FROM TURBO HIDDEN MARKOV MODELS TO TURBO STATE-SPACE MODELS

Florent Perronnin and Jean-Luc Dugelay

Institut Eurécom
Multimedia Communications Department
BP 193, 06904 Sophia Antipolis Cedex, France
{florent.perronnin, jean-luc.dugelay}@eurecom.fr

ABSTRACT

We recently introduced a novel approximation of the intractable two-dimensional hidden Markov model (2-D HMM), the turbo-HMM (T-HMM), which consists of a set of interconnected horizontal and vertical 1-D HMMs. In this paper, we consider the extension of this framework to the continuous state HMM, generally referred to as the state-space model (SSM). We provide efficient approximate answers to the three following problems: 1) how to compute the likelihood of a set of observations, 2) how to find the sequence of states that best "explains" a set of observations and 3) how to estimate the model parameters given a set of observations. The application of this work to the challenging problem of face recognition in the presence of large illumination variations will illustrate the potential of our approach.

1. INTRODUCTION

While the hidden Markov model (HMM) has been extensively applied to 1-D problems [1], the complexity of its extension to 2-D grows exponentially with the data size and is hence intractable in most applications of practical value. We recently introduced a novel approximation of the 2-D HMM [2], the turbo-HMM (T-HMM), which consists of a set of interconnected horizontal and vertical 1-D HMMs that "communicate" through an iterative process by inducing prior probabilities on each other.

Although in most applications the state variable is discrete, some problems are best described with a continuous state HMM, generally referred to as the state-space model (SSM) [3]. One of the most famous examples of such a problem is the estimation of the instantaneous position and speed of an object in the space. Another problem, that is investigated in this paper, is the separation of an image into its reflectance and luminance. However, the growth in complexity that plagues the 2-D HMM also arises in the case of the 2-D SSM and approximations are necessary.

The goal of this paper is to extend the T-HMM framework to the continuous state HMM. We introduce the turbo-SSM (T-SSM) as an approximation of the 2-D SSM and provide efficient approximate answers to the three following problems [1]. If O denotes a set of observations and λ the model parameters,

- How to compute $P(O|\lambda)$?
- How to find the state sequence Q that best "explains" O?
- How to adjust λ to maximize $P(O|\lambda)$?

This work was supported in part by France Telecom Research and Development.

The remainder of this paper is organized as follows. In the next section, we provide a brief review of the approximations underlying the T-HMM framework and consider its extension to the T-SSM. In the three following sections, we provide answers to the three previously listed problems. Finally, in section 6, we apply the T-SSM framework to the challenging problem of face recognition in the presence of illumination variation and present experimental results.

2. FROM THE T-HMM TO THE T-SSM

2.1. Approximating a 2-D HMM with a T-HMM

We first introduce a set of notations that will be used throughout this paper. Let $O = \{o_{i,j}, i = 1, \ldots, I, j = 1, \ldots, J\}$ be the set of all observations. We also introduce $o_i^{\mathcal{H}}$ and $o_j^{\mathcal{V}}$ for the i-th row and j-th column of observations, respectively. Similarly, $Q = \{q_{i,j}, i = 1, \ldots, I, j = 1, \ldots, J\}$ denotes the set of all states, while $q_i^{\mathcal{H}}$ and $q_j^{\mathcal{V}}$ denote the i-th row and j-th column of states. Finally, let λ be the set of all model parameters, and let $\lambda_i^{\mathcal{H}}$ and $\lambda_j^{\mathcal{V}}$ be the respective rows and columns of parameters.

The joint likelihood of O and Q given λ can be expressed as:

$$\begin{aligned} P(O,Q|\lambda) &= P(O|Q,\lambda)P(Q|\lambda) \\ &= \textstyle\prod_{i,j} P(o_{i,j}|q_{i,j},\lambda)P(q_{i,j}|q_{i,j-1},q_{i-1,j},\lambda) \end{aligned}$$

Note that the transition probability $P(q_{i,j}|q_{i,j-1}, q_{i-1,j}, \lambda)$ reduces to $P(q_{1,j}|q_{1,j-1}, \lambda)$ if $i = 1$, to $P(q_{i,1}|q_{i-1,1}, \lambda)$ if $j = 1$ and to the initial occupancy probability $P(q_{1,1}|\lambda)$ if $i = j = 1$.

