Efficient Face Verification Under Makeup Using Few Salient Facial Regions

Khouloud Ferchichi, Haythem Ghazouani and Walid Barhoumi

July 12, 2020

Efficient Face Verification Under Makeup Using Few Salient Facial Regions

Khouloud Ferchichi[1], Haythem Ghazouani[1,2], and Walid Barhoumi[1,2]

khouloudferchichi1@gmail.com, haythem.ghazouani@enicar.u-carthage.tn, walid.barhoumi@enicarthage.rnu.tn
[1]Research Team on Intelligent Systems in Imaging and Articial Vision (SIIVA)
LR16ES06 Laboratoire de Recherche en Informatique, Modélisation et Traitement de l' Information et de la Connaissance (LIMTIC), Université de Tunis El Manar, Institut Supérieur d' Informatique, Tunis, Tunisia
[2]Ecole Nationale d'Ingénieurs de Carthage, Université de Carthage, Tunis, Tunisia

***Abstract*—Automatic recognition of persons has attracted the attention of many researchers during the last years due to its many applications in various fields. However, this task faces several challenges related to many changes that can affect the human face. In particular, make-up faces represent a major challenge for facial recognition and verification. To deal with this issue, we propose an efficient salient patch-based method for verifying faces under makeup variation. Firstly, we use Mutli-Task Cascaded Convolutional Neural Networks (MTC-CNN) to jointly, detect and align the face with five landmarks. The Histogram of Oriented Gradients (HOG) descriptor and Local Binary Patterns (LBP) are then adopted to represent the face by concatenating their histogram features in few salient regions around the detected landmarks. Finally, we calculate the similarity measure between the extracted features to compare the two faces and determine whether they are for the same person or not. The performance of the proposed method is validated on the challenging YMU (YouTube Makeup dataset) and $MIFS$ (Makeup Induced Face Spoofing) datasets. The obtained results proved the superiority of the proposed method against multi-patch based method from the state of the art.**

***Index Terms*—Face verification, Mutli-task Cascaded Convolutional Networks, Histogram Oriented of Gradients , Local Binary Patterns**

I. Introduction

During the past few years, a great number of face recognition methods were developed in both computer vision and computer security communities [4], [12], [13], [16]. Typical applications include face recognition [4], [14], [19], [21], [27], [32], [33], [38], facial age estimation [5], [15], [16], [38], [40], facial expression recognition [2], [3], [9], [39], facial gender identification [24], [25], facial sketch recognition [36], [37], and human ethnicity recognition from facial images [18], [26]. The face recognition problem is within the most challenging within the framework of method dealing with human faces because of the diversity caused by expressions, gender, pose, illumination, cosmetics, etc. In particular, makeup can significantly alter the appearance of the face and complicate the verification task. More precisely, makeup can change the perceived facial shape and appearance by modifying contrast levels in the mouth and eye region and altering skin texture. Recent studies have demonstrated that makeup can significantly degrade face matching accuracy. Indeed, the influence of makeup to the human perception has been discussed in [28], [29]. Dantcheva et al. [11] showed that makeup changes can weaken the performance of current face recognition schemes. A significant decrease has been observed when the verification is conducted between the face with and without cosmetics. Furthermore, few works have made efforts on robust face recognition with makeup changes [20]. For instance, Chen et al. [6] introduced a patch-based ensemble learning method, which uses multiple sub-spaces generated by sampling patches from before-makeup and after-makeup face images. Li et al. [22] proposed a learning from generation approach for makeup-invariant face verification by introducing a bi-level adversarial network (BLAN).

Zhang et al. [35] proposed a discriminative marginal metric learning (DMML) method to learn a robust metric space such that facial images with makeup relations are mapped closely and facial images without makeup relations are separated from each other as far as possible. In [20], Hu et al. used Canonical Correlation Analysis (CCA) to learn the meta subspace, which can maximize the correlation of feature vectors belonging to the same individual. Guo et al. [17] proposed to use Partial Least Square (PLS) to learn the correlation between different facial parts separately. Wa et al. [30] proposed a unified Face Morphological Multi-branch Network (FM 2 u-Net) for makeup-invariant face verification, which can simultaneously synthesize many diverse makeup faces through face morphology network (FM-Net) and effectively learn cosmetics-robust face representations using attention-based multi-branch learning network (AttM-Net). The experimental results show that the correlation learning is an essential step to improve the verification performance with cosmetic changes.

Generally, by going over the literature of facial verification under makeup, we notice that the works designed to improve the accuracy rate have significantly increased the response time. Finding the right trade-off between computational cost and verification accuracy, when dealing with makeup faces, is the main motivation of this work. Indeed, in this paper, we perform face verification using only few salient face regions. These regions are based on five landmarks robustly extracted using a Mutli-Task Cascaded Convolutional Neural Networks (MTC-CNN). Two well known descriptors, namely Histogram of Oriented Gradients (HOG) and Local Binary Patterns, are tested when extracting features from the salient regions. The reason we selected these features is that they have shown good performance in recent kinship verification research [23], [31], [41]. In addiction, since the HOG is edge-based and the LBP is texture-based, we investigate the makeup impact in facial texture and facial component contours. Experimental results on two real-world makeup face datasets are used to show the possibility of verifying the under makup faces using a limited number of salient facial regions using classical descriptors.

