

● Rapport Final - Projet I ●

Reconnaissance de Visage

Auteurs du Rapport : Fotsing Sikadie Gervais S., Ginel Dorleon, Eliodor Ednolson Guy M. (P21)

Superviseur : HO Tuong Vinh

Module : Reconnaissance de Forme - M2 - IFI

Mots clés : [Détection de visage, Reconnaissance de visage, CNN, Descripteurs binaires, Keras, TensorFlow, Deep Learning, Fisher, Motif local]

I. DESCRIPTION DU PROBLÈME

La reconnaissance visuelle de l'image ou de la vidéo est aujourd'hui un sujet populaire dans la recherche scientifique. Beaucoup de lieux publics ont généralement des caméras de surveillance pour la capture vidéo et ces caméras ont leur valeur significative à des fins de sécurité. C'est largement reconnu que la reconnaissance du visage a joué un rôle important dans le système de surveillance car il n'a pas besoin de la coopération de l'objet. Les avantages réels de la reconnaissance faciale par rapport aux autres domaines (dans la biométrie par exemple) sont l'unicité et l'acceptation. La première étape et la plus importante de la reconnaissance du visage est la détection de visage. En fait, détecter les visages humains et extraire du visage les caractéristiques d'une image sans contrainte sont un processus difficile. Il est très difficile de localiser les positions des visages dans une image avec précision. Le visage humain est un objet tridimensionnel (3-D), et pourrait être sous une perspective déformée et d'une illumination irrégulière. Il existe plusieurs variables qui nuisent les performances de la détection, y compris le port de lunettes, coloration cutanée, genre, poils faciaux et expressions faciales, la variation de la pose (avant, non frontale), l'occlusion, l'orientation de l'image. La détection de visage avec une forte précision est un problème difficile dans la vision par ordinateur et les autres domaines liés ; ce qui fait que la précision et la vitesse de l'identification sont un problème majeur. Le but de cet rapport est de présenter une méthode implémentée pour la reconnaissance de visage, d'évaluer cette méthode en terme d'erreur et de taux de calcul.

II. INTRODUCTION

La reconnaissance de visage est l'une des applications les plus pertinentes de l'analyse d'image. C'est un véritable défi de construire un système automatisé qui équivaut à la capacité humaine pour reconnaître les visages. Bien que les humains identifient assez bien les visages connus, nous ne sommes pas très qualifiés quand nous devons faire face à une grande quantité de visages inconnus. Les ordinateurs, avec une mémoire et une vitesse de calcul presque illimitée, devrait

surmonter les limitations humaines. La reconnaissance du visage est un domaine évolutif, changeant et qui s'améliore constamment. Au cours de ces dernières années, la reconnaissance du visage a attiré beaucoup d'attention et ses recherches ont rapidement augmenté non seulement pour les ingénieurs mais aussi pour les neuro-scientifiques, puisqu'elle comporte de nombreuses applications potentielles dans la communication par vision par ordinateur et le système automatique de contrôle d'accès. En particulier, la détection de visage est une partie importante, première étape, de la reconnaissance du visage. C'est dans cette même ordre d'idée que nous nous sommes proposés d'approfondir dans ce domaine dans le but de comprendre les différents algorithmes, les méthodes anciennes, les travaux réalisés afin d'implémenter au final une solution pouvant démontrer réellement l'importance de la reconnaissance de visage comme sous-domaine de la reconnaissance de formes. Dans ce rapport, nous présentons notre démarche pour la construction d'un système de reconnaissance de visage.

III. DÉFIS DE LA RECONNAISSANCE DE VISAGE

La reconnaissance faciale permet d'effectuer une vérification biométrique dans toutes les situations. Elle fait l'objet de nombreuses recherches pour des applications liées à la sécurité (identification de criminels, vérification de l'identité des voyageurs, contrôle de sécurité dans les entreprises). Pour les applications commerciales, la reconnaissance faciale devrait permettre la sécurisation des transactions en ligne.

Elle repose sur la reconnaissance automatique des caractéristiques des visages (écartement des yeux, des arêtes du nez, des commissures des lèvres, des oreilles, menton, etc.) avec comparaison avec celles déposées au préalable dans une base de données. Le développement de capteurs 3D, des logiciels associés à l'étude des visages en mouvement, aux vues de profil, aux modèles de vieillissement ouvre encore l'espoir de nombreux perfectionnements.