The T-HMM relies on two major approximations. We first assume that $P(q_{i,j}|q_{i,j-1}, q_{i-1,j}, \lambda)$ is separable, i.e.:

$$P(q_{i,j}|q_{i,j-1},q_{i-1,j},\lambda) \propto P(q_{i,j}|q_{i,j-1},\lambda_i^{\mathcal{H}})P(q_{i,j}|q_{i-1,j},\lambda_j^{\mathcal{V}})$$

We then modify the condition in $P(q_{i,j}|q_{i,j-1}, \lambda_i^{\mathcal{H}})$:

$$P(q_{i,j}|q_{i,j-1},\lambda_i^{\mathcal{H}}) \approx P(q_{i,j}|o_i^{\mathcal{H}},\lambda_i^{\mathcal{H}})$$

This leads to the following approximation of the joint likelihood:

$$P(O,Q|\lambda) \approx \prod_j \left[P(o_j^{\mathcal{V}}, q_j^{\mathcal{V}}|\lambda_j^{\mathcal{V}}) \prod_i P(q_{i,j}|o_i^{\mathcal{H}},\lambda_i^{\mathcal{H}}) \right] \qquad (1)$$

that we will denote $P^{\mathcal{V}}(O,Q|\lambda)$. Each term $P(o_j^{\mathcal{V}}, q_j^{\mathcal{V}}|\lambda_j^{\mathcal{V}})$ corresponds to a 1-D vertical HMM. Note that $\prod_i P(q_{i,j}|o_i^{\mathcal{H}}, \lambda_i^{\mathcal{H}})$ is in effect a horizontal prior for column j. A symmetric horizontal quantity $P^{\mathcal{H}}(O,Q|\lambda)$ can be derived. For extensive details on the two previous approximations the reader can refer to [2].

2.2. The T-SSM model

We now assume that the approximations that lead to formula (1) are valid in the continuous case. To keep the mathematical analysis tractable, we choose the emission probabilities, horizontal and vertical transition probabilities and initial occupancy probabilities to be Gaussians. In this paper, we will only present results for the case where the observations and states are uni-dimensional. Using the state-space formalism, the emission probability (or *measurement model*) can be expressed by the following equation:

$$o_{i,j} = f_{i,j} q_{i,j} + u_{i,j} \, , i = 1, \ldots, I, j = 1 \ldots J$$

where $u_{i,j} \sim \mathcal{N}(0, {\sigma_{i,j}}^2)$ is the *measurement noise*. The horizontal and vertical transition probabilities (or *process models*) can be written as:

$$\begin{aligned} q_{i,j} &= g^{\mathcal{H}}_{i,j} q_{i,j-1} + v^{\mathcal{H}}_{i,j} \, , i = 1, \ldots, I, j = 2 \ldots J \\ q_{i,j} &= g^{\mathcal{V}}_{i,j} q_{i-1,j} + v^{\mathcal{V}}_{i,j} \, , i = 2, \ldots, I, j = 1 \ldots J \end{aligned}$$

where $v^{\mathcal{H}}_{i,j} \sim \mathcal{N}(0, {s^{\mathcal{H}}_{i,j}}^2)$ and $v^{\mathcal{V}}_{i,j} \sim \mathcal{N}(0, {s^{\mathcal{V}}_{i,j}}^2)$ are respectively the horizontal and vertical *process noises*. Finally, we introduce the horizontal and vertical initial occupancy probabilities:

$$\begin{aligned} q_{i,1} &= \mu^{\mathcal{H}}_i + v^{\mathcal{H}}_{i,1} \, , i = 1, \ldots, I \\ q_{1,j} &= \mu^{\mathcal{V}}_j + v^{\mathcal{V}}_{1,j} \, , j = 1, \ldots, J \end{aligned}$$

To sum-up, the parameters of our system are $f_{i,j}$, ${\sigma_{i,j}}^2$, $g^{\mathcal{H}}_{i,j}$, ${s^{\mathcal{H}}_{i,j}}^2$, $g^{\mathcal{V}}_{i,j}$, ${s^{\mathcal{V}}_{i,j}}^2$, $\mu^{\mathcal{H}}_i$ and $\mu^{\mathcal{V}}_j$.