This paper is organized as follows: Section II presents the proposed method. Section III, describes the validation data and the experimental protocol, the Section IV details the experimental results, and Section V concludes the work while proposing some ideas for future work.

II. Proposed method

The proposed method is mainly composed of three steps. Firstly, given two input images of the same person under different makeup situations (I1 is an image with makeup and I2 is an image without makeup), faces and five keypoints are located. The detected keypoints are left eye corner, right eye corner, left mouth corner, right mouth corner and the highest point of the nose. In fact, the choice of these keypoints is not arbitrary. We assume that the five regions around these landmarks are very relevant for face verification. Then, to describe the faces within the defined regions, we investigated the use of HOG and LBP features. To verify whether the two images belong to the same person or not, two metrics (Euclidean and Chi-squared distances) are tested to measure the similarity between the resulting features. Finally, an empirical threshold is used to make the final decision.

A. Face and keypoint detection

Face detection and localization from images is a necessary first step in face verification systems. To deal with this issue, we applied Multi-Task Cascaded Convolutional Networks (MTCNN) [34] in order to locate salient regions, while cropping and aligning the facial areas. In fact, MTCNN is a CNN-based face detection method, which uses three cascaded CNNs for fast and accurate joint face detection and alignment. This method detects five facial landmarks (two eyes, two mouth corners and the highest point of the nose). The candidate regions are produced in a first stage and refined in the two latter stages. The five landmarks detection results are produced in the third stage.

Figure 2 illustrates the output of the first step of the proposed method, where Figure 2.a represents the input image, Figure 2.b represents the face detected and key points and Figure 2.c illustrates the bounding box, that is formatted as [x, y, width, height] under the key 'box'. The confidence is the probability for a bounding box to be matching a face, and each keypoint is identified by a pixel position (x, y). It is clear that the face is correctly detected. Indeed, the face detection rate is almost 100% while testing on $MIFS$ and YMU datasets, which are composed of 364 images and 212 images, respectively .

We use the detected landmarks to locate the salient facial regions around the main face components (eyes, noise and mouth). We used 4-region and 5-region-based verification as shown in Figure 3 and 4, respectively. In the 5-region-based verification we defined rectangular patches around the 5 detected landmarks, however for the 4-region-based verification we have determined the center of the mouth using the two mouth corner landmarks and defined a region around it to cover all the mouth.

B. Face representation

To represent the facial images using the defined salient regions, we investigate the application of HOG and LBP descriptors. The use of the HOG [10], which is an edge-based descriptor is to study the robustness of the facial component contour representation against makeup changes. However, the use of LBP descriptor [1] is to investigate how the makeup can affect a robust textural representation. The main reason behind the selection of these features is that they have shown good performance in recent face verification works. On the one hand, HOG is based on the computation of local histograms of the orientations of the image gradient in a grid. The underlying idea is that an object appearance can be characterized by the local distribution of its edge orientation. The HOG feature is robust and has no sensitivity to both light and geometric changes, and its computational complexity is low. In this work, we compute HOG histogram for each salient region. This is done by applying a 1D centered point discrete derivative mask in both the horizontal and vertical directions. The gradient is then transformed to polar coordinates, with the angle constrained to be between 0 and 180 degrees. Then, orientation binning consist to create the histograms for each region, such that the histogram bins are uniformly supplemented from 0° to 180° with a gap of 20°. Every pixel in the region casts a weighted vote to one of the 9 histogram bins for the orientation it belongs to. After that, histograms are normalized to diminish the effect of contrast changes due to makeup. Finally, all region histograms are concatenated in order to construct a 36-dimensional final feature vector that will represent the face. On the other hand, LBP is computed by comparing the intensity value of the center pixel with its surrounding neighbor intensity values. A small neighborhood (3×3) is used to extract the LBP feature in this work. An eight-bit string coding the intensity differences between the 8 neighbors and the central pixel is then extracted and transformed to its equivalent decimal number, which corresponds to the vote in the 256-bin histogram. A LBP histogram is computed for each salient region and the final feature vector is formed by concatenating all histograms $(256 \times 4 = 1024$ bins). The same feature extraction scheme is followed for extracting the features for the 5-region-based face verification (Figure 4). The

