# A Model of Illumination Variation for Robust Face Recognition

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 approach to face recognition which consists in modeling the set of possible transformations between face images of the same person. While our previous work focused on geometric transformations to model facial expressions, in this article we consider feature transformations as a means to compensate for illumination variations. Although this approach requires to learn the set of possible illumination transformations through a training phase, we will show experimentally that the trained parameters are very robust. Even in the challenging case where the databases used to train the transformation model and to assess the performance of the system are very different, the proposed approach results in large improvements of the recognition rate.*

## 1. Introduction

Pattern classification deals with the general problem of inferring classes from observations [1]. Hence, the success of a pattern classification system is based on its ability to distinguish between inter- and intra-class variabilities. Face recognition is a very challenging task as different faces have the same global shape while face images of the same person are subject to a wide range of variabilities including facial expressions, pose, illumination conditions, presence or absence of eyeglasses and facial hair, aging, occlusion, etc. Illumination, which will be the focus of this paper, remains one of the toughest variabilities to cope with as shown during the FERET evaluation [2] and the facial recognition vendor test 2000 [3].

It is possible to deal with the illumination at three different stages: during the *preprocessing*, the *feature extraction* or the *classification*.

Preprocessing algorithms for illumination compensation include general image processing tools such as histogram equalization and gamma correction [4]. A simple but very effective preprocessing, which is based on Weber's law, consists in applying a logarithm transform to the image intensity [5, 6]. Another class of preprocessing algorithms consists in separating an image into its reflectance and illumination fields [7]. An assumption which is generally made for this type of approach is that the luminance varies slowly across the image while sharp changes can occur in the reflectance.

At the feature extraction stage, the goal is to derive features that are invariant to illumination. Edge maps, derivatives of the gray level and Gabor features were compared in [5] and an empirical study showed that none of these features was sufficient to overcome the variations due to changes in the direction of illumination. Another idea is to *learn* features which are insensitive to illumination variations such as the Fisherfaces [8].

Finally, various algorithms have been proposed to cope with the illumination variation at the classification stage. The idea underlying [9] is that the set of images of an object in fixed pose, but under all possible illumination conditions, is a convex cone in the space of images that can be approximated by low dimensional linear subspaces. [10] proposed an approach based on 3D morphable models which encode both shape and texture information and an algorithm that recovers these parameters from a single face image.

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 [11]. While our previous work focused on *geometric* transformations to model facial expressions, we introduce in this article *feature* transformations as a means to compensate for illumination variations. This approach to illumination compensation, which works at the classification stage, involves a training phase to learn the set of possible illumination transformations. While approaches based on learning can suffer from poor generalization when the training and test sets are different, we will show experimentally the good generalization ability of our approach.

The remainder of this paper is organized as follows. A brief review of the probabilistic model of face transforma-

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

tion is given in the next section. Section 3 introduces our model of illumination transformation. Section 4 focuses on how to find jointly the best set of geometric and feature transformations between two face images. Finally, section 5 summarizes experimental results for a face identification task. While it is common to train and test a system on the same database, to assess the performance of our novel illumination compensation algorithm we used two very different databases. We think this is a much more realistic approach as, in practice, one never has access at training time to the exact test conditions. Even in this challenging case the proposed approach results in large improvements of the recognition rate.

## 2. A model of face transformation

### 2.1. Framework

While most face recognition techniques directly model the face, [11] models the set of possible *transformations* between face images of the same person. 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 consistency costs are embedded within the probabilistic framework of a 2D HMM. At any position on the query face image, the system is in one of a finite set of states 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.

A major assumption in our system is that the intra-class variability is the same for all classes and, thus, that the model of face transformation is *shared* by all individuals. Hence, it can be trained on pairs of images of persons that are not enrolled in the system.

### 2.2. Local Transformations

Let us assume that we have two face images: a template image $\mathcal{F}_T$ and a query image $\mathcal{F}_Q$. Feature vectors are extracted on a sparse grid from $\mathcal{F}_Q$ and on a dense grid from $\mathcal{F}_T$. We then apply a set of local transformations at each position $(i, j)$ of the sparse grid. In our previous work, these transformations were limited to geometric transformations and, more precisely, to translations. Each translation maps a feature vector of $\mathcal{F}_Q$ with a feature vector in $\mathcal{F}_T$.

Let $o_{i,j}$ be the observation extracted from $\mathcal{F}_Q$ at position $(i, j)$ and let $q_{i,j}$ be the associated state (i.e. local deformation). If $\tau$ is a translation vector, the probability that at position $(i, j)$ the system emits observation $o_{i,j}$, knowing that it is in state $q_{i,j} = \tau$, is $b^{\tau}_{i,j}(o_{i,j}) = P(o_{i,j}|q_{i,j} = \tau, \lambda)$ where $\lambda = (\lambda_T, \lambda_{\mathcal{M}})$. We separate $\lambda$ into *face dependent* (FD) parameters $\lambda_T$ which are extracted from $\mathcal{F}_T$ and *face independent transformation* (FIT) parameters $\lambda_{\mathcal{M}}$, i.e. the parameters of the shared transformation model $\mathcal{M}$. The emission probability $b^{\tau}_{i,j}(o_{i,j})$ represents the cost of matching $o_{i,j}$ with the corresponding feature vector in $\mathcal{F}_T$ that will be denoted $m^{\tau}_{i,j}$. $b^{\tau}_{i,j}(o_{i,j})$ is modeled with a mixture of Gaussians as linear combinations of Gaussians have the ability to approximate arbitrarily shaped densities:

$$b^{\tau}_{i,j}(o_{i,j}) = \sum_k w^k_{i,j} b^{\tau,k}_{i,j}(o_{i,j})$$

$b^{\tau,k}_{i,j}(o_{i,j})$'s are the component densities and the $w^k_{i,j}$'s are the mixture weights and must satisfy the following constraint: $\forall (i,j)$, $\sum_k w^k_{i,j} = 1$. Each component density is a $D$-variate Gaussian function of the form:

$$b^{\tau,k}_{i,j}(o_{i,j}) = \frac{\exp\left\{-\frac{1}{2}(o_{i,j} - \mu^{\tau,k}_{i,j})^T \Sigma^{k(-1)}_{i,j}(o_{i,j} - \mu^{\tau,k}_{i,j})\right\}}{(2\pi)^{\frac{N}{2}} |\Sigma^k_{i,j}|^{\frac{1}{2}}}$$

where $\mu^{\tau,k}_{i,j}$ and $\Sigma^k_{i,j}$ are respectively the mean and covariance matrix of the Gaussian, $D$ is the size of feature vectors and $|.|$ is the determinant operator. We use a bi-partite model which separates the mean into additive FD and FIT parts:

$$\mu^{\tau,k}_{i,j} = m^{\tau}_{i,j} + \delta^k_{i,j} \tag{1}$$

where $m^{\tau}_{i,j}$ is the FD part of the mean. $w^k_{i,j}$, $\delta^k_{i,j}$ and $\Sigma^k_{i,j}$ are FIT parameters. Intuitively, $b^{\tau}_{i,j}$ should be approximately centered and maximum around $m^{\tau}_{i,j}$.

### 2.3. Neighborhood Consistency

The neighborhood consistency of the local transformations is ensured via the transition probabilities of the 2D HMM. We explain in the next section that a 2D HMM can be approximated by a set of interdependent horizontal and vertical 1D HMMs. The transition probabilities of the horizontal and vertical 1D HMMs are $P(q_{i,j} = \tau | q_{i,j-1} = \tau', \lambda)$ and $P(q_{i,j} = \tau | q_{i-1,j} = \tau', \lambda)$. They model respectively the horizontal and vertical elastic properties of the face at position $(i, j)$ and are part of the face transformation model $\mathcal{M}$.

### 2.4. Turbo-HMMs

While HMMs have been extensively applied to 1D problems, the complexity of their extension to 2D grows exponentially with the data size and is intractable in most cases of interest. [12] introduced Turbo-HMMs (T-HMMs), in reference to the turbo error-correcting codes, to approximate the computationally intractable 2D HMMs. A T-HMM consists of horizontal and vertical 1D HMMs that "communicate" through an iterative process by inducing prior probabilities on each other. The T-HMM framework provides efficient formulas to 1) compute efficiently $P(\mathcal{F}_Q|\mathcal{F}_T, \mathcal{M})$, i.e. the probability that $\mathcal{F}_T$ and $\mathcal{F}_Q$ belong to the same person knowing the face transformation model $\mathcal{M}$, and 2) train automatically all the parameters of $\mathcal{M}$.

The computation of $P(\mathcal{F}_Q|\mathcal{F}_T, \mathcal{M})$ is based on a modified version of the forward-backward algorithm which is applied successively and iteratively on the horizontal and vertical 1D HMMs until they reach agreement.

The *Maximum Likelihood Estimation* (MLE) of the parameters of $\mathcal{M}$ is based on a modified version of the *Baum-Welch* algorithm. To train $\mathcal{M}$, we present pairs of pictures (a template and a query image) that belong to the same persons and optimize the transformation parameters $\lambda_{\mathcal{M}}$ to maximize the likelihood of the pairs of pictures.

## 3. Modeling the illumination variation

In this section, we will first show how to transform the illumination into an additive variability in the feature domain and then, how to constrain the illumination variation.

### 3.1. The illumination as an additive variability

The starting point of our approach is the well-known assumption that an image $I$ can be seen as the product of a reflectance $R$ and an illumination $L$ [13]:

$$I(x, y) = R(x, y) \times L(x, y)$$

Applying the logarithm operator, we obtain:

$$\log I(x, y) = \log R(x, y) + \log L(x, y)$$

and the illumination turns into an additive term in the pixel domain. If the feature extraction involves only linear operators, such as the convolution, the illumination remains additive in the feature domain. Denoting $F_d$ the linear feature extraction operator for the $d$-th dimension of the feature vectors and $o_{i,j} = \{o_{i,j}[1], ...o_{i,j}[D]\}$ the feature vector extracted at position $(i, j)$, we get:

$$\begin{aligned} o_{i,j}[d] &= F_d\{\log I(x, y)\} \\ &= F_d\{\log R(x, y)\} + F_d\{\log L(x, y)\} \end{aligned}$$

Hence, if the illumination was constant in each feature component across the whole face, subtracting in each component the average value $\bar{o}[d]$ would be a simple approach to removing the undesired additive illumination term. However, the illumination is unlikely to be perfectly constant in each component. Moreover, when subtracting $\bar{o}[d]$, one may also discard useful reflectance information. Nevertheless, this simple combination of logarithm transform in the pixel domain and mean normalization in the feature domain, that will be referred to as the *Log-Mean Normalization* (or *LM-Norm*), and which, to the best of our knowledge, has never been suggested, will be tested in the section on experimental results.

