Un Modèle Probabiliste de Transformation entre Images Appliqué à la Reconnaissance de Visages

Florent Perronnin * Jean-Luc Dugelay

Institut Eurécom, Multimedia Communications Department
BP 193, 06904, Sophia Antipolis Cedex, France

{florent.perronnin, jean-luc.dugelay}@eurecom.fr

Résumé

Nous introduisons dans cet article une nouvelle mesure de "distance" entre images de visages qui nécessite d'estimer l'ensemble des transformations possibles entre visages d'une même personne. La transformation globale, que nous supposons trop complexe pour être modélisée directement, est approximée par un ensemble de transformations locales sous la contrainte que des transformations voisines doivent rester cohérentes entre elles. Transformations locales et contraintes de voisinage sont incorporées dans le cadre probabiliste d'un modèle de Markov caché bidimensionel. Nous nous appliquons plus particulièrement à modéliser deux sources de variabilité : les expressions faciales et les changements d'illumination. Nous montrons à l'aide d'un corpus constitué de 3 bases de données que la méthode proposée est extrêmement compétititve par rapport à une des méthodes les plus significatives de l'état de l'art.

Mots clefs

Biométrie, reconnaissance de visages, reconnaissance des formes, traitement d'images, modèle de Markov caché.

1 Introduction

La biométrie consiste à déterminer l'identité d'un individu à partir de ses caractéristiques physiques (visage, géométrie de la main, empreintes digitales, etc.) ou comportementales (parole, signature, etc.). Un système biométrique est donc un système typique de reconnaissance des formes. En mode identification, étant donné une liste d'identités possibles, le but est d'associer à une observation l'identité la plus plausible. En mode vérification, étant donné une identité, le but est de déterminer si l'observation correspond oui ou non à l'identité proposée.

Dans cet article, nous nous intéressons à la reconnaissance de visages. C'est un problème extrêmement difficile car les visages de personnes différentes ont globalement la même forme alors que les images d'un même visage peuvent fortement varier du fait des expressions faciales, des conditions d'illumination, de la pose, de la présence ou de l'absence de lunettes, moustaches ou barbes, etc. Bien que d'énormes progrès aient été accomplis au cours de ces 30 dernières années, la reconnaissance de visages n'est pas encore un problème résolu comme l'ont montré les évaluations conduites par NIST [1, 2].

Pour les systèmes biométriques, le manque de données lors de l'enrôlement est un problème crucial. Généralement, lorsqu'un nouvel utilisateur est enrôlé dans un système de reconnaissance de visages, seules quelques images sont acquises de manière à limiter la durée de l'enrôlement et donc la gêne pour l'utilisateur. Il est alors difficile d'estimer de manière robuste un modèle de variabilité à partir des observations disponibles. Dans le cas où une seule observation est disponible, la variabilité est même impossible à estimer. En général, l'image d'enrôlement est alors directement utilisée comme modèle et le score de reconnaissance est une distance entre les images d'enrôlement et de test. Toute la difficulté consiste donc à définir une distance pertinente.

Pour définir une telle distance, il convient de formaliser la relation qui existe entre observations d'une même classe, c'est-à-dire, entre les images de visage d'une même personne. Soient I_t une image d'enrôlement ("template") et I_q une image de test ("query"). Si $\mathcal{R}$ est la relation entre images d'une même personne, alors notre mesure de distance peut s'exprimer de la manière suivante :

$$P(I_q|I_t, \mathcal{R}) \quad (1)$$

Dans cet article, nous considérons une nouvelle distance probabiliste entre images. Cette distance nécessite d'estimer l'ensemble des transformations possibles entre images d'une même personne. La transformation globale étant trop difficile à modéliser directement, nous l'approximons à l'aide d'un ensemble de transformations locales sous la contrainte que des transformations locales doivent rester cohérentes entre elles. Les transformations locales et les contraintes de cohérence sont incorporées dans le cadre probabiliste des modèles de Markov cachés bi-dimensionnels (MMC 2-D). Les états de notre MMC sont les transformations locales autorisées. Les probabilités d'émission modélisent le coût de la mise en correspondance d'une région dans l'image d'enrôlement avec une autre région dans l'image de test et les probabilités de tran-

*Ces travaux sont en partie soutenus par France Telecom R & D.