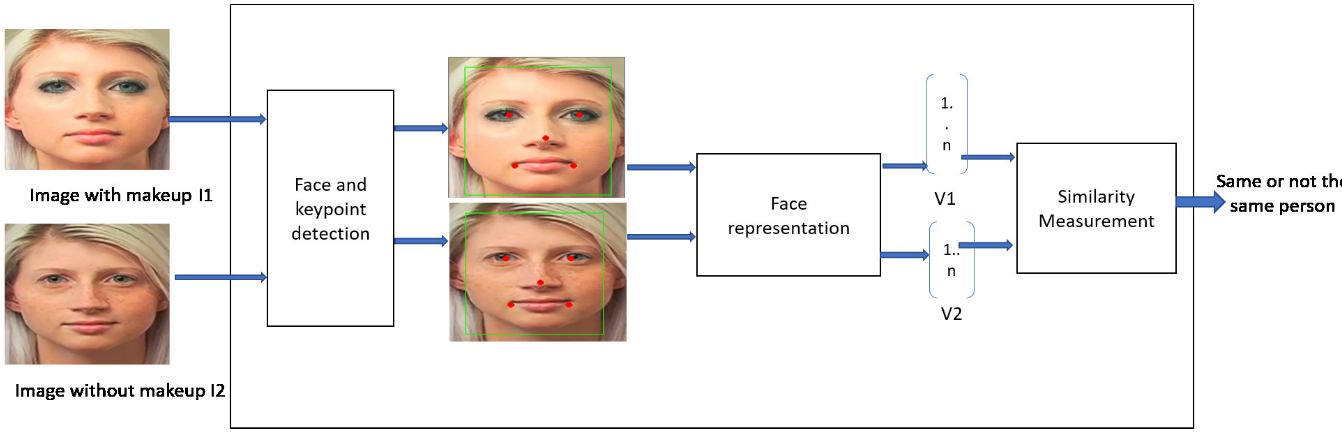


Fig. 1: Flowchart of the proposed method for face verification

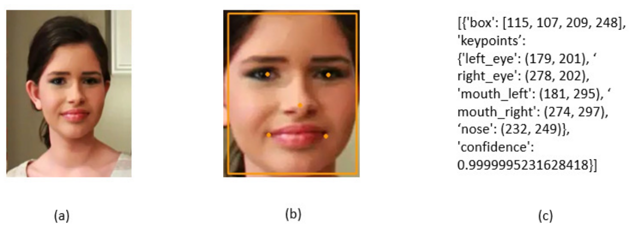


Fig. 2: Exemple of face and key points detection

final HOG and LBP-based feature vectors are 45-dimensional and 1280-dimensional, respectively.

C. Feature Similarity Measure

The objective of face verification is to determine whether the two face images are of the same person or not. Formally, two face images of the same person are called a similar pair; otherwise, two face images of different persons are called a dissimilar pair or a different pair. Since face verification needs an appropriate way to measure the difference or similarity between two images, two similarity metrics are investigated. The first one is the Euclidean distance Euc (1) which is a standard measurement while the second one is the Chi-squared distance χ^2 (2), which is a semantic measurement.

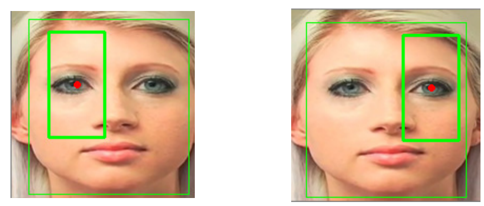
$$Euc(V_1, V_2) = \sqrt{\sum_{i=1}^{n} (V_1^i - V_2^i)^2} \quad (1)$$

$$\chi^2(V_1, V_2) = \sqrt{\sum_{i=1}^{n} \frac{(V_1^i - V_2^i)^2}{(V_1^i + V_2^i)}} \quad (2)$$

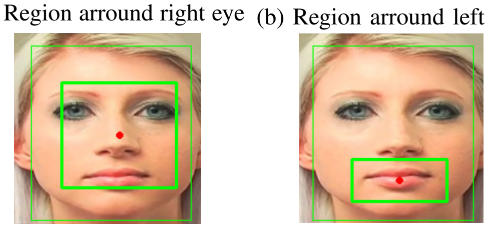
Where V_1 and V_2 are two n-dimensional feature vectors. Finally, a threshold is empirically set while finding the best trade-off between a high positive rate and a low negative rate.

III. Validation Data and Experimental Protocol

We tested the performance of the proposed method experimentally on the following datasets: YouTube Makeup (YMU)

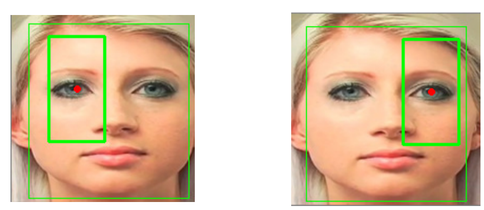


(a) Region arround right eye (b) Region arround left eye

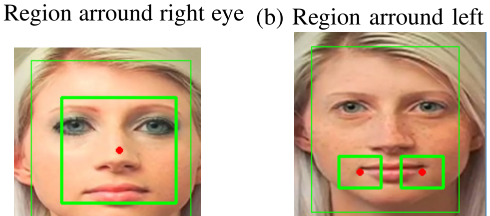


(c) Region arround noise (d) Region arround mouth

Fig. 3: The differents regions under four key points



(a) Region arround right eye (b) Region arround left eye



(c) Region arround noise (d) Region arround two corners of mouth

Fig. 4: The differents regions under five key points

dataset and Makeup Induced Face Spoofing ($MIFS$) dataset. The verification is validated using the HOG and LBP descriptors while testing the Euclidean and the chi-squared distances as feature similarity measures. Experiments were conducted using a 64 bit Windows operating system with Intel Core i5-2430M CPU at 2.40 GHz and 8GB RAM. In this study, we used the $MIFS$ dataset [7], which contains the before and after makeup images to evaluate the performance of the proposed scheme. Examples are shown in Figure 5. This dataset contains 642 individuals with 107 makeup transformations taken from random You-Tube makeup video tutorials. Each subject is attempting to spoof a target identity, which makes this dataset particularly challenging. Indeed, three sets of face images are provided: images of a subject before makeup; images of the same subject after makeup with the intention of spoofing and images of the target subject who is being spoofed. In addition to the $MIFS$ dataset, we carried out experiments on the YMU YouTube makeup dataset introduced by Dantcheva et al. [11], which contains the before and after makeup images of 151 Caucasian female subjects taken from YouTube makeup tutorials. Samples from YMU dataset are shown in Figure 6 (after face cropping and alignment). For the majority of the subjects, there are four shots per subject, two shots before the application of makeup and two shots after the application of makeup. For some subjects, there is either only one shot or three shots each before and after the application of makeup. The total number of images in the YMU dataset is 604, with 302 makeup images and 302 no-makeup images. It is worth noting that the degree of makeup in this dataset varies from subtle to heavy. The dataset is relatively unconstrained, exhibiting variations in facial expression, pose and resolution.

