

Many tracking applications require the simultaneous tracking of several objects, thus implying problems related to their appearance and disappearance in/from the scene, which can be analyzed over extended periods of time. The complexity of the tracked features and all the variables involved has led to the development of new technologies, such as high-speed cameras, and innovative computational approaches, that have been increasingly integrated in laboratories progressively dedicated to movement analysis, which has allowed for new insight into the tracking of features throughout image sequences. In fact, automated movement visual analysis computational systems can provide a number of significant advantages as is the case of more reliable events assessment, seeing that the computational algorithms always use and apply the same criteria, in addition to the fact that the systems do not suffer issues of fatigue or drifts, thereby permitting the processes to operate almost indefinitely and continually. However, the computational tracking of features in images is not a self-contained problem, as it involves several complex issues, such as image segmentation, an issue which is not addressed by this paper, but can be further developed in [Gonçalves, Tavares and Natal (2008); Raut, Raghuvanshi, Dharaskar and Raut (2009); Zhang, Fritts and Goldman (2008)].

1.1 Related Work

Many strategies have been proposed so as to address the difficulties associated with the visual tracking of features. In the following section, some of the key contributions produced over the last number of years, which are closely related to the topic of this paper are pointed out.

From single object to multi-target methods, first tracking approaches could not overcome problems related to the occlusion [Sethi and Jain (1987)], entrance, or disappearance of features [Rangarajan and Shah (1991)]. In the meantime, new solutions were presented; namely, [Salari and Sethi (1990)] the management of these

problems by first establishing the correspondence between the points detected, and subsequently, by extending the tracking process of the missing features by adding a number of hypothetical points [Rangarajan and Shah (1991)]. The problem of missed features is addressed by predicting their position based on an assumption of constant velocity. In turn, [Intille, Davis and Bobick (1997)] contend with the change in the number of features by examining specific regions in the image to detect appearances/disappearances before computing the correspondence results. [Rosales and Sclaroff (1998)] propose an occlusion detection routine which deals with occlusion problems by predicting the future locations of features, based on current 3D velocity and position estimates, and assumptions relating to the characteristics of the shape of the objects, and the manner in which they evolve over time.

[Veenman, Reinders and Backer (2001)] contribute to previous studies by introducing a common motion constraint for correspondence. The adopted constraint provides a severe restriction for the coherent tracking of points that lie on the same object. However, it is unsuitable for points lying on isolated objects moving in different directions. The algorithm in question assumes that the number of objects remains invariable along the image sequences.

[Arnaud, Memin and Cernuschi-Frias (2005)] use two points trackers: the first is a linear tracker well-suited for image sequences exhibiting global dominant motion; the latter is a nonlinear tracker, implemented by a conditional particle filter, which permits the tracking of points whose motion may be described locally. Hence, the proposed methodology deals with noisy sequences, abrupt changes of trajectories, occlusion cases and cluttered background.

In [Tissainayagam and Suter (2005)] objects are tracked through the image sequence by using a multiple hypothesis tracking algorithm coupled with a multiple model Kalman filter from the first frame onwards. The tracking process is carried out by predicting the position of each object's centroid in the next frame, and then by analyzing a region of interest surrounding the centroid so as to identify key points. This process is continued by matching the key points with the associated object's contour within the region under analysis. The correspondence between the extracted measurements and the predictions is established on a Mahalanobis distance basis.

Another key characteristic of a successful tracking system is its ability to effectively search for the pursued features in each frame of the image sequence under analysis and to establish the correct correspondences with the features being tracked.

A common approach for the detection of objects is to use the information from a single frame. However, some of the methods used to detect objects in image se-

quences make use of the temporal information computed from the sequence under analysis in order to reduce the number of false detections. The tasks of detecting the objects and establishing the correspondence between their instances across image frames can either be performed separately or jointly. In the former case, possible objects' regions in every image frame are obtained by using an object detection algorithm. Subsequently, the tracker matches the objects across the image frames. For instance, [Shafique and Shah (2003)] propose a multiframe approach to preserve the temporal coherency of velocity and position. The authors consider the correspondence issue as being a graph theoretic problem. Multiple image frame correspondence is concerned with identifying the best unique path for each point. In the cases of misdetection or occlusion, the path will consist of the missing positions in the corresponding image frames. This approach uses a window of image frames in the establishment of the correspondences in order to successfully handle occlusion cases whose durations are shorter than the defined temporal period. In the later case, the objects' regions and their correspondences are jointly estimated by iteratively updating the information of the previous frames in terms of the locations and regions of the objects being tracked, [Arnaud, Memin and Cernuschi-Frias (2005); Raut, Raghuvanshi, Dharaskar and Raut (2009); Rosales and Sclaroff (1998)].

Another possible approach is based on a Track Before Detect (TBD) setup, which deals with the tracking problem by assuming unthresholded measurements. The TBD is especially suitable for tracking weak objects, i.e. objects which in a classical setting will not often lead to a successful detection, [Boers and Driessen (2004); Salmond and Birch (2001)].

In order for a tracking system to present an appropriate performance, the most probable potential features' locations obtained should be used to update the features' state estimator. This is usually a data association problem. The probability of a given measurement being correct can be established by a distance function between the predicted state of the feature and the associated measured feature. The fact that the features' state may also consist of several characteristics such as color, size or shape, or even the composition of signals from heterogeneous sensors, such as in [Zhou, Taj and Cavallaro (2008)] should also be noted. This gains significant importance when distinguishing the features that are to be tracked as they may appear close to each other or even overlap one another.

The simplest correspondence implies the use of the nearest neighbor approach. However, in the case of the objects being close to each other, there is always a high probability that the correspondence obtained will be incorrect. An associated measurement which is incorrect can prevent the filter from successfully converging. Several statistical data association techniques exist to overcome this problem:

for example, Joint Probability Data Association Filtering (JPDAF) and Multiple Hypothesis Tracking (MHT) are two techniques which are widely used to solve problems related to data association. A detailed review and comparison of these techniques can be found, for example, in [Cox (1993)] and [Drummond (1995)].

Different approaches may be used to incorporate the updated measurements obtained by the tracking method. The Kalman filter is a widespread technique used in the tracking of objects throughout image sequences. However, it has recently been substituted by particle filters, [Arulampalam, Maskell, Gordon and Clapp (2002); Sitz, Schwarz and Kurths (2004)]. The Kalman filter is based on the assumption that disturbances and the initial features' state vector are distributed normally. It has been proven that the statistical mean obtained for the conditional distribution of a state is an optimal estimator in the sense that it minimizes the mean square error. However, if the assumption of normality is overlooked, there is no guarantee that the Kalman filter will provide the conditional mean of the state vector, [Maybeck (1979)].

