

Illumination Invariant Facial Pose Classification

Ajay Jaiswal
School of Computer &
Systems Sciences,
JNU, New Delhi – India

Nitin Kumar
School of Computer &
Systems Sciences,
JNU, New Delhi – India

R. K. Agrawal
School of Computer &
Systems Sciences,
JNU, New Delhi – India

ABSTRACT

In this paper, we compared the performance of various combinations of edge operators and linear subspace methods to determine the best combination for pose classification. To evaluate the performance, we have carried out experiments on CMU-PIE database which contains images with wide variation in illumination and pose. We found that the performance of pose classification depends on the choice of edge operator and linear subspace method. The best classification accuracy is obtained with Prewitt edge operator and Eigenfeature regularization method. In order to handle illumination variation, we used adaptive histogram equalization as a preprocessing step resulting into significant improvement in performance except for Roberts operator.

General Terms

Pattern Recognition, Classification.

Keywords

Pose Classification, Edge detection, Linear Subspace Methods.

1. INTRODUCTION

Human head orientation is being used to express important inter-personal information. For example, a person usually points the direction of his head to indicate the intended target in conversation. Similarly in a dialogue, head movement indicates the listener to switch roles in conversation. There is also importance of the head movement as a form of gesture in a discussion. People may nod to indicate that they agree to what is being said, and use other gestures to indicate disagreement.

Humans possess inherent ability to understand such movements of head or face of a person conveniently which allows interpreting the intention or movement of other persons without any verbal form of communication. But to perform such pose estimation automatically is a challenging task. Hence, it has drawn the attention of the research community in last few years [1]. In methods for automatic head/face pose estimation, we need a process that can infer pose information from digital images. The method used for head pose estimation should be robust to changes in factors affecting images such as illumination, facial expression, environmental conditions, foreign objects *etc.* Human head orientation can be performed along three axes shown in Fig. 1 [1]. These orientations are defined as pitch, yaw and roll [1]. Although head pose can be estimated in any dimension, but most of the research work of pose estimation has been done when the movement of head is towards left or right which is known as yaw.

Pose estimation is also one of the preprocessing steps in face recognition across pose. In recent methods for face recognition involving pose variation, first the pose of the test input face is

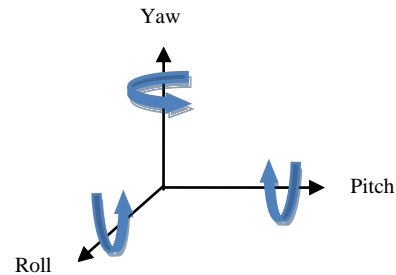


Figure 1: Directions of human head orientation

estimated and then it is transformed in to an already known reference pose [2,3]. When multiple view of a facial image are available, pose estimation can be used to reduce the search space, as first pose is estimated then probe image is compared with gallery images in the similar pose [4]. Pose estimation can be carried out in 3D image space or 2D image space. In 3D context, the human face is represented along three dimensions which require more storage and greater computational complexity. These limitations are overcome by the 2D models. In 2D context, various methods for face pose estimation are classified into two categories [5]: Landmark feature detection based and Appearance based subspace methods. Landmark feature detection methods [6] use certain localized landmarks points on the image and tries to estimate the pose information by modeling the displacement of these points across different poses. However, it is very sensitive to accurate localization of landmarks and also assumes that the ratios of these points do not change significantly in multiple poses. Appearance based subspace methods treat the whole face as one feature vector in some feature subspace [7] and pose estimation is carried out in the transformed feature subspace. Subspace methods circumvent these problems of landmarks localization and modeling. However, these methods are based on assumption that inter-pose variations are always larger than intra-pose variations. This condition is not satisfied always since different individuals in same pose may have large variations in appearance due to e.g. glasses, expressions, illumination and skin color. A hybrid model is also proposed for pose estimation [8] which exploits the advantages of the above mentioned two approaches. To overcome above mentioned limitations, a novel discriminative feature descriptor, Local Energy based Shape Histogram (LESH), is proposed in [5]. In this approach, a pose similarity feature space (PSFS) is generated with the help of feature descriptor that transforms the multi-class problem into two-class by using inter-pose and intra-pose similarities.

