PERGAMON

Pattern Recognition 35 (2002) 2711–2718

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Improving the robustness of parametric shape tracking with switched multiple models ☆

J.C. Nascimento*, J.S. Marques

Electrotecnia e Computadores, Instituto Superior Tecuico (IST), University of Tecnica de Lisboa, Torre Norte, IST, Av. Rovisco Pais, 1049-001 Lisboa, Portugal

Received 31 October 2001; accepted 31 October 2001

Abstract

This paper addresses the problem of tracking objects with complex motion dynamics or shape changes. It is assumed that some of the visual features detected in the image (e.g., edge strokes) are outliers i.e., they do not belong to the object boundary. A robust tracking algorithm is proposed which allows to efficiently track an object with complex shape or motion changes in clutter environments. The algorithm relies on the use of multiple models, i.e., a bank of stochastic motion models switched according to a probabilistic mechanism. Robust filtering methods are used to estimate the label of the active model as well as the state trajectory.

Keywords: Robust multi-model tracker (RMM); Kalman multi-model tracker (KMM); Dynamic models; Gaussian; Outliers; Strokes

1. Introduction

Object tracking has various applications in the scope of medical diagnosis, surveillance and human–machine interface. It is usually assumed that the object shape and motion slowly vary during the observation interval, being described by a finite set of parameters.

Typically, a stochastic difference equation is used to describe the evolution of the shape and motion parameters. These assumptions are not valid in the presence of abrupt changes or complex parameter trajectories which can only be described by nonlinear dynamic systems. Two examples concern the estimation of the lips boundaries e.g., for speech recognition or face animation and the estimation of the human motion [1].

To deal with these difficulties, the use of switched dynamical models was recently advocated i.e., a set of models switched according to a probabilistic rule, each of them being tailored to a specific motion regime or shape evolution [2–4]. Two problems have to be addressed if we want to use switched dynamical models for shape tracking. First, given a video sequence we have to determine which model is active at each instant of time. This is the labeling problem. Second, we have to estimate the state of the active model using the available data. This amounts to estimating the shape and motion parameters of the object to be tracked. These problems have been addressed either by nonparametric techniques [2] or by parametric ones, based on the propagation of Gaussian mixtures [3,4].

Although switched models are able to describe complex motion and shape evolution, they fail in the presence of outliers i.e., if the image measurements contain invalid data. Typically, the tracker loses the object boundary when wrong edge points are detected in the image, e.g., edge points belonging to the background or to inner regions of the object to be tracked. This is a major drawback which prevents the application of such models in many tracking problems. This difficulty is addressed here. A robust tracking algorithm is presented which extends the method described in Refs. [3,4]. The proposed tracker is based on two key concepts. First, middle level features (strokes) are used instead of low level ones (edge points). Second, two labels (valid/invalid) are considered for each stroke. These techniques have been used

☆ This work was partially supported by FCT under project TMO.

* Corresponding author. Tel.: +351-21-841-82-70; fax: +351-21-841-82-91.

E-mail address: pcjan@pop.isr.ist.utl.pt (J.C. Nascimento).

0031-3203/02/$22.00
PII: S0031-3203(01)00244-8

with success when a single model is adopted to track the object [5]. The proposed algorithm allows a robust performance of the switched multiple model tracker in the presence of outliers.

2. Multiple dynamical models

In order to estimate the object position and deformation, three steps are considered [6]: contour prediction, image measurement and contour update. The first step predicts the position of the object boundary in the next image. The second step computes image features in the vicinity of the predicted contour e.g., by sampling the predicted contour at equally spaced points. The third step uses the image measurements to update the contour estimate. It is assumed that image features (edges points) either belong to the boundary of the object to be tracked or they are produced by the background or inner edges (outliers). The main difficulty lies in the presence of false alarms or detection failures which produce undesirable effects. One way to deal with this situation is by considering that each feature is either valid or invalid. This approach is not practical since it involves 2^N hypotheses (data interpretations), N being the number of detected features (sometimes hundreds). We adopt a different approach to reduce the number of hypotheses. The edge points are linked in M strokes, and each stroke is classified either as valid or invalid. This reduces the number of hypotheses to 2^M, with $M \ll N$.