Particle filters have been presented as representing a good alternative for the Kalman filter; mainly, because they represent a conditional distribution with several particles, which allows for multimodal state distributions, [Blake, Curwen and Zisserman (1993)]. However, particle filters have also revealed some serious problems, such as difficulties in tracking multiple objects as well as articulated objects. Additionally, if the modeled system has reduced noise or if the measured features have a very low variance, then the particle filter may not perform successfully or even collapse. To overcome these difficulties, several variations of the particle filter have been proposed, such as the Path Relinking Particle Filter, the Scatter Search Particle Filter, [Pantrigo, Sanchez, Gianikellis and Montemayor (2005)], the Kernel Particle Filter, [Chang and Ansari (2005)] and the Annealed Particle Filter, [Deutscher, Blake, North and Bascle (1999)]. Nevertheless, particle filters continue to represent an expensive computational solution, [Petrie (2004)].

1.2 Proposed Tracking Framework

In order to track the movement of feature points along image sequences, this work proposes the use of difference equations to model the features' trajectories, which are updated with the measurements obtained at discrete instances (in every image frame). This is achieved by using, a well-known statistical modeling approach: the Kalman filter, [Arulampalam, Maskell, Gordon and Clapp (2002); Welch and Bishop (1995)]. Thus, the tracking framework which has been developed benefits from the advantages of a statistical approach which has been properly formulated: it has the flexibility to adequately represent the undergoing movement in time series, in addition to allowing for the prediction of future observations. The Kalman

filter estimates a dynamical system by using a form of feedback control: the filter estimates the system's state at a particular point in time and then obtains feedback in the form of (noisy) measurements.

As has been previously indicated, the drawbacks of the Kalman filter are related to its relatively severe restrictive assumptions, [Arulampalam, Maskell, Gordon and Clapp (2002)]. To track the movement of features throughout image sequences, the tracking framework which has been developed combines the Kalman filter with optimization techniques for data association, in order to improve the filters' robustness whenever cases of the occlusion of features and non-linear movements are concerned. The correspondence between the predicted features and the measured features is based on the Mahalanobis distance minimization. The Mahalanobis distance ensures that the correspondence is performed according to the behavior of each tracked feature which has been previously identified. Its approximation to the χ -square distribution allows for the choice of a significance level, which represents the minimum value for a possible match. Therefore, even in the case of the Kalman filter restrictions not being satisfied, a frequent reality in many tracking applications, the results obtained by the proposed tracking framework may be corrected by the adopted matching solution.

In the proposed tracking framework, a management model has also been employed, as proposed in [Tavares and Padilha (1995)], which assures a successful resolution of problems related to the appearance, disappearance and occlusion of features, which has proven to be especially useful when the tracking is performed throughout image sequences of a considerable length. The model is capable of making the decision to continue tracking each tracked feature by taking into account its previous behavior. Features which continue to successfully appear in the image scene will obviously continue to be tracked; On the other hand in the case of no measured feature being associated with a feature in an image frame, then its tracking may cease, depending on the feature's previous behavior and on the number of existing image frames in which the measured feature has not been successfully associated with that feature.

Therefore, the approach which has been adopted by us and which is explained in detail in the subsequent sections of this paper, represents a novel and unified framework to track feature points along image sequences, which has the advantage of being extremely robust and computationally efficient, as can be verified by the experimental results which have been included and analyzed in detail throughout this paper.

1.3 Paper Overview

This paper is organized as follows: In the following section, an introduction is provided on the Kalman filter. Section 3 presents the solution which has been adopted for overcoming the correspondence problem, which is based on the optimization and Mahalanobis distance. Next, in section 4, an explanation as to the manner in which the integrated management model deals with the features being tracked, as well as their appearance and disappearance along image sequences is provided. Subsequently, some experimental tracking results obtained by using the proposed tracking framework on synthetic and real image sequences are presented and discussed. Finally, in the last section, some main conclusions are presented and future work is suggested.

2 The Kalman Filter

The Kalman filter is an optimal recursive Bayesian stochastic method. It provides *optimal* estimates that minimize the mean of the squared error of the modeled system. From a Bayesian stochastic viewpoint, the filter propagates the conditional probability density of the system's state conditioned by the knowledge of the actual data on the tracked features acquired by the measuring devices.

The equations of the Kalman filter fall into two main processes: time update (or prediction) and measurement update (or correction). Time update equations are responsible for projecting forward (in time) the current system's state and error covariance estimates so as to obtain the *a priori* system's estimates for the subsequent time step. In turn, measurement update equations deal with the system's feedback; that is, new measurements are incorporated into the *a priori* system's estimates to obtain improved *a posteriori* system's values, [Welch and Bishop (1995)].

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

$$X_t^- = \Phi X_{t-1}^+,$$

where Φ represents the system's state X_{t-1}^+ at the previous time step $t - 1$ to the system's state X_t^- in the current step t . The superscripts $+$ and $-$ indicate if the measurement data has or not been respectively incorporated. The related uncertainty is obtained by:

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

where P is the covariance matrix and Q models the system's noise.

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

$$K_t = P_t^- H^T [HP_t^- H^T + R_t]^{-1},$$

$$X_t^+ = X_t^- + K_t [U_t - HX_t^-],$$

$$P_t^+ = [I - K_t H] P_t^-,$$

where K is chosen to represent the filter's gain, that minimizes the a posteriori error covariance equation, H processes the transformation of coordinates between the predicted and the measurement spaces, R_t is the measurement noise, and I is the identity matrix, [Arulampalam, Maskell, Gordon and Clapp (2002); Welch and Bishop (1995)].

In the developed tracking framework, each features' state, X_t , is composed by its position $[x_t \ y_t]^T$ in the image frame, as well as its velocity $[v_{x_t} \ v_{y_t}]^T$ and acceleration $[a_{x_t} \ a_{y_t}]^T$. Although the measurements update, U_t , only comprehends the features' positions, the associated velocity and acceleration can be derived:

$$X_t = [x_t \ y_t \ v_{x_t} \ v_{y_t} \ a_{x_t} \ a_{y_t}]^T,$$

where:

$$v_{x_t} = x_t - x_{t-1}, \quad v_{y_t} = y_t - y_{t-1}, \quad a_{x_t} = v_{x_t} - v_{x_{t-1}}, \quad a_{y_t} = v_{y_t} - v_{y_{t-1}}.$$

In the developed tracking framework, if a tracked feature finds no correspondent in the set of new measurements, then its prediction would function as its measurement but with greater uncertainty (in the experimental examples presented in this paper, the uncertainty of a missed feature is doubled). Thus, using a binary variable $z_t^{(i)}$ which returns 1 (one) if a new measured feature has been successfully matched with feature i :

$$U_t = \begin{cases} [x_t \ y_t]^T & \text{if } z_t^{(i)} = 1 \\ [x_t^- \ y_t^-]^T & \text{otherwise} \end{cases},$$

with:

$$P_t^- = \begin{cases} P_t^- & \text{if } z_t^{(i)} = 1 \\ 2P_{t-1}^- & \text{otherwise} \end{cases}.$$

Thereafter, the tracking of missed features continues as usual, unless the features are discarded by the management model delineated further on. Throughout the image sequence, whenever a new feature is detected, its tracking is begun by initializing a new Kalman filter for it.

