Spatiotemporal-Boosted DCT Features for Head and Face Gesture Analysis