Let $x(t) \in \Re^n$ be a vector containing the shape parameters of the object to be tracked (e.g., control points of a spline curve). We assume that the state vector is generated by a set of stochastic difference equations [7]

$$x(t) = A_{k(t-1),k(t)} x(t-1) + w(t), \tag{1}$$

where $w(t) \sim \mathcal{N}(0, Q_{k(t-1),k(t)})$ is a white Gaussian noise, $k(t) \in \{1, \ldots, m\}$ is the label of the active model at instant t and m is the number of steady-state models (see Fig. 1). Additional models are considered at transitions (matrices A, Q depend on $k(t)$ and $k(t-1)$). It is assumed that the label sequence $k(t)$ is a first order Markov process with the transition probability

$$T_{rq} = p(k(t) = q \mid k(t-1) = r), \tag{2}$$

where $r, q \in \{1, \ldots, m\}$, and m is the number of steady-state models. Switched dynamic models were studied in control theory and aeronautics to deal with abrupt changes in dynamic systems (e.g., see Refs. [7,8]). The application of these models in object tracking has been considered in Refs. [3,4], assuming that all the data points are valid. In this paper the available observations are the strokes detected in the image. However, we do not know which strokes belong to the object boundary and should therefore be considered as valid. Since this information is not available, a label (valid/invalid) is assigned to each stroke and all the label

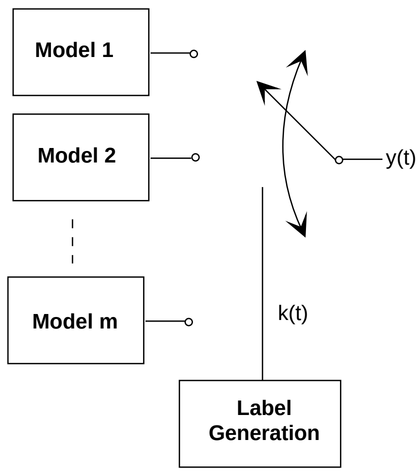

Fig. 1. Bank of switched models.

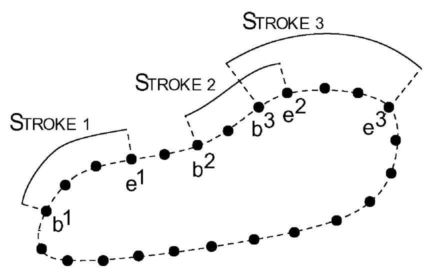

Fig. 2. Predicted shape and image strokes.

sequences are considered. Each label sequence is denoted as a *data interpretation*. An interpretation I_i is defined as a binary sequence $I_i^1, \ldots, I_i^M$, where $I_i^j \in \{0, 1\}$ is the label of the jth stroke in the ith interpretation.

Let $y(t)$ be the vector of all image features detected at instant t and let $y_i(t)$ be a vector with all valid features according to the ith interpretation, $(y_i(t) \subset y(t))$. It will be assumed that the sensor model for the ith interpretation is given by

$$y_i(t) = C_i x(t) + \eta(t), \tag{3}$$

where $\eta(t) \sim \mathcal{N}(0, R_i)$ is a white Gaussian noise and C_i is the observation matrix associated to the ith interpretation.

Fig. 2 shows an example in which there are 2^3 interpretations. A possible interpretation is $I_i = (1\,1\,0)$. In this case, matrix C_i includes the rows associated with the indices $\{b^1, \ldots, e^1, b^2, \ldots, e^2\}$ of the image features which are considered as true. The observation matrices C_i, C_j associated with two interpretations I_i, I_j are different since the observation vectors y_i, y_j contain different data features and they usually have different dimensions.

The state of switched multiple model is characterized by the transition density $p(x(t), k(t) \mid x(t-1), k(t-1))$, which can be split as follows:

$$p(x(t),k(t)\,|\,x(t-1),k(t-1))$$
$$=p(x(t)\,|\,k(t),x(t-1),k(t-1))$$
$$p(k(t)\,|\,x(t-1),k(t-1)). \quad (4)$$

The first factor depends on the dynamic Eq. (1) while the second is an element of the transition matrix of the Markov chain $T_{k(t-1),k(t)}$.

