

Utilisation de Séquences Vidéo avec Critères de Qualité pour la Reconnaissance Faciale

Aurélien Mayoue, Anouar Mohamed Mellakh, Dianle Zhou, Dijana Petrovska-Delacrétaz et Bernadette Dorizzi

TELECOM & Management SudParis
Département Electronique et Physique
9 rue Charles Fourier, 91011 Evry cedex, France

{Aurelien.Mayoue, Mohamed.Anouar_mellakh, Dianle.Zhou}@it-sudparis.eu
{Dijana.Petrovska, Bernadette.Dorizzi}@it-sudparis.eu

Résumé Cet article présente un système de reconnaissance faciale automatique qui a été développé et évalué dans le cadre de la compétition MBGCv1 – Portal Challenge. Le scénario d'évaluation consiste à comparer une image de référence à une séquence vidéo où la personne marche à travers un portail métallique. Le système exploite les caractéristiques de Gabor pour la reconnaissance faciale. 4 mesures liées à la qualité du visage sont exploitées pour écarter les trames non pertinentes de la séquence vidéo.

Mots clés Biométrie, Reconnaissance faciale, Mesures de qualité, Séquence vidéo.

1 Introduction

Au cours des dernières années, on observe un intérêt grandissant autour de la biométrie. Ainsi, les gouvernements font de plus en plus appel à cette technologie pour sécuriser leurs frontières. L'apparition du passeport biométrique constitue une nouvelle étape dans leurs volontés de contrôler les flux migratoires. En effet, à l'intérieur de ce passeport, se trouve une puce RFID (Radio Frequency Identification) contenant les données numériques (empreintes et photo d'identité) de son possesseur. Ceci devrait ainsi favoriser l'utilisation des systèmes de reconnaissance biométrique basés sur les empreintes et le visage pour vérifier l'identité des personnes aux postes frontières.

Cet article se focalise sur la vérification de l'identité par le visage. La reconnaissance faciale constitue un domaine de recherche très actif du fait du caractère non intrusif et sans contact, voire sans coopération. En effet, contrairement à l'utilisation des empreintes digitales où la personne doit poser ses doigts sur un capteur pour s'authentifier, la vérification d'identité par un système de reconnaissance facial ne nécessite absolument aucun effort de la part de la personne contrôlée. Cette technologie peut donc s'avérer très utile pour réduire par exemple le temps d'attente avant l'embarquement dans les aéroports.

Dans le scénario proposé, les personnes sont filmées pendant leur traversée d'un portique de sécurité. Ensuite, cette séquence vidéo est comparée à une photo de référence (de type photo d'identité) pour vérifier l'identité de la personne. Les données biométriques utilisées pour simuler ce scénario proviennent de l'expérience *Portal Challenge* de l'évaluation *Multiple Biometric*

Grand Challenge version 1 (MBGCv1) organisée par le *National Institute of Standards and Technology* (NIST) au cours de l'année 2008 [23].

Cet article présente pour la première fois les performances obtenues par un système de reconnaissance faciale sur ce nouveau jeu de données. Le système proposé est entièrement automatique et est constitué des modules suivants:

- Détection du visage, des yeux et de la bouche.
- Rejet des trames non pertinentes (pour la reconnaissance faciale) en termes de qualité liée à la netteté, luminosité et résolution du visage. De plus une nouvelle mesure de qualité liée à la précision de la détection des yeux est introduite.
- Reconnaissance faciale basée sur une approche originale exploitant la fusion des réponses d'amplitude et de phase des filtres de Gabor.

Dans la suite de l'article, la section 2 rappelle succinctement les principaux travaux réalisés dans le domaine de la reconnaissance faciale tandis que les sections 3 et 4 sont consacrées à la description des différents modules de notre système. Ensuite, la section 5 définit précisément le scénario d'évaluation et les données biométriques utilisées. La section 6 est dédiée aux résultats expérimentaux et souligne l'influence de l'utilisation de mesures de qualité sur les performances de notre système. Enfin, avant de conclure, les perspectives sont évoquées dans la section 7.

