Emotion Recognition using discrete cosine transform and discrimination power analysis

Shikha Surendran
Dept. of Electronics and Communication
MIT, Anjarakandy, Kannur Kerala
shikhasurendrann@gmail.com

ABSTRACT

The paper states the technique of recognizing emotion using facial expressions is a central element in human interactions. Discrete cosine transform (DCT) is a powerful transform to extract proper features for face recognition. The proposed technique uses video input as input image taken through the webcam in real time on which discrete cosine transform is performed. After applying DCT to the entire face images, some of the coefficients are selected to construct feature vectors. Most of the conventional approaches select coefficients in a zigzag manner or by zonal masking. In some cases, the low-frequency coefficients are discarded in order to compensate illumination variations. Since the discrimination power of all the coefficients is not the same and some of them are discriminant than others, so we can achieve a higher true recognition rate by using discriminant coefficients (DCs) as feature vectors. Discrimination power analysis (DPA) is a statistical analysis based on the DCT coefficients properties and discrimination concept. It searches for the coefficients which have more power to discriminate different classes better than others.. The proposed method can be implemented for any feature selection problem as well as DCT coefficients. One-against-one classification method is adopted using these features. Evaluation of the procedure is performed in MATLAB using an image database of 20 people and each subject have 6 different facial expressions. After training about many epochs system achieved approximately 94% accuracy.

Key Words:

Discrimination power analysis (DPA), Discrete cosine transform (DCT), face recognition, Coefficient selection (CS).

1. INTRODUCTION

A persons face changes according to emotions or internal states of the person. Face is a natural and powerful communication tool. Analysis of facial expressions through moves of face muscles leads to various applications. Facial expression recognition plays a significant role in Human Computer Interaction systems. Humans can understand and interpret each other's facial changes and use this understanding to response and communicate. A machine capable of meaningful and responsive communication is one of the main focuses in robotics. There are many other areas which benefit from the advances in facial expression analysis such as psychiatry, psychology, educational software, video games, animations, lie detection or other practical real-time applications. A wide variety of approaches have been developed by researchers for face recognition problem. From one point of view, these various approaches are categorized into two general groups, namely feature-based and holistic approach. In feature-based approaches ,shape sand geometrical relationship of the individual facial features including eyes, mouth and nose are analyzed. But in holistic approaches, the face images are analyzed as two-dimensional holistic patterns. Feature-based approaches are more robust against rotation, scale and illumination variations, but their success depends on the facial feature detection .Due to difficulties in facial feature detection, holistic approaches are considered much more frequently than feature-based approaches . Extracting proper features is crucial for satisfactory design of any pattern classifier system. In this way, two types of discrete transforms, statistical and deterministic, have been widely used for feature extraction and data redundancy reduction. The basis vectors of the statistical transforms depend on the statistical specification of the database and different basis vectors are possible for different databases. Deterministic transforms have invariant basis vectors which are independent of the database. Although statistical transforms have a great ability to remove correlation between data, they have high computational complexity. Also computation of the basis vectors for each given database is needed. Among statistical approaches, principal component analysis (PCA) and linear discriminant analysis (LDA) are two powerful statistical tools for feature extraction and data representation. Because of some limitations of the PCA and LDA, a variety of modifications have been proposed by authors. Combination of statistical and deterministic transforms constructs a third type of feature extraction approaches with both advantages. In this type, DCT reduces the dimension of data to avoid singularity and decreases the computational cost of PCA and LDA. Studies showed that using the PCA or the LDA in the DCT domain yields the same results as the one obtained from the spatial domain. Various combinations of the DCT, PCA and LDA have been surveyed. Some used a combination of the DCT, PCA and the characteristics of the human visual system for encoding and recognition of faces. After applying the DCT to an image, some coefficients are selected and others are discarded in data dimension reduction. The selection of the DCT coefficients is an important part of the feature extraction process. In most of the approaches which utilize the DCT, not enough attention was given to coefficients selection (CS). The coefficients are usually selected with conventional methods such as zigzag or zonal masking. These conventional approaches are not necessarily efficient in all the applications and for all the databases. Discrimination power analysis (DPA) is a novel approach which selects features (DCT coefficients) with respect to their discrimination power. DPA utilizes statistical analysing of a database, associates each DCT coefficient discrimination to a number, and generates a CS mask. The proposed feature extraction approach is database dependent and is able to find the best discriminant coefficients for each database.

