2014 IEEE International Conference on Vehicular Electronics and Safety (ICVES)
December 16-17, 2014. Hyderabad, India

Face Pose Estimation for Driver Distraction Monitoring by Automatic Clustered Linear Discriminant Analysis

Hari C.V.[1] and Praveen Sankaran[2]

***Abstract*— Smooth varying data is hard to classify/divide to separate classes since there is small separation. Large number of close and adjacent poses create smooth varying manifolds. Manual class formation by selecting different data points from entire database into different training classes will affect the error rate in smooth varying data classification. This paper proposes classification of smooth varying data based on clustering and discriminant analysis. The clustering process results in different clusters which can be used for classification based on discriminant analysis. The automated class formation based on the data points in the manifold reduces effort of manual clustering and it gives very comparable results. This pose estimation can be used as a measure of driver distraction monitoring.**

I. INTRODUCTION

Separation between continuous face pose data points is assumed to be very small. Generally such points are considered as belonging to a smooth varying manifold. Classification of such manifolds is a challenging problem since class separation between them would be very small. Cluster formation during training is also a challenging task for getting minimum error during classification. We propose in this paper a clustered Linear Discriminant Analysis (LDA) method, that will group smooth data points into optimal classes. LDA features are now obtained using these classes. This pose estimation measure can be taken as one of the cues to determine driver distraction. Development of an automated system that would alert the driver of a vehicle to distractions of self will reduce the number of motor vehicle accidents because nearly 40 - 50 % accidents are due to driver distractions.

Chutorian and Trivedi[1] gives a detailed survey on different head pose estimation methods and discuss various advantages and disadvantages of several methods. They also compared different methods in terms of classification accuracy or mean absolute error. They classified the manifold embedding technique in to three different categories - Linear Subspaces, Kernelized Subspaces and Nonlinear Subspaces. In pose estimation, after the modelling of manifold, dimensionality reduction methods and regression analysis can be used for low dimensional embedding of test data.

Mckenna and Gong [2] used the most prominent linear dimensionality reduction technique Principal Component Analysis (PCA) for pose changes visualization in low dimensional manifold subspaces. They used gabor wavelets for real time pose estimation. If the number of samples per classes increases, PCA performance decreases. Then LDA [3] can be used for getting more class separability. LDA maximizes ratio of between class variance to the within class variance. Tangkuampien and Suter [4] showed the dimensionality reduction method Kernel Principal Component Analysis (KPCA) [5] and a manifold embedding method Image Euclidean Distance [6] can be used for 3D pose estimation.

Non linear dimensionality reduction methods - Isometric feature mapping (Isomap) [7], Locally Linear Embedding (LLE) [8] and Laplacian Eigenmaps [9] are also using for pose estimation. But, the main disadvantage of these methods is the unavailability of projection matrices for new test samples. So, regression analysis, neural networks or any other similar methods are used for pose estimation after these non linear dimensionality reductions. Balasubramanian *et.al* [10] used Biased Manifold Embedding (BME), which uses the pose angle information for calculating the neighbourhoods in the feature space which are then used for dimensionality reduction.

Xu and Wunsch [11] prepared a survey on different clustering algorithms used in different parts of science. They also discussed different applications and validations of clustering. Clustering algorithms try

[1]Research Scholar, Department of Electronics & Communication Engineering, National Institute of Technology Calicut, Kerala, India - 673 601 haricv@gmail.com
[2]Assistant Professor, Department of Electronics & Communication Engineering, National Institute of Technology Calicut, Kerala, India - 673 601 psankaran@nitc.ac.in

TABLE I: Manual Clustering - Cluster Information

N	Number of Poses/Persons per Class	Positive Side		Negative Side	
		Number of Clusters	Cluster Information	Number of Clusters	Cluster Information
6	30	3	1^{st} class consists of pose angles from 1 to 30, 2^{nd} class consists of pose angles from 31 to 60 and so on	3	1^{st} class consists of pose angles from -1 to -30, 2^{nd} class consists of pose angles from -31 to -60 and so on
12	15	6	1^{st} class consists of pose angles from 1 to 15, 2^{nd} class consists of pose angles from 16 to 30 and so on	6	1^{st} class consists of pose angles from -1 to -15, 2^{nd} class consists of pose angles from -16 to -30 and so on
18	10	9	1^{st} class consists of pose angles from 1 to 10, 2^{nd} class consists of pose angles from 11 to 20 and so on	9	1^{st} class consists of pose angles from -1 to -10, 2^{nd} class consists of pose angles from -11 to -20 and so on
36	5	18	1^{st} class consists of pose angles from 1 to 5, 2^{nd} class consists of pose angles from 6 to 10 and so on	18	1^{st} class consists of pose angles from -1 to -5, 2^{nd} class consists of pose angles from -6 to -10 and so on

