Learning Switching Dynamic Models for Objects Tracking

Gilles Celeux, Jorge S. Marques, Jacinto Nascimento

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Learning Switching Dynamic Models for Objects Tracking

Gilles Celeux, Jorge S. Marques, Jacinto Nascimento

N° 4863

Juillet 2003

THÈME 4

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4863--FR+ENG

Learning Switching Dynamic Models for Objects Tracking

Gilles Celeux, Jorge S. Marques, Jacinto Nascimento

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet IS2

Rapport de recherche n° 4863 — Juillet 2003 — 27 pages

Abstract: Many recent tracking algorithms rely on model learning methods. A promising approach consists of modelling the object motion with switching autoregressive models. This article is involved with parametric switching dynamical models governed by an hidden Markov chain. The maximum likelihood estimation of the parameters of those models is described. The formulas of the EM algorithm are detailed. Moreover, the problem of choosing a good and parsimonious model with BIC criterion is considered. Emphasis is put on choosing a reasonable number of hidden states. Numerical experiments on both simulated and real data sets highlight the ability of this approach to describe properly object motions with sudden changes. The two appplications on real data concern object and heart tracking.

Key-words: Auto Regressive Model, Hidden Markov Chain, EM Algorithm, BIC Criterion, Image Processing.

Estimation de modèles dynamiques cachés pour le suivi d'objets

Résumé : Beaucoup d'algorithmes récents de suivi d'objets reposent sur des modèles d'apprentissage. Une approche prometteuse consiste à considérer que le mouvement des objets peut être régi par des modèles autorégressifs différents. Dans cet article, nous supposons que le passage d'un modèle autorégressif paramétrique à l'autre se fait par une chaîne de Markov cachée. Nous présentons l'estimation des paramt̀res de ce type de modèles par le maximum de vraisemblance. Pour ce faire, nous utilisons l'algorithme EM dont nous détaillons le formulaire. De plus, nous nous intéressons au problème du choix de modèle, et notamment au problème du choix de l'ordre de la chaîne de Markov cachée par le critère BIC. Des expérimentations sur des données simulées et des données réelles illustrent la capacité de ce type de modèles autorégressif à sauts à réaliser efficacement le suivi d'objets présentant des changements de mouvement abrupts. Les deux applications sur des données réelles concernent le suivi de gestes et le suivi de mouvements du cœur.

Mots-clés : Modèle autorégressif, modèle de Markov caché, algorithme EM, critère BIC, analyse d'images.

1 Introduction

It was recently proposed the use of switched dynamical systems to describe the evolution of object contours in video sequences (Isard and Blake, 1998, Marques and Lemos, 2001, North *et al.*, 2001). This type of models allows to describe complex motion dynamics with abrupt changes by using a set of dynamical systems, each of them being tailored to a specific type of motion or shape configuration. For example, a traffic monitoring system in which cars are being tracked could use different models to track different types of motion and manoeuvres. It is assumed that at each instant of time only one dynamical system is active i.e., the observed data is produced by one of the systems at each instant of time but we do not know which. Switching between two different models can occur at any moment. A discrete Markov process is used to model the switching process.

The estimation of the state vector and active model can be performed by using non parametric methods (e.g., particle filters Isard and Blake 1998) or parametric methods (e.g., based on mixtures of Gaussian distributions, Marques and Lemos, 2001 or Nascimento and Marques 2002). In many applications, the switched models are supposed to be known or are identified from previous learning samples. A notable contribution, in the object tracking domain, to identify and exploit a switching model in the same exercise is the paper of North *et al.* (2001) which proposes a Monte Carlo EM algorithm (see Tanner 1991) to identify a multi-class dynamical model. Their model is analogous to the one we present here. But their so-called CONDENSATION algorithm is painfully slow since it involves a large amount of sampling operations in order to obtain a reliable identification of the model dynamics. They proposed complexity improvements by using importance sampling (North *et al.*) (2001), but by its very nature, Monte Carlo EM remains a time consuming algorithm. They were obliged to consider a Monte Carlo approximation of the EM algorithm because they were facing a difficult problem where the observed data are noisy and mainly where there is no one to one mapping between the observed data and the true signal. In the present paper, we restrict attention to situations where there is such a one to one mapping between the observations and the sequence of mixed states. This is a less general situation than the one considered by North *et al.* (2001), but it is a situation recovering most of the applications. We model the switching dynamical system with an hidden Markov chain, the unobserved Markov chain being the sequence of unobserved labels indicating the active model at each instant. Thus, taking profit of the incomplete data structure, we derive the exact EM algorithm for maximum likelihood estimation of the model parameters without needing Monte Carlo approximations. By the way, we provide a much faster algorithm for an efficient identification of such multi-class systems. Moreover, we give elements to answer the question of choosing the number of models to be switched for a specific application by using penalized likelihood criteria as the BIC criterion (see for instance Burnham and Anderson, 1998).

The paper is organized as follows. In Section 2, we describe the hidden Markov model (HMM) used in this paper. In Section 3, the formulas of the EM algorithm for this particular HMM are detailed. In Section 4, the model selection problem is considered and we derived possible penalized criteria for choosing among several specific models, focusing essentially on the number of hidden states. In Section 5, we present illustrative numerical experiments on both simulated and real data sets and a short discussion section ends the paper.

2 The model

Let $(C_t), t \in [0,T]$, be the contour of an object to be tracked in a video sequence and let $\mathbf{x}_t$ be a vector of parameters which characterizes the object contour at the time instant t (see Blake and Isard,1998, for details). In this paper we shall consider a system identification problem: given a set a sequence $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, we wish to estimate a switched dynamical model which describes the data.

Thus, we consider an observed sequence $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in $\mathbf{R}^d$ and a hidden sequence $z_1, \ldots, z_n$ in $\{1, \ldots, m\}$ such that

$$\mathbf{x}_t = A^{z_t}\mathbf{x}_{t-1} + \mathbf{w}_t \tag{1}$$

where the $\mathbf{w}_t$ are independent random variables following a normal distribution with mean 0 and variance matrix S^{z_t}, and $\mathbf{w}_t$ and $\mathbf{x}_\tau$ are independent variables if $t > \tau$. Moreover the sequence (z_t) is supposed to be an ergodic Markov chain with transition matrix B of dimension m where

$$B(i,j) = P(z_t = j | z_{t-1} = i).$$

Let $\theta = (A_i, S_i, i = 1, \ldots, m; B)$ be the vector-valued parameter of the model.

The parameters are usually unknown and need to be estimated from the data.

3 Learning algorithm

This model is an hidden Markov model where the missing data are the hidden labels $z_1, \ldots, z_n$. However there are some differences with respect to the most common hidden Markov chain models used for instance in speech recognition (see Rabiner 1993) since the observed data are not conditionally independent, given the hidden Markov states. But, as it happens with the ordinary hidden Markov chains it is advantageous to make use of the EM algorithm to estimate the parameter through the maximum likelihood approach (see for instance McLachlan and Krishnam 1997). In this section, we detail the EM algorithm for the switching dynamic model we consider.

An iteration of this algorithm consists of computing the expected value of the complete loglikelihood of θ with respect of the missing data $\mathbf{z} = (z_1, \ldots, z_n)$, knowing the observed data $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and a current value of the parameter (E step) and then deriving the parameter value maximizing this expected loglikelihood (M step).