3. Density propagation

The problem to be solved can be formulated as follows: given a set of observations $Y^t=\{y(1),\ldots,y(t)\}$ which may contain outliers, what is the best estimate of the state and model label $\hat{x}(t)$, $\hat{k}(t)$. This is a nonlinear filtering problem. If the joint probability density function, conditioned on the observations is evaluated $p(x(t),k(t)\,|\,Y^t)$, estimates of the unknown parameters $(\hat{x}(t),\hat{k}(t))$ can be obtained by using the maximum a posteriori (MAP) method

$$(\hat{x}(t),\hat{k}(t))=\arg\max_{x(t),k(t)} p(x(t),k(t)\,|\,Y^t). \quad (5)$$

Using the law of total probabilities, the a posteriori density becomes

$$p(x(t),k(t)\,|\,Y^t)=\sum_{K^{t-1}} p(x(t),k(t),K^{t-1}\,|\,Y^t)$$
$$=\sum_{K^{t-1}} p(x(t)\,|\,K^t,Y^t)p(K^t\,|\,Y^t)$$
$$=\sum_{K^{t-1}} c_{K^t}\,p(x(t)\,|\,K^t,Y^t), \quad (6)$$

where $c_{K^t}=p(K^t\,|\,Y^t)$ and $K^t=\{k(1),\ldots,k(t)\}$ is the model label sequence up to instant t. Since $p(x(t)\,|\,K^t,Y^t)$ is a Gaussian density, the joint density $p(x(t),k(t)\,|\,Y^t)$ defined in Eq. (6) is a mixture of Gaussians, each of them being associated to a different label sequence K^t.

The computation of the mixture modes depends on the method being used. If all the observations are valid, each $p(x(t)\,|\,K^t,Y^t)$ (Gaussian component) could be updated by Kalman filtering and this is the optimal solution [4]. However, when $y(t)$ is contaminated with outliers robust filtering methods must be adopted. In fact, assuming that the model sequence K^t is known, the mean and covariance matrix can be computed using the S-PDAF method, recently proposed in Ref. [5] inspired by the work of Bar-Shalom and Fortmann [9] in the context of target tracking. To update the coefficients c_{K^t} in the switched model case, a new update law is required. After straightforward manipulation (see Appendix A)

$$c_{K^t}=\gamma c_{K^{t-1}}T_{k(t-1)k(t)}\sum_i \alpha_i(t)\prod_{j=1}^{M}\prod_{n=b^j}^{e^j}\mathscr{E}_i^j(s_n,t), \quad (7)$$

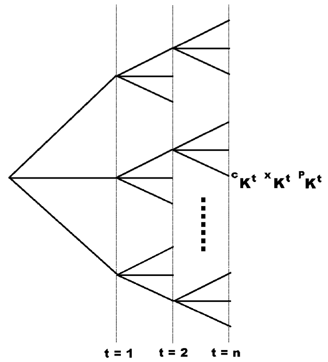

Fig. 3. Tree structure of RMM tracker ($m=3$).

where γ is a normalization constant, $c_{k(t-1)}$ is the predicted mixture coefficient, $T_{k(t-1)k(t)}$ is an element of the transition matrix of the Markov chain, $\alpha_i(t)$ is the association probability assigned to the data interpretation $I_i(t)$, M is the number of strokes, b^j, e^j are the indices of the jth stroke, $\mathscr{E}$ is a normal or uniform distribution, depending on the stroke j being considered as valid/invalid on the interpretation $I_i(t)$.

The Kalman filter is a particular case of S-PDAF (see Appendix A) since a single model is used and all the data is considered as valid. Therefore, $\mathscr{E}$ becomes independent of j and i. In this case, Eq. (7) can be written as

$$c_{K^t}=\gamma c_{K^{t-1}}T_{k(t-1)k(t)}\prod_{n=1}^{L}\mathscr{E}(s_n,t). \quad (8)$$

The state mean and covariance matrix estimates are updated by S-PDAF, and given by (see Ref. [5] for details)

$$\hat{x}_{K^t}=\hat{x}(t\,|\,t-1)+\sum_{i=1}^{m_t}\alpha_i(t)K_i(t)v_i(t), \quad (9)$$

$$P_{K^t}=\left[I-\sum_{i=1}^{m_t}\alpha_i(t)K_i(t)C_i\right]P(t\,|\,t-1)$$
$$+\sum_{i=0}^{m_t}\alpha_i(t)x_i(t)x_i(t)^{\mathrm{T}}-\hat{x}(t\,|\,t)\hat{x}(t\,|\,t)^{\mathrm{T}}, \quad (10)$$

where $K_i(k)$, $v_i(k)$ are the Kalman gain and innovation associated to the interpretation $I_i(k)$. The filter defined in Eqs. (6)–(10) will be denoted as Robust Multi-Model tracker (RMM).

