www.ijecs.in

International Journal Of Engineering And Computer Science ISSN:2319-7242

Volume 3 Issue 10 October,2014 Page No.8765-8771

Temporal Segmentation of Facial Behavior in Static Images Using HOG & Piecewise Linear SVM

Preeti Saraswat [1], Srikanth G [2]

[1]Masters in Computer Science,Dept. of CSE, GITS, Udaipur, India

preetisaraswat17@gmail.com

2 Masters in Power System and Automation, Dept. of EEE, GITAM Hyderabad, Telangana, India

srikanthgvr1@gmail.com

Abstract: Temporal segmentation of facial gestures in spontaneous facial behavior recorded in real-world settings is an important, unsolved, and relatively unexplored problem in facial image analysis. Several issues contribute to the challenge of this task. These include non-frontal pose, moderate to large out-of-plane head motion, large variability in the temporal scale of facial gestures, and the exponential nature of possible facial action combinations. To address these challenges, we propose a two-step approach to temporally segment facial behavior. The first step uses spectral graph techniques to cluster shape and appearance features invariant to some geometric transformations. The second step groups the clusters into temporally coherent facial gestures. We evaluated this method in facial behavior recorded during face-to-face interactions. The video data were originally collected to answer substantive questions in psychology without concern for algorithm development. The method achieved moderate convergent validity with manual FACS (Facial Action Coding System) annotation. Further, when used to preprocess video for manual FACS annotation, the method significantly improves productivity, thus addressing the need for ground-truth data for facial image analysis. Moreover, we were also able to detect unusual facial behavior. This paper consists of efficient facial detection in static images using Histogram of Oriented Gradients (HOG) for local feature extraction and linear piecewise support vector machine (PL-SVM) classifiers. Histogram of oriented gradient (HOG) gives an accurate description of the contour of image. HOG features are calculated by taking orientation of histogram of edge intensity in a local region. PL-SVM is nonlinear classifier that can discriminate multi-view and multi-posture from the images in high dimensional feature space. Each PL-SVM model forms the subspace, corresponding to the cluster of special view. This paper consists of comparison of PL-SVM and several recent SVM methods in terms of cross validation accuracy.

Keywords: Facial detection, histogram of oriented gradients, classification, support vector machine.

1. Introduction

The detection of humans in images and videos especially is an important problem for computer vision and pattern recognition. Temporal segmentation of facial behavior from video is an important unsolved problem in automatic facial image analysis. With few exceptions, previous literature has treated video frames as if they were independent, ignoring their temporal organization. Facial actions have an onset, one or more peaks, and offsets, and the temporal organization of these events is critical to facial expression understanding and perception [1, 2, 3]. For automatic facial image analysis, temporal segmentation is critical to decomposing facial behavior into action units (AUs) and higher-order combinations or expressions [2], to improving recognition performance of facial expression recognizers, and to detecting unusual expressions, among other applications. Several factors make the task of recovering the temporal structure of facial behavior from video a challenging topic, especially when video is obtained in realistic settings characterized by non-frontal pose, moderate out-of-plane head motion, subtle facial actions, large variability in the temporal scale of facial actions (both within and between event classes) and an exponential number of possible facial action combinations. To address these challenges, we propose a two-step approach to temporally segment facial behavior. The first step uses spectral graph techniques to cluster shape and appearance features. The resultant clusters are invariant to some geometric transformations. The second step groups the clusters into temporally coherent facial gestures (Fig. 1).

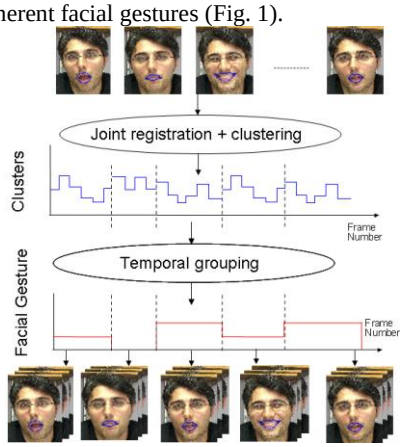

Figure 1. Temporal segmentation of facial gestures.

Facial expression is a challenging task in many fields because as humans are highly deformable objects whose appearance depends on numerous factors:

- variability of appearance of face due to the size, color that pedestrians may carry

- irregularity of shape: pedestrians may have different features
- variability of the environment in which they appear (usually pedestrians exist in a cluttered background in complex scenarios whose look is influenced by illumination or by weather conditions)
- Variability of the actions they may perform and positions they may have (run, walk, stand, shake hands etc).