Assuming a specific initial condition for $\mathbf{x}_1$ and z_1 the complete loglikelihood of θ is

$$L(\theta|\mathbf{x}, \mathbf{z}) = \sum_{t=2}^{n} [\ln p(\mathbf{x}_t | z_t, \mathbf{x}_{t-1}, \theta) + \ln B(z_{t-1}, z_t)] + \ln p(\mathbf{x}_1 | z_1, \theta) + \ln p(z_1)$$

where $p(\mathbf{x}_t|z_t, \mathbf{x}_{t-1}, \theta)$ is the density of a Gaussian distribution with mean $A_{z_t}\mathbf{x}_{t-1}$ and variance matrix S_{z_t}, $p(\mathbf{x}_1|z_1, \theta)$ is the initial distribution of the state vector and $p(z_1)$ is the initial distribution of the hidden state. Thus up to an additive constant, the complete loglikelihood of θ is

$$L(\theta|\mathbf{x}, \mathbf{z}) = \sum_{t=2}^{n} \left[\ln p(\mathbf{x}_t|z_t, \mathbf{x}_{t-1}, \theta) + \ln B(z_{t-1}, z_t)\right]. \tag{2}$$

Since it is difficult to maximize directly the observed likelihood $L(\mathbf{x}, \theta) = \ln P(\mathbf{x}|\theta)$, it is convenient to take profit of the missing data structure of the model with the EM algorithm which consists of an iterated maximization of the expectation of the complete loglikelihod knowing the observed data

$$Q(\theta|\theta') = E[L(\theta|\mathbf{x}, \mathbf{z})|\mathbf{x}, \theta')].$$

In the present context, the essential task of the E step is computing the conditional probabilities of the hidden states $\mathbf{z}$ knowing $\mathbf{x}$ and a current value of the parameter θ. This can be performed using a forward-backward algorithm as described in Muri (1997). Actually, using Markov property, we get

$$P(\mathbf{z}|\mathbf{x}, \theta) = \left[\prod_{t=2}^{n} P(z_t|z_{t-1}, \mathbf{x}, \theta)\right] P(z_1|\mathbf{x}, \theta)$$

and, for all $t = 1, \dots, n$,

$$P(z_t|z_{t-1}, \mathbf{x}, \theta) = \frac{P(z_t, z_{t-1}|\mathbf{x}, \theta)}{P(z_{t-1}|\mathbf{x}, \theta)}.$$

Both terms are computed in the forward-backward algorithm described hereafter in the E step of the EM algorithm. In the following the upper index r will denote the current parameter estimates at the rth iteration of EM. Moreover, we will denote $\mathbf{x}_1^t = (\mathbf{x}_1, \dots, \mathbf{x}_t)$ and $\mathbf{z}_1^t = (z_1, \dots, z_t)$.

E step: This step consists of computing the conditional distribution of the hidden sequence $\mathbf{z}$ given $\mathbf{x}$ and θ^r. It is decomposed into two recursions, the forward and the backward recursions and is leading to the computation of $P(z_{t-1}, z_t|\mathbf{x}, \theta^r)$ and $P(z_t|\mathbf{x}, \theta^r)$ by the formulas (4) and (5) derived below.

Forward recursion. This recursion consists of solving the *prediction* equation and the *filtering* equation to update the distribution of z_t given past and current measurements.

The prediction equation is as follows for all $j = 1, \dots, m$ and $t = 1, \dots, n$

$$\begin{aligned} P(z_t = j|\mathbf{x}_1^{t-1}, \theta^r) &= \sum_{i=1}^{m} P(z_{t-1} = i, z_t = j|\mathbf{x}_1^{t-1}, \theta^r) \\ &= \sum_{i=1}^{m} P(z_t = j|z_{t-1} = i, \mathbf{x}_1^{t-1}, \theta^r) P(z_{t-1} = i|\mathbf{x}_1^{t-1}, \theta^r) \\ &= \sum_{i=1}^{m} B^r(i,j) P(z_{t-1} = i|\mathbf{x}_1^{t-1}, \theta^r). \end{aligned}$$

The filtering equation is as follows for all $j = 1, \dots, m$ and $t = 1, \dots, n$

$$\begin{aligned} P(z_t = j|\mathbf{x}_1^{t}, \theta^r) &= P(z_t = j|\mathbf{x}_t, \mathbf{x}_1^{t-1}, \theta^r) \\ &= \frac{P(z_t = j, \mathbf{x}_t|\mathbf{x}_1^{t-1}, \theta^r)}{P(\mathbf{x}_t|\mathbf{x}_1^{t-1}, \theta^r)} \\ &= \frac{P(z_t = j, \mathbf{x}_t|\mathbf{x}_1^{t-1}, \theta^r)}{\sum_{i=1}^{m} P(z_t = i, \mathbf{x}_t|\mathbf{x}_1^{t-1}, \theta^r)} \\ &= \frac{P(\mathbf{x}_t|z_t = j, \mathbf{x}_1^{t-1}, \theta^r) P(z_t = j|\mathbf{x}_1^{t-1}, \theta^r)}{\sum_{i=1}^{m} P(\mathbf{x}_t|z_t = i, \mathbf{x}_1^{t-1}, \theta^r) P(z_t = i|\mathbf{x}_1^{t-1}, \theta^r)}. \end{aligned}$$

Thus, using Markov property for $(\mathbf{x}_t)$, we get

$$P(z_t = j|\mathbf{x}_1^{t}, \theta^r) = \frac{P(\mathbf{x}_t|z_t = j, \mathbf{x}_{t-1}, \theta^r) P(z_t = j|\mathbf{x}_1^{t-1}, \theta^r)}{\sum_{i=1}^{m} P(\mathbf{x}_t|z_t = i, \mathbf{x}_{t-1}, \theta^r) P(z_t = i|\mathbf{x}_1^{t-1}, \theta^r)}. \tag{3}$$

Backward recursion. This recursion consists of solving the *smoothing* equation to obtain $P(z_{t-1}, z_t|\mathbf{x}, \theta^r)$ and $P(z_t|\mathbf{x}, \theta^r)$.
We have

$$P(z_{t-1} = i, z_t = j|\mathbf{x}, \theta^r) = P(z_{t-1} = i|z_t = j, \mathbf{x}, \theta^r) P(z_t = j|\mathbf{x}, \theta^r)$$

and from the independence of the future observations given z_t

$$P(z_{t-1} = i, z_t = j|\mathbf{x}, \theta^r) = P(z_{t-1} = i|z_t = j, \mathbf{x}_1^{t-1}, \theta^r) P(z_t = j|\mathbf{x}, \theta^r).$$

Thus by Bayes formula

$$P(z_{t-1} = i, z_t = j|\mathbf{x}, \theta^r) = \frac{P(z_t = j|z_{t-1} = i, \mathbf{x}_1^{t-1}, \theta^r) P(z_{t-1} = i|\mathbf{x}_1^{t-1}, \theta^r) P(z_t = j|\mathbf{x}, \theta^r)}{P(z_t = j|\mathbf{x}_1^{t-1}, \theta^r)},$$

and using the independence of z_t and $\mathbf{x}_\tau$, $t > \tau$

$$P(z_{t-1} = i, z_t = j|\mathbf{x}, \theta^r) = \frac{B^r(i,j) P(z_{t-1} = i|\mathbf{x}_1^{t-1}, \theta^r) P(z_t = j|\mathbf{x}, \theta^r)}{P(z_t = j|\mathbf{x}_1^{t-1}, \theta^r)}. \tag{4}$$

And we have

$$P(z_{t-1}=i|\mathbf{x},\theta^r)=\sum_{j=1}^{m}P(z_{t-1}=i,z_t=j|\mathbf{x},\theta^r)$$

which, from (4), can be written

$$P(z_{t-1}=i|\mathbf{x},\theta^r)=\sum_{j=1}^{m}\frac{B^r(i,j)P(z_{t-1}=i|\mathbf{x}_1^{t-1},\theta^r)P(z_t=j|\mathbf{x},\theta^r)}{P(z_t=j|\mathbf{x}_1^{t-1},\theta^r)},$$

that is

$$P(z_{t-1}=i|\mathbf{x},\theta^r)=P(z_{t-1}=i|\mathbf{x}_1^{t-1},\theta^r)\sum_{j=1}^{m}\frac{B^r(i,j)P(z_t=j|\mathbf{x},\theta^r)}{P(z_t=j|\mathbf{x}_1^{t-1},\theta^r)}. \tag{5}$$

Remark: There is no need to compute the joint probability $P(z_{t-1}=i,z_t=j|\mathbf{x},\theta^r)$ in order to compute $P(z_{t-1}=i|\mathbf{x},\theta^r)$. (The backward recursion (5) is self content.)

Finally the E step consists of

1. Starting from $P(z_1=j|\mathbf{x}_1,\theta^r)$ for $j=1,\dots,m$ (computed using the initial conditions of the processes $(\mathbf{x}_t)$ and (z_t)) compute $P(z_t=j|\mathbf{x}_1^{t-1},\theta^r)$ and $P(z_t=j|\mathbf{x}_1^{t},\theta^r)$ for $j=1,\dots,m$ using the forward step.
2. Then it is possible to derive $P(z_n=i|\mathbf{x},\theta^r)$ for $i=1,\dots,n$ using the final result of the forward step. And, using the backward step, we get $P(z_{t-1}=i,z_t=j|\mathbf{x},\theta^r)$ and $P(z_{t-1}=i|\mathbf{x},\theta^r)$ for $i=1,\dots,m$ and $j=1,\dots,m$.

