Head Pose Estimation for recognizing Face Images using Collaborative Representation based Classification

Srija Chowdhury and Jaya Sil
Computer Science and Technology,IIEST,Shibpur

***Abstract*—Real time face recognition is challenging due to time taken for searching the test image within a wide variation of training images. We propose an efficient face recognition method by applying collaborative repersentation based classification (CRC) technique in two steps. Using CRC first we select the images closer to the pose of the test image compare to others in the training set. The test image is then searched among the selected training images only, instead of all, thus reducing the search space and time. For person recognition we calculate Gabor wavelets of the eye regions of the test and the selected training images as the basis functions containing detail edge information. The images are reconstructed as a linear span of the basis vectors using CRC and the sparse coefficient vectors are used as feature vectors. To obtain the best match image we apply $l2norm$ on the feature vectors corresponding to the test and the selected images. The proposed head pose estimation based face recognition method has been validated using PIE and Head Pose databases and obtain comparable accuracy with other methods at a much lesser storage and time complexity.**

***Index Terms*—Sparsity, Face Recognition, Pose Estimation**

I. Introduction

Head pose estimation and person identification has immense scope in real time applications like surveillance systems. However, the existing techniques of face recognition [1], [2], [3], [4] are hardly effective for real time applications due to large search space and time complexity. Dimension reduction based face recognition methods include principal component analysis (PCA), linear discriminant analysis (LDA), Independent Component Analysis (ICA), Multi Dimensional Scaling (MDS) and Isomap [5] have been applied over the periods by the researchers to extract important features only in order to reduve the system complexity. However, PCA [5], and LDA though effective for dimension reduction, not very suitable for face recognition of pose variant images. MDS cannot capture non-linearity in feature space whereas Isomap is not efficient for new data points. In Kernel PCA [6] based methods, choosing the kernel is a major problem. Face recognition under different lightning conditions, sizes and positions [7], [8] is also challenging. However, in all feature based face recognition methods, accuracy depends on the feature descriptors [9]. Choosing the right feature descriptor is a major task which depends on the application or the problem we are dealing with. In collaborative representation based classification (CRC) [10], a linear combination of the training samples are used to represent the query sample. However, in this method a large number of training samples are required to properly represent the query sample. With higher sampling rate we can obtain sufficient number of pose variant images from the surveillance camera and apply CRC for person identification in real time. The aim of the paper is to recognize the persons from their pose variant images in a reduced search space without sacrificing accuracy. The training images are assigned apriori into different class labels based on the head pose measurement. We apply CRC to identify the selected head pose images in the training set closest to the test image by utilizing sparsity in the context of the pose variation in images. For person recognition the eye region of the test and the selected images are cropped because eye is the most informative feature for face recognition. We obtain forty Gabor filter responses for each eye region as the basis functions(or dictionary) to obtain the sparse coefficient vector, chosen as feature vector of size $[40x1]$. The best match query image was found by applying *l2-norm* between the feature vectors of the test and the training images. The main contribution of the paper is to reduce the search space containing only nearest pose training group instead of the entire database. For person identification instead of creating handcrafted features we use sparse coefficients as feature vectors using rotation invariant Gabor filters as basis functions. These two factors account for the importance of the proposed method in face recognition at a very low complexity. The proposed head pose estimation based face recognition method on PIE and Head Pose databases and obtain comparable accuracy with other methods.

Rest of the paper is organized as follows- *Section 2* presents collaborative representation based head pose estimation and face recognition, *Section 3* deals with results and discussions while conclusions are arrived at *Section 4*.

II. Face Recognition:using CRC

Classification is the problem of identifying the class of a query sample on the basis of the training instances having different class labels assigned apriori. The dataset contains images of various persons labelled according to the poses and we have manually grouped the images according to poses. However, the chance of belongingness of the query sample into all training classes is not equal, therefore, searching the whole training data set is unnecessary. The problem becomes critical in real time computer vision applications where time is a major factor.