sition mettent en relation les états de régions voisines et modélisent le coût des contraintes de cohérence.
Pour pouvoir estimer les paramètres du MMC, et donc de notre distance, de manière robuste nous supposons que la variabilité intra-classe est la même pour toutes les classes ce qui revient à utiliser la même distance pour tous les individus. Il est ainsi possible d'entraîner notre système avec des individus autres que les utilisateurs ce qui simplifie la phase d'enrôlement.
Nous proposons d'appliquer ce cadre théorique général à deux types de variabilités : les expressions faciales et l'illumination. Pour modéliser les déformations élastiques causées par les expressions, nous utilisons des transformations géométriques discrètes (section 2). Pour modéliser les variations d'illumination, nous utilisons des transformations continues sur les vecteurs caractéristiques extraits de l'image (section 3). Des expériences en mode identification conduites sur les bases de données FERET [1], PIE [3] et ARDB [4] ont montré que notre distance était particulièrement robuste aux variations d'expression, de pose et d'illumination et que les résultats obtenus étaient compétitifs avec l'une des méthode les plus significatives de l'état de l'art appelée BIC [5].

2 Expressions faciales

Les déformations élastiques causées par les expressions faciales peuvent être modélisées à l'aide de transformations géométriques discrètes. Nous restreignons l'ensemble des transformations géométriques possibles aux translations dans la mesure ou une transformation affine globale de faible amplitude peut être approximée à l'aide d'un ensemble de translations locales.

2.1 Transformations géométriques

Supposons que les vecteurs d'observation sont extraits de l'image I_q sur une grille. Soient o_{ij} l'observation extraite à la position (i, j) et q_{ij} l'état associé. Si τ est un vecteur de translation, la probabilité d'émission, c'est-à-dire, la probabilité qu'à la position (i, j) le MMC émette l'observation o_{ij} sachant qu'il est dans l'état $q_{ij} = \tau$, est notée $b_{ij}^{\tau} = P(o_{ij}|q_{ij} = \tau, \lambda_{\mathcal{R}})$, où $\lambda_{\mathcal{R}}$ est l'ensemble des paramètres de notre modèle de relation $\mathcal{R}$ entre images d'une même personne.
Une translation τ met en correspondance le vecteur o_{ij} dans I_q avec le vecteur noté m_{ij}^{τ} dans I_t et la probabilité d'émission modélise le coût de la mise en correspondance de o_{ij} avec m_{ij}^{τ} (c.f. figure 2.1). Nous modélisons b_{ij}^{τ} à l'aide d'une mixture de Gaussiennes. Ce choix est motivé par le fait qu'une combinaison linéaire de Gaussiennes peut approximer n'importe quelle densité de probabilité :

$$b_{ij}^{\tau} = \sum_k w_{ij}^k b_{ij}^{\tau k} \tag{2}$$

Les $b_{ij}^{\tau k}$ sont les composantes et les w_{ij}^k sont les poids des composantes. Chaque composante $b_{ij}^{\tau k}$ est une Gaussienne de moyenne $\mu_{ij}^{\tau k}$ et de covariance Σ_{ij}^k. Nous écrivons $\mu_{ij}^{\tau k}$

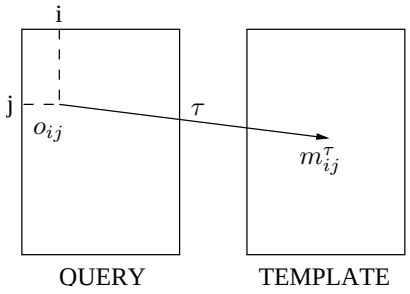

Figure 1 – *Mise en correspondance locale.*

comme une fonction de m_{ij}^{τ} : $\mu_{ij}^{\tau k} = f_{ij}^k(m_{ij}^{\tau})$. Dans [6], nous nous sommes limités aux variabilités additives. Il est intéressant de donner une interprétation graphique des précédentes équations : alors que la forme de b_{ij}^{τ}, qui dépend de w_{ij}^k, de Σ_{ij}^k et des paramètres de f_{ij}^k, est la même pour toutes les personnes, sa moyenne qui est approximativement égale à m_{ij}^{τ} est dépendante de la personne à reconnaître. Intuitivement, b_{ij}^{τ} modélise la variabilité intra-classe autour de la position (i, j).