In existing facial detection methods, feature representation and classifier design are two main problems being investigated. Visual feature descriptors have been proposed for facial detection including Haar-like features , HOG, v-HOG, Gabor filter based cortex features , covariance features , Local Binary Pattern (LBP) [4] , HOG-LBP [5], Edgelet [6], Shapelet [7], Local Receptive Field (LRF) [8], Multi-Scale Orientation (MSO) [9], Adaptive Local Contour [10], Granularity-tunable Gradients Partition (GGP) descriptors [11], pose-invariant descriptors [12], Practical Swarm Optimization[13] .

Recently, histogram of oriented gradients (HOG) and region covariance features are preferred for pedestrian detection. It has been shown that they outperform those previous approaches. HOG is a gray level image feature formed by a set of normalized gradient histograms. Shape and appearance of object can be well defined by the distribution of local intensity gradients or intensity gradients. HOG features are calculated by taking orientation histogram of edge intensity in a local region or block [14]. A HOG feature vector represents the local shape of object, giving edge information at plural cell. For the flatter regions like ground or wall of a building, the histogram of oriented gradients has the flatter distribution. In the border between object and background, one of the elements in the histogram has larger value and it indicates the direction of edge. The total numbers of HOG features are more and redundant. Hog features can be applied effectively for the classification of object having specific shape or appearance such as face, humans, bicycle, motor car etc, because they are based on the information on the edge.

Linear SVM is the most popular classifier with several reported landmark works for human detection. The reasons we selection of SVM classifiers is that, it is easy to train and, unlike neural networks, the global optimum is guaranteed. The extracted features on labeled samples are usually fed into a classifier for training. However, when we need to detect multi-view and multi-posture humans simultaneously in a video system, the performance of a linear SVM often drops significantly. It is observed in experiments that humans of continuous view and posture variations form a manifold, which is difficult to be linearly classified from the negatives. An algorithm that requires multi-view and multi-posture humans to be correctly classified by a linear SVM in the training process often leads to over-fitting. Some non-linear classification methods such Kernel SVMs are options to handle this problem, but they are generally much more computationally expensive than linear methods [15]. The PL-SVM used the piecewise discriminative function to construct the non-linear classification boundary that can discriminate the multiple positive subclasses from the negative class. PL-SVM is group of several linear SVM.

Figure 2: *An overview of our feature extraction and object detection chain. The detector window is tiled with a grid of overlapping blocks in which Histogram of Oriented Gradient feature vectors are extracted. The combined vectors are fed to a linear SVM for object/non-object classification. The detection window is scanned across the image at all positions and scales, and conventional non-maximum suppression is run on the output pyramid to detect object instances, but this paper concentrates on the feature extraction process.*

2. Overview of Method

There has been substantial effort devoted to automatic facial image analysis over the past decade. Major topics include facial feature tracking, facial expression analysis, and face recognition [16, 17, and 18]. Facial expression refers to both emotion-specified expressions (e.g., happy or sad) and anatomically based facial actions [19]. Comprehensive reviews of automatic facial may be found in [20, 16, 18, 21]. Here we briefly review literature most relevant to the current study. The pioneering work of Black and Yacoob [22] recognizes facial expressions by fitting local parametric motion models to regions of the face and then feeding the resulting parameters to a nearest neighbor classifier for expression recognition. De la Torre et al. [23] use condensation and appearance models to simultaneously track and recognize facial expression. Chang et al. [24] use a low dimensional Leipschitz embedding to build a manifold of shape variation across several people and then use I-condensation to simultaneously track and recognize expressions. Lee and Elgammal [25] use multi-linear models to construct a non-linear manifold that factorizes identity from expression. Littleworth et al. [26] learn an appearance classifier for facial expression recognition. Shape and appearance features are common to most work on this topic. More recently, investigators have proposed use of dynamic features in addition to those of shape and appearance to recognize facial expressions and actions [27, 28, 29]. In a pioneering study, Mase and Pentland [30] found that zero crossings in the velocity contour of facial motion are useful for temporal segmentation of visual speech. Recently, Hoey [16] present a multilevel Bayesian network to learn the dynamics of facial expression. Irani and Zelnik [14] propose a modification of structure-from-motion factorization to temporally segment rigid and non-rigid facial motion. These approaches all assume accurate registration prior to segmentation. Accurate registration of non-rigid facial features, however, is still an open research problem. Navneet Dalal and Bill Triggs algorithm on Histogram of Oriented Gradients (HoG) is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid [31]. The basic idea is that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. In practice this is implemented by dividing the image window into small spatial area called as 'cell', for each cell accumulating a local 1-D histogram of gradient directions or edge orientations over the pixels of the cell. The combined histogram entries form the representation. For better invariance to illumination, shadowing, etc., it is also useful to contrast-normalize the local responses before using them. This can be done by accumulating a measure of local histogram energy over the larger spatial regions called as 'block'. We will refer to the normalized descriptor blocks as Histogram of Oriented Gradient(HOG) descriptors.Especially for 2D image data, factorizing rigid from nonrigid motion is a challenging problem. To solve this problem without recourse to 3D data and modeling, we propose a clustering algorithm that is invariant to specific geometric transformations. This is the first step toward temporal segmentation of facial actions. We then propose an algorithm to group clusters effectively into temporally coherent chunks. We show the benefits of our approach in two novel applications. In one, we detect unusual or rare facial expressions and actions; in the other, we use the method to preprocess video for manual FACS coding. By temporally segmenting facial behavior, we increase the efficiency and reliability of manual FACS annotation.