Fig. 5: An illustration of the facial images of $MIFS$.

The detailed comparison of makeup face database is illustrated in Table 1 For point detection and alignment training, we used samples from these two datasets. For $MIFS$ dataset we used 364 individuals however for YMU dataset we used 212 individuals.

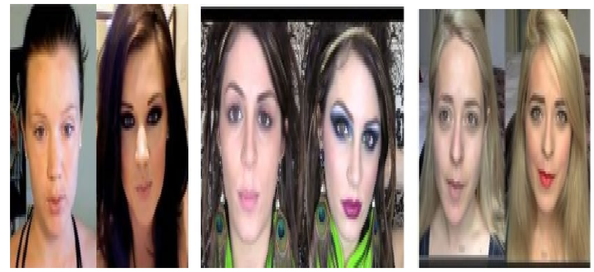


Fig. 6: An illustration of the facial images of YMU

properties	$MIFS$	YMU
Images	642	604
Subjects	107	151
Female	107	151
Male	0	0

TABLE I: Comparison of existing makeup face dataset

IV. Experimental results

This section discusses experiments performed to demonstrate the HOG and LBP effectiveness for face verification with makeup variations using the defined salient regions. In order to evaluate the performance of the proposed face matching, we examine the face verification using the defined 4 and 5 salient regions shown in Figure 3 and 4, respectively. For all the experiment runs, we calculate the accuracy rate, the area under the ROC curve (AUC) and the Equal Error Rate (EER) to evaluate the performance of the suggested method. The face verification is investigated for the LBP and HOG descriptor. Note that the accuracy and AUC are two different measures, which can be differently interpreted for the verification results. The different metrics, with different thresholds using the two descriptors for matching scenarios on the YMU and $MIFS$ datasets are summarized in Table 2 . Th ROC curve, describing the verification performance by computing the true-positive and false positive rates, are calculated using many different threshold values. The Auc, accuracy rate and EER of different feature representations on the two makeup face dataset with different distances are also shown in Table 2. As shown in this table, our approach obtains better performance with four regions than five regions. This is because the mouth is well represented using the 4 regions. Indeed, the AUC is $72,12\%$ when five regions are used and it reaches $74,94\%$ using four regions. Moreover, the LBP performs worst than HOG method on YMU using the Euclidean distance. As seen in table 2, LBP performs better on the dataset YMU than HOG with Euclidean distance while the reverse is correct for $MIFS$ dataset. Both datasets include some expression and pose variation however the $MIFS$ dataset includes also illumination and resolution changes. We can conclude that, in controlled environment, LBP can perform well than HOG when it comes to makeup face verification. However, HOG is more suitable for in-the-wild face verification. The verification accuracy on the YMU dataset are generally higher than those obtanied on $MIFS$, which means that the makeup face verification on $MIFS$ is more difficult than on YMU. Indeed, the size of $MIFS$ is larger than that of YMU which can explain the unbalanced results.

Moreover, we have plotted the receiver operating characteristic (ROC) curves of the two feature representations with the two distances in Figure 7 and Figure 8 on YMU and $MIFS$ datasets, respectively. As shown in Figure 7, using four-region-based verification achieved much better performance than five-regions-based verification as reported also later from the AUC and accuracy rates in Table 2.

As works on face verification and recognition under makeup are not very abundant in the literature, the proposed method is evaluated against the makeup face verification and recognition method proposed in [6] since the author used the YMU dataset for the validation. It is worth noting, that for the comparison results, the computational time is recorded using the hardware configuration reported in the original papers. The authors in [6] used a more powerful configuration than the configuration used in this work (i7-2600 CPU at 3.40 GHz against i5-2430M CPU at 2.40 GHz). In [6], each face image is tessellated into patches and each patch is represented by a set of features namely viz., Local Gradient Gabor Pattern (LGGP) [8], Histogram of Gabor Ordinal Ratio Measures (HGORM) and Densely Sampled Local Binary Pattern (DS-LBP). An improved Random Subspace Linear Discriminant Analysis (SRS-LDA) is used to perform ensemble learning by sampling patches and constructing multiple common subspaces between before-makeup and after-makeup facial images. Finally, Collaborative based and Sparse-based Representation Classifiers are used to compare feature vectors in this subspace and the resulting scores are combined via the sum-rule. As seen in Table 3, the proposed method, used with the LBP descriptor on 4 salient regions, slightly outperformed the method proposed in [6] in terms of accuracy.

It is worth noting that this method uses 2560 patches and jointly 4 descriptors. Indeed, using regions in the face that are not prominent for face verification can lead to a decrease in the accuracy rate. Moreover, using a large number of descriptors can lead to miss-verification results due to the information redundancy on the features they generate. The proposed method performs well using only one descriptor because it discards regions that can lead to unnecessary features. The difference in accuracy is not very significant between the suggested method and the method proposed in [6], however the gain in computational time is around 4.11 seconds which is very significant when it comes to real-time applications such as security, face recognition on smartphone systems, etc.