978-1-5090-2029-4/16/$31.00 @2016 IEEE

A. Collaborative Representation Based Classification

Say, there are k classes of instances represented by the matrix A = [A_1 A_2 ... A_k] where $A_i \epsilon \Re^m$ (i = 1 . . . k). In sparse representation based classifier (SRC), the instance of $i-th$ class is represented by the vector A_i and a query sample $y \epsilon \Re^m$ is coded as $y = A\alpha$, where α=[α_1 α_2...α_k] is the sparse coding vector and α_i is associated with class i. We can classify the query sample y more accurately into a particular class compared to the samples of other classes. However, in order to represent the class of the query sample accurately, there must be enough training samples of each class.

Say, we have m training classes and n instances in each class. The training set of instances is represented as a matrix A, where A_{ij} represents *j-th* training example of *i-th* class containing mxn training samples of different classes. Using CRC approach we can reconstruct a query sample y by a linear combination of the training samples, as given in *equation* (1),

$$y = AX \tag{1}$$

where X is the coefficient vector, each element of which represents contribution of the training samples in a particular class for reconstructing the query sample.

The sparsity in coding reduces space complexity using the same principle of SRC while representing the query sample in terms of the coefficient vector X, (evaluated using suitable-norm) [11]. *Equation* (2) is used to obtain the coefficients x_i in vector X for each training sample.

$$X = (A^T A)^{-1} A^T y \tag{2}$$

For complete A, we can faithfully represent any query sample y using A. In general, more is the number of training samples, better reconstruction of the query sample.

B. Pose Estimation

We apply CRC approach to identify the head pose image in the training set close to the test image. First we estimate pose of the test image either in absolute value or with respect to the nearest head pose images in the training samples. Suppose the test images and the training images are of size pxq and there are n training images in each of m classes. Therefore, each class represents a head pose and n number of different persons in each class constitute the training set. If mxn is sufficiently large, we can faithfully recontruct the test sample using CRC method.

For pose estimation we need sufficient number of poses with large variability in a complete training set considering individuals. When large number of training samples having close variation in poses (not more than 15 degrees) are available, CRC can recognise the person faithfully. As the gap between two successive poses of the same person increases, their similarity decreases and recognition of the person becomes difficult. In real time, we might not always have many closely oriented poses of the same person in the training set. So, we build a over complete training set consisting of pose variant images of different persons and as a first step identify the images having same or nearest pose to that of the test image. In the proposed method, the training images are labelled with respect to (w.r.t.) the pan angle. For a given test image, the contribution of each class specific images is calculated using *equation* (2) and select the training images from each class contributes most compare to others in terms of estimated pose of the test image. *Fig. 1* shows the test images and the corresponding image in the training set contributes most for estimating the pose of the test image.

Fig. 1: **(*center*)Test image($0°$) and few corresponding training images from best match head pose training set($+22.5°$) in CMU PIE (average pose estimation error=** $6°$**)**

C. Person Recognition

1) Gabor Filter: Complex Gabor function was first introduced by Dennis Gabor[12] where he proposed a *"Quantum Principle"* for information, similar to *Heisenberg's Uncertainity Principle* in Quantum Mechanics. The work of Gabor was extended in 2D form by J.G.Daughman, to represent filters with an optimal localization in 2D spatial and frequency domain[13]. This band limited signal extracts multi-resolutional, spatially located features of a confined frequency band. Forty Gabor Filters are chosen for the proposed method, shown in *Fig. 2*.

