

A Hybrid Model for Frontal View Human Face Detection and Recognition

Mayuresh Nayak^{#1}, Lalit P. Saxena^{*2S}

[#]Software Engineer, Software Development Unit, Mumbai MH India

^{*}Research Scientist, Applied Research Section, Combo Consultancy, Obra UP India

Abstract—The human face detection and recognition search for human faces in images incorporated in computer-vision and real-time imaging applications. For an example, an attendance record keeping system is an immediate application, which extensively involves face recognition technologies, whereas their other applications that too are based on face recognition. This paper presents a hybrid model developed for the recognition of frontal view of human faces in an image. The proposed hybrid model uses an integral sum of images and template matching that confines the development of a human face detection and recognition model. The system takes as input the human face image, detects the frontal face, crops the selection and normalize it for post processing. First, the normalized images are stored for the training purposes and later the testing is done with these database images and real time imaging. This model extracts features from an input image of a human face, and match these features with the features stored in the database, previously recorded during model training. The obtained results confirmed the efficacy of the proposed model and favoured for the real-time implementation of the system for identification, recognition and security purposes.

Keywords—Human faces, pre-processing, detection, feature extraction, matching, recognition.

I. INTRODUCTION

The human face detection and recognition research search for human faces in images incorporated in several computer-vision and real-time imaging applications. Frontal-view human face detection and recognition carried out worldwide. Human face is one of the most complicated topics to be selected for recognition purpose as there are many variations and distinguishing features (Taigman et al., 2014). There are various applications incorporating face detection and recognition module in the overall functioning of an application. To name such an application, attendance record keeping is an immediate application driven out of this scope. Face recognition is evident in following application areas, such as, biometric attendance, content-based image retrieval, gate passkey codes, video coding, video conferencing, crowd surveillance, security checkpoints, and intelligent human-computer interfaces (Sharma et al., 2012).

There is a need to know more about human faces and how two images of face of a same person can differ and to what extent, and likewise. In any face recognition application, detecting a face is a basic step, which locates and separates the face from the image background. The Viola-Jones algorithm is a famous and effective attempt in face recognition from the images (Viola & Jones, 2004). The concept of integral images used in face recognition employ both local mean and local variance to binarize gray scale images, without any restrictions on window size (Viola & Jones, 2004; Bradley & Roth, 2007). According to (Viola & Jones, 2004; Bradley & Roth, 2007), an integral image is defined as an image with pixel intensity at each (at a specific position) pixel is the summation of the pixels intensities at both positions that is, above and left of that specific position. As far as the work is done in this area, the frontal human face images are passed as an input, the face through the features are saved and matched with the test input image based on the matching criteria, and finally the matching is done. The images are to be stored with the selected features of a person's face and specific labels are assigned for an easy access. There are certain considerations needs to be taken in this research area are; images types, face location in image, image background, and others.

The organization of the paper is as follows. Section 2 presents state of the art methods in face detection and recognition. The description of the proposed hybrid model is provided in section 3. Section 4 discusses about

the database and the experimental results. Section 5 concludes the paper and presents future insights for improvement.

II. PREVIOUS WORK

A human face has always been considered as a complex multidimensional structure, thus developing computational models is quite challenging. Since decades, the research in human face detection and recognition have considerably expanded in this area. The detection of a human face is a difficult problem in machine vision applications, because of its appearance variations and conditional changeability. In order to overcome this difficulty, there are several algorithms developed that include edge-detection algorithms to pattern recognition techniques. A survey by Hjelmas and Low (2001) is an attempt, which classified algorithms based on either features or images and presented their technical implementation and performance measures. This work reported comparison of results on the available common datasets, however in lack of standardized tests, this didn't provide a comprehensive comparative evaluation. Another attempt to review face recognition techniques done by Zhao, Chellappa, Philips, and Rosenfeld, (2003), presents a critical survey of still- and video-based face recognition research. Further, this work also covered several insightful analyses in recognition techniques, like psychophysical studies, system evaluation, and issues of illumination and pose variation. The two main problems faced in face recognition are the illumination variances and changes in pose, both at outdoor and indoor situations.