Our goal is now to alleviate the unrealistic constraint of a constant illumination in each frequency band. As the system described in section 2 is designed to model additive variabilities, as expressed by equation (1), a first idea would be to train the Gaussian mixtures parameters, i.e. $w$'s, $\delta$'s and $\Sigma$'s, not only to model the facial expression variations, but also the various possible illumination conditions. Although this approach might first sound appealing, we believe it is suboptimal for two main reasons :

- A very large number of Gaussians would be necessary to model all the possible variabilities, increasing unreasonably the memory and CPU requirements.
- The choice of Gaussians at adjacent positions would be unconstrained, which is not satisfying as the illumination cannot vary in an arbitrary manner over the face.

However, the performance of this approach will also be evaluated in the section on experimental results and will serve as a baseline for our novel model of illumination transformation.

### 3.2. Constraining the illumination variation

The idea is to introduce feature transformations to model the illumination variation and to enforce consistency between feature transformations at adjacent positions in the same manner we enforced consistency between geometric transformations. Hence, our states which represent both local geometric and feature transformations are now doubly indexed: $q_{i,j} = (q^1_{i,j}, q^2_{i,j})$. $q^1_{i,j}$ is the geometric transformation part of the state and $q^2_{i,j}$ is the feature transformation part. If $q_{i,j} = (\tau, \phi)$, the emission probability $b^{\tau,\phi}_{i,j}$ is still modeled with a mixture of Gaussians:

$$b^{\tau,\phi}_{i,j} = \sum_k w^k_{i,j} b^{\tau,\phi,k}_{i,j}$$

where the $b^{\tau,\phi,k}_{i,j}$'s are $D$-variate Gaussians with means $\mu^{\tau,\phi,k}_{i,j}$ and covariance matrices $\Sigma^k_{i,j}$. The new means are of the form:

$$\mu^{\tau,\phi,k}_{i,j} = \mu^{\tau,k}_{i,j} + \phi = m^\tau_{i,j} + \delta^k_{i,j} + \phi$$

In [11] we only separated parameters into FD and FIT parameters. Here, we go one step further by separating the FIT parameters into geometrical transformation parameters and feature transformation parameters.

If we assume that geometric and feature transformations model respectively differences in facial expression and illumination between images, and that facial expression and illumination *variations* are mostly independent (i.e. a facial expression change between two adjacent positions has a limited impact on the illumination change between the same positions and vice versa), then the horizontal and vertical transition probabilities can be separated as follows:

$$\begin{aligned} P(q_{i,j}|q_{i,j-1}) &= P(q^1_{i,j}|q^1_{i,j-1}) \times P(q^2_{i,j}|q^2_{i,j-1}) \\ P(q_{i,j}|q_{i-1,j}) &= P(q^1_{i,j}|q^1_{i-1,j}) \times P(q^2_{i,j}|q^2_{i-1,j}) \end{aligned}$$

While the choice of a discrete number of geometric transformations is natural due to the discrete nature of the feature extraction grid of the template image, it is easier to deal with the illumination with an *infinite continuous* set of illumination states. We choose the horizontal and vertical illumination components of the transition probabilities to be $D$-variate Gaussians:

$$\begin{aligned} &P(q^2_{i,j} = \phi|q^2_{i,j-1} = \phi') = P(q^2_{i,j} = \phi|q^2_{i-1,j} = \phi') \\ &\qquad = \frac{\exp\left\{-\frac{1}{2}(\phi-\phi')^T S^{(-1)}(\phi-\phi')\right\}}{(2\pi)^{\frac{N}{2}}|S|^{\frac{1}{2}}} \end{aligned}$$

In the following we will assume that the covariance matrix $S$ is diagonal and therefore, that the components of the feature vectors are independent from each other. $S$ is the only parameter of our illumination transformation model.

## 4. Finding the best transformation

Let $O = \{o_{i,j}\}$ and $Q = \{q_{i,j}\}$ denote respectively the set of all observations and states, with $i \in [1, I]$ and $j \in [1, J]$. Finding the best transformation between two face images requires to find the sequence of states $Q^*$, which satisfies:

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

where $Q = (T, \Phi)$ and $T = \{\tau_{i,j}\}$ and $\Phi = \{\phi_{i,j}\}$ correspond respectively to the set of geometric and feature transformations. 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.