M step: The M step consists of updating the estimate of the unknown parameter θ by maximizing the current expectation of the complete likelihood

$$\theta^{r+1}=\arg\max_{\theta}Q(\theta|\theta^r)$$

where

$$Q(\theta|\theta^r)=E[L\theta|\mathbf{x},\mathbf{z})|\mathbf{x},\theta^r)].$$

We have from (2)

$$\begin{aligned}Q(\theta|\theta^r)&=\sum_{t=2}^{n}\sum_{i=1}^{m}P(z_t=i|\theta^r,\mathbf{x})\ln p(\mathbf{x}_t|z_t=i,\mathbf{x}_{t-1},\theta)\\&+\sum_{t=2}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m}P(z_t=i,z_{t-1}=j|\theta^r,\mathbf{x})\ln B(j,i).\end{aligned}$$

Thus, denoting $\mathbf{A}=(A^i,i=1,\dots,m)$ and $\mathbf{S}=(S^i,i=1,\dots,m)$, $Q(\theta|\theta^r)$ is of the form $Q_1(\mathbf{A},\mathbf{S}|\theta^r)+Q_2(B|\theta^r)$ and we have

$$(\mathbf{A}^{r+1},\mathbf{S}^{r+1})=\arg\max_{\mathbf{A},\mathbf{S}}Q_1(\mathbf{A},\mathbf{S}|\theta^r)$$

and

$$B^{r+1} = \arg\max_{B} Q_2(B|\theta^r).$$

And, thus, we deduce straightforwardly the update values $\mathbf{A}^{r+1}$ and $\mathbf{S}^{r+1}$ from the m.l. equations given in North *et al.* (2001).

$$\begin{gathered}(A^i)^{r+1} R_{11}^{ir} = R_{01}^{ir} \\ (S^i)^{r+1} = R_{00}^{ir} - 2(A^i)^{r+1}(R_{01}^{ir})^T + (A^i)^{r+1}(R_{11}^{ir})((A^i)^{r+1})^T, \qquad (6)\end{gathered}$$

where

$$\begin{aligned} R_{00}^{ir} &= \frac{1}{D}\sum_{t=2}^{n} P(z_t = i|\theta^r, \mathbf{x})\, \mathbf{x}_t \mathbf{x}_t^T \\ R_{11}^{ir} &= \frac{1}{D}\sum_{t=1}^{n-1} P(z_t = i|\theta^r, \mathbf{x})\, \mathbf{x}_t \mathbf{x}_t^T \\ R_{01}^{ir} &= \frac{1}{D}\sum_{t=2}^{n} P(z_t = i|\theta^r, \mathbf{x})\, \mathbf{x}_t \mathbf{x}_{t-1}^T \\ R_{10}^{ir} &= \frac{1}{D}\sum_{t=2}^{n} P(z_t = i|\theta^r, \mathbf{x})\, \mathbf{x}_{t-1} \mathbf{x}_t^T \end{aligned}$$

with

$$D = \sum_{t=2}^{n} P(z_t = i|\theta^r, \mathbf{x}).$$

The equation (6) can be rewritten

$$(S^i)^{r+1} = R_{00}^{ir} - (A^i)^{r+1}(R_{01}^{ir})^T.$$

In the same manner, the update of B is derived straightforwardly in this HMM context (see for instance Muri 1997 or Marques 1999)

$$B^{r+1}(i,j) = \frac{1/(n-1)\sum_{t=2}^{n} P(z_{t-1} = i, z_t = j|\mathbf{x}, \theta^{r+1}}{1/n\sum_{t=1}^{n} P(z_t = i|\mathbf{x}, \theta^{r+1})}$$

for $i = 1, \ldots, m$ and $j = 1, \ldots, m$.

Remark: The number of parameters to be estimated with the present model, where a different variance matrix S^i is assumed for each hidden state i is $\nu = md^2 + \frac{md(d+1)}{2} + m(m-1)$. More parsimonious models are possible by assuming

- $S^i = S$ for $i = 1, \ldots, m$,
- diagonal A^i and $S^i, i = 1, \ldots, m$,

- diagonal A^i and $S^i = \sigma_i^2 I, i = 1, \ldots, m$.

For those alternative models, it is possible to derive the update estimation of S^i in the M step without difficulty.

It is often useful to know the most probable sequence of hidden dynamic models given the observed data $\mathbf{x}$ and the estimate $\hat{\theta}$ obtained with the above EM algorithm. This can be achieved with a Viterbi-like algorithm that we now describe. Following Marques (1999), it takes the following form. Denoting for $j = 1, \ldots, m$ and $t = 1, \ldots, n$

$$\delta_t(j) = \max_{\mathbf{z}_1^t / z_t = j} P(\mathbf{x}_1^t, \mathbf{z}_1^t | \hat{\theta}),$$

$\hat{\theta}$ being the m.l. estimate of θ obtained at the convergence of the EM algorithm, we proceed to the following forward and backward steps.

Forward step: Starting from

$$\delta_1(j) = P(z_1 = j) P(\mathbf{x}_1 | z_1 = j, , \hat{\theta})$$

compute for $j = 1, \ldots, m$ and $t = 1, \ldots, n$

$$\delta_t(j) = P(\mathbf{x}_t | \mathbf{x}_{t-1}, z_t, \theta^r) \max_i \delta_{t-1}(i) B^r(i, j).$$

Backward step: Starting from $z_n = \arg\max_{i=1,\ldots,n} \delta_n(i)$ compute for $t = n-1, \ldots, 1$,

$$z_t = \arg\max_i \delta_t(i) B(i, z_{t+1}).$$

4 Choosing an honest model

An important question when using a multi-class model, is to choose a reliable model, containing enough parameters to ensure a realistic fit to the learning data set, but not too much to avoid overfitting and poor performances for future use of the model. In our context, a particular type of variance matrices S^i, $i = 1, \ldots, m$, was chosen to reduce the number of free parameters. More important is to determine a reliable number m of hidden states. A classical way to deal with this dimension selection problem is to minimize a penalized loglikelihood criteria, of the form "Deviance+Penality", where

$$\text{Deviance} = -2L(\hat{\theta} | \mathbf{x}),$$

$L(\hat{\theta} | \mathbf{x})$ being the observed loglikelihood evaluated at the parameter value maximizing it. The deviance is a measure of the fit of the model to the learning sample $\mathbf{x}$. The Penalty term is an increasing function of the number of free parameters in the model (see for instance Burnham and Anderson 1998). Thus the art of model selection lies on choosing a good penalty function. For instance the BIC criterion, first proposed by Schwarz (1978), and which can be regarded, in the Bayesian framework, as an approximation of the integrated loglikelihood $L(\mathbf{x}) = \int L(\theta|\mathbf{x})\pi(\theta)d\theta$, $\pi(\theta)$ being a non informative prior distribution on the parameter θ (see Kass and Raftery, 1995), is as follows for a model M

$$\mathrm{BIC}(M) = -2L(\hat{\theta}|\mathbf{x}) + \nu_M \ln \nu_M, \tag{7}$$

ν_M denoting the number of free parameters of the model M. It has been proved to be reliable on an empirical ground for mixture models (see for instance Roeder and Wasserman, 1997) and, on a theoretical ground, it has been proved to provide a consistent estimate of the dimension of a hidden Markov model under reasonable assumptions (Gassiat, 2002). Another well-known criterion is the AIC criterion of Akaike (1974) which makes use of the less stringent penalty term $2\nu_M$ and which, in many circumstances, is expected to select too complex models (see Burnham and Anderson 1998). In the present paper, we use the BIC criterion to select the number of hidden states of the models we experimented.