Most of 2D image-based methods handle either illumination or pose variation. Hence, these methods do not perform well when face images incorporate both illumination and pose variations.

Sang *et al.* [4] have proposed a method which involves three phases for pose classification with illumination variation. In phase one, well known Sobel operator is used for finding edges

from images having light and pose variations. PCA is applied to reduce the dimensions in phase two and linear discriminant analysis (LDA)/ LDA for regression (LDA-r) [9] is used to extract features to estimate the pose in phase three. The research work [4] have also employed commonly used canny operator in phase one but the performance with sobel operator is found superior in comparison to canny. In literature, other edge detection operators e.g. Roberts, LoG (Laplacian of Gaussian), Prewitt *etc.*[10]are also suggested. It is also known that the choice of edge detection operator depends on underlying application. Hence, there is a need to investigate the better choice of edge operator. Similarly various subspace methods such as LDA /QR [11], Null-space LDA [12], Dual-space LDA [13], Eigenfeature regularization [14], SVM Based Feature extraction[15] *etc.* have been suggested for feature extraction which perform better in comparison to LDA/ LDA-r. In this paper, we investigate the performance of various combinations of edge operators and subspace methods to determine the best combination for pose classification. The performance is evaluated in terms of classification accuracy.

The rest of the paper is organized as follows: Section 2 provides an overview of various edge operators. Section 3 briefly describes various linear subspace methods. Section 4 describes the experimental set up and results. Conclusions and future work are included in section 5.

2. EDGE DETECTION

Edge detection is an important step in image processing as edges contain a major part of information in images. The objective of edge detection is to locate the homogeneous region boundaries in an image based on characteristics such as intensity and texture. Edges in an image can be step, line and junctions edges [16]. In the last few decades various methods for edge detection have been proposed. The performance of a particular edge detection method depends on underlying application. These methods can be differently classified on the basis of different criteria [16]. From the technical point of view, edge detection methods are divided into two classes viz. search-based and zero-crossing based. Search-based methods work on the principle that local edge strength represents local maxima of the absolute value of the first derivative while zero-crossing methods find the location of zero-crossing in the second derivative of an image. From the conceptual point of view, these methods can be divided into contextual and non-contextual approaches. The contextual methods exploit a priori scene information. These methods perform better only in specific context. On the other hand, non-contextual methods are independent of the context. Therefore these methods can be used in large number of applications. Another classification of edge detection methods divides these methods in to nine categories: classical (e.g. Sobel), Gaussian based (e.g. LoG), multi-resolution [17], non-linear [18], wavelet based [19], statistical [20], machine learning based [21], contextual [22] and line edge [23] methods.

Typically edge detection methods involve three key steps: differentiation, smoothing and labeling. Differentiation involves computation of the derivative of the image. Smoothing is used to reduce the effect of noise, while labeling incorporates edge localization and suppressing of false edges. Different edge detection methods may or may not involve all the aforementioned steps.

Several first order derivative approximations along two perpendicular directions have been proposed, and some of the most commonly used are follows:

The **Roberts** [10] operator is calculated using a set of axes rotated 45 degrees to the usual orientation of the column and row, and considers only four neighboring pixels to determine an

edge pixel. The two kernels used by Roberts operator are shown below:

+1	0
0	-1

0	+1
-1	0

It is sensitive to noise because only four pixels are used to evaluate gradients. Most real-world images contain noise in them and the Roberts algorithm will highlight this noise as edges.