3. LIKELIHOOD OF A SET OF OBSERVATIONS

The goal of this section is to estimate $P(O|\lambda)$. Integrating formula (1) over all states Q, we get:

$$\begin{aligned} P(O|\lambda) &= \int_Q P(O, Q|\lambda) dQ \\ &\approx \int_Q \prod_j \left[P(o^{\mathcal{V}}_j, q^{\mathcal{V}}_j | \lambda^{\mathcal{V}}_j) \prod_i P(q_{i,j} | o^{\mathcal{H}}_i, \lambda^{\mathcal{H}}_i) \right] dQ \\ &\approx \prod_j \int_{q^{\mathcal{V}}_j} \left[P(o^{\mathcal{V}}_j, q^{\mathcal{V}}_j | \lambda^{\mathcal{V}}_j) \prod_i P(q_{i,j} | o^{\mathcal{H}}_i, \lambda^{\mathcal{H}}_i) dq^{\mathcal{V}}_j \right] \end{aligned}$$

We note $P^{\mathcal{V}}_j = \int_{q^{\mathcal{V}}_j} P(o^{\mathcal{V}}_j, q^{\mathcal{V}}_j | \lambda^{\mathcal{V}}_j) \prod_i P(q_{i,j} | o^{\mathcal{H}}_i, \lambda^{\mathcal{H}}_i) dq^{\mathcal{V}}_j$. $P^{\mathcal{V}}_j$'s can be computed with a modified version of the forward-backward algorithm as in [2]. However, while in [2] the forward-backward equations relate the forward and backward probabilities, in the continuous case these quantities are probability density functions (pdf's) and we are interested in the equations that relate the *parameters* of these pdf's.

We introduce the following vertical forward, backward and occupancy probabilities:

$$\begin{aligned} \alpha^{\mathcal{V}}_{i,j}(q_{i,j}) &= P(o_{1,j}, ... o_{i,j}, q_{i,j} | \lambda^{\mathcal{V}}_j) \\ \beta^{\mathcal{V}}_{i,j}(q_{i,j}) &= P(o_{i+1,j}, ... o_{I,j} | q_{i,j}, \lambda^{\mathcal{V}}_j) \\ \gamma^{\mathcal{V}}_{i,j}(q_{i,j}) &= P(q_{i,j} | o^{\mathcal{V}}_j, \lambda^{\mathcal{V}}_j) \end{aligned}$$

Defining the corresponding horizontal quantities is straightforward. As the emission, transition and initial occupancy probabilities are Gaussians, if we also initialize the occupancy probabilities γ's in a Gaussian manner, one can show that $\alpha^{\mathcal{V}}_{i,j}$ and $\beta^{\mathcal{V}}_{i,j}$ have the following form:

$$\begin{aligned} \alpha^{\mathcal{V}}_{i,j}(q_{i,j}) &= \frac{c^{\alpha\mathcal{V}}_{i,j}}{\sigma^{\alpha\mathcal{V}}_{i,j} (2\pi)^{\frac{1}{2}}} \exp\left\{ -\frac{(q_{i,j} - \mu^{\alpha\mathcal{V}}_{i,j})^2}{2{\sigma^{\alpha\mathcal{V}}_{i,j}}^2} \right\} \\ \beta^{\mathcal{V}}_{i,j}(q_{i,j}) &= \frac{c^{\beta\mathcal{V}}_{i,j}}{\sigma^{\beta\mathcal{V}}_{i,j} (2\pi)^{\frac{1}{2}}} \exp\left\{ -\frac{(q_{i,j} - \mu^{\beta\mathcal{V}}_{i,j})^2}{2{\sigma^{\beta\mathcal{V}}_{i,j}}^2} \right\} \end{aligned}$$

and that $\gamma^{\mathcal{V}}_{i,j}$ is a Gaussian with mean $\mu^{\gamma\mathcal{V}}_{i,j}$ and variance ${\sigma^{\gamma\mathcal{V}}_{i,j}}^2$. Introducing the notations $\mu^{b\mathcal{H}}_{i,j}$, ${\sigma^{b\mathcal{H}}_{i,j}}^2$ and $c^{b\mathcal{H}}_{i,j}$:

$$\mu^{b\mathcal{H}}_{i,j} = \frac{f_{i,j} o_{i,j} {\sigma^{\gamma\mathcal{H}}_{i,j}}^2 + \mu^{\gamma\mathcal{H}}_{i,j} \sigma^2_{i,j}}{{f_{i,j}}^2 {\sigma^{\gamma\mathcal{H}}_{i,j}}^2 + \sigma^2_{i,j}} \quad {\sigma^{b\mathcal{H}}_{i,j}}^2 = \frac{{\sigma^{\gamma\mathcal{H}}_{i,j}}^2 \sigma^2_{i,j}}{{f_{i,j}}^2 {\sigma^{\gamma\mathcal{H}}_{i,j}}^2 + \sigma^2_{i,j}}$$

$$c^{b\mathcal{H}}_{i,j} = \frac{\exp\left\{ -\frac{1}{2} \frac{(o_{i,j} - f_{i,j} \mu^{\gamma\mathcal{H}}_{i,j})^2}{({\sigma_{i,j}}^2 + {f_{i,j}}^2 {\sigma^{\gamma\mathcal{H}}_{i,j}}^2)} \right\}}{(2\pi)^{\frac{1}{2}} ({\sigma_{i,j}}^2 + {f_{i,j}}^2 {\sigma^{\gamma\mathcal{H}}_{i,j}}^2)^{\frac{1}{2}}}$$

we can estimate $\mu^{\alpha\mathcal{V}}_{i,j}$, $\mu^{\beta\mathcal{V}}_{i,j}$, $\mu^{\gamma\mathcal{V}}_{i,j}$, ${\sigma^{\alpha\mathcal{V}}_{i,j}}^2$, ${\sigma^{\beta\mathcal{V}}_{i,j}}^2$, ${\sigma^{\gamma\mathcal{V}}_{i,j}}^2$ and $c^{\alpha\mathcal{V}}_{i,j}$.

- *Forward α variable*:
  - Initialization:

$$\mu^{\alpha\mathcal{V}}_{1,j} = \frac{\mu^{b\mathcal{H}}_{1,j} {s^{\mathcal{V}}_{1,j}}^2 + \mu^{\mathcal{H}}_1 {\sigma^{b\mathcal{H}}_{1,j}}^2}{{s^{\mathcal{V}}_{1,j}}^2 + {\sigma^{b\mathcal{H}}_{1,j}}^2} \quad {\sigma^{\alpha\mathcal{V}}_{1,j}}^2 = \frac{{s^{\mathcal{V}}_{1,j}}^2 {\sigma^{b\mathcal{H}}_{1,j}}^2}{{s^{\mathcal{V}}_{1,j}}^2 + {\sigma^{b\mathcal{H}}_{1,j}}^2}$$

$$c^{\alpha\mathcal{V}}_{1,j} = \frac{c^{b\mathcal{H}}_{i,j} \exp\left\{ -\frac{1}{2} \frac{(\mu^{b\mathcal{H}}_{1,j} - \mu^{\mathcal{V}}_j)^2}{{\sigma^{\mathcal{V}}_j}^2 + {\sigma^{b\mathcal{H}}_{1,j}}^2} \right\}}{(2\pi)^{\frac{1}{2}} ({s^{\mathcal{V}}_{1,j}}^2 + {\sigma^{b\mathcal{H}}_{1,j}}^2)^{\frac{1}{2}}}$$

  - Recursion:

$$\begin{aligned} \mu^{\alpha\mathcal{V}}_{i+1,j} &= \frac{g^{\mathcal{V}}_{i+1,j} \mu^{\alpha\mathcal{V}}_{i,j} {\sigma^{b\mathcal{H}}_{i+1,j}}^2 + \mu^{b\mathcal{H}}_{i+1,j} ({s^{\mathcal{V}}_{i+1,j}}^2 + {g^{\mathcal{V}}_{i+1,j}}^2 {\sigma^{\alpha\mathcal{V}}_{i,j}}^2)}{{\sigma^{b\mathcal{H}}_{i+1,j}}^2 + {s^{\mathcal{V}}_{i+1,j}}^2 + {g^{\mathcal{V}}_{i+1,j}}^2 {\sigma^{\alpha\mathcal{V}}_{i,j}}^2} \\ {\sigma^{\alpha\mathcal{V}}_{i+1,j}}^2 &= \frac{{\sigma^{b\mathcal{H}}_{i+1,j}}^2 ({s^{\mathcal{V}}_{i+1,j}}^2 + {g^{\mathcal{V}}_{i+1,j}}^2 {\sigma^{\alpha\mathcal{V}}_{i,j}}^2)}{{\sigma^{b\mathcal{H}}_{i+1,j}}^2 + {s^{\mathcal{V}}_{i+1,j}}^2 + {g^{\mathcal{V}}_{i+1,j}}^2 {\sigma^{\alpha\mathcal{V}}_{i,j}}^2} \\ c^{\alpha\mathcal{V}}_{i+1,j} &= \frac{c^{\alpha\mathcal{V}}_{i,j} c^{b\mathcal{H}}_{i+1,j} \exp\left\{ -\frac{1}{2} \frac{(\mu^{b\mathcal{H}}_{i+1,j} - g^{\mathcal{V}}_{i+1,j} \mu^{\alpha\mathcal{V}}_{i,j})^2}{{\sigma^{b\mathcal{H}}_{i+1,j}}^2 + {s^{\mathcal{V}}_{i+1,j}}^2 + {g^{\mathcal{V}}_{i+1,j}}^2 {\sigma^{\alpha\mathcal{V}}_{i,j}}^2} \right\}}{(2\pi)^{\frac{1}{2}} ({\sigma^{b\mathcal{H}}_{i+1,j}}^2 + {s^{\mathcal{V}}_{i+1,j}}^2 + {g^{\mathcal{V}}_{i+1,j}}^2 {\sigma^{\alpha\mathcal{V}}_{i,j}}^2)^{\frac{1}{2}}} \end{aligned}$$

  - Termination: $P^{\mathcal{V}}_j = c^{\alpha\mathcal{V}}_{I,j}$
- *Backward β variable*:
  - Initialization: $\mu^{\beta\mathcal{V}}_{I,j} = 0 \qquad {\sigma^{\beta\mathcal{V}}_{I,j}}^2 \to \infty$
  - Recursion:

$$\begin{aligned} \mu^{\beta\mathcal{V}}_{i,j} &= \frac{1}{g^{\mathcal{V}}_{i+1,j}} \left(\frac{\mu^{b\mathcal{H}}_{i+1,j} {\sigma^{\beta\mathcal{V}}_{i+1,j}}^2 + \mu^{\beta\mathcal{V}}_{i+1,j} {\sigma^{b\mathcal{H}}_{i+1,j}}^2}{{\sigma^{\beta\mathcal{V}}_{i+1,j}}^2 + {\sigma^{b\mathcal{H}}_{i+1,j}}^2} \right) \\ {\sigma^{\beta\mathcal{V}}_{i,j}}^2 &= \frac{1}{{g^{\mathcal{V}}_{i+1,j}}^2} \left({s^{\mathcal{V}}_{i+1,j}}^2 + \frac{{\sigma^{b\mathcal{H}}_{i+1,j}}^2 {\sigma^{\beta\mathcal{V}}_{i+1,j}}^2}{{\sigma^{b\mathcal{H}}_{i+1,j}}^2 + {\sigma^{\beta\mathcal{V}}_{i+1,j}}^2} \right) \end{aligned}$$

- *Occupancy probability γ*:

$$\mu^{\gamma\mathcal{V}}_{i,j} = \frac{\mu^{\alpha\mathcal{V}}_{i,j} {\sigma^{\beta\mathcal{V}}_{i,j}}^2 + \mu^{\beta\mathcal{V}}_{i,j} {\sigma^{\alpha\mathcal{V}}_{i,j}}^2}{{\sigma^{\alpha\mathcal{V}}_{i,j}}^2 + {\sigma^{\beta\mathcal{V}}_{i,j}}^2} \quad {\sigma^{\gamma\mathcal{V}}_{i,j}}^2 = \frac{{\sigma^{\alpha\mathcal{V}}_{i,j}}^2 {\sigma^{\beta\mathcal{V}}_{i,j}}^2}{{\sigma^{\alpha\mathcal{V}}_{i,j}}^2 + {\sigma^{\beta\mathcal{V}}_{i,j}}^2}$$

Symmetric formulas can be derived for the corresponding horizontal quantities. The steps of the algorithm are very similar to the steps of the modified forward-backward for the T-HMM. Suppose we start the iterative process with row operations. The $\gamma_{i,j}^{\mathcal{V}}$ pdf's have first to be initialized. In the absence of any prior information, we set $\mu_{i,j}^{\gamma\mathcal{V}} = 0$ and $\sigma_{i,j}^{\gamma\mathcal{V}^2} \to \infty$, $\forall(i,j)$. Then the modified forward-backward algorithm is applied successively and iteratively on the rows and columns until they reach agreement. We used as a measure of convergence the average *Kullback's symmetric divergence* [4] between the distributions $\gamma_{i,j}^{\mathcal{H}}$ and $\gamma_{i,j}^{\mathcal{V}}$. This algorithm is clearly linear in the size of the data modulo the number of iterations.