2. Feature extraction

In this section, various DCT feature extraction approaches are considered and a new efficient approach is proposed. DCT feature extraction consists of two stages. In the first stage, the DCT is applied to the entire image to obtain the DCT coefficients, and then some of the coefficients are selected to construct feature vectors in the second stage. Dimension of the DCT coefficient matrix is the same as the input image. In fact the DCT, by itself, does not decrease data dimension; so it compresses most signal information in a small percent of coefficients

.2.1. DCT and coefficients selection

For an M X N image, where each image corresponds to a 2D matrix, DCT coefficients are calculated as follows:

$$F(u,v)=\frac{1}{\sqrt{MN}}x(u)x(v) \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y) X Cos(\frac{(2x+1)u\pi}{2M}) X Cos(\frac{(2x+1)u\pi}{2M})$$

u = 0; 1; . . . ;M; v= 0; 1; . . . ;N

where $x(\omega)$ is defined by

$$x(\omega)=\begin{cases} \frac{1}{\sqrt{2}} & \omega = 0 \\ 1 & otherwise \end{cases}$$

F(x,y) is the image intensity function and f(u,v) is a 2d matrix of dct coefficients .block-based implementation and the entire image are the two implementations of the dct . Entire image dct has been used in this paper. The dct is applied to the entire image to obtain the frequency coefficient matrix of the same dimension. Fig. 1a and b shows a typical face image and its dct coefficients image. In general ,the dct coefficients are divided into three bands (sets), namely low frequencies, middle frequencies and high frequencies. Fig. 1c visualizes these bands. Low frequencies are correlated with the illumination conditions and high frequencies represent noise and small variations. Middle frequencies coefficients contain useful information and construct the basic structure of the image .from the above discussion, itseems that the middle frequencies coefficients are more suitable candidates in face recognition**.**

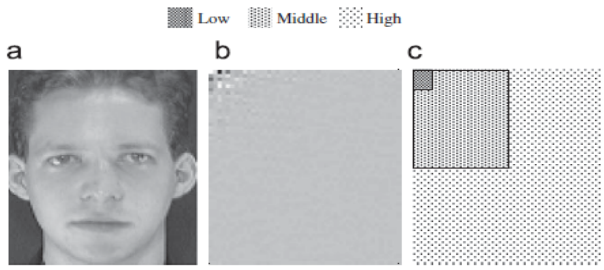

Fig. 1. (a) A typical face image; (b) its DCT transformed image and (c) typical division of the DCT coefficients into low, middle and high frequencies.

2.2. Basis of data-dependent approach

Conventional CS approaches select the

fixed elements of the DCT coefficients matrix. We call these conventional CS approaches deterministic. Zigzag, zonal masking and their modifications are the examples of the deterministic approaches. Two popular deterministic approaches are shown in Fig. 2. Although the deterministic approaches are simple, they are not necessarily efficient for all the databases. Using a DCT coefficient in a feature vector, which improves the true recognition rate in a database, may deteriorate results in another database. We claim that an adaptive CS method will improve the accurate recognition rate. In this way, we propose another group of CS approaches namely, data-dependent approaches. The main idea behind the data-dependent approaches is that all of the DCT coefficients do not have the same ability to discriminate various classes. In other words, some coefficients, namely discriminant coefficients, separate classes better than others. These discriminant coefficients are dependent on database specifications. Our aim is to find these discriminant coefficients of a database. We propose a statistical approach to solve this problem. Our approach statistically analyzes all images of a database to associate each DCT coefficient with a number related to its discrimination power (DP). We name this analysis discrimination power analysis (DPA). Experimental results show that DPA as a data-dependent approach improves the accurate recognition rate in exchange for a little increase in computations. A discriminant coefficient has small variation within a class, and also large variation between the classes. It is described completely in the following subsection. The idea of data-dependent approach and DPA can be implemented for any feature selection problem as well as DCT coefficients. DPA is different from other similar approaches such as PCA and LDA, etc., which utilize between- and within- class variances. These approaches are trying to obtain a transform which maximizes the discrimination of the features in the transformed domain, while DPA searches for the best discriminant features in the original domain. Also combination of DPA and premasking makes it robust against the limited number of the training samples. DPA has no singularity problem which is common among LDA-based methods. DPA can be used as a stand-alone feature reduction

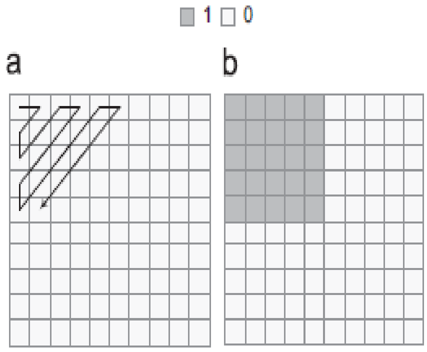

Fig. 2. Scheme of the DCT coefficients selection with deterministic approaches, (a) zigzag and (b) zonal masking.