Tan and Triggs (2007) tried to deal with illumination variations while face recognition using local texture descriptors and distance transforms based on similarity metric. This method introduced ternary patterns in facial images to preserve the local details, less sensitive to noise in uniform regions. For experimentation, method tested three standard datasets with the other methods focusing on difficult illumination conditions. A hybrid neural-network model combining local image sampling, a self-organizing map (SOM) neural network, and a convolutional neural network is proposed by Lawrence, Giles, Tsoi, and Back (1997). In the model, sampling provides dimensionality reduction, SOM quantizes image samples, and whereas the larger features are extracted in a hierarchical set by convolutional network. This work claimed to outperform eigenfaces approaches when tested over standard databases, however specifically, it rapidly performs accurate classifications taking only approximately normalized and pre-processed images.

In an extreme implementation, 3D shapes and textures estimation from single images using computer graphical algorithm is designed by Blanz and Vetter, (2003). This work developed a face identification framework, includes a morphable model of 3D faces of images uses statistical fitting for 3D shape and texture estimation. The results show the effectiveness of this framework towards publicly available datasets and it assured in solving pose variations through profile views and frontal views surpassing the illumination effects of cast shadows and specular reflections problems. Timotius, Setyawan, and Febrianto (2010) proposed a combination of kernel principal component analysis (KPCA) and support vector machine (SVM) based face recognition method. In this method, KPCA extracts features and SVM is applied to classify the input images based on these features. The authors observed through a comparison of methods, using KPCA and SVM combined obtained better results than "nearest neighbour classifier, support vector machine, and the combination of kernel principal component analysis and nearest neighbour classifier."

III. PROPOSED HYBRID FACE DETECTION AND RECOGNITION MODEL

This section describes the proposed hybrid face detection and recognition model based on the integral sum of images and template matching, the model diagram is as follows (see figure 1):

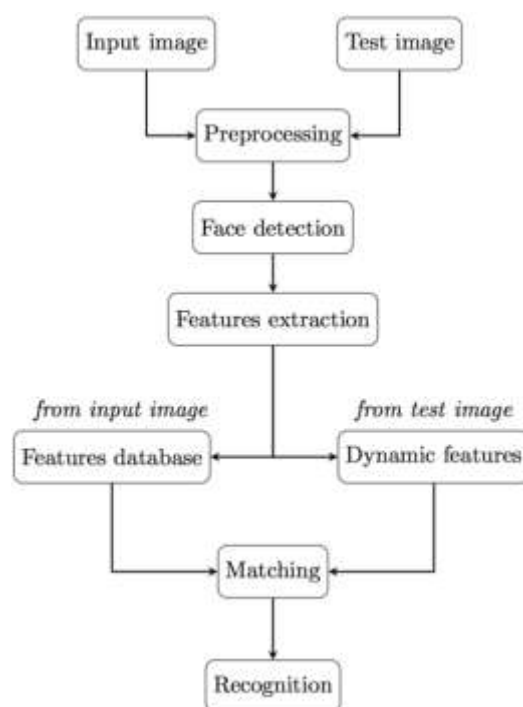


Fig 1. A hybrid human face detection and recognition model

The current research work provides a hybrid face detection and recognition model enabling efficient human facial recognition. This model takes human face images only, detect face features and store in the database. In the recognition phase, a test is given as input and based on the stored features matching is done, if the features are matched then the image is recognized, otherwise, no match is found is displayed.

In the first phase, the proposed hybrid model takes as input the human face image, initially the image is processed for preparing the database. However, in the testing phase, the images are taken as input too, but here the processed results are dynamic and used only in the matching for recognition. The pre-processing step checks for the images format and size to be processed through the model. In this step, images are normalized for making it suitable for face detection step, precisely the nose, mouth, a pair of eyes. The detection phase uses the Viola Jones algorithm in detecting human face using a window scanning through the pixels value over the entire image. The window is capable enough to detect pixels value in the face region distinguishing it from the other regions in the image. Further, this algorithm is tuned to detect facial features like, nose, mouth, a pair of eyes, and the face in whole. The features extraction step uses integral sum of images in distinguishing nose, mouth, and a pair of eyes of a face together and face in whole.