2 Situation par rapport aux travaux précédents

Les premières études dans le domaine de la reconnaissance faciale remontent au début des années 90. Depuis, de nombreuses techniques de reconnaissance de visages à partir d'images fixes ont été proposées [1]. Ainsi, les évaluations FRVT 2006 [2] ont permis d'atteindre un Taux de Faux Rejets (False Rejection Rate - FRR) de 0,01% pour un Taux de Fausses Acceptations (False Acceptance Rate - FAR) de 0,001% lorsque les conditions d'acquisition sont idéales (i.e. illumination contrôlée, visage haute résolution, de face et sans expression). Parmi les méthodes les plus répandues, on peut citer l'Analyse en Composantes Principales (Principal Components Analysis - PCA) [3] ou l'Analyse Discriminante Linéaire (Linear Discriminant Analysis - LDA) [4] comme exemples d'approche globale et la Correspondance Élastique de Graphes (Elastic Graph Matching - EGM) [5] comme exemple d'approche locale. Notre système de reconnaissance faciale exploite les caractéristiques issues des filtres de Gabor qui permettent de modéliser le système visuel humain [6]. Ces caractéristiques ont déjà été utilisées avec succès dans le domaine de la biométrie (empreintes digitales [7], iris [8] et visages [9, 10]). Tandis que la plupart des systèmes de reconnaissance faciale exploitent uniquement la réponse en amplitude des filtres de Gabor, nous proposons ici une approche originale basée sur la fusion de l'amplitude avec la phase. L'apport d'une telle fusion est discuté dans [11, 26] à travers une évaluation comparative réalisée sur FRGCv2 [24] (base biométrique contenant plus de 35,000 images de visages).

Contrairement aux images, les séquences vidéo utilisées pour la reconnaissance faciale sont généralement jugées de faible qualité car elles sont sujettes aux variations de pose, d'expression et d'illumination. En revanche, la redondance des données peut être utilisée pour améliorer les performances de reconnaissance. Ainsi, sont apparues des méthodes utilisant les modèles statistiques pour exploiter l'information provenant de toutes les trames. Un Modèle de Markov Caché (Hidden Markov Models - HMM) construit à partir des caractéristiques issues de la PCA ainsi qu'un Modèle de Mélange de Gaussiennes (Gaussian Mixture Model - GMM) basé sur les ondelettes de Gabor sont proposés respectivement dans [12] et [13]. Toutefois, ces méthodes, appliquées directement à la vidéo, présentent l'inconvénient qu'elles utilisent toutes les trames de

la séquence, y compris les trames de mauvaise qualité (visage non frontal, faible illumination...). Or dans [14], il a été montré que le fait d'écarter les trames de faible qualité d'une séquence permettait d'améliorer les performances de reconnaissance. Pour identifier les trames de bonne (ou mauvaise) qualité, de nombreuses mesures relatives à la qualité du visage ont ainsi été proposées ces dernières années. Les plus utilisées concernent la netteté, la luminosité, la résolution et la rotation du visage [15, 16]. Riopka et Boulton ont établi dans [17] qu'une détection des yeux non précise est une source importante d'erreurs pour la reconnaissance faciale. Ainsi, en plus des mesures de qualité classiques liées à la netteté, luminosité et résolution, nous présentons dans cet article une nouvelle mesure qui nous renseigne sur la précision de notre détection des yeux.

3 Méthode basée sur l'analyse globale du visage par filtrage de Gabor

Cette section décrit les principaux modules de notre système de reconnaissance faciale.

Détection du visage, des yeux et de la bouche: la première étape de notre système consiste à localiser la position du visage au cours de la séquence. Pour cela, nous reprenons la cascade de classificateurs proposée dans [18] et entraînée à partir des filtres de Haar grâce à l'algorithme Adaboost. Ce détecteur de visages est réputé pour être robuste et très rapide. Ensuite, un Modèle d'Apparence Active [19] est appliqué sur la zone de l'image correspondant au visage pour déterminer avec précision les centres des yeux et de la bouche. Au cours de cette étape, les trames des vidéos sont traitées de façon indépendante.

Prétraitement: grâce aux centres des yeux et de la bouche, une normalisation géométrique appliquée à la zone du visage de l'image globale permet de fournir une image réduite (128x128 pixels) du visage dans laquelle la position des yeux ((32,42) et (96,42)) et de la bouche (64,102) est prédéfinie. Ensuite, pour convertir l'image réduite en niveau de gris, nous utilisons la composante Valeur de l'espace de couleurs TSV (Teinte, Saturation, Valeur) qui exprime mieux le niveau d'intensité des couleurs que le système RVB [20]. Finalement, un lissage anisotropique [21] s'avère plus efficace qu'une simple normalisation d'histogramme pour réduire l'influence de la variation d'illumination dans des conditions non contrôlées (cf. Fig. 1).