Note that we do not obtain one estimate of $P(O|\lambda)$ but two: a horizontal one $P^{\mathcal{H}}(O|\lambda) = \prod_i P_i^{\mathcal{H}}$ and a vertical one $P^{\mathcal{V}}(O|\lambda) = \prod_j P_j^{\mathcal{V}}$. Combining these two scores is a classical problem of decision fusion. One can show that the optimal estimate of $P(O|\lambda)$ based on a divergence criterion is:

$$P(O|\lambda) \propto \sqrt{P^{\mathcal{H}}(O|\lambda)P^{\mathcal{V}}(O|\lambda)} \tag{2}$$

4. MOST LIKELY SEQUENCE OF STATES

The goal of this section is to find the sequence of states Q^* that "best" explains the set of observations O:

$$Q^* = \arg\max_Q P(Q|O,\lambda) = \arg\max_Q P(O,Q|\lambda)$$

We first describe a direct answer to this problem that does not make use of the T-SSM framework and explain the shortcomings of this approach. We then provide an approximate solution based on the T-SSM framework.

4.1. A direct solution

If we assume that the transition probabilities of the 2-D SSM are separable, the joint likelihood $P(O,Q|\lambda)$ can be written as a product of emission probabilities and horizontal and vertical transition probabilities. To find the best sequence of states Q^*, we set $\partial \log P(O,Q|\lambda)/\partial q_{i,j} = 0, \forall(i,j)$ and obtain a system of $I \times J$ linear equations with $I \times J$ unknowns.

If equations are ordered correctly, this is a banded system with bandwidth $\min(I,J)$. Hence, the complexity of solving this system is in $\mathcal{O}((I \times J) \times \min(I,J)^2)$ [5]. While this is much lower than the case of a general linear system, whose complexity is in $\mathcal{O}((I \times J)^3)$, it might be too demanding if I and J are large.

Finally, if ${s_{i,j}^{\mathcal{H}}}^2 \ll {\sigma_{i,j}}^2$ and ${s_{i,j}^{\mathcal{V}}}^2 \ll {\sigma_{i,j}}^2$ this system of equations is *ill-conditioned* [5], i.e. a very small perturbation on the observations $o_{i,j}$ (due to noise) or on the parameters (due to estimation errors) might lead to completely different solutions.

4.2. An approximate solution

Because of the complexity of the previous approach and to its potential instability, we explored an alternative approach based on our modified forward-backward algorithm, as applied to the T-SSM. We define $\gamma_{i,j}(q_{i,j}) = P(q_{i,j}|O,\lambda)$. To find the states that best explain the observation data, we choose the sequence of locally optimal states:

$$q_{i,j}^* = \arg\max_{q_{i,j}} \gamma_{i,j}(q_{i,j})$$

Although choosing the sequence of locally optimal states may not lead to the sequence of globally optimal states, this approximation is valid in the case where the best sequence of states accounts for most of the total probability, i.e. in the case where the distribution of state sequences is sharply peaked.

However, we do not have access to $\gamma_{i,j}$ but to its estimates $\gamma_{i,j}^{\mathcal{H}}$ and $\gamma_{i,j}^{\mathcal{V}}$. As both $\gamma_{i,j}^{\mathcal{H}}$ and $\gamma_{i,j}^{\mathcal{V}}$ are Gaussians, we can assume that $\gamma_{i,j}$ is Gaussian with mean $\mu_{i,j}^{\gamma}$ and thus $q_{i,j}^* = \mu_{i,j}^{\gamma}$. Using one more time a criterion based on the Kullback-Leibler divergence, the optimal combination rule is:

$$\mu_{i,j}^{\gamma} = \frac{\sigma_{i,j}^{\gamma\mathcal{V}^2}\mu_{i,j}^{\gamma\mathcal{H}} + \sigma_{i,j}^{\gamma\mathcal{H}^2}\mu_{i,j}^{\gamma\mathcal{V}}}{\sigma_{i,j}^{\gamma\mathcal{V}^2} + \sigma_{i,j}^{\gamma\mathcal{H}^2}}$$

Moreover, one can show from the equations derived in the previous section that $|\mu_{i,j}^{\gamma\mathcal{H}} - \mu_{i,j}^{\gamma\mathcal{V}}|$ converges toward 0. While we have not shown that $\mu_{i,j}^{\gamma\mathcal{H}}$ and $\mu_{i,j}^{\gamma\mathcal{V}}$ actually converge, we found experimentally that it was the case.

