Optimisation conjointe de la taille de stockage et des performances de modèles de classification pour l'authentification de visages

Pierre BONAZZA, Johel MITERAN, Dominique GINHAC, Julien DUBOIS

Laboratoire Le2i, Arts et Métiers, FRE CNRS 2005 Univ.Bourgogne Franche-Comté, 9 rue Alain Savary, 21000 Dijon, France
pierre.bonazza@u-bourgogne.fr, miteranj@u-bourgogne.fr, dginhac@u-bourgogne.fr, julien.dubois@u-bourgogne.fr

Résumé – Depuis 2016, afin de minimiser les risques pour la vie privée, la Commission Nationale de l'Informatique et des Libertés (la CNIL) privilégie les dispositifs de sécurité biométrique garantissant à l'usager la maîtrise de ses données personnelles. Ce papier présente une étude d'adéquation algorithme et architecture appliquée au problème de stockage de données utiles à l'authentification de visages sur une carte sans contact de faible capacité. Les solutions proposées utilisent des méthodes standards, alliant les EigenFaces [2] aux Forêts Aléatoires (RF) ou les Machines à Vecteurs Supports (SVM). Appliquée à une large base de donnée, notre méthode montre un couple sensibilité et spécificité atteignant respectivement 96,14% et 99,16% pour les SVM, et 84,90% et 97,99% pour les RF. Différents algorithmes de compression ont été étudiés afin de réduire les données jusqu'à une dizaine de kilos octets.

Abstract – Since 2016, in order to reduce the risks to privacy, the National Commission on Computer Technology and Freedom (the CNIL) recommands systems that guarantee to the user the mastering of his data. This paper deals with an algorithm and architecture matching study, applied to the problem of storing face authentication useful data on a low capacity card. The proposed solutions use standards methods, combining EigenFaces [2] with Random Forest (RF) or Support Vector Machine (SVM). The performances on our dataset shows a couple sensitivity/specificity of respectively 96,14 % and 99,16 % for the SVM, and 84,90 % and 97,99 % for the RF. Various compression algorithms are studied in order to decrease the data to a decade of Kilo Bytes.

1 Introduction

Dans les domaines liés à la sécurité, la biométrie est de nos jours un outil privilégié. On différencie l'authentification, contrôlant que la mesure correspond au gabarit enregistré (modèle d'apprentissage), de l'identification, cherchant une correspondance de la mesure à une des classes enregistrées. Les données biométriques étant liées à la vie privée des usagers, une autorité administrative indépendante française (la CNIL) est chargée de veiller à ce que les systèmes s'en servant soient encadrés, exigeant une justification de leur utilisation et la légitimité, s'il y a lieu, de la conservation des données sur serveurs.

Dans le cadre d'un projet FUI *(Nuc Track - FUI 17)* consistant à apporter à faible coût une solution à la protection du stockage d'éléments sensibles, nous avons été amenés à proposer un dispositif biométrique permettant une authentification du personnel par leur visage. Les contraintes fixées par l'application de ce projet sont de respecter les normes imposées par la CNIL et de stocker le modèle (résultat de l'apprentissage) sur une carte sans contact à capacité très limitée (RFID MIFARE 13,56 MHz, EEPROM 8 Ko). L'enrôlement est également une étape importante qui doit être rapide et sans contraintes pour l'utilisateur. Cet article présente donc une démarche d'adéquation algorithme et architecture dans le sens où nous avons dû optimiser conjointement les performances de classification et la taille du modèle, un point rarement étudié dans la littérature.

Les algorithmes de Deep Learning dépassent les capacités humaines en termes reconnaissance de visages, atteignant jusqu'à 99,47% [8] de précision en environnement non contrôlé. Toutefois ces algorithmes, utilisant une très grande quantité de données, nécessitent à l'heure actuelle un temps d'apprentissage conséquent dépassant souvent l'heure de calcul, ce qui les rend incompatible avec les contraintes exposées. Les Machines à Vecteurs Supports (SVM) ainsi que les Random Forest (RF) ont un apprentissage rapide et utilisent peu de données en entrée, tout en ayant des performances intéressantes avec une précision allant respectivement jusqu'à 92% et 90%, selon leur paramétrage, la complexité des images utilisées et les méthodes d'analyses de données [3, 4].