to separate data into different subsets with maximum internal similarity. Cluster analysis can be divided into four steps - Feature selection or extraction, Clustering algorithm design or selection, Cluster validation and Results interpretation. The different clustering algorithms are classified as distance and similarity measures based, hierarchical, squared error (Vector Quantization) based, pdf estimation via mixture densities, graph theory based, combinatorial search techniques, fuzzy, neural networks based, kernel based, sequential data and large scale data sets methods like CLARA, CURE, CLARANS, BIRCH . . . *etc*. K means [12] clustering algorithm is a well known squared error type clustering and in this method the cluster centroids are recomputed when new sample joins a cluster.

In distance and similarity measures clustering standard distance measurements (Minkowski distance, Euclidean distance, City block distance . . . *etc*) are used. Based on the distance between different data points the entire data is divided into different clusters. In hierarchical clustering, the data organizes in to larger groups, which contain smaller groups and so on. The different hierarchical clustering algorithms are single linkage or nearest neighbour method, complete linkage or farthest neighbour method, average linkage and median linkage methods [12].

II. Driver Distraction Monitoring

Driver distraction is the diversion or change in attention of the driver from his/her driving tasks. This distraction occurs because the driver temporarily performs additional tasks and this causes change in attention and accidents [13]. The main objective of this work is the development of an automated system that would alert the driver of a vehicle to distractions of self. We can monitor the driver by attaching a camera inside the vehicle. After the separation of the video in to frame by frames, the Viola Jones method [14] will identify the face region. Head orientation will form one of the cues to determine driver distraction. For example, if the driver is distracted by a co-passenger, sitting on the back side of the driver, then the driver may turn his/her head. The second cue the work would focus on is the eyes. If the driver is sleepy during the driving, it will also cause an accident. So, by monitoring the eyes, especially the amount of iris that can be detected from the image/video under consideration, and then alerting the occupants, this type of accident can be reduced. Based on this distraction cue measurements, a driver alert monitoring system will give an alert to the driver and co-passengers for any distraction detected. The driver distraction monitoring system is as shown in Fig1. This research would develop an automated system based on image data collected from inside the to vehicle to alert the occupants of a distracted driver.

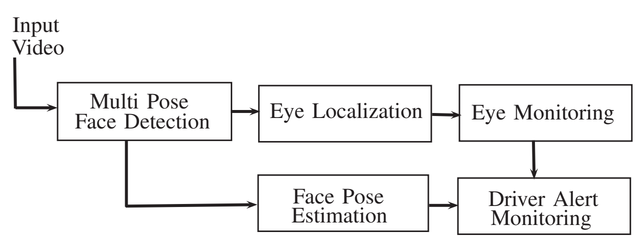

Fig. 1: Driver Alert Monitoring

III. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis is a dimensionality reduction method which can be used for classification problems if the number of training samples of each

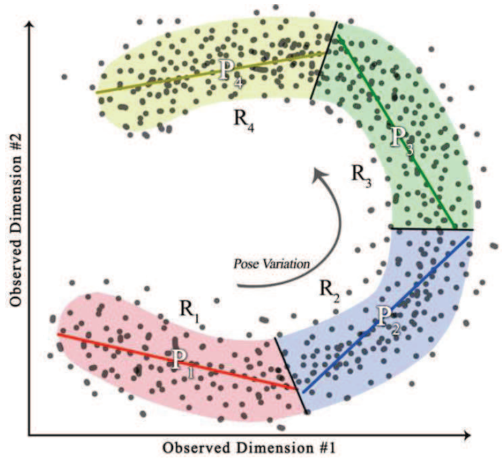

Fig. 2: Smooth Varying Manifolds in to separable Clusters - Conceptual Drawing[15].