Let $Q_n = (T_n, \Phi_n)$ be the best set of states after the $n$-th iteration. Assuming for instance that we start by decoding geometric transformations, the steps of the algorithm are as follows:

1. Initialize $\Phi_0$: $\forall(i,j)$, $\phi_{i,j} = 0$, i.e. we assume there is no illumination variation between the two images.

2. $T_n = \arg\max_T \log P(O,T|\Phi_{n-1},\lambda)$, i.e. $T_n$ maximizes the joint probability of observations and geometric transformations knowing $\Phi_{n-1}$, the set of previously obtained feature transformations.

3. $\Phi_n = \arg\max_\Phi \log P(O,\Phi|T_n,\lambda)$, i.e. $\Phi_n$ maximizes the joint probability of observations and feature transformations knowing $T_n$, the set of geometric transformations previously obtained.

4. Go back to step 2 until $T_n$ and $\Phi_n$ converge.

We will now detail the steps 2 and 3 of this algorithm.

### 4.1. Finding $T_n$

To find the best sequence of geometric transformations $T_n$, one applies the modified version of the forward-backward algorithm introduced in [12] and estimates the occupancy probabilities $\gamma_{i,j}(t) = P(q^1_{i,j} = t|O,\Phi_{n-1},\lambda)$, i.e. the probability of being in state $q^1_{i,j} = t$ at position $(i,j)$. At each position $(i,j)$, we look for the best state $\tau$:

$$\tau = \arg\max_t \gamma_{i,j}(t)$$

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.

If $\gamma_{i,j}(\tau,n)$ is the probability of being in state $\tau$ with the $n$-th mixture component accounting for $o_{i,j}$, the best Gaussian index $k$ is given by:

$$k = \arg\max_n \gamma_{i,j}(\tau,n)$$

If $\tau$ and $k$ are respectively the indexes of the best state and Gaussian at position $(i,j)$, we introduce the quantity $\Psi^{\tau,k}_{i,j} = (o_{i,j} - \mu^{\tau,k}_{i,j})$ which can be interpreted as the variability that is left unexplained by the geometric transformations. Let $\Sigma^k_{i,j}$ be the covariance of the best Gaussian at position $(i,j)$. In the following, for simplicity, we will drop the $\tau$ and $k$ indexes and replace the notation $\Psi^{\tau,k}_{i,j}$ with $\Psi_{i,j}$ and $\Sigma^k_{i,j}$ with $\Sigma_{i,j}$.

### 4.2. Finding $\Phi_n$

To find the best sequence of feature transformations $\Phi_n$, we can pursue two different approaches: either apply directly the *Viterbi* algorithm, or a modified version of the *forward-backward*. In both cases, as $\Sigma_{i,j}$ and $S$ the covariances of the emission and transition probabilities are assumed diagonal, it it simple to show that finding the best state sequence $\Phi$ can be done independently in each of the $D$ dimensions. Therefore, if $\Psi_{i,j} = [\psi_{i,j}[1], ...\psi_{i,j}[D]]^T$, $\Sigma_{i,j} = \text{diag}\{\sigma_{i,j}[1]^2, ...\sigma_{i,j}[D]^2\}$ and $S = \text{diag}\{s[1]^2, ... s[D]^2\}$ in the following, we drop the dimension indexes and use the notations $\psi_{i,j}$, $\sigma^2_{i,j}$ and $s^2$.

### 4.2.1. Viterbi variant

We assume that transition probabilities are separable, i.e.:

$$P(q^2_{i,j}|q^2_{i-1,j}, q^2_{i,j-1}) \propto P(q^2_{i,j}|q^2_{i-1,j})P(q^2_{i,j}|, q^2_{i,j-1})$$

(see [12] for more details on this approximation). The joint likelihood $P(O, \Phi|T_n, \lambda)$ can be written as a product of emission probabilities and horizontal and vertical transition probabilities. For one given dimension, to find the best sequence of states $\Phi_n$, we set $\partial \log P(O, \Phi|T_n, \lambda)/\partial \phi_{i,j} = 0$ , $\forall(i, j)$ and obtain:

$$\begin{gathered}\phi_{i-1,j} + \phi_{i+1,j} + \phi_{i,j-1} + \phi_{i,j+1} - \\ \phi_{i,j}\left(\frac{s^2}{\sigma^2_{i,j}} + 4\right) = -\psi_{i,j}\left(\frac{s^2}{\sigma^2_{i,j}}\right) , \forall(i,j)\end{gathered}$$

with obvious modifications for $i = 1$ or $I$ and $j = 1$ or $J$. This is a linear system of $I \times J$ equations with $I \times J$ unknowns. If equations are ordered properly, this system is banded with bandwidth $\min(I, J)$. Hence, the complexity of solving this system is in $\mathcal{O}((I \times J) \times \min(I, J)^2)$. We recall that there are $D$ such systems to solve, one per dimension of the feature vectors.

At training time, to find the optimal $s^2$ which maximizes $\log P(O, \Phi|T_n, \lambda)$, we set $\partial \log P(O, \Phi|T_n, \lambda)/\partial s^2 = 0$ and obtain:

$$\hat{s}^2 = \frac{\sum_{i,j}\left[(\phi_{i,j} - \phi_{i-1,j})^2 + (\phi_{i,j} - \phi_{i,j-1})^2\right]}{(I-1) \times J + I \times (J-1)}$$

In the previous formula, $s^2$ is estimated with one pair of images. The extension to multiple pairs of images is straightforward.