5. MODEL PARAMETERS ESTIMATION

The goal of this section is to adjust the model parameters λ to maximize $P(O|\lambda)$. Let the Q function be defined as [1]:

$$Q(\lambda,\lambda') = \int_Q P(O,Q|\lambda') \log P(O,Q|\lambda) dQ$$

If λ' is the initial estimate, then it was proven that the maximization of Q with respect to λ leads to an increased likelihood. Introducing $Q^{\mathcal{H}}(\lambda,\lambda') = \int_Q P(O,Q|\lambda') \log P^{\mathcal{H}}(O,Q|\lambda) dQ$ and $Q^{\mathcal{V}}(\lambda,\lambda') = \int_Q P(O,Q|\lambda') \log P^{\mathcal{V}}(O,Q|\lambda) dQ$ and using approximation (2), we get:

$$Q(\lambda,\lambda') \approx \frac{1}{2}\left[Q^{\mathcal{H}}(\lambda,\lambda') + Q^{\mathcal{V}}(\lambda,\lambda')\right] + C$$

where C is a constant which is independent of λ. The parameter estimation is done by setting $\partial Q(\lambda,\lambda')/\partial\lambda = 0$.

6. EXPERIMENTAL RESULTS

We recently introduced a novel approach to face recognition which consists in modeling the set of possible transformations between face images of the same person [6]. The global face transformation is approximated with a set of local transformations under the constraint that neighboring transformations must be consistent with each other. Local transformations and neighboring constraints are embedded within the probabilistic framework of the T-HMM. At any position on the face, the system is in a state where each state represents a local transformation. Emission probabilities model the cost of local transformations and transition probabilities relate states of neighboring regions and implement the consistency rules.

While [6] focuses on *discrete geometric transformations* to model facial expressions, [7] considers *continuous feature transformations* to compensate for illumination variations. The idea is that the illumination cannot vary in an arbitrary manner over the face and, thus, that the illumination variation should be constrained. In this section, we summarize extensive experimental results presented in [7] (while this paper used the T-SSM, it did not give any detail about its mathematical framework).

The states of the system are doubly indexed and can be split into geometric and illumination parts. A central idea in our approach is to apply iterative passes to find successively the geometric and feature transformations that best explain the transformation between the two face images. The geometric transformations are estimated using the T-HMM framework and the feature transformations using the T-SSM framework.

[7] shows that, if we apply a log transform in the pixel domain and if the feature extraction step involves only linear operators, such as the convolution, then the illumination is additive in the feature domain. In all the following experiments, a log transform was applied in the pixel domain before the feature extraction. [7] also introduces the *Log-Mean Normalization* (or *LM-Norm*) which consists in applying a log in the pixel domain before the feature extraction and a mean normalization in each feature component. If the illumination was constant in each feature component across the whole face, LM-Norm would be a simple approach to removing the undesired additive illumination term. It will thus serve as a baseline for our novel illumination compensation algorithm.

We now detail our model of illumination variation. Due to the additive nature of the illumination term, we choose $f_{i,j} = 1$. Moreover, we assume that $g^{\mathcal{H}}_{i,j} = g^{\mathcal{V}}_{i,j} = 1$, $\forall(i,j)$ and that the initial occupancy probabilities are maximally non-informative, i.e. ${s^{\mathcal{H}}_{i,1}}^2 \rightarrow \infty$, $\forall i$ and ${s^{\mathcal{V}}_{1,j}}^2 \rightarrow \infty$, $\forall j$. The transition probabilities variances, i.e. ${s^{\mathcal{H}}_{i,j}}^2$'s and ${s^{\mathcal{V}}_{i,j}}^2$'s which model the speed of variation of the illumination are the only parameters than need to be trained in our illumination variation model. In the following, we will assume that ${s^{\mathcal{H}}_{i,j}}^2 = {s^{\mathcal{V}}_{i,j}}^2 = s^2$, $\forall(i,j)$.