2) Feature Vectors: For person recognition, eye region of the test image and training images in the nearest pose group are cropped out. Experimentally we consider forty Gabor Filters ($G_1, G_2, ..., G_{40}$), collectively called as G(say) to act as basis functions and linearly span the cropped out images. Gabor filter responses are used to extract detail of edge information from the images. The coding of an image into the dictionary of Gabor Filter is obtained using CRC and the sparse coefficients (*equation 2*) serve as the feature vector for the image. The

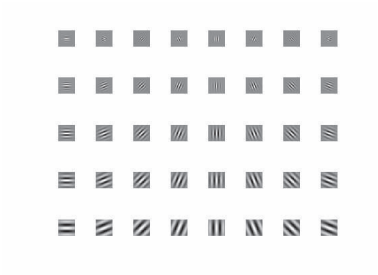

Fig. 2: **Real parts of Gabor Filters at varying scale and frequency**

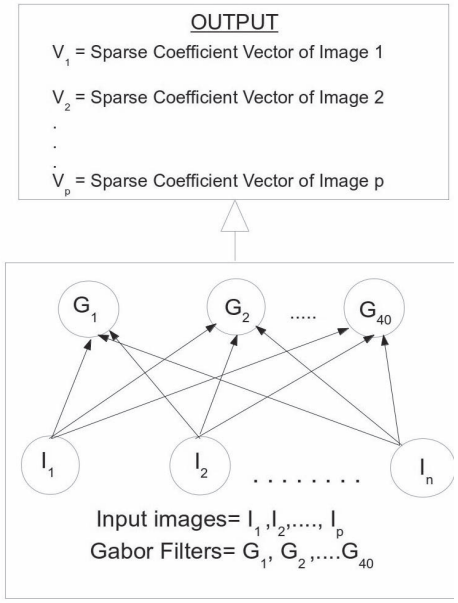

Fig. 3: **Sparse Vector Formation from input Images**

feature vectors (V_1, V_2,...) for the training images (I_1,I_2,...) and V_{test} for the test image (say I_{test}) are generated and the best match is found using *l2-norm*. Best match image is obtained using equation (3) and the procedure is depicted in *Fig. (3)*.

$$Best\ Match = arg\ min_i ||V_i - V_{test}||_2 \tag{3}$$
$$V_i = (G^T G)^{-1} G^T I_i$$
$$V_{test} = (G^T G)^{-1} G^T I_{test}$$

The entire process is depicted in *Algorithm 1*

ALGORITHM 1: PERSON RECOGNITION

POSE ESTIMATION Input: Test images and all training images
Procedure: For each test image apply CRC using the training images as the dictionary
OUTPUT: Nearest Pose Group (Training set images at nearest head pose)

PERSON RECOGNITION INPUT: Nearest Pose group pf that test image as obtained from POSE ESTIMATION
Step 1: Say nearest pose group of a test image (I_{test}) has $'p'$ training images (I_1, I_2..., I_p).
Step 2: Crop out eye region of the test image as well as of the selected training images. All cropped out regions are of size 1600x1
Step 3: Code each cropped out image into the dictionary of 40 Gabor filters (dictionary size = 1600x40)
Step 4: Use CRC to get 40x1 coefficient vector (fearure vector) for each training image in the nearest pose group (say, V_1, V_2,.. ,V_t) as well for the test image (say, V_{test}).
Step 5: Classify the test image (for person recognition) using *equation (3)*.
OUTPUT: Best Match among selected images (Recognised Person)

III. RESULTS AND DISCUSSIONS

A. Databases

For experimentation two databases are built using CMU PIE [14] and Head Pose image database [15]. CMU PIE image database consists of 43168 images of 68 persons taken under 13 different poses and 43 different illuminations. From *Database 1*, one randomly chosen pose of each of the persons are in the test set and the training set comprises of the remaining poses of each person. The images are grouped according to poses for the experiment. Our method works well if the test image pan varies from $-67.5°$ to $+67.5.°$ *Fig. 4* depicts different pose variant images of an individual such as $-67.5°$, $-45°$, $-22.5°$, $0°$, $22.5°$, $45°$ and $67.5°$. In the experiment, the training and test sets are absolutely different with respect to both pose and person, achieving more than 90% accuracy using the proposed method.
Database 2 is built from Head Pose image database [15] which consists of 2790 images of 15 persons with pan and tilt varying from $-90°$ to $+90°$. Each person has two sets of images for each pose, varying in either expression or presence (absence) of spectacles. Only one set is chosen for the experiment, part of it shown in *Fig. 5*. *Database 2* is built by taking 91 images of each of the 15 persons with pan varying from $-90°$ to $+90°$ ($15°$ difference between each pose) in each level of tilt ($-60°$, $-30°$, $-15°$, $0°$, $+15°$, $+30°$ and $+60°$). Ten fold cross-validation technique has been applied on *Database 2* with varying total number of images (training and test together), as shown in *table I*. *Table 1* also shows the pan error, tilt