2.1 Kalman Filter Initialization

In the proposed tracking framework, when the tracking process is initialize for a feature, Q , P_0^+ , R_1 are defined as the identity matrix and Φ is associated with a constant acceleration model:

$$\Phi = \begin{bmatrix} 1 & 0 & \Delta t & 0 & \frac{\Delta t^2}{2} & 0 \\ 0 & 1 & 0 & \Delta t & 0 & \frac{\Delta t^2}{2} \\ 0 & 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 1 & 0 & \Delta t \\ 0 & 0 & 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha \end{bmatrix}.$$

In the experimental examples presented in this paper, Δt has been made equal to 1(one) and α to 0.1. Additionally, X_0^+ and U_1 are defined by the measurements obtained from the first image frame and $H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$.

3 Correspondence with Optimization techniques and Mahalanobis Distance

During the updating step of the Kalman filter, in order to associate the new measurement data which has been acquired (measure features) with the previously tracked features (predicted features), a criteria of correspondence (matching) must be adopted. That is, during this step, it is assumed that for each feature being tracked, one measurement should be considered for the correction of its predicted state.

According to the usual Kalman approach, the search area for each feature's position in the image plane under analysis is provided by an ellipse centered on its predicted position, whose axes are determined by the eigenvectors of the associated covariance reduced matrix, and its rays are derived by the related eigenvalues, [Correia, Campilho and Padilha (1995); Tavares and Padilha (1995); Tissainayagam and Suter (2005)]. As the filter is converging, it provides more accurate estimates and consequently, the size of the search areas successively decreases, and the involved computational cost is thereby reduced, [Correia, Campilho and Padilha (1995); Tavares and Padilha (1995)]. However, this usual approach may raise a number of inconveniences: there may not be any measured feature in the searching area; alternatively, there might be several measured features in the same area; and even if there is only one measured feature for each estimated feature in the associated search area, there is no guarantee that the best set of correspondences has been achieved. Further on in this paper, an approach which is capable of surpassing such ambiguities is outlined: an optimization technique is used to obtain the most

favorable association between the predictions and the actual set of measurements; and the cost of each correspondence is provided by the normalized squared Mahalanobis distance. Thus, this matching approach permits one to obtain the best global set of correspondences between the estimates and the measurements in a Mahalanobis distance sense.

3.1 Optimization in the Measurement Update

To optimize the set of correspondences found between the filters' predictions and the acquired measurements, various optimization algorithms may be used. In the proposed tracking framework, the Simplex algorithm has been applied. This algorithm is a widespread iterative algebraic procedure used to determine at least one optimal solution for each problem, [Bastos and Tavares (2006); Hillier and Lieberman (2001); Press, Teukolsky, Flannery and Vetterling (2002)].

As a linear optimization method, the Simplex algorithm optimizes a function, which is subject to some restrictions. In the case of tracking the movement of points along image sequences, the main goal adopted has been to minimize the global cost of the association between the set of acquired measurements and the estimates provided by the Kalman filter. According to the approach in question, it has been assumed that for each estimate, there will be one measurement at most, and thus each new measurement will thereby correspond to a feature's position. In the proposed tracking framework, this has been accomplished through the use of the assignment formulation of the Simplex algorithm, [Hillier and Lieberman (2001)].

3.2 Mahalanobis Distance

To optimize the correspondences between the set of measured features and the set of estimated features, a cost has been associated to each correspondence which is provided by the squared Mahalanobis distance.

The Mahalanobis distance between two features is scaled by the statistical variation in each component of the entity. Therefore, if X_E is an estimated feature's position and X_M represents the measured feature's position, then their squared Mahalanobis distance is obtained by:

$$d_{\chi^2} = \frac{(X_M - X_E)^T (V_M + V_E)^{-1} (X_M - X_E)}{2},$$

where V_M is the covariance matrix of the measurements, and V_E is the covariance matrix associated with the feature's prediction, [Tavares and Padilha (1995)]. The Mahalanobis distance has been chosen, instead of the Euclidean distance, essentially because it takes into account the correlations of the data set in addition to the fact that it is scale-invariant.

In the tracking framework which has been proposed, the covariance of each estimate is considered to be the reduced prediction error covariance matrix provided by the Kalman filter, and the measurement covariance matrix is calculated by considering all the measurements acquired from the associated image frame.

According to the above expression, not only does the Mahalanobis distance depend on the actual measurement and estimate of that tracked feature, but also on its previous behavior as its covariances are also contemplated. Therefore, the Mahalanobis distance values will be inversely proportional to the quality of the prediction/measurement association; consequently, to optimize the correspondences, the related cost function should be minimized.

The squared Mahalanobis distance can be approximated, in this case, by a χ -square distribution with 2 degrees of freedom. Thus, if a significance level of 90% is chosen, then all correspondences should be inferior to the threshold value of 4.6052. If a feature does not satisfy this condition, then it will have no correspondence, [Tavares and Padilha (1995)].

3.3 Problems of Object Occlusion, Appearance and Disappearance

One of the restrictions of the assignment formulation of the Simplex algorithm is the “one to one” correspondence between each measured and estimated feature by the Kalman filter. However, when features are occluded or disappear off the scene definitively, or alternatively, new features appear on the scene, this restriction does not stand, due to the fact that the numbers of estimated and measured features are not equal. To overcome this difficulty, a standard procedure has been applied: fictitious features are added in order for the number of estimated features to be equal to the number of the measured features. The cost of each correspondence that has been made with a fictitious variable is considered to be nil. Subsequently, each estimated feature that has been matched with a fictitious feature is considered to be unmatched, [Bastos and Tavares (2006); Hillier and Lieberman (2001); Press, Teukolsky, Flannery and Vetterling (2002)]. Additionally, as has been described in the previous section, estimated features are also considered to be unmatched if the associated minimal squared Mahalanobis distance is greater than a given threshold value.

Thus, if a previously tracked feature does not find any correspondence among the set measured features, then its tracking may be stopped (depending on the adopted management model that the following section will describe), or may continue with greater uncertainty, as previously explained in section 2. If a measured feature does not find any correspondence among the set of estimated features, then it will be considered to be a new feature and its tracking will be initialized, Figure 1.

