ISSN: 2321-7782 (Online)
Volume 1, Issue 6, November 2013
International Journal of Advance Research in Computer Science and Management Studies
Research Paper
Available online at: www.ijarcsms.com

A new approach to Face Verification: Local Sparse Representation

Shashank .S. Lokhande[1]
Student M. Tech (E & TC)
Dr. Babasaheb Ambedkar Technological University Lonere
Raigad - India

Narendra .S. Jadhav[2]
Assistant Professor, Department of E & TC
Dr. Babasaheb Ambedkar Technological University Lonere
Raigad - India

Abstract: Due to development and advanced techniques, face verification has become an important tool now a day. Since it has been used increasingly in different applications which have given a boost to the research, Different techniques have been developed day by day. Here we are going to present one of the budding technique i.e. sparse techniques which are used now a day. Due to sparse coding, local descriptor has been highlighted i.e. image descriptor formed by summation of sparse coefficient vectors. Different database has been used from previous available tested techniques and possible efforts are made to develop a robust technique for verification.

Keywords: Sparse, Local Descriptor, Database, Robust.

I. Introduction

Face recognition research is known for its different applications in areas such as public security, human computer interaction, finance security etc. Two main tasks involved in face recognition are identification and verification. In identification matching plays a key role. In verification two different types of images of same person are taken and the work is to detect wether the two images belong to the same person or not.

When a pair of face images is gbyethe same personerson to determine wether the image corresponthe same personerson or not, therthe possibilityility of getang an set of image not belonging to the training set under consideration , so no other information is useful for verification instead of this, just a general measures are required to verify. This can be done by determining the feature vectors of both the images and then calculating the distance between the two images which will make verification easier.

Face & fiducial point detection → Pose estimation → Component extraction → Local descriptor → Component representation

Fig.1 Block Diagram of proposed face verification system

As shown in fig 1 from the pair of images given as input the face and fiducial points are detected, then the pose is determined according to the fiducial points. Suppose if the image is projected is different views then it is first projected in the frontal view and then different components like eyes, nose, eyebrow etc. are extracted. Further in local descriptor, the local structure of these extracted components is determined based on the trained dictionary. Finally the similarity between the two descriptors of given image pair is measured.

II. Related Work

Here we review a related work done on face verification that has been evaluated on LFW benchmark and also work carried on with sparse representation. The similarity measure between a pair of images is a key component in face verification. Different approaches such a Patterns of oriented Edge magnitude (POEM), LDML & MKNN, k- Nearest neighbor classifier & cosine similarity metric learning and many more have been used.

In most of the approaches mentioned above require a training set but this imposes restriction on verification since they require a training set so there is need of training free face verification and evaluated in a supervised setting. One of such type of technique is Local Adaptive regression kernel. To increase the roboustness of sparse coding we formulate the local sparse coding based face verification frame work.

III. Proposed Approach

A. Fiducial point detection :

A Discriminative filtering technique using PCA is used to detect the fiducial points. In discriminative filtering filters are desighned using closed – form expression. Using PCA with discriminative filtering does not affect the detection algorithm even if there are small changes in the target patterns. The role of discriminative filter $\Theta_{M\times M}$ is to maximize the energy of a coefficient c(m,n) in a matrix C_{M*M} obtained when the paterrn U_{M*M} is convolved with Θ. Different types of discriminative filters are desighned for each direction of highest energy, from the pattern set we want to detect.The unit used to determine energy of c(m,n) is $DSNR_2$ (Two Dimentional Signal to Noise ratio) given by :

$$(DSNR_2)_{(m,n)} = \frac{c(m,n)^2}{\sum_{i=1}^{M}\sum_{j=1}^{M} c(i,j)^2 - c(m,n)^2} \tag{1}$$