Fig. 4: **Pose Variation from -67.5° to $+67.5^\circ$ in Database 1 with a gap of 22.5° between successive images of a particular person**

Fig. 5: **Different Pose Variant Images in *Database 2***

TABLE I: **Recognition accuracy by pose estimation method with varied pan and tilt error using *Database 2***

No. of persons	Num of images	pan error	tilt error	Accu racy
6	546	5°	10°	95 %
8	728	6°	11°	92 %
10	910	6°	12°	85 %
12	1092	8°	15°	82 %
15	1365	10°	16°	80 %

error and person recogition accuracy with varying number of images of different persons. Due to closeness in pose variant images, head pose estimation procedure alone identifies the best match training image to recognize the face of the test image. However, in general performance improves if we apply the person identification method after pose estimation. For example in PIE database, due to less variation in the number of poses as well as illumination problem, head pose estimation method is unable to recognise the test images. So we apply the proposed person recognition method after pose estimation of test images.

B. Comparisons

Comparison of recognition accuracy of the proposed method with other existing face recognition methods[3], [4] are listed in *table 2* while ROC curves are shown in *Fig. 6* and *Fig. 7*. Tables and figures reveal that our method gives better recognition rate in a reduced search space in comparison to other methods and classifiers.

TABLE II: **Comparison of methods using CMU-PIE database**

Methods	Accuracy(%)
PCA	58
Gabor PCA	54.5
LDA	59.5
Gabor LDA	65.5
ICA	59
Gabor Supervised LPP	74.3
Global DCT	44.1
Local DCT + Feature Fusion	70.9
Local DCT + Decision Fusion	68.5
NN	78.8
LRC	81.9
SVM	78.1
S-SRC	90.0
CRC-RLS	89.3
Proposed Method	**91.43**

IV. Conclusions

The proposed CRC based person identification method can recognise faces in a much reduced space compare to others. It has been proved that more the number of training images, higher is the recognition accuracy since more number of images of a person reconstruct the face of the person more accurately. The proposed method is efficient and robust. The proposed method works well when number of pose variant images is less in a database and also when there is contrast or illumination problem. For handling illumination problem, histogram equalization has been applied. The method works better if there is a wide variation in headpose in the dataset. If number of test images is m and number of training images in nearest pose training set is n and that in entire Dataset is N, then complexity of searching in proposed method = $O(n)$ where $n << N$. Our future work comprises of using sparse Deep Boltzman Machine (DBM) or sparse Convolutional