The **Prewitt** method [10] models the edges in an image as ramp function. This can be implemented with the help of convolution kernels whose response is maximal to edge pixels by giving equal weights to all the neighboring pixels in horizontal and vertical directions. The major advantage of Prewitt operator is its simplicity but it is also sensitive to noise and produces thick edges in the images.

-1	0	+1
-1	0	+1
-1	0	+1

+1	+1	+1
0	0	0
-1	-1	-1

Fig 2. Convolution kernels for Prewitt operator

The **Sobel** method [10] slightly changes the convolution kernel by giving more weight to central pixel in comparison to Prewitt operator. This method also suffers from thick edges and is sensitive to noisy images.

-1	0	+1
-2	0	+2
-1	0	+1

+1	+2	+1
0	0	0
-1	-2	-1

Fig 3. Convolution kernels for Sobel operator

Laplacian of Gaussian (LoG) [10] method finds edges by finding zero crossings after filtering image with a Laplacian of Gaussian filter. The Laplacian $L(x,y)$ of an image with pixel intensity values $I(x,y)$ is given by:

$$L(x, y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad (1)$$

Image is smoothed using Gaussian filter before applying Laplacian filter to reduce the noise. LoG is a linear and non-directional operator. The advantage of LoG lies in its capability to determine localized edges. Disadvantage of LoG operator is that it does not find the orientation of edges. Also it does not perform better at the corners and curves.

Canny edge detector [24] is proposed to achieve the following performance criteria: low error rate, localized edge points and unique response to an edge. Low error rate is important as edge detector should not miss any edge point and do not respond to

non-edges. Localized edge points minimize the distance between the edge pixels as obtained by the detector and the actual edge. A third criterion is to have unique response to a single edge.

The canny method finds edges by looking for local maxima of the gradient of smoothed image. The gradient is calculated using the derivative of a Gaussian filter in two orthogonal directions and then finds the edge strength and direction of edges in terms of two components of gradient. Candidate edge pixels are recognized as the pixels that survive a thinning process called non-maximal suppression. Finally, hysteresis is used to eliminate streaking. It uses two different thresholds, a high and a low. All candidate edge pixels above the high

threshold are marked as edge pixels and all pixels above the lower threshold that can be connected to any pixel above the high threshold through a chain of edge pixels are also marked as edge pixels. All pixels below lower threshold are non-edge pixels. The advantages of Canny operator is its ability to detect edges in images with noise and true weak edges. Its disadvantage includes complexity of implementation and high computation requirement in comparison to other operators.

3. LINEAR SUBSPACE METHODS

Here we briefly describe the linear subspace methods used for evaluating the performance in conjunction with various edge operators for feature extraction. In Linear discriminant analysis (LDA) [25], the objective is to seek the transformation matrix W such that the following criterion function

$$J(W) = \frac{W^T S_b W}{W^T S_w W} \quad (2)$$

is maximized, thereby maximizing the between-class scatter and minimizing the within-class scatter of the projected images simultaneously.

Here S_b and S_w are between-class scatter and within-class scatter matrix respectively. If S_w is non-singular then we solve

$$S_w^{-1} S_b W = \lambda W \quad (3)$$

Hence in case of LDA, the transformation matrix $W = \{w_i | i=1, 2, \dots, m\}$ is represented in terms of eigenvectors of matrix $U = S_w^{-1} S_b$ corresponding to m largest eigenvalues. Note that $c-1$ is an upper bound on m because the rank of S_b can be at most $c-1$ or less, where c is the number of classes [25].