3. Facial detection and classification

The detection of human body based on HOG includes the following six steps: gamma correction and normalization in RGB space, gradient calculation, statistical analysis of gradients of a cell, normalization of block, generation of vector, and classification based on SVM.

3.1. Gamma and Color Normalization: We use exponential gamma correction function to remove the effect of ambient disturbance gradients of a cell, normalization of block, generation of vector, and classification based on SVM.

3.2. Gradient Computation: For gradient computation, first the gray scale image is filtered to obtain x and y derivatives of pixels using conv2 (image,filter,' same') method with those kernels:

Ix = [-1 0 1]

and

Iy = [-1 0 1] T

After calculating x, y derivatives (**Ix** and **Iy**), the magnitude and orientation of the gradient is also computed:

Magnitude= $|G| = (Ix^2 + Iy^2)^{0.5}$

Θ = arctan (Iy / Ix)

One thing to note is that, at orientation calculation **rad2deg (atan2 (val))** method is used, which returns values between [-180°,180°]. Since unsigned orientations are desired for this implementation, the values which are less than 0° is summed up with 180°.

3.3. Orientation Binning: The next step is the fundamental nonlinearity of the descriptor. Each pixel calculates a weighted vote for an edge orientation histogram channel based on the orientation of the gradient element centered on it, and the votes are accumulated into orientation bins over local spatial regions that we call cells. Cells can be either rectangular or radial (log-polar sectors). The orientation bins are evenly spaced over 0-1800 ("unsigned" gradient) or 0-3600 ("signed" gradient) [2].

To reduce aliasing, votes are interpolated bilinearly between the neighboring bin centers in both orientation and position. The vote is a function of the gradient magnitude at the pixel, either the magnitude itself, its square, its square root, or a clipped form of the magnitude representing soft presence/ absence of an edge at the pixel.

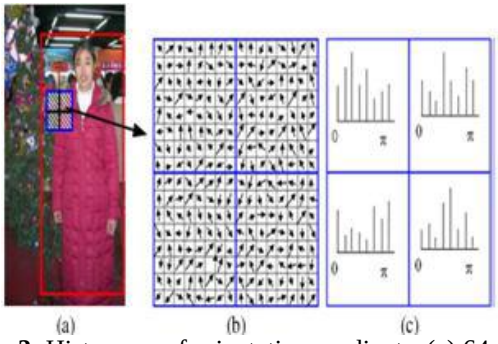

Figure 3: Histogram of orientation gradients. (a) 64 × 128 detection window (the biggest rectangle) in an image. (b) 16 × 16 blocks consists of four cells. (c) Histograms of orientation gradients corresponding to the four cells.

3.4. Descriptor Blocks: In order to account for changes in illumination and contrast, the gradient strengths must be locally normalized, which requires grouping the cells together into larger, spatially connected blocks [2]. The HOG descriptor is then the vector of the components of the normalized cell histograms from all of the block regions. These blocks typically overlap, meaning that each cell contributes more than once to the final descriptor.

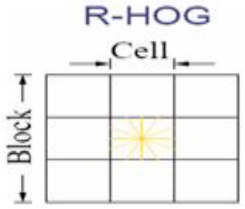

Figure 4: Rectangular HOG

Two main block geometries exist: rectangular R-HOG blocks and circular C-HOG blocks. R-HOG blocks are generally square grids, represented by three parameters: the number of cells per block, the number of pixels per cell, and the number of channels per cell histogram [6]. In this implementation we have used 2x2 cell blocks of 8x8 pixel cells with 9 bin histogram channels.

3.5. Detector Window: Detector window size is 64x128 pixels. Our detection window result in 8 x 16 cells and 7 x 15 R-HOG blocks, since the blocks are overlapping. Each R-HOG block has 2 x 2 cells, which also has 1 x 9 histogram vector each. So the overall size of window is 7 x 15 x 2 x 2 x 9. Therefore, the feature vector size is of 3780. Our detection window includes about 16 pixels of margin around the person on all four sides. This border provides a significant amount of context that helps detection.