### 4.2.2. Forward-backward variant

A complexity in $\mathcal{O}((I \times J) \times \min(I, J)^2)$ is much lower than the complexity of solving a general linear system of $I \times J$ equations with $I \times J$ unknowns which is in $\mathcal{O}((I \times J)^3)$. However it might still be too demanding if $I$ and $J$ are large. Therefore, we explored an alternative approach which is based on our modified forward-backward algorithm, as applied to T-HMMs [12]. The extension from discrete states HMMs to continuous states HMMs (also referred to as *state space models* or *SSMs*) consists mainly in replacing sums with integrals.

We define $\gamma_{i,j}(\phi) = P(q^2_{i,j} = \phi|O, T_n, \lambda)$, i.e. the probability of being in state $\phi$ at position $(i, j)$. To find the states that best explain the illumination transformation, we choose the sequence of locally optimal states $\Phi$, i.e.:

$$\phi_{i,j} = \arg\max_{\phi} \gamma_{i,j}(\phi)$$

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

$$\begin{aligned}\alpha^{\mathcal{V}}_{i,j}(\phi) &= P(o_{1,j}, ...o_{i,j}, q^2_{i,j} = \phi|T_n, \lambda) \\ \beta^{\mathcal{V}}_{i,j}(\phi) &= P(o_{i+1,j}, ...o_{I,j}|q^2_{i,j} = \phi, T_n, \lambda) \\ \gamma^{\mathcal{V}}_{i,j}(\phi) &= P(q^2_{i,j} = \phi|o_{1,j}, ...o_{I,j}, T_n, \lambda)\end{aligned}$$

Defining the corresponding horizontal quantities is straightforward. As the emission and transition probabilities are Gaussians, if we initialize the occupancy probabilities $\gamma$'s in a Gaussian manner, one can show that the forward, backward and occupancy probabilities are Gaussian shaped. The parameters of these Gaussians, i.e. their means and variances, will be respectively denoted $\mu^{\alpha\mathcal{V}}_{i,j}$, $\mu^{\beta\mathcal{V}}_{i,j}$, $\mu^{\gamma\mathcal{V}}_{i,j}$ and $\sigma^{\alpha\mathcal{V}^2}_{i,j}$, $\sigma^{\beta\mathcal{V}^2}_{i,j}$, $\sigma^{\gamma\mathcal{V}^2}_{i,j}$. It is easy to show that we have:

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

Successive horizontal and vertical passes of our modified forward-backward (extended to T-HMMs with an infinite continuous set of states) are applied iteratively to estimate $\mu^{\alpha\mathcal{V}}_{i,j}$, $\mu^{\beta\mathcal{V}}_{i,j}$, $\sigma^{\alpha\mathcal{V}^2}_{i,j}$ and $\sigma^{\beta\mathcal{V}^2}_{i,j}$ until convergence of the $\gamma^{\mathcal{H}}_{i,j}$ and $\gamma^{\mathcal{V}}_{i,j}$ probability densities. As we do not have access to $\gamma_{i,j}$ but to $\gamma^{\mathcal{H}}_{i,j}$ and $\gamma^{\mathcal{V}}_{i,j}$, a simple combination rule based on the minimum divergence criterion is to set:

$$\phi_{i,j} = \frac{\sigma^{\gamma\mathcal{V}^2}_{i,j}\mu^{\gamma\mathcal{H}}_{i,j} + \sigma^{\gamma\mathcal{H}^2}_{i,j}\mu^{\gamma\mathcal{V}}_{i,j}}{\sigma^{\gamma\mathcal{V}^2}_{i,j} + \sigma^{\gamma\mathcal{H}^2}_{i,j}}$$

The complexity of this algorithm is clearly in $\mathcal{O}(I \times J \times N)$ where $N$ is the number of horizontal and vertical passes.

The optimal parameter $s^2$ is given by:

$$\hat{s}^2 = \frac{\sum_{i,j}\int_{\phi,\phi'}(\phi - \phi')^2\left[\xi^{\mathcal{H}}_{i,j}(\phi, \phi') + \xi^{\mathcal{V}}_{i,j}(\phi, \phi')\right]d\phi d\phi'}{(I-1) \times J + I \times (J-1)}$$

where $\xi^{\mathcal{H}}_{i,j}(\phi, \phi') = P(q^2_{i,j-1} = \phi, q^2_{i,j} = \phi'|O, T_n, \lambda)$ and $\xi^{\mathcal{V}}_{i,j}(\phi, \phi') = P(q^2_{i-1,j} = \phi, q^2_{i,j} = \phi'|O, T_n, \lambda)$. Introducing the notations $\rho^{\alpha\mathcal{H}}_{i,j} = s^2/(s^2 + \sigma^{\alpha\mathcal{H}^2}_{i,j})$ and $\rho^{\alpha\mathcal{V}}_{i,j} = s^2/(s^2 + \sigma^{\alpha\mathcal{V}^2}_{i,j})$, we get:

$$\begin{aligned}\hat{s}^2 &= \frac{\sum_{i,j}\left[(\mu^{\gamma\mathcal{H}}_{i,j} - \mu^{\gamma\mathcal{H}}_{i,j-1})^2 + (\mu^{\gamma\mathcal{V}}_{i,j} - \mu^{\gamma\mathcal{V}}_{i-1,j})^2\right]}{(I-1) \times J + I \times (J-1)} \\ &+ \frac{\sum_{i,j}\left[\rho^{\alpha\mathcal{H}}_{i,j-1}\sigma^{\alpha\mathcal{H}}_{i,j-1}{}^2 + \rho^{\alpha\mathcal{H}}_{i,j-1}{}^2\sigma^{\gamma\mathcal{H}^2}_{i,j}\right]}{(I-1) \times J + I \times (J-1)} \\ &+ \frac{\sum_{i,j}\left[\rho^{\alpha\mathcal{V}}_{i-1,j}\sigma^{\alpha\mathcal{V}}_{i-1,j}{}^2 + \rho^{\alpha\mathcal{V}}_{i-1,j}{}^2\sigma^{\gamma\mathcal{V}^2}_{i,j}\right]}{(I-1) \times J + I \times (J-1)}\end{aligned}$$

The term $(\mu^{\gamma\mathcal{H}}_{i,j} - \mu^{\gamma\mathcal{H}}_{i,j-1})^2 + (\mu^{\gamma\mathcal{V}}_{i,j} - \mu^{\gamma\mathcal{V}}_{i-1,j})^2$ corresponds to $(\phi_{i,j} - \phi_{i,j-1})^2 + (\phi_{i,j} - \phi_{i-1,j})^2$ in the re-estimation formula of the Viterbi variant (c.f. the previous section). The additional terms are due to the fact that the forward-backward algorithm integrates over all paths to estimate $s^2$ while Viterbi only takes into account the best path.

## 5. Experimental results

In this section, we will first introduce the databases used to train and test our system and briefly describe Gabor features. We will then evaluate the performance of the LM-Norm introduced in section 3.1 and finally the performance of our novel model of illumination transformation.

### 5.1. Databases

#### 5.1.1. The FERET face database

To train our transformation model, we used the FERET face database [2]. 500 individuals were extracted from the FAFB set which contains frontal views that exhibit large variations in facial expressions but very little variability in terms of illumination. There are two images per person in the FAFB set. We also used the 200 individuals in the FAFC set which contains frontal views that exhibit large variations in illumination conditions and facial expressions. There are three images per person in the FAFC set. All the FERET images were pre-processed to extract 128x128 pixels normalized facial regions.

#### 5.1.2. The YALE B face database

The YALE B face database [9] was used to assess the performance of our system. It contains the images of 10 subjects under 9 different poses and 64 illumination conditions. As the focus of this paper is on illumination compensation, 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. The same pre-processing that was applied to the FERET images was applied to the Yale B face images.

### 5.2. Gabor features

In our experiments, we used Gabor features which have long been successfully applied to face recognition and facial analysis. Assuming polar coordinates $(\rho, \theta)$, the spectral half plane is partitioned into $M$ frequency and $N$ orientation bands [14]:

$$G_{i,j}(\rho,\theta) = \exp\left\{-\frac{1}{2}\left[\frac{(\rho-\omega_{\rho_i})^2}{\sigma^2_{\rho_i}} + \frac{(\theta-\omega_{\theta_j})^2}{\sigma^2_{\theta_i}}\right]\right\}$$

$$\text{with } i \in [1,M] \text{ and } j \in [1,N]$$

The parameters $\omega_{\rho_i}$, $\sigma_{\rho_i}$, $\omega_{\theta_j}$ and $\sigma_{\theta_i}$ are defined as follows:

$$\omega_{\rho_i} = \omega_{min} + \sigma_0 \tfrac{(f+1)f^{i-1}-2}{f-1} \qquad \sigma_{\rho_i} = \sigma_0 f^{i-1}$$

$$\omega_{\theta_j} = \tfrac{(j-1)\pi}{N} \qquad \sigma_{\theta_i} = \tfrac{\pi\omega_{\rho_i}}{2N}$$

After preliminary experiments, we chose $\omega_{min} = \pi/24$, $\omega_{max} = \pi/3$, $f = \sqrt{2}$, $M = 4$ and $N = 6$, which resulted in 24 dimensional feature vectors. Gabor responses are obtained through the convolution of an image and the Gabor wavelets. We use the modulus of these responses as feature vectors which introduces a non-linearity in the computation of our features. Thus, the illumination cannot be considered as a perfectly additive term in the feature domain.

Feature vectors were extracted every 16 pixels of the query images and every 4 pixels of the template images in both horizontal and vertical directions.

### 5.3. Performance of the LM-Norm

The goal of this section is to assess the performance of the LM-Norm introduced in 3.1. In this first set of experiments, we applied straightforwardly the face transformation model introduced in [11] which does not make use of feature transformations.