During the training phase, these features are extracted and stored in the database with unique labels assigned to each of the faces and its respective features. Caution is recommended while storing these features in the database, specific labels is to be assigned to a particular facial feature. Practically, in testing of the facial image, an input image called as test image, is passed through the model and steps till feature extraction are followed as it is. Here, the features extracted are referred as dynamic features, since these are not stored in the database, while arbitrarily stored for matching and recognition purposes. Based on the defined matching criteria, the recognition results are obtained, that is, on successful or unsuccessful recognition of a face it produces as matched or failed results as output, respectively.

For facial features extraction the integral sum of images is used over the input image having a human face (Bradley & Roth, 2007; Viola & Jones, 2004). This is employed to extract four basic types of features based on the criteria of easy computation. Rectangular features are computed easily using an intermediate representation for the image called as the integral image. Thus, an integral image at a position (x, y) is the summation of the pixels intensities at (x, y) including both above and to the left of (x, y) , is defined by:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y')$$

where $\overline{ii}(x, y)$ is defined as an integral image and $\overline{i}(x, y)$ is the test image. Using the following pair of recurrences:

$$\overline{s}(x, y) = \overline{s}(x, y - 1) + i(x, y)$$

$$\overline{ii}(x, y) = \overline{ii}(x - 1, y) + \overline{s}(x, y)$$

(where $\overline{s}(x, y)$ is referred as the cumulative row sum, calculating $\overline{s}(x, -1) = 0$, and $\overline{ii}(-1, y) = 0$) over the test image together assist in formulating an integral image in a single stance. This enables the model to extract features from sub-images that is using a window scanning face in an image.

For recognition of the faces based on the stored features in the database to a respective face, the model uses template matching. The template matching procedure is as follows: a pixel in the search image with coordinates (x_s, y_s) has intensity $I_s(x_s, y_s)$ and a pixel in the template coordinates (x_t, y_t) has intensity $I_t(x_t, y_t)$. Thus, the absolute difference in the pixel intensities is defined as $Diff(x_s, y_s, x_t, y_t) = |I_s(x_s, y_s) - I_t(x_t, y_t)|$ (Yuen et al., 2009).

$$SAD(x, y) = \sum_{i=0}^{T_{rows}} \sum_{j=0}^{T_{cols}} Diff(x + i, y + j, i, j)$$

The mathematical representation of the idea about looping through the pixels in the search image as we translate the origin of the template at every pixel and take the \overline{SAD} measure is the following:

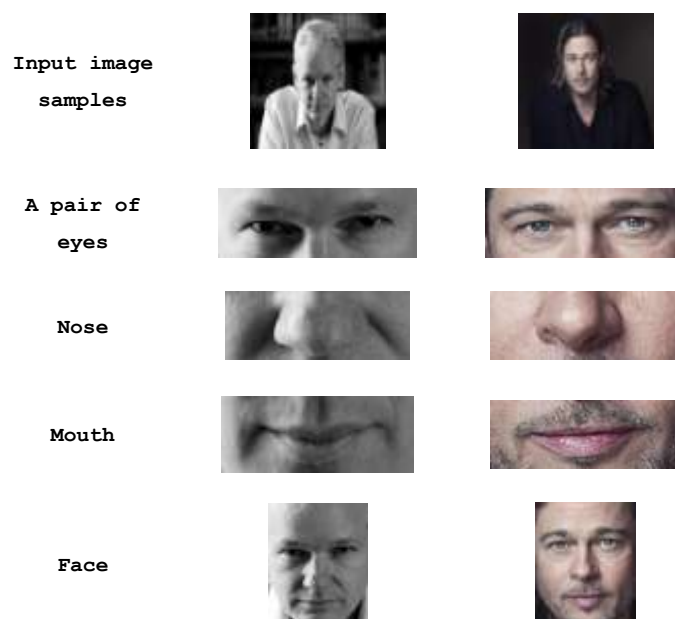
$$\sum_{x=0}^{S_{rows}} \sum_{y=0}^{S_{cols}} SAD(x, y)$$

S_{rows} and S_{cols} denote the rows and columns of the search image and T_{rows} and T_{cols} denote the rows and columns of the template image, respectively. In this method the lowest SAD score gives the estimate for the best position of template within the search image. The method is the simple to implement and understand.