3.6 Algorithms for clustering: In this section we review the state of the art in clustering algorithms. In this review, we use a new matrix formulation that enlightens the connections between clustering methods.

3.6.1Connectivity based clustering (hierarchical clustering): These algorithms connect "objects" to form "clusters" based on their distance. A cluster can be described largely by the maximum distance needed to connect parts of the cluster. At different distances, different clusters will form, which can be represented using a dendrogram, which explains where the common name "hierarchical clustering" comes from: these algorithms do not provide a single partitioning of the data set, but instead provide an extensive hierarchy of clusters that merge with each other at certain distances. In a dendrogram, the y-axis marks the distance at which the clusters merge, while the objects are placed along the x-axis such that the clusters don't mix.

3.6.2 Centroid-based clustering: Most k-means-type algorithms require the number of clusters - k - to be specified in advance, which is considered to be one of the biggest drawbacks of these algorithms.K-means has a number of interesting theoretical properties. On the one hand, it partitions the data space into a structure known as a Voronoi diagram. On the other hand, it is conceptually close to nearest neighbor classification, and as such is popular in machine learning. Third, it can be seen as a variation of model based classification, and Lloyd's algorithm as a variation of the Expectation-maximization algorithm

3.6.3Distribution-based clustering: Distribution-based clustering is a semantically strongmethod, as it not only provides you with clusters, but also produces complex models for the clusters that can also capture correlation and dependence of attributes. However, using these algorithms puts an extra burden on the user: to choose appropriate data models to optimize, and for many real data sets, there may be no mathematical model available the algorithm is able to optimize (e.g. assuming Gaussian distributions is a rather strong assumption on the data).

3.6.4 Density-based clustering: On a data set consisting of mixtures of Gaussians, these algorithms are nearly always outperformed by methods such as EM clustering that are able to precisely model this kind of data.Mean-shift is a clustering approach where each object is moved to the densest area in its vicinity, based on kernel density estimation. Eventually, objects converge to local maxima of density. Similar to k-means clustering, these "density attractors" can serve as representatives for the data set, but mean-shift can detect arbitrary-shaped clusters similar to DBSCAN. Due to the expensive iterative procedure and density estimation, mean-shift is usually slower than DBSCAN or k-Means.

3.7 Types of Evaluations

3.7.1 Internal evaluation: When a clustering result is evaluated based on the data that was clustered itself, this is called internal evaluation. These methods usually assign the best score to the algorithm that produces clusters with high similarity within a cluster and low similarity between clusters.

The following methods can be used to assess the quality of clustering algorithms based on internal criterion:

- **Davies–Bouldinindex**:The Davies–Bouldin index can be calculated by the following formula:

$$DB=\frac{1}{n}\sum_{i=1}^{n}\max_{i\neq j}(\sigma_i + \sigma_j/d(c_i,c_j)$$

where n is the number of clusters, c_x is the centroid of cluster x, σ_x is the average distance of all elements in cluster x to centroid c_x , and d (c_i c_j) is the distance between centroids c_i and c_j . Since algorithms that produce clusters with low intra-cluster distances (high intra-cluster similarity) and high inter-cluster distances (low inter-cluster similarity) will have a low Davies–Bouldin index, the clustering algorithm that produces a collection of clusters with the smallest Davies–Bouldin index is considered the best algorithm based on this criterion.

- **Silhouette coefficient**:The silhouette coefficient contrasts the average distance to elements in the same cluster with the average distance to elements in other clusters. Objects with a high silhouette value are considered well clustered; objects with a low value may be outliers. This index works well with k-

means clustering, and is also used to determine the optimal number of clusters.

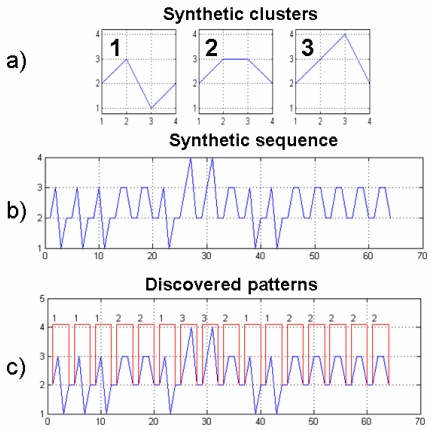

Figure 5. a) 3 synthetic clusters b) Synthetic sequence c) Temporalclusters found by our algorithm.

3.7.2 External evaluation: In external evaluation, clustering results are evaluated based on data that was not used for clustering, such as known class labels and external benchmarks. Such benchmarks consist of a set of pre-classified items, and these sets are often created by human (experts). Thus, the benchmark sets can be thought of as a gold standard for evaluation. These types of evaluation methods measure how close the clustering is to the predetermined benchmark classes. However, it has recently been discussed whether this is adequate for real data, or only on synthetic data sets with a factual ground truth, since classes can contain internal structure, the attributes present may not allow separation of clusters or the classes may contain anomalies. Additionally, from a knowledge discovery point of view, the reproduction of known knowledge may not necessarily be the intended result.