Ces considérations nous ont conduits à choisir une méthode standard de type Eigenfaces [2], combinée avec un classifieur (SVM ou RF). Toutefois, dans leur version initiale, les EigenFaces sont connus pour un certain nombre de limitations, telles qu'une sensibilité au changement de luminance ou d'orientation de l'éclairage. Comme l'ont montré X. Tan et B. Triggs [9], il est nécessaire d'appliquer des pré-traitements qui seront étudiés dans cet article. Dans la section 2, nous décrirons les protocoles d'enrôlement, d'authentification et d'évaluation des méthodes retenues. La section 3 sera consacrée à l'étude et la présentation des résultats des pré-traitements optimums, et de l'adéquation entre performances de classification et taille des modèles issus de l'apprentissage.

2 Méthode

2.1 Protocoles du système d'authentification

Lors de l'apprentissage (enrôlement) et de l'authentification, la personne autorisée est automatiquement détectée dans la scène grâce à l'algorithme de Viola-Jones [1] appliqué en deux étapes : recherche d'un visage, puis, dans la région correspondante, localisation des yeux, du nez et de la bouche afin de confirmer la détection. Les étapes suivantes sont appliquées si et seulement si un seul visage est présent dans la scène.

Apprentissage : plusieurs images de la même personne sont acquises ; des caractéristiques sont extraites, puis un algorithme de classification est appliqué afin de construire un modèle, représentant cette personne, qui sera enregistré sur la carte en respectant les contraintes de taille spécifiées précédemment. L'ensemble de l'opération ne doit pas excéder quelques minutes.

Authentification : la personne autorisée présente sa carte ; son nom est transmis au système et son modèle à l'algorithme de classification, conjointement à l'image de son visage acquise selon les modalités décrites ci-dessus. Le résultat de la classification permet de valider ou non l'accès à la zone sécurisée.

2.2 Protocoles de mise au point et évaluation

La méthode des EigenFaces utilisée en authentification nécessite la construction préalable d'un espace de visages dans lequel les images à apprendre ou à reconnaître seront projetées [2]. Les images de visages doivent être scindées en trois parties principales et redimensionnées au format 100x100 pixels, en niveaux de gris. Des extraits sont représentés Figure 1.

FIGURE 1 – Bases de visages utilisées *: ORL, ESSEX, LFW, Le2i*

- **La première partie** E compose un espace de visages, ne contenant pas d'images de la personne à authentifier, créé à partir de la base ORL (AT&T FACE DATABASE) [7] composée de 400 images de 40 sujets distincts acquises en laboratoire. Sa taille étant importante on ne peut pas le stocker sur la carte.

- **La deuxième partie** A est constituée des images de la personne autorisée ($n_a^A = 40$ images utilisées pour l'apprentissage et $n_t^A = 100$ pour le test).

- **La troisième partie** N contient des images de personnes non autorisées ($n_a^N = 1500$ images pour l'apprentissage et $n_t^N = 18070$ pour le test). N est créée à partir des bases FACE94 (ESSEX FACE DATABASE) [6] formée de 3060 images de 153 sujets différents acquises en laboratoire ; LABELED FACES IN THE WILD [5] « alignée », composée de 13233 images de 5749 « personnalités » en environnement non contraint ; Le2i construite dans notre laboratoire avec 1777 images de 12 sujets distincts soumis parfois à divers éclairages.

2.3 Attributs et opérateurs de classification

Comme préconisé par M. Turk et A. Pentland [2], nous réduisons la dimension du modèle à environ 90% de l'inertie de ses valeurs propres. Le nombre de valeurs propres conservées représente donc la dimension des vecteurs attributs utilisés à l'entrée du système de classification. Dans cette étude, nous observerons également l'effet d'une réduction du modèle à 70% et 50% de l'inertie de ses valeurs propres. Le système considéré est donc un classifieur supervisé à deux classes, pour lequel un modèle est enregistré par personne à authentifier. Nous avons exploré deux approches de classification : les RF et les SVM. Des expériences préliminaires nous ont montré qu'il était préférable d'équilibrer les classes avant l'apprentissage des RF, les images de la personne autorisée ont donc été dupliquées 38 fois pour obtenir $n_a^A \simeq n_a^N$. Les SVM ont été entraînés avec les quantités définies dans la section précédente, utilisant un noyau RBF en apprentissage automatique pour déterminer les paramètres optimums convenant à nos données.