2.3. Discrimination power analysis

DP of a coefficient depends on two attributes: large variation between the classes and small variation within the classes(and it means large discrimination power).Thus in one possible way, the DP can be estimated by the division of the between-class variance to the within-class variance. In this way, it is expected that one can estimate large values for discriminant coefficients. For an image size of M X N having C classes and S training images for each class, totally C X S training images are present. DP of each coefficient) can be estimated as follows

1. Construct the train set matrix, Aij, by choosing the DCT coefficients of the positions i and j for all classes and all training images:
2. Calculate the mean value of each class:
3. Calculate variance of each class:
4. Average the variance of all the classes
5. Calculate the mean of the all training samples
6. Calculate the variance of all the training samples
7. Estimate the DP for location(i,j):

The high value of DP means high discrimination ability of the corresponding coefficient. Their values are normalized between zero and one. As expected, high-frequency coefficients have small power of discrimination. The middle frequency coefficients have approximately large power of discrimination. Also discrimination power of low-frequency coefficients varies based on the database.

2.4. Premasking

Because DPA is a statistical analysis, the number of training samples influence its validation. Ideally, when enough number of samples for each class is available, DPA is an optimum CS approach. But when a limited number of the training samples is available, it degrades. Suppose that a training set of a database does not have any illumination variations. Therein, a large DP is calculated for the DC coefficient. So it is a discriminant coefficient. Any illumination variation in the test set may lead to incorrect recognition. A large number of training samples which contains various conditions of the face images increases the DPA validity . As considered in the previous section, the DCT domain is divided into three bands. Middle frequencies are suitable candidates for the correct recognition among these bands. Our experiments show that large DP may be calculated on high frequencies, especially with small size of training set, but it is not reasonable because of the noisy nature of these coefficients. The same phenomenon happens for the low frequencies, except that large DP will be reasonable if there is no illumination variation. By attention to our experiments and properties of the three bands, our proposed method limits the search area of the discriminant coefficients to avoid reaching an unreasonable results, which is a consequence of the small number of the training samples. Furthermore, limitation of the search area decreases the computational cost, and it is sufficient to analyse the DP for the coefficients inside the area. A premask(pm) ,like the one shown in Fig. 3e, limits the search area of the discriminant coefficients. The coefficients outside the mask are not used for recognition .Large DP values outside the premask most probably result from the small number of the training samples . Suitable size and shape of the premask depends on the database specifications. Size of the premask is determined by the number of coefficients which is desired for selection.Very small size premask loses the discriminant coefficients,and very large premask has no significant effect and increases the computational complexity. As an experimental rule, the number of coefficients inside the premask must be 1.5–2times greater than the number of coefficients which is desired for selection.High frequencies always masked, but masking of low frequencies depends on the illumination variation of the database.

2.5. Connection with PCA and LDA

Principal component analysis (PCA) and linear discriminant analysis (LDA) are two powerful tools for feature extraction and data representation. They have been successfully used in human face recognition. Their work immediately led to the PCA Eigen face technique . LDA extracts features that are most efficient for discrimination, while PCA extracts features that are most efficient for representation, which may not be very useful at all from a classification standpoint Toper form PCA or LDA ,the two-dimensional (2-D) image matrices must be previously transformed into one-dimensional (1-D)vectors. The resulting vector space is usually of a very high dimensionality :An image of 256X256 resolutions thenbecomesa65,536-dimensionalvector.The high dimensionality makes PCA and LDA difficult to implement. On the one hand, PCA and LDA both involve eigen value decomposition, which is extremely time consuming for high dimensional vectors. On the other hand, if the dimensionality exceeds the number of data points, LDA becomes a singularity problem and cannot be solved directly. To avoid the implementation difficulties of PCA and LDA, it is necessary to decrease data dimensions with a preprocessing step.For this purpose, PCA and LDA can be implemented in the DCT domain. In this way, data dimension is reduced by selecting some of the coefficients with a zigzag manner. Also, it is a good idea to use DPA as a preprocessing step before PCA and LDA, finding discriminant coefficients and data dimensions reduction.We call these modifications of PCA and LDA in the DCT domain as DPA–PCA and DPA–LDA, respectively. PCA and LDA transfer data from original space to feature space against DPA, which finds discriminant coefficients in the original space.