4 Tracked Features Management Model

As has been stressed, new features can arise in any image frame of an image sequence, but they may also disappear temporarily or even definitively. Thus, in the very lengthy image sequences obtained, for example, from surveillance systems, one has to decide if a missed feature should be kept in the tracking process due to it having been temporally occluded, or alternatively, if its tracking process should be stopped, and the associated computational resources freed, as it could have definitively disappeared from the image scene. This decision is even of greater importance if many features are being tracked and the available computational resources are reduced.

In the proposed tracking framework, a management model which associates a confidence value, $\lambda_t^{(i)}$, to each tracked feature has been used: while a feature is being tracked, in each image frame t , if it has successfully been matched with a measure feature, then its confidence value will be increased if this is lower than an upper threshold value, λ_{\max} ; on the other hand, if it is not matched with any measure feature, its confidence value will decrease, and if it is inferior to a lower threshold confidence value, λ_{\min} , then the feature will be considered to have definitively disappeared from the scene in which case its tracking should be stopped and its computational resources freed:

$$\lambda_t^{(i)} = \begin{cases} \lambda_{t-1}^{(i)} - 1 & \text{if } \lambda_{\min} < \lambda_{t-1}^{(i)} \wedge z_t^{(i)} \neq 1 \\ \lambda_{t-1}^{(i)} + 1 & \text{if } \lambda_{t-1}^{(i)} < \lambda_{\max} \wedge z_t^{(i)} = 1 \\ \lambda_{\max} & \text{if } \lambda_{t-1}^{(i)} = \lambda_{\max} \wedge z_t^{(i)} = 1 \\ \text{tracking of } i \text{ is stopped} & \text{if } \lambda_{t-1}^{(i)} = \lambda_{\min} \wedge z_t^{(i)} \neq 1 \end{cases},$$

where $z_t^{(i)}$ is a binary variable that returns 1 (one) if feature i has been successfully matched in frame t .

Therefore, if a feature disappears for a reduced number of consecutive image frames, its tracking process will be continued without losing any data. However, if the number of consecutive images in which the tracked feature has not been successfully matched with a measured feature is higher than a predefined value, its tracking will be stopped and the feature discarded by the management model and consequently, its computational resources will be freed. If a discarded feature reappears later, it will be considered to be a new feature and its tracking initialized.

With the followed management strategy, the proposed tracking framework can continually track lengthy image sequences containing several features and maintain the computational resources used as reduced as possible, which can be essential in applications with severe restrictions in terms of the computational resources available.

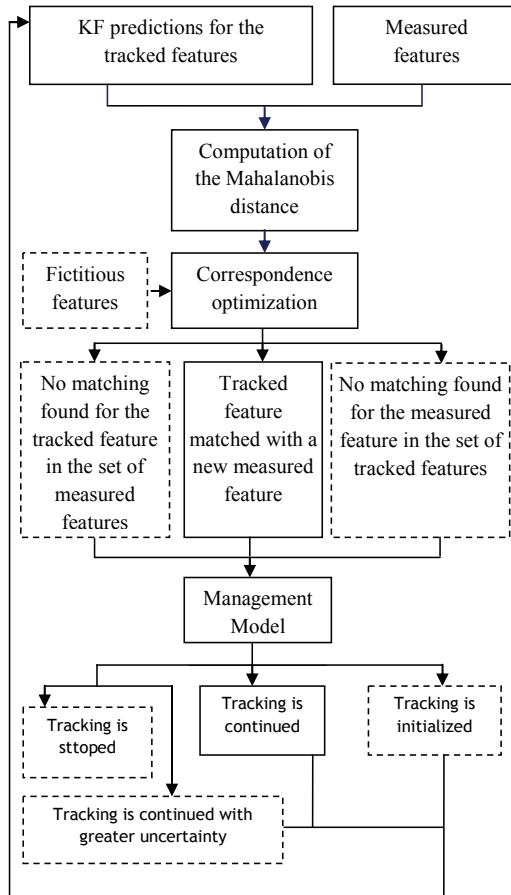


Figure 1: Integrated solution in the proposed tracking framework for the matching and management of the tracked features.

The results presented in this paper were obtained using integer confidence values between 0 (zero) and 5, and all features have been initialized with a confidence value of 3, [Tavares and Padilha (1995)].

5 Experimental Results

In this section, the use of the proposed tracking framework is exemplified and discussed with the aid of three experimental image sequences.

In each frame of the examples presented, the predicted position is indicated with a red +, the uncertainty area is circumscribed in black, each measurement is the center of the detected green contour, and the corrected position is represented by a blue x. The association between each prediction/measurement is depicted by a black line segment.

For the first example, Figure 2, a synthetic sequence of 15 image frames has been considered. At the beginning of the sequence, the two blobs are visible. Then, the circular blob disappears definitively, yet the management model continues its tracking during the subsequent image frames, gradually increasing its uncertainty, until it stops its tracking (it should be noted that in image frame (e), the uncertainty region surpasses the image border). In the 2nd image frame, a triangular blob appears, and in the 3rd image frame the square blob instantly disappears. In the 4th image frame, the acquired blobs overlap each other, and with the image processing techniques that have been used, only one measured feature is obtained and matched with one tracked blob. However, both blobs continue to be subsequently correctly tracked. From the 7th image frame onwards 25 blobs are successfully tracked. In the 10th image frame, the 23rd rectangular blob disappears, and the management model proceeds as previously described in the case of the circular blob. Their track is discontinued after the 14th image frame thereby freeing the associated computational resources which, in turn, favorably implies the computational performance achieved by the proposed tracking framework, Figure 3.

The confidence values associated with the tracked features using the tracking management model presented are indicated in Table 1.

As previously mentioned, in order to associate the measured features and the tracked features, a global optimization criterion is used which is based on the Mahalanobis distance. In this first example, all the distances are lower than 1.5, Figure 4, and consequently, the threshold provided by χ -square never has to be used. The low Mahalanobis distances are a guarantee of a high matching confidence.

In this first experimental example, one can notice that if a feature disappears, the related uncertainty value increases, yet the approach adopted will keep on trying to track it for several image frames, at which time it will be definitively discarded. As has already been referred to, this may be helpful in application cases in which some features can be occluded or incapable of being detectable for short periods of time.

In the second experimental example, the tracking of people in images from a surveillance system in a shopping centre has been analyzed, Figure 5 (images from [EC-Funded-CAVIAR-project and 2001-37540 (2004)]). The features that were to be tracked were obtained by manual segmentation; however, their detection could have

been done automatically by using suitable image processing techniques.