The computation of Eqs. (7), (9) and (10) is organized in a tree structure, each branch being characterized by (see Fig. 3), x_{K^t}, P_{K^t} and c_{K^t}. The structure illustrated in Fig. 3 suggests that the number of leaves (Gaussian mixtures) increases as time passes by. Assuming that we have m label values, the mixture will have m^t modes at time t. It is crucial to limit the growth, in order to obtain a practical solution. Several strategies can be applied to achieve this goal, e.g., by using mode merging and elimination [3]. In this paper, the second method is adopted by discarding the mixture components with small coefficients.

Let us now consider the estimation of the unknown variables $x(t)$, $k(t)$. The model label is estimated using the MAP method as follows:

$$\hat{k}(t) = \arg\max_q P\{k(t) = q \mid Y^t\}$$

$$= \arg\max_q \sum_{K^{t-1}} p\{k(t) = q, K^{t-1} | Y^t\} \quad (11)$$

$$= \arg\max_q \sum_{K^t:\, k(t)=q} c_{K^t} \quad (12)$$

In the case of the state vector, the mean square method was used instead, for the sake of simplicity (see Appendix B), leading to

$$q\hat{x}_{K^t} = \gamma \sum_{K^t:\, k(t)=q} c_{K^t} \sum_i \alpha_i(t) x_i(t \mid t). \quad (13)$$

The state estimate is a weighted sum of the estimates associated to the tree paths ending with the q label.

Shape representation and feature detection are described in the next section.

4. Object tracking

4.1. Contour shape representation

To represent a moving object in a given frame t, it is assumed that the object boundary is a transformed version of a reference shape plus shape deformation [6]. Let $r(s): I \to \Re^2$ be a parametric representation of the object boundary. It is assumed that [1]

$$r(s) = \mathcal{T} r_r(s) + d(s) + v(s), \quad (14)$$

where $\mathcal{T}$ is a geometric transformation (e.g., affine transformation), r_r, d and v are the parametric descriptions of the reference shape, deformation and measurement noise, respectively. For the sake of simplicity, these curves are described by B-splines. It will be assumed that $\mathcal{T}$, d can be expressed in terms of a small number of parameters which are updated by the RMM filter and v is a white noise process.

Several transforms can be considered (e.g., translation, Euclidean similarities, affine transform) [6]. The affine transform is a flexible solution, since it allows to represent the motion of planar objects in 3D space. In this case the object boundary is

$$r_1(s_i) = a_1 r_{r1}(s_i) + a_2 r_{r2}(s_i) + a_3 + d_1(s_i) + v_1(s_i),$$

$$r_2(s_i) = a_4 r_{r1}(s_i) + a_5 r_{r2}(s_i) + a_6 + d_2(s_i) + v_2(s_i), \quad (15)$$

[1] Other works assume that $r_k = \mathcal{G}_k(r_r + d) + v$, [1,10]. Both approaches have advantages and disadvantages.

where $r(s) = (r_1(s), r_2(s))$, $r_r(s) = (r_{r1}(s), r_{r2}(s))$, $a_1, \ldots, a_6$ are the motion parameters at instant t and $v(s)=(v_1(s), v_2(s))$ is the measurement noise curve. Furthermore, it will be assumed that shape deformation is a linear combination of N_c deformation modes i.e.,

$$d(s) = \sum_{k=1}^{N_c} d_k \phi_k(s), \quad (16)$$

where $\phi_k(s)$ are known deformation curves and $d_1, \ldots, d_{N_c}$ are 2D vectors.

Assuming that the object moves during the acquisition process, dynamical equations have to be devised to describe the evolution of shape and motion parameters. Let $x(t)$ denote the vector of unknown motion and shape parameters

$$x = [a_1, \ldots, a_6, d_{x1}, \ldots, d_{xN_c}, d_{y1}, \ldots, d_{yNc}]^{\mathrm{T}}, \quad (17)$$

and let y be a $2L \times 1$ vector obtained by sampling the object boundary at L equally spaced points

$$y = [r_1(s_1), \ldots, r_1(s_L), r_2(s_1), \ldots, r_2(s_L)]^{\mathrm{T}}. \quad (18)$$

Eq. (15) can be written as follows:

$$y(t) = Cx(t) + v(t), \quad (19)$$

where

$$C = \begin{bmatrix} M & O & B & O \\ O & M & O & B \end{bmatrix}, \quad (20)$$

$$M = \begin{bmatrix} r_{r1}(s_1) & r_{r2}(s_1) & 1 \\ r_{r1}(s_2) & r_{r2}(s_2) & 1 \\ \vdots & \vdots & \vdots \\ r_{r1}(s_L) & r_{r2}(s_L) & 1 \end{bmatrix}, \quad B = \begin{bmatrix} \phi_1(s_1) \cdots \phi_{Nc}(s_1) \\ \phi_1(s_2) \cdots \phi_{Nc}(s_2) \\ \cdots\cdots\cdots\cdots \\ \phi_1(s_L) \cdots \phi_{Nc}(s_L) \end{bmatrix}. \quad (21)$$

In Eq. (20) M is a $L \times 3$ matrix, B is a $L \times N_c$ B-spline interpolation matrix, and O is a null matrix of appropriate dimensions. Similar expressions can be derived for other motion models.