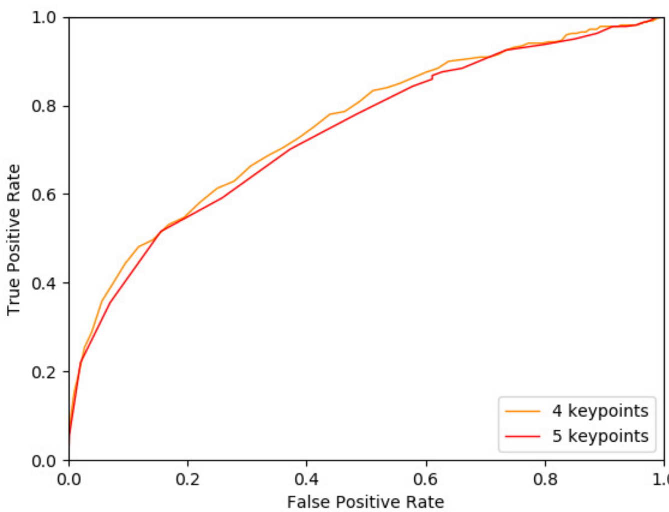
To end with, there are usually billions of facial images on various social websites, and millions of images are added to the websites everyday. One key problem is how to automatically manage such large-scale images. In this problem, there are two challenges to be tackled: (1) who the people in images are and (2) how to recognize the facial images with makeup. Previous face recognition techniques may be effective to tackle the first problem however makeup face verification should be a useful technique to alleviate the second challenge. When the facial makeup relation is known, it is possible to automatically organize the images according to the subject identities.

V. Conclusion

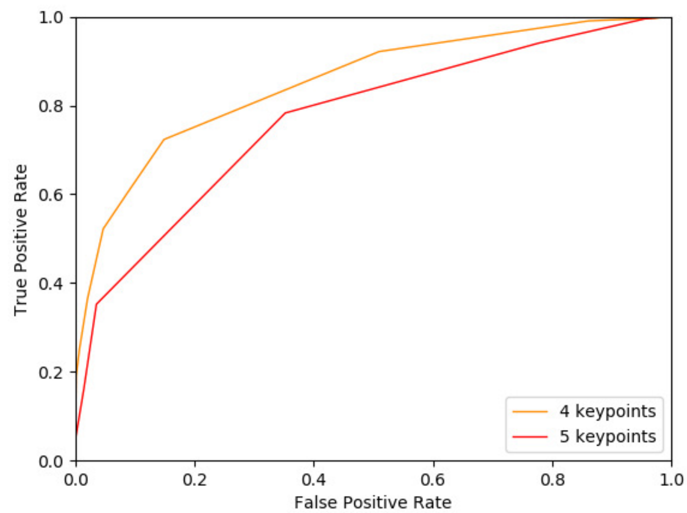
In this paper, we have studied the makeup face verification problem in the wild using real datasets. This work investigated the use of the HOG and LBP descriptors for the face verification under different conditions including makeup changes. We have proven in this work, that using limited number of salient regions can be more effective than representing the face as a whole using multiple patches. Both investigated descriptors have given good verification results. According to the experiments, LBP has shown to be more effective on controlled environments however HOG have shown more robustness against in-the-wild conditions. Like all research works, this work is not without flaws. In fact, the salient region sizes and the verification thresholds were set manually. This leaves room for further improvement in precision by automatically setting these parameters. Moreover, the HOG and LBP descriptors were used separately and their late and early fusion can be investigated in future works.

TABLE II: Auc, accuracy and EER for the proposed method with different number of patches, thresholds and distance metrics

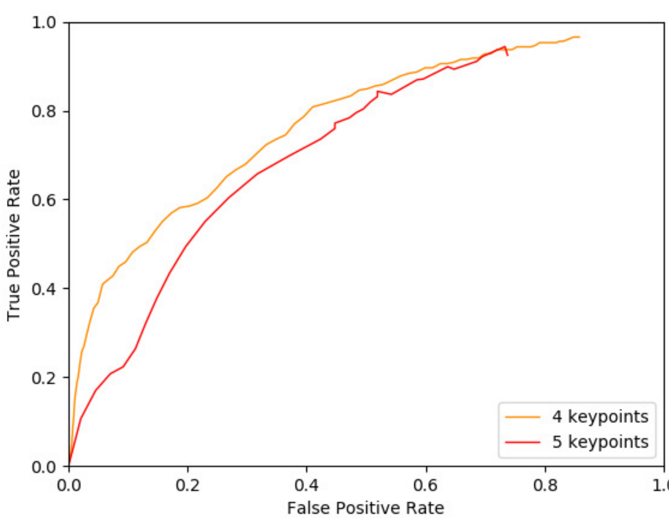
Descriptor	**Threshold**	Keypoints	**Dataset**	**distance**	**AUC**	**Accuracy**	**EER**
HOG	0.5	4	YMU	Chisquare	63.3%	68.57%	0,315
HOG	0.61	5	YMU	Chisquare	46.4%	67.12%	0.328
HOG	0.24	4	YMU	Euclidean	74.94%	74.94%	0.32
HOG	0.33	5	YMU	Euclidean	72.12%	72.12%	0.33
LBP	0.06	4	YMU	Chisquare	65.48%	68,05%	0.32
LBP	0.08	5	YMU	Chisquare	63.39%	52.68%	0.373
LBP	0.17	4	YMU	Euclidean	85.18%	89.68%	0.23
LBP	0.2	5	YMU	Euclidean	76.99%	76.99%	0.28
HOG	0.48	4	$MIFS$	Chisquare	68.49%	67.77%	0.323
HOG	0.42	5	$MIFS$	Chisquare	49.7%	65.47%	0.346
HOG	0.21	4	$MIFS$	Euclidean	72.19%	66.38%	0.33
HOG	0.31	5	$MIFS$	Euclidean	70.7%	65.68%	0.342
LBP	0.08	4	$MIFS$	Chisquare	62.67%	62.81%	0.37
LBP	0.1	5	$MIFS$	Chisquare	53.33%	49.7%	0.51
LBP	0.17	4	$MIFS$	Euclidean	78%	67.4%	0.327
LBP	0.25	5	$MIFS$	Euclidean	70%	67.35%	0.32



