Boosting child abuse victim identification in Forensic Tools with hashing techniques

Rubel Biswas
Departamento IESA
Universidad de León
Researcher at INCIBE
rbis@unileon.es

Victor González-Castro
Departamento IESA
Universidad de León
Researcher at INCIBE
vgonc@unileon.es

Eduardo Fidalgo Fernández
Departamento IESA
Universidad de León
Researcher at INCIBE
efidf@unileon.es

Deisy Chaves
Departamento IESA
Universidad de León
Researcher at INCIBE
dchas@unileon.es

***Abstract*—In this work, we present a scheme to identify occluded faces using perceptual image hashing. Most of the existing methods for this problem focus on occlusion detection and further removal of the occluded area by training a facial model. In this paper, we propose a new hashing method which does not require prior training. Our model combines frequency coefficients and statistical image information to increase the recognition accuracy of occluded faces. The proposed method aims to improve face recognition accuracy in forensic tools such as victim identification in Child Sexual Abuse (CSA) materials. Experimental results showed that the proposed method outperforms the results obtained with perceptual image hashing for occluded face identification using the LFW database.**

***Index Terms*—Face identification, Face recognition, Perceptual hashing, CLOSIB, pHash, NMF**

Type of contribution: Ongoing research

I. Introduction

Automatic face identification or recognition is widely used in many real-time applications such as forensics, surveillance or criminal identification among others. In recent years, deep learning techniques have achieved a considerable development in this area [1]. Nevertheless, there are still some open issues during face identification. One of the most challenging problems is the occlusion of the face, which can be caused by several reasons, such as self-occlusion (e.g. non-frontal position), accessories (e.g. glasses, masks or hair) or adversarial attacks (i.e. image faces modified by adding small changes to make difficult the identification).

In the literature, we can find works that deal with the automatic Child Sexual Abuse (CSA) material detection [2]. However, after detecting such kind of material, Law Enforcement Agencies (LEA) may face the problem of identifying the victims. Children usually are presented dressed with customs, accessories that cover their faces and also sometimes the CSA material receives eye adversarial attacks before uploading the material to the Web. Therefore, occluded face identification in CSA images remains a challenging task for LEAs.

To address the problem of identifying occluded faces we proposed a face recognition method based on perceptual image hashing [3] to avoid the training of facial models and represent an image content as a fixed-length vector. This research is part of the European project Forensic Against Sexual Exploitation of Children (4NSEEK), and our primary goal is to enhance the Forensic Tools ability to recognize occluded faces in child pornography materials.

II. Related Work

Due to the availability of high-end GPU cards and training data, nowadays, deep learning methods achieve state of the art results in the task of face recognition [4]. Hongjun and Aleix [5] used SVM to find a hyperplane which is parallel to the affine subspace of occluded data. While Min et al. [6] trained a SVM classifier to detect the occluded part and use the non-occluded area to match with a corresponding part of a faces gallery.

On the other hand, many perceptual image hashing methods have been proposed and applied to the field of multimedia security [7]. Researchers have focused on image hashing schemes based on the concepts of deep hashing, Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Discrete Fourier Transform (DFT), and non-negative matrix factorization (NMF), among others.

We have noticed that occluded face recognition requires a trained model with occluded face features. To overcome the need for training a model, we designed an image hashing method to identify a person when the face image is partially occluded which can be applied in CSA cases.

III. Methodology

Fig. 1 represents the proposed perceptual hashing scheme, which is two-fold: (1) computation of the hash code of a non-occluded face image and (2) verification of the cropped occluded face.

In the first stage, Multi-Task Cascade CNN (MTCNN) method [8] is applied to detect a face contained in an image. Then, the face is cropped using the detected bounding box coordinates and re-sized it to 120 $\times$120 pixels. Next, the perceptual hash [3] and CLOSIB descriptors [9], [10] are computed to extract 64 coefficients and 128 statistical features of the face image, respectively. Afterwards, NMF [11] is applied to reduce the 128 CLOSIB features into 64 values. Finally, an element-wise multiplication between the 64 pHash coefficients and the 64 CLOSIB features is carried out to attain the final feature vector, i.e. the hash code, for the face, which is stored. In the second stage, an occluded face is cropped manually from an image and resized into 120 $\times$120 pixels. After that, the image hash codes are obtained using the process previously described. Finally, the similarity score is computed between the occluded face hash code and the ones stored with a correlation coefficient function. If the score is greater than a threshold, T, it is considered that a similar face is found in the internal storage of the module.

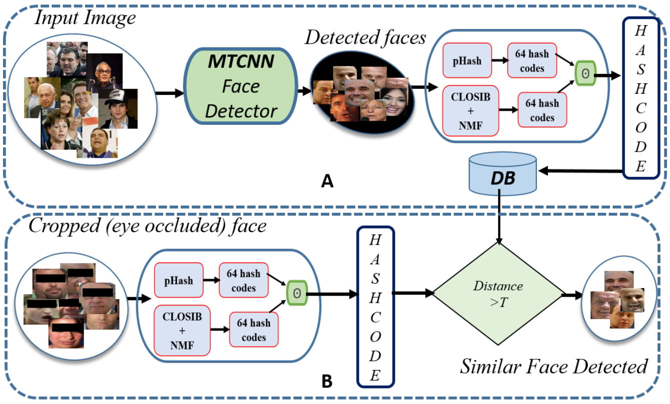

Figure 1. Occluded face verification scheme. A) Non-occluded face hash code computation and storage. B) Occluded face identification by comparing face hash code against previously stored hash codes

IV. Experimental Results and Discussion

During the experimentation, a PC with 128GB RAM and 12GB Titan Xp GPU is used. For testing purposes, we created an occluded face dataset by modifying the images in the LFW dataset. The resulting dataset, *occluded-LFW*, comprises 13,233 images with mostly frontal views of occluded faces that correspond to 5,749 identities.