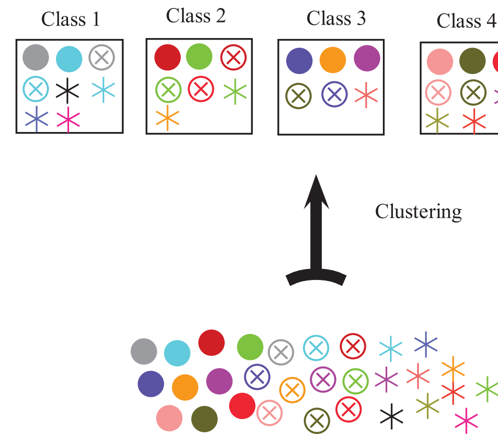

Fig. 3: Clustering - Strategy 1 Conceptual Drawing. (Entire database is used for clustering)

class are large. The objective of this method is to find a projection matrix which maximizes the ratio of between class variance to the within class variance for getting maximum class separation.

The between class scatter matrix is defined as

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T \tag{1}$$

and the within class scatter matrix is defined as

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \tag{2}$$

where μ_i is mean image of class X_i and N_i is the number of samples in class X_i and c is the number of classes. If S_w is non singular, the optimal projection W_{opt} is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between class scatter matrix of the projected samples to the determinant of the within class scatter matrix of the projected samples, where

$$W_{opt} = arg\ max \frac{|W^T S_B W|}{|W^T S_W W|} = [w_1\ w_2\ \ldots\ w_m] \tag{3}$$

where $\{w_i \mid i = 1, 2, \ldots m\}$ is the set of generalized eigen vectors of S_B and S_W corresponding to the m largest eigen values $\{\lambda_i \mid i = 1, 2, \ldots m\}$, i.e.,

$$S_B w_i = \lambda_i S_W w_i, \quad i = 1, 2, \ldots m \tag{4}$$

The maximum value of m is $c - 1$ [16].

A. Face Pose Estimation

The face pose estimation based on clustering and discriminant analysis consists of two steps. Initially the entire data set is divided into different classes for training. This can be done manually or automatically. Clustering methods can be used for automatic class formation for training. Then LDA projects the data in to a new plane and euclidean distance measurement is used for the classification in this plane.

We used $'K'$ means [12], [17], Fuzzy $'C'$ means [12] and Hierarchical [12] algorithms for clustering. We divide the smooth manifold into multiple disjointed, locally linear, separable clusters as shown in conceptual drawing in Fig.2.

Two different types clustering strategies are followed here. In strategy 1, pose images of all persons are clustered and forming different classes as shown in Fig.3. In strategy 2, pose images of a single person clustered as shown in Fig.4. Then, add same pose images of all other persons into corresponding classes as shown in Fig.5 and this final classes are used for clustering. Strategy 1 is more time consuming than strategy 2 because the first strategy contains larger number of data points for clustering.

In both cases, we initially clustered the data into various classes. LDA is applied on these classes and nearest neighbour algorithm with euclidean distance measurement is used for classification. The training image which is closest to the test image based on the euclidean distance measurement is the winner. The pose of the winning image is taken as the detected pose. Leave-one-out testing strategy is followed here. One

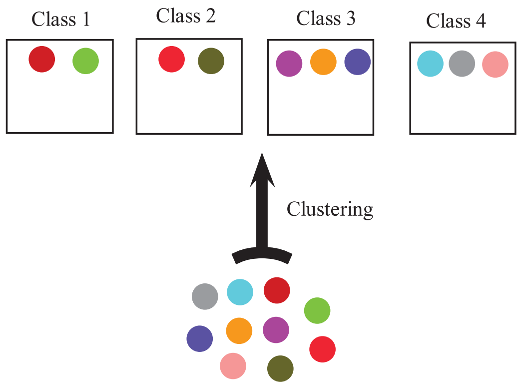

Fig. 4: Clustering - Strategy 2 Conceptual Drawing. (Pose images of a single person is used for clustering)

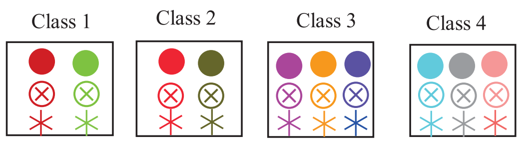

Fig. 5: Clustering - Strategy 2 Final Class Formation Conceptual Drawing.
(Pose images of all other persons into corresponding classes)

person's data is taken out from the database for testing, and the remaining data points are used for clustering and training. In each iteration, different data is kept out for testing. We iterated till all data points are used for testing. We used different clustering algorithms with varying the number of clusters.