L'efficacité de la reconnaissance faciale dépend de plusieurs facteurs clés, dont la qualité de l'image, des algorithmes d'identification et de la disponibilité de bases de données fiables. L'application la plus porteuse sur le plan commercial semble être celle de l'authentification pour les achats en ligne, à la place de la communication d'un mot

de passe. Mais il demeure encore des réticences à vaincre, notamment celle de protection de la vie privée. [1]

IV. TRAVAUX EXISTANTS

Notre travail se concentre sur la reconnaissance du visage, un problème qui a reçu une attention significative dans la recherche scientifique. Parmi les nombreuses méthodes proposées dans la littérature, nous distinguons ceux qui n'utilisent pas le deep learning, que nous appelons «peu profond», de ceux qui le font, que nous appelons «profond». Les méthodes peu profondes commencent par extraire une représentation de l'image du visage en utilisant des descripteurs d'image locaux fabriqués à la main tels que SIFT, LBPH, HOG [2, 3]; Ensuite, ils agrègent ces descripteurs locaux en un descripteur de visage global en utilisant un mécanisme de mise en commun, par exemple Fisher Vector. Il existe une grande variété de méthodes qui ne peuvent pas être décrites en détail ici. Les méthodes dites profondes se caractérisent par l'utilisation d'un extracteur de fonctions CNN, une fonction apprenable obtenue en composant plusieurs opérateurs linéaires et non linéaires. Un système représentatif de cette classe de méthodes est DeepFace. Cette méthode utilise une CNN profonde formée pour classer les visages utilisant un ensemble de données de 4 millions d'exemples couvrant 4000 identités uniques. Il utilise également une architecture de réseau siamois, où le même CNN est appliqué sur des paires de faces pour obtenir descripteurs qui sont ensuite comparés en utilisant la distance euclidienne. [4,5] À l'aide d'une très grande quantité de données d'entraînement, DeepFace utilise un ensemble de CNN, ainsi comme phase de pré-traitement dans laquelle les images de visage sont alignées sur une pose canonique en utilisant une 3D maquette. Lorsqu'il a été introduit, DeepFace a réalisé les meilleures performances sur les visages étiquetés sur la base Life Face in the Wild (LFW), ainsi que les visages de Youtube in the Wild (YFW).

V. SOLUTION PROPOSÉE : 3 APPROCHES

Dans notre étude nous nous intéressons à la reconnaissance basée sur les images fixes prises en entrée dans notre base de données. Nous présentons trois approches. Une implémentations avec l'algorithme de Fisher, une 2e avec LBPH et la 3e avec les réseaux d neurones convolutifs (CNN). Donc notre implémentation se fait avec deux approches dites "peu-profondes", à savoir FISHER et LBPH, et une approche dite profonde qui est le CNN.

1 - Méthode d'Analyse Discriminante Linéaire de Fisher

La méthode de Fisher trouve une description mathématique des caractéristiques les plus dominantes de l'ensemble de formation dans son ensemble. Cette approche se fait en plusieurs phases :

a) La phase d'apprentissage : C'est la phase où le système apprend le visage à partir d'une ou plusieurs images du

même visage, elle s'effectue en utilisant l'algorithme de Fisher. [6,7]

b) La phase de test : Elle consiste à identifier un visage de la base de test à partir des images qui se trouvent dans la base d'apprentissage.

c) Approche : L'Analyse Discriminante Linéaire de Fisher permet d'effectuer une réduction de la dimension spécifique. Inventée par le grand statisticien Sir R. A. Fisher. Il l'a utilisée avec succès pour classer les fleurs dans son article de 1936. Afin de trouver la combinaison de fonctionnalités qui sépare le mieux entre les classes, l'Analyse Discriminante Linéaire maximise le rapport entre les classes et la dispersion des classes, au lieu de maximiser la dispersion globale. L'idée est simple : les mêmes classes devraient se regrouper ensemble, tandis que les classes différentes sont aussi loin que possible les unes des autres dans la représentation de la dimension inférieure. Contrairement à l'ACP, il permet d'effectuer une véritable séparation de classes. Cette approche cherche les directions qui sont efficaces pour la discrimination entre les données.

d) Descripteurs utilisés : Les combinaisons résultantes peuvent être employées comme classificateur linéaire, ou généralement dans la réduction de caractéristiques avant la classification postérieure. Avec l'Analyse Discriminante de Fisher, chaque visage, qui se compose d'un grand nombre de pixel, est réduit à un plus petit ensemble de combinaisons linéaires avant la classification. La méthode FisherFaceReconigzer permet d'extraire les caractéristique en se basant sur une combinaison linéaire des valeurs de pixel, qui forment un Template. Les combinaisons obtenues en utilisant LDA s'appellent les Fisherfaces.

