

Dynamic tracking re-adjustment: a method for automatic tracking recovery in complex visual environments

Anastasios Doulamis

Published online: 29 September 2009
© Springer Science + Business Media, LLC 2009

Abstract Detection and analysis of events from video sequences is probably one of the most important research issues in computer vision and pattern analysis society. Before, however, applying methods and tools for analyzing actions, behavior or events, we need to implement robust and reliable tracking algorithms able to automatically monitor the movements of many objects in the scene regardless of the complexity of the background, existence of occlusions and illumination changes. Despite the recent research efforts in the field of object tracking, the main limitation of most of the existing algorithms is that they are not enriched with automatic recovery strategies able to re-initialize tracking whenever its performance severely deteriorates. This is addressed in this paper by proposing an *automatic tracking recovery* tool which improves the performance of any tracking algorithm whenever the results are not acceptable. For the recovery, non-linear object modeling tools are used which probabilistically label image regions to object classes. The models are also time varying. The first property is implemented in our case using concepts from functional analysis which allow parametrization of any arbitrary non-linear function (with some restrictions on its continuity) as a finite series of known functional components but of unknown coefficients. The second property is addressed by proposing an innovative algorithm that optimally estimates the non-linear model at an upcoming time instance based on the current non-linear models that have been already approximated. The architecture is enhanced by a decision mechanism which permits verification of the time instances in which tracking recovery should take place. Experimental results on a set of different video sequences that present complex visual phenomena (full and partial occlusions, illumination variations, complex background, etc) are depicted to demonstrate the efficiency of the proposed scheme in proving tracking in very difficult visual content conditions. Additionally, criteria are proposed to objectively evaluate the tracking performance and compare it with other strategies.

Keywords Tracking recovery · Object detection · Event analysis

A. Doulamis (✉)
Technical University of Crete, Polytechnioupolis Campus, Kounoupidiana, Chania, Crete, Greece
e-mail: adoulam@ergasya.tuc.gr

1 Introduction

Event analysis and area supervision from video sequences is a critical task for many applications dealing with service quality assurance (adherence to predefined procedures of a production), security/safety (prevention of actions that may lead to hazardous situations), crisis management in public service areas (e.g., train stations, airports), etc. However, the traditional approaches for event detection in videos assume well structured environments and they fail to operate in largely unsupervised way under adverse and uncertain conditions from those on which they have been trained. Another drawback of the current methods is the fact that they focus on narrow domains using specific concept detectors such as “human faces”, “cars”, “buildings” and so on. These limitations make the current surveillance systems inconvenient, since automatic video monitoring fails in case of environmental changes, such as occlusions, appearance/disappearance of objects, and inaccurate which leads to a human assistive supervision. However, manual video surveillance is an expensive and highly subjective solution. In particular, recent studies have proven that the attention of the operators of current surveillance systems is mainly attracted by the appearance of the monitored individuals and not by their behavior [27]. Thus, re-configurable and re-adjusting tools able to adapt their response to the new environmental conditions have attracted great research interest in computer vision society [9].

Detection and tracking of moving objects is one of the key components of an area surveillance architecture [14, 22]. Motion detection aims at segmenting foreground regions corresponding to moving objects from the background. Probably the most popular techniques for moving objects segmentation are the background subtraction and temporal differencing [15, 29, 30]. Background subtraction detects moving objects in an image by evaluating the difference of pixel features of the current scene image against the reference background image [15, 30]. On the other hand, temporal differencing calculates the difference of pixel features between consecutive scene frames in an image sequence [29]. These methods, however, present several limitations; they do not work in case that the background changes from time to time, they are very sensitive to noise and illumination variations, they fail in case of occlusions (either partial or full) and they usually recover only some parts of the moving objects.

However, in real-life scenarios there exists several complex visual phenomena which requires more intelligent detection/tracking algorithms [24]. Examples include the (i) agile motion which is a sustained object movement that exceeds a tracker’s dynamic prediction abilities, (ii) distraction which is a phenomenon that in a scene there is another object of similar appearance to the object being tracked and (iii) occlusion which is the situation when another object is interposed between the camera and the tracked object.

To address these difficulties complicated moving object detection/tracking algorithms have been proposed in the computer vision society, which can be discriminated into three main categories; *motion models, search methods and appearance-based techniques* [2]. In motion models, motion information is exploited to predict the new location of an object. The simplest approaches assume a linear relationship between the object movement and model describing through affine transformations [20]. Non-linear approaches have been also adopted but they are suitable only for specific environments and motion models [5]. The main drawback of these approaches, however, is that their accuracy is dropped in the existence of agile motion, distraction and occlusions. The search techniques exploit the assumption that the appearance of an object does change from time to time and thus it presents similar properties within adjacent frames of a video sequence. Approaches towards this direction are the methods of [26] and [3]. These techniques iteratively search for a

region in a video frame that maximizes the similarity between this frame and the target one. Such approaches, however, are sensitive to background distractors, clutter, and occlusions issues.

To overcome these problems, stochastic methods have been reported in the literature. One classical example is the Kalman filter which exploits the randomness generated by a linear dynamic operator perturbed by Gaussian noise [21]. Superior techniques to Kalman filters are the particle filters, which also support the assumptions of linear dynamic and Gaussian observations using, however, nonparametric density estimation and multiple hypotheses [13]. Particle filters' simplicity, robustness and effectiveness make them successful models in many challenging tasks. They are able to simultaneously track multiple hypotheses (objects) and they can recursively approximate the posterior probability density function (pdf) in the state space with a set of random samples (particles) [1, 23].

The performance of a particle filter algorithm actually depends on appearance models and the similarity measures used for object matching. The appearance of an object is in fact represented by using visual features. Many features have been proposed in the literature to characterize both rigid and non-rigid object appearances such as color histograms [12], contours [16, 35] and texture [25]. In real-life environments, however, the appearance models can change over time due to illumination variations, complex objects' motion, occlusions, image distortion phenomena, etc [2]. To improve object appearance, and especially its robustness in time, statistical models are used such as linear prediction schemes [34], Gaussian mixture models [11, 33], kernel density methods [19], Hidden Markov Models [28, 32] or deformable models [37]. However, these techniques fail in the case when partial/full occlusions occur and when the background color properties contain similar colors or textures to the tracked objects. A dynamic spatial bias appearance model, called DSBAM, is proposed in [2]. The model exploits on online learning strategies to improve robustness of object tracking by capturing the spatial coherence of the object appearance dynamically using local region confidences. Thus, the model is robust to partial occlusions and similar backgrounds.

The main, however, limitation of all the above-mentioned methods is that *there are no mechanisms to re-initialize the tracking algorithm in an automatic framework whenever its performance severely deteriorates*. This means that these methods suffer from re-adjusting and reconfiguration. Despite the effectiveness of the objects' appearance models, it is difficult to consider all possible variations of color distributions and texture properties since color clothes or paintings can be of any type. To overcome the difficulties we need an automatic recovery mechanism able to re-initialize the tracker each time its performance is unacceptable. Towards this direction, methods that combine object detection and tracking have been studied in the literature. In particular, a technique for unsupervised video segmentation that consists of two phases—the initial segmentation and the temporal tracking—is presented in [31]. The method initially applies a segmentation algorithm (either in color, or in motion domain) and then tracks the objects under a unsupervised framework. Thus, this method is only applicable for very simple visual content like video conferencing applications and it fails in complex motion phenomena usually encountering in surveillance applications. A multitarget tracking algorithm which exploits the meanshift descriptor combined with an automatic initialization process by a particle filter is shown in [36]. The method switches between the particle filter based detection and the mean-shift tracking and yields good results in simple outdoor-captured video sequences. Again this approach assumes a well structured environment. In the same framework, one of our previous works (the [7]) combines face detectors and depth information to efficiently

track humans applicable only in video conferencing sequences or generic objects in three-dimensional stereoscopic sequences. Thus, the method of [7] is not applicable to complicated sequences derived from surveillance applications. It also fails in case of complicated motions and occlusions. Similarly, in [18] tracking of multiple objects is accomplished using a coupled optimization problem. More specifically, the method is formulated in a Minimum Description Length hypothesis selection framework, which allows our system to recover from mismatches and temporarily lost tracks. Despite its ability, the method of [18] suffers from reconfiguration which permits automatic tracker initialization necessary in broad domain application scenarios.