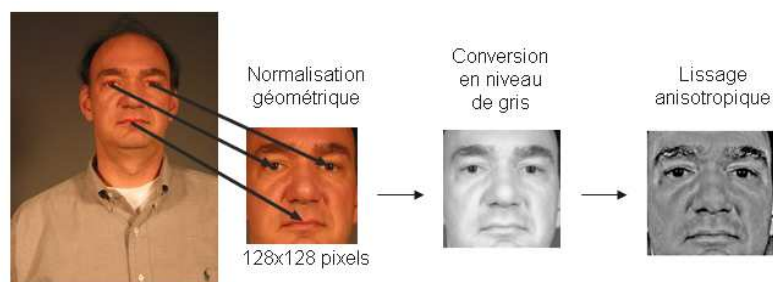


Figure 1 : Illustration des différentes étapes du prétraitement.

Extraction des caractéristiques: l'analyse globale du visage est basée sur la convolution de l'image prétraitée de visage avec une famille de filtres de Gabor caractérisés par différentes orientations et résolutions. Le résultat de convolution est un nombre complexe défini par son amplitude et sa phase pour chaque pixel de l'image (cf. Fig. 2). Pour notre approche, la combinaison de la phase et de l'amplitude est motivée par le fait que l'information utile liée à la texture se situe dans la phase de l'analyse par filtrage de Gabor. En pratique, les réponses en amplitude et en phase de chaque convolution sont sous-échantillonnées, normalisées (centrées et

réduites) puis concaténées en un seul vecteur. Enfin, une réduction de dimensionnalité discriminante des vecteurs caractéristiques est réalisée par LDA (l'algorithme DLDA [22] est utilisé pour l'apprentissage direct de l'espace de réduction).

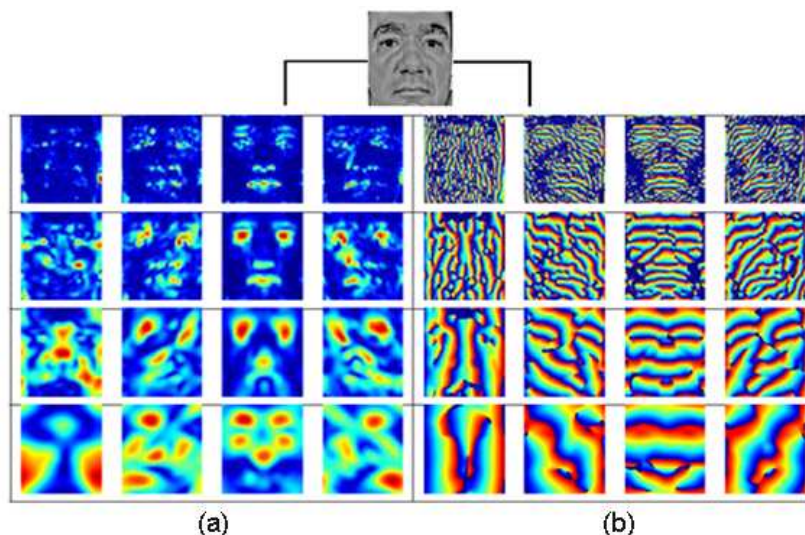


Figure 2 : Résultat de la convolution d'une image avec une famille de filtres de Gabor de 4 orientations (horizontales) et 4 résolutions (verticales). Les réponses en amplitude (a) et en phase (b) sont représentées.

4 Mesures de qualité

Cette section est dédiée aux mesures de qualité et à la stratégie employée pour écarter, au sein d'une séquence vidéo, les trames non pertinentes pour la reconnaissance faciale. En plus des mesures traditionnelles liées à la résolution, luminosité et netteté du visage [15], nous introduisons une nouvelle mesure relative à la précision de notre détection des yeux.

Résolution: le nombre de pixels entre les yeux du visage dans l'image originale nous renseigne sur la précision de l'acquisition. Plus ce nombre est important et plus l'image sera jugée de bonne qualité pour la reconnaissance faciale.

Luminosité: pour estimer la luminosité du visage, nous moyennons la valeur de la composante d'illumination de tous les pixels dans l'image (en niveau de gris) du visage normalisé. Cette mesure permet d'écarter les trames contenant les visages les plus sombres.

Netteté: soient I l'image (en niveau de gris) du visage normalisé et LI cette même image après filtrage passe-bas. La mesure $Q_{\text{netteté}}$ qui estime la netteté du visage est définie par l'Equation 1:

$$Q_{\text{netteté}} = \|I - LI\| \quad (1)$$

Le visage sera d'autant plus net que la valeur de $Q_{\text{netteté}}$ est élevée.