IV. DATABASE PREPARATION AND EXPERIMENTAL RESULTS

The human face features are extracted using the proposed hybrid face detection and recognition model. The images of 50 people producing around 200 facial images in jpeg format are taken from the standard database of ORL website images, used for face detection and recognition. In order to fulfil the matching criterion template matching along with correlation coefficient is employed in the matching and recognition steps. The features namely, mouth, nose, a pair of eyes and face in whole, are detected from an input image and each of the respective feature is stored in database assigned with a unique label, see table 1. In case of a test image, same said features are intended to be extracted, however intentionally, this time these features are not stored in the database. The test image features are matched with the features stored in the database, and based on the matching criterion the recognition results are obtained.

Table1. Sample images stored in the database of the hybrid face recognition model



Based on the recognition results for 200 facial images, it is observed that the hybrid model obtained better results than other methods. The hybrid model used template matching along with correlation coefficient, given by the formula:

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}}$$

where $\bar{A} = \text{mean}(A)$ and $\bar{B} = \text{mean}(B)$, A and B are the training and testing images respectively.

In the research, it is sometimes observed that the images of a same are not matched correctly. Then, this certainly needs attention and features of such images are to be revisited in the extraction steps. The similarity based on template matching and correlation coefficient both are taken in the hybrid; however, it is not sufficient for every facial image taken in the varying conditions. This is left for future improvement of this hybrid model and work is progressing in this direction too.

V. CONCLUSIONS

This paper presented a hybrid face detection and recognition model based on integral sum of images and template matching. For precise recognition results, the model employed correlation coefficient for similarity measures of the facial images. In experiments, 200 facial images of 50 people were used for training and testing purposes. The proposed hybrid model obtained better results than other methods under given similarity criterion. In the future advancements, the varying lighting conditions are the major issues, and currently work is underway to tackle these situations.

REFERENCES

- [1] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
- [2] Y. Taigman, M. Yang, M.A. Ranzato, L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," In *proc. CVPR 2014*. http://www.cv-foundation.org/openaccess/content_cvpr_2014/papers/Taigman_DeepFace_Closing_the_2014_CVPR_paper.pdf
- [3] A. Sharma, M.A. Haj, J. Choi, L.S. Davis, D.W. Jacobs, "Robust pose invariant face recognition using coupled latent space discriminant analysis," *Computer Vision and Image Understanding*, Vol. 116, pp. 1095–1110, 2012.
- [4] P. Viola, M.J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, Vol. 57, pp. 137–154, 2004.

- [5] D. Bradley, G. Roth, "Adaptive thresholding using the integral image," *Journal of Graphics Tools*, Vol. 12, pp. 13–21, 2007.
- [6] E. Hjelmås, B.K. Low, "Face detection: a survey," *Computer Vision and Image Understanding*, Vol. 83, pp. 236–274, 2001.
- [7] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, Vol 35, pp. 399–458, 2003.
- [8] X. Tan, B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *AMFG*, Vol. LNCS 4778, pp. 168–182, 2007.
- [9] S. Lawrence, C.L. Giles, A.C. Tsoi, A.D. Back, "Face recognition: a convolutional neural-network approach," *IEEE Transactions on Neural Networks*, Vol. 8, pp. 98–113, 1997.
- [10] V. Blanz, T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, pp. 1–12, 2003.
- [11] I.K. Timotius, I. Setyawan, A.A. Febrianto, "Face recognition between two person using kernel principal component analysis and support vector machines," *International Journal on Electrical Engineering and Informatics*, Vol. 2, pp. 53–61, 2010.
- [12] C.T. Yuen, M. Rizon, W.S. San, T.C. Seong, "Facial features for template matching based face recognition," *American Journal of Applied Sciences*, Vol. 6, pp. 1897–1901, 2009.