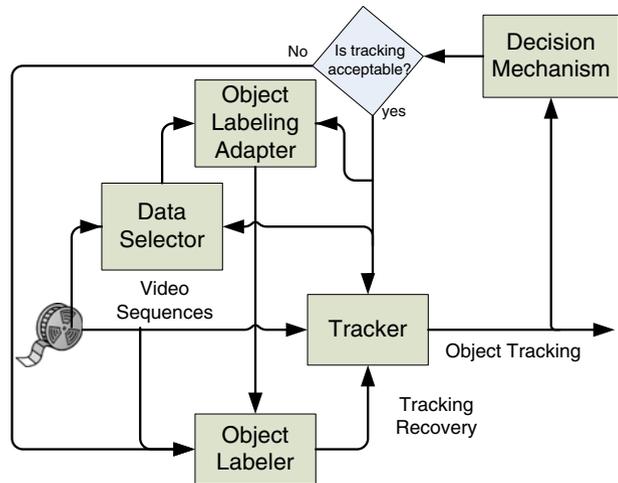
In this paper the aforementioned difficulties are addressed by proposing a novel framework able to *automatically recover* the results of a tracking algorithm whenever its performance is not acceptable. Recovery is accomplished by using adaptable non-linear object labeling methods. Labeling exploits appropriate visual descriptors so as to assign to image regions probabilities of belonging to one of the available objects in a scene. Two main difficulties, however, are encountered in object labeling. The first deals with the fact that the non-linear function that relates the visual descriptors with the desired classification outputs of the image regions is actually unknown, while the second with the fact that this function is also time varying due to the environmental changes of the visual content. To solve the first problem, in this paper, we exploit concepts derived from functional analysis to model any arbitrarily non-linear function (with some restrictions on its continuity) as a finite series of known functional components but of unknown coefficients [4, 17]. Thus, the problem is equivalent to the estimation of the unknown coefficients of the non-linear known functional components. The second issue is addressed by imposing the non-linear model to be varied through time. In this case, we introduce recursive parameter (coefficient) estimation methods, able to optimally update the non-linear object models to the new environmental conditions. In particular, the recursive strategy is implemented in a way that a) the non-linear model trusts as much as possible the current conditions, and (b) a minimal degradation of the already obtained knowledge is achieved. A decision mechanism is incorporated to activate tracking recovery. Special emphasis is given on the design of a computationally efficient recursive strategy so that it can be applied to real-life application scenarios.

This paper is organized as follows: Section 2 describes the architecture of the proposed tracking recovery algorithm. Section 3 discusses the particle filter methodology used for object tracking, while the formulation of the object labeler is shown in Section 4. The adaptation strategy of the object labeler is presented in Section 5, while Section 6 discusses the optimal data selection procedure. The decision mechanism is reported in Section 7. Experimental results are shown in Section 8 while Section 9 concludes the paper.

2 Dynamic tracking readjusting architecture

The goal of this paper is to improve the performance of an multiple object tracking algorithm in complex visual conditions, (such as illumination changes, occlusions (full or partial), background variations, non-rigid object motion), by proposing an architecture which enables *automatic tracking recovery* whenever unacceptable results are encountered. For this reason, we introduce an architecture presented in Fig. 1. The architecture consists of five (5) different modules; the tracker, the object labeler, the object labeling adapter, the data selector and the decision mechanism.

Fig. 1 The proposed tracking recovery architecture



Tracker This module is responsible for tracking multiple objects in a scene. Theoretically, any tracking algorithm can be incorporated in the proposed architecture, but in our case a particle-filter tracker has been implemented due to its efficiency and robustness in complex visual environments.

Object Labeler This module labels image regions as objects by taking into account appropriate visual descriptors. Object labeler is activated whenever the Decision Mechanism ascertains that the tracking performance is not acceptable. In the proposed architecture, non-linear relationships are incorporated in a recursive implementation so that the non-linear object models are self learnt from the environment and the tracking information. Then, the object labeler is used for tracking recovery by re-initializing the object samples at the positions where it is more probable to locate a tracked object.

Object Labeling Adapter Objects' visual characteristics are changing from time (frame) to time (frame). Thus, the use of a static (though non-linear) relationship that visually characterizes objects does not yield appropriate results, significantly deteriorating the tracker's performance. This is more evident for real-life surveillance applications in which occlusions, illumination variations, motion in the background, etc. are encountered. The role of this module is to dynamically readjusting the non-linear object model to fit the current environmental conditions. This module is activated whenever reliable tracked objects are identified. In particular, in this case, data selector is activated in order to describe the current visual conditions. Then, recursive learning strategies are activated to update the models of the object labeler module.

Data Selector The adapter should take into account information about the current visual content so as to update the models of the object labeler performance to the current conditions. This is achieved by this module, which uses an automatic process that picks up the most confident image regions from a set of reliable tracked objects.

Decision Mechanism This module is responsible for detecting those time instances (frames) in which the tracker performance cannot be considered as acceptable and thus recovery

should take place. The mechanism exploits the probabilistic nature of the tracker as well as the evolution of the tracker through time.

3 Particle filter tracker

In this paper, a particle filter approach is adopted for object tracking due its reliability and robustness in complicated visual environments. In the following, we briefly describe the particle filter methodology using the concepts of Sequential Importance Sampling [1].

Let us denote as $z_{1:k} = \{z_m, m = 1, 2, \dots, k\}$ a sequence of observable states. Let us also denote as $s_{0:k} = \{s_m, m = 0, 1, \dots, k\}$ a sequence of unobservable states of the target. Then, the probabilistic tracker is estimated by calculating the posterior conditional probability density function (pdf) $p(s_{0:k}|z_{1:k})$ of the states $s_{0:k} = \{s_m, m = 0, 1, \dots, k\}$ up to time k given the observable states $z_{1:k}$. Using the Bayes law, this function can be expressed as

$$p(s_{0:k}|z_{1:k}) = p(s_{0:k-1}|z_{1:k-1}) \frac{p(z_k|s_{0:k}, z_{1:k-1})p(s_k|s_{0:k-1}, z_{1:k-1})}{p(z_k|z_{1:k-1})} \tag{1}$$

In most practical problems, the state can be represented as a first order Markov process. Thus, $p(s_k|s_{0:k-1}, z_{1:k-1}) = p(s_k|s_{k-1})$. Similarly, the $p(z_k|s_{0:k}, z_{1:k-1}) = p(z_k|s_k)$. As a result Eq. 1 can be written as

$$p(s_{0:k}|z_{1:k}) = p(s_{0:k-1}|z_{1:k-1}) \frac{p(z_k|s_k)p(s_k|s_{k-1})}{p(z_k|z_{1:k-1})} \tag{2}$$

Exploiting the particle filter theory, the posterior $p(s_{0:k}|z_{1:k})$ can be represented by a set of weighted particles (samples), i.e., $\{s_{0:k}^i, q_k^i\}_{i=1}^N$. Each particle $s_{0:k}^i$ represents a potential trajectory of the state sequence and q_k^i denotes its likelihood estimated from the sequence of observations up to time k [23].

Then the posterior density could be approximated as:

$$p(s_{0:k}|z_{1:k}) \approx \sum_{i=1}^N q_k^i \delta(s_{0:k} - s_{0:k}^i) \tag{3}$$

where $\delta(x)$ represents Dirac function.

Since sampling directly from the posterior is usually impossible, the weights are chosen using the principle of Importance Sampling (IS) [23]. It should be noted that IS is a general technique for estimating the properties of a particular distribution $p(s_{0:k}|z_{1:k})$ while only having samples generated from a different distribution $r(s_{0:k}|z_{1:k})$ rather than the distribution of interest. Then, the proposed weights q_k^i in (3) can be given as

$$q_k^i \propto \frac{p(s_{0:k}^i|z_{1:k})}{r(s_{0:k}^i|z_{1:k})} \tag{4}$$

Assuming a factorize form for the $r(s_{0:k}|z_{1:k})$ as

$$r(s_{0:k}|z_{1:k}) = r(s_k|s_{0:k-1}, z_{1:k})r(s_{0:k-1}|z_{1:k}) \tag{5}$$

We can obtain the following recursive update equation [1]

$$q_k^i = \frac{\tilde{q}_k^i}{p(z_k|z_{1:k})} \text{ with } \tilde{q}_k^i = q_{k-1}^i \frac{p(z_k|s_{0:k}^i, z_{1:k-1})p(s_k^i|s_{0:k-1}^i, z_{1:k-1})}{r(s_k^i|s_{0:k-1}^i, z_{1:k})} \tag{6}$$

The factor \tilde{q}_k^i are the unnormalized weights for the i -th particle. In addition, the factor $p(z_k|z_{1:k})$ can be approximated by the sum $\sum_{i=1}^N \tilde{q}_k^i$ so that the weights q_k^i are indeed normalized.

Taking into account, the first order Markov process assumptions, as we have described above, we can re-write Eq. 6 as

$$\tilde{q}_k^i = q_{k-1}^i \frac{p(z_k|s_k^i)p(s_k^i|s_{k-1}^i)}{r(s_k^i|s_{0:k-1}^i, z_{1:k})} \text{ and } q_k^i = \frac{\tilde{q}_k^i}{\sum_{m=1}^N \tilde{q}_k^m} \tag{7}$$

In high dimensional spaces, i.e., for high values of variable k , sampling is inefficient since this leads to a continuous increase of the weight variance resulting in a selection of few particles only [6]. To solve this problem, we need to apply another re-sampling methodology that aims at eliminating the effect of particles with low importance weights and multiple particles that correspond to high weights values.

The efficiency of a particle filter algorithm relies on the definition of a good proposal distribution. One possible strategy is the one that minimizes the weight variance of the new samples at time k , given the observations $z_{1:k}$ and the particles $s_{1:k-1}^i$, In [6] it can be shown that in this case,

$$r(s_k^i|s_{0:k-1}^i, z_{1:k}) = r(s_k^i|s_{k-1}^i, z_{1:k}) = p(s_k^i|s_{k-1}^i, z_k) \propto p(z_k|s_k)p(s_k|s_{k-1}^i) \tag{8}$$

which leads to the following weight update

$$q_k^i \propto q_{k-1}^i p(z_k|s_{k-1}^i) \tag{9}$$

with the assumption that $\sum_{i=1}^N \tilde{q}_k^i = 1$.