Manual clustering does not explain or provide any logical or scientific base for forming classes from a smooth manifold. Hence we need to move to an automated process that would be based on some sound mathematical principle.

IV. Experimental Analysis

The proposed method is tested and evaluated on CUbiC FacePix database [18], [19]. Sample images from this database are as shown in Fig.6. This database consists of three different sets (each set consisting of 181 pose images) of face pose images of 30 people in an interval of $1°$. In positive side the angle varies from $1°$ to $90°$ and in negative side it varies from $-1°$ to $-90°$

First we manually created a $'N'$ number of clusters / classes from 180 poses. The different N values and its cluster formation details is as shown on Table I. We used LDA on these classes and calculated the mean

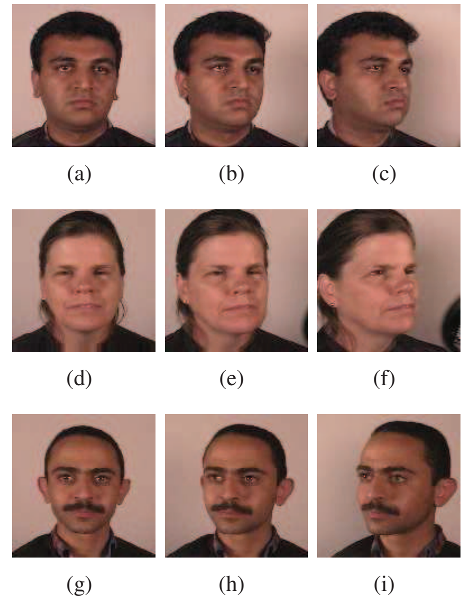

Fig. 6: Sample Test/Training Images

absolute error (MAE) based on the leave-one-out cross validation testing. It is defined as

$$MAE = \frac{1}{n}|\hat{\theta}_i - \theta_i| \tag{5}$$

where $\hat{\theta}_i$ is the estimated pose of the test input, θ_i is the original pose of the test image and n is the total number of samples used for testing.

Then, instead of manual clustering we used different clustering algorithms for class creation. We used K means, fuzzy C means and hierarchical clustering. We used single, average, complete and median linkage types in hierarchical clustering.

A. Results and Discussions

As discussed in the previous sessions, manual clustering followed by LDA is used on the data points initially. We used all data (total 30 persons) points for training and testing. Its MAE is as shown in Table II. From this analysis, it is very clear that the error is decreasing by increasing the number of classes for training.

Manual clustering is not possible for creating more number of clusters if the training images are large. Total number of data points used here are 5400 (30 persons $\times$ 180 poses $=$ 5400). In the automated class formation, we initially used all data points for clustering. Total number of data points used for

TABLE II: Manual Clustering and LDA - MAE (in degrees)

Number of Classes	6	12	18	36
MAE (in degrees)	16.74	6.72	6.79	5.52

clustering is 5220 (29 persons $\times$ 180 poses = 5220) because the remaining 180 pose images of a person is used for testing. The data points used for testing is not used for clustering (leave-one-out testing). We varied the number of clusters and iterated till all data points used for testing. We tested and evaluated this method using K means clustering and the result of this analysis is as shown in Table III. The main disadvantage of this method is that it is very time consuming.

