

Pose Invariant Face Recognition Using Neuro-Biologically Inspired Features

Pramod Kumar Pisharady and Martin Sauerbeck

Abstract—A face recognition algorithm using biologically inspired C_2 standard model features is proposed. The C_2 feature extraction system, which is based on the standard model of the ventral stream of visual cortex, is modified for the extraction of facial features. The S_2 units in the learning stage are tuned to frontal and profile views of the faces to provide pose invariant recognition. A new learning function is utilized to extract discriminative facial features. The spatial invariance is limited both in the local and global pooling stages. The reduced spatial invariance in the local pooling preserves the minute shape and texture details of the facial components, whereas that in the global pooling stage preserves the component location information. The faces are classified using a support vector machines classifier with linear kernel. The proposed algorithm is tested using two face datasets (MIT-CBCL, FEI) with wide variations in profiles, and compared with existing algorithms (the original C_2 feature extraction algorithm and MPCALDA algorithm). The experimental results provide evidence of the better performance of the proposed approach for pose invariant face recognition, by at least 6.5%.

Index Terms—Face recognition, pose invariance, human-computer interaction, pattern recognition, biologically inspired features, C_2 features.

I. INTRODUCTION

Automatic recognition of human faces has been an active research area for the last two decades. Unresolved challenges in this field still keep face recognition a hot area of research. The research efforts in face processing include face detection, face recognition, face tracking, pose estimation, and expression recognition [1]. The major recognition methods applied to face images are eigenface, neural network, dynamic link architecture, hidden Markov model, geometrical feature matching, and template matching. The surveys on face recognition [2] and face detection [1] provide the details of these methods.

The mainstream computer vision research has always been challenged by human vision, and the mechanism of human visual system is yet to be understood well. The human visual system rapidly and effortlessly recognizes a large number of diverse objects in cluttered, natural scenes and identifies specific patterns, which inspired the development of computational models of biological vision systems [3–6]. A case study in unconstrained face recognition on Facebook [7] shows the utility of biologically inspired features for face recognition. However the models in the case study

encapsulate the properties of only visual area V1 and the representations utilized are not optimized or modified specifically for face recognition.

Human faces are characterized with specific shapes and textures. The addition of skin texture features with shape features improves the performance of face recognition algorithms. The C_2 features proposed by Serre *et al.* [4], [5] have the capability to capture the shape and texture features in spite of the position and scale variations of objects. The model encapsulates the properties of visual areas V1, V4 and inferotemporal cortex. The utility of C_2 features in face and expression recognition is less studied [8], [9]. Singh *et al.* [8] utilized the C_2 features with slight modification (used log polar Gabor filter instead of the normal Gabor filter) for extracting features of a mosaic of face images. The focus of the algorithm presented by Pramod *et al.* [9] is on the classification aspect. The C_2 feature extraction system is utilized as it is in their study.

The present paper investigates possible modifications of the C_2 feature extraction system to make it suitable for face recognition applications. It proposes a modified C_2 feature extraction system for pose invariant face recognition. A support vector machines (SVM) classifier with linear kernel is utilized to classify the faces using the extracted modified C_2 features. The performance of the proposed algorithm is tested and compared using two face datasets, the MIT-CBCL face recognition database [10] and the FEI face database [11].

II. C_2 FEATURES FOR POSE INVARIANT FACE RECOGNITION

This section briefly reviews the C_2 feature extraction system and proposes modifications for using C_2 features for pose invariant face recognition.

A. C_2 feature Extraction System

Riesenhuber and Poggio proposed a hierarchical model of ventral visual object-processing stream in the visual cortex [3]. Serre *et al.* implemented a computational model of the system, and utilized it for robust object recognition [4], [5]. The features extracted by this model are known as C_2 standard model features (SMFs).

The C_2 feature extraction system consists of four layers.

Layer 1: Layer 1 (S_1) consists of a battery of Gabor filters with 12 orientations (0° to 175° , in steps of 15°) and 16 sizes (divided into 8 bands). The S_1 layer imitates the simple cells in the primary visual cortex (V1), detecting edges and bars at different orientations (Fig. 1).

Layer 2: Layer 2 (C_1) models the complex cells in V1, by applying a *MAX* operator locally (over different scales and

Manuscript received April 7, 2012; revised May 21, 2012.

The author is with the Institute of High Performance Computing, A*STAR, Singapore (e-mail: pramodkumarpisharady@gmail.com, Tel.: +65 6419 1206, fax: +65 6463 1452).

positions) to the first layer's outputs. This operation provides tolerance to different object projection sizes, positions, and rotations in the 2-D plane of the visual field.