In the first 8 image frames (Figure 5(a-h)) 3 persons are successfully tracked. In the 10th image frame (Figure 5(j)) another person begins to be tracked. In the 12th frame (Figure 5(l)) one of the previously tracked persons starts to enter a store and thus the management model stops his tracking in the 16th image frame (Figure 5(p)). However, if he had left the store earlier, the management model could have maintained its previous tracking data, in which case, when he left the store the proposed tracking framework would initialize his tracking as a new feature. Once again, it should be noted that in the proposed tracking framework, the maximum number of image frames during which a feature continues to be tracked without any acquired measurement being associated is user-determined (in this case, 5 image frames have been stipulated).

Table 1: Tracking the blobs in the first 9 image frames of Figure 1: Features' confidence values ([0, 5]).

Features/Frames	1	2	3	4	5	6	7	8	9
Circular	3	2	1	0	-	-	-	-	-
Square	3	4	3	4	5	5	5	5	5
Triangular	-	3	4	5	4	5	5	5	5
23x Rectangular	-	-	-	-	-	-	3	4	5

In the third, and final, experimental example, a sequence of 414 image frames with 3 mice in a lab environment, have been considered, Figures 6 and 7. Several difficulties are associated with the tracking of the center points of the mice's silhouettes in the image sequence in question. One of which entails the rapid movement of the mice, as they may move back and forth, drastically changing direction at any time, Figure 6, or alternatively, may move very quickly in an invariable direction, Figure 7.

The non-linear behavior verified in this third example, which is not successfully dealt with by the usual Kalman filter approach, can give rise to variations of up to 45 pixels between the estimated positions and the associated measurement (in 320x240 images), both along the xx and yy axis, Figure 8¹. In this figure, the error represented is due to the difference between the positions of the features predicted by the Kalman filter and the matched measured features. Despite the discrepancies of the mice's movements, the proposed tracking framework always recovers well,

¹ The noise data considered by the graph of this figure is due to a noisy measurement which was instantly acquired in image frame 293, but it was not successfully matched with a new measurement, and consequently, it was discarded by the management model in use.

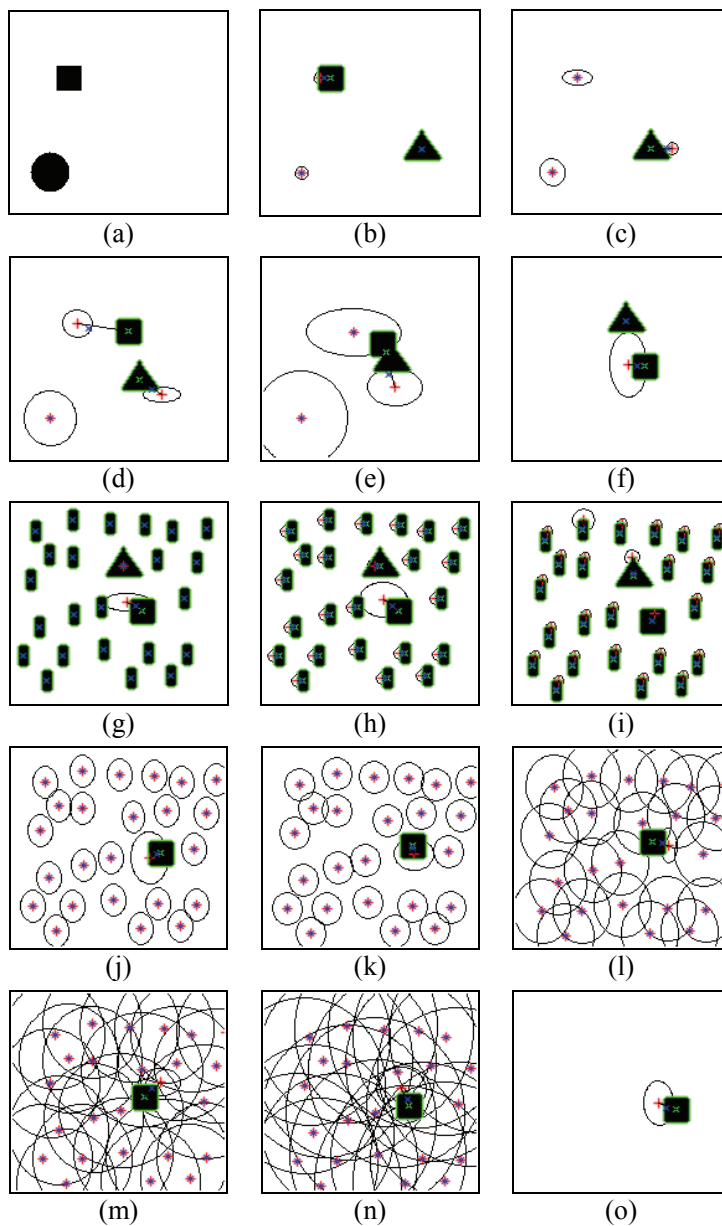


Figure 2: Tracking blobs along a 15 image frame sequence: (a) - original 1st frame; (b)-(o) - Kalman Filtering results: search area defined by solid ellipses, the predicted position for each blob is indicated by +, and the corrected positions indicated by x.

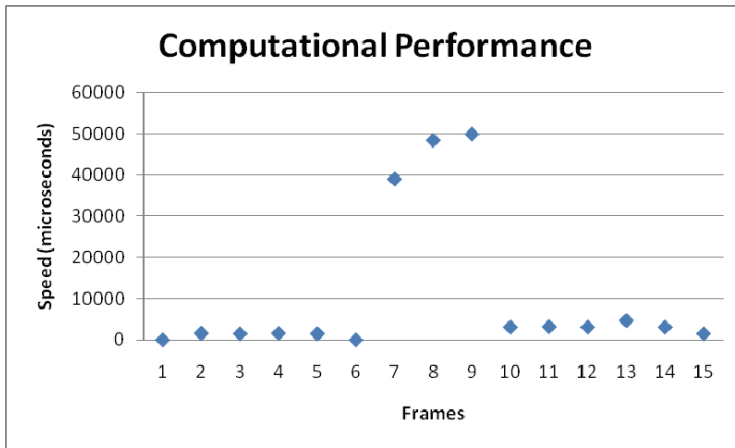


Figure 3: Tracking the blobs in the image sequence of Figure 1: Processing time in an Intel(R) Core(TM)2 Duo CPU at 2.00 GHz and 2038 MB RAM running Microsoft Windows Vista.