Précision de la détection des yeux: l'idée de cette mesure de qualité est d'écarter les visages mal normalisés (géométriquement) et donc sources d'erreurs potentielles pour la vérification d'identité. Dans notre scénario, les personnes marchent en direction de la caméra de façon régulière au cours du temps (voir Section 5). La distance entre les yeux devrait ainsi augmenter de façon quasi-linéaire au cours de l'enregistrement. Soit $d(n)$ la distance entre les yeux à l'instant n . Par régression linéaire, la fonction $d:n \rightarrow d(n)$ est approximée par une droite d'équation: $d_{\text{lin}}(n) = a.n + b$.

Finalement, la mesure $Q_{\text{précision}}$ relative à la précision de la détection des yeux est définie par l'Equation 2:

$$Q_{\text{précision}}(n) = |d(n) - d_{\text{lin}}(n)| \quad (2)$$

Plus la valeur de $Q_{\text{précision}}(n)$ est importante et plus la qualité de la normalisation du visage correspondant à la trame n est faible (cf. Fig. 3).

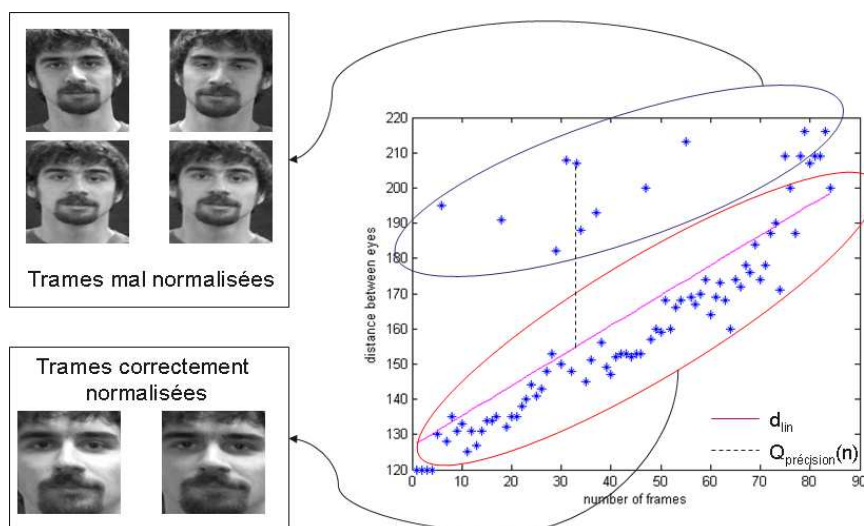


Figure 3 : Exemple pour une séquence vidéo de l'évolution de la distance entre les yeux au cours du temps. Les visages mal normalisés (i.e. dont les yeux ont été mal localisés) peuvent être identifiés grâce à $Q_{\text{précision}}$.

Les quatre mesures de qualités que nous venons de présenter sont utilisées pour écarter au sein d'une séquence vidéo les trames de plus mauvaise qualité. En pratique, pour chaque critère de qualité, 10% des trames sont écartées de la séquence.

5 Scénario d'évaluation et données biométriques

Le scénario étudié dans cet article consiste à comparer une image de référence à une séquence vidéo où la personne marche à travers un portail métallique en direction de la caméra. Une application concrète de ce scénario peut être la vérification d'identité aux postes frontières. Les données biométriques utilisées pour cette expérimentation proviennent de l'évaluation MBGCv1 – Portal Challenge [23] à laquelle notre système a été soumis dans le cadre du projet ANR-07-CSOSG KIVAOU. Cette base contient les données de 111 personnes parmi lesquelles 29 personnes ont participé à deux sessions d'acquisition. Au cours d'une session, une image de référence et une séquence vidéo sont enregistrées (cf. Fig. 4). Ainsi, l'ensemble d'enrôlement est constitué de 140 images de référence tandis que l'ensemble de test contient 140 séquences vidéo. Les images d'enrôlement (1504x1000 ou 998x1500 pixels) ont été acquises dans des conditions d'illumination contrôlée et contiennent exclusivement des visages frontaux avec en moyenne 110 pixels entre les yeux. Les vidéos de test (1080x1440 pixels / frame) durent en moyenne trois secondes. La résolution entre les yeux varie de 50 à plus de 200 pixels au fur et à mesure que la personne se rapproche de l'objectif de la caméra. Le protocole d'évaluation propose 200 tests client et 5592 tests imposteur. A notre connaissance, aucun résultat n'a précédemment été publié sur cette base de

données. Les lecteurs qui souhaitent avoir plus d'information sur les évaluations MBGCv1 ou qui veulent se procurer la base de données peuvent visiter le site web du NIST [23].