2.2 Cohérence de voisinage

La cohérence de voisinage est assurée par les probabilités de transition du MMC. Si celui-ci est Markovien du premier ordre, la probabilité de transition est de la forme $P(q_{ij}|q_{i,j-1}, q_{i-1,j})$ ce qui permet de prendre en compte à la fois la corrélation horizontale et verticale. Cependant, la complexité du MMC 2-D étant exponentielle dans la taille des données, nous l'approximons à l'aide d'un ensemble de MMC 1-D horizontaux et verticaux comme expliqué dans la section suivante. Les probabilités de transition horizontales et verticales correspondantes sont : $a_{ij}^{\mathcal{H}}(\tau';\tau) = P(q_{ij} = \tau'|q_{i,j-1} = \tau)$ et $a_{ij}^{\mathcal{V}}(\tau';\tau) = P(q_{ij} = \tau'|q_{i-1,j} = \tau)$. Pour rendre notre distance invariante aux translations globales nous imposons :

$$a_{ij}^{\mathcal{H}}(\tau+\delta\tau;\tau) = a_{ij}^{\mathcal{H}}(\delta\tau) \qquad a_{ij}^{\mathcal{V}}(\tau+\delta\tau;\tau) = a_{ij}^{\mathcal{V}}(\delta\tau) \tag{3}$$

2.3 Turbo MMC

Alors que les MMC 1-D ont été appliqué avec succès à des problèmes uni-dimensionnels, la complexité de leur extension à deux dimensions croît de manière exponentielle avec la taille des données ce qui les rend inutilisables en pratique. Nous avons introduit dans [7] les Turbo MMC (T-MMC) en référence aux codes correcteurs d'erreur. Un T-MMC consiste en un ensemble de MMC 1-D horizontaux et verticaux qui communiquent au travers d'un processus itératif. Les T-MMCs fournissent des algorithmes efficaces pour résoudre les deux problèmes suivants :
– L'estimation de $P(I_q|I_t, \mathcal{R})$, notre mesure de distance probabiliste.
– L'estimation des paramètres $\lambda_{\mathcal{R}}$, à savoir les w_{ij}^k, Σ_{ij}^k, les paramètres de f_{ij}^k ainsi que les $a_{ij}^{\mathcal{H}}$ et $a_{ij}^{\mathcal{V}}$.
Pour plus de détails, le lecteur peut se référer à [7].

3 Conditions d'illumination

De toute évidence, les transformations géométriques précédemment introduites ne sont d'aucune utilité pour modéliser les variations d'illumination. L'idée est donc d'introduire un nouveau type de transformations (sur les vecteurs caractéristiques), et donc d'états, et d'assurer la cohérence entre états voisins.

3.1 Transformations photométriques

Le point de départ pour modéliser l'illumination est l'hypothèse qu'une image I peut être séparée en un produit de deux termes : sa réflectance R et son illumination L [8]. En appliquant l'opérateur logarithme nous obtenons donc :

$$\log I = \log R + \log L \tag{4}$$

et l'illumination devient additive dans le domaine des pixels. En supposant maintenant que l'opérateur d'extraction des paramètres $\mathcal{F}$ est linéaire, nous obtenons :

$$\mathcal{F}\{\log I\} = \mathcal{F}\{\log R\} + \mathcal{F}\{\log L\} \tag{5}$$

et l'illumination reste additive dans le domaine des vecteurs caractéristiques. Si nous introduisons des états "photométriques", nos états sont maintenant doublement indexés : $q_{ij} = (\tau_{ij}, \phi_{ij})$. τ_{ij} et ϕ_{ij} sont respectivement les états "géométriqes" et "photométriques" du système. La probabilité d'émission notée $b_{ij}^{\tau\phi}$ est toujours modélisée à l'aide d'une mixture de Gaussiennes :

$$b_{ij}^{\tau\phi} = \sum_k w_{ij}^k b_{ij}^{\tau\phi k} \tag{6}$$

où les $b_{ij}^{\tau\phi k}$ sont des Gaussiennes de moyenne $\mu_{ij}^{\tau\phi k}$ et de covariance Σ_{ij}^k. Si l'état "photométrique" ϕ représente aussi la contribution additive dans le domaine des vecteurs caractéristiques, nous obtenons :

$$\mu_{ij}^{\tau\phi k} = \mu_{ij}^{\tau k} + \phi = f_{ij}^k(m_{ij}^\tau) + \phi \tag{7}$$