When the LM-Norm is associated to Gabor features, the feature extraction consists of 3 steps:

1. logarithm transform in the pixel domain
2. Gabor features extraction
3. mean normalization in each frequency band

Gabor features combined with LM-Norm will be denoted *LM-GB* features. We compared the performance of these features to Gabor features that will be referred to as *GB* features and to features that combine steps 1 and 2 and that will be denoted *L-GB* features.

The face transformation model was trained on the FAFB data only. Hence, no information on illumination variations could be learned at training time. The transformation model was trained as described in [11] up to 8 Gaussians per mixture (Gpm). Figure 1 shows the results.

Averaging the performance over the 3 subsets, the identification rate is 68.0% for GB features compared to 74.0% for L-GB features and 84.8% for the LM-GB features. Note

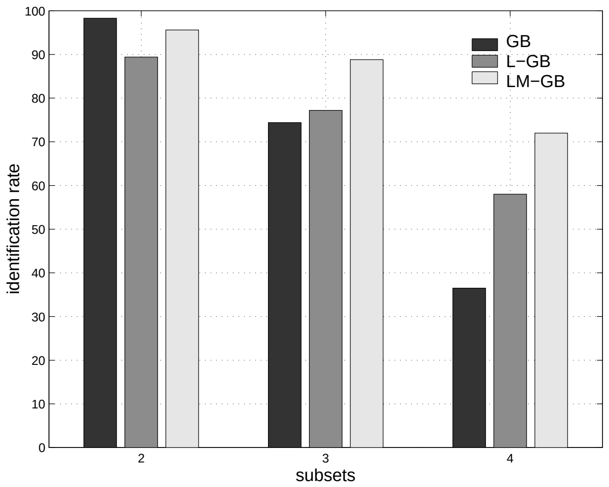

**Fig. 1**. Performance of GB (Gabor), L-GB (log + Gabor) and LM-GB (log + Gabor + mean normalization) features when the transformation model is trained solely on FAFB.

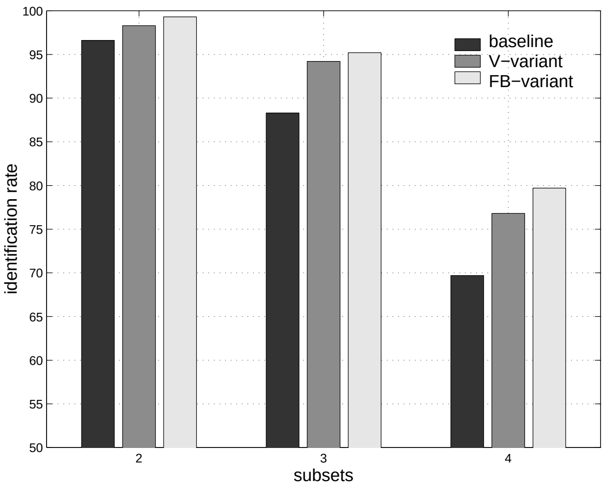

**Fig. 2**. Performance of the baseline system compared to the Viterbi and forward-backward variants (resp. V- and FB-variant) of our novel illumination compensation algorithm.

that with L-GB features the performance decreases significantly compared to GB features on the simple $\mathcal{S}_2$ subset which seems to indicate that the log transform has a negative impact on the recognition when there is little illumination variation.

We performed similar tests (not shown in this paper) with the GB, L-GB and LM-GB features on the popular Eigenfaces [15] and Fisherfaces [8] algorithms and observed similar trends. We would like to underline that, although we tested the combination of Gabor features and LM-Norm, we believe that LM-Norm could benefit to other "linear" features such as DCT features.

### 5.4. Performance of our novel approach

The goal of this second set of experiments is not only to assess the performance of our novel model of illumination transformation but also to assess the performance of the simple approach discussed in section 3.1, which is based solely on the transformation model introduced in [11] and which does not make use of any feature transformation. The latter algorithm will be referred to as the baseline.

For both algorithms, we applied a logarithm transform in the pixel domain prior to the extraction of Gabor features (L-GB features) as both methods require the illumination to be an additive term in the feature domain.

For our novel approach, we first trained our system up to 8 Gpm using only the FAFB data as explained in [11]. Then, using this model, we trained the covariance matrix $S$, which is the only parameter of the illumination transformation model, on the FAFC data only. The assumption is that, as the transformation model trained on FAFB already accounted for variations due to facial expressions, all the variability that remained unexplained was due to illumination. The diagonal elements of $S$ were initialized to values close to 0 and then, 3 training iterations were applied. At both training and test time, the number of iterations of the decoding process described in section 4 was set to 3. To find $\Phi_n$ with the forward-backward variant of the algorithm described in section 4.2, we applied 5 horizontal and vertical passes.

For the baseline, we simply trained the system on both the FAFB and FAFC data up to 16 Gpm, instead of 8 Gpm, as more data was available.

Figure 2 shows the performance of the baseline compared to the Viterbi and forward-backward variants of our novel approach (resp. *V-variant* and *FB-variant*). Comparing Figures 1 and 2, one can see that adding the FAFC data increases on the average the identification rate of the baseline system from 74.0% to 84.1%. However, both variants of our novel approach clearly outperform the baseline, especially for the harder $\mathcal{S}_3$ and $\mathcal{S}_4$ subsets.