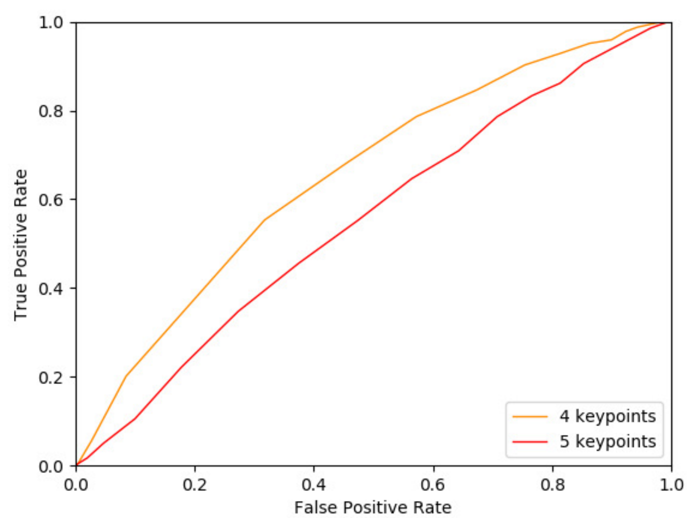
(a) HOG with Euclidean distance



(b) LBP with Euclidean distance

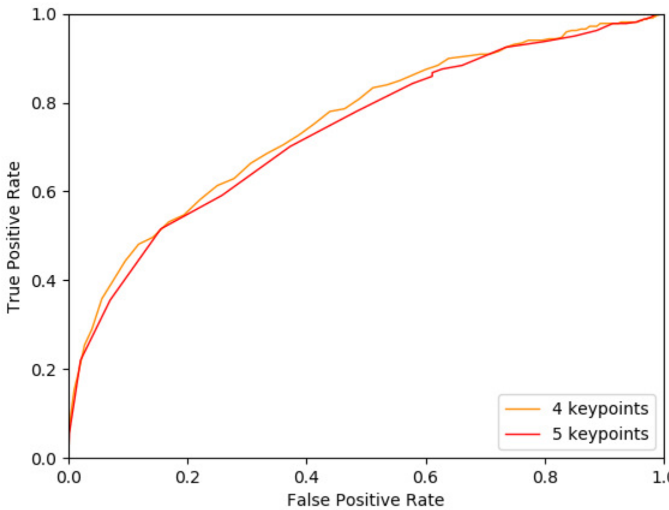


(c) HOG with chisquared distance

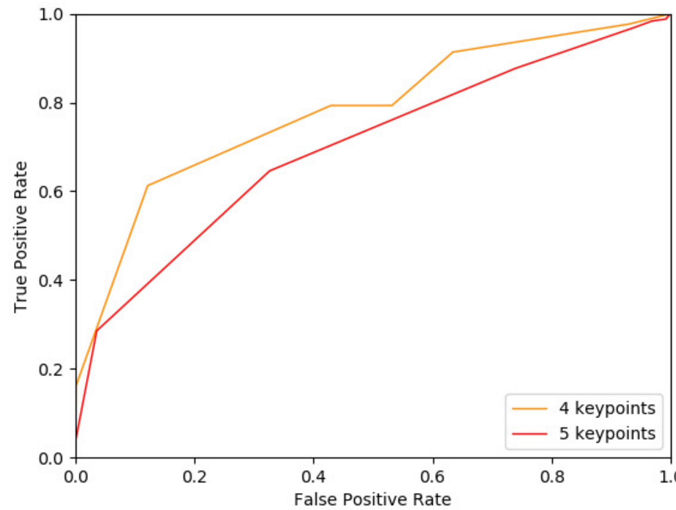


(d) LBP with chisquared distance

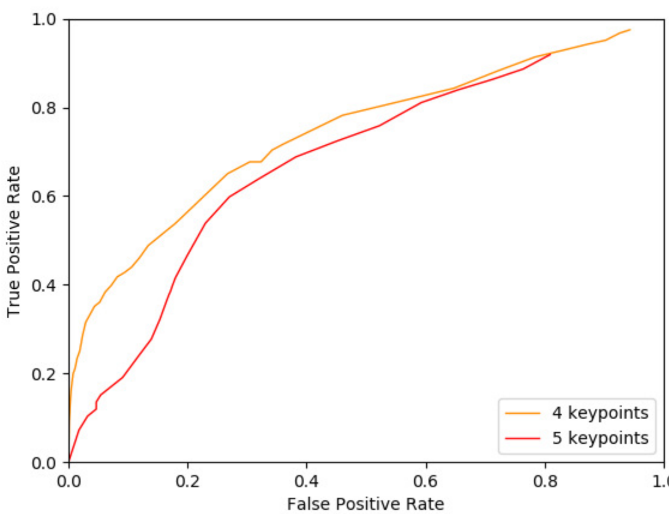
Fig. 7: ROC CURVE OF FACE VERIFICATION USING FEATURE REPRESENTATION HOG AND LBP RESPECTIVELY ON YMU WITH EUCLIDEAN AND CHISQUARED DISTANCE



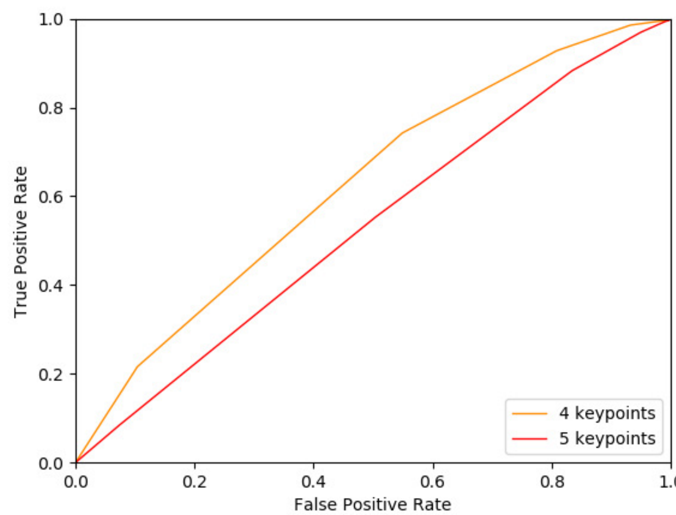
(a) HOG with Euclidean distance



(b) LBP with Euclidean distance



(c) HOG with chisquared distance



(d) LBP with chisquared distance

Fig. 8: ROC CURVE OF FACE VERIFICATION USING FEATURE REPRESENTATION HOG AND LBP RESPECTIVELY ON $MIFS$ WITH EUCLIDEAN AND CHISQUARED DISTANCE

TABLE III: Comparison of the face recognition results and computational time on the YMU datset.

Method	**Accuracy**	**computational time (s)**
(SRS-LDA)+COTS-2+COTS-3+HGORM (2560 patches)	89.40%	4.18
Proposed (LBP+4 patches)	**89.68%**	**0.07**

References

[1] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12):2037–2041, 2006.

[2] Hela Bejaoui, Haythem Ghazouani, and Walid Barhoumi. Fully automated facial expression recognition using 3d morphable model and mesh-local binary pattern. pages 39–50, 11 2017.

[3] Hela Bejaoui, Haythem Ghazouani, and Walid Barhoumi. Sparse coding-based representation of lbp difference for 3d/4d facial expression recognition. *Multimedia Tools and Applications*, 78:22773–22796, 08 2019.

[4] Peter N. Belhumeur, João P Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):711–720, 1997.

[5] Enrico Bondi, Pietro Pala, Stefano Berretti, and Alberto Del Bimbo. Reconstructing high-resolution face models from kinect depth sequences. *IEEE Transactions on Information Forensics and Security*, 11(12):2843–2853, 2016.