Layer 3: In layer 3 (S_2), radial basis functions (RBFs) are utilized to imitate the visual area V4 and posterior inferotemporal (PIT) cortex. Layer 3 aids shape and texture recognition by comparing the C_1 images with prototypical C_1 image patches. The prototypical C_1 image patches (the

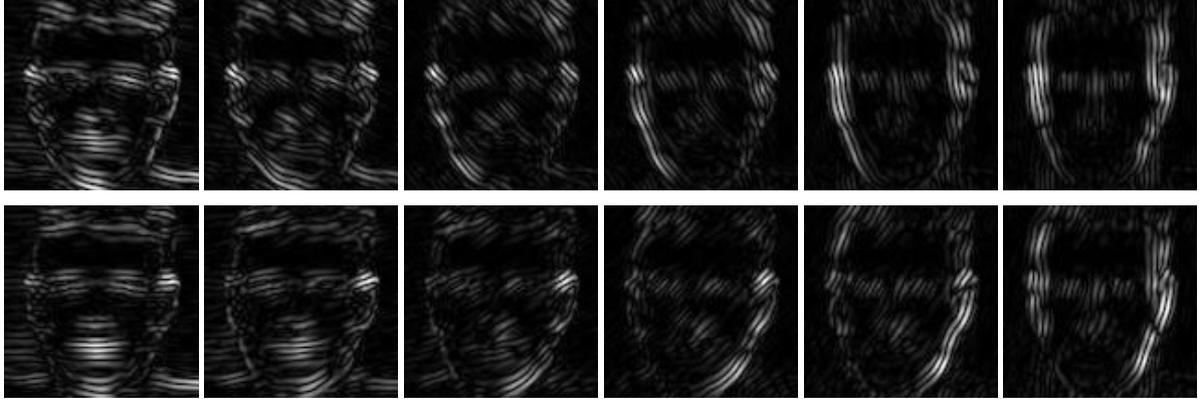


Fig. 1. The S_1 responses of a face image at different orientations. The orientations vary from 0° (left top) to 175° (left bottom) in steps of 15° , increasing in clockwise direction.

B. Proposed Modifications

The proposed modifications in the C_2 feature extraction system for pose invariant face recognition are described in this sub-section.

1) Limiting the position invariance

The local pooling over position in the C_1 layer provides invariance to position. However some useful information (in shape, texture, position) is lost due to this local maximization. This is significant in the context of face recognition as the basic shape of different faces are similar, and the major differences are in the shapes and relative positions of minute facial components (e.g. eyes, nose, mouth). In order to avoid this loss, the local invariance is limited to 8×8 units (at the highest scale, instead of 22×22 in the original system).

2) Tuning of the S_2 units for the recognition of frontal and profile faces

Simple cells in the RBF stage (layer 3, S_2) combine bars and edges in the image to more complex shapes. Each S_2 unit response depends on the Euclidean distance between crops of the C_1 image and the stored prototype patch. In order to recognize the faces in spite of the profile variations, each S_2 unit is tuned to different views of the same face, the frontal, left profile, and the right profile views. A battery of prototype patches with 3 layers (corresponding to the 3 views) is extracted from the geometrically significant and textured positions of the face images¹ (Fig. 2). The C_2 features corresponding to all 3 layers are extracted and the maximum of these values (1) are utilized for the recognition².

$$C_2 = \max_i \{C_2^i\}, \quad (1)$$

where C_2 represents the C_2 feature component corresponding

¹ Images shown are taken from MIT-CBCL face database [10].

² The proposed algorithm utilizes the edge/shape information (C_1 response) from different profile views. The pose invariance is achieved by finding the best similarity.

prototype patches) are learned and stored during the training (in humans, these patches correspond to learned patterns of previously seen visual images and are stored in the synaptic weights of the neural cells).

Layer 4: The fourth layer (C_2) imitates the inferotemporal cortex. It applies a *MAX* operator globally to the outputs of layer S_2 , resulting in the C_2 feature representation that expresses the best similarities with the prototype patches.

to a battery of patches, and C_2^i represents that corresponding to a layer in the battery.

