Discrete Cosine Transformer for Facial Expression Recognition Using Support Vector Machine

G.U.Kharat and S.V.Dudul

Abstract--**This study aims at developing intelligent computers or robots that are mind implemented. As a first step we investigate the emotionally intelligent computers which can perceive human emotions. In this research paper we have implemented Discrete Cosine Transformer (DCT) Using Support Vector Machine (SVM) to recognize universally recognized six principal emotions namely angry, disgust, fear, happy, sad and surprise along with neutral. Physical parameters such as mean, entropy, energy, Variance etc. are extracted from the emotion expressers. Recognition rate of 100% for all six principal** ***emotions*** **along with neutral is obtained on training data set and 94.28% on testing data set (unseen examples)**

Keywords-- **Supprot Vector Machine (SVM), Confusion Matrix, Facial expression detection, Feature extraction, Discrete Cosine Transformer (DCT), Mean Square Error (MSE).**

I. INTRODUCTION

IN the present time of computer implemented machine civilization, many kind of machine systems or equipments offering services to human beings have been indispensable for leading our daily life, with increasing tendency in future. In fact not only the machine computational speed but also machine intelligence is playing an important role. It is argued that computer to be able to interact with human beings, it need to have ability to understand the emotional state of a person. Facial expression is the best channel for emotion recognition and making machine intelligent. Hence a concept of Active Human Interface (AHI) is emerged as a new paradigm for developing the technology to realize a heart to heart communication between human beings and intelligent machine [1] – [6]. As Mehrabian [7] reported that message transfer of spoken words is only 7%, voice information 38% and facial expression 55%, it has become crucial to use facial expression as a message transfer media between human beings and intelligent machine.

II. RELATED WORK

The problem of recognizing facial expressions had attracted the attention of Computer-Vision community [8]-[15]. Optical flow computation for recognizing and analyzing facial expressions is used by [15, 9, 11, 13]. Maze [11] approached facial expression recognition from both top to bottom and bottom-up directions. In both cases the focus was on computing the motion of facial muscles rather that the motion of facial features. Four facial expressions were studied: Surprise, Anger, Happiness, and Disgust. Essa and Petland [16] and Essa [17] have proposed a physical based approach for modeling and analyzing facial expressions. They proposed extending FACS model to the temporal dimensions thus calling FACS+, to allow combine spatial and temporal modeling of facial expression. They assumed that a mesh is originally overlaid on the face and then tracked all the vertices based on optical flow field throughout the sequence. The emphasis was on accuracy of capturing facial changes which is most essential to synthesis. Recognizing results were reported [17] on six subjects displaying four expressions and eyebrow rising. The top row of figure 1 shows AU2 (raising eyebrow) from the FACS model and linear actuation profile of the corresponding geometric points. The bottom row shows the observed motion of these control points of the expression raising eyebrow by FACS+.

Yaser Yaccob and Larry S. Devis [15] proposed an approach to analyze and classify facial expressions from optical flow .This approach was based on qualitative tracking of principal regions of the face & flow computation at high density gradient points. The mid level representation is commutated from the spatial and temporal motion fields. Authors carried out experiments on 30 subjects in a laboratory environment

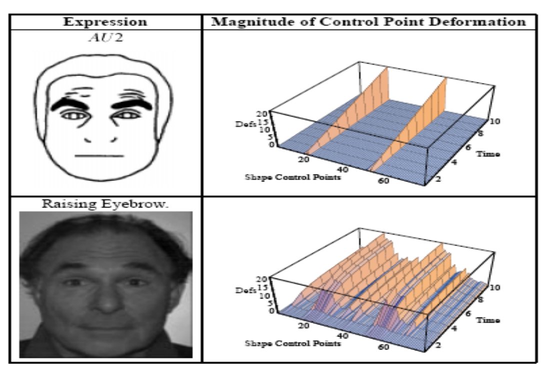

Fig1: FACS/CANDIDE deformation vs. Observed deformation for the Raising Eyebrow expression. Surface plots (top) show deformation over time for FACS actions AU2, and bottom for an actual video sequence of raising eyebrows.