In practice, $p(z_k|s_{k-1}^i)$ is only achievable in particular cases, such as Gaussian noise and linear observation models [1, 6]. Thus, alternatively, we can select the priori as importance function, i.e.,

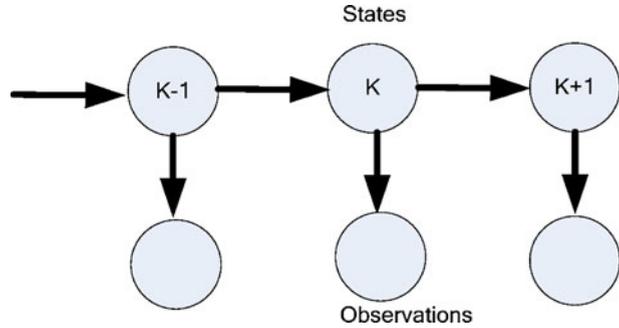
$$q_k^i \propto q_{k-1}^i p(z_k|s_k^i) \tag{10}$$

A graphical representation of such a model is shown in Fig. 2.

4 Object labeler

As we have mentioned in Section 1, despite the efficiency of a tracking algorithm, its performance can severely deteriorate due to visual complexities, such as abrupt motions, full/partial occlusions, motion in the background and/or illumination variations in the scene, etc. In this section, we propose a novel automatic mechanism able to improve the tracker performance whenever a severe deterioration takes place. In particular, each time the tracking performance is considered as unacceptable—this is defined by a decision mechanism described in Section 7—the object labeler is activated to recover tracking taking into account the current conditions. Object labeling takes as input visual descriptors and then it classifies image regions as objects with respect to these descriptors.

Fig. 2 A graphical representation for the adopted particle filter model



The proposed object labeling model satisfies two properties:

- (i) *A non-linear relationship* between the visual descriptors and the object models and
- (ii) *time varying object models* able to dynamically update the non-linear relationship of descriptors-objects to fit changes of the environment

Let us suppose that at the t -th video frame of a sequence, the proposed object labeling is activated since at this frame tracking yields erroneous results. In Section 7, we define how these time instance (frames) are defined. Let us also assume that the t -th frame have been divided into R regions, (e.g., blocks), and each of them is assigned to one of L available objects. Extracting for each image region, M descriptors $\mathbf{x}_i(t) \in R^M$ with $i=1,2,\dots,R$ M -dimensional vectors are formed. Let us denote as $O_j^{(t)}(\mathbf{x}_i(t)), j=1,2,\dots,L$ the probability of the i -th image region at t -th frame to be assigned to the j -th tracked object. Superscript (t) of $O_j^{(t)}$ expresses that probabilities are time varying.

Function $O_j^{(t)}(\mathbf{x}_i(t))$ is actually unknown since in real-life situations it is impossible to find an analytical non-linear relationship between descriptors and objects. For this reason, we initially parametrize the unknown function so that is can be expressed as a finite series of known functional components but of unknown coefficients [17]. That is,

$$O^{(t)}(\mathbf{x}_i(t)) \approx \sum_{k=1}^K v_k(t)\beta_k(\mathbf{x}_i(t)) \tag{11}$$

In Eq. 11, we have omitted subscript j for simplicity. The $v_k(t)$ are the unknown coefficients for this expansion, while $\beta_k(\mathbf{x}_i(t))$ are the known functional components which take as inputs the M -dimensional descriptor vectors $\mathbf{x}_i(t)$. Finally, K defines the approximation degree of such expansion. Larger values of K yield better approximation at an extent of an increase of the parameters number. Actually, the number K and the number of descriptors extracted M define the number of unknown coefficients in (11). Let us denote as D this number.

Usually, the functional components $\beta_k(\mathbf{x}_i(t))$ are considered to be of constant type. In this case, a scale parameter σ_k is introduced to modify the components, meaning that $\beta_k(\mathbf{x}_i(t)) = \beta(\sigma_k(t), \mathbf{x}_i(t))$. One common choice for the scaling parameter is through the inner product $\sigma_k(t) \cdot \mathbf{x}_i(t)$. In other words, the known functional components can be written as

$$\beta_k(\mathbf{x}_i(t)) = \beta(\sigma_k(t) \cdot \mathbf{x}_i(t)) = \beta\left(\sum_{m=1}^M \sigma_{k,m}(t) x_{i,m}(t)\right) \tag{12}$$

where $x_{i,m}(t)$ is the m -th element of vector $\mathbf{x}_i(t)$ and $\sigma_{k,m}(t)$ the respective m -th element of $\sigma_k(t)$.

Let us form in the following a matrix $\Sigma(t) = [\sigma_1(t) \cdots \sigma_K(t)]^T$. Then, the output $O^{(t)}(\mathbf{x}_i(t))$ is given as

$$O^{(t)}(\mathbf{x}_i(t)) \approx \sum_{k=1}^K v_k(t) \beta_k(\mathbf{x}_i(t)) = \mathbf{v}^T(t) \cdot \boldsymbol{\beta}(\Sigma(t) \cdot \mathbf{x}_i(t)) \tag{13}$$

In Eq. 13, $\mathbf{v}^T(t)$ is the vector that contains all the coefficients elements $v_k(t)$ while T denotes the transpose matrix. The $\beta(\Sigma(t) \cdot \mathbf{x}_i(t))$ is a vector-valued function given as $\beta(\Sigma(t) \cdot \mathbf{x}_i(t)) = [\beta(\sigma_1^T(t) \cdot \mathbf{x}_i(t)) \beta(\sigma_2^T(t) \cdot \mathbf{x}_i(t)) \cdots \beta(\sigma_K^T(t) \cdot \mathbf{x}_i(t))]^T$. As a result, $\beta(\Sigma(t) \cdot \mathbf{x}_i(t))$ returns a vector each element of which is the output of the functional component of the same input but for different scaling parameters.

The unknown components of Eq. 13 are the elements of vector $\mathbf{v}^T(t)$ and the scaling parameters $\Sigma(t) = [\sigma_1(t) \cdots \sigma_K(t)]^T$, which are also time varying. Time variation represent the fact that different relationships between descriptors-objects are encountered for different environments. In the following, we propose a novel adaption strategy which, based on the results of object labeling for the previous frames, the new unknown coefficients are recursively estimated.

5 Object labeling adapter

Assuming a slight modification of the non-linear function from time to time we can relate the model parameters as follows.

$$\mathbf{v}(t + 1) = \mathbf{v}(t) + d\mathbf{v} \text{ and } \Sigma(t + 1) = \Sigma(t) + d\Sigma \tag{14}$$

where $d\mathbf{v}$ and $d\Sigma$ are small perturbations of parameters \mathbf{v} and Σ .

Let us also assume that at the (t) frame, a reliable mask for all the L available objects is derived through the tracking algorithm. Then, the labels for all the L tracked objects and the background can be considered as known. Thus,

$$O^{(t+1)}(\mathbf{x}_i(t)) = I_i(t) \tag{15}$$

where $I_i(t)$ are the labels (IDs) for the i th image region at the t -th frame. Thus, $I_i(t)$ takes values in the range $[1 L]$ since we have assumed that L objects are available. In (15), the superscript $(t+1)$ means that the labeler output is calculated using the new model parameters, i.e., the $\mathbf{v}(t+1)$, $\Sigma(t+1)$.

Exploiting Eq. 14, we can linearize Eq. 13 using a first order Taylor series expansion. Then we can prove the following theorem.

Theorem 1 *The difference in object labeling of an image region using the coefficients $\mathbf{v}(t+1)$ and $\Sigma(t+1)$ and $\mathbf{v}(t)$, $\Sigma(t)$ is linearly related with the small perturbations $d\mathbf{v}$ and $d\Sigma$ while the parameters of the linear model only depend on the previous coefficients $\mathbf{v}(t)$, $\Sigma(t)$*

Proof Using the first order Taylor series expansion, we can express $\beta(\Sigma(t + 1) \cdot \mathbf{x}_i(t))$ in relation with $\beta(\Sigma(t) \cdot \mathbf{x}_i(t))$ as

$$\begin{aligned} \boldsymbol{\beta}(\Sigma(t + 1) \cdot \mathbf{x}_i(t)) &= \boldsymbol{\beta}((\Sigma(t) + d\Sigma) \cdot \mathbf{x}_i(t)) = \boldsymbol{\beta}(\Sigma(t) \cdot \mathbf{x}_i(t) + d\Sigma \cdot \mathbf{x}_i(t)) = \\ &= \boldsymbol{\beta}(\Sigma(t) \cdot \mathbf{x}_i(t)) + \mathbf{W} \cdot d\Sigma \cdot \mathbf{x}_i(t) \end{aligned} \tag{Pr1}$$

In Eq. Pr1, \mathbf{W} is a diagonal matrix that contains the first derivatives of $\beta(\Sigma(t) \cdot \mathbf{x}_i(t))$ with respect to the coefficients $\Sigma(t)$.