2- Méthode Motif Binaire Local, LBPH (Local Binary Patterns Histograms)

Dans la section précédente, nous avons présenter la méthode de reconnaissance avec Fisher. Les deux méthodes effectuent la reconnaissance en comparant le visage à reconnaître avec un ensemble de formation de visages connus. Dans l'ensemble de formation, nous fournissons les visages de l'algorithme et racontons à quelle personne ils appartiennent. Lorsque l'algorithme est invité à reconnaître un visage inconnu, il utilise l'ensemble de formation pour faire la reconnaissance.

Cependant, LBPH analyse chaque visage dans l'ensemble de formation séparément et de façon indépendante. Dans LBPH, chaque image est analysée de manière indépendante, tandis que la méthode des propres formes examine l'ensemble de données dans son ensemble. La méthode LBPH est un peu plus simple, dans le sens où nous caractérisons chaque image dans l'ensemble de données localement ; Et lorsqu'une nouvelle image inconnue est fournie, nous effectuons la même analyse et comparons le résultat à chacune des images dans l'ensemble de données. La façon dont nous analysons les images est en caractérisant les motifs locaux dans chaque emplacement de l'image.[8,9]

LBPH - Algorithme utilisé

Sous sa forme la plus simple, le vecteur de caractéristique LBPH, est créé de la manière suivante :

- Diviser la fenêtre examinée dans les cellules (par exemple 16x16 pixels pour chaque cellule).
- Pour chaque pixel dans une cellule, comparez le pixel à chacun de ses 8 voisins (à gauche, à gauche, au milieu, à gauche, à droite, etc.).
- Suivez les pixels le long d'un cercle, c'est-à-dire dans le sens des aiguilles d'une montre ou dans le sens inverse des aiguilles d'une montre.
- Lorsque la valeur du pixel central est supérieure à la valeur du voisin, écrivez "0". Sinon, écrivez "1". Cela donne un nombre binaire à 8 chiffres (qui est généralement converti en décimale pour plus de commodité).
- Calculer l'histogramme, sur la cellule, de la fréquence de chaque "nombre" se produisant (c'est-à-dire chaque combinaison dont les pixels sont plus petits et plus grand que le centre). Cet histogramme peut être considéré comme un vecteur caractéristique à 256 dimensions.
- Normalisez l'histogramme.
- Concaténer (normalisé) les histogrammes de toutes les cellules. Cela donne un vecteur de fonctionnalité pour la fenêtre entière.

3 - Reconnaissance avec les Réseaux de Neurons Convolutionnels (RNC ou CNN). (Keras + TensorFlow)

La deuxième approche proposée est celle de reconnaissance de visage avec les réseaux de neurones convolutifs. L'approche avec CNN a été testée sur la base de données MINST et fournit un taux d'erreur de 0,23% en Février 2012 et 97, 6% de performance. Les réseaux de neurones convolutifs sont basés sur le perceptron multicouche. C'est-à-dire que le CNN est basé sur un classifieur linéaire de type réseau neuronal formel organisé en plusieurs couches au sein desquelles une information circule de la couche d'entrée vers la couche de sortie uniquement ; il s'agit donc d'un réseau de type feedforward. Chaque couche du réseau est alors constituée d'un nombre variable de neurones, les neurones de la couche de sortie correspondant toujours aux sorties du système.[10]

Le réseau de neurones convolutionnels utilisé ici fonctionne comme un réseau de reconstruction ; il réalise une projection non linéaire du visage présenté en entrée sur un sous - espace puis reconstruit un visage de référence choisi au préalable. Dans l'implémentation, nous avons utilisé les bibliothèques Keras et TensorFlow. L'avantage de Keras est qu'elle permet d'envelopper à la fois TensorFlow et Theano.[11]