Mase proposed an emotion recognition system that uses the major direction of specific facial mussels [18]. With 11 windows manually located in the face, the muscles movements were extracted by the use of optical flow. For classification K-nearest neighbor rule was used with an

G.U.Kharat, A.P., Department of EXTC, Anuradha Engineering College, Chikhli (Maharashtra), India. Email: gukharat@rediffmail.com
Dr. S.V.Dudul Prof & Dean (Research & Development), VIIT, Pune, (Maharashtra), India. dudulsv@rediffmail.com

accuracy of 80% with four emotions happiness, anger, disgust and surprise. Black [19] used parametric models to extract the shape and movements of the mouth, eye and eyebrows. They also build a mid and high level representation of facial actions with 89% accuracy. Tian [20] attempted to recognize action units developed by Ekman and Frisen [17], using permanent and transient features such as lip, nasolabial, furrow, and wrinkles. Geometric models were used to locate the shapes and appearances of these features.

We propose to use Discrete Cosine Transformer (DCT) to extract the features for facial expression and use of Support Vector Machine (SVM) for this pattern recognition problem.

III. Facial Expression Database

Database of facial expressions in seven emotions (including neutral) is collected for Japanese females. Ten expressers posed 3 to 4 examples of each of the seven emotions for a total of 219 images of facial expressions. This data was prepared when expresser look into the semi reflective plastic sheet towards camera. Hairs were tied away to expose all expressive zones of the face. Tungsten lights were positioned to create even illumination on the face. The box enclosed the region between camera and plastic sheet to reduce back reflections. The images were printed in monochrome and digitized using flatbed scanner. Sample images are shown in figure 2. This authentic database is collected from data repositories. Seventy images are selected for our experiment which uses ten expressers in seven emotions. Preprocessing is done by deblurring the images. Mean image is obtained to nullify the effect of difference in background effect.

IV. Support Vector Machine

Machine learning algorithms receive input data during a training phase, build a model of the input and output a hypothesis function that can be used to predict future data. Given a set of labeled training examples

$$S = ((x_1, y_1), \ldots, (x_i, y_i)),\ y_i \in \{-1, 1\}$$

Learning systems typically try to find a decision function of the form

$$h(x) = \mathrm{sgn}((w \cdot x) + b)$$

That yields a label $\in \{-1, 1\}$ (for the basic case of binary classification) for a previously unseen example x.

Support Vector Machines [27, 28] are based on results from statistical learning theory, pioneered by Vapnik [29, 30], instead of heuristics or analogies with natural learning systems. These results establish that the generalization performance of a learned function on future unseen data depends on the complexity of the class of functions it is chosen from rather than the complexity of the function itself. By bounding this class complexity, theoretical guarantees about the generalization performance can be made .SVMs perform an implicit embedding of data into a high dimensional feature space, where linear algebra and geometry may be used to separate data that is only separable with nonlinear rules in input space. To do so, the learning algorithm is formulated to make use of kernel functions, allowing efficient computation of inner products directly in feature space, without need for explicit explicit embedding. Given a nonlinear mapping Φ that embeds input vectors into feature space, kernels have the form

$$K(x, z) = (\Phi(x) \cdot \Phi(z))$$

SVM algorithms separate the training data in feature space by a hyperplane defined by the type of kernel function used. They find the hyperplane of maximal margin, defined as the sum of the distances of the hyperplane from the nearest data point of each of the two classes. The size of the margin bounds the complexity of the hyperplane function and hence determines its generalization performance on unseen data. The SVM methodology learns nonlinear functions of the form:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{l} \alpha_i y_i K(x_i \cdot x) + b\right)$$

where the α_i are Lagrange multipliers of a dual optimization problem. It is possible to show that only some of the α_i are non-zero in the optimal solution, namely those arising from training points nearest to the hyperplane, called support vectors. These induce sparseness in the solution and give rise to efficient approaches to optimization. Once a decision function is obtained, classification of an unseen example x amounts to checking on what side of the hyperplane the example lies.