4.2. Feature detection

Feature detection is performed by searching along the normal direction at specific contour points as suggested in Refs. [1,6,10]. The length of the inspection interval depends on the uncertainty associated with the predicted contour and it is given by

$$\rho(s_i, t) = \delta \sqrt{n(s_i)^{\mathrm{T}} S(s_i, t) n(s_i)}, \quad (22)$$

where $n(s_i)$ is the unit normal at s_i, and

$$S(s_i, t) = C(s_i) \left(\sum_{k(t)} c_{k(t) \mid k(t-1)} P(k(t) \mid k(t-1)) \right) \times C(s_i)^{\mathrm{T}} + R(s_i) \quad (23)$$

is a covariance matrix of the predicted boundary point $s = s_i$ considering all the available models. In Eq. (23), $C(s_i)$ is a matrix formed by lines i and $i + L$ of C.

Each feature is detected by comparing the image profile with shifted versions of a 1D template T. This procedure is based on the minimization of a cost function given by

$$\mathscr{J}(t_0) = \int_t |p(t) - T(t - t_0)|^2 \, \mathrm{d}t, \tag{24}$$

where $p(t)$ is the image profile along the direction orthogonal to the object boundary, t is the distance to the object boundary and $T(t)$ is the template. All the local minima below a given threshold are considered as features. Therefore, several features may be detected for a single contour sample.

4.3. Experimental results

The RMM tracker was used in the estimation of the boundaries of objects with significant shape changes. An example of lip tracking is presented. A comparison between the proposed method and the Kalman multi-model (KMM) filter described in Ref. [4] is given.

Two dynamic models were used to describe lip motion. The first model (model 1) performs a vertical contraction of the object boundary using the boundary estimate computed at the previous frame. The second model (model 2) expands the object contour (see Fig. 4). The first model is tailored to describe the evolution of the lip contour while the mouth is closing while the second model is useful when the mouth is opening.

In Fig. 5, we can see the performance of the KMM and RMM trackers using the same input data. It is clear that the outliers produced by the nose and teeth introduce strong distortions in the shape estimates obtained with the KMM tracker, leading to useless results after a few frames. The robust tracker described in this paper overcomes this difficulty and exhibits good tracking performance in this experiment.

A more difficult situation is presented in Fig. 6. In this case, the KMM tracker loses the boundary of the lips and fails to estimate the correct dynamic model. Fig. 6 shows the results given by RMM filter (second row) showing a remarkable robustness with respect to outliers. We have even increased the search area during the feature detection phase, therefore allowing more outliers. The algorithm selects the expansion model in these frames since it is the one which describes best the opening of the mouth. The robustness of the RMM is shown even in the presence of a large number of clutter features.

Fig. 4. Contour prediction using models 1 (dots); and 2 (dashed line).

Fig. 7 show the performance of the RMM algorithm in the presence of sudden shape changes. Three consecutive frames are shown in this figure. The use of multiple models allows us to track sudden changes of motion or shape deformation. The expansion model is selected in this example to track the opening of the mouth. The predicted contours obtained by both models is also displayed showing that the expansion model performs better in these three frames.

5. Conclusions

A new algorithm has been described for tracking of moving objects in video sequences. This allows the use of switched dynamic models, modeled by a bank of stochastic

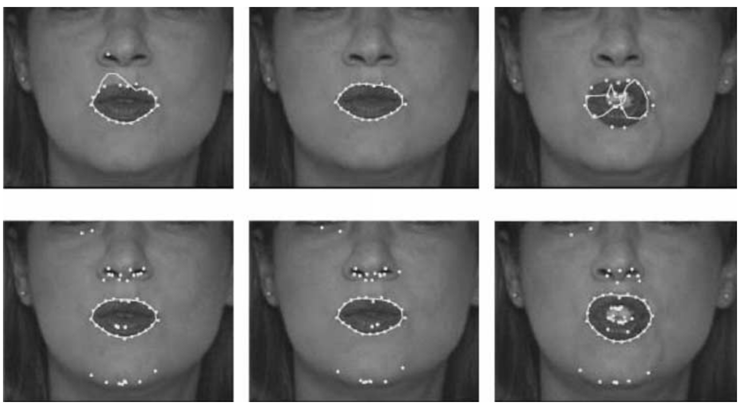

Fig. 5. Lip tracking with KMM tracker (first row, active model: 2 1 1) and RMM tracker (second row, active model: 2 1 2), (frames 8, 9, 13).