Taking into account Eqs. 14 and 15, we can relate object labeling using the new coefficients $\mathbf{v}(t+1)$ and $\Sigma(t+1)$, i.e., the $O^{(t+1)}(\mathbf{x}_i(t))$ with the previous ones as follows

$$\begin{aligned} O^{(t+1)}(\mathbf{x}_i(t)) &= \mathbf{v}^T(t+1) \cdot \beta(\Sigma(t+1) \cdot \mathbf{x}_i(t)) = \mathbf{v}^T(t+1) \cdot (\beta(\Sigma(t) \cdot \mathbf{x}_i(t)) + \mathbf{W} \cdot d\Sigma \cdot \mathbf{x}_i(t)) = \\ &= \mathbf{v}^T(t) \cdot \beta(\Sigma(t) \cdot \mathbf{x}_i(t)) + \mathbf{v}^T(t) \cdot \mathbf{W} \cdot d\Sigma \cdot \mathbf{x}_i(t) + d\mathbf{v}^T \cdot \beta(\Sigma(t) \cdot \mathbf{x}_i(t)) = \\ &= O^{(t)}(\mathbf{x}_i(t)) + \mathbf{v}^T(t) \cdot \mathbf{W} \cdot d\Sigma \cdot \mathbf{x}_i(t) + d\mathbf{v}^T \cdot \beta(\Sigma(t) \cdot \mathbf{x}_i(t)) \end{aligned} \tag{Pr2}$$

where in (Pr2) we have ignored second order terms since they minimally contribute to the total amount. Matrix \mathbf{A} in (Pr2) is the gradient of function β with respect to the previous model parameters.

As a result, the difference in object labeling for an image region, when the new parameters $\mathbf{v}(t+1)$, $\Sigma(t+1)$ are used is linearly related with the labeling using the previous parameters and the small perturbations $d\mathbf{v}$, $d\Sigma$.

Thus, from (Pr2), we can derive that

$$\begin{aligned} O^{(t+1)}(\mathbf{x}_i(t)) - O^{(t)}(\mathbf{x}_i(t)) &= I_i(t) - O^{(t)}(\mathbf{x}_i(t)) \\ &= \mathbf{v}^T(t) \cdot \mathbf{W} \cdot d\Sigma \cdot \mathbf{x}_i(t) + d\mathbf{v}^T \cdot \beta(\Sigma(t) \cdot \mathbf{x}_i(t)) \end{aligned} \tag{Pr3}$$

Taking for granted (15) and Theorem 1 [see Eq. Pr3], we can relate new object labels with the current ones through a linear equation of the form

$$O^{(t+1)}(\mathbf{x}_i(t)) = O^{(t)}(\mathbf{x}_i(t)) + \mathbf{H} \cdot d\mathbf{g} \tag{16}$$

where \mathbf{H} is a matrix including elements coming from the current coefficients $\mathbf{v}(t)$, $\Sigma(t)$ and $d\mathbf{g}$ a vector that contains all small perturbations $d\mathbf{v}$ and $d\Sigma$.

In order to reliably estimate the coefficients for the adaptive non-linear model, we should take into account the effect of all image regions R for all the available objects L (including the background). Thus, $W=R \times L$ linear equations of the form of (16) are created and the optimal values for the small increments $d\mathbf{g}$ can be estimated by solving the above mentioned linear system, i.e.,

$$d\mathbf{g} = \mathbf{H}^{-1} \cdot \mathbf{e} \tag{17}$$

where vector \mathbf{e} contains the object labeling differences for all regions and objects, i.e., $\mathbf{e} = [a_{1,1}, \dots, a_{R \times L}]^T$ where $a_{i,j} = I_{i,j}(t) - O_j^{(t)}(\mathbf{x}_i(t))$ for $i=1,2,\dots,R$, (we recall that R is the number of image regions in a frame) and $j=1,2,\dots,L$ (we recall that j is the number of available objects in a scene).

5.1 Estimating the new model coefficients

However, the unknown variables involved in Eq. 17 actually depend on the approximation order K of (11) and the number M of descriptors used in $\mathbf{x}_i(t) \in R^M$ to represent the visual content of an image region and we recall that this number is denoted as D . As a result, three difference cases can be obtained, which are examined in the following subsections. In particular, for a high number of descriptors and model parameters, required to achieve a low approximation error, it is probable that the number of number D of unknowns of (17) to be

greater than the number of linear equations W , (under-determined case). Instead, for low values of K and descriptor number M is to quite probable the number D of unknown to be smaller than the number of linear equations W , (over-determined case). Finally, when the number of unknowns equals the number of linear equations then, the small increments can be straightforwardly estimated by solving the linear system of (17). Both cases are low computational complexity.

5.1.1 Under-determined case

In this case, the number of unknowns is greater than the number of equations. As a result, we need additional constraints requirements for the coefficients in order to guarantee one probable solution, since otherwise, there is an infinity number of coefficients that can satisfy (17). As additional constraint, in this case, we impose the minimal deviation of the new coefficients from the current ones. This is expressed as

$$\min \|\mathbf{g}(t+1) - \mathbf{g}(t)\|_2 = \|\mathbf{d}\mathbf{g}\|_2 \quad (18a)$$

subject to

$$\mathbf{d}\mathbf{g} = \mathbf{H}^{-1} \cdot \mathbf{e} \quad (18b)$$

where \mathbf{g} is vector that contains all coefficients at frame instance t and $t+1$ denoted similarly to $\mathbf{d}\mathbf{g}$.

Taking into account this additional constraint and equation (17), we can obtain the optimal solution at the adaptation as

$$\mathbf{d}\mathbf{g} = (\mathbf{H}^T \cdot \mathbf{H})^{-1} \mathbf{H}^T \cdot \mathbf{e} \quad (19)$$

The solution of (19) is actually the minimal distance from the origin to the constraint hyper-surface of $\mathbf{e} - \mathbf{H} \cdot \mathbf{d}\mathbf{g} = 0$.

5.2 Over-determined case

This is the case, when the number of unknowns is smaller than the number of linear equations. Such a system has usually no solution. Thus, the goal is to find the values of the unknown parameters $\mathbf{d}\mathbf{g}$ which “best” fit the equations, in the sense of solving the quadratic minimization problem defined as

$$\min \sum \left| \sum h_{ij} \cdot \mathbf{d}g_j - e_i \right|^2 \quad (20)$$

where h_{ij} , $\mathbf{d}g_j$, and e_i are elements of matrix \mathbf{H} , and vectors $\mathbf{d}\mathbf{g}$ and \mathbf{e} respectively. This minimization problem has a unique solution, provided that the columns of the matrix \mathbf{H} are linearly independent. This solution coincides with the solution of (19).

6 Data set selector

The small modification of the model parameters $\mathbf{d}\mathbf{g}$ requires the calculation of the matrix \mathbf{H} and vector \mathbf{e} . Matrix \mathbf{H} depends only on current model parameters, which have been already estimated from the previous steps of the algorithm as expressed through the parameters $\mathbf{v}(t)$, $\Sigma(t)$. Vector \mathbf{e} is the difference between the model output when no adaptation takes place

(i.e., using the current model parameters) and approximate labels of the objects $I_i(t)$ used to describe the current visual environment. Labels $I_i(t)$ can be supervisedly (manually) provided but in such a case we loose the automatic operation of the proposed architecture. For this reason, in this section, we describe an algorithm for selecting the most reliable image regions of a frame as objects regions and then exploiting these labels to update in an automatic fashion the model parameters.

The proposed data selection algorithm exploits the tracking performance at a previous frame in which reliable results are derived (the probability of (1) takes high values). Since, however, this mask does not coincide with the mask of the current visual environment, a refinement mechanism is included in the process to discard regions that present low confidence from being part of an object and simultaneously retain regions that are characterized by high confidence. The introduction of the refinement mechanism is necessary since otherwise it is high probable to select vague regions as object labels diluting the efficiency of the adaptation.

To localize the lowest and highest confident image regions, we follow the procedure described next. Initially, we detect all image regions that have been assigned to an object by the tracker (at a previous reliable time instance), and then we estimate the region that is closest to the center of gravity of the tracking output. Let us denote as μ_x, μ_y the x and y -coordinate of this region. Then, we tag all objects' regions as the output of a two-dimensional independent Gaussian probability density, the mean value of which is the μ_x, μ_y coordinates.

$$p_g(r_x, r_y) = \frac{1}{2\pi \cdot a_h \cdot a_v} \exp\left(-\frac{(r_x - \mu_x)^2}{2a_h^2}\right) \cdot \exp\left(-\frac{(r_y - \mu_y)^2}{2a_v^2}\right) \tag{21}$$

where r_x, r_y is the x and y coordinates of an image region in the respective block and p_g the probability density function of the Gaussian.

The standard deviation a_h, a_v are estimated as follows. Let us denote as h_l, h_r the most left and right image region of an object, with $h_l < h_r$. Let us also denote as v_t, v_b the respective most top and bottom region of the object, with $v_t < v_b$. Then, we assume that the area $h_l - h_r$ and $v_t - v_b$ is within the pdf with a confidence interval (CI) of $q\%$. Taking into account that properties of the Gaussian pdf, we can express that the cumulative distribution between $\mu_x - na_h$ ($\mu_y - na_v$) and $\mu_x + na_h$ ($\mu_y + na_v$) where n is any arbitrary number such that

$$\int_{\mu_x - na_h}^{\mu_x + na_h} p_g dr_x = \text{erf}\left(\frac{n}{\sqrt{2}}\right) \text{ x - dimension} \tag{22a}$$

$$\left(\int_{\mu_y - na_v}^{\mu_y + na_v} p_g dr_y = \text{erf}\left(\frac{n}{\sqrt{2}}\right) \right) \text{ y - dimension} \tag{22b}$$

and erf is the error function. Thus, when $n=4$, the confidence interval (CI) that we derive is 99.9936657516%. Since the inverse error function erf can be also defined, we are able to find a value for n that satisfies q . In particular, in case that we assume that the most left and right part of the image regions are within the pdf with an interval of 99.99%, then, the value of $n=3.8906$, and thus, $2 \cdot 3.8906 \cdot a_h = h_l - h_r \Rightarrow a_h = \frac{h_l - h_r}{7.7812}$. Similarly, and for the same confidence interval the $a_v = \frac{v_t - v_b}{7.7812}$.