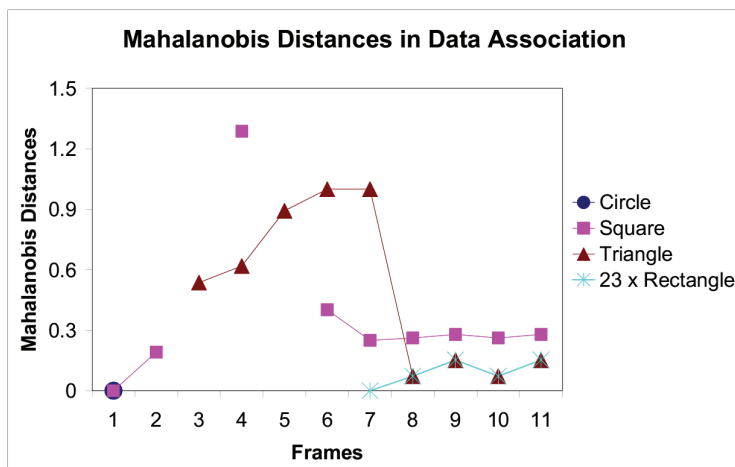


Figure 4: Values of the Mahalanobis distances used in the matching of the blobs tracked during the first 9 image.

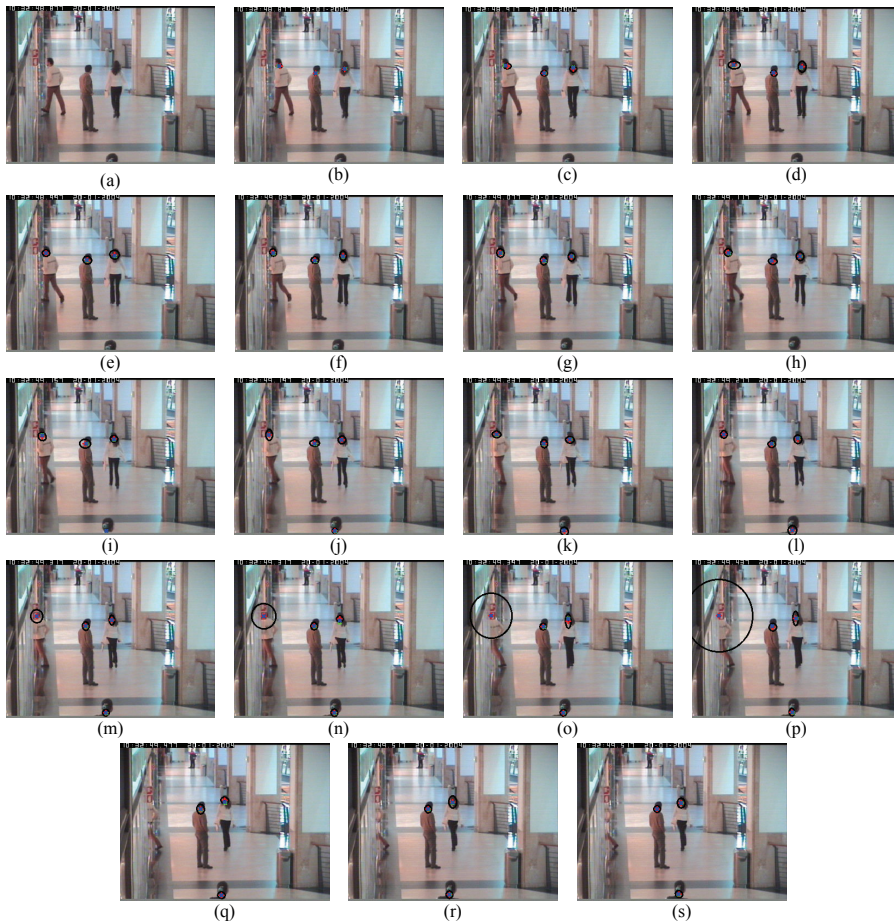


Figure 5: Tracking persons in a shopping centre: (a) - original 1st image frame; (b) to (s) - the management model in use allows for the successful tracking of features during extensive image sequences.

as can be verified in Figure 8 where the relative maximum errors are quite often followed by relatively low errors (below 10 pixels).

By calculating the Root of the Mean Square Errors (RMSE) for each of the mice tracked, the results indicated in Table 2 are obtained, which are quite low bearing in mind the complexity of the movement involved.

Another difficulty related to the third experimental example is related to the mice's segmentation: in some image frames the mice are so close to each other that there are only two segmented tracked features, or even only one, as can be verified in

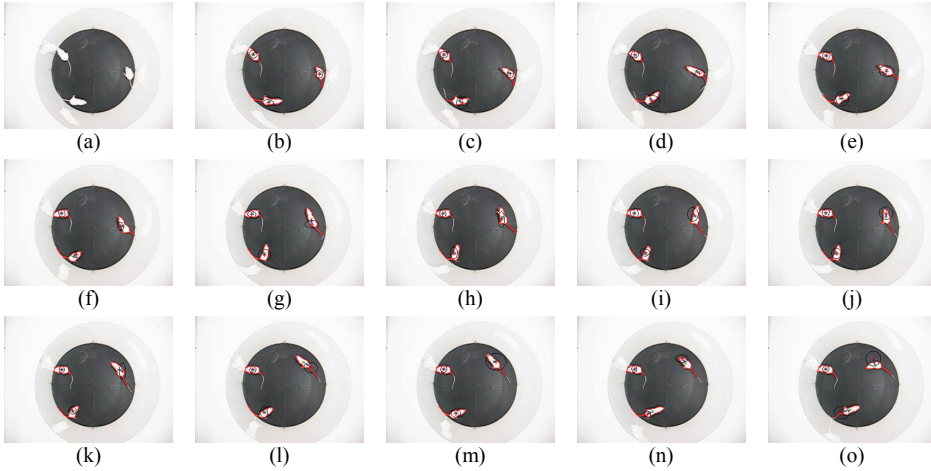


Figure 6: Tracking mice in a lab environment over 414 image frames: (a) - original 205th image frame; (b) - (o) - even with significant changes in the tracked movement, the proposed tracking framework can perform the tracking correctly.



Figure 7: Tracking mice in a lab environment over 414 image frames: quick movement with severe direction changes can be correctly tracked by the proposed tracking framework.

Figure 9. In this third case, the segmentation task has been achieved by background subtraction and then the center of the mass of each blob detected has been found.

This experimental example also proves that the proposed tracking framework can recover the tracking of features which are not visible or non-detected during some image frames well, due to the features' management policy defined in the integrated features' management model.

6 Conclusions and Future Work

This paper has aimed to present a novel and integrated computational tracking framework capable of performing the tracking of feature points throughout image sequences in a robust and efficient manner. In such a framework, the Kalman filter

has been used to predict and correct the position of the tracked features' position, as well as their velocity and acceleration, throughout the image sequences. To accomplish the most efficient matching in each new image frame between the predicted features and the measured features, optimization techniques and Mahalanobis distance have been employed. This matching approach allows one to overcome the cases in which the measured features lie beyond the searching areas considered by the default Kalman approaches, as well as when the movement in question presents high non-linearity, enhancing the robustness and flexibility of the proposed tracking framework.