The SVM approach is highly modular, allowing domain specific selection of the kernel function used. Table 1 gives the kernels used during our evaluation of SVM-based expression classification. In contrast to previous "black box" learning approaches, SVMs allow for some intuition and human understanding. They deal with noisy data and over fitting (where the learned function perfectly explains the training set but generalizes poorly) by allowing for some misclassifications on the training set. This handles data

Table 1
Typically used kernel functions. C, Γ, degree are parameters used to define each particular kernel from the family given.

Kernel	Formula
Linear	$x \cdot z$
Polynomial	$(\gamma x \cdot z + c)^{degree}$
Radial Basis Function (RBF)	$\exp(-\gamma \lvert x - z\rvert^2)$
Sigmoid	$\tanh(\gamma x \cdot z + c)$

that is linearly inseparable even in higher space. Multi-class classification is accomplished by a cascade of binary classifiers together with a voting scheme. SVMs have been successfully employed for a number of classification tasks such as text categorization [22], genetic analysis [23] and face detection [24]. Their high classification accuracy for small training sets and their generalization performance on data that is highly variable and difficult to separate make SVMs particularly suitable to a real time approach to expression recognition in video. They perform well on data that is noisy due to pose variation, lighting, etc. and where often minute differences distinguish expressions corresponding to entirely different emotions.

V. RECOGNIZING FACIAL EXPRESSIONS

In our research work various neural networks namely Multilayer Perceptrons (MLP), Recurrent Feedback Network (RBF) & Support Vector Machine (SVM) are used for training the networks. Vigorous experimentation is done by varying training parameters. It is found that SVM provides optimal result. For the classification of emotions, total 29 parameters, including 22 Eigen values of a Eigen vector and seven physical parameters namely mean, standard deviation, entropy, contrast, energy, homogeneity and correlation are inputted to the SVM network. So SVM uses 29 processing elements in input layer. Output is seven emotions hence in out put layer seven processing elements are used. Sensitivity analysis is performed on input data and from histogram it is observed that 15 parameters are most sensitive, eight eigen values of eigen vector and seven physical parameters. Hence an optimal vector is obtained containing 15 parameters as inputs. SVM used for training and testing the network consists of 15 input neurons and 7 output neurons. 50 % examples are used for training and remaining 50% for testing. By varying step size, maximum number of epochs, number of runs, an optimal network is obtained.

VI. EXPERIMENT

The block schematic of facial expression recognition system is given in figure1. We have developed a program in MATLAB to obtained DCT and physical parameters of images in the data set. An intelligent Vector is obtained containing optimal DCT values and seven physical parameters of the images in the data set. Vigorous experimentation is done by selecting proper number of epochs, number of runs, learning rate, step size on randomize data set to generalize the problem.

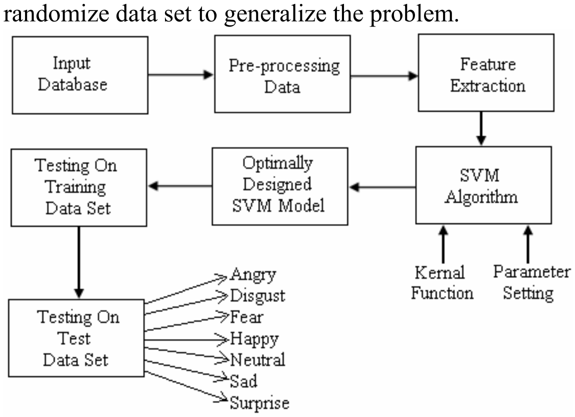

Fig2: Procedure for Facial Emotion Recognition.