LDA for regression (LDA-r) [9] is a variant of LDA, which solves regression problem. If we consider a set of predictor/response pairs $\{(x_i, y_i)\}_{i=1}^n$ where $x_i \in R^{d \times 1}$, $y_i \in R$ and n represents the total number of predictor/response pairs. Here, d corresponds to the dimensionality of the input and y_i is the class label. The computation of between-class scatter matrix S_b and within-class scatter matrix S_w is modified as follows [9]:

$$S_w = \frac{1}{n_w} \sum_{(i,j) \in A_w} f(y_i - y_j)(x_i - x_j)(x_i - x_j)^T \quad (4)$$

$$S_b = \frac{1}{n_b} \sum_{(i,j) \in A_b} f(y_i - y_j)(x_i - x_j)(x_i - x_j)^T \quad (5)$$

where n_w and n_b are cardinality of sets A_w and A_b respectively. A_w and A_b are given by

$$A_w = \{(i, j) : |y_i - y_j| < \tau, i, j \in \{1, \dots, n\}, i \neq j\}$$

$$A_b = \{(i, j) : |y_i - y_j| \geq \tau, i, j \in \{1, \dots, n\}, i \neq j\}$$

The function $f(\cdot)$ is a weight function with positive values and τ determines a boundary to separate the points of two different sets A_w and A_b . The weight function used in the research work [4] is $f(x) = \sqrt{|x| - \tau|}$ and $\tau = \alpha * \text{std}(y)$ where $\alpha = 0.1$ as in [9].

LDA/QR proposed by Ye *et al.* [11] is another extension of the classical LDA which involves two stages. The first stage maximizes the separation between various classes via QR decomposition. The QR decomposition method decompose a matrix A into two matrix Q and R such that $A = QR$ where Q is an orthogonal matrix and R is an upper triangular matrix. The

transformation matrix so obtained can be used independently for dimensionality reduction. The distinguishing property of first stage of LDA/QR is its low time/space complexity. The second stage takes into consideration the issue of within-class distance and also maintains low time/space complexity.

The LDA is performed in the principal subspace of S_w , in which $W^T S_w W \neq 0$. However, the null space of S_w , in which $W^T S_w W = 0$, also contains much discriminative information. It is possible to find some projection matrix W satisfying $W^T S_w W = 0$ and $W^T S_b W \neq 0$ which also maximizes Fisher criterion function in (2). This modification of LDA is termed as Null space LDA [12].

In the null-space LDA, the discriminative information present outside the null space is discarded. However, the discriminative information to distinguish any two faces is available in both subspaces (null space and principal space). Hence, the feature vector obtained using only the null-space LDA may not be suitable for better classification. Wang and Tang [13] proposed a Dual-space LDA approach to take full advantage of the discriminative information present in both subspaces. The eigenvalue spectrum in the null space of within-class scatter matrix is estimated based on a probabilistic visual model. The discriminant analysis is simultaneously applied in both null and the principal spaces of the within-class scatter matrix. The two sets of discriminative features so obtained are merged together for feature extraction.

Jiang *et al.* [14] proposed Eigenfeature regularization and extraction method. In this approach, image space spanned by eigenvectors of within-class scatter matrix is decomposed into three subspaces: i) a reliable subspace which is spanned mainly by the variation in face, ii) an unstable subspace due to noise and finite number of training samples and iii) null space. Eigenfeatures are then regularized differently in these three subspaces based on an eigenspectrum model to alleviate the problems of instability, overfitting, or poor generalization. This also allows the discriminant evaluation to be performed in the whole space. Feature extraction or dimensionality reduction occurs only at the final stage after the discriminant assessment. This facilitates a discriminative and a stable low-dimensional feature representation of the face image.

There are some inherent limitations in LDA which includes the small sample size (SSS) problem and a limited dimensionality because the rank of S_b is bounded by $c-1$ where c is number of classes. To overcome these limitations of LDA, Sang *et al.* [15] proposed a new feature extraction algorithm, called SVM-based discriminant analysis. They proposed computation of between-class scatter matrix based on the principle of SVM margins in conjunction with regularization process [14] to resolve the above mentioned limitations to determine distinctive features. In this method, SVM based modification of Fisher's criterion is obtained [15]. In first step, eigenfeature regularization is carried out as suggested by Jiang *et al.* [14]. Then, the regularized within class scatter matrix is obtained as:

$$\hat{S}_w = V_w \hat{\Lambda} V_w^T \quad (6)$$

where $\hat{\Lambda} = \text{diag}[\hat{\lambda}_1^w, \hat{\lambda}_2^w, \dots, \hat{\lambda}_{N_w}^w]$ is a matrix with regularized eigenvalues in its diagonal, and $V_w = [v_1^w, \dots, v_{N_w}^w]$

is the matrix of corresponding eigenvectors. Since SVM is originally proposed for a two-class problem, it is required to modify the scatter matrix for multi-class problem. The modified S_b is given by:

$$S_b = \sum_{i=1}^{c-1} \sum_{j=i+1}^c 2p_i p_j \frac{1}{2\|\omega\|} \left(\frac{\omega}{\|\omega\|} \right) \left(\frac{\omega}{\|\omega\|} \right)^T = \sum_{i=1}^c \sum_{j=1}^c p_i p_j \frac{\omega \omega^T}{\|\omega\|} \quad (7)$$

where ω is the hyperplane parameter of the SVM classifier, which separates classes. The rank of this between-class scatter matrix is bounded by $\min(c(c-1)/2, N_{SV})$ instead of $c-1$, where N_{SV} is the number of support vectors.

Now, we need to maximize the modified Fisher's criterion given by:

$$J(v) = \arg \max \frac{v^T S_b v}{v^T S_w v} \quad (8)$$

Generally in the case of lower number of samples, the matrix S_w is not a full rank matrix. However, the discriminant information corresponding to the zero eigenvalues of S_w have the most discriminative power. But, LDA cannot extract discriminant information only from the principal subspace of S_w . Zhang *et al.* [26] have proposed another approach based on the matrix exponential and is known as Exponential Discriminant Analysis (EDA). To extract discriminative information, they had replaced the eigenvalues of S_w i.e. λ_{wi} , by $\exp(\lambda_{wi})$ and λ_{bi} i.e. the eigenvalue of S_b , by $\exp(\lambda_{bi})$ and represent

$$\exp(\Lambda_b) = \text{diag}(e^{\lambda_{b1}}, e^{\lambda_{b2}}, \dots, e^{\lambda_{bn}}) \quad (9)$$

$$\exp(\Lambda_w) = \text{diag}(e^{\lambda_{w1}}, e^{\lambda_{w2}}, \dots, e^{\lambda_{wn}}) \quad (10)$$

Then the LDA criterion is modified as below [26]:

$$J(W) = \arg \max_W \left| \frac{W^T (\Phi_b \exp(\Lambda_b) \Phi) W}{W^T (\Phi_w \exp(\Lambda_w) \Phi) W} \right| \\ = \arg \max_W \left| \frac{W^T \exp(S_b) W}{W^T \exp(S_w) W} \right| \quad (11)$$

However, it should be noted that we must normalize scatter matrices S_b and S_w before carrying out further transformation because $\exp(S_b)$ and $\exp(S_w)$ may involve large numbers. This method captures high-order mixed central moments which can reveal the intereffect between many factors.

4. EXPERIMENTAL SETUP AND RESULTS

To determine best combination of edge operator and linear subspace method for pose classification, we have used five edge detection operators (Sobel, Canny, Prewitt, LoG and Roberts) in phase one and, eight linear subspace methods (LDA, LDA-r, LDA-QR, Null-space LDA, Dual-space LDA, Eigenfeature regularization, SVM based feature extraction and Exponential LDA) in phase three. To investigate the performance of various combinations, we have performed experiments on a widely used CMU-PIE face database [27]. In preprocessing, we have cropped face images containing central portion of face and resized each image to 60x50. We have selected images of 65 persons and 7 pose classes (c22, c02, c05, c27, c29, c14, c34) with 21 different lighting conditions for each pose. The pose classes are arranged sequentially from left profile to right profile. The training set was constructed by randomly choosing three images from each pose for each individual.