In the proposed tracking framework, a tracked features' management model has been integrated. The model used associates each tracked feature to a confidence value that is used to distinguish cases of a feature's temporal occlusion from its definitive disappearance. When a tracked feature is merely temporally occluded, its tracking is maintained. However, when it has disappeared definitively, its tracking is ceased and the associated computational resources are freed, thereby enhancing the computational efficiency of the proposed tracking framework, which can be extremely attractive in applications of low computational resources.

Table 2: RMSE in xx and yy axis associated with the tracking results of the mice case.

Mouse	N ^o of image frames	RMSE in xx	RMSE in yy
1	395	7.42	7.51
2	394	6.59	7.05
3	366	8.36	8.38

This work can be continued by carrying out a comparison between the tracking results obtained by the Kalman filter and those accomplished by other stochastic filters, such as the Unscented Kalman filter and Particle filters. Additionally, the comparison of the proposed tracking framework with other tracking systems applied in the tracking of features throughout complex image sequences, involving the appearance and disappearance of the tracked features as well as interactions between them, would be of great importance and interest.

Acknowledgement: The authors would like to express their deep gratitude to Dr. Vanessa Esteves from the University of Porto – Portugal for her valuable help with the English version of this paper.

Additionally, the authors would like to thank Prof. Hemerson Pistori and Prof. João Bosco Monteiro (at Dom Bosco Catholic University - Brazil) for kindly having shared the mice images used in the final example presented in this paper.

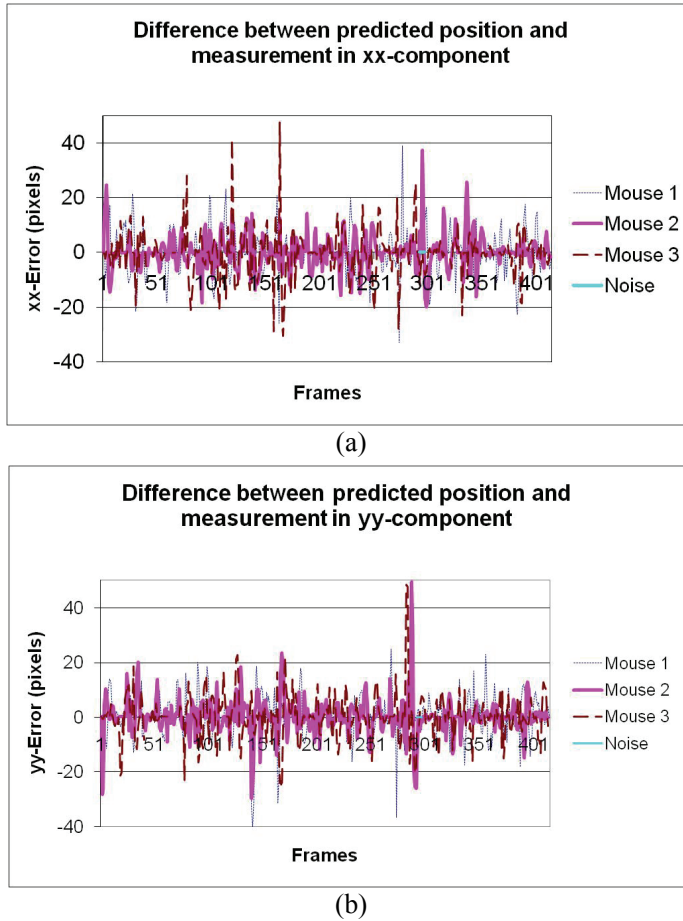


Figure 8: Tracking mice in a lab environment throughout 414 image frames: differences between predicted positions and associated measurements in xx and yy axis, (a) and (b) respectively.

This work was partially done within the scope of the following project: *Segmentation, Tracking and Motion Analysis of Deformable (2D/3D) Objects using Physical Principles*, reference POSC/EEA-SRI/55386/2004, financially supported by the Portuguese FCT - Fundação de Ciência e Tecnologia.

The first author would like to express their gratitude for and acknowledge the support of the PhD grant SFRH / BD / 12834 / 2003 of the FCT.

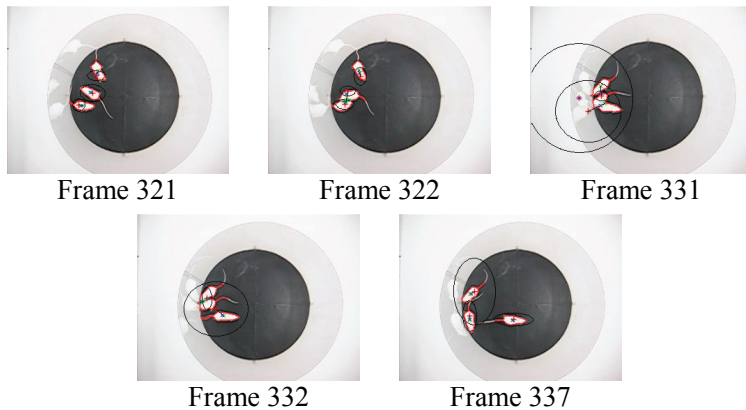


Figure 9: Tracking mice in a lab environment throughout 414 image frames: the tracked features may not be continually and successfully matched, but the proposed tracking framework always recovers their tracking adequately.

References

- Aggarwal, J.; Cai, Q.** (1999): Human Motion Analysis: A Review. *Computer Vision and Image Understanding* 73:428/440.
- Arnaud, E.; Memin, E.; Cernuschi-Frias, M.** (2005): Conditional filters for image sequence based tracking - Application to point tracking. *IEEE Transactions on Image Processing* 14:63/79.
- Arulampalam, M.; Maskell, S.; Gordon, N.; Clapp, T.** (2002): A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing* 50:174/188.
- Azarbayejani, A.; Wren, C.; Pentland, A.** (1996): Real-Time 3D Tracking of the Human Body. IMAGE'COM 96. Bordeaux, France.
- Bastos, L.; Tavares, J.** (2006): Matching of Objects Nodal Points Improvement using Optimization. *Inverse Problems in Science and Engineering* 14:529/541.
- Blake, A.; Curwen, R.; Zisserman, A.** (1993): A Framework for Spatio-temporal Control in the Tracking of Visual Contours. *International Journal of Computer Vision* 11:127/145.
- Boers, Y.; Driessen, J.** (2004): An Interacting Multiple Model Particle Filter. *IEE Proceedings Radar, Sonar and Navigation* 151:351/357.
- Chang, C.; Ansari, R.** (2005): Kernel Particle Filter for Visual Tracking. *IEEE Signal Processing Letters* 12:242/245.