- Démarche avec le CNN

Notre démarche avec le CNN prend en entrée une image de taille fixe et la passe à travers une succession de couches de convolution C_i , de sous- échantillonnage S_i et de neurones complètement connectés F_i . Chaque couche

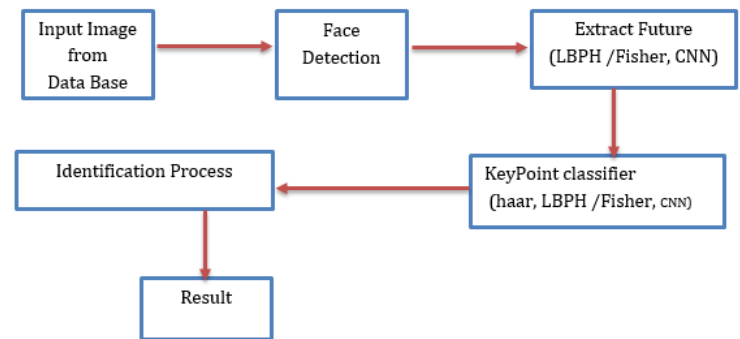
(nous utilisons 2 couches) calcule son opération spécifique (convolution et sous- échantillonnage) avec son vecteur de poids, ajoute ensuite un biais puis passe le résultat dans une fonction d'activation. La sortie du réseau est une image, de même taille que l'entrée, qui est reconstruite par la dernière couche du réseau. Cette démarche a été inspirée par les travaux de Lecun et al. et adaptée à notre problème.[12]

VI. ARCHITECTURE DU SYSTÈME

L'architecture de notre système tourne autour d'un algorithme très simple résumant notre processus de reconnaissance de visage.

Pour chaque image d'entrée, Encodons une image à l'aide de l'algorithme pour créer une version simplifiée de l'image. En utilisant cette image simplifiée, trouvez la partie de l'image qui ressemble le plus au codage générique d'un visage. Découvrez la pose du visage en trouvant les principaux points de repère dans le visage. Une fois que nous trouvons ces repères, utilisez-les pour déformer l'image afin que les yeux et la bouche soient centrés. Passez l'image de visage centrée à travers un réseau neuronal qui sait comment mesurer les caractéristiques du visage. Le système, en regardant tous les visages que nous avons mesurés dans le passé, trouver (le système) quelle personne a les mesures les plus proches des mesures du visage.

Fig1. Architecture du système



VII. OUTILS

Le système a été implémenté sous environnement Linux, (Ubuntu 17.04) sur une machine de 4 processeurs, à l'aide de la plate-forme de développement interactif Spyder d'Anaconda, OpenCV3.2.0, Python version 3.5, Jupyter Notebook, Google Tensorflow API, Keras.

VIII. EXPÉRIMENTATION

- Données utilisées pour l'expérimentation

Nous avons réalisé l'expérimentation de notre système en se basant sur les données réelles de la base " The ORL Data of Faces" [13] Cette base contient 40 sujets distincts dont dix images différentes pour chacun des sujets. Pour

certaines sujets, les images ont été prises à différents moments, variant l'éclairage, les expressions faciales (yeux ouverts / fermés, souriant / pas souriant) et les détails du visage (lunettes / pas de lunettes). Toutes les images ont été prises contre un fond noir homogène avec les sujets en position verticale et frontale (avec une tolérance pour certains mouvements latéraux).

- Apprentissage

Les données d'apprentissage de chaque modèle présenté ici sont constituées de 200 images avec toujours et chaque fois les 5 premières images de chaque sujet.

- Test

On effectue le test de chaque modèle avec 200 images qui sont les 5 dernières images de chaque sujet.

IX. RÉSULTATS ET ANALYSE

Le tableau et le diagramme suivants présentent les différents résultats obtenus.

On constate alors que le CNN avec la combinaison des bibliothèques KerasTensorFlow pour l'implémentation le CNN donne le meilleur taux de précision avec 99%. LBPH donne le 2e meilleur taux de 93.4%.

Fig2. Tableau de performance

Methode	Precision	Erreur
Fisher	86.81%	13.19%
LBPH	93.40%	6.60%
CNN (Keras + TensorFlow)	99.50%	0.50%

Le diagramme suivant présente les résultats, taux de précision et taux d'erreur des 3 méthodes implémentées.