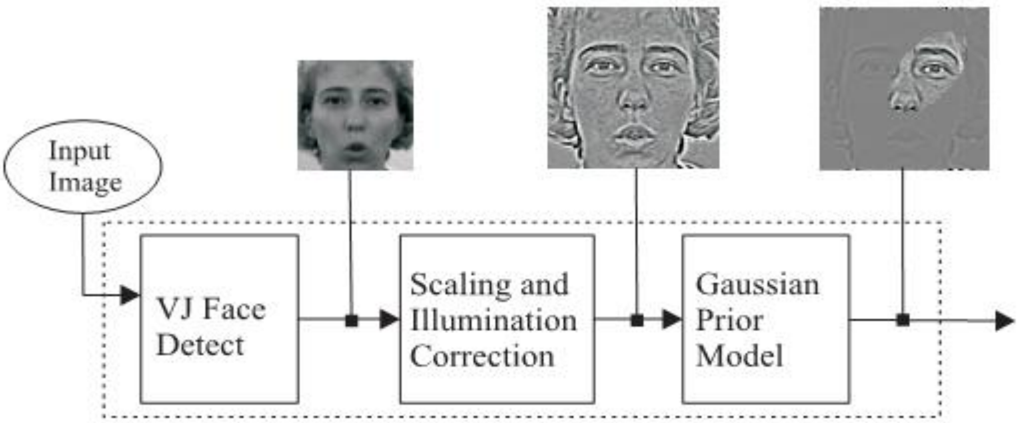

Fig 2. Pre- Processing steps.

As shown in above fig 2 four pre-processing steps are applied to each image. Viola jones face detection algorithm is initially applied to each input image. After that the image is scaled to predefined resolution and then illumination correction is performed. In the final step a Gaussian prior model is used for fiducial points.

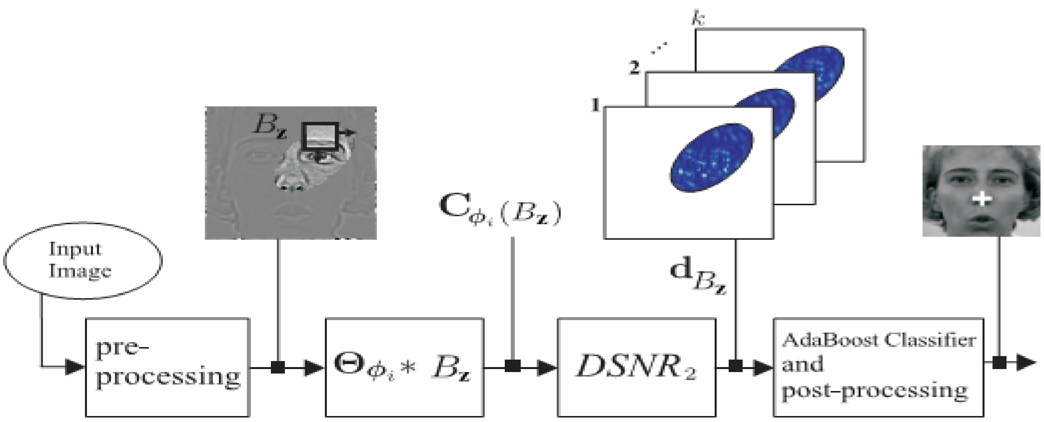

Figure 3

The image after Preprocessing will be processed through a sliding window B_z, where z is the co-ordinate corresponding to the center of the block . $C_{\phi_i}(B_z)$ is obtained by filtering B_z with each Θ_{ϕ_i} where i = {1,……,k}, k being number of principal components. The output of the box $DSNR_2$ provides the $DSNR_2$ (as per eqn 1) . The classifier generated with the AdaBoost algorith will use d_{Bz} to classiby B_z . Finally a post processing step is performed consisting of scaling to original image size and a simple clustering algorithm to group the close labels.

After completing the process on the different images considerating 11 fiducial points True positive (TP) and False Positive (FP) rates were computed. Candidate is a true positive when its distance to the orignal labeling is similar than 10% of the face intra-ocular (between the pupil) distance. Below figure shows the detection of different fiducial points.