A number of measures are adapted from variants used to evaluate classification tasks. In place of counting the number of times a class was correctly assigned to a single data point (known as true positives), such pair counting metrics assess whether each pair of data points that is truly in the same cluster is predicted to be in the same cluster.

Some of the measures of quality of a cluster algorithm using external criterion include:

- **Rand measure (William M. Rand 1971):**The Rand index computes how similar the clusters (returned by the clustering algorithm) are to the benchmark classifications. One can also view the Rand index as a measure of the percentage of correct decisions made by the algorithm. It can be computed using the following formula:

RI= $\frac{TP+TN}{TP+FP+FN+TN}$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. One issue with the Rand index is that false positives and false negatives are equally weighted. This may be an undesirable characteristic for some clustering applications. The F-measure addresses this concern, as does the chance-corrected adjusted Rand index.

- **Jaccard index**: The Jaccard index is used to quantify the similarity between two datasets. The Jaccard index takes on a value between 0 and 1. An index of 1 means that the two dataset are identical and an index of 0 indicate that the datasets have no common elements. The Jaccard index is defined by the following formula:

J (A, B) =$\frac{|A\cap B|}{|A\cup B}$ = $\frac{TP+TN}{TP+FP+FN}$

This is simply the number of unique elements common to both sets divided by the total number of unique elements in both sets.

The Mutual Information is an information theoretic measure of how much information is shared between a clustering and a ground-truth classification that can detect a non-linear similarity between two clusterings. Adjusted mutual information is the corrected-for-chance variant of this that has a reduced bias for varying cluster numbers.

- **Confusion matrix**: A confusion matrix can be used to quickly visualize the results of a classification (or clustering) algorithm. It shows how different a cluster is from the gold standard cluster.

4. Support vector machine

SVM are based on optimal hyperplane for linearly separable patterns but can be extended to patterns that are not linearly separable by transformations of original data to map into new space. They are explicitly based on a theoretical model of Learning and come with theoretical guarantees about their performance. They also have a modular design that allows one to separately implement and design their components and are not affected by local minima.

Support vectors are the elements of the training set that would change the position of the dividing hyper plane if removed. Support vectors are the critical elements of the training set. The problem of finding the optimal hyperplane is an optimization problem and can be solved by optimization techniques [32,33].

When data is assumed to get separated perfectly, then it can be optimized as

Minimize $\| w \|^2$, subject to

$$(w.x_i + b) \geq 1, if y_i = 1$$

$$(w.x_i + b) \leq -1, if y_i = -1$$

The last two constraints can be compacted to :

$$y_i(w.x_i + b) \geq 1$$

This is a quadratic program. The overall SVM classification procedure is depicted in Figure 2.

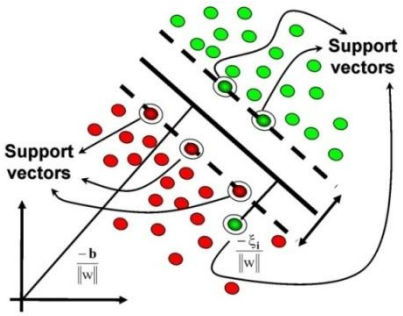

Figure 6: The classification process of SVM

5. Experimental Results

We evaluated the algorithm two ways. One, we tested its ability to temporally segment facial gestures and identify ones that occur rarely. Two, we used it to preprocess video of spontaneous facial expression intended for FACS annotation.

5.1. Temporal segmentation of mouth events

In this experiment, we have recorded a video sequence in which the subject spontaneously made five different facial gestures (sad, sticking out the tongue, speaking, smiling, and neutral). We use person-specific Active Appearance Models [34,35] to track the non-rigid/rigid motion in the sequence (see fig. 7).

Figure 7. AAM tracking across several frames.

By eliminating consecutive frames that have the same cluster label, sequence length is reduced to 20% of the original length (see fig. 8.a and 8.b). Then, the temporal segmentation algorithm discovers the facial gestures shown in 8.c. Observe that there are some time windows that remain unclassified. These windows correspond to gestures lasting only a single frame or ones that are unusual or infrequent.