It is also interesting to notice that the FB-variant outperforms the V-variant. Actually, the latter one is optimal in the *Maximum-Likelihood* framework while our modified forward-backward based on the T-HMM framework is not guaranteed to be optimal. However, while Viterbi only takes into account the best path, i.e. the one that best explains the data, the forward-backward algorithm integrates over all paths. As explained in 4.2.2, this choice has an impact on the re-estimation of $S$ and we believe that the difference in performance is mainly due to the difference in the re-estimation formula. The average identification rate of the V-variant and FB-variant over the three subsets are respectively 89.1% and 90.8%.

We also compared our novel approach with the eigenfaces [15] and Fisherfaces [8]. Especially Fisherfaces were shown to compensate for illumination variations if trained with the appropriate data. To carry out a fair comparison, we did not apply these algorithms directly on the gray level images but on their LM-GB representations. A feature vector was extracted every four pixels of the images in both horizontal and vertical directions. The eigen- and Fisher-spaces were trained on the FAFB and FAFC sets as was done for our baseline system. The best identification rates we obtained for eigenfaces and Fisherfaces are respectively 87.1% and 83.1%. The fact that eigenfaces outperform Fisherfaces is not surprising considering the small number of training observations per class and the mismatch between training and test conditions [16].

Finally, we would like to stress the fact that our novel algorithm is very efficient as it takes on the average to our best system less than 25 ms to compare two images on a 2 GHz Pentium 4 with 1 GB RAM.

## 6. Conclusion and future work

In this paper, we introduced a novel approach to illumination compensation, which consists in modeling the set of possible illumination transformations between face images of the same person. This approach is naturally embedded in a face recognition system which already models transformations between face images due to facial expressions. We showed experimentally that, even in the challenging case where we trained and tested our system on two different databases, our novel approach to illumination compensation resulted in large improvements of the recognition rate. Note that our results are competitive with state of the art results recently published on the YALE B database [7].

However, much work remains to be done to perfectly compensate for illumination variations. For the challenging $\mathcal{S}_4$ subset, the best identification rate we obtain is close to 80%. Although this corresponds to an almost 70% relative error rate reduction compared to the same system without any illumination compensation, we are still far from the almost perfect recognition rate we get for the simpler $\mathcal{S}_2$ subset. We believe that one limitation of our current approach is the fact that the covariance matrix $S$ in our illumination transformation model is fixed for all pairs of images. We think that $S$ should incorporate both some a priori knowledge learned off-line through a training phase, as is currently the case, but also some information which is dependent on the pairs of images that need to be compared.

Finally, we would like to point out that, while our model of illumination compensation has been introduced in the context of face recognition, it could benefit to other research areas. As our original approach to face recognition has a lot in common with motion estimation algorithms, and especially MAP estimation of dense motion [4], we think that our approach could be applied to the difficult problem of motion estimation in the presence of illumination variations.

## 7. References

[1] J. Schürmann, *Pattern classification, a unified view of statistical and neural approaches*, John Wiley & Sons, Inc., 1996.

[2] 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.

[3] D. M. Blackburn, M. Bone and P. J. Phillips, “Face recognition vendor test 2000: evaluation report,” Tech. Rep., 2001.

[4] A. Bovik, *Handbook of image and video processing*, Academic Press, 2000.

[5] Y. Adini, Y. Moses and S. Ullman, “Face recognition: the problem of compensating for changes in illumination direction,” *IEEE Trans. on PAMI*, vol. 19, no. 7, pp. 721–732, July 1997.

[6] M. Savvides and V. Kumar, “Illumination normalization using logarithm transforms for face authentication,” in *IAPR AVBPA*, 2003, pp. 549–556.

[7] R. Gross and V. Brajovic, “An image preprocessing algorithm for illumination invariant face recognition,” in *IAPR AVBPA*, 2003, pp. 10–18.

[8] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, “Eigenfaces vs. fisherfaces: recognition using class specific linear projection,” *IEEE Trans. on PAMI*, vol. 19, no. 7, pp. 549–556, July 1997.

[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.

[10] V. Blanz, S. Romdhani and T. Vetter, “Face identification across different poses and illuminations with a 3d morphable model,” in *IEEE AFGR*, 2002.

[11] F. Perronnin, J.-L. Dugelay and K. Rose, “Deformable face mapping for person identification,” in *IEEE ICIP*, 2003.

[12] 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.

[13] B. K. P. Horn, *Robot Vision*, Mc Graw-Hill, New-York, 1986.

[14] B. Duc, S. Fischer and J. Bigün, “Face authentication with gabor information on deformable graphs,” *IEEE Trans. on PAMI*, vol. 8, no. 4, pp. 504–516, April 1999.

[15] M. A. Turk and A. P. Pentland, “Face recognition using eigenfaces,” in *IEEE CVPR*, 1991, pp. 586–591.

[16] A. M. Martìnez and A. C. Kak, “Pca versus lda,” *IEEE Trans. on PAMI*, vol. 23, no. 2, pp. 228–233, Feb 2001.