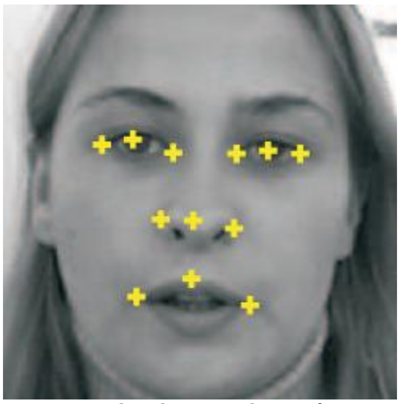

Figure. 4 Fiducial points in human face.

B. Pose Estimation :

One of the main hurdles in face recognition is to obtain robustness to variations in facial pose. Many existing face recognition systems yield good results when comparing faces in frontal pose, but the performance decreases when the faces are in profile. Recently some systems have been proposed which give better result. Ultimately, when given an image of a single face, we will need to estimate all six pose parameters: x and y location, scale, yaw (rotation around the neck, from the left profile to right profile), pitch (rotation up and down), and roll (rotation in the image plane). A seventh parameter, focal-length of the camera lens, may or may not be required.

A sparse representation has been designed for human faces, which captures the unique signatures of a human face effectively, while facilitating the estimation of the head position and pose. The representation is a collection of projections to a number of randomly generated possible configurations of the human face. Each projection corresponds to a pose of the head along with a configuration of its facial features. The projections should respond to change in pose and feature configuration, while largely ignoring other image variations.

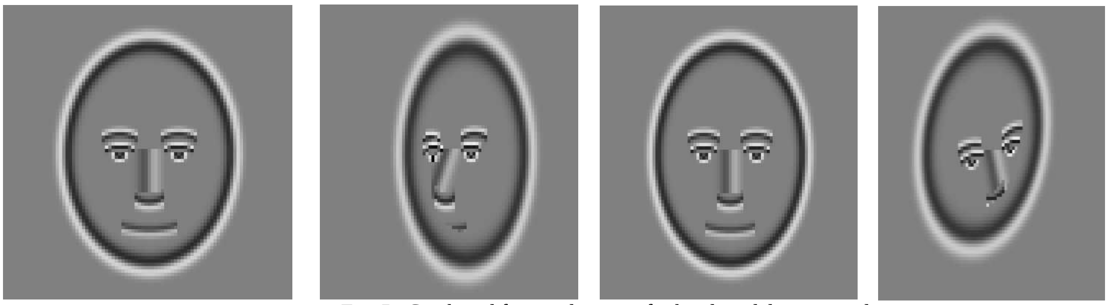

Fig. 5. Combined feature detectors for head models in several pose

A simple parametric model of a head is used, in which a parameter vector determines the head shape and the locations of facial features upon it. For a given pose and parameter vector, we predict the locations of 10 straight and curved edge gradients in the image, and compute a specific filter for each. The correlation between the image and this filter gives a measure of how likely the image is to contain that edge, and therefore how likely it is to match a face with the given pose and model parameters. Some sets of 10 filters generated using the 3D model of the head and perspective projection are shown in Figure 5.

A large number of samples $\{X^n | n = 1, 2, \cdots, N\}$ that represent the pose of the model and the position and shapes of the facial features are generated using support vector regression (SVR). Each vector X_n then constructs the set of shape filters that will compute the image responses

$$R_n = \{eyel_n, eyer_n, brol_n, bror_n, irisl_n, irisr_n, nose_n, \; noseprofile_n, mouth_n, head_n\} \quad (2)$$

Where each of eyeln , eyern , etc. is a 2D pattern that matches one predicted edge gradient. There are 10 predicted edges, so we obtain 10N filters. Note that a filter matched to the head boundary (to yield the response headn) is also used to compare the relative positions of the features to the head. Computing the correlation between an image and all these 10N filters is a linear transformation. We found, however, that the absolute values of these correlations produced better pose estimates.