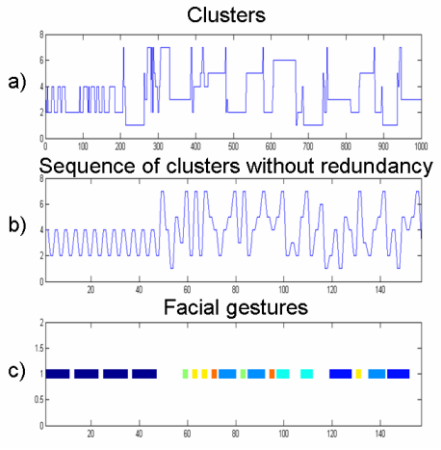

Figure 8. a) Original sequence of clusters. b) Sequence of clusters with just the transitions. c) Discovered facial gestures.

Accuracy of the clustering approach was confirmed by visual inspection. Fig. 9 shows one frame of the output video resulting from finding the temporal clusters in the video sequence. Each frame of the video contains three columns, the first column shows the original image fitted with a person-specific AAM [34] model. The second column represents a prototype of each of the clusters found by the algorithm. The third column shows all facial gestures found in the video. In each frame, the cluster and temporal gesture that corresponds to the image is highlighted.

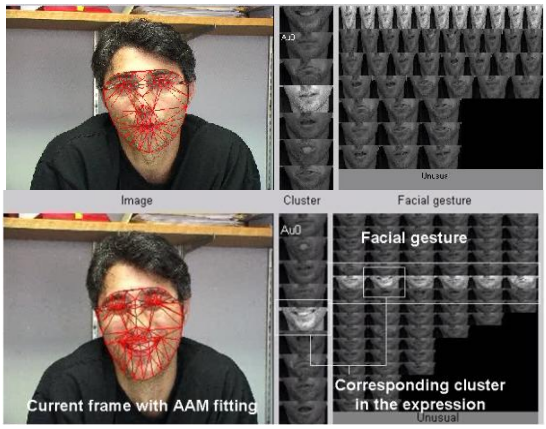

Figure 9. Frame of the output video.

We use subject 25 from the DS180 database [36]. The DS180 is a deception scenario in which 26 young adults must convince an interviewer of their honesty whether or not they are guilty of having taken a sum of money. In the observational scenario, subjects entered a room in which there was or was not a check for a specified amount (typically $100). Subjects were instructed that they could take the check if they wished and then would be interrogated about their actions. The subject's task then was to convince the interrogator that they had not taken the check whether or not they had. We have tracked the facial features of subject 25 with AAMs [34,35].. This manual FACS coding provides ground truth for the analysis. From the shape and appearance data, we compute the affinity matrix K with shape and appearance informationand compute the first 20 eigenvectors. We run 100 iterations of k-means in the embedded space and keep the solution with smallest error. To compute the accuracy of the results for a c cluster case with the ground truth, we compute a c−by−c confusion matrix C, where each entry cij is the number of samples in clusters i, which belong to class j. It is difficult to compute the accuracy by only using the confusion matrix C because we do not know which cluster matches which class. An optimal way to solve for the correspondence [38] is to compute the following maximization problem:

maxtr(CP) | P is a permutation matrix (6)

and the accuracy is obtained by dividing the results for the number of data points to be clustered. To solve eq. 6, we use the classical Hungarian algorithm [38]. Table 1 shows the accuracy results. The clustering approach achieved 70% agreement with manual annotation, which is comparable to the inter-observer agreement of manual coding (80%). It is interesting to notice that the clustering results depend on the shape and appearance parameters, Figure 10 shows the accuracy as a function of these two parameters, and we can observe that it is stable over a large range of values.

Subject	Accuracy	# of Clusters	# of frames
25	70%	21	558

Table 1 Clustering Accuracy

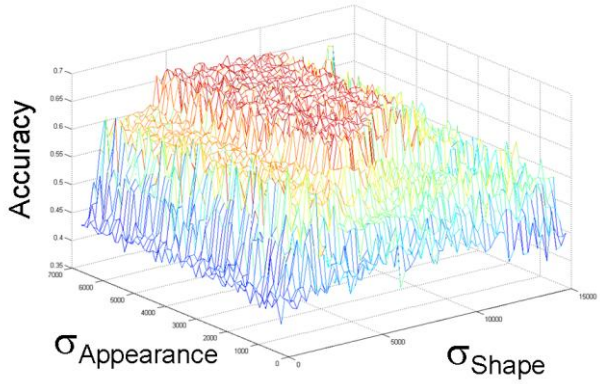

Figure 10. Accuracy variation versus σ_a and σ_s

Database consists of 2416 positive training and 912 negative training images. It consists of 1126 positive and 300 negative testing images. The performance of detector is evaluated for various performance parameters and cross validation accuracy is calculated. The image is split in to the 2×2 pixel cell, 4×4 pixel cell and 8×8 pixel cell. It gives better results for the 8×8 pixel cell because it better describes the orientation of gradients at the edge. The cross validity accuracy is more in 64×128 pixels detection window than other detection window .The comparison of the cross validation accuracy for PL-SVM is shown in the result tables. We have compared the cross validation accuracy for the various SVM kernels and compared the results with PL-SVM. Non-linear SVM kernels perform best when the features are less. Problem of linear SVM is tackled by PL-SVM to get better cross validation accuracy.