2.4 Prétraitements

Afin d'améliorer la robustesse des Eigenfaces aux transformations citées dans l'introduction, nous avons, suivant l'étude de X. Tan et B. Triggs [9], évalué l'influence de plusieurs pré-traitements. D'une part en appliquant leur méthode utilisant une correction gamma, un filtrage par différences de gaussiennes et une égalisation de contraste, dans le but de réduire l'influence de l'éclairage. D'autre part, nous avons confronté au précédent traitement un simple filtre de Sobel qui présente l'avantage de mettre en avant les hautes fréquences de l'image tout en étant très rapide au niveau de l'exécution. Un aperçu de ces pré-traitements est donné sur la Figure 2.

FIGURE 2 – Aperçus des filtres *: Sobel et Tan-Triggs*

Il aurait été possible de construire un vecteur d'attributs mêlant l'image originale $I_{Originale}$ et l'image filtrée I_{Filtre} par simple concaténation des deux informations. Toutefois, ceci aurait conduit à une augmentation significative de la taille de stockage du modèle appris, ce que nous cherchons à éviter. Nous avons donc choisi de construire la nouvelle image I_{Fusion} par combinaison linéaire des précédentes :

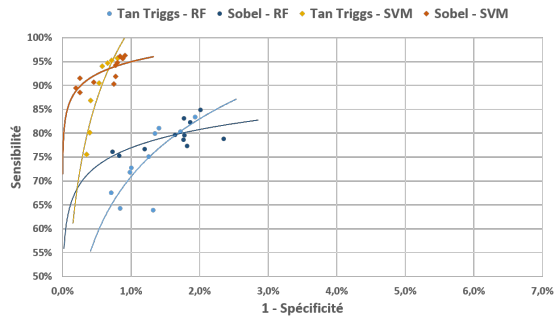
$$I_{Fusion} = \frac{100-\beta}{100} \cdot I_{Originale} + \frac{\beta}{100} \cdot I_{Filtre} \quad | \quad \beta \in [0;100]$$

Afin de rechercher la combinaison optimale nous faisons ensuite varier β entre 0 et 100. Pour chacune de ses valeurs, nous avons évalué les différentes grandeurs caractérisant les performances (spécificité, sensibilité et précision).

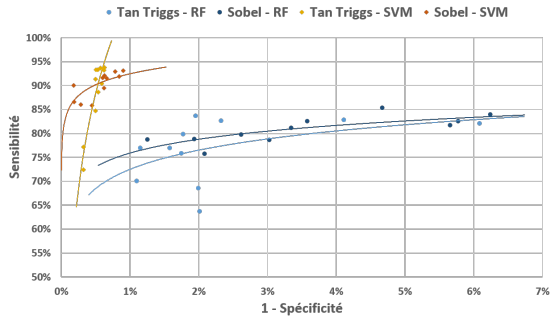
3 Résultats et analyses

3.1 Performances générales de classification

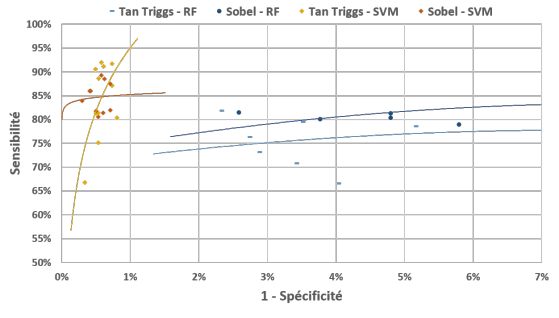
Nous avons procédé à l'enrôlement de 12 personnes avec les classifieurs précédemment cités, selon les différentes combinaisons de filtrages énoncés et suivant le protocole défini dans la section 2.2. Afin de comparer les performances des SVM et des RF sur les mêmes jeux de données, nous avons étudié trois espaces de visages dont nous avons réduit les dimensions, en conservant les vecteurs propres représentant respectivement 90%, 70% et 50% (soit en moyenne 200, 80 et 40 vecteurs propres) de l'inertie des données après ACP (Analyse en Composantes Principales). La Figure 3 expose donc les moyennes de leur sensibilité en fonction de leur spécificité (ainsi que leurs courbes de tendances).



(a) 90% de l'inertie des données après ACP



(b) 70% de l'inertie des données après ACP



(c) 50% de l'inertie des données après ACP

FIGURE 3 – Sensibilité des RF et SVM en fonction de leur spécificité pour les prétraitements énoncés

Les SVM donnent très clairement de meilleurs résultats dans tous les cas. Nous utiliserons par la suite la limitation du nombre de vecteurs propres pour diminuer la taille des modèles résultants de l'apprentissage, tout en optimisant les performances de classifications. Les méthodes présentent un taux de rejet peu élevé des personnes autorisées, pour un taux d'intrusion très faible. Les taux d'intrusion les plus faibles sont obtenus pour les valeurs de β les plus élevées, concordant en contre partie avec les taux de faux rejets les plus élevés.