The performance of SVR applied to our sparse representations is clearly superior to SVR applied tothe raw pixel dataa. With sparse representations, we successfully estimate poses to within 15◦of the annotation in about 90% of the natural images. Without sparse representations, the success rate at this error tolerance is about 13% worse. Below fig shows the examples of estimated poses.

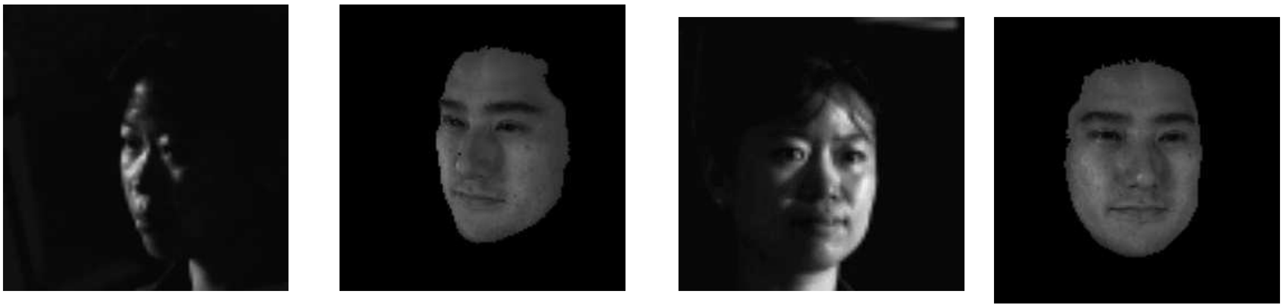

Fig. 6. Examples of estimated poses

C. Component and Feature Extraction :

After cropping and resizing the faces, each sample is decomposed into blocks and then a set of low level features descriptors are extracted from each block. Various methods are used to capture different features. Histogram of oriented gradients (HOG) captures information related to shape, Local Binary pattern is used to compute texture, intensity calculates colour information and sailent visual properties are evaluated by Gabor filters.

D. Local Sparse representation :

Descriptor was developed by applying random projection tree and PCA to encode the normalized feature vectors into discrete codes. The descriptor when used in face verification yield good results so after studying this descriptor we have developed a novel descriptor based on sparse coding of local patches by using learned dictionary. In order to obtain representative features of the face, we first learn a dictionary from the collected facial images. For each image, the face region is cropped. A sliding window moves on the facial region to sample the local patches. Normalization is performed for every cropped patch.

There are several choices for patch normalization: mean subtraction with unit vector normalization, mean subtraction only, unit vector normalization only and no normalization. The last column of Fig. 7 shows the learned dictionaries with different normalization schemes. We can see that the learned dictionary is composed by some local structure and texture pat- terns. The reconstruction images from different normalization schemes are depicted in the first column of Fig. 7. One can notice that since the normalization is performed on each patch, some local structure may be amplified in homogeneous area. The first column is the reconstructed images by mean subtraction with unit-vector normalization, mean subtraction only, no normalization and unit-vector normalization only, respectively in each row The 2nd to 4th columns are sparse coefficient vectors of three sequential patches