3.2 Cohérence de voisinage

Nous supposons que les probabilités de transition sont séparables de la manière suivante :

$$P(q_{ij}|q_{i,j-1}) = P(\tau_{ij}|\tau_{i,j-1}) \times P(\phi_{ij}|\phi_{i,j-1}) \tag{8}$$
$$P(q_{ij}|q_{i-1,j}) = P(\tau_{ij}|\tau_{i-1,j}) \times P(\phi_{ij}|\phi_{i-1,j}) \tag{9}$$

Alors que le choix d'un nombre discret de transformations géométriques est naturel de par la nature discrète de la grille d'extraction des vecteurs caractéristiques, il est plus simple de traiter l'illumination comme une variable continue. Nous choisissons le modèle suivant pour contraindre la variation d'illumination :

$$P(\phi_{ij} = \phi|\phi_{i,j-1} = \phi') = P(\phi_{ij} = \phi|\phi_{i-1,j} = \phi') = \frac{\exp\left\{-\frac{1}{2}(\phi - \phi')^T S^{(-1)}(\phi - \phi')\right\}}{(2\pi)^{\frac{D}{2}}|S|^{\frac{1}{2}}} \tag{10}$$

La matrice de covariance S est le seul paramètre de notre modèle de variation d'illumination. Si S est diagonale et si les vecteurs caractéristiques représentent une information fréquentielle, alors chaque élément de la diagonale modélise la rapidité de la variation de l'illumination dans la bande de fréquence correspondante.

3.3 Turbo MMC à états continus

Les MMC à états continus sont généralement appelés modèles états-espace. L'acroissement de la complexité du passage à deux dimensions que connaît le MMC à états discrets existe aussi pour le MMC à états continus. Nous avons donc étendu dans [9] le T-MMC aux états continus. Ceci nous permet de résoudre les deux problèmes suivants :
– Trouver la meilleure séquence d'états d'illumination.
– Estimer les paramètres de la matrice de covariance S.
Pour plus de détails, le lecteur peut se référer à [9, 10].

4 Validation expérimentale

Nous allons comparer l'approche proposée (notée par la suite PMLT pour "Probabilistic Mapping with Local Transforms") avec l'approche dite BIC (Bayesian Inter/Intrapersonal Criterion) [5], qui est l'un des algorithmes de reconnaissance de visage les plus compétitifs [1]. Nous nous attacherons notamment à évaluer la robustesse de ces deux approches vis à vis des variations d'expression et d'illumination, mais aussi de pose.

Nous avons utilisé 3 bases de données pour réaliser nos expériences :
– La base de données FERET [1], qui contient 1,199 individus, sert de base d'entrainement aussi bien pour PMLT que pour BIC.
– La base de données AR [4], qui contient 126 individus, permet notamment d'évaluer l'impact des variations d'expression et d'illumination.
– La base de données PIE [3], qui contient 68 individus, permet notamment d'évaluer la robustesse vis-à-vis des variations de pose et d'illumination.

Pour toutes les images les coordonnées des yeux et de la bouche ont été localisées manuellement de manière à extraire des images normalisées de taille 128×128.

Nous avons utilisé la même paramétrisation pour PMLT et BIC puisque notre but est de comparer ces deux classificateurs. Les étapes de la paramétrisation sont les suivantes. Tout d'abord l'opérateur logarithme est appliqué sur l'image de manière à transformer l'illumination en une variabilité additive. Ensuite, les vecteurs caractérisitiques sont extraits en chaque point d'une grille à l'aide d'un banc de filtres de Gabor (4 échelles et 6 orientations) de manière à obtenir une représentation fréquentielle locale de l'image.

Toutes nos expériences ont été menées en mode identification. Nous avons fait pour chacune d'elle un test de McNemar afin de déterminer si la différence de performance observée entre BIC et PMLT était significative. Si la différence est significative avec plus de 95% de confiance, alors le score du meilleur algorithme est mis en gras.

4.1 Expressions faciales

Pour évaluer la robustesse de BIC et PMLT vis-à-vis des expressions faciales, nous avons mené des expériences sur la base de données AR. Les images correspondant à l'expression "neutre" sont utilisées pour l'enrôlement et les images correspondant aux expressions "sourire", "colère" et "cri" sont utilisées comme images de test. Les résultats sont présentés dans le tableau 1.