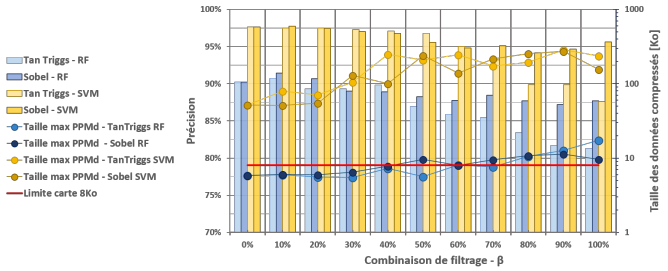
3.2 Minimisation de la taille des modèles

L'espace des visages E occupe quelques dizaines de Mo, quel que soit le filtrage, ce qui ne présente pas d'inconvénient car, étant construit sans images de l'utilisateur, il n'est pas nécessaire de le stocker sur la carte. Selon le pré-traitement et la réduction de dimension de l'espace de visage, les modèles issus de l'apprentissage atteignent des extrémums de 9 Ko et 1200 Ko pour les RF, et de 10 Ko et 1000 Ko pour les SVM. En l'état, aucun ne peut être directement enregistré sur le support de stockage défini (8 Ko). Nous avons donc évalué plusieurs méthodes de compression de données. Les taux de compressions moyens résultants, dépendant directement de la structure des modèles d'apprentissage, sont présentés dans le tableau 1. Même si leurs performances sont voisines, les meilleurs taux de compression sont atteints par les algorithmes BZip2 et PPMd.

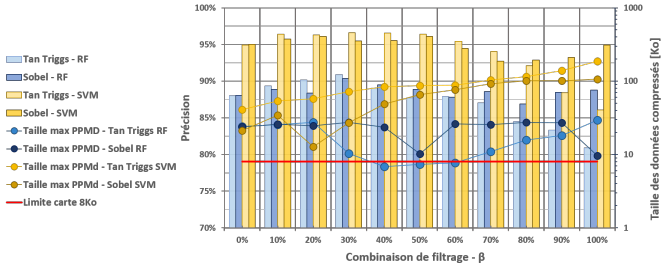
	BZip2	Deflate	Deflate 64	LZMA	LZMA 2	PPMd	Rar
RF	**97,07 %**	95,37 %	95,31 %	95,57 %	95,56 %	**97,15 %**	95,05 %
SVM	**72,56 %**	69,84 %	69,98 %	70,83 %	70,83 %	**72,87 %**	65,71 %

TABLE 1 – Taux de compression selon le type de classifieur

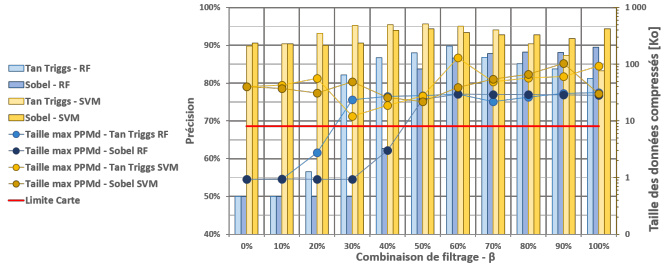
Nous avons représenté sur la Figure 4, les moyennes des précisions et les tailles maximales des modèles RF et SVM compressés, pour les différents types de filtrage, en fonction de β. Cette étude a été réalisée respectivement pour 90% (Figure 4a), 70% (Figure 4b) et 50% (Figure 4c) de l'inertie de l'information après ACP. Pour 90% d'inertie, l'optimum est le RF, avec une précision de 91,45% pour $\beta = 10\%$, le filtre de Sobel et une taille de modèle maximale de 6,5 Ko. Puis pour 70% d'inertie, l'optimum est le RF, avec une précision de 89,03% pour $\beta = 40\%$, le filtre de Tan-Triggs et une taille maximale de 6,9 Ko. Enfin, pour 50% de l'inertie, les performances des RF rejoignent celles des précédents apprentissages au-delà de $\beta = 50\%$ mais aucun ne convient du fait de l'occupation mémoire trop importante. Le compromis optimal des SVM, dans ce cas, est donné par le filtre de Tan-Triggs pour $\beta = 30\%$ donnant une précision de 95,28% pour une taille maximale de 9,8 Ko. Cependant, bien que s'approchant des limites du système, cette combinaison ne répond pas à nos contraintes. Les SVM donnent une précision systématiquement plus élevée, mais nécessitent une taille de stockage systématiquement plus importante que la capacité disponible. Une solution répondant aux contraintes de notre application est donc obtenue par les RF, basés sur 90% de l'inertie. Cette conclusion serait à ajuster dans le cas de l'utilisation d'une carte de capacité supérieure.