The optimally designed SVM is fed with 90% data for learning and 10% data for cross validation. Optimally designed SVM is -

Input Neurons = 71
Output Neurons. = 07
Input Exemplars = 189

For output layer –

Neurons = 07
Step size = 0.01

Supervised Learning Control –

Maximum epochs = 1000
Weight update = Batch

Train N Times –

Epochs = 1000
Number of runs = 03

Termination after 100 epochs without improvement. Network is tested on training data set and testing data set.

VII. RESULTS.

The optimally design Support Vector Machine is tested on the training dataset. The results obtained are excellent. We got 100% recognition rate for all the six principal emotions namely Angry, Disgusts, Fear, Happy, Sad and Surprise along with Neutral. Results are shown in table 2 and 3. Finally the network is tested on the unseen examples (Test Data Set) we obtained 100% recognition result for Angry, Disgusts, Fear, Happy, Neutral, Sad and 66% for surprise emotion. Results are shown in table 4 and 5.The overall recognition rate on testing data set is 94.28%.

VIII. CONCLUSION AND DISCUSSION

In this research paper we proposed SVM for classification of emotions using Discrete Cosine Transformer and Physical Parameters. We achieved 100% classification result for all principal emotions along with Neutral on training data set and 100% results for Angry, Disgusts, Fear, Happy, Neutral and Sad. The recognition result for surprise emotion is 66%. It indicates that the proposed SVM model using DCT is less efficiency in recognizing surprise emotion.

We form a group of 45 peoples to recognize emotions manually from Japanese female data set. 15 students, 15 professors and 15 supporting staff members of different sex and age are included in this group. Different photographs from data set are shown in random fashion to all these 45 peoples, each photograph for 30 seconds. They are asked to recognize the emotions. It is observed that for a particular expresser in a dataset failure rate is very high for surprise emotion. It indicates that the said person in the data set is unable to express the surprise emotion correctly. This might be a cause for reduction in recognition rate of surprise emotion in test data set. In future we plan to perform semantic rating for classification of emotions. Again a large set of data will be prepared to generalize the results and results will be compared with authentic data sets available in data repositories.

KA.AN3.41	KM.HA3.6	KM.HA3.6	KL.DI3.173	NM.FE3.112	NA.SA3.207	TM.NE3.179

Figure 3: Images of Japanese females in various emotions from database.

TABLE 2
CONFUSION MATRIX FOR TRAINING DATA SET

Output / Desired	*Oa*	*Od*	*Of*	*Oh*	*On*	*Os*	*Osu*
Oa	28	0	0	0	0	0	0
Od	0	26	0	0	0	0	0
Of	0	0	28	0	0	0	0
Oh	0	0	0	26	0	0	0
On	0	0	0	0	29	0	0
Os	0	0	0	0	0	27	0
Osu	0	0	0	0	0	0	25

TABLE 3
PERFORMANCE PARAMETERS FOR TRAINING DATA SHEET

Perfor mance	***Oa***	***Od***	***Of***	***Oh***	***On***	***Os***	***Osu***
MSE	0.004573094	0.004941307	0.005142552	0.005426207	0.005125951	0.005980763	0.005362856
NMSE	0.036236801	0.041648994	0.040749133	0.045736084	0.039462089	0.048842896	0.046723557
MAE	0.060578007	0.060623876	0.064835764	0.06277692	0.064641815	0.067253203	0.061942866
Min Abs Error	0.001386537	0.000248042	0.001571136	0.003599081	0.000107029	0.000126913	0.002192829
Max Abs Error	0.138233983	0.153507803	0.199924032	0.191471734	0.162957982	0.192632016	0.199459873
r	0.994994142	0.991914549	0.994673533	0.991711097	0.993992509	0.991561998	0.994784949
Percent Correct	100	100	100	100	100	100	100

TABLE 4
CONFUSION MATRIX FOR TEST DATA SET.