3. Multiclass Classification

SVM was used for classification of extracted features into different expression categories. SVM is a popular machine learning algorithm which maps the feature vector to a different plane, usually to a higher dimensional plane, by a non-linear mapping, and finds a linear decision hyper plane for classification of two classes.

4. The whole procedure

In order to evaluate the proposed method, a simplified face recognition system has been used. The block diagram is illustrated step-by-step in the following:

(1) Divide the database into training and test sets randomly.

(2) Compute the DCT coefficients of all the images of the training set .

(3) Normalize the DCT coefficients as previously described

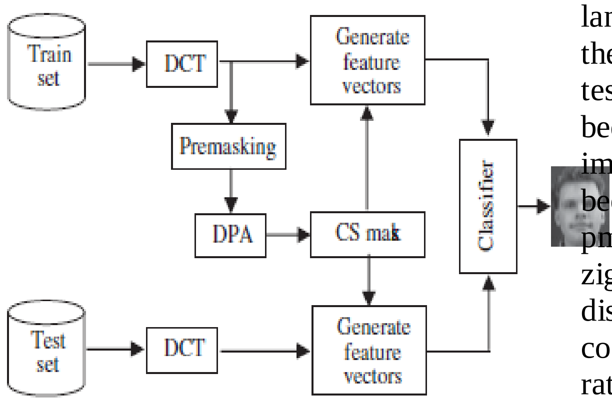

Fig. 3. Procedure of proposed approach

(4) Limit the search area of the discriminant coefficients by a premasking and set the elements of the DP matrix by zero for the coefficients which are outside the premask.

(5) Use DPA to estimate DP for all of the coefficients inside the premask.

(6) Find the n (n is the desired number of features) largest values of DP matrix and mark their corresponding positions. Construct an MXN CS mask matrix and set the marked positions to one and others to zero.

(7) Multiply DCT coefficients of each image by the CS mask, which is developed in previous step, and construct feature vectors by converting 2Dmatrices to 1Dvectors.

(8) Classify the test set using SVM.

For each SVM classifier, a model for probability estimates is trained. Scaling before applying SVM is very important to avoid features with huge ranges from dominating others. SVM with a radial basis function (RBF) kernel is used . RBF kernel transforms the features into a higher dimensional space, so that they can be then linearly separated by a hyperplane.

5. Experimental results and discussion

In order to evaluate the proposed technique, experiments are performed on facial databases created. I created video sequences for 6 basic emotions plus neutral. The database contains videos of high quality gathered from 20 different individual including changes in illumination conditions, subjects wearing eye glasses and different facial expressions. The experiments are programmed in the MATLAB language(Version 7). In all of the simulations, the database is randomly divided to train and test sets. Six images of each individual, have been used as the training set and the other images as the test set. Two premasks have been used in our simulations as:premask1 or pm1 and premask2 or pm2. Also the notation zigzag(dn) represents the zigzag manner by discarding the first n coefficients. In order to compare the approaches, the false recognition rate is computed in each simulation with the given formula

false recognition rate=$\frac{n_m}{n_t}$

where n_m is the number of misclassified images and n_t is the number of the test set images.|

5.1. Number of training images

The number of the training images has an important effect on the performance of each pattern recognition system. The overall performance of the approach deteriorates by decreasing the number of training images and vice versa. DPA is most sensitive to the number of training images. Limiting the search area of DPA reduces its sensitivity. The preference of pm2+DPA over DPA becomes more distinctive when the number of training images is decreased. Twenty-five coefficients have been used in this simulation. This database involves 200 frontal facial images, with 10 images of 20 individuals(anger-40, disgust-40, fear40, sadness-40, happy-20, surprise-20). The size of each image is 243X320 with256graylevels. Each image is scaled down to the size of 60X80 pixels. Simulation was repeated with different number of coefficients and similar results were achieved.