The occluded-LFW dataset is used to evaluate the pHash algorithm and pHash combined with CLOSIB through NMF. We computed the correlation coefficient scores between occluded face hash codes and stored non-occluded face ones to assess their similarity. Finally, the proposed scheme retrieved as the identity the one with the maximum similarity score greater than a threshold, T. Table I presents the identification accuracy and the average processing time for occluded face recognition. In the experimentation, firstly, the hash codes of 13,233 non-occluded faces are extracted from LFW dataset and saved into the system storage. Secondly, in the occluded face identification stage, the hash code of each occluded face from the *occluded-LFW* dataset is computed and compared with the stored hash codes to identify the person. We have selected the threshold $T = 0.98$, because it reduces the false positive rate.

We have observed that the proposed combination of pHash and CLOSIB with NMF yields an accuracy of 69.89% in this experiment and outperformed pHash which obtained an accuracy of 40.3%. Presumably, because NMF allows reducing the feature vector dimension and helps to combine pHash and CLOSIB successfully to get a richer face representation. Finally, we found that the processing time of the combination of pHash and CLOSIB with NMF, in average 0.017 seconds, is higher than the one observed for pHash, in average 0.002 seconds. However the proposed method is still suitable for real-time applications such as forensics tools and it does not need to train any model.

V. Conclusions and Future Work

In this work, we proposed a perceptual image hashing scheme for occluded face recognition aiming to improve the occluded faces recognition in child pornography materials. The proposed approach is based on the combination of statistical features and frequency coefficients of an image

Table I
Partial occluded face identification accuracy and average processing time for occluded-LFW dataset

pHash		pHash and CLOSIB-NMF	
Accuracy (%)	Avg. time (Sec.)	Accuracy (%)	Avg. time (Sec.)
40.30	0.002	69.89	0.017

instead than a hard training. However, to achieve accurate results at least a non-occluded face (in a similar scenario) must be stored in the system before performing the occluded face identification. Moreover, we demonstrated that the proposed hashing method obtained the highest accuracy in this task, 69.89%, outperforming pHash, which obtained an accuracy of 40.30%. However, the processing time for the proposed schema is higher than that of pHash, but it presents a better trade-off between accuracy and the processing time.

In future work, we will attempt the design of new hashing techniques and evaluate them on different state-of-the-art datasets with similar characteristics to the problem domain.

Acknowledgement

This research has been funded with support from the European Commission under the 4NSEEK project with Grant Agreement 821966. This publication reflects the views only of the author, and the European Commission cannot be held responsible for any use which may be made of the information contained therein. This work is also supported by the framework agreement between the University of León and INCIBE (Spanish National Cybersecurity Institute) under Addendum 01. We acknowledge NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

[1] N. Crosswhite, J. Byrne, C. Stauffer, O. Parkhi, Q. Cao and A. Zisserman: "Template adaptation for face verification and identification", in *Image and Vision Computing*, vol. 79, pp. 35–48, 2018.
[2] A. Gangwar, E. Fidalgo, E. Alegre, V. Gonzáles-Castro, Pornography and Child Sexual Abuse Detection in Image and Video: A Comparative Evaluation, in: 8th International Conference on Imaging for Crime Detection and Prevention (ICDP), 2017, pp. 37–42
[3] C. Zauner: "Implementation and Benchmarking of Perceptual Image Hash Functions", in *Master's thesis, University of Applied Sciences Hagenberg, Austria*, 2010.
[4] F. Schroff, D. Kalenichenko and J. Philbin: "Facenet: A unified embedding for face recognition and clustering", in *IEEE conference on computer vision and pattern recognition*,pp. 815–823, 2015.
[5] J. Hongjun and M. M. Aleix: "Support vector machines in face recognition with occlusion", in *Proceedings of the IEEE Conference Computer Vision and Pattern Recognition*, pp. 136–141, 2009.
[6] M. Rui, H. Abdenour and D. Jean-Luc: "Improving the recognition of faces occluded by facial accessories", in *Automatic Face and Gesture Recognition and Workshops, IEEE*, pp. 442–447, 2011.
[7] C. Qin, X. Chen, J. Dong and X. Zhang: "Perceptual image hashing with selective sampling for salient structure features", in *Displays*, Vol. 45, pp. 26-37, 2016.
[8] K. Zhang, Z. Zhang, Z. Li and Y. Qiao: "Joint face detection and alignment using multitask cascaded convolutional networks", in *IEEE Signal Processing Letters*, Vol. 23, n. 10, pp. 1499–1503, 2016.
[9] O. García-olalla; L. Fernández-Robles; E. Alegre; M. Castejón-Limas; E. Fidalgo. Boosting Texture-Based Classification by Describing Statistical Information of Gray-Levels Differences. Sensors 2019, 19, 1048.
[10] O. García-olalla; Alegre, E.; Fernández-Robles, L.; Fidalgo, E.; Saikia, S. Textile retrieval based on image content from CDC and webcam cameras in indoor environments. Sensors 2018, 18, 1329.
[11] J. Pan and J. Zhang: "Large Margin Based Nonnegative Matrix Factorization and Partial Least Squares Regression for Face Recognition", in *Pattern Recognition Letters*, vol. 32, no. 14, pp. 1822-1835, 2011.