We used two face databases to assess the performance of our novel approach: the FERET database [8] to train the system and the YALE B database [9] to test it. 500 individuals were extracted from the FAFB set of FERET which contains frontal views that exhibit large variations in facial expressions but very little variability in terms of illumination. We also used the 200 individuals in the FAFC set which contains frontal views that exhibit large variations in illumination conditions and facial expressions. The YALE B face database contains the images of 10 subjects under different poses and illumination conditions. We used only the set which contains frontal face images. We divided the database into the four traditional subsets $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$ and $\mathcal{S}_4$ according to the angle the light source makes with the axis of the camera (less than $12°$, between $12°$ and $25°$, between $25°$ and $50°$ and between $50°$ and $77°$). For each person, the 7 images in $\mathcal{S}_1$ were successively used as the enrollment image and the images in $\mathcal{S}_2$, $\mathcal{S}_3$ and $\mathcal{S}_4$ were used as test images which made a total of 26,600 comparisons.

We performed three tests. The first experiment, which will be referred to as the baseline, is based only on the model introduced in [6] which does not make use of any feature transformation. LM-Norm was performed on the features and the system was trained on the FAFB and FAFC data. In the second and third experiments, we used the illumination compensation approach introduced in [7]. We first trained the geometric transformation part of the system only on the FAFB data as described in [6]. Then using the FAFC data, the illumination transformation part of the system was trained. The difference between the second and third experiments, is in the use of the direct solution described in 4.1 or of the approximate solution based on the T-SSM described in 4.2.

Results are presented on Figure 1. The average identification rates are respectively 84.4% for the LM-Norm baseline, 89.1% for the variant of our system that uses the direct solution and 90.8% for

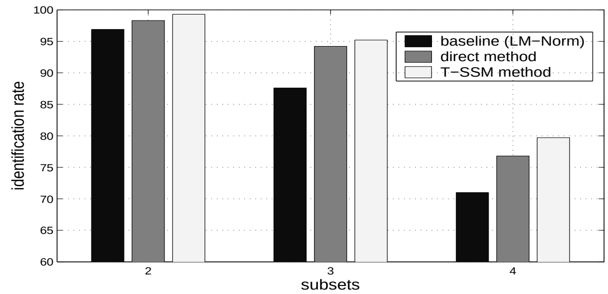

Fig. 1. Performance of the baseline system with LM-Norm compared to the direct (c.f. 4.1) and T-SSM (c.f. 4.2) variants of the illumination compensation algorithm.

the variant that uses the T-SSM solution. We believe these results show the potential of the T-SSM framework.

7. CONCLUSION

In this paper, we extended the discrete state T-HMM to the continuous state T-SSM. We provided efficient approximate answers to the three following problems: 1) how to compute the likelihood of a set of observations, 2) how to find the sequence of states that best "explains" a set of observations and 3) how to estimate the model parameters given a set of observations. This work was applied to a face recognition system to compensate for illumination variations.

Future work on the T-SSM will concentrate on alleviating the constraint of Gaussian emission and transition probabilities.

8. REFERENCES

[1] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications," *Proc. of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb 1989.

[2] F. Perronnin, J.-L. Dugelay and K. Rose, "Iterative decoding of two-dimensional hidden Markov models," in *IEEE ICASSP*, 2003, vol. 3, pp. 329–332.

[3] T. Kailath, A. H. Sayed and B. Hassibi, *Linear estimation*, Prentice Hall, 2000.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1993.

[5] A. Quarteroni, R. Sacco and F. Saleri, *Numerical Mathematics*, Springer-Verlag, 2000.

[6] F. Perronnin, J.-L. Dugelay and K. Rose, "Deformable face mapping for person identification," in *IEEE ICIP*, 2003, vol. 1, pp. 661–664.

[7] F. Perronnin and J.-L. Dugelay, "A model of illumination variation for robust face recognition," in *MMUA workshop*, 2003, pp. 157–164.

[8] P. J. Phillips, H. Moon, S. A. Rizvi and P. J. Rauss, "The feret evaluation methodology for face recognition algorithms," *IEEE Trans. on PAMI*, vol. 22, no. 10, pp. 1090–1104, Oct 2000.

[9] A. S. Georghiades, P. N. Belhumeur and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. on PAMI*, vol. 23, no. 6, pp. 643–660, June 2001.