Emotion Recognition from Video Sequence

Saransh Srivastava
Atique Firoz
Advisor: Dr. Amitabh Mukherjee
{saranshs,atiquef,amit}@iitk.ac.in
Dept. of Computer Science and Engineering

April 10, 2012

Abstract

Over the last many years a lot of work is being done to make human-computer interaction more human-like experience. Areas like neural networks have gathered immense interest within the academia. In this work, we are working on detecting emotions from a video clip We use Constrained Local Models (CLM) for extracting the shape feature vectors from each video frame. On these vectors we applied an unsupervised k means clustering algorithm to form clusters of different emotions from a video clip. We tested the algorithms on the GEMEP - FERA 2011 Dataset.

1 Introduction

Human vision can experience emotion as associated with mood, temperament, personality and disposition. Computer Vision seeks to emulate the human vision by analyzing digital image as input. The fact that world is three- dimensional while computer vision is two-dimensional is basically one of the main problems that complicate Computer Vision.

With great advancement in technology in terms of different techniques of people interacting with each other it is quite necessary that one should be aware of current emotions of the person he/she is interacting.

Emotion recognition is not a new problem, Principal Component Analysis (PCA), Independent Component Analysis (ICA), Active Appearance Model (AAM) based emotion recognition has been done in the last decade. The advantages of using Principal Component Analysis and Independent Component Analysis were in the use of dimensionality reduction of data image. AAM based approach was used by Eric Patterson and Matthew S. Ratliff in 2008.

We are using CLM [1] based emotion recognition on the dataset SSPNET GEMEP-FERA that includes multiple facial expressions in a single video clips. In the training dataset it has 7 actors with different emotions with time (total 155 videos). Emotions such as anger, fear, joy, relief and sadness can be classified.

As overview, CLM fitting is done on the each frame of video clips to generate the feature vector of the feature points on the face, which is then clustered using unsupervised method of kmeans clustering to get the mean frame of each cluster.

2 Sequence of Events

2.1 CLM Fitting

Constrained Local Model Fitting makes independent predictions regarding locations of models landmark points, which are combined by enforcing a prior over their joint motion. Most fitting methods employ a linear approximation to how the shape of a non-rigid object deforms, coined the Point Distribution Method (PDM). It models non-rigid shape variations linearly and composes it with a global rigid transformation, placing the shape in image frame:

$$x_i = sR(\widehat{x_i} + \phi_i q) + t$$

Where x_i denotes the 2D-location of the PDMs ith landmark point and p = s, R, t, q denotes the parameters of PDM, which consist of Global Scaling s, a rotation R, a translation t, and a set of non-rigid parameters q.

CLM basically consist of two goals, which are (i) Perform an exhaustive local search for each PDM landmark around their current estimate using feature detectors, and (ii) optimize the PDM parameters such that the detection responses over all of its landmarks are jointly maximized. In the first step of CLM fitting, a likelihood map is generated for each landmark positions by applying local detectors to constrained regions around the current estimate. Once the responses for each landmark have been found, by assuming conditional independence, optimization proceeds by maximizing.

2.2 Kmeans Clustering

kmeans is an unsupervised learning algorithms that solve the well-known clustering problem. The main idea is to define k centroids, one for each cluster.

Algorithm is to place K points into the space represented by the objects that are being clustered. These points represent initial group centroids. Assign each object to the group that has the closest centroid. It is same as minimizing an objective function.

When all the objects have been assigned, reallocation of the positions of the K centroids is done, the same is repeated until the centroid no longer moves. This produces a separation of the objects into groups from which the metric to be minimized can be calculated. We use this clustering method to cluster the frames of approximate same emotional status into different clusters.

3 Result

We tested the algorithm on random videos from the internet having multiple emotions and on subset of SSPNET GEMEP-FERA-2011 dataset which has single emotion in any video clip. The emotion in the videos are anger, fear, joy, relief and sadness. The CLM algorithm was able to detect the face of the person in the video clip and track down all the 66 landmark points. The emotion in the video clip formed a cluster when passed through the k-means clustering algorithm. The resulting clusters gave an accurate result for 57 out of 71 videos tested.