(a) 90% de l'inertie des données après ACP



(b) 70% de l'inertie des données après ACP



(c) 50% de l'inertie des données après ACP

FIGURE 4 – Précision des RF et SVM et leurs tailles compressées en fonction des combinaisons de prétraitement

Sur un PC disposant de 4 cœurs cadencés à 3.6 GHz, les temps de calculs sont de l'ordre de la seconde pour l'apprentissage des RF et de la minute pour les SVM. L'étape de vérification ne nécessite que quelques millisecondes. Les temps de compression (utilisée uniquement lors de l'enrôlement) et de décompression (utilisée une fois par passage de carte) ne dépassent pas le dixième de seconde.

4 Conclusion et perspectives

Les deux classifieurs étudiés présentent des performances convaincantes. Cette étude nous a conduit à choisir les modèles issus de l'apprentissage par les RF qui peuvent être stockés sur le support cible, bien qu'ils soient légèrement moins performants que les SVM. Il existe d'autres types de supports personnels à plus haute capacité, conduisant à un système global plus onéreux, et nécessitant une étude de fiabilité et de compatibilité aux normes en vigueurs. La conclusion de cette étude serait à adapter à ces nouveaux supports, la démarche générale restant identique et les résultats de comparaison pouvant être directement exploités.

Afin de pallier aux changements d'apparence du visage dans le temps, la rapidité d'apprentissage et de compression permet une mise à jour régulière du modèle en entraînant un nouveau classifieur avec les images acquises lors de la vérification.

Les perspectives de ces travaux sont d'étudier une nouvelle modalité en combinant des données biométriques acquises dans les domaines visible et infrarouge, afin de renforcer la sécurité tout en restant neutre vis à vis de la vie privée. Nous développons dans ce but, et dans le même esprit que F. Lamare [10] et N. Miura et al. [11], un prototype de système biométrique proposant une méthode d'acquisition simultanée des empreintes digitales et des vaisseaux sanguins.

Remerciements

Ces travaux ont été réalisés au Le2i à l'Université de Bourgogne, dans le cadre du projet Nuc-Track, FUI 17, co-financé par le Conseil Régional de Bourgogne.

Références

[1] P. Viola et M. Jones, *Robust Real-Time Face Detection*, International Journal of Computer Vision (57)(2004) pp.137-154.

[2] M. Turk et A. Pentland, *Eigenfaces for Recognition*, Journal of Cognitive Neuroscience (3)(1991) pp.71-86.

[3] A. Bouzalmat, J. Kharroubi et A. Zarghili *Comparative Study of PCA, IDA, LDA using SVM Classifier*, Journal of Emerging Technologies in Web Intelligence (6)(2014) pp.1275-1286.

[4] A.Z. Kouzani, S. Nahavandi et K. Khoshmanesh, *Face classification by a random forest*, TENCON 2007-2007 IEEE Region 10 Conference (2007) pp.1-4.

[5] G. Huang, M. Ramesh, T. Berg et E. Learned-Miller *Labeled Faces in the Wild : A Database for Studying Face Recognition in Unconstrained Environments*, Workshop on Faces in 'Real-Life' Images : Detection, Alignment, and Recognition (2008).

[6] Essex collection of facial images, `http://cswww.essex.ac.uk/mv/allfaces/index.html`.

[7] ORL collection of facial images, `http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html`.

[8] Y. Sun, X. Wang et X. Tang, *Deeply Learned face Representations Are Sparse, Selective and Robust*, arXiv :1412.1265 (2014).

[9] X. Tan et B. Triggs, *Enhanced Local Texture Features Sets for Face Recognition Under Difficult Lighting Conditions*, Analysis and Modeling of Faces and Gestures (2007) pp.168-182.

[10] F. Lamare, *OCT en phase pour la reconnaissance biométrique par empreintes digitales et sa sécurisation*, Traitement du signal et de l'image. Institut National des Télécommunications (2016).

[11] N. Miura, A. Nagasaka, et T. Miyatake, *Personal identification device and method*, Google Patents (2012).