TABLE III: All data points for clustering - MAE (in degrees)

Number of Clusters	10	18	25	75	150
MAE (in degrees)	14.26	13.58	13.60	13.42	15.27

TABLE IV: Single person's data points for clustering - MAE (in degrees)

Number of Classes		10	18	25	75	150
K Means		10.47	7.58	6.47	5.30	5.05
Fuzzy C Means		9.62	6.95	6.35	5.55	5.07
Hierarchical	**Single**	18.34	12.92	11.97	7.08	5.13
	Average	10.83	7.71	6.96	5.26	5.04
	Complete	11.50	7.95	6.29	5.28	5.03
	Median	11.19	7.87	6.33	5.22	5.02

So to avoid the time consuming problem and low accuracy, we used a single person's data for clustering. It forms different clusters. Then add the same poses of other person's in to the corresponding classes. It will reduce the operation time drastically and gives better results than the previous method. Here also, the leave-one-out testing strategy is used for testing. The person whose database is used for testing is not used for clustering and class formation. This method is tested and evaluated with K means, fuzzy C means and Hierarchical clustering algorithms by using entire database. Similar to the previous case, we calculated the mean absolute error and it is as shown in Table IV. It is very clear that, by increasing the number of clusters, error rate in the pose classification is reducing. Also, we can avoid the effort of manual clustering because it needs much more time to select each and every poses of each persons from a database contains more than 5000 images.

V. CONCLUSIONS

This paper proposed an automatic class formation based on clustering followed by discriminant analysis for human face pose estimation. The automated class formation for clustering reduces the effort of manual process. If we use all person's data, clustering could fail because the presence of multiple people creates a more complex manifold structure and hence clusters need not indicate pose variation. If single person's poses are used for clustering, it gives comparable results with manual clustering. It can also be noted that the former method is more time efficient. This face pose estimation can be used as a cue measurement of the automatic driver distraction monitoring.

REFERENCES

[1] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 4, pp. 607–626, 2009.

[2] S. J. McKenna and S. Gong, "Real-time face pose estimation," *Real-Time Imaging*, vol. 4, no. 5, pp. 333–347, 1998.

[3] S. Balakrishnama and A. Ganapathiraju, "Linear discriminant analysis-a brief tutorial," *Institute for Signal and information Processing*, 1998.

[4] T. Tangkuampien, D. Suter, *et al.*, "3d object pose inference via kernel principal component analysis with image euclidian distance (imed)." in *BMVC*. Citeseer, 2006, pp. 137–146.

[5] B. Schölkopf, S. Mika, A. Smola, G. Rätsch, and K.-R. Müller, "Kernel pca pattern reconstruction via approximate pre-images," in *ICANN 98*. Springer, 1998, pp. 147–152.

[6] L. Wang, Y. Zhang, and J. Feng, "On the euclidean distance of images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1334–1339, 2005.

[7] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[8] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[9] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[10] V. N. Balasubramanian, J. Ye, and S. Panchanathan, "Biased manifold embedding: A framework for person-independent head pose estimation," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–7.

[11] R. Xu, D. Wunsch, *et al.*, "Survey of clustering algorithms," *Neural Networks, IEEE Transactions on*, vol. 16, no. 3, pp. 645–678, 2005.

[12] E. Gose, R. Johnsonbaugh, and S. Jost, *Pattern recognition and image analysis*. Prentice-Hall, Inc., 1996.

[13] R. S. Committee *et al.*, "Inquiry into driver distraction-report of the road safety committee on the inquiry into driver distraction," *Parliamentary Paper No. 209 Session 2003-2006*, 2006.

[14] M. Jones and P. Viola, "Fast multi-view face detection," *Mitsubishi Electric Research Lab TR-20003-96*, vol. 3, p. 14, 2003.

[15] J. Foytik and V. K. Asari, "A two-layer framework for piecewise linear manifold-based head pose estimation," *International journal of computer vision*, vol. 101, no. 2, pp. 270–287, 2013.

[16] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 711–720, 1997.

[17] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. California, USA, 1967, pp. 281–297.

[18] J. A. Black, M. Gargesha, K. Kahol, P. Kuchi, and S. Panchanathan, "A framework for performance evaluation of face recognition algorithms," in *Proc. SPIE*, vol. 4862, 2002, p. 164.

[19] D. Little, S. Krishna, A. John Jr, and S. Panchanathan, "A methodology for evaluating robustness of face recognition algorithms with respect to variations in pose angle and illumination angle." in *ICASSP (2)*. Citeseer, 2005, pp. 89–92.