In order to derive penalized loglikelihood criteria as BIC, it is necessary to calculate the observed likelihood value of the parameter estimate $\hat{\theta}$,

$$\ell(\hat{\theta}|\mathbf{x}) = p(\mathbf{x}_1, \ldots, \mathbf{x}_n|\hat{\theta}).$$

It is possible to compute this likelihood directly from the output of the EM algorithm as follows. We have

$$\ell(\hat{\theta}|\mathbf{x}) = p(\mathbf{x}_n|\mathbf{x}_1^{n-1}, \hat{\theta})p(\mathbf{x}_{n-1}|\mathbf{x}_1^{n-2}, \hat{\theta}) \ldots p(\mathbf{x}_2|\mathbf{x}_1, \hat{\theta})p(\mathbf{x}_1|\hat{\theta}).$$

Remarking that for $t = 1, \ldots, n$, we have for any $j = 1, \ldots, m$

$$p(\mathbf{x}_t|\mathbf{x}_1^{t-1}, \hat{\theta}) = \sum_{i=1}^{m} P(\mathbf{x}_t|z_t = i, \mathbf{x}_{t-1}, \hat{\theta})P(z_t = i|\mathbf{x}_1^{t-1}, \hat{\theta}),$$

the observed likelihood can be derived straightforwardly from the filtering equation (3) at the last iteration of EM.

5 Experimental Results

This section presents experimental results with synthetic and real images to test the performance of the proposed algorithm.

5.1 Synthetic Data

Example 1 - Model Learning

The first example is an experiment in which we try to estimate the parameters of a switched dynamic model from a sequence of synthetic images. This example considers an object contour moving according to three different dynamics. The object shape is assumed to be known and only translation motions are considered. A state vector sequence $\mathbf{x}_t$ with 3500 samples was generated according to (1). At each time the state vector contains the displacement coordinates, thus $\mathbf{x}_t \in \Re^2$. This experiment was performed with the following dynamic matrices

$$A^1 = \begin{bmatrix} 0.95 & 0 \\ 0 & 0.95 \end{bmatrix}, \; A^2 = \begin{bmatrix} 0.9 & 0 \\ 0.3 & -0.98 \end{bmatrix}, \; A^3 = \begin{bmatrix} -0.8 & 0.5 \\ 0 & -0.8 \end{bmatrix}. \qquad (8)$$

and variance $S^i = I, i = 1, 2, 3$. In this example the number of models is assumed to be known and the following initializations were used

$$\hat{A}^1 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}, \; \hat{A}^2 = \begin{bmatrix} -0.5 & 0 \\ 0 & 0.8 \end{bmatrix}, \; \hat{A}^3 = \begin{bmatrix} -0.5 & 0 \\ 0 & -0.5 \end{bmatrix}. \qquad (9)$$

$$\hat{S}^1 = \hat{S}^2 = \hat{S}^3 = I \qquad (10)$$

The goal is to estimate the different dynamics as well as the covariance matrices which characterize the trajectory of the object.

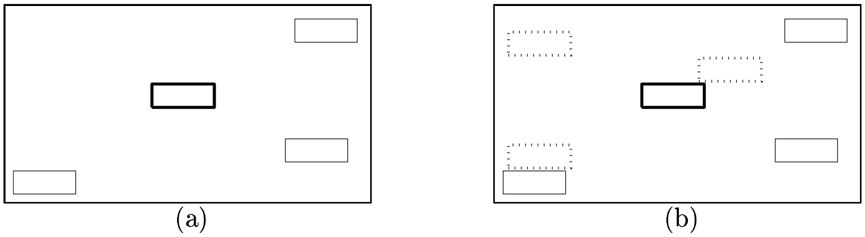


Figure 1: (a) Initial shape (bold line) and three predicted shapes(thin lines); (b) initial and improved predictions of the object (dashed lines).

One way to visualize different dynamical models in the context of shape tracking consists of showing the reference shape at a given position and orientation together with the predicted shapes according to the different dynamical models. The parameters of the predicted shapes are given by $\hat{\mathbf{x}}_t = A^i \mathbf{x}_{t-1}$, where i denotes the index of the active model. Fig. 1(a) shows the reference shape at position $(1, 1)$ (bold solid line) and the predicted shapes obtained from (8) according to the dynamic models (thin solid lines). Fig. 1(b) shows the initializations obtained from (9) (dashed lines). The goal, is thus, to approximate the solid lines (true positions of the object) by the dashed lines (estimated positions of the object).

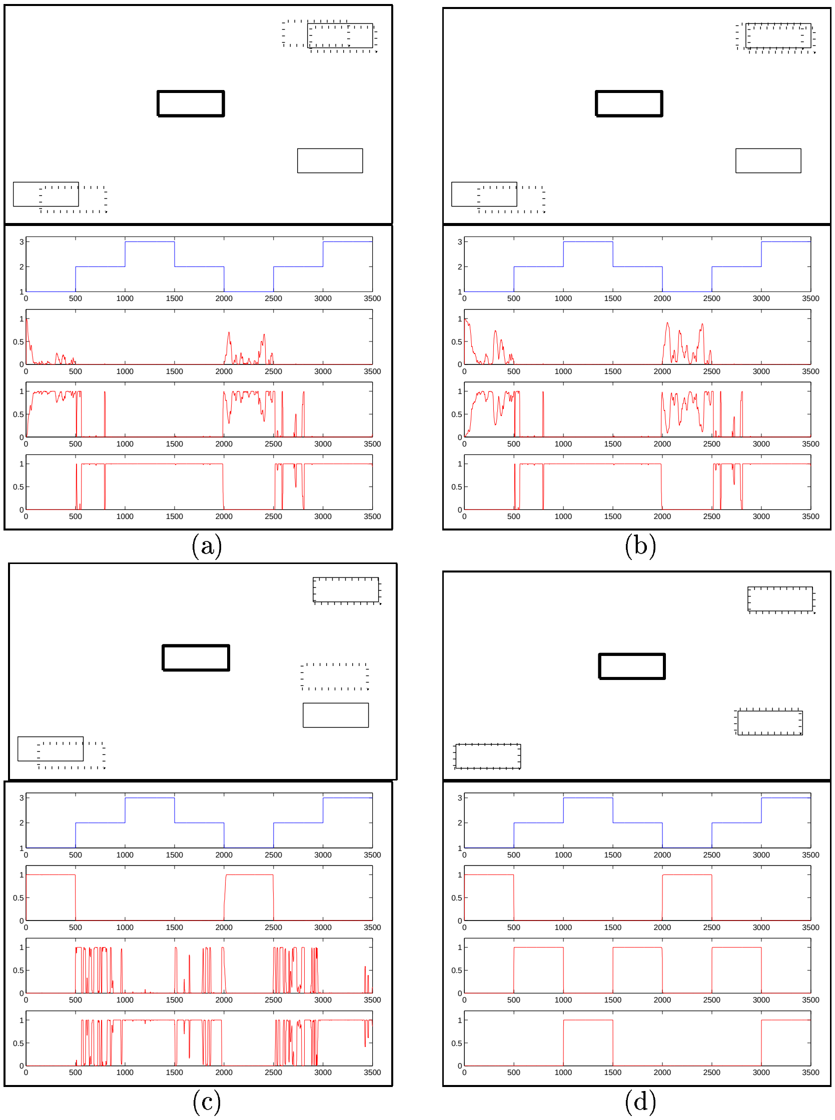
To illustrate the evolution of the model estimates obtained by the EM method, Fig. 2 shows the estimates at iteration 5, 10, 12 and 20, as well as the evolution of the weights. Each figure depicts the correct model label and the weights associated to each model.

It can be concluded that a good estimation of all model weights is achieved in the E step. A data conflict appears at iteration 5 (see fig. 2 (a)) where two prototypes compete to represent the same position of the object. Therefore, the convergence of the weights associated to the competing models is slower.

The estimate of the dynamic and covariance matrices obtained in this experiment are the following:

$$\hat{A}^1 = \begin{bmatrix} 0.9516 & 0.0119 \\ -0.0071 & 0.9445 \end{bmatrix}, \hat{A}^2 = \begin{bmatrix} 0.8869 & 0.0000 \\ 0.3129 & -0.9823 \end{bmatrix}, \hat{A}^3 = \begin{bmatrix} -0.8087 & 0.5007 \\ 0.0066 & -0.8132 \end{bmatrix}, \quad (11)$$

$$\hat{S}^1 = \begin{bmatrix} 0.9751 & -0.0296 \\ -0.0296 & 0.9943 \end{bmatrix}, \hat{S}^2 = \begin{bmatrix} 1.0504 & -0.0124 \\ -0.0124 & 1.0146 \end{bmatrix}, \hat{S}^3 = \begin{bmatrix} 1.0126 & -0.0267 \\ -0.0267 & 1.0406 \end{bmatrix}. \quad (12)$$



(a) (b) (c) (d)

Figure 2: Evolution of the contour estimates (top), the model labels and weights (bottom): (a)iteration 5, (b) iteration 10, (c) iteration 15, (d) iteration (20).

Example 2 - Estimation of the number of models

This example illustrates the estimation of the number of switched dynamical models that should be used to represent a trajectory of an object. We wish to determine the number of models which describes the trajectory best.

For this purpose we consider two different situations with three dynamic models, depicted in Fig. 3. In the first case (Fig. 3(a)) the models are the same as in the example 1. In the second case (Fig. 3(b)) two of the models are very close to each other in such a way that the object predictions are overlapped.

In this example, it is assumed that the number of models is unknown. To achieve an honest guess for the number of models we vary the number of models from one up to four models. Two criteria for choosing the number of models are considered: the BIC criterion and the mean square error criterion given by

$$\hat{e}^i = \frac{1}{N-1} \sum_{t=2}^{n} \|\mathbf{x}_t^i - \hat{\mathbf{x}}_t^i\|^2 \tag{13}$$

where $\hat{\mathbf{x}}_t^i = \mathbf{x}_{t-1} \sum_{i=1}^{m} P(z_t = i \mid \theta, \mathbf{x}) \hat{A}^i$.

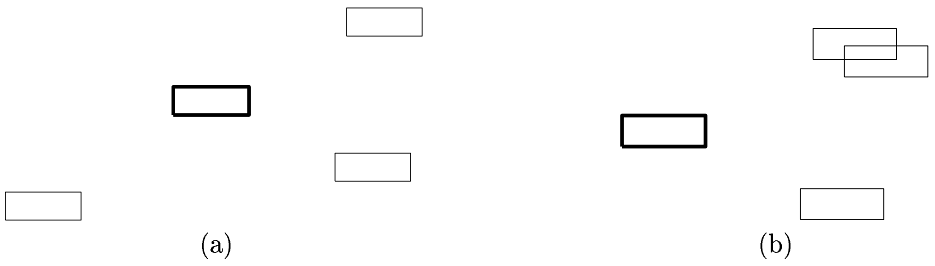


Figure 3: Objects predictor: (a) non overlapping case (b) overlapping case.

Table 1 shows the results of three experiments. In the first case the BIC criterion chooses 3 models in all the experiments. This is the best classification, since the use of one or two models is not enough to represent the data, and choosing four models corresponds to an overfitting.

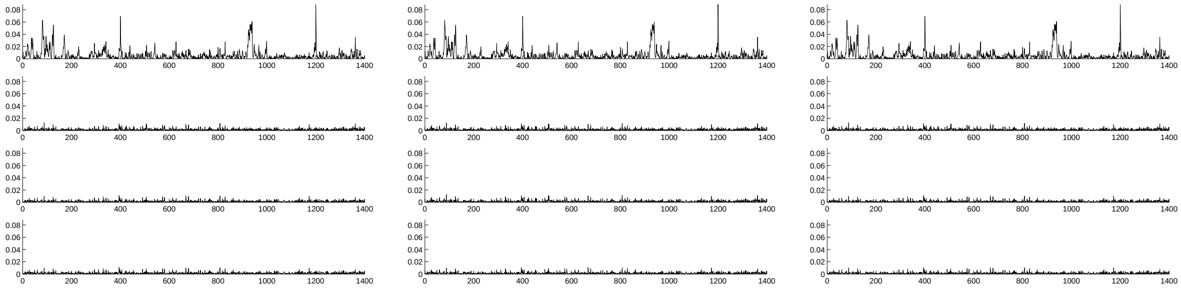
In the second case, the BIC criterion chooses two models. A single model is used to represent the dynamical models with similar parameters.

Fig. 4 shows the results obtained by the mean square criterion. It appears clearly that using a single model leads to a larger error. The number of models cannot be obtained by computing the minimum of the mean square error since it is a decreasing function. But, it can be found by detecting an elbow shape of this function if it exists.

To assess the previous results Monte Carlo experiments were performed. We have generated 100 different data experiments according to (1), each one with 3500 samples for the

	Example 1			Example 2		
Number of models	I_1	I_2	I_3	I_1	I_2	I_3
1 model	1378.4	inf	1379.3	inf	inf	1328.2
2 models	1509.3	1494.4	1429.6	**1271.5**	**1271.5**	**1271.5**
3 models	**1268.6**	**1268.6**	**1268.6**	1286.5	1314.4	1286.5
4 models	1323.6	1323.8	1324.2	1351.4	1341.9	1340.2

Table 1: BIC histograms for the two examples with three different initializations computed from 100 experiments.

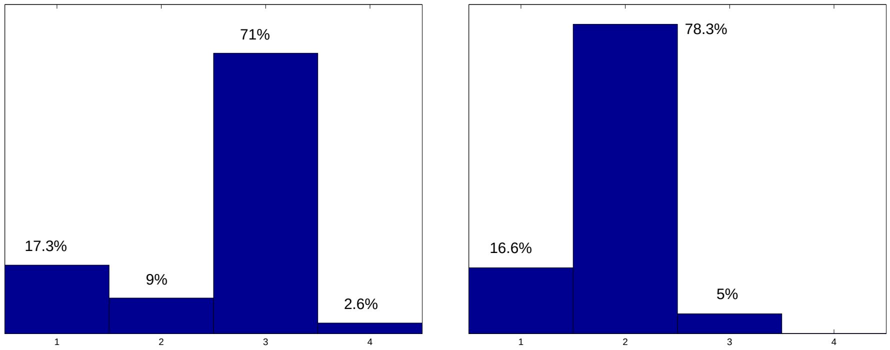


(a) (b) (c)

Figure 4: Computation of the prediction error (second example) for three different initializations. From top to bottom: one to four models.

state vector **x**. Three random initializations were adopted for each experiment and the best number of models according to BIC criterion was chosen. The covariance matrices were initiated as follows: $S^i = kI, i = 1 \dots 4$ and the dynamics A^i were initialized with diagonal matrices with random diagonal values uniformly distributed in (-1,1).

Fig. 5 shows the choice frequencies for $K = 1, 2, 3, 4$ using BIC for the non overlapping case (a) and the overlapping case (b). The BIC criterion provides a reliable choice of the number of models in the non overlapping case, and, as expected, BIC prefers two models most of the time in the overlapping case.



(a) (b)

Figure 5: Frequencies of choosing each number of models with BIC, for the non overlapping case (a), and overlapping case (b).

5.2 Real Data

Here we present several experiments with image sequences of a hand moving and a heart beating. In both cases, the shape or motion has abrupt changes. Thus, multiple models are needed to cope with these changes. As in the previous examples, we estimate first the models involved in the data generation (learning) and we perform an exact estimation of the number of models involved in the sequence by using BIC criterion.

In the following experiments we have used the *Robust Multi Model Tracker* (RMMT) proposed by Nascimento and Marques (2002) which allows to track moving objects with multiple models in the presence of outliers. The method works very well if the dynamic models are known in advance. Here we extend this method to the more general case in which the number of models and the model parameters are unknown. The output of this tracker provides an estimate of the object shape evolution throughout the video sequence. This information is contained in the state vector $\mathbf{x}_t$ which represents the parameters of a time-varying shape at time t. This information will be the input to the learning algorithm proposed herein. From this information it is possible to learn the dynamics as well as to infer the number of models which are useful to describe the object boundary.

5.2.1 Shape Modeling

In the following examples it is assumed that the object shape is a transformed version of a (known) template plus an additional deformation, both of them being described by B-splines.

Let $v(s)$ be a mapping in $I \to \Re^2$ representing the object boundary, where $I \subset \Re$ and let $v^r(s) : I \to \Re^2$ be a reference shape. It is assumed that

$$v(s) = \mathcal{G}(v^r(s)) + v^d(s) \tag{14}$$

where $v(s)$, $\mathcal{G}(v^r(s))$ and $v^d(s)$ represent the observed shape, the transformed version of the reference shape and the shape deformation, respectively. These curves are modeled by

$$v(s) = \sum_{n=1}^{M} \theta_n \phi_n(s) \tag{15}$$

where $\phi_n(s) : I \rightarrow \Re, n = 1, \dots, M$ is a set of B-spline basis functions, $\theta_n \in \Re^2$ is the nth control point of the B-spline.

The curves defined in (14) are sampled at N equally spaced points, leading to following curve discretization

$$v = [v_x(s_1), \dots, v_x(s_N), v_y(s_1), \dots, v_y(s_N)]^T. \tag{16}$$

Equation (14) can be written as

$$v = \mathcal{G}(v^r) + v^d \tag{17}$$

where v, $\mathcal{G}(v^r)$ and v^d, are $2N \times 1$ vectors. In the following examples we model the shape of the object by using two transformations: a translation (2 degrees of freedom) or an Euclidean similarity (4 degrees of freedom). Since the object may deform locally we need to take into account the deformation of the boundary curve. Thus, we can define a state vector containing both kind of information

$$\mathbf{x} = [x_1, \dots, x_D, \theta^d_{x1}, \dots, \theta^d_{xM}, \theta^d_{y1}, \dots, \theta^d_{yM}]^T \tag{18}$$

where $\mathbf{x}$ is a $(D + 2M) \times 1$ vector. The first D coefficients define the global transformation, the remaining coefficients define the shape deformation.

Equation (17) can be written in a matrix notation as

$$v = C\,x + e \tag{19}$$

where C and e are a $2N \times (D + 2M)$ and $2N \times 1$ shape matrices, respectively, depending on the transformation being considered in the tracking experiments.

5.2.2 Gesture tracking

The first example considers a hand oscillating in front of a shirt. In this example the state vector $\mathbf{x}_t$ is 2 dimensional containing the displacement coordinates of the translation vector (see Fig. 6).

The shape matrices are

$$C = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}, \qquad e = [v^r_x(s_1), \dots, v^r_x(s_N), v^r_y(s_1), \dots, v^r_y(s_N)]^T, \tag{20}$$

where $\mathbf{1}, \mathbf{0}$, and e are $2N \times 1$ vectors, e containing the coordinates of the reference shape.

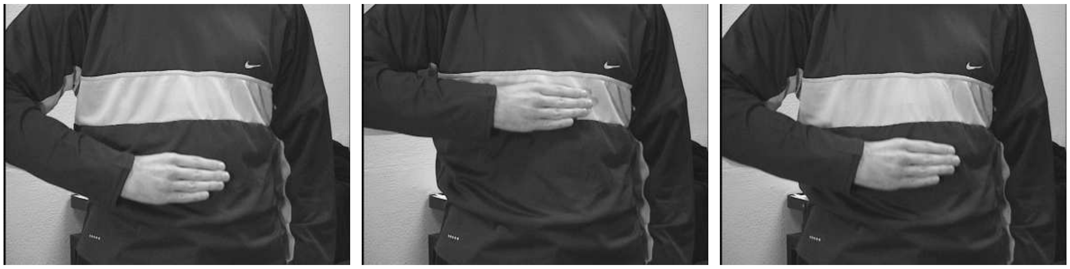


Figure 6: Hand motion (upwards and downwards).

Since the prediction of the state vector is given by

$$\hat{\mathbf{x}}_t = A^{z_t}\hat{\mathbf{x}}_{t-1} \tag{21}$$

the upwards and downwards hand motion can be obtained adopting different dynamics for each model. Two dynamical models were used in this experiment with

$$A^m = \begin{bmatrix} 1 & 0 \\ 0 & \Delta_m \end{bmatrix}, \qquad m = 1, 2 \tag{22}$$

where $\Delta_1 = 0.91$, and $\Delta_2 = 1.10$. Thus, the first model describes the upper motion by allowing a scale in the y-coordinate ($\mathbf{x}_t$ is affected by this factor applying in equation (1)), while the second describes the down motion.

In Fig. 7 it is shown the estimated results provided by the RMMT. In Fig. 7 (a) the first row shows the best model label, the second row shows the translation in x-coordinate (almost constant during the sequence), the third row illustrates the y-coordinate translation. The last figure shows that there is a sudden change in the hand motion (frame 7) corresponding to a switching between the models. Fig. 7(b) shows the evolution of the hand boundary provided by the RMMT tracker.

The estimated matrices are the following

$$\hat{A}^1 = \text{diag}(0,91,1.21)\ \hat{A}^2 = \text{diag}(0,99,0.93) \tag{23}$$

The first diagonal coefficient represents the displacement in x. From Fig. 7(b) we may conclude that there is a small drift during the experiment, meaning that the hand does not have a pure vertical motion. The second diagonal coefficients describe the vertical motion.

To find an honest model we vary the number of models from one up to three models and compute the BIC criterion in each case. In this task, we generated ten different initializations and we selected the best BIC. The diagonal entries of the dynamic matrices were randomly initialized. Fig. 8 depicts the BIC for each case, the dashed line corresponds to the BIC

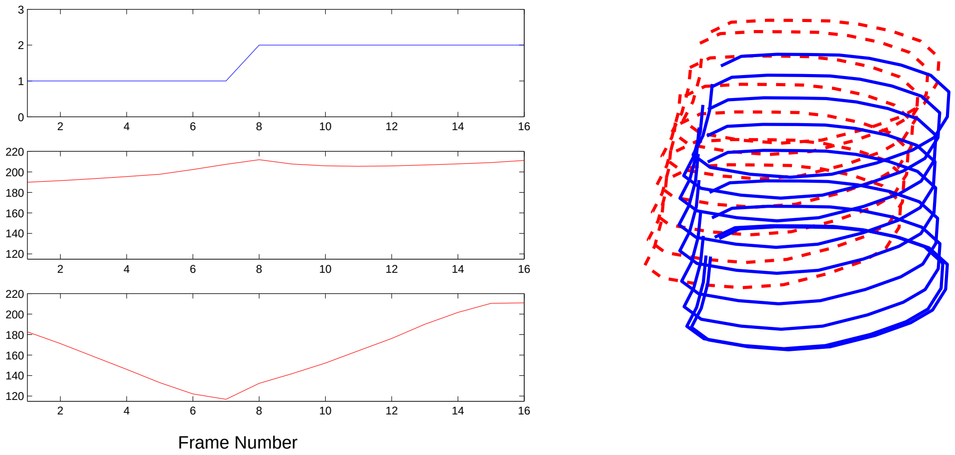


Figure 7: Evolution of the label model and the state vector coefficients during a sequence of 16 frames (a); Evolution of the boundary hand in the video sequence (dashed line: upper motion, solid line: down motion) (b).

computed using three models (upper dashed line) and using one model (down dashed line). The bold line is the BIC values using two models. It is clearly seen that the best BIC is reached with two models. Higher values of this BIC correspond to poor initializations (eighth and ninth values).

5.2.3 Heart tracking

In this example we estimate the motions and the number of models needed to track the left ventricle in a sequence of ultrasound images. Fig. 9 displays two ultrasound images showing the four heart cavities in two different phases. Fig. 9(a) shows an ultrasound image which corresponds to the systole phase. The walls of the left ventricle are contracted. Fig. 9(b) shows an image in the diastole phase, in which the walls of the left ventricle are expanded.

We have considered the group of Euclidean transformations to model the motion of the ventricle. Therefore, we have four degrees of freedom ($D = 4$). To model the deformation, a B-spline sampled at 48 points with $M = 25$ control points was used. The observation model for this case is given by

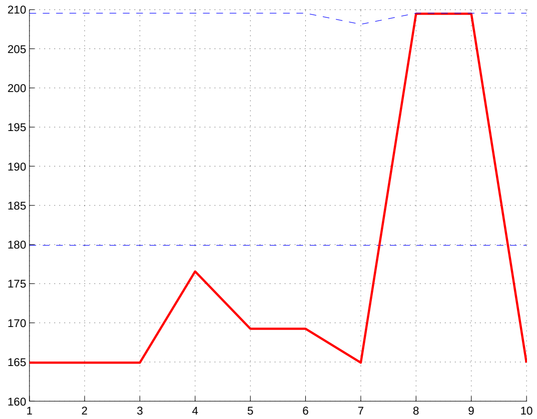
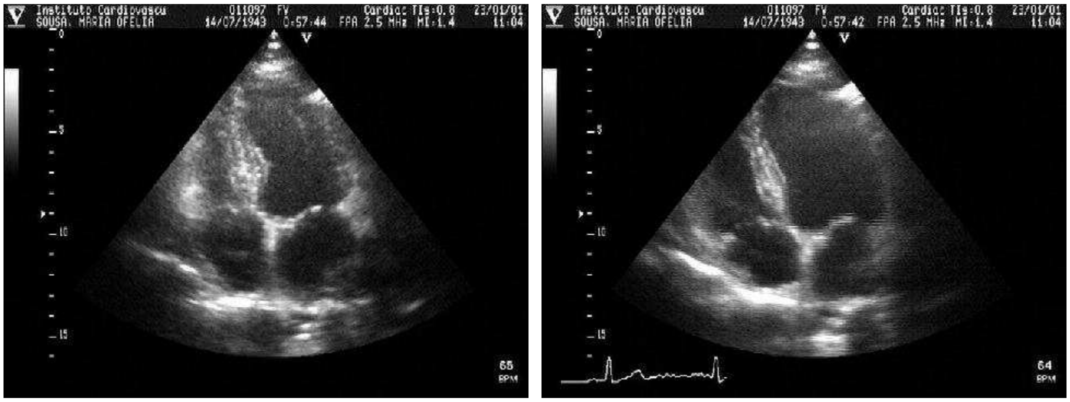


Figure 8: BIC results for 10 initializations in using : one model (down dashed line), two models (bold line), and three models (upper dashed line).



(a) (b)

Figure 9: Ultrasound images showing the typical boundaries on the cardiac cycle: systole phase (a) diastole phase (b).

$$C = \begin{bmatrix} M_x & -M_y & \mathcal{B}_{N\times M} & \mathbf{O}_{N\times M} \\ M_y & M_x & \mathbf{O}_{N\times M} & \mathcal{B}_{N\times M} \end{bmatrix}, \qquad \text{with } M_x = \begin{bmatrix} v_x^r(s_1) & 1 \\ \vdots & \vdots \\ v_x^r(s_N) & 1 \end{bmatrix}, M_y = \begin{bmatrix} v_y^r(s_1) & 0 \\ \vdots & \vdots \\ v_y^r(s_N) & 0 \end{bmatrix}, \tag{24}$$

where C is a $2N\times(D+2M)$ matrix, $\mathcal{B}$ is a $N\times M$ interpolation B-spline matrix, $\mathbf{O}$ is a null matrix with the appropriate dimensions, M_x, M_y are $N\times 2$ matrices with the coordinates of the reference shape.

The motion dynamics used to represent the two different phases are the following

$$A_m = \mathrm{diag}(l_m, \mathbf{1}), \qquad m = 1, 2, \tag{25}$$

with size $(D+2M)\times(D+2M)$, $\mathbf{1}$ is a $2M\times 1$ vector, $l_m = [\Delta_m \;\; 1 \;\; 1 \;\; 1]$ is a $D\times 1$ vector with $\Delta_1 = 0.9$, $\Delta_2 = 1.1$. In this case the different matrices provide a different scale on the boundary of the ventricle. The first one produces a contraction ($\Delta_1 = 0.9$) while the second one produces an expansion ($\Delta_1 = 1.1$).

Applying the RMMT to this example it is possible to collect the motion in the state vector $\mathbf{x}_t$ (results of this tracker are presented in Nascimento *et al.* (2003). In this case the dimension of $\mathbf{x}_t$ is 54×1 (containing the global and local parameters). In the learning process of multiple motion classes, we do not take in consideration the local parameters (deformation on the curve) of $\mathbf{x}_t$. This can be illustrated in Fig. 10. This figure shows the estimated evolution of the left ventricle contour through a sequence of images in two cases: without deformation in $\mathbf{x}_t$ (a) and with deformation (b). It can be seen that the clusters of the contours remain similar, which is the information needed to learn the dynamics as well as to infer the number of motion models. In conclusion, the local information in $\mathbf{x}_t$ can be neglected if we wish to learn the dynamic model of the motion parameters. In this way we can reduce the complexity of the problem.

Fig. 11 shows the evolution of the weights during the iteration process. The estimation of the model weights is achieved by using a transition matrix for the Markov chain that keep the switching rate slow. In fact, the EM learning method has more difficulties to learn when the switching rate increases. The first row of the Fig. 11 shows the label of the active model. The low level corresponds to the systole phase, whereas the high level corresponds to the diastole phase. The second and third rows of this figure show the evolution of the models weights at different iteration of the EM algorithm.

Initial matrices A are diagonal matrices with random coefficients in the interval $(-1, 1)$ and the estimated values are the following

$$\hat{A}^1 = \mathrm{diag}(0.98\;0.99\;0.82\;0.75), \;\; \hat{A}^2 = \mathrm{diag}(1.02\;1.00\;0.72\;0.61) \tag{26}$$

The first value in each dynamic matrix is the scaling factor which accounts for the contractions (matrix $\hat{A}^1$), and expansions (matrix $\hat{A}^2$) of the ventricle walls.

This example can however be simplified, reducing the computational complexity. Fig. 12 depicts the evolution of the four coefficients contained in the state vector during three cardiac cycles. The second and fourth rows (first and third coefficient of $\mathbf{x}_t$) describe the evolution of scaling and rotation coefficients. The third and fifth rows (second and fourth coefficient of $\mathbf{x}_t$) show horizontal and vertical translations respectively. Given the label model illustrated

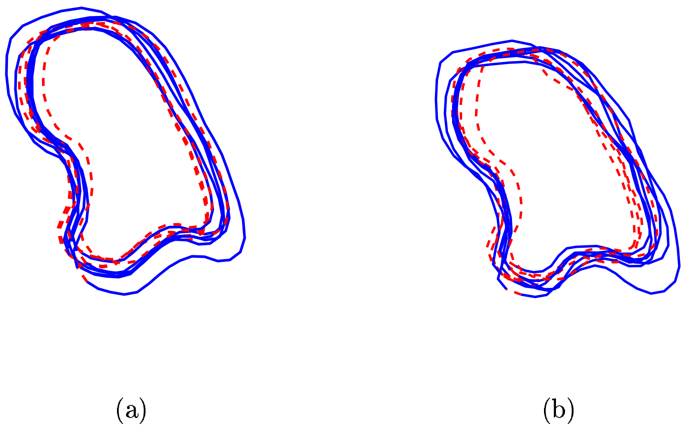


Figure 10: Estimated evolution of the left ventricle contour in two situations: with deformation (a) and without deformation (b) in the state vector. Contours provided by the contraction model are in dashed line and expansion model in solid line.

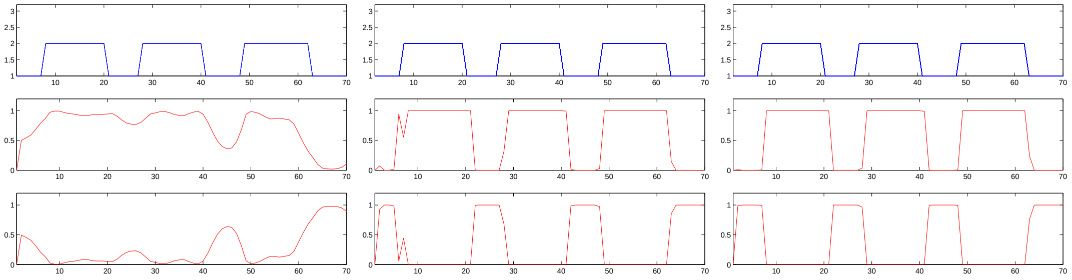


Figure 11: Evolution of the model weights in the E step. From left to right: iteration 3, 16 and 30.

in the first row, we can conclude that the scaling contains valuable information about the heart motion cycle.

Under this reasoning we repeated the experience in the EM framework. The initializations were random in the interval $(-1, 1)$, and the obtained estimates were $\hat{A}^1 = 0.97$ and $\hat{A}^1 = 1.01$ which, corresponds to the desired characterizations of the motion classes. Fig. 13 shows the evolution of the weights during the iteration process.

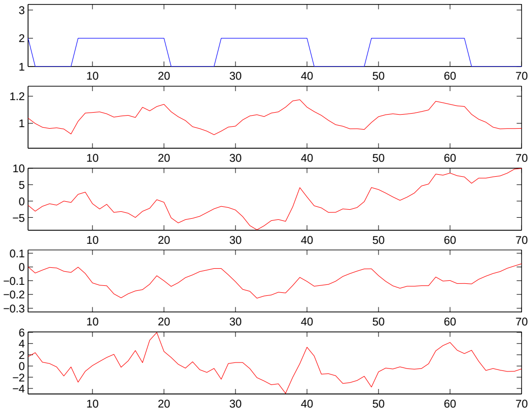


Figure 12: Evolution of the components of the state vector during three cardiac cycles. First row: label model; second row: scale and rotation; third row: horizontal translation; fourth row: rotation; fifth row: vertical translation.

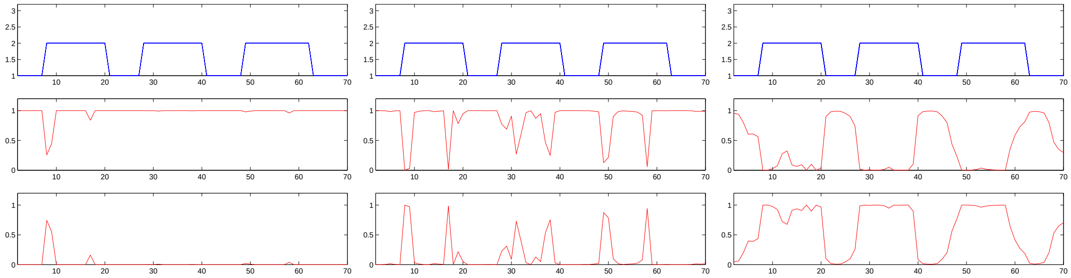
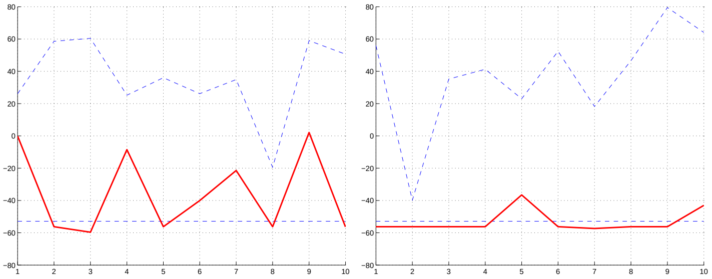


Figure 13: Evolution of the model weights in the simple case. From left to right: iteration 10, 25 and 40.

To choose a reasonable number of switching models we use the same procedure as in the previous example. Fig. 14(a) depicts the BIC values using 10 different initializations. Here the dashed lines corresponds to the BIC computed using just one model (dashed line at the bottom) and using three models (upper dashed line). In Fig. 14(a) we see that the best BIC is obtained by using two models (five times), this is already expected since the contour of the left ventricle suffers from abrupt shape changes leading to two classes of shapes (see Fig. 12). These results were obtained with random diagonal matrices. If we improve the initialization using prior information, namely the fact that the only random entry is the one related to the scale coefficient, it appears that the superirity of the two models solution is more pronounced with BIC criterion, as illustrated in Fig. 14(b).



(a) (b)

Figure 14: BIC values from two different types of initialization. The four coefficients are random variables (a); the scale coefficient is random (b).

6 Conclusions

In this article a new algorithm to learn complex motions with abrupt shape or motion changes is presented. The technique proposed herein is based on a hidden Markov model for object tracking and maximum likelihood estimation of the model parameters is derived.

The motion trajectory is represented by a sequence of dynamic models where discontinuous changes are allowed. Thus, the object undergoes motion or shape deformation during the observation period. The data is provided by the so called Robust Multi Model Tracker (RMMT) (Nascimento and Marques 2002), which was already proposed in context of multiple models. This tracker deals with sudden changes in shape or motion by describing the data as the output of a bank of non linear filters equipped with a switching mechanism. The tracking data is then used to estimate a set of switched dynamic models using the EM algorithm. It was experimentally observed, in synthetic and real images, that the proposed algorithm efficiently estimates the dynamical models involved in the object motion with complexity reduction. The discussion was extended to determine the number of models that describe best the trajectory. This is incorporated in the same framework, where the BIC is directly obtained from the E-step. It was concluded that a satisfactry estimation of the number of switching models can be achieved as well.

One of the most important future directions of this work is the improvement of the initial conditions. Sometimes the initial position may appear to lead to a poor solution, and there is the need to find a good initialisation strategy for the EM algorithm. This point is important since the highest maximized likelihood enters the definition of many criteria, including BIC, aiming to select a good switching model and especially to choose a relevant number of hidden states. Automatic strategies as those described in Biernacki et al. 2003 or Berchtold (2003). **Acknowledgements**

The ultrasound images were kindly provided by Prof. João Sanches of IST and Prof. Fausto Pinto of Instituto Cardio Vascular de Lisboa.

This work was partially supported by FCT and GRICES.

References

[1] Akaike, H. (1974), A new look at the statististical identification model, *IEEE Trans. Auto. Control*, **19**, 716-723.

[2] Berchtold, A. (2003), Optimisation of Mixture Models: Comparison of Different Strategies. Forthcoming in *Computational Statistics.*

[3] Biernacki, C., Celeux, G. and Govaert, G. (2003), Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics and Data Analysis*, **41**, 561-575.

[4] Blake, A. and Isard, M. (1998) *Active Contours*, Springer-Verlag. 1998.

[5] Burnham K. P. and Anderson D. R. (1998), *Model Selection and Inference: a Practical Information-Theoretic Approach.* Springer-Verlag.

[6] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), Maximum Likelihood for Incomplete Data via the EM Algorithm (with discussion), *Journal of the Royal Statistical Society, B*, **39**, 1-38.

[7] Gassiat, E. (2002), Likelihood ratio inequalities with applications to various mixtures, *Ann. Inst. Henri Poincaré*, **38**, 897-906.

[8] Isard, M. and Blake, A. (1998), A mixed-state Condensation tracker with automatic model-switching. *ICCV*, pp. 107-112.

[9] Kass, R. and Raftery, A. (1995), Bayes factor and model uncertainty, *Journal of the American Statistical Association*, **90**, 773-795.

[10] McLachlan, G. J. and Krishnam, T. (1997), *The EM algorithm and Extensions*, New York: Wiley.

[11] McLachlan, G. J. and Peel, D. (2000), *Finite Mixture Models*, New York: Wiley.

[12] Marques, J. S. (1999), *Reconhecimento de Padrões: Métodos Estatísticos e Neuronais.*, Lisbon: IST Press.

[13] Marques, J. S. and Lemos, J. M. (2001), Optimal and Suboptimal Shape Tracking Based on Multiple Switched Dynamic Models. *Image and Vision Computing*, 539-550, June.

[14] Muri, F. (1997), Comparaison d'algorithmes d'identification de chaînes de Markov cachées et application à la détection de régions homogènes dans les séquences ADN. PhD Thesis université René Descartes, Paris 5.

[15] Nascimento, J. and Marques, J. S. (2002), Robust Multi-Model Filter for Shape Tracking in the Presence of Outliers. *Pattern Recognition*, vol. 35, pp. 2711-2718, December 2002.

[16] J. Nascimento, J. S. Marques, J. Sanches. (2003). Estimation of Cardiac Phases in Echographic Images Using Multiple Models. In *IEEE Int. Conf. on Image Processing*, Barcelona, September 2003.

[17] North, B., Blake, A., Isard, M. and Rittscher J. (1999), Learning and classification of complex dynamics. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, 1016–1034.

[18] Rabiner, L. (1989), A tutorial on Hidden Markov Models and Selected Application in Speeche Recognition, *Proceedings of the IEEE*, **77**, 257-285.

[19] Roeder, K. and Wasserman, L. (1997), Practical Bayesian Density Estimation Using Mixtures of Normals, *Journal of the American Statistical Association*, **92**, 894-902.

[20] Schwarz, G (1978), Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.

[21] Tanner, M. A. (1991), *Tools for Statistical Inference, Observed Data and Data Augmentation Methods.* Lectures Notesin Statistics, Springer-Verlag.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l’Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399