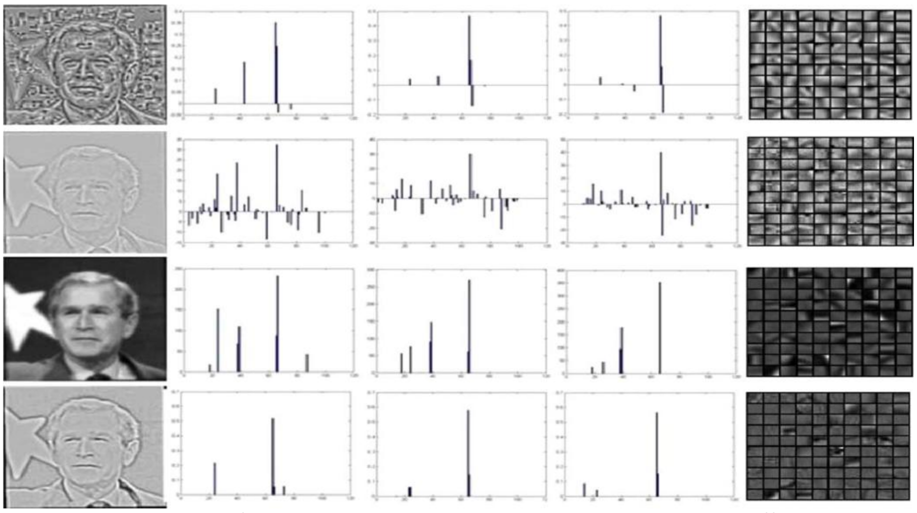

Fig. 7. Examples of the reconstructed images and the corresponding sparse coefficients

E. Local descriptor and component representation:

In the feature representation, we first partition the face image into several components, like the eyes, nose, eyebrow, etc. Each component is then divided into sub-blocks. The sparse coefficients are computed for every patch in a sliding window manner from each sub-block. Since the sparse coefficients form a vector, not like LBP which is coded in a specific value, we have tried several ways to count the contribution of each dictionary atom in a sub-block. Intuitively, the coefficient vectors of all patches in the sub-block can be summed up directly. It is more reliable to represent the feature by counting the occurrence of the dictionary atoms used in the linear combination. This is because the variations from the values of sparse coefficients are larger than the occurrence of the dictionary atoms used for neighboring patches. This gives the histogram of the occurrence of the dictionary atoms. Here, such histogram representation is exploited and referred to as a Feature Vector (FV). Once we have the FVs from all sub-blocks, all FVs are concatenated as a *Local Descriptor* for one face component. The flow is shown in the bottom half of Fig. 8. We selected 9 components in face images, including two eyes, two eyebrows, two cheek, nose, mouth and forehead. The positions of these components are determined by the fiducial points with predefined settings.