This results into a total of $3*7*65=1365$ images for training and $18*7*65=8190$ for testing. The experiment was repeated 20 times with different training samples to determine average classification accuracy. We applied edge operator followed by PCA for dimensionality reduction. Finally, linear subspace methods are used for feature extraction. K-nearest neighbor classification (K=1) method with Euclidean distance is used as a classifier. The results for average classification accuracy for different linear subspace methods over various edge detectors are shown in Table 1. For a given edge detector, the maximum classification accuracy achieved by a linear subspace method is shown in bold.

We observe the following from Table1:

1. The performance in terms of classification accuracy with Sobel and Prewitt edge operator is similar using all linear subspace methods.
2. The performance in terms of classification accuracy with Sobel/Prewitt edge operator is significantly better in comparison to other edge operators using all linear subspace methods.
3. Eigenfeature regularization method significantly outperforms other linear subspace methods except for Canny edge detector.
4. The standard deviation of the classification accuracies obtained using 20 different training and testing datasets shows that the classification accuracy is highly dependent on the choice of training samples. This is due to the fact that the face database contains wide illumination variations.
5. Although Canny edge detector is considered to be a better edge operator in literature but experimental results show that it does not perform better for pose classification problem.

Prewitt edge operator in combination with eigenfeature regularization method outperforms all other combinations.

Table 1: Comparison of average pose classification accuracy and standard deviation

	LDA	LDA-r	LDA/QR	NLDA	Dual_LDA	Eig_reg	Eig_SVM	Exp_LDA
Roberts	82.44 (10.48)	76.75 (9.86)	81.09 (14.37)	81.87 (12.19)	83.08(10.57)	87.42 (11.16)	83.62 (10.67)	81.74 (11.13)
Prewitt	84.14 (10.6)	79.66 (9.92)	84.24 (13.73)	82.86 (12.12)	84.41(10.76)	89.12 (10.73)	84.39 (10.79)	82.79 (11.98)
Sobel	84.33 (10.81)	79.46 (10.02)	83.22 (14.06)	82.56 (11.93)	84.75(10.84)	88.46 (10.43)	84.76 (10.66)	83.09 (11.95)
LoG	79.49 (10.41)	73.06 (9.00)	78.77 (14.94)	77.15 (12.12)	80.51(10.47)	81.16 (10.94)	80.1(10.24)	76.68 (9.97)
Canny	78.32 (10.24)	67.40 (10.10)	70.32 (15.56)	71.1 (10.3)	79.26(10.77)	75.51 (12.45)	78.91(10.67)	77.28 (12.42)

In order to overcome the effect of illumination variation present in face database, we applied adaptive histogram equalization prior to edge operator and used eigenfeature regularization in phase three of pose classification problem. Table 2 contains the comparison of classification accuracy with and without adaptive histogram equalization.

Table 2: Comparison of classification accuracy with and without adaptive histogram equalization