- Chen, C. W.; Huang, T. S.; Arrott, M.** (1998): Modeling, Analysis, and Visualization of Left Ventricle Shape and Motion by Hierarchical Decomposition. *IEEE Deformable Models In Medical Image Analysis* 16:342/356.
- Correia, M.; Campilho, A.; Padilha, A.** (1995): Motion Estimation for Traffic Analysis. 7^o Congresso Português de Reconhecimento de Padrões. Aveiro, Portugal, p 831/838.
- Cox, I.** (1993): A review of statistical data association techniques for motion correspondence. *International Journal of Computer Vision* 10:53/66.
- 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, p 27/32.
- Deutscher, J.; Blake, A.; North, B.; Bascle, B.** (1999): Tracking through Singularities and Discontinuities by Random Sampling. 7th Conference on Computer Vision, vol 2, p 1144/1149.
- Deutscher, J.; Reid, I.** (2005): Articulated Body Motion Capture by Stochastic Search. *International Journal of Computer Vision* 61:185/205.
- Drummond** (1995): Multiple Sensor Tracking with Multiple Frame, Probabilistic Data Association. *Proceedings of the SPIE: Signal and Data Processing of Small Targets* 2561:322/336.
- EC-Funded-CAVIAR-project and 2001-37540, I.** (2004): <http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/>.
- Feldman, A.; Balch, T.** (2003): Automatic Identification of Bee Movement Using Human Trainable Models of Behaviour. Mathematics and Algorithms of Social Insects. Atlanta, USA.
- Fish, D.; Nielsen, J.** (1993): Clinical Assessment of Human Gait. *Prosthetics and Orthotics Journal* 5:39/48.
- Gonçalves, P.; Tavares, J.; Natal, R.** (2008): Segmentation and Simulation of Objects Represented in Images using Physical Principles *CMES: Computer Modeling in Engineering & Sciences* 32:45/55.
- Han, K.; Feng, Y.; Owen, D.** (2007): Numerical Simulations of Irregular Particle Transport in Turbulent Flows Using Coupled LBM-DEM. *CMES: Computer Modeling in Engineering & Sciences* 18:87/100.
- Hillier, F.; Lieberman, G.** (2001): *Introduction to Operations Research*. McGraw-Hill International Editions.
- Intille, S.; Davis, J.; Bobick, A.** (1997): Real-time closed-world tracking. IEEE Conference on Computer Vision and Pattern Recognition. San Juan, Puerto Rico,

p 697/703.

Maybeck, P. (1979): *Stochastic Models, Estimation, and Control*. Academic Press.

Pantrigo, J. J.; Sanchez, A.; Gianikellis, K.; Montemayor, A. (2005): Combining Particle Filter and Population-based Metaheuristics for Visual Articulated Motion Tracking. *Electronic Letters on Computer Vision and Image Analysis* 5:68/83.

Petrie, T. (2004): Tracking Bouncing Balls using Kalman Filters and Condensation. Technical Report. University of Colorado.

Press, W.; Teukolsky, S.; Flannery, B.; Vetterling, W. (2002): *Numerical Recipes in C++: The Art of Scientific Computing*. Cambridge University Press.

Rangarajan, K.; Shah, M. (1991): Establishing motion correspondence. Computer Vision and Pattern Recognition Proceedings, p 103/108.

Raut, S.; Raghuvanshi, M.; Dharaskar, R.; Raut, A. (2009): Image Segmentation - A State-Of-Art Survey for Prediction. International Conference on Advanced Computer Control Singapore, p 420/424.

Rosales, R.; Sclaroff, S. (1998): Improved Tracking of Multiple Humans with Trajectory Prediction and Occlusion Modelling. Proceedings IEEE Conference on Computer Vision and Pattern Recognition, Workshop on the Interpretation of Visual Motion. Santa Barbara, USA.

Salari, V.; Sethi, I. (1990): Feature point correspondence in the presence of occlusion, . *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12:87/91.

Salmond, D.; Birch, H. (2001): A particle filter for track-before-detect. American Control Conference. Arlington, USA, p 3755/3760.

Sethi, I.; Jain, R. (1987): Finding trajectories of feature points in a monocular image sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 9:56/73.

Shafique, K.; Shah, M. (2003): A non-iterative greedy algorithm for multi-frame point correspondence. IEEE International Conference on Computer Vision. Nice, France, p 110/115.

Sitz, A.; Schwarz, U.; Kurths, J. (2004): The Unscented Kalman Filter, a Powerful Tool for Data Analysis. *International Journal of Bifurcation and Chaos* 14:2093/2105.

Tavares, J.; Padilha, A. (1995): Matching Lines in Image Sequences with Geometric Constraints. 7º Congresso Português de Reconhecimento de Padrões. Aveiro, Portugal.

Tissainayagam, P.; Suter, D. (2005): Object tracking in image sequences using point features. *Pattern Recognition* 38:105/113.

Veenman, C.; Reinders, M.; Backer, E. (2001): Resolving motion correspondence for densely moving points. *Transactions on Pattern Analysis and Machine Intelligence* 23:54/72.

Veeraraghavan, A.; Roy-Chowdhury, A.; Chellappa, R. (2005): Matching Shape Sequences in Video with Applications in Human Movement Analysis. *IEEE Transactions on Pattern analysis and Machine Intelligence* 27:1896-1909.

Vieth, M. (2007): Reliable Recognition and Tracking of Multiple Persons in Work Safety Relevant Environments. MSc Thesis. Fachhochschule Bonn-Rhein-Sieg.

Wang, J.; Singh, S. (2003): Video Analysis of Human Dynamics - A Survey. *Real-time Imaging Journal* 9:320/345.

Welch, G.; Bishop, G. (1995): An Introduction to Kalman Filter. University of North Carolina at Chapel Hill.

Zhang, H.; Fritts, J.; Goldman, S. (2008): Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding* 110:260/280.

Zhang, Y.; Cheng, B.; Oh, C.; Spehar, J.; Burgess, J. (2008): Kinematic Analysis of Lumbar Spine Undergoing Extension and Dynamic Neural Foramina Cross Section Measurement. *CMES: Computer Modeling in Engineering & Sciences* 29:55/62.

Zhou, H.; Hu, H. (2008): Human motion tracking for rehabilitation - A survey. *Biomedical Signal Processing and Control* 3:1/18.

Zhou, H.; Taj, M.; Cavallaro, A. (2008): Target detection and tracking with heterogeneous sensors. *IEEE Journal of Selected Topics in Signal Processing* 2:503/513.