[6] Cunjian Chen, Antitza Dantcheva, and Arun Ross. An ensemble of patch-based subspaces for makeup-robust face recognition. *Information fusion*, 32:80–92, 2016.

[7] Cunjian Chen, Antitza Dantcheva, Thomas Swearingen, and Arun Ross. Spoofing faces using makeup: An investigative study. In *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, pages 1–8. IEEE, 2017.

[8] Cunjian Chen and Arun Ross. Local gradient gabor pattern (lggp) with applications in face recognition, cross-spectral matching, and soft biometrics. In *Biometric and Surveillance Technology for Human and Activity Identification X*, volume 8712, page 87120R. International Society for Optics and Photonics, 2013.

[9] Ira Cohen, Nicu Sebe, Ashutosh Garg, Michael S Lew, and Thomas S Huang. Facial expression recognition from video sequences. In *Proceedings. IEEE International Conference on Multimedia and Expo*, volume 2, pages 121–124. IEEE, 2002.

[10] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.

[11] Antitza Dantcheva, Cunjian Chen, and Arun Ross. Can facial cosmetics affect the matching accuracy of face recognition systems? In *2012 IEEE Fifth international conference on biometrics: theory, applications and systems (BTAS)*, pages 391–398. IEEE, 2012.

[12] Changxing Ding, Jonghyun Choi, Dacheng Tao, and Larry S Davis. Multi-directional multi-level dual-cross patterns for robust face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 38(3):518–531, 2015.

[13] Changxing Ding and Dacheng Tao. A comprehensive survey on pose-invariant face recognition. *ACM Transactions on intelligent systems and technology (TIST)*, 7(3):1–42, 2016.

[14] Qingxiang Feng, Chun Yuan, Jeng-Shyang Pan, Jar-Ferr Yang, Yang-Ting Chou, Yicong Zhou, and Weifeng Li. Superimposed sparse parameter classifiers for face recognition. *IEEE transactions on cybernetics*, 47(2):378–390, 2016.