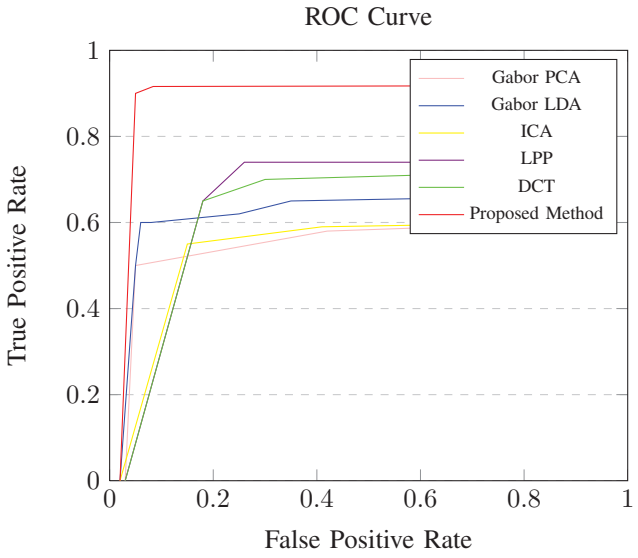

Fig. 6: **Comparison of ROC using *Database 1***

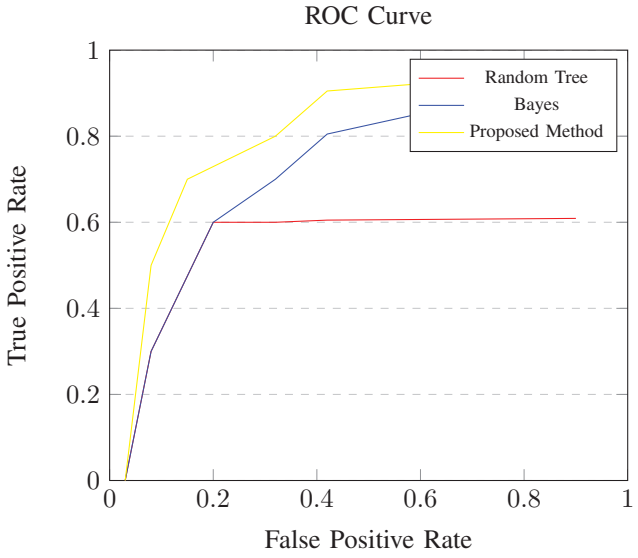

Fig. 7: **Comparison of ROC using *Database 2***

Neural Network (CNN) in a reduced search space for face recognition.

REFERENCES

[1] X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recognition*, vol. 42, no. 11, pp. 2876–2896, 2009.

[2] S. Chitra and D. G. Balakrishnan, "A survey of face recognition on feature extraction process of dimensionality reduction techniques," *Journal of Theoretical and Applied Information Technology*, vol. 36, no. 1, pp. 92–100, 2012.

[3] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," *Image Processing, IEEE Transactions on*, vol. 16, no. 7, pp. 1716–1725, 2007.

[4] Z. Zheng, F. Yang, W. Tan, J. Jia, and J. Yang, "Gabor feature-based face recognition using supervised locality preserving projection," *Signal Processing*, vol. 87, no. 10, pp. 2473–2483, 2007.

[5] S. Singh, M. Sharma, and N. S. Rao, "Accurate face recognition using pca and lda," in *International Conference on Emerging Trends in Computer and Image Processing*. Citeseer, 2011, pp. 62–68.

[6] M.-H. Yang, "Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods," in *fgr*. IEEE, 2002, p. 0215.

[7] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[8] F. Gianfelici, C. Turchetti, and P. Crippa, "A non-probabilistic recognizer of stochastic signals based on klt," *Signal Processing*, vol. 89, no. 4, pp. 422–437, 2009.

[9] C. Geng and X. Jiang, "Face recognition using sift features," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE, 2009, pp. 3313–3316.

[10] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang, "Collaborative representation based classification for face recognition," *arXiv preprint arXiv:1204.2358*, 2012.

[11] X. Li, Y. Pang, and Y. Yuan, "L1-norm-based 2dpca." *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics: a publication of the IEEE Systems, Man, and Cybernetics Society*, vol. 40, no. 4, pp. 1170–1175, 2010.

[12] D. Gabor, "Theory of communication. part 1: The analysis of information," *Electrical Engineers-Part III: Radio and Communication Engineering, Journal of the Institution of*, vol. 93, no. 26, pp. 429–441, 1946.

[13] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *JOSA A*, vol. 2, no. 7, pp. 1160–1169, 1985.

[14] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*. IEEE, 2002, pp. 46–51.

[15] N. Gourier, D. Hall, and J. L. Crowley, "Estimating face orientation from robust detection of salient facial structures," in *FG Net Workshop on Visual Observation of Deictic Gestures*. FGnet (IST–2000–26434) Cambridge, UK, 2004, pp. 1–9.