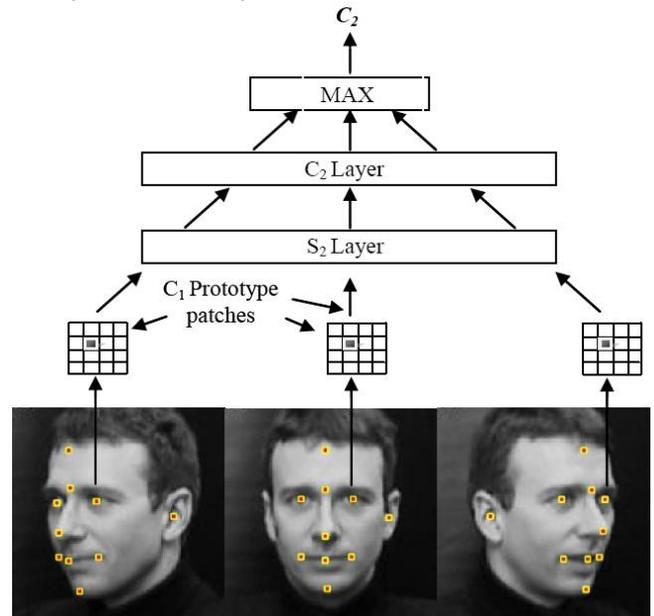


Fig. 2. Three layers of prototype patches are extracted from three different views (frontal, left and right profiles) of a person's face. Each layer consists of patches extracted around the geometrically significant and textured positions of the face as shown (the patches are extracted from the C_1 response of input images). The maximum of the responses forms a component of the final C_2 feature vector. If a particular facial part is not visible in a view (say left ear is not visible in right profile view) the corresponding patch is extracted from the symmetrically opposite side.

3) Discriminative learning function

The learning function (2) is utilized in earlier approaches [4, 12], where the objective is to recognize general objects.

$$R(X, P) = \exp\left(-\frac{\|X - P\|^2}{2\sigma^2}\right),$$

where,

$R(X, P)$ the S_2 response corresponding to the C_1 image

patch X and prototype patch P ,

$\|X - P\|$ the Euclidian distance (which is a measure of the similarity) between X and P ,

σ the standard deviation, =1.

Fig. 3 shows the variation of the S_2 response $R(X, P)$ with variation in the Euclidian distance $\|X - P\|$. The experimental results show that the majority of distance values $\|X - P\|$ falls between 0-0.5. The lesser distance values are due to the lesser interclass shape difference of faces (compared to that of general objects). In order to extract discriminative facial features in spite of the smaller interclass shape differences, the learning function is modified (3).

$$R(X, P) = \frac{1}{k \left(\frac{\|X - P\|^2}{2\sigma^2} \right) + 1},$$

where,

k a tunable parameter, equal to 80/7 in our case (obtained empirically).

Fig. 3 also shows a comparison of the learning functions (2) and (3). Clearly the utilization of the new learning function (3) extracts better discriminative features $R(X, P)$.

4) Component based approach: Utilizing the location information

The proposed algorithm utilizes a component based approach for face recognition. The information on the locations of the prototype patches (which represents the facial components) is utilized during the pooling in layer 4. The pooling space is limited to $S_p\%$ of the space around the original patch location (instead of global maximization, $S_p=100\%$, in the original C_2 feature extraction system). Fig. 4 shows the variation in accuracy with S_p for MIT-CBCL dataset. The algorithm provided best accuracy when S_p is 23%. Lesser value of S_p (say 10%) led to lesser accuracy as it does not provide much invariance. At the same time, when the value of S_p is higher (say 80%) the accuracy again declined as the component information is lost.

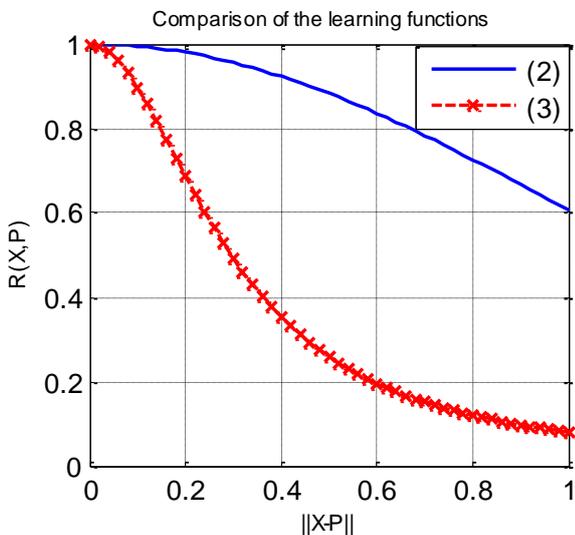


Fig. 3. Comparison of the learning functions represented by (2) and (3). The utilization of (3) led to better discriminative features.