[15] Xin Geng, Zhi-Hua Zhou, and Kate Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 29(12):2234–2240, 2007.

[16] Guodong Guo, Yun Fu, Charles R Dyer, and Thomas S Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Transactions on Image Processing*, 17(7):1178–1188, 2008.

[17] Guodong Guo, Lingyun Wen, and Shuicheng Yan. Face authentication with makeup changes. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(5):814–825, 2013.

[18] Srinivas Gutta, Jeffrey RJ Huang, P Jonathon, and Harry Wechsler. Mixture of experts for classification of gender, ethnic origin, and pose of human faces. *IEEE Transactions on neural networks*, 11(4):948–960, 2000.

[19] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi, and Hong-Jiang Zhang. Face recognition using laplacianfaces. *IEEE transactions on pattern analysis and machine intelligence*, 27(3):328–340, 2005.

[20] Junlin Hu, Yongxin Ge, Jiwen Lu, and Xin Feng. Makeup-robust face verification. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2342–2346. IEEE, 2013.

[21] Effrosini Kokiopoulou and Yousef Saad. Orthogonal neighborhood preserving projections. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*, pages 8–pp. IEEE, 2005.

[22] Yi Li, Lingxiao Song, Xiang Wu, Ran He, and Tieniu Tan. Learning a bi-level adversarial network with global and local perception for makeup-invariant face verification. *Pattern Recognition*, 90:99–108, 2019.

[23] Jiwen Lu, Xiuzhuang Zhou, Yap-Pen Tan, Yuanyuan Shang, and Jie Zhou. Neighborhood repulsed metric learning for kinship verification. *IEEE transactions on pattern analysis and machine intelligence*, 36(2):331–345, 2013.

[24] Baback Moghaddam and Ming-Hsuan Yang. Gender classification with support vector machines. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pages 306–311. IEEE, 2000.

[25] Baback Moghaddam and Ming-Hsuan Yang. Learning gender with support faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):707–711, 2002.

[26] Yongsheng Ou, Xinyu Wu, Huihuan Qian, and Yangsheng Xu. A real time race classification system. In *2005 IEEE International Conference on Information Acquisition*, pages 6–pp. IEEE, 2005.

[27] Chuan-Xian Ren, Zhen Lei, Dao-Qing Dai, and Stan Z Li. Enhanced local gradient order features and discriminant analysis for face recognition. *IEEE transactions on cybernetics*, 46(11):2656–2669, 2015.

[28] Gillian Rhodes, Alex Sumich, and Graham Byatt. Are average facial configurations attractive only because of their symmetry? *Psychological science*, 10(1):52–58, 1999.

[29] Sayako Ueda and Takamasa Koyama. Influence of make-up on facial recognition. *Perception*, 39(2):260–264, 2010.

[30] Wenxuan Wang, Yanwei Fu, Xuelin Qian, Yu-Gang Jiang, Qi Tian, and Xiangyang Xue. Fm2u-net: Face morphological multi-branch network for makeup-invariant face verification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5730–5740, 2020.

[31] Haibin Yan, Jiwen Lu, and Xiuzhuang Zhou. Prototype-based discriminative feature learning for kinship verification. *IEEE Transactions on cybernetics*, 45(11):2535–2545, 2014.

[32] Shuai Yang, Jiaying Liu, Yuming Fang, and Zongming Guo. Joint-feature guided depth map super-resolution with face priors. *IEEE transactions on cybernetics*, 48(1):399–411, 2016.

[33] Weiwei Yu, Xiaolong Teng, and Chongqing Liu. Face recognition using discriminant locality preserving projections. *Image and Vision computing*, 24(3):239–248, 2006.

[34] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.

[35] Lining Zhang, Hubert PH Shum, Li Liu, Guodong Guo, and Ling Shao. Multiview discriminative marginal metric learning for makeup face verification. *Neurocomputing*, 333:339–350, 2019.

[36] Shengchuan Zhang, Xinbo Gao, Nannan Wang, and Jie Li. Robust face sketch style synthesis. *IEEE Transactions on Image Processing*, 25(1):220–232, 2015.

[37] Shengchuan Zhang, Xinbo Gao, Nannan Wang, Jie Li, and Mingjin Zhang. Face sketch synthesis via sparse representation-based greedy search. *IEEE transactions on image processing*, 24(8):2466–2477, 2015.

[38] Wenyi Zhao, Rama Chellappa, P Jonathon Phillips, and Azriel Rosenfeld. Face recognition: A literature survey. *ACM computing surveys (CSUR)*, 35(4):399–458, 2003.

[39] Wenming Zheng, Xiaoyan Zhou, Cairong Zou, and Li Zhao. Facial expression recognition using kernel canonical correlation analysis (kcca). *IEEE transactions on neural networks*, 17(1):233–238, 2006.

[40] Li Zhifeng, Gong Dihong, Li Xuelong, et al. Aging face recognition: A hierarchical learning model based on local pattern selection. *IEEE Transactions on Image Processing*, 25(5):2146–2154, 2016.

[41] Xiuzhuang Zhou, Junlin Hu, Jiwen Lu, Yuanyuan Shang, and Yong Guan. Kinship verification from facial images under uncontrolled conditions. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 953–956, 2011.