Output / Desired	*Oa*	*Od*	*Of*	*Oh*	*On*	*Os*	*Osu*
Oa	2	0	0	0	0	0	0
Od	0	4	0	0	0	0	1
Of	0	0	2	0	0	0	0
Oh	0	0	0	4	0	0	0
On	0	0	0	0	1	0	1
Os	0	0	0	0	0	3	0
Osu	0	0	0	0	0	0	3

TABLE 5
PERFORMANCE PARAMETERS FOR TEST DATA SHEET

Performance	Oa	Od	Of	Oh	On	Os	Osu
MSE	0.029818304	0.045804112	0.046595851	0.065397661	0.05985745	0.085148886	0.097772501
NMSE	0.346049259	0.297053139	0.540757117	0.424123067	1.319856769	0.695382566	0.538970914
MAE	0.136381569	0.164640674	0.180957422	0.196586838	0.21184576	0.212855451	0.248942027
Min Abs Error	0.001582893	0.000859621	0.013854721	0.000672441	0.000543104	0.003794222	0.065073572
Max Abs Error	0.400283368	0.482096299	0.422336925	0.570485048	0.425769503	0.837479803	0.815861978
r	0.877222777	0.87419135	0.822805022	0.860489337	0.578051815	0.605617008	0.779557963
Percent Correct	100	100	100	100	100	100	60

IX. REFERENCES

[1] H. Kobayashi and F. Hara: The Recognition of Basic Facial Expressions by Neural Network, *International Joint Conference on Neural Network, pp.460/466* (1991-1 1)

[2] H. Kobayashi and F. Hara: Recognition of Six Basic Facial Expressions and Their Strength by Neural Network, *FEE International Workshop on Robot and Human Communication, pp.381/386* (1992-09)

[3] H. Kobayashi and F. Hara: Recognition of Mixed Facial Expressions by Neural Network, *IEEE International Workshop on Robot and Human Communication, pp.381/386*(1992-09)

[4] H. Kobayashi and F. Hara: Dynamic Recognition of Basic Facial Expressions by Discretetime Recurrent Neural Network, *Proceedings of International Joint Conference on Neural Network, pp. 1 SS/158* (1993)

[5] F. Hara and H. Kobayashi: Computer Graphics for Expressing Robot-Artificial Emotions, *IEFE International Workshop on Robot and Human Communication, pp. 155/160* (1992)

[6] H. Khayashi and F. Hara: Study on Face Robot for Active Human Interface, *IEEE International Workshop on Robot and Human Communication, pp.276/281* (1993-1 1)

[7] Amehrabian, Communication without words, *Psychology today, vol.2, no.4, pp.53-56*, 1968.

[8] D. Beymer, A. Shashua, and T. Poggio, *Example Based Image Analysis and Synthesis,* M.I.T. A.I. Memo No. 1431, 1993.

[9] I.A. Essa and A. Pentland, A Vision System for Observing and Extracting Facial Action Parameters, Prai. *IEEE CVPR , pp.76-83,*1994.

[10] H. Li, P. Roivainen, and R. Forcheimer, 3D Motion Estimation in Model-Based Facial Image Coding, *IEEE Trans. Pattern Analysis and Machine Iritelligence, vol. 15, pp. 545-555,*1993.

[11] K. Mase, Recognition of Facial Expression from Optical Flow, *IEICE Trans., vol. E 74, pp. 3,474-3,483,* 1991.

[12] K. Matsuno, C. Lee, and S. Tsuji, Recognition of Human Facial Expressions without Feature Extraction, *Puoc. ECCV, pp. 513-520*, 1994.

[13] M. Rosenblum, Y. Yacoob, and L.S. Davis, Human Emotion Recognition from Motion Using a Radial Basis Function Netbvork Architecture, *lEEE Workshop Motion of Non-Rigid and Articulated Objects, Austin, Texas, pp. 43-49,* Nov. 1994.