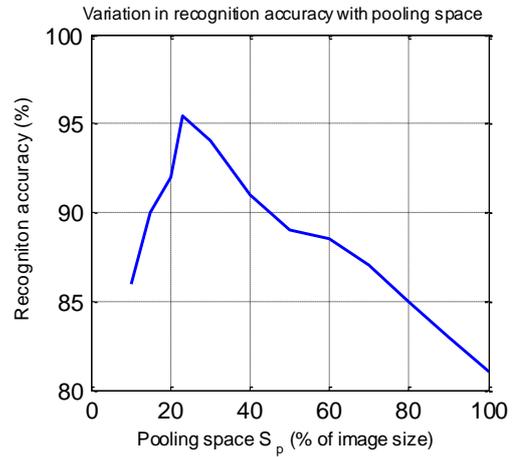


Fig. 4. The variation in recognition accuracy with pooling space S_p . The algorithm provided the best performance when S_p is 23%.

III. IMPLEMENTATION OF THE FACE RECOGNITION SYSTEM

Fig. 5 shows the training phase of the face recognition algorithm. The prototype patches are selected using the feature selection algorithm based on SVM normals [12]. The C_2 features are extracted using the proposed modified algorithm, and are normalized between -1 and +1. An SVM classifier with linear kernel is trained using these features and the class labels.

Fig. 6 shows the testing phase. The C_2 features are extracted and normalized using the data from the training phase. The classification is done using the trained SVM classifier model.

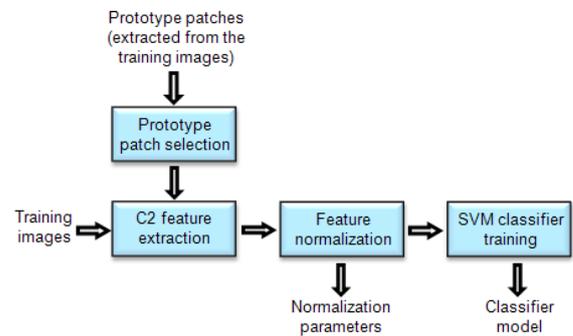


Fig. 5. The training phase.

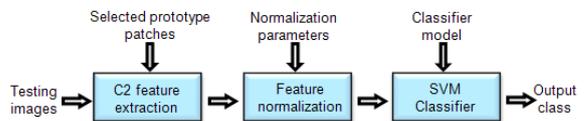


Fig. 6. The face recognition system (the testing phase).

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed face recognition algorithm is tested using two different face datasets, the MIT-CBCL face recognition database [10] and FEI face database [11]. Ten class subsets of these datasets with wide variations in pose are utilized for the evaluation (Fig. 7(a) and (b)). The details of the subsets are provided in Table 1. The performance of the proposed system, which is based on modified C_2 features, is compared with that of original C_2 features based system and MPCALDA (Multilinear Principle Component Analysis plus Linear Discriminant Analysis) [13].



Fig. 7. Sample images from the datasets utilized, (a) MITCBCL face dataset [10] and (b) FEI face dataset [11].

A. Results

The MIT-CBCL subset consists of 2000 face images. The algorithm is trained using 150 images (15 per class with left, right and frontal views) and tested using the rest 1850 images. The reported results are the average accuracies over 10 different runs of the algorithm (with different train and test sets). The proposed algorithm provided an accuracy of 95.40% and it outperformed the conventional C_2 features and MPCALDA systems (Table 1).

The FEI face database is a Brazilian face database taken in

an upright frontal position with profile rotation of up to about 180 degrees. The subset contains 10 classes of face images (5 male and 5 female), with variations in pose. The training set consists 60 images and testing set consists 80 images. The reported results are the average accuracies over 3 different runs of the algorithm (the training-testing sets are partitioned in such a way that all the images are included in the training and testing sets at least once). The proposed algorithm provided a recognition accuracy of 98.25%, outperforming the other two compared methods (Table 1).