5.2. Combining with PCA and LDA for feature extraction

DPA+PCA and DPA+LDA are the new modifications of conventional PCA and LDA approaches. In these new modifications, DPA selects discriminant coefficients and provides the input vectors for PCA and LDA .These approaches have a difference on the CS section and using DPA decreases the false recognition rate. Among the PCA with conventional CS approaches, zigzag(d3) reaches the minimum false recognition rate, but pm2+DPA reaches a smaller error with less computations. DPA+LDA significantly decreases the false recognition rate, but limiting the search area has a small reduction of error rate.

5.3 SVM Classification

SVM was used for classification of extracted features into different expression categories. SVM is a popular machine learning algorithm which maps the feature vector to a different plane, usually to a higher dimensional plane, by a non-linear mapping, and finds a linear decision hyper plane for classification of two classes. Since SVM is a binary classifier, we implemented one-against-one (OAO) technique for multi-class classification . In OAO approach, a classifier is trained between each pair of classes; hence KC2 number of classifiers were constructed in total, where K is the number of classes. Using voting strategy, a vector can be classified to the class having the highest number of votes. After several experiments with linear, polynomial, and radial basis function (RBF) kernels, RBF kernels was selected for its superior classification performance.

5.4. Summary of the results

The difference of false recognition rate between various approaches is a function of feature numbers. DPA- based approaches achieve the performance of PCA/LDA or better with less complexity. An overall accuracy of 91.8 percent was obtained. The system performed worst for sadness expression as it misclassified sadness as anger.

Confusion Matrix

	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	38	0	2	0	0	0
Disgust	1	35	4	0	0	0
Fear	2	0	38	0	0	0
Happy	0	0	0	20	0	0
Sad	10	2	0	0	28	0
Surprise	0	0	0	0	0	20

Conclusion

This paper has presented a computationally efficient facial expression recognition system for accurate classification of the six universal expressions using features derived from DCT coefficients. DPA is a data- dependent approach which utilizes the statistical analysis in order to find the most discriminant coefficients. A pm limits the search area and reaps two benefits: first is the better performance especially with a small number of training images, and second is the computational cost reduction. Also a new modification to original PCA and LDA in the DCT domain was proposed, namely DPA+PCA and DPA+LDA. The experimental results on the facial database showed the advantage of DPA approach over conventional approaches. The proposed feature extraction method is as powerful as PCA/LDA. Using the appearance features , the system performs the one-against-one classification task and determines the expression based on majority vote. Experimental results showed that DPA-based approaches achieve the performance of PCA/LDA or better with less complexity. The proposed method can be implemented for any feature extraction and emotion recognition problem .

ACKNOWLEDGEMENT

First and foremost I thank the Almighty God for all the blessings. Special thanks to our principal, Prof. Madhavan Nambiar for providing us with excellent library which were absolutely necessary for the completion of seminar. I extend my heartfelt gratitude to Asst.Prof. Shruti K for her whole hearted support, advice and suggestions as a guide to the project.

REFERENCES

[1] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Computing Surveys 35 (2003) 399–458.

[2] Saeed Dabbaghchian MasoumehP. Ghaemmaghami, Ali Aghagolzadeh Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology. , Pattern Recognition 43 (2010) 1431–1440

[3] M. Kirby, L. Sirovich, Application of the Karhunen–Loeve procedure for the characterization of human faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1) (1990) 103–108.

[4] M. Turk, A. Pentland, Eigenfaces for recognition, International Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.

[5]K. Etemad, r. Chellappa, discriminant analysis for recognition of human face images, journal of the optical society of america a 14 (8) (1997) 1723

[6]J. Whitehill, M. S. Bartlett, and J. Movellan, “Automatic facial expression recognition,” in Social Emotions in Nature and Artifact. London, U.K.: Oxford Univ. Press, 2013.

[7] Fasel. B and Luettin. J, (2003) Automatic Facial Expression analysis: A survey, Pattern Recognition, , 36: 259–75.

[8] H. K. Ekenel. A Robust Face Recognition Algorithm for Real-World Applications.
PhD thesis, Universitat Karlsruhe (TH), Karlsruhe, Germany,
Feb. 2009.

[9] Gehrig, T., Ekenel, H.K., A Common Framework for Real-Time Emotion
Recognition and Facial Action Unit Detection IEEE Workshop on CVPR
for Human Communicative Behavior Analysis, June 2011.

[10] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. A Survey of Affect
Recognition Methods: Audio, Visual, and Spontaneous Expressions. IEEE
Transactions on Pattern Analysis and Machine Intelligence, 31(1):3958,
2009.