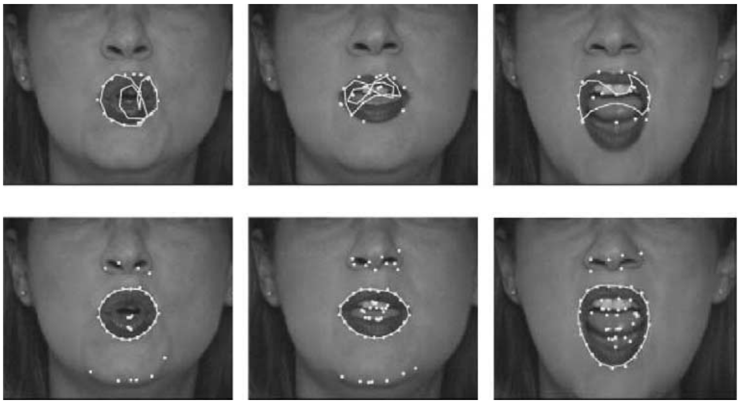

Fig. 6. Lip tracking with KMM tracker (first row, active model: 2 1 1) and RMM tracker (second row, active model: 2 2 2), (frames 16, 27, 46).

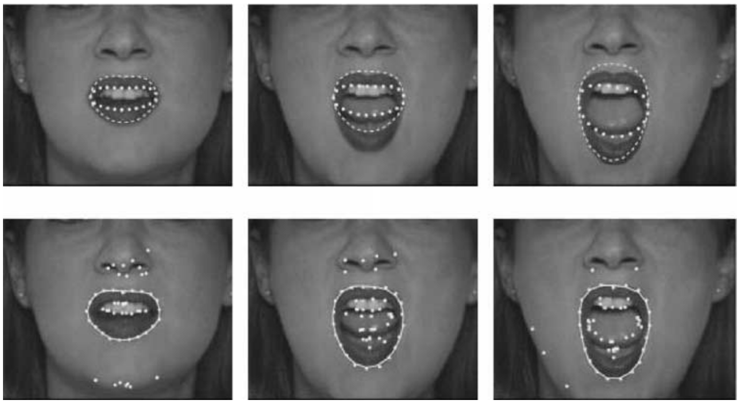

Fig. 7. Lip tracking with RMM: predicted contours (first row) and estimated contours (second row), (frames 45, 46, 47), active model: 2 2 2).

difference equations. Furthermore, it is assumed that the visual features detected in the image contain outliers, i.e., invalid features which do not belong to the object boundary. A robust filtering algorithm is proposed which is able to deal with multiple dynamics and invalid observations. This is accomplished by computing the propagation of the a posteriori density using Gaussian mixtures. Experimental results presented in the paper show that significant improvements are achieved, compared to the results obtained by the Kalman MM filter which was recently proposed in Refs. [3,4]. The algorithm was tested in lip tracking tasks. It was experimentally observed that the proposed method efficiently copes with the presence of abrupt shape changes and noisy measurements corrupted by outliers. This is clearly seen in some test sequences in which the mouth suddenly opens or closes. The RMM tracker still manages to estimate the lips contours well in these cases.

Appendix A

A.1. Mixture coefficients for S-PDAF model

$$c_{K^t} \triangleq \frac{p(K^t, Y^t)}{p(Y^t)} = \frac{1}{P(Y^t)} \int p(y(t) \mid K^t, Y^{t-1}, x(t)) \times p(K^t, Y^{t-1}, x(t))\, \mathrm{d}x(t)$$

$$= \frac{1}{p(Y^t)} \int \sum_i p(y(t) \mid I_i(t), K^t, Y^{t-1}, x(t)) \times p(I_i(t) \mid k(t), Y^{t-1}, x(t))\, p(K^t, Y^{t-1}, x(t))\, \mathrm{d}x(t)$$

$$= \frac{1}{p(Y^t)} \int \sum_i \alpha_i(t)\, p(y(t) \mid I_i(t), K^t, Y^{t-1}, x(t)) \times p(k(t) \mid K^{t-1}, Y^{t-1}, x(t))\, p(K^{t-1}, Y^{t-1}, x(t))\, \mathrm{d}x(t)$$