[14] D. Terzopoulos and K. Waters, Analysis and Synthesis of Facial lmage Sequences Using Physical and Anatomical Models, *lEEE Trans. Pattern Analysis and Machine Intelligence, vol. 15, pp. 569-579,* 1993.

[15] Y.Yacob and L Devis. Recognizing Human facial expression from long image sequences using optical flow. *IEEE transaction on Pattern Analysis and Machine Intelligence [PAMI], 18{6}:636-642,* 1996.

[16] Irfan A. Essa and Alex P. Pentland, Coding, Analysis, Interpretation, and Recognition of Facial Expressions. *IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 19, No. 7,* July 1997.

[17] Analysis, Interpretation and Synthesis of Facial Expressions by Irfan Aziz Essa B.S., Illinois Institute of Technology (1988) S.M., Massachusetts Institute of Technology (1990) Submitted to the Program in Media Arts and Sciences, MIT Media Laboratory, in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY February 1995.

[18] Mase K., Recognition of facial expression from optical flow. *IEICE Transc., E. 74(10):3474–3483,* 0ctober 1991.

[19] Black, M. J. and Yacoob, Y. Tracking and recognizing rigid and non-rigid facial motions using local parametric model of image motion. In Proceedings of the *International Conference on Computer Vision, pages 374–381. IEEE Computer Society, Cambridge, MA,* 1995.

[20] Tian, Ying-li, Kanade, T. and Cohn J., Recognizing Lower Face Action Units for Facial Expression Analysis. Proceedings of the *4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00) , pp. 484 – 490.* March 2000.

[21] C. Darwin, *The expression of the emotions in man and animals.* (University of Chicago Press, 1965. Original work published in 1872).

[22] T. Joachims. Text categorization with support vector machines: Learning with many relevant features. In *Proceedings of ECML-98, 10th European Conferenceon Machine Learning*, pages 137–142, Heidelberg, DE,1998. Springer Verlag.

[23] A. Zien, G. R.atsch, S. Mika, B. Scholkopf, T. Lengauer, and K. Muller. Engineering support Vector machine kernels that recognize translation Initiation sites. *Bioinformatics*, 16(9):799–807, 2000.

[24] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: An application to face detection. In Proceedings *of Computer Vision and Pattern Recognition '97*, pages 130–36, 1997.

[25] B. Boser, I. Guyon, and V. Vapnik, A training algorithm for optimal margin classifiers. In D. Haussler, editor, Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, pages 144–152, 1992.

[26] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines and other Kernel-based Learning Methods. (Cambridge University Press, Cambridge, UK, 2000).

[27] V. N. Vapnik, The Nature of Statistical Learning Theory (Springer-Verlag New York, Inc., New York, NY, USA, 1995).

[28] V. N. Vapnik, *Statistical Learning Theory* (Wiley, New York, NY, USA, 1998).

X. BIOGRAPHY:

G. U. Kharat, born on February 13th, 1966 in Chikhli (Maharashtra State) India. He received B. E. degree in Electronics from Amravati University (Maharashtra) in 1989 and M. E. degree in Control System from Shivaji University Kolhapur (Maharashtra) in 1995 in distinction. He is having 18yrs of teaching experience in Engineering colleges in Amravati University & Pune University.

Dr. S. V. Dudul, born on August 28th, 1964. He is a gold medalist in B. E. (Electronics & Power) from Nagpur University (Maharashtra) in 1986. He received M. E. (Electronics) in distinction from Shivaji University Kolhapur(Maharashtra) and Ph.D. in Computational intelligence in 2003 from Amravati University (Maharashtra). His having 20yrs. Of teaching experience, 5yrs of administrative experience and 8yrs of research experience. He has 59 research papers/publications. He received Eminent Engineer Award in 2005 by the Institution of Engineer's (India) Amravati region