Fig3. Diagramme de performance

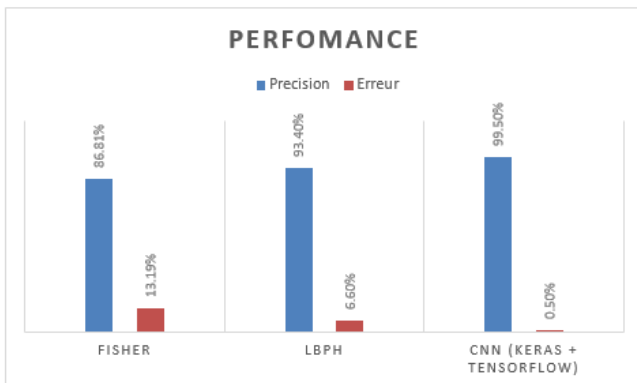


Fig4. Résultat avec Fisher

```

Untitled Last Checkpoint: a few seconds ago (unsaved changes)
View Insert Cell Kernel Widgets Help
114 print ("Taux d'erreur",1-bon/tous)
115 # Show the image and check for ESC being press
116 #cv2.imshow('Reconnaissance', image1)
117 #cv2.waitKey(0)
118 cv2.destroyAllWindows()
6 is Correctly Recognized with confidence 710.8111
6 is Correctly Recognized with confidence 931.8605
7 is Correctly Recognized with confidence 704.1806
7 is Correctly Recognized with confidence 950.1701
7 is Correctly Recognized with confidence 651.8871
7 is Correctly Recognized with confidence 938.3205
7 is Correctly Recognized with confidence 711.4188
8 is Correctly Recognized with confidence 899.5798
8 is Correctly Recognized with confidence 465.1854
8 is Correctly Recognized with confidence 759.4851
8 is Correctly Recognized with confidence 482.8124
8 is Correctly Recognized with confidence 781.4995
9 is Correctly Recognized with confidence 883.6569
9 is Incorrectly Recognized as 40
9 is Incorrectly Recognized as 23
9 is Correctly Recognized with confidence 867.4257
9 is Correctly Recognized with confidence 444.6352
Taux de bon classement 0.8681318681318682
Taux d'erreur 0.13186813186813184
1

```

Fig5. Résultat avec LBPH

```

Untitled Last Checkpoint: 16 hours ago (unsaved changes)
View Insert Cell Kernel Widgets Help
114 print ("Taux d'erreur",1-bon/tous)
115 # Show the image and check for ESC being press
116 #cv2.imshow('Reconnaissance', image1)
117 #cv2.waitKey(0)
118 cv2.destroyAllWindows()
6 is Correctly Recognized with confidence 69.587106398
6 is Correctly Recognized with confidence 80.554937014
7 is Correctly Recognized with confidence 62.612414434
7 is Correctly Recognized with confidence 52.542243714
7 is Correctly Recognized with confidence 53.337307143
7 is Correctly Recognized with confidence 58.896360942
7 is Correctly Recognized with confidence 55.804671065
8 is Correctly Recognized with confidence 63.627442443
8 is Correctly Recognized with confidence 49.570975203
8 is Correctly Recognized with confidence 59.02411644
8 is Correctly Recognized with confidence 47.524764537
8 is Correctly Recognized with confidence 68.284385689
9 is Correctly Recognized with confidence 62.709508508
9 is Correctly Recognized with confidence 61.118471555
9 is Incorrectly Recognized as 3
9 is Correctly Recognized with confidence 55.627143297
9 is Correctly Recognized with confidence 48.509892987
Taux de bon classement 0.9340659340659341
Taux d'erreur 0.06593406593406592
1

```

Fig6. Résultat avec CNN

```

Jupyter Untitled Last Checkpoint: a minute ago (unsaved changes)
Edit View Insert Cell Kernel Widgets Help
Entraînement...
(182, 22500)

Layer (type) Output Shape Param #
-----
dense_7 (Dense) (None, 512) 11520512
activation_7 (Activation) (None, 512) 0
dropout_5 (Dropout) (None, 512) 0
dense_8 (Dense) (None, 512) 262656
activation_8 (Activation) (None, 512) 0
dropout_6 (Dropout) (None, 512) 0
dense_9 (Dense) (None, 40) 28520
activation_9 (Activation) (None, 40) 0
-----
Total params: 11,803,688
Trainable params: 11,803,688
Non-trainable params: 0

Taux de bon classement 0.99509848858
Taux d'erreur 0.00490151142
[ ]