TABLE I: SUMMARY OF EXPERIMENTAL EVALUATION: DATASETS UTILIZED AND COMPARISON OF RECOGNITION ACCURACIES

Dataset	# training images	# testing images	Recognition accuracy (%)		
			Proposed algorithm	Original C_2 features [4]	MPCALDA [13]
MIT-CBCL dataset [10]	150	1850	95.40	87.05	88.53
FEI dataset [11]	60	80	98.25	91.75	89.84

B. Analysis of the Performance Improvement by Proposed Modifications

We conducted experiments to estimate the contribution of every modification. All the four modifications are found to have notable effect on improving the recognition accuracy (see Table 2 reporting the results of individual modifications activated, for MIT-CBCL database). With all of the modifications activated we achieved 8.35% accuracy improvement. Furthermore we tested all possible combinations of modification activations to study the interaction effects. The interaction effects can be explained by analyzing misclassified images as follows: 1) The new learning function enhanced the recognition of both frontal and profile faces. 2) The limited position invariance and component based approach are more relevant for the recognition of frontal faces (than for a combination of frontal and profile faces). 3) The modification for profile invariance is more relevant for the recognition of a combination of frontal and profile faces (than for only frontal faces).

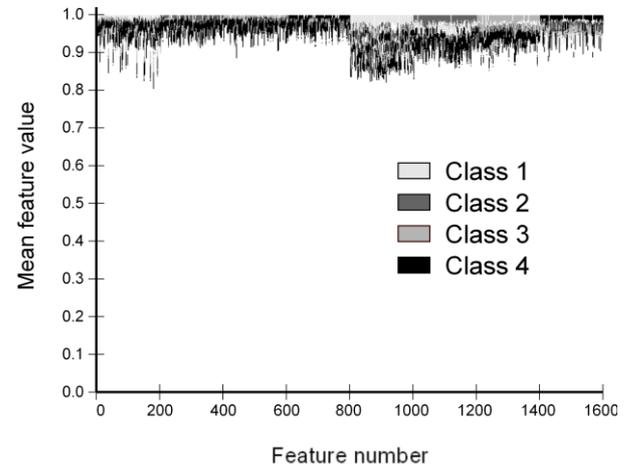
TABLE II: PERFORMANCE IMPROVEMENT ANALYSIS

Modification	Accuracy	Improvement
None (original C_2 features)	87.05	-
Limited position invariance	89.23	2.18
Profile invariance	93.45	6.40
New learning function	92.38	5.33
Component based approach	92.81	5.76
All modifications together	95.40	8.35

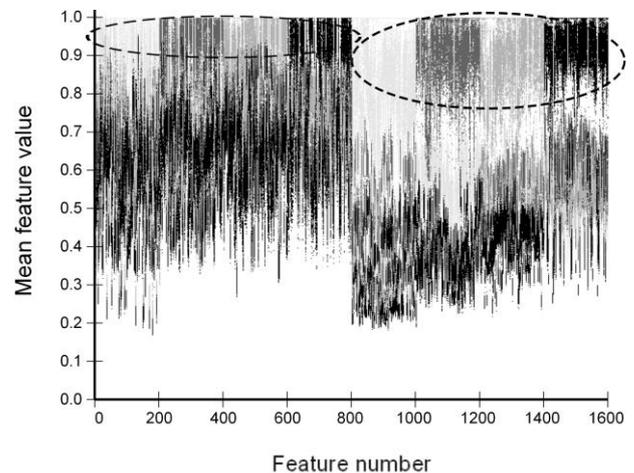
C. Discussion

Fig. 8 shows a comparison of the interclass separation of features extracted by three systems, *a*) System with learning function (2) and class specific prototype patches, *b*) System with learning function (3) and class specific prototype patches (proposed system), and *c*) System with learning function (3) and universal dictionary of features [4]

(prototype patches extracted from natural images). On comparison the proposed algorithm provided the best interclass separation (*b*).



(a)



(b)

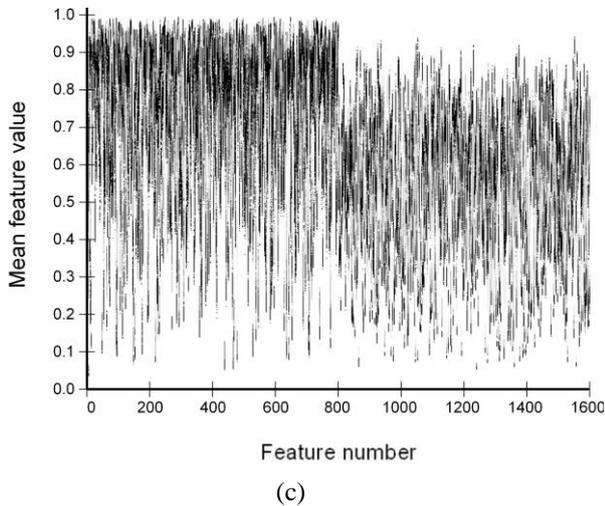


Fig. 8. Comparison of interclass separation of features. Plotted are the mean feature values for all 1600 features (mean of class1 feature1, class2 feature1, ..., class3 feature1600, class4 feature1600). The x-axis represents the feature number and y-axis represents the mean feature value (for clarity only 4 classes, from MIT-CBCL dataset, are shown). (a) shows the distribution when the features are extracted using equation (2) and (b) shows that when features are extracted using (3). The marked regions in (b) show the better interclass separation capability of the proposed algorithm. (c) shows the distribution when the features are extracted using a universal dictionary of features [4]. The utilization of universal dictionary of features is not effective in face recognition as the resultant class features are more overlapped.