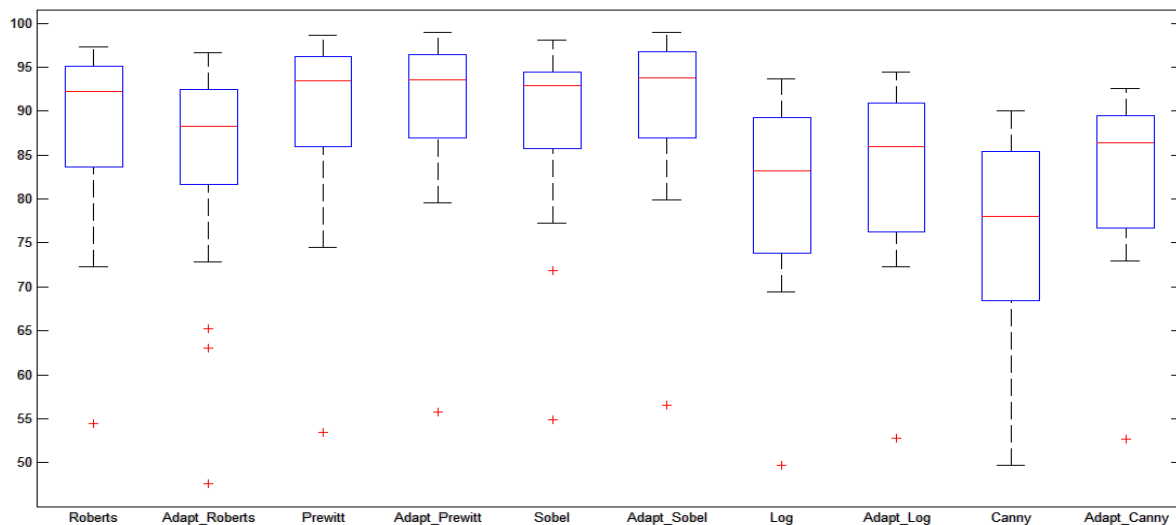
	Eig_reg (without adaptive histogram)	Eig_reg (with adaptive histogram)
Roberts	87.42 (11.16)	84.31 (12.70)
Prewitt	89.12 (10.73)	89.99 (10.03)
Sobel	88.46 (10.43)	90.24 (9.87)
LoG	81.16 (10.94)	83.17 (10.48)
Canny	75.51 (12.45)	82.75 (9.54)

From Table 2 it can be observed that classification accuracy increases with the use of adaptive histogram equalization except for Robert edge operator. The classification accuracy has improved significantly in case of Canny operator with the use of adaptive histogram equalization. It can also be noted that standard deviation of classification accuracies has decreased with all operators except Robert operator. It signifies that adaptive histogram equalization reduces the effect of illumination variation to estimate pose. Similar observations can be made from box plot

shown in Figure 2.

5. CONCLUSION

Pose classification is a challenging problem in computer vision. Various methods have been proposed for identity independent pose classification. One of appearance based subspace methods for pose estimation involves three phases. An edge operator is used for finding edges from images, PCA is applied to reduce the dimensions, and finally a linear subspace feature extraction is used for pose classification. In literature, various edge operators and linear subspace method for feature extraction are suggested. It is well known that the choice of edge operator and linear subspace methods depends on underlying application. In this paper, we have investigated the performance of various combinations of edge operators and subspace methods to determine the best combination for pose classification. The performance is evaluated in terms of classification accuracy. For evaluating the performance, we have carried out experiments on CMU-PIE database which contains images with wide variation in illumination and pose. We found that the performance of pose classification depends on the choice of combination of edge operator and linear subspace method. The performance of Sobel and Prewitt edge operator is more suitable in comparison to other edge operators. We obtained best classification accuracy with Prewitt operator for edge detection with Eigenfeature regularization method. It is also found that the performance of method is sensitive to variation in illumination. In order to overcome this, we used adaptive histogram equalization as preprocessing step and noticed significant improvement in performance except for Roberts operator.



Figure

2: Box plot for classification accuracy in Table2

In the future work, we are planning to investigate robustness of best combination for pose classification on other publicly available face databases. We would also explore scale-space invariant approaches to tackle illumination variation in pose classification.