Figure 4 : Exemple de données biométriques d'une même session extraites de la base MBGCv1 – Portal Challenge.

6 Conditions d'expérimentation et résultats

Pour l'expérimentation, une famille de filtres de Gabor de 8 orientations et 4 résolutions (résultant en un ensemble de 32 filtres) a été sélectionnée. Les données utilisées pour la construction de l'espace de réduction par DLDA proviennent de l'ensemble de développement de la base FRGCv2 [24]. Notre détecteur de visages exploite le code source d'OpenCV [25]. Au niveau du calcul des scores, la distance cosinus est utilisée pour estimer la similarité entre deux vecteurs caractéristiques. Etant donné que la comparaison entre une image d'enrôlement et une vidéo de test nous fournit n distances, le score final sera le minimum de ces n distances. Nous reportons dans le Tableau 1 les performances de notre système de reconnaissance faciale en termes de Taux d'Erreur (Equal Error Rate - EER) et de Taux de Vérification (Verification Rate - VR) à FAR=0,1%.

Tableau 1 : Performances de vérification faciale sur MBGCv1 – Portal Challenge.

Stratégie	EER	VR à FAR=0,1%
Utilisation de toutes les trames	5,62% [\pm 0,97]	49,65% [\pm 5,82]
Rejet des trames de mauvaise qualité	5,49% [\pm 0,96]	55,66% [\pm 5,78]

Ces résultats mettent en évidence que le fait d'écarter pour chaque critère de qualité 10% des trames les moins pertinentes permet d'améliorer les performances de reconnaissance par rapport à une stratégie qui consisterait à prendre en compte toutes les trames. De plus, l'extraction de caractéristiques n'est pas réalisée sur les trames écartées ce qui a pour conséquence de diminuer considérablement le temps de traitement. En considérant les intervalles de confiance, il est à noter que nos résultats devront être validés sur une base présentant plus de tests client et imposteur. Néanmoins, ces premiers résultats encourageants, obtenus avec des mesures de qualité simples, incitent à explorer davantage la qualité des images pour améliorer les performances des systèmes de reconnaissance faciale.

7 Conclusions et perspectives

Cet article reporte les performances d'un système de reconnaissance faciale entièrement automatique et évalué dans le cadre de la compétition MBGCv1 – Portal Challenge. Notre système

présente l'originalité qu'il exploite conjointement les réponses en amplitude et en phase des filtres de Gabor. Au cours de notre étude, nous avons souligné l'importance des mesures de qualité pour traiter les séquences vidéo et ainsi écarter les trames non pertinentes pour la reconnaissance faciale. Ceci permet à la fois de diminuer les temps de calculs et d'améliorer légèrement les performances de vérification de notre système.

Jusqu'à présent, seules les mesures de qualité liées à la séquence de test ont été exploitées pour écarter les trames de mauvaise qualité. Il semblerait judicieux d'utiliser également la qualité de l'image d'enrôlement. Une stratégie consisterait ainsi à extraire dans chaque séquence les trames qui seraient les plus proches en termes de qualité de l'image d'enrôlement. Ceci permettrait de diminuer les problèmes liés notamment à la variation d'illumination ou à la déformation du visage dû à la normalisation géométrique. En effet, on peut espérer trouver dans chaque séquence vidéo au moins une trame dans laquelle le visage possède une illumination et une résolution proches de celles de l'image d'enrôlement. Afin d'évaluer les performances maximales que notre système actuel pourrait atteindre avec une telle stratégie, nous avons réalisé un post-traitement. Dans un premier temps, nous avons extrait pour chaque séquence la trame qui donne le meilleur score de vérification pour un test client. En principe, il s'agit de la trame qui ressemble le plus à l'image de référence. Dans un deuxième temps, nous avons relancé les tests clients et imposteurs en remplaçant les séquences vidéo par les trames ainsi sélectionnées. Nous obtenons dans ces conditions un Taux d'Erreur de 3,97% et un Taux de Vérification (à FAR=0,1%) de 75,11%. Cela nous donne une bonne idée de la marge de progression qu'on peut espérer atteindre en modifiant notre stratégie actuelle de sélection de trames.