Having defined the mean and the standard deviation of (21), we can tag all image regions of an object with respect to their distance from the center of gravity of a reliably estimated mask at a previous time instance. Then, we select as the most confident regions within an object the ones whose confidence interval is within 66% i.e., the regions that are within $[\mu_x - a_h, \mu_x + a_h]$ and $[\mu_y - a_v, \mu_y + a_v]$.

Figure 3 presents a graphical representation of the proposed method adopted for optimal data selection. In this case, a reliable tracked mask has been detected and the most left, right, bottom and top lines of the region have been detected. Then, the center of gravity of the region is calculated and the standard deviation to achieve 99.99% confidence interval for those lines is estimated. In the following, we select as data the ones lying within a 66% confidence interval.

7 The decision mechanism

The goal of the decision mechanism is to automatically detect those time instance (frames) which tracking recovery should take place since the performance of the tracking algorithm cannot be considered as acceptable. Upon such a decision, the object labeler is activated to classify image regions as objects and then it exploits these results to re-initialize the tracker. For the object labeling, an optimal selection strategy is required to pick up the most reliable image regions of an object able to represent the current visual conditions. The adapter is also activated to improve object labeling in case of complex visual changes of the environment.

To yield a reliable outcome of the the decision mechanism, two conditions are taken into account in this paper. The first exploits the probabilistic nature of the tracker while the second exploits the fact that, between two successive frames, the position of an object does not significantly change.

For the first condition, we use the results of Eq. 2. In particular, in case that the probability values of (2) are low, we result in a low confident tracking and thus recovery is more probably. Instead, if the probability values of (2) are high, the confidence in tracking

Fig. 3 A graphical representation of the proposed optimal data selection algorithm

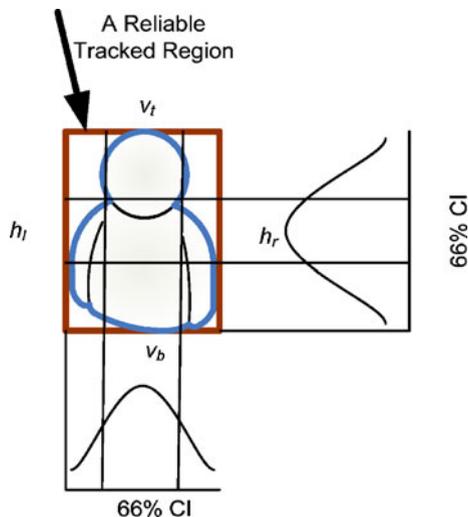


Table 1 Algorithmic form of the proposed tracking recovery scheme

Summary of the Proposed Tracking Recovery Algorithm

1. If the decision mechanisms considers that the tracker performance is not acceptable using Eq. 22a, 22b then
2. Estimate the object labels $I_i(t)$ using the Gaussian probability of (21) and a confidence interval of 66%.
3. Estimate vector e as the difference between $I_i(t)$ and the outputs of the labeler using the previous model parameters $O_j^{(t)}(\bar{x}_i(t))$.
4. Estimate matrix \mathbf{H} as a function of previous model parameters through (Pr2) and (16).
5. Estimate the small modification of the model parameters using Eq. (17).
6. Update the model parameters as in (14).
7. Re-initialize the samples of the tracker using the object labels as in (13) with the new model parameters

is also high and thus it is more probable to need no recovery. That is,

$$DM1 = \begin{cases} 1 & p(s_{0:k}|z_{1:k}) < T_1 \text{ and recovery is probable} \\ 0 & p(s_{0:k}|z_{1:k}) \geq T_1 \text{ and recovery is not probable} \end{cases} \quad (23a)$$

Similarly, let us denote as $U_j^{(t)}$ the set of pixels that has been assigned to the j -th object at the t -th time instance through the application of the tracker. If this set significantly deviates from the set at the previous time instance ($t-1$)th then recovery is probable since a dramatic change from one frame to the other has been detected. On the contrary, if the sets $U_j^{(t-1)}$ and $U_j^{(t)}$ contain almost the same regions, a consistent monitoring of the objects is expected. That is, the second condition for the decision mechanism is

$$DM2 = \begin{cases} 1 & F(U_j^{(t)}, U_j^{(t-1)}) > T_2 \text{ and recovery is probable} \\ 0 & F(U_j^{(t)}, U_j^{(t-1)}) \leq T_2 \text{ and recovery is not probable} \end{cases} \quad (23b)$$

The case that both $DM1$ and $DM2$ are 1 occurs when a significant visual change of the environment takes place with a simultaneous low confidence of the tracker. In this case, undoubtedly, tracking recovery should take place. On the opposite case that both $DM1$ and $DM2$ are 0, no recovery is activated. The vague case of $DM1$ is 1 and $DM2$ is 0 occurs



Fig. 4 Tracking results for a characteristic shot of PETS sequence using without the proposed recovery strategy



Fig. 5 Tracking results for a characteristic shot of PETS sequence using with the use of the proposed recovery strategy

when the tracker monitors an object with low confidence without, however, a significant evident of an environmental change. This actually indicates a non-reliable tracking and recovery should take place if this situation is repeated for a certain number of frames. In case that $DM1$ is 0 and $DM2$ is 1, a significant visual change takes place but the tracker is still able to follow the objects with high confidence. Despite the fact that this case does not actually require a recovery process, in our implementation, we re-initialize the tracker if this case continues for a certain number of frames just to achieve a more reliable performance.

Table 1 summarizes the main step of the proposed tracking recovery methodology.

8 Experimental results

In the following, we evaluate the performance of the proposed tracking recovery algorithm in a set of different video sequences, which present complex phenomena such as occlusions, illumination changes, presence of multiple objects, etc. In particular, in Section 8.1 we discuss the video sequences used in this paper to assess the performance of the proposed scheme, while experimental results of real-life video objects, moving in complex conditions are depicted in Sections 8.2 and 8.3. The results have been evaluated using either subjective criteria (see Section 8.2), by depicting the tracking performance before and after the recovery, or objective measurements (see Section 8.3).

8.1 Video sequences details

A set of different video sequences have been included in this paper to evaluate the efficiency of the proposed tracking recovery architecture. The sequences include indoor and outdoor environments. The latter presents high illuminations fluctuations. Some sequences

Fig. 6 The results of the adaptable object labeling module before, during and after the occlusion



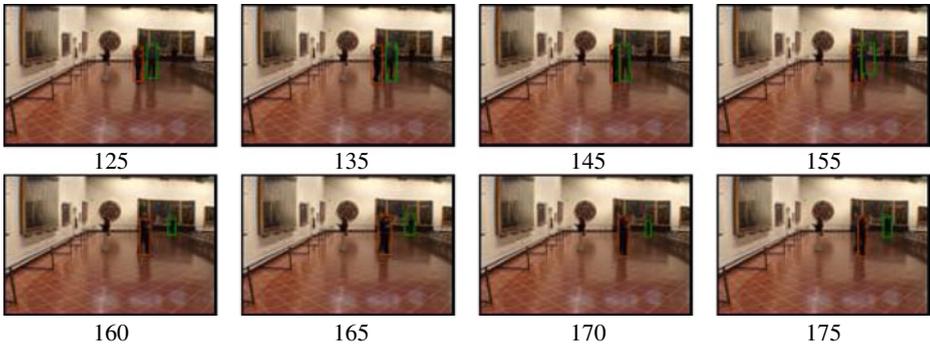


Fig. 7 Tracking results for a characteristic shot of Uffizzi sequence using without the proposed recovery strategy

are publicly available, such as the PETS one, so as to compare our results under a common framework. Some of them have been recorded under the framework of European Union funded research projects (such as POLYMNIA [8] and SCOVIS [10]) and present complex situations, like partial or full occlusions, background movements, illumination changes and presence of multiple objects. This way, we are able to evaluate the performance of the proposed tracking recovery scheme in real-life complex conditions.