6. REFERENCES

- [1] E. M. Chutorian and M. M. Trivedi. Head Pose Estimation in Computer Vision: A Survey, *IEEE Trans. on PAMI*, 31(4): 607 – 626, 2009.
- [2] S.Z. Li and A.K. Jain. *Hand book of face recognition*. Springer-Verlag, 2005.
- [3] H. S. Lee and D. Kim. Generating frontal view face image for pose invariant face recognition, *PR letters*, 27(7):747-754, 2006.
- [4] S. Choi, C. Choi and N. Kwak. Face recognition based on 2D images under illumination and pose variations, *PR Letters*, 32(4):561–571, 2011.
- [5] M.S. Sarfraz and O. Hellwich. Head Pose Estimation in Face Recognition Across Pose Scenarios, in *Proc. VISAPP (1)*:235-242, 2008.
- [6] S. Zhao and Y. Gao. Automated Face Pose Estimation Using Elastic Energy Models, *The 18th ICPR*, 4:618-621, 2008.
- [7] S. Gong, S. McKenna, and J. J. Collins. An investigation into face pose distributions, In *FG.*, 265, 1996.
- [8] M. Gründig and O. Hellwich. 3D Head Pose Estimation with Symmetry Based Illumination Model in Low Resolution Video, *DAGM, LNCS*, 45-53, 2004.
- [9] S. Choi, C. Choi and N. Kwak. Feature extraction for regression problems and an example application for pose estimation of a face, In *Proc. fifth International Conference on Image Analysis and Recognition – ICIAR*, 435–444, 2008.
- [10] R. Gonzalez and R. Woods. *Digital Image Processing*, Prentice Hall, second edition, 1992.
- [11] J. Ye and Q. Li. A two-stage linear discriminant analysis via qr-decomposition, *IEEE Trans. on PAMI*, 27:929-941, 2005.
- [12] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu. A New LDA Based Face Recognition System Which can Solve the Small Sample Size Problem, *Journal of Pattern Recognition*, 33(10): 1713-1726, 2000.
- [13] X. Wang and X. Tang. Dual-Space Linear Discriminant Analysis for Face Recognition, *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, 2: 564-569, 2004.
- [14] X. Jiang, B. Mandal, and A. Kot. Eigenfeature Regularization and Extraction in Face Recognition, *IEEE Trans. on PAMI*, 30(3): 383-394, 2008.
- [15] S.K. Kim, Y. J. Park, K.A. Toh, and S. Lee. SVM-based feature extraction for face recognition. *Pattern Recognition*, 43(8): 2871-2881, 2010.
- [16] M. A. Oskoei and H. Hu. A Survey on Edge Detection Methods, Technical report, 2010.
- [17] B. G. Schunck. Edge detection with Gaussian filters at multiple scales, in *Proc. IEEE Comp. Soc. Work. Comp. Vis.*, 208–210, 1987.
- [18] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion, *IEEE Trans. on PAMI*, 12:629–639, 1990.
- [19] D. Heric and D. Zazula. Combined edge detection using wavelet transform and signal registration, *Elsevier Journal of Image and Vision Computing*, 25:652–662, 2007.
- [20] S. Konishi, A. L. Yuille, J. M. Coughlan, and S. C. Zhu. Statistical Edge Detection: Learning and Evaluating Edge Cues, *IEEE Transactions on PAMI*, 25(1):57-74, 2003.
- [21] J. Wu, Z. Yin and Y. Xiong. The Fast Multilevel Fuzzy Edge Detection of Blurry Images, *IEEE Signal Processing Letters*, 14(5):344–347, 2007.
- [22] Y. Yu and C. Chang. A new edge detection approach based on image context analysis, *Elsevier Journal of Image and Vision Computing* 24:1090–1102, 2006.
- [23] G. Giraudon. Edge Detection from Local Negative Maximum of Second Derivative. In *Proceedings of IEEE, International Conference on CVPR*, 643-645, 1985.
- [24] J. Canny. A computational approach to edge detection, *IEEE Trans. on PAMI*, 8 (6): 679–698, 1986.
- [25] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [26] T. Zhang, B. Fang, Y. Y. Tang, Z. Shang and B. Xu. Generalized Discriminant Analysis: A Matrix Exponential Approach, *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40(1):186-197, 2010.
- [27] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression (PIE) database of human faces. Technical Report CMU-RITR- 01-02, The Robotics Institute, Carnegie Mellon University, 2001.