		BIC	PMLT
AR	sourire	94%	**99%**
	colère	86%	**98%**
	cri	56%	**71%**

Tableau 1 – *Robustesse vis-à-vis des expressions faciales.*

4.2 Conditions d'illumination

Pour évaluer la robustesse de BIC et PMLT vis-à-vis des conditions d'illumination, nous avons mené des expériences sur les bases de données AR et PIE. Pour AR, la même image que précédemment a été utilisée pour l'enrôlement. Il y a deux types de conditions d'illumination pour les images de test : avec une lumière allumée (du côté droit ou gauche) ou avec les deux lumières allumées. Pour la base de donnée PIE, l'image d'enrôlement correspond au visage éclairé par une lumière ambiante. Les images de test correspondent au visage éclairé par un flash, avec ou sans lumière ambiante. Les résultats sont présentés dans le tableau 2.

		BIC	PMLT
AR	lumière g. ou d.	94%	**100%**
	lumières g. et d.	48%	54%
PIE	avec lum. amb.	100%	100%
	sans lum. amb.	54%	**65%**

Tableau 2 – *Robustesse vis-à-vis de l'illumination.*

4.3 Pose

Pour évaluer la robustesse de BIC et PMLT vis-à-vis de la pose, nous avons mené des expériences sur la base de données PIE. L'image d'enrôlement est la même que celle utilisée précédemment. Les images de test correspondent respectivement à des rotations de la tête de 15^o en haut ou en bas, ou bien de 22^o et 45^o à droite ou à gauche. Les résultats sont présentés dans le tableau 3.

		BIC	PMLT
PIE	$\pm 15^o$ h./b.	69%	**94%**
	$\pm 22^o$ d./g.	70%	**90%**
	$\pm 45^o$ d./g.	32%	**57%**

Tableau 3 – *Robustesse vis-à-vis de la pose.*

5 Conclusion

Nous avons introduit dans cet article une nouvelle mesure de "distance" entre visages qui nécessite d'estimer l'ensemble des transformations possibles entre images d'une même personne. La transformation globale est approximée par un ensemble de transformations locales sous la contrainte que des transformations voisines doivent rester cohérentes entre elles. Transformations locales et contraintes de voisinages sont incorporées dans le cadre probabiliste d'un modèle de Markov caché bi-dimensionel. Nous nous sommes attachés plus particulièrement à modéliser deux sources de variabilité : les expressions faciales et les changements d'illumination.
Nous avons évalué la robustesse de l'approche proposée vis-à-vis des expressions faciales, des conditions d'illumination et de la pose et nous avons comparé nos résultats avec une approche de l'état de l'art : BIC. Nous avons montré au cours de nos expériences que, sur les bases de données utilisées, l'approche proposée était plus robuste aux variations d'expression et de pose et, dans une moindre mesure, aux conditions d'illumination.

Références

[1] P. Phillips, H. Moon, S. Rizvi, et P. Rauss. The feret evaluation methodology for face recognition algorithms. *IEEE PAMI*, 22(10) :1090–1104, Oct 2000.

[2] M. Bone D. M. Blackburn et P. J. Phillips. Face recognition vendor test 2000 : evaluation report. Rapport technique, 2001.

[3] T. Sim, S. Baker, et M. Bsat. The CMU pose, illumination, and expression (PIE) database. Dans *IEEE AFGR*, 2002.

[4] A. Martìnez et R. Benavente. The AR face database. Rapport technique 24, CVC, 1998.

[5] B. Moghaddam et A. Pentland. Probabilistic visual learning for object representation. *IEEE PAMI*, 19(7) :696–710, 1997.

[6] F. Perronnin, J.-L. Dugelay, et K. Rose. Deformable face mapping for person identification. Dans *IEEE ICIP*, volume 1, pages 661–664, 2003.

[7] F. Perronnin, J.-L. Dugelay, et K. Rose. Iterative decoding of two-dimensional hidden Markov models. Dans *IEEE ICASSP*, volume 3, pages 329–332, 2003.

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

[9] F. Perronnin et J.-L. Dugelay. From turbo hidden Markov models to turbo state-space models. Dans *IEEE ICASSP*, 2004.

[10] F. Perronnin et J.-L. Dugelay. A model of illumination variation for robust face recognition. Dans *MMUA workshop*, pages 157–164, 2003.