Figure 4 shows a characteristic shot from the PETS sequence in which a full occlusion takes place. The sequence depicts persons discussing in a meeting room (indoor environment of almost constant illumination). Another two characteristic examples of indoor sequences of almost constant illumination are the ones presented in Figs. 7 and 10. The first have been recorded inside the famous Gallery of the Uffizzi Museum in Florence for surveillance purposes under the framework of POLYMNIA project. The dataset depicts multiple visitors of the museum, looking at the exhibitions and yielding complex visual occlusions. Additionally, the objects of the individuals are very small compared to the frame size, making tracking a difficult process. The second sequence has been recorded in Demokritos research laboratory for the purposes of SCOVIS project and visually presents very complex motions of the persons individuals. Finally, Fig. 12 shows a characteristic shot of an outdoor sequence in which, as depicted, illumination variations are noticeable. This sequence has been recorded under the framework of POLYMNIA project in the

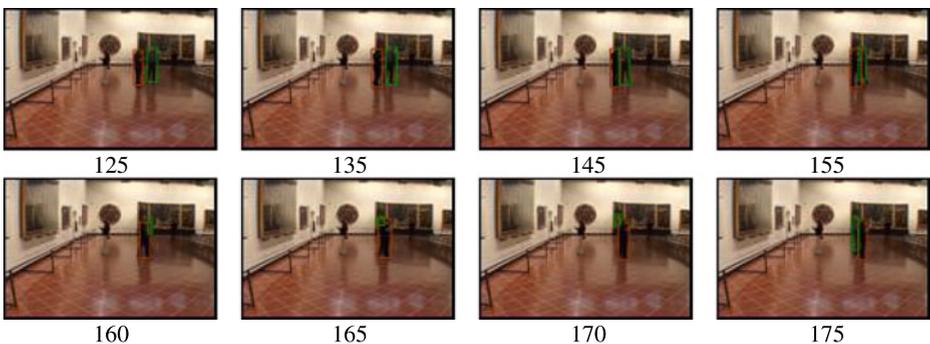


Fig. 8 Tracking results for a characteristic shot of Uffizzi sequence using with the use of the proposed recovery strategy



Fig. 9 The results of the adaptable object labeling module before, during and after the occlusion

premises of an open thematic park. The sequence presents persons coming for a ride to the bumper cars.

8.2 Tracking re-adjustment—subjective evaluation

Figure 4 shows the results of the adopted particle filter–based tracking for the shot PETS sequence. The specific shot depicts 19 frames in which a full occlusion is encountered. As we observe, the tracker performance deteriorates in the occluded regions since it is difficult in this case to monitor the correct trajectory of the objects. We also notice that tracking is deteriorated after the occlusion since the algorithm cannot initialize correct the samples at the previous video frames. The results after the proposed tracking recovery scheme are shown in Fig. 5. We observe a significant improvement of the tracking performance, robust to the full occlusion.

Tracking recovery is assisted through the proposed adaptable object labeler. In particular, whenever the tracker performance is considered as unacceptable by the Decision mechanism either due to the fact that the probabilistic tracking is not reliable or due to a dramatic change of the environment (see Section 7) the object labeler is activated to re-initialize the objects regions that are to be tracked. Figure 6 shows the results of the adaptable object labeling at frames before, during and after the full occlusion in which the tracking performance deteriorates. In all cases, blocks 8×8 have been detected as image regions, while the DC along with some of the 9 zig-zag scanned AC coefficients of each block are used as appropriate visual descriptors. We notice that correct labeling is accomplished even for this complex visual content case.

The same results for the Uffizzi Museum sequence are depicted in Figs. 7 and 8. In this particular case, we have selected a shot consisting of 53 frames in which a partial occlusion

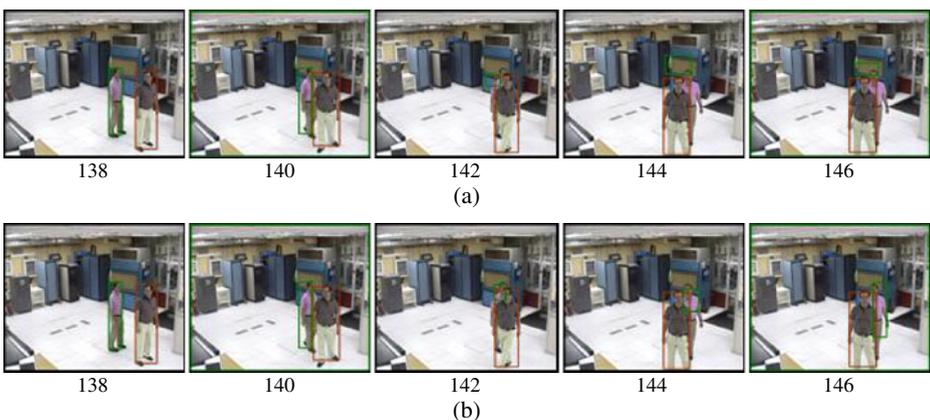


Fig. 10 Tracking results for a characteristic shot of Demokritos sequence **a** without and **b** with the use of the proposed recovery strategy

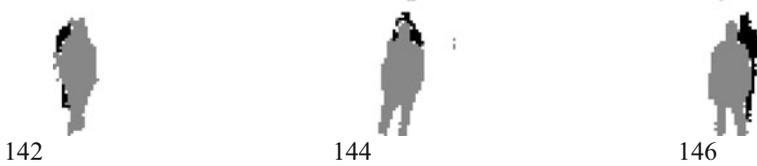


Fig. 11 The results of the adaptable object labeling module before, during and after the occlusion

between two persons takes place. In the shot, there is also two other persons that they are almost still. Fig. 7 shows tracking results for the two moving persons, one Lady and a Man without the use of the recovery mechanism, while the results after the recovery are shown in Fig. 8. As we observe tracking is almost correctly accomplished for the Lady who stands in front of the Man, while the tracking of the Man deteriorates just before, during and after the partial occlusion. In those time instance, the Decision Mechanism is activated to recover tracking by exploiting the results of the adaptable object labeling scheme. The results of the labeler after the automatic adaptation to the current environment are shown in Fig. 9. The results indicate that the labeler correctly classifies the image blocks as objects. As descriptors the same extracted in Fig. 6 have been selected.

The tracking recovery results for a characteristic shot of 9 frames of the Demokritos sequence without and with tracking recovery is also depicted in Fig. 10. Again, the proposed automatic recovery scheme significantly improves the tracking performance. The object labeling results for this case are shown in Fig. 11. Correct classification is accomplished again.

Finally, Fig. 12(a) shows the results for a single object person for the outdoor sequence of POLYMNIA without the use of the recovery strategy. Due to the illumination variations, tracking performance deviates from the moving object and sometimes is tracked to other neighboring objects of similar visual characteristics. The results after recovery are shown in Fig. 12(b) in which significant improvement is accomplished.

8.3 Tracking re-adjustment-objective evaluation

The previous results are based on a subjective evaluation of the proposed tracking recovery scheme by depicting the results in several visually complex, either indoor or outdoor video

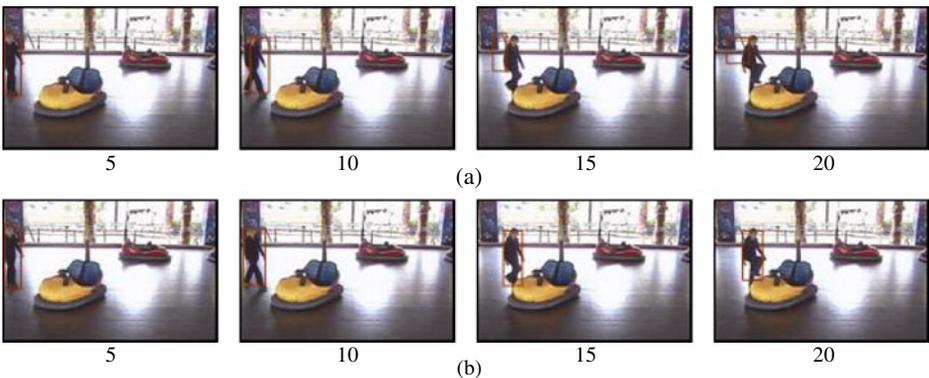


Fig. 12 Tracking results for a characteristic shot of POLYMNIA sequence **a** without and **b** with the use of the proposed recovery strategy

sequences. In this section, we introduce objective criteria able to assess the tracking performance of the proposed scheme and compare it with other traditional approaches.

Let us denote as R the reference mask of the actual object region. Let us also denote as T the tracked image region. Then, the

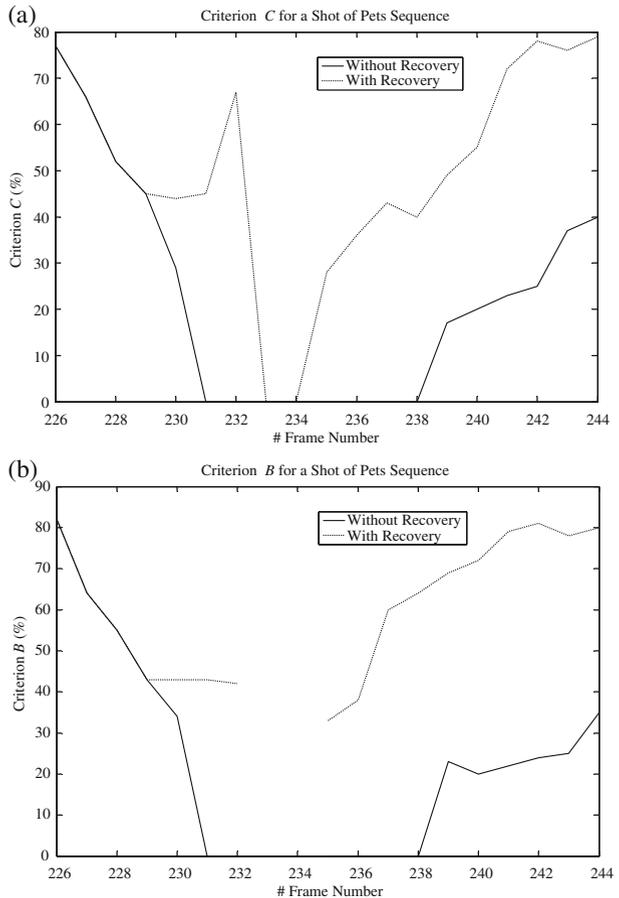
$$C = \frac{R \cap T}{T} \tag{24}$$

expresses how close is the tracked mask with the reference one. As a result, values of C close to 1 indicate that the tracked region coincides with the object. Otherwise, the tracked region is far away from the object.

The criterion of (24), however, is not adequate since it is possible large parts of the reference actual object to be located outside the tracked mask even though when C takes values close to one. This is for example the case when the tracked mask coincides with a part (even small) of the object. For this reason, there is the need for another criterion, defined as

$$B = 1 - \frac{R \cap T^c}{R} \tag{25}$$

Fig. 13 The performance of the shot of Fig. 4 (Pets) for the objective criteria **a** C and **b** B



Equation 25 presents the percentage of the reference object that is located within the tracked mask. Again, values of B close to one correspond to the case that only a small part of the actual object is found outside the tracked area.

In case that both criteria take high values, this corresponds to the case that a correct tracking performance is accomplished since the tracked mask contains the largest proportion of the object without leaving outside a significant part of it. Instead, the case that C is low while B is high refers to the fact that, although the tracked mask mostly contains the reference object, it is much largest than the reference one including other image regions. Similarly, when C is high and B is low, the tracked mask locates only a subset (probably small) of the actual reference object. When both criteria take low values, the tracked mask is far away from the object.

Figures 13–16 shows the results of both criteria C and B for the shots of Pets, Uffizzi, Demokritos and Polymnia sequence respectively. The results are depicted with and without tracking recovery. It is clear that the proposed recovery strategy significantly improves the tracking performance especially in case that complex visual phenomena such as occlusions and illumination changes are encountered.

These criteria have been also extracted for 15,000 frames of different sequences and their average values are presented in Table 2. It is clear that the proposed tracking recovery

Fig. 14 The performance of the shot of Fig. 7 (Uffizzi) for the objective criteria **a** C and **b** B

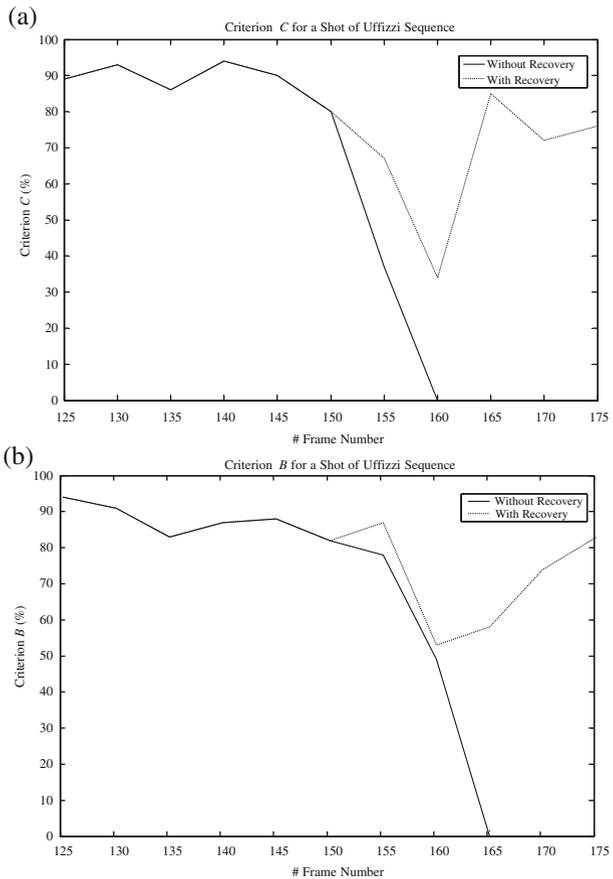
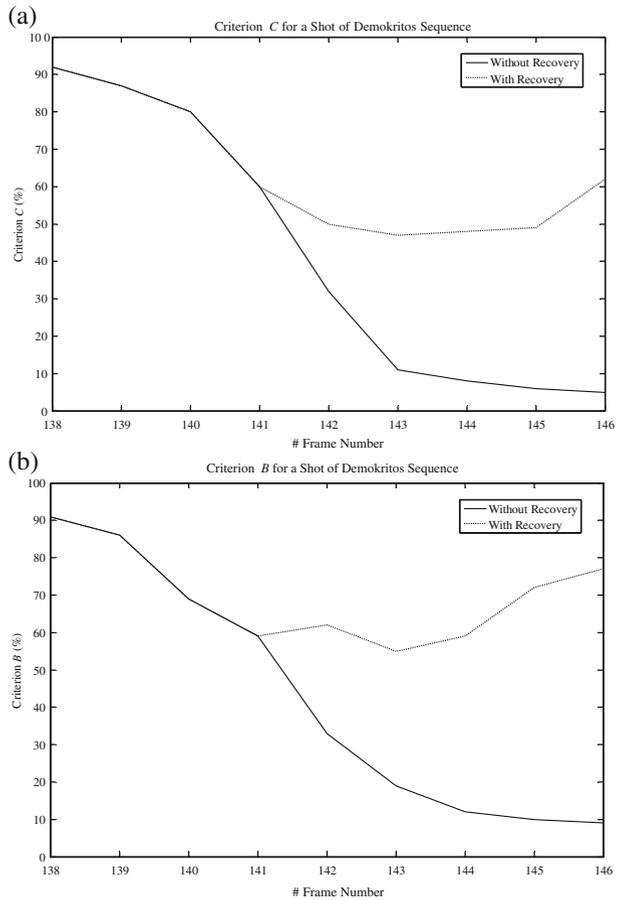


Fig. 15 The performance of the shot of Fig. 10 (Demokritos) for the objective criteria **a** C and **b** B



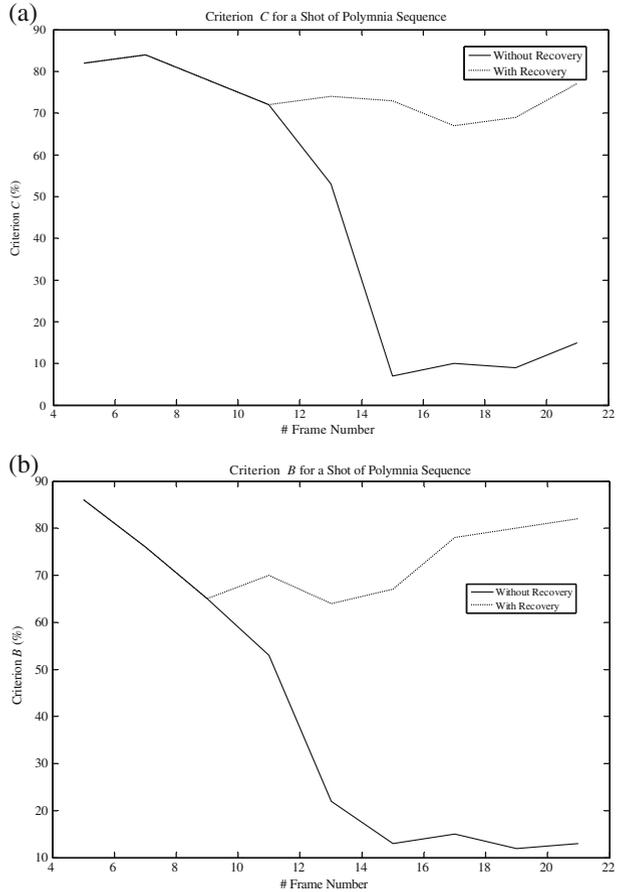
scheme improves the performance but this improvement is more evident in complex visual environments.

Computational Complexity The proposed tracking re-adjustment method requires low computational complexity since the adopted algorithm concludes to a convex minimization problem subject to linear constraints. In real life applications, in which computational complexity is crucial, one can approximate the solution of the proposed method by early stopping the algorithm. Although an early stopping would result to a deviated solution from the optimal one, the tracking performance will be better than the one provided without the use of re-adjustment.

9 Conclusions

Action, behavior and/or workflow analysis from video data, is a research attractive topic nowadays especially due to the rapid increase of surveillance systems and the increasing need for monitoring crucial infrastructures for security, safety or product quality purposes. Motion analysis and detection/tracking of moving objects is probably one of the most important issues towards event analysis since the moving objects are the ones that act on

Fig. 16 The performance of the shot of Fig. 12 (Polymnia) for the objective criteria **a** *C* and **b** *B*



the environment. However, the main difficulties in accurately tracking of moving objects in real-life visual environments are due to complex visual phenomena, such as occlusions (full or partial), illumination variations, agile motion, complex background, etc.

The research effort in object tracking can be discriminated into three categories, (i) the motion models (ii) the search methods and (iii) the appearance-based techniques. Despite however the technique used, we need an automatic recovery mechanism able to re-initialize the tracker each time its performance is unacceptable. This is addressed in this paper by proposing a novel tracking recovery technique which automatically labels image regions as objects using non-linear object models. The adopted non-linear models are time-varying since the visual characteristics of the objects change from time to time. For this reason, concepts derive from functional analysis are adopted to parametrize the non-linear model

Table 2 Average Values of Criteria *C* and *B* over several different video sequences

Without Recovery		With Recovery	
C	B	C	B
46.78%	73.53%	48.67%	75.44%

and then linearization tools are applied to find the new model parameters to fit the current visual conditions. The architecture is enhanced with a decision mechanism able to verify the time instances in which tracking recovery from take place.

The efficiency and robustness of the proposed scheme has been tested on a set of real-life video sequences in which complex motions (full and partial occlusions), illumination changes and presence of multiple objects in the scene are encountered. The evaluation has been performed subjectively by comparing the results among effective tracking methods (like the particle filter one) with the proposed recovery methodology. Additionally, two criteria are presented to objectively assess the tracking recovery performance and compare it with other approaches presented in the literature.

As future work, we are going to apply adaptable neural network models for modeling the non-linear object labeler mechanisms. The advantage of such an implementation is that it can be combined with a hardware implementation making the system applicable in embedded architectures that require low processing and memory requirements.

Acknowledgement This work is supported by the European Union funded project SCOVIS “Self Configurable Cognitive Video Supervision” supported by the Seventh Framework Programme (FP7/2007–2013) under grant agreement no 216465.

References

1. Arulampalam S, Maskell S, Gordon N, Clapp T (2002) A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking. *IEEE Trans Signal Process* 50(2):174–188
2. Chen D, Yang J (2007) Robust object tracking via online dynamic spatial bias appearance models. *IEEE Trans Pattern Anal Mach Intell* 29(12):2157–2169
3. Comanicu D, Ramesh V, Meer P (2000) Real-time tracking of non-rigid objects using mean shift. *Proc Int'l Conf Computer Vision and Pattern Recognition*, pp 142–149
4. Cybenko G (1989) Approximation by superpositions of a sigmoidal function. *Math Control, Signal Syst* 2:303–314
5. Davatzikos C, Prince J, Bryan R (1996) Image registration based on boundary mapping. *IEEE Trans Med Imag* 15(1):112–115
6. Doucet A, Godsill S, Andrieu C (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat Comput* 10(3):197–208
7. Doulamis A, Doulamis N, Ntalianis K, Kollias S (2003) An efficient fully unsupervised video object segmentation scheme using an adaptive neural-network classifier architecture. *IEEE Trans Neural Netw* 14(3):616–630
8. Doulamis A, Kosmopoulos D, Christogiannis C, Varvarigou D (2004) Polymnia: personalised leisure and entertainment over cross media intelligent platforms. *European Workshop on Integration of Knowledge, Semantics and Digital Media Technology*, London, UK, 25–26 November 2004
9. Doulamis A, van Gool L, Nixon M, Varvarigou T, Doulamis N (2008) First ACM international workshop on analysis and retrieval of events, actions and workflows in video streams. *16th ACM International Conference on Multimedia*, Vancouver, Canada, October 2008
10. Doulamis A, Kosmopoulos D, Sardis E, Varvarigou T (2008) An architecture for a self configurable video supervision. *ACM Workshop on Analysis and Retrieval of Events, Actions and Workflows in Video Streams*, Vancouver, Canada, October 2008
11. Grimson W.E.L., Stauffer C (1999) Adaptive background mixture models for real-time tracking. *Proc Int'l Conf Computer Vision and Pattern Recognition*, pp 22–29
12. Heisele B, Kressel U, Ritter W (1997) Tracking non-rigid, moving objects based on color cluster flow. *Proc Int'l Conf Computer Vision and Pattern Recognition*, pp 253–257
13. Isard M, Blake A (1998) Condensation c conditional density propagation for visual tracking. *Int J Comput Vis* 29(1):5–28
14. Jeyakar J, Venkatesh Babu R, Ramakrishnan KR (2008) Robust object tracking with background-weighted local kernels. *Comput Vis Image Underst* 112:296–309

15. Jodoin P-M, Mignotte M, Konrad J (2007) Statistical background subtraction using spatial cues. *IEEE Trans Circuits Syst Video Technol* 17(12):1758–1763
16. Kass M, Witkin A, Terzopoulos D (1988) Snakes: active contour models. *Int J Comput Vis* 1(4):321–331
17. Kreyszig E (1989) *Introductory functional analysis with applications*. Wiley, New York
18. Leibe B, Schindler K, Cornelis N, Van Gool L (2008) Coupled object detection and tracking from static cameras and moving vehicles. *IEEE Trans Pattern Anal Mach Intell* 30(10):1683–1698
19. Leichter I, Lindenbaum M, Rivlin E (2009) Tracking by Affine Kernel transformations using color and boundary cues. *IEEE Trans Pattern Anal Mach Intell* 31(1):164–171
20. Lucas B, Kanade T (1981) An iterative image registration technique with an application to stereo vision. *Proc DARPA Image Understanding Workshop*, pp 121–130
21. Medeiros H, Park J, Kak A (2008) Distributed object tracking using a cluster-based Kalman filter in wireless camera networks. *IEEE J Sel Topics Signal Process* 2(4):448–463
22. Nascimento JC, Marques JS (2006) Performance evaluation of object detection algorithms for video surveillance. *IEEE Trans Multimedia* 8(4):761–774
23. Odobez J-M, Gatica-Perez D, Ba SO (2006) Embedding motion in model-based stochastic tracking. *IEEE Trans Image Process* 15(11):3515–3531
24. Rasmussen C, Hager GD (2001) Probabilistic data association methods for tracking complex visual objects. *IEEE Trans Pattern Anal Mach Intell* 23(6):560–576
25. Shahrokni A, Drummond T, Fua P (2004) Texture boundary detection for real-time tracking. *Proc European Conf Computer Vision*, pp 566–577
26. Shi J, Tomasi C (1994) Good features to track. *Proc Int'l Conf Computer Vision and Pattern Recognition*, pp 593–600
27. Smith GJD (2004) Behind the Screens: examining constructions of deviance and informal practices among CCTV control room operators in the UK. *Surveill Soc* 2(2/3):376–395
28. Stenger B, Ramesh V, Paragios N, Coetzee F, Bouhman J (2002) Topology free hidden markov models: application to background modeling. *Proc Int'l Conf Computer Vision*, pp 294–301
29. Tekalp M (1995) *Digital video processing*. Prentice Hall PTR, ISBN 0131900757
30. Tsai D-M, Lai S-C (2008) Independent component analysis-based background subtraction for indoor surveillance. *IEEE Trans Image Process* 18(1):158–167
31. Wang D (1998) Unsupervised video segmentation based on watersheds and temporal tracking. *IEEE Trans Circuits Syst Video Technol* 8(5):539–546
32. Wang P, Ji Q (2005) Multi-view face tracking with factorial and switching HMM. *Seventh IEEE Workshops on Application of Computer Vision, 2005. WACV/MOTIONS 1*:401–406
33. Wang H, Suter D, Schindler K, Shen C (2007) Adaptive object tracking based on an effective appearance filter. *IEEE Trans Pattern Anal Mach Intell* 29(9):1661–1667
34. Yang J, Waibel A (1996) A real-time face tracker. *Proc Workshop Computer Vision*, pp 142–147
35. Yilmaz A, Li X, Shah B (2004) Contour based object tracking with occlusion handling in video acquired using mobile cameras. *IEEE Trans Pattern Anal Mach Intell* 26(11):1531–1536
36. Yonemoto S, Sato M (2008) Multitarget Tracking Using Mean-shift with Particle Filter based Initialization. *IEEE 12th International Conference Information Visualization*, pp. 521–526
37. Zhong Y, Jain AK, Dubuisson-Jolly M-P (2000) Object tracking using deformable templates. *IEEE Trans Pattern Anal Mach Intell* 22(5):544–549



Anastasios D. Doulamis received the Diploma degree in Electrical and Computer Engineering from the National Technical University of Athens (NTUA) in 1995 with the highest honor. In 2000, he has received the PhD degree in electrical and computer engineering from the NTUA. From 1996–2000, he was with the Image, Video and Multimedia Lab of the NTUA as research assistant. From 2001 to 2002, he serves his mandatory duty in the Greek army in the computer center department of the Hellenic Air Force, while in 2002, he join the NTUA as senior researcher. His PhD thesis was supported by the Bodosakis Foundation Scholarship. In 2006, he is Assistant professor in the Technical University of Crete in the area of multimedia systems.

Dr. Doulamis has received several awards and prizes during his studies, including the Best Greek Student in the field of engineering in national level in 1995, the Best Graduate Thesis Award in the area of electrical engineering with A. Doulamis in 1996 and several prizes from the National Technical University of Athens, the National Scholarship Foundation and the Technical Chamber of Greece. In 1997, he was given the NTUA Medal as Best Young Engineer. In 2000, he received the best Phd thesis award by the Thomaidion Foundation in conjunction with N. Doulamis.

In 2001, he served as technical program chairman of the VLBV'01. He has also served as program committee in several international conferences and workshops. He is reviewer of IEEE journals and conferences as well as and other leading international journals. He is author of more than 200 papers in the above areas, in leading international journals and conferences.

His research interests include, non-linear analysis, neural networks, multimedia content description, intelligent techniques for video processing.