Cell size	**Error rate(%)**
2x2 pixel	2.15%
4x4 pixel	1.72%
8x8 pixel	1.22%

Table 2: Cross validation results for various cell sizes

Bin size	Error rate(%)
4 bin(0-180 degree)	5.63%
4 bin(0-180 degree)	3.52%
4 bin(0-180 degree)	1.55%

Table 3:Cross validation results for various orientation bin sizes

SVM Kernel Function	Error rate(%)
Linear -SVM	3.63%
Polynomial -SVM	8.52%
quadratic –SVM	8.35%
MLP-SVM	11.10%
PL-SVM	1.22%

Table 4: Cross validation results for various SVM Kernels

6. Conclusion

In this paper we have presented a method for temporal segmentation of facial behavior and illustrate its usefulness in two novel applications. The method is invariant to geometric transformations, which is critical in real-world settings in which head motion is common. The method clusters similar facial actions, identifies unusual actions, and could be used to increase the reliability and efficiency of manual FACS annotation.

The current implementation is for the mouth region detection system using Histogram of Oriented Gradients features (HoG) and Piecewise Linear Support Vector Machine algorithm. This is the most challenging region in that the degrees of freedom of facial motion are largest in this region. The densest concentration of facial muscles is in the mouth region and the range of motion includes horizontal, lateral, and oblique [38]. Also, because of higher concentration of contra lateral innervations in the lower face, the potential for asymmetric actions is much greater than for the rest of the face [39]. To be useful, a system must include all facial regions. The problem of the multi-view and multi-posture detection can be tackled by PL- SVM. We have proposed a PL-SVM training algorithm that can automatically divide the feature space and train the PL-SVM with the margins of the linear SVMs increased iteratively. Current work expands clustering to include eye, midrace, and brow features.

References

[1] Z. Ambadar, J. Schooler, and J. F. Cohn. Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. Psychological Science,16:403–410, 2005.

[2] J. F. Cohn, Z. Ambadar, and P. Ekman. Observer-based measurement of facial expression with the facial action coding system. J. Coan and J. Allen (Eds). The handbook of emotion elicitation and assessment. Oxford University Press Series in Affective Science. NY: Oxford., 2006.

[3] J. F. Cohn and T. Kanade. Use of automated facial image analysis for measurement of emotion expression. The handbook of emotion elicitation and assessment. Oxford University Press Series in Affective Science., New York: Oxford., 2007.

[4] Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou, "Discriminative Local binary patterns for human detection in personal album," in Proc. IEEE Int Conf. Comput. Vis. Pattern Recognit., Jun. 2008, pp. 1–8.

[5] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in Proc. IEEE Int. Conf. Computer. Vis., Oct. 2009, pp. 32–39.

[6] B. Wu and R. Nevatia, "Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors," in Proc. IEEE Int. Conf. Comput. Vis., vol. 1. Oct. 2005, pp. 90–97.

[7] P. Sabzmeydani and G. Mori, "Detecting pedestrians by learning shapelet features," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2007, pp. 1–8.

[8] S. Munder and D. M. Gavrila, "An experimental study on pedestrian classification," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 11, pp. 1863–1868, Nov. 2006.

[9] Q. Ye, J. Jiao, and B. Zhang, "Fast pedestrian detection with multi-scale orientation features and two-stage classifiers," in Proc. IEEE 17th Int. Conf. Image Process., Sep. 2010, pp. 881– 884.

[10] W. Gao, H. Ai, and S. Lao, "Adaptive contour features in Oriented granular space for human detection and segmentation," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 1786–1793.

[11] Y. Liu, S. Shan, W. Zhang, X. Chen, and W. Gao, "Granularity-tunable gradients partition (GGP) descriptors for human detection," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 1255–1262.

[12] Z. Lin, L. Davis, and D. Doermann, "Hierarchical part-template Matching for pedestrian detection

[13] Sung Tae An, Jeong Jung Kim, and Ju Jang Lee, "SDAT-Simulteneous Detection and Tracking of Humans using Partical Swarm Optimization", IEEE International Conference on Mechatronics and Automation, Aug.2010, pp. 483-488.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE Int. Conference on Computer Vision Pattern Recognition, Jun.2005, pp. 886–893.

[15] Qixiang Ye, Zhenjun Han, Jianbin Jiao, Jianzhuang Liu, "Human Detection in Images via Piecewise Linear Support Vector" IEEE Transaction on Image Processing , VOL. 22, NO. 2, February, 2013.