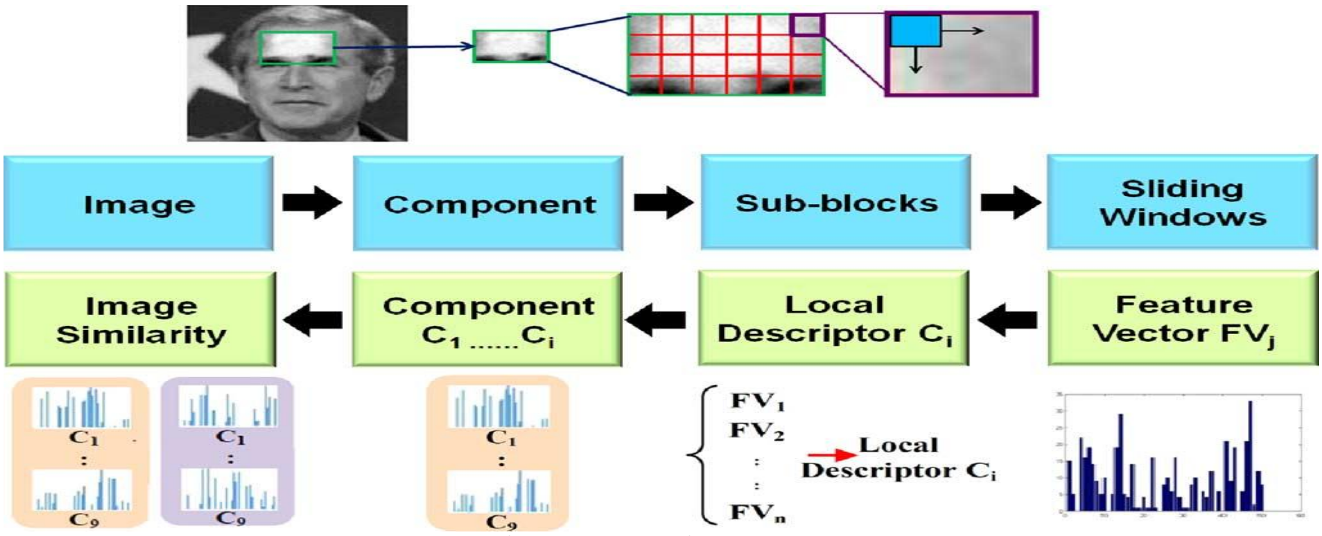

Fig. 8. Feature representation obtained from the hierarchy of the whole image, component level to sliding window.

F. LFW dataset:

The LFW (Labeled Faces in the Wild) data set contains large amount of face images of different people, including pose, occlusion, resolution and illumination. The LFW dataset provides sets for validation: development set and testing set. The experiments include feature representation at component level and for the whole face.

IV. EXPERIMENTAL RESULTS

We provide experimental results on different datasets. First, extensive experiments were conducted to find the parameter set. In all of our experiments, the face images were scaled in a 100 by 100 bounding box. Images were aligned by three feature points (eye corners and the nose tip). We scaled the distance of two outer eye corners to 65 pixels and rotated the image to align the two eye corners to be horizontal Using PCA the accuracy can be improved by 3 to 5 % whereas by WPCA 8 to 10 % improvement is achieved over original feature Finally, we evaluated the accuracy of the proposed face verification system on the very challenging LFW benchmark. The final results of the validation of the LFW testing set is shown in Table I.

TABLE I
FACE VERIFICATION ACCURACIES BY USING DIFFERENT DESCRIPTORS ON LFW TESTING DATASET

Method	Orignal	PCA	WPCA
LBP	0.697	0.748	0.795
LSR_W	0.712	0.757	0.818
LSR_O	0.702	0.725	0.815

V. CONCLUSION

Here we proposed a novel feature representation using learned dictionary to describe the local image characteristics with the sparse representation. The sparse coefficients from local patches are fused as a histogram to describe the local structure and characteristic of face components. Different techniques have been used that are flexible to use to make the approach easy and interesting.

Acknowledgement

I would like to thank all of them who have directly or indirectly in helped me and guided me at each step in making this paper.

References

1. Guo, R. Wang, J. Choi, and L. S. Davis, "Face verification using sparse representation," in IEEE Conf. Computer Vision and Pattern Recognition, Workshp on Biomet. (CVPRW), 2012.
2. Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face recognition with learning- based descriptor," in CVPR, 2010, pp. 2707–2714.
3. "Estimating Facial Pose from a Sparse Representation " by hanky Moon and Matt L. Miller.
4. V. Kruger, S. Bruns, and G. Sommer. "Efficient head pose estimation with gabor wavelet networks". In British Ma- chine Vision Conference, 2000.
5. J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation", IEEE Trans. PAMI, 31(2) : 210-227, 2009, 1,2,3,4.
6. Yalda Amidi, Mohammad Tagi Sadegi "Face verification using Sparse Representation Techniques" in 6th International Symposium on Telecommunications (IST'2012).

AUTHOR(S) PROFILE

Mr. Shashank S. Lokhande, received B.Tech degree in Electronics & Telecommunication from Department of Technology, Shivaji University, Kolhapur. Presently he is pursuing his M. Tech in Dr. Babasaheb Ambedkar Technological University, Lonere, Raigad.

Prof. Narendra Jadhav, received the B.Tech and M.Tech degree from the Dr. BATU Lonere, Raigad Maharashtra, INDIA, in 2004, in Electronics and Telecommunication engineering. His research interests include gesture recognition, artificial intelligence, and FPGA System in Signal Processing