```

X. CONCLUSION

La reconnaissance de visage est un domaine de recherche en pleine expansion et qui propose de multiples applications pratiques. Les méthodes utilisées pour sa mise en œuvre sont diverses et se basent pour certaines directement sur les caractéristiques de l'image et pour d'autre sur les algorithmes d'apprentissage ou encore des approches statistiques. Le présent rapport présente 3 approches de la reconnaissance que nous avons implémentées. L'approche avec CNN donne le meilleur taux de précision de 99%. Il n'existe pas encore de méthodes ou approche parfaite. Fisher a l'avantage d'une bonne classification de différentes classes images. LBPH se fait en divisant l' image en plusieurs petites régions dont les fonctions sont extraites, ce qui donne l'avantage de pouvoir mieux reconnaître les faces par rapport à Fisher. Dans ce rapport, les méthodes de prédiction, l'extraction de descripteurs par Local Binary Patterns, Histogram of Oriented Gradients et Convolutional Neural Networks, tous prévus pour une étude de performances de notre système ont été discutés. Les algorithmes impliqués dans chacun des cas ont été présentés de manière sommaire pour mieux comprendre leur mode de fonctionnement et les défis impliqués. L'ensemble des techniques envisagées ont été choisies dans le but de produire un système de reconnaissance de visage de plus en plus performant, ce qui nécessite d'explorer plusieurs voies d'amélioration qui pourraient être prometteuses ou non selon le cas visé du système de surveillance automatisé. La suite du projet consistera donc à l'élaboration de quelques améliorations ainsi que d'autres solutions qui pourraient être découvertes par la suite

Références

- [1] Évaluation de méthodes de reconnaissance de visages pour l'identification d'individus à partir d'une image de référence Francis C. Migneault, professeur-superviseur : Éric Granger Département de génie de la production automatisée, Laboratoire d'imagerie de vision et d'intelligence artificielle École de Technologie Supérieure, Université du Québec, Montréal, Canada - 2016
- [2] Fontaine, X., Achanta, R., Süsstrunk, S. (2017). Face Recognition in Real-world Images. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (No. EPFL-CONF-224338)
- [3] Lawrence, A., Ashwin, N. M., Manikantan, K. (2017). Face Recognition Using Background Removal Based on Eccentricity and Area Using YCbCr and HSV Color Models. In Proceedings of the International Conference on Signal, Networks, Computing, and Systems (pp. 33-43). Springer India.
- [4] Sun, X., Wu, P., Hoi, S. C. (2017). Face detection using deep learning : An improved faster rcnn approach. arXiv preprint arXiv :1701.08289.
- [5] Y. LeCun, L. Bottou, Y. Bengio, and P. Hader. Gradient-based learning applied to document recognition. In Intel ligent Signal Proerssing, pages 306351. IEEE Press, 2001.
- [6] P. N. Belhumeur, J.P. Hespanha etc, "Eigenfaces vs Fisher faces : r Analysis and Machine Intelligence, vol.20, no. 7, pp. 711-720, 1997.

- [7] M. Gunther et al. The 2013 face recognition evaluation in mobile environment. In " The 6th IAPR International Conference on Biometrics, June 2013.
- [8] A. J. O'Toole, P. J. Phillips, F. Jiang, J. Ayyad, N. Penard, and H. Abdi. Face recognition algorithms surpass humans matching faces over changes in illumination. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29 :1642–1646, 2007.
- [9] C. Lu and X. Tang. Learning the face prior for Bayesian face recognition. In Computer Vision ECCV 2014, volume 8692 of Lecture Notes in Computer Science. Springer International Publishing, 2014.
- [10] D. Cox and N. Pinto. Beyond simple features : A large-scale feature search approach to unconstrained face recognition. In Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, pages 8–15, March 2011.
- [11] H. Wang, S. Z. Li, and Y. Wang. Face recognition under varying lighting conditions using self quotient image. In FG, pages 819–824. IEEE Computer Society, 2004.
- [12] S. R. Arashloo and J. Kittler. Class-specific kernel fusion of multiple descriptors for face verification using multiscale binarised statistical image features. IEEE Transactions on Information Forensics and Security, 9(12) :2100–2109, 2014.
- [13] <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>