Références

- [1] S.Z. Li and A.K. Jain, Handbook of Face Recognition. *Springer*, 2005.
- [2] J. Phillips, T. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott and M. Sharpe, FRVT 2006 and ICE 2006: Large-Scale Results, March 2007.
<http://www.frvt.org/FRVT2006/docs/FRVT2006andICE2006LargeScaleReport.pdf>
- [3] M. Turk and A. Pentland, Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71-86, 1991.
- [4] D.L. Swets and J.J. Weng, Using Discriminant Eigenfeatures for Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):831-836, 1996.
- [5] L. Wiskott, J.-M. Fellous, N. Krüger and C. von der Malsburg, Face Recognition by Elastic Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775-779, 1997.
- [6] J. Daugman, Uncertainty Relation for Resolution in Space, Spatial Frequency and Orientation Optimized by Two-Dimensional Visual Cortical Filters. *Journal of the Optical Society of America*, 2(7):1160, 1985.
- [7] C.-J. Lee and S-D Wang, Fingerprint Feature Extraction Using Gabor Filters. *Electronics Letters*, 35(4):288-290, 18 February 1999.
- [8] J. Daugman, How Iris Recognition Works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21-30, January 2004.
- [9] M. Zhou and H. Wei, Face Verification Using Gabor Wavelets and Adaboost. In *International Conference on Pattern Recognition*, pages 404-407, 2006.
- [10] L.L. Shen and L. Bai, Gabor Feature Based Face Recognition Using Kernels Methods. In *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 170-175, 2004.

- [11] M. Tistarelli, M. Bicego, J.L. Alba Castro, D.G. Jiménez, M.A. Mellakh, D. Petrovska-Delacrétaz and B. Dorizzi, 2D Face Recognition. In *Guide to Biometric Reference Systems and Performance Evaluation* (ed. D. Petrovska-Delacrétaz, G. Chollet and B. Dorizzi), Springer, 2009.
- [12] X. Liu and T. Chen, Video-Based Face Recognition Using Adaptative Hidden Markov Models. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 340-345, vol.1, 18-20 June 2003.
- [13] J.L. Alba Castro, D.G. Jiménez, E.A. Rúa, E.G. Agulla and E.O. Muras, Pose-corrected Face Processing on Video Sequences for Webcam-based Remote Biometric Authentication. *Journal of Electronic Imaging*, 17, January 2008.
- [14] S-A Berrani and C. Garcia, Enhancing Face Recognition from Video Sequences using Robust Statistics. In *Proceedings of the IEEE International Conference on Video- and Signal-based Surveillance*, Como, Italy, September 2005.
- [15] K. Nasrollahi and T.B. Moeslund, Face Quality Assessment System in Video Sequences. In *Proceedings of the First European Workshop on Biometrics and Identity Management (BIOID 2008)*, Roskilde, Denmark, May 2008.
- [16] E.A. Rúa, J.L. Alba Castro and C.G. Mateo, Quality-based Score Normalization and Frame Selection for Video-based Person Authentication. In *Proceedings of the First European Workshop on Biometrics and Identification Management (BIOID 2008)*, Roskilde, Denmark, May 2008.
- [17] T. Riopka and T. Boulton, The Eyes Have It. In *Proceedings of the 2003 ACM SIGMM Workshop on Biometrics Methods and Applications*, 9-16, 2003.
- [18] R. Lienhart and J. Maydt, An Extended Set of Haar-like Features for Rapid Object Detection. In *IEEE International Conference on Image Processing*, vol.1, pages 900-903, September 2002.
- [19] T. Cootes and C. Taylor, Statistical Models of Appearance for Computer Vision. Technical Report, University of Manchester, March 2004.
- [20] R.C. Gonzalez and R.E. Woods, *Digital Image Processing* (2nd Edition). Prentice Hall, January 2002.
- [21] R. Gross and V. Brajovic, An Image Preprocessing Algorithm for Illumination Invariant Face Recognition. In *4th International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, Springer, June 2003.
- [22] H. Yu and J. Yang, A Direct LDA Algorithm for High-Dimensional Data - with Application to Face Recognition. *Pattern Recognition*, 34(10):2067-2070, 2001.
- [23] Multiple Biometric Grand Challenge (MBGC), <http://face.nist.gov/mbgc/>.
- [24] J. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek, Overview of the Face Recognition Grand Challenge. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, June 2005.
- [25] <http://sourceforge.net/projects/opencvlibrary/>
- [26] M.A. Mellakh, Reconnaissance des visages dans les conditions dégradées. *Thèse de doctorat, Institut National des Télécommunications*, 2009.