$$= \frac{1}{p(Y^t)} T_{k(t-1)k(t)} \sum_i \alpha_i(t)$$

$$\times \int p(y(t) \mid I_i(t), k(t), Y^{t-1}, x(t)) p(x(t) \mid K^{t-1}, Y^{t-1})$$

$$\times p(K^{t-1}, Y^{t-1}) \, \mathrm{d}x(t)$$

$$= \gamma T_{k(t-1)k(t)} c_{K^{t-1}} \sum_i \alpha_i(t)$$

$$\times \int p(y(t) \mid I_i(t), k(t), Y^{t-1}, x(t)) p(x(t) \mid K^{t-1}, Y^{t-1}), \tag{A.1}$$

with $\gamma = p(Y^{t-1})/p(Y^t)$.

Since $y(k)$ may contain some gaps along the contour, it depends on the localization of the strokes detected in the image. Thus, the probability $p(y(t) \mid I_i(t), k(t), Y^{t-1})$ can be written as

$$p(y(t) \mid I_i(t), k(t), b, e, M, Y^{t-1}), \tag{A.2}$$

where $b = \{b^1, \ldots, b^M\}$, $e = \{e^1, \ldots, e^M\}$ defines the beginning and the end of the strokes. Assuming that all features are independently generated, i.e.

$$p(y(t) \mid I_i(t), k(t), b, e, M, Y^{t-1})$$

$$= \prod_{j=1}^{M} \prod_{n=b^j}^{e^j} p(y^j(s_n, t) \mid I_i^j(t)), \tag{A.3}$$

where $y^j(s_n, t)$ is the feature point belonging to the jth stroke detected in the vicinity of s_n. It is assumed that visual features have uniform distributions in the search area if they are classified as unreliable ($I_i^j = 0$) and they are generated with a Gaussian distribution if they are classified as reliable. We define

$$\mathscr{E}_i^j(s_n, t) = p(y^j(s_n, t) \mid I_i^j(t))$$

$$= \begin{cases} V(s_n, t)^{-1} & \text{if } I_i^j(t) = 0, \\ \rho^{-1} \mathcal{N}(v^j(s_n, t); 0, S(s_n, t)) & \text{otherwise.} \end{cases} \tag{A.4}$$

where $V(s_n, t)$ is the volume of the search area, ρ is the normalization constant, $v^j(s_n, t) = y^j(s_n, t) - C(s_n)x(t \mid t-1)$ is the innovation associated to the jth stroke and $S(s_n, t) = C(s_n)P(t \mid t-1)C(s_n)^{\mathrm{T}} + R(s_n)$ is the covariance of the innovation vector where $C(s_n)$ and $R(s_n)$ are the output matrix and noise covariance associated to the nth sample of the object contour. Replacing Eq. (A.4) by Eq. (A.3), in Eq. (A.1)

$$c_{K^t} = \gamma c_{K^{t-1}} T_{k(t-1)k(t)} \sum_i \alpha_i(t) \prod_{j=1}^{M} \prod_{n=b^j}^{e^j} \mathscr{E}_i^j(s_n, t). \tag{A.5}$$

A.2. Mixture coefficients for Kalman model

The Kalman model is a particular case of S-PDAF, which corresponds to assuming that all the observed data is valid. Eq. (A.1) can be written as

$$c_{K^t} = \gamma T_{k(t-1)k(t)} c_{K^{t-1}} \int p(y(t) \mid k(t), Y^{t-1}, x(t))$$

$$\times p(x(t) \mid K^{t-1}, Y^{t-1}). \tag{A.6}$$

Assuming independence of the L features along the contour we can write

$$c_{K^t} = \gamma c_{K^{t-1}} T_{k(t-1)k(t)} \prod_{n=1}^{L} \mathscr{E}(s_n, t). \tag{A.7}$$

$\mathscr{E}(s_n, t)$ is similar to Eq. (A.4) and it is defined as

$$\mathscr{E}(s_n, t)$$

$$= \begin{cases} V(s_n, t)^{-1} & \text{if no features detected,} \\ \rho^{-1} \mathcal{N}(v(s_n, t); 0, S(s_n, t)) & \text{otherwise.} \end{cases} \tag{A.8}$$

$v(s_n, t)$, $S(s_n, t)$ have the same meaning as before, however the superscript j and subscript i are suppressed since we do not have strokes interpretations (all strokes are valid).