[16] W. Zhao and R. Chellappa. (Editors). Face Processing: Advanced Modeling and Methods. Elsevier, 2006.

[17] A Martinez. Matching expression variant faces. Vision Research, 43(9):1047–1060, 2003.

[18] S. Li and A. Jain. Handbook of face recognition. New York: Springer., 2005.

[19] P. Ekman and W. Friesen. Facial action coding system: A technique for the measurement of facial movement. Consulting Psychologists Press., 1978.

[20] M. Pantic, N. Sebe, J. F. Cohn, and T. Huang. Affective multimodal human-computer interaction. In ACM International Conference on Multimedia, pages 669–676, 2005.

[21] Y. Tian, J. F. Cohn, and T. Kanade. Facial expression analysis. In S. Z. Li and A. K. Jain (Eds.). Handbook of face recognition. New York, New York: Springer., 2005.

[22] F. De la Torre, Y. Yacoob, and L. Davis. A probabilistic framework for rigid and non-rigid appearance based tracking and recognition. In Int. Conf. on Automatic Face and Gesture Recognition, pages 491–498, 2000.

[23] C. Hu, Y. Chang, R. Feris, and M. Turk. Manifold based analysis of facial expression. In CVPRW'04: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Volume 5, page 81, Washington, DC, USA, 2004. IEEE Computer Society.

[24] C. Lee and A. Elgammal. Facial expression analysis using nonlinear decomposable generative models. In IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pages 17–31, 2005.

[25] G. Littlewort, M. Bartlett, I. Fasel, J. Chenu, and J. Movellan. Analysis of machine learning methods for real-time recognition of facial expressions from video. In Computer Vision and Pattern Recognition, 2004.

[26] J. F. Cohn. Automated analysis of the configuration and timing of facial expression. In P. Ekman and E. Rosenberg, editors, What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS), Oxford University Press Series in Affective Science, pages 388–392. October 2005.

[27] J. F. Cohn and K. Schmidt. The timing of facial motion in posed and spontaneous smiles. International Journal of Wavelets, Multiresolution and Information Processing, 2:1 – 12, March 2004.

[28] M. Pantic and I. Patras. Dynamics of Facial Expression: Recognition of Facial Actions and their Temporal Segments from Face Profile Image Sequences. IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, 36(2):433–449, April 2006.

[29] K. Mase and A. Pentland. Automatic lipreading by computer. (J73-D-II(6)):796–803, 1990.

[30] J. Hoey. Hierarchical unsupervised learning of facial expression categories. In IEEE Workshop on Detection and Recognition of Events in Video, pages 99–106, 2001.

[31] L. Zelnik-Manor and M. Irani. Temporal factorization vs. spatial factorization. In ECCV, pages 434–445, 2004.

[32] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE Int. Conference on Computer Vision Pattern Recognition, Jun.2005, pp. 886–893.

[33] Fen Xu, ming Gao, " Human detection and tracking based on HOG and particle filter", IEEE International Congress on Image and Signal Processing, CISP-2010, pp. 1503-1507

[34] S. Q. Ren, D. Yang, X. Li, and Z. W. Zhuang, "Piecewise support vector machines," Chin. J. Comput., vol. 32, no. 1, pp. 77– 85, 2009.

[35] H. B. Cheng, P.-N. Tan, and R. Jin, "Efficient algorithm for localized support vector machine," IEEE Trans. Knowl. Data Eng., vol. 22, no. 4, pp. 537–549, Apr. 2010.

[36] M. Frank and P. Ekman. The ability to detect deceit generalizes across different types of high-stakes lies. Journal of Personality and Social Psychology, 72(6):1429–1439., 1997.

[37] D. E. Knuth. The Standford GraphBase. Addison-Wesley Publishing Company, 1993.

[38] P. Ekman andW. Friesen. Facial action coding system (facs): Manual. In Consulting Psychologists Press, Palo Alto, CA, USA, 1978.

[39] W. Rinn. The neuropsychology of facial expression. Psychological Bulletin, 95:52–77, 1984.

Author Profile

Preeti SaraswatB.Tech,withfirst-class honors,in Information TechnologyfromRajasthan Technical University and pursuingM.tech (CSE) fromGeetanjali Institute of Technical Studies, Udaipurunder Rajasthan Technical University (kota) as well. She has published someNational conferencepapers.Her area of interestisinusingcomputer visionandmachinelearningtoenablemachines to understand emotions more efficiently like humans.

G.Srikanth B.Tech JITS from J.N.T University Hyderabad, of Electrical and Electronics department and pursuing M.Tech Power Systems and automation from GITAM University-Hyderabad He has published some International and National conferencepapers.His area of interest is on Power Quality Improvements, MicroGrids and Computer Vision.