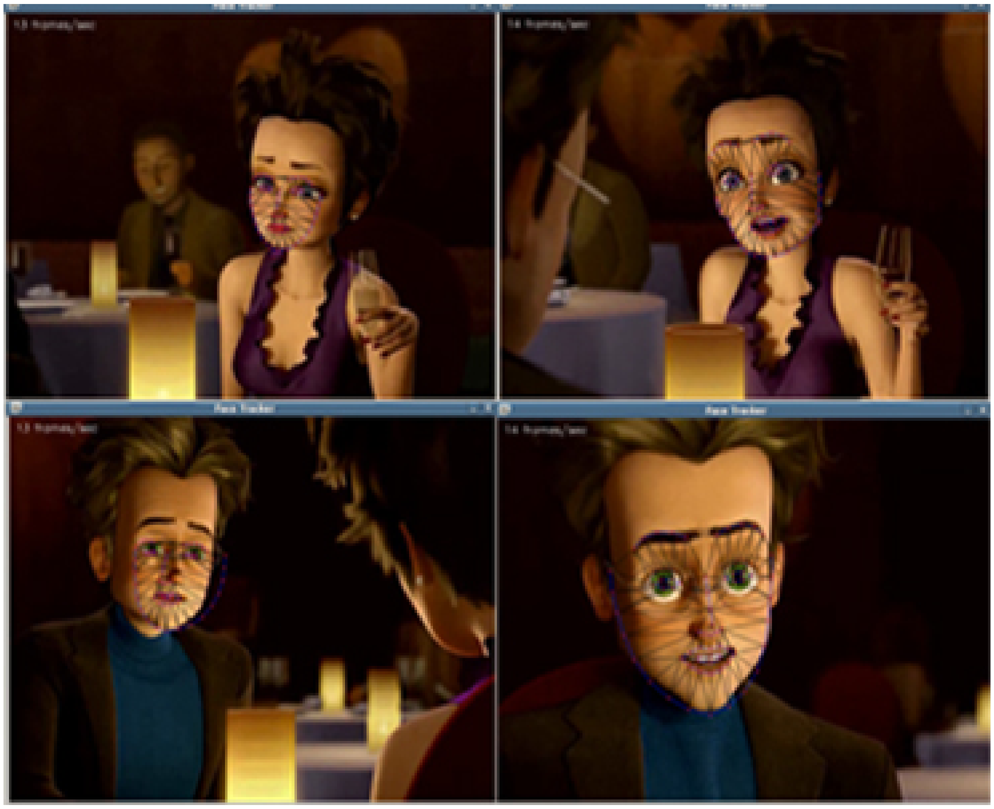

fig:1 CLM fitting on the faces, frames corresponds to the four different facial expressions
ref: Megamind video clip

In the video clip there are four different expressions which includes the neutral face of the girl, happy face of girl, neutral face of Megamind(boy) and excited face of Megamind. Based on the expressions in the video frames kmeans clustering gives a total of four different clusters corresponding to each expressions of the boy and girl.

In the graph below the green and yellow clusters are close enough because the shape vectors corresponding to the neutral faces of boy and girl have approximately the same value. Similarly the red and blue clusters are very far because the shape vectors have a large variation due to the changed expression of faces.

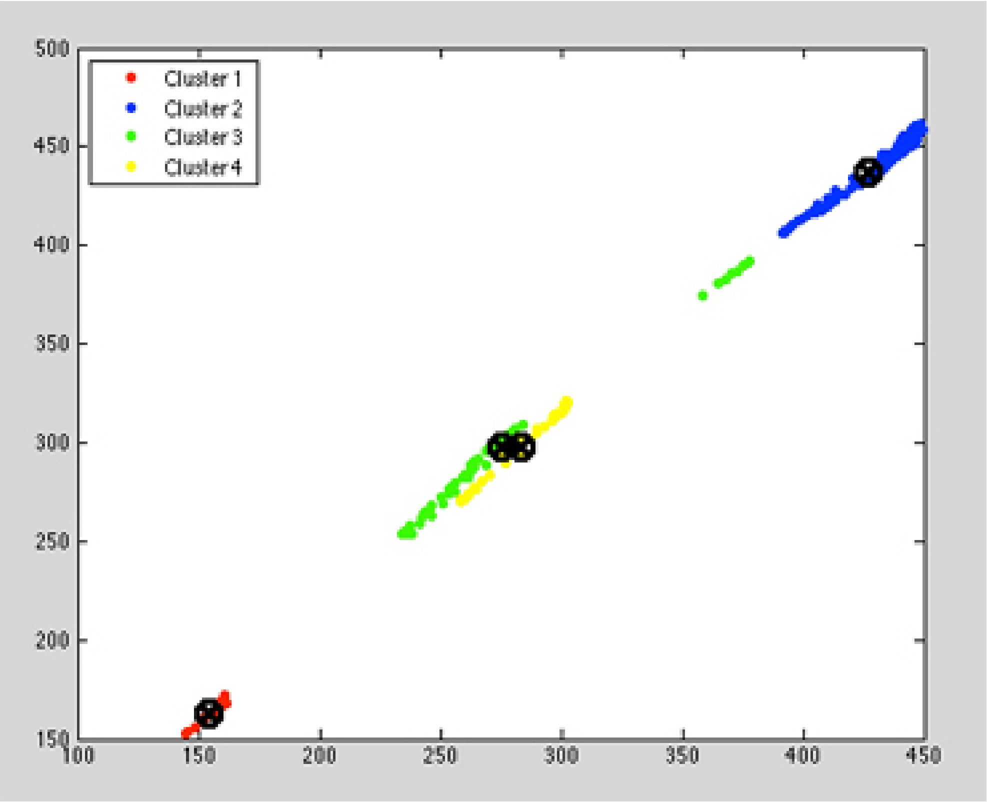

fig:2 unsupervised kmeans clustering of the frames gives four distinct clusters, two nearby clusters (yellow and green) correspond to the neutral faces of the boy and girl,whereas other two distinct clusters(red and blue) correspond to the excited faces of boys and girl.

4 Conclusion and Future Work

As part of a course project, due to time constraints the entire process of supervised clustering of the emotions using PHOG [2] algorithm could not be executed. We plan to complete the rest of the steps of the process described in the paper [3] in due course of time.

References

[1] J.M. Saragih, S. Lucey, and J.F. Cohn. Face alignment through subspace constrained mean-shifts. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1034 –1041, 29 2009-oct. 2 2009.

[2] Anna Bosch, Andrew Zisserman, and Xavier Munoz. Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM international conference on Image and video retrieval*, CIVR '07, pages 401–408, New York, NY, USA, 2007. ACM.

[3] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon. Emotion recognition using phog and lpq features. In *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 878 –883, march 2011.