The number of prototype patches (Section 2.2.2), the constant k in (3) (Section 2.2.3), and the pooling space S_p in layer 4 (Section 2.2.4) are tunable parameters in the proposed algorithm. The reported values of these parameters are found empirically. The increase in the number of prototype patches improves the recognition accuracy at the expense of reduced computational speed. The constant k decides the steepness of the learning function. The value of S_p is selected to get an optimum performance between invariance and selectivity (Fig. 4).

Manual annotation of training images is done in the proposed algorithm to identify the geometrically significant and textured positions. This can be automated using an algorithm to identify the facial parts. If a facial part is not visible in a profile view, the corresponding part in the symmetrically opposite side is considered (exploiting the bilateral symmetry property of human faces [14]). Experiments are also run by neglecting the facial component which is not visible, and the algorithm still provided good performance. The proposed algorithm mainly considered the horizontal (left-right) profile variations and didn't consider the vertical (up-down) profile variations. The algorithm can be extended to provide up-down profile invariance.

V. CONCLUSION

A pose invariant face recognition algorithm is proposed

using modified C_2 standard model features. The C_2 feature extraction system is modified to extract discriminative shape and texture features of facial parts. The prototype patches are extracted from the geometrically significant and textured parts of different profile views of faces. The algorithm is tested using two profile face datasets and its performance is compared with that of the original C_2 features and MPCALDA based features. On comparison the proposed algorithm provided more than 6.5% improvement in the recognition accuracy.

As a future work, we plan to apply the proposed algorithm for facial expression recognition. The expressions are to be recognized irrespective of the face identity.

REFERENCES

- [1] M. H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.
- [2] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces - a survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705–740, 1995.
- [3] M. Poggio T. Riesenhuber, "Hierarchical models of object recognition in cortex," *Nature Neuroscience*, vol. 2, no. 11, pp. 1019–1025, 1999.
- [4] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 411–426, 2007.
- [5] T. Serre, L. Wolf, and T. Poggio, "Object recognition with features inspired by visual cortex," *International Conference on Computer Vision and Pattern Recognition*, San Diego, CA, 2005, pp. 994–1000.
- [6] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 300–312, February, 2007.
- [7] P. Nicolas, S. Zak, Z. Todd, and C. David, "Scaling up biologically inspired computer vision: A case study in unconstrained face recognition on facebook," in *IEEE Computer Vision and Pattern Recognition, Workshop on Biologically Consistent Vision*, Colorado Springs, CO, June, 2011.
- [8] R. Singh, M. Vatsa, A. Ross, and A. Noore, "A mosaicing scheme for pose-invariant face recognition," *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol.37, no.5, pp. 1212–1225, October, 2007.
- [9] P. Pramod Kumar, P. Vadakkepat, and A. P. Loh, "Hand posture and face recognition using a fuzzy-rough approach," *International Journal of Humanoid Robotics*, vol. 07, no. 03, pp. 331–356, September, 2010.
- [10] B. Weyrauch, J. Huang, B. Heisele, and V. Blanz, "Component-based face recognition with 3D morphable models," in *Proc. of CVPR Workshop on Face Processing in Video (FPIV'04)*, Washington DC.
- [11] C. E. Thomaz and G. A. Galdi, "A new ranking method for principal components analysis and its application to face image analysis," *Image and Vision Computing*, vol. 28, no. 06, pp. 902–913, June, 2010.
- [12] J. Mutch and D. G. Lowe, "Object class recognition and localization using sparse features with limited receptive fields," *International Journal of Computer Vision*, vol. 80, no. 1, pp. 45–57, 2008.
- [13] L. Haiping, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 18–39, January, 2008.
- [14] F. T. Nikolaus and H. B. Heinrich, "How is bilateral symmetry of human faces used for recognition of novel views," *Vision Research*, vol. 38, no. 01, pp. 79–89, January, 1998.