Appendix B

B.1. State update assuming the model $k(t) = q$

$$q\hat{x}_{K^t} \triangleq E\{x(t) \mid Y^t, k(t) = q\}$$

$$= \int x(t) p(x(t) \mid Y^t, k(t) = q) \, \mathrm{d}x(t)$$

$$= \int \frac{x(t) p(x(t), k(t) = q \mid Y^t)}{p(k(t) = q)} \, \mathrm{d}x(t)$$

$$= \frac{1}{p(k(t) = q)} \int x(t) \sum_{K^{t-1}} p(x(t), k(t)$$

$$= q, K^{t-1} \mid Y^t) \, \mathrm{d}x(t) \tag{B.1}$$

$$= \frac{1}{p(k(t) = q)} \int x(t) \sum_{K^t: k(t) = q} c_{K^t}$$

$$\times \sum_i p(x(t), I_i(t) \mid Y^t) \, \mathrm{d}x(t)$$

$$= \gamma \sum_{K^t: k(t) = q} c_{K^t}$$

$$\times \sum_i \int x(t)$$

$$p(x(t) \mid I_i(t), Y^t) p(I_i(t) \mid Y^t) \, \mathrm{d}x(t) \tag{B.2}$$

$$= \gamma \sum_{K^t:\, k(t)=q} c_{K^t} \sum_i \alpha_i(t) \int x(t)$$

$$p(x(t) \mid I_i(t), Y^t)\, \mathrm{d}x(t), \qquad \text{(B.3)}$$

where $\alpha_i(t) \triangleq p(I_i(t) \mid Y^t)$ is the a posteriori association probability of the ith interpretation assigned to the model k. Since

$$x_i(t \mid t) = E\{x(k) \mid I_i(t), Y^t\}. \qquad \text{(B.4)}$$

Replacing Eq. (36) in Eq. (35)

$$q\hat{x}_{K^t} = \gamma \sum_{K^t: k(t)=q} c_{K^t} \sum_i \alpha_i(t) x_i(t \mid t). \qquad \text{(B.5)}$$

References

[1] A. Baumberg, D. Hogg, Learning deformable models for tracking the human body, in: R. Jain, M. Shah (Eds.), Motion Based Recognition, Kluwer, Dordrecht, 1997, pp. 39–60.

[2] M. Isard, A. Blake, A mixed-state condensation tracker with automatic model-switching, Proceedings of the International Conference on Computer Vision, 1998, pp. 107–112.

[3] J.S. Marques, J.M. Lemos, Shape tracking based on switched dynamical models, Proceedings of the IEEE International Conference on Image Processing, Kobe, 1999, pp. 954–958.

[4] J.S. Marques, J.M. Lemos, Optimal and suboptimal shape tracking based on switched dynamic models, Image Vision Comput. (2001) 539–550.

[5] J. Nascimento, J.S. Marques, Robust shape tracking in the presence of cluttered background, Proceedings of the IEEE International Conference on Image Processing, Vol. 3, Vancouver, 2000, pp. 82–85.

[6] A. Blake, M. Isard, Active Contours, Springer, Berlin, 1998.

[7] J. Tugnait, Detection and estimation for abruptly changing systems, Automatica 18 (1982) 607–615.

[8] C. Chang, M. Athans, State estimation for discrete systems with switching parameters, IEEE Trans. Aerosp. Electronic. Syst. 14 (1978) 418–425.

[9] Y. Bar-Shalom, T. Fortmann, Tracking and Data Association, Academic Press, New York, 1988.

[10] T. Cootes, C. Taylor, D. Cooper, J. Graham, Active shape models—their training and application, Comput. Vision Image Understanding 61 (1) (1995) 38–59.

About the Author—JACINTO NASCIMENTO received the E.E. degree from Instituto Superior de Engenharia de Lisboa and the M.S. degree from Instituto Superior Técnico, Lisbon, in 1995 and 1998, respectively. Presently, he is pursuing the Ph.D. degree at IST where he is affiliated with Institute for Systems and Robotics (ISR). His research interests are in image processing and shape tracking.

About the Author—JORGE S. MARQUES received the E.E., M.S. and Ph.D. degrees from Instituto Superior Técnico (IST), Lisbon, in 1981, 1984, 1990, respectively. He is associate professor at IST and a researcher at Institute for Systems and Robotics (ISR), Lisbon. He has published over 70 papers in international journals and conferences. His research interests are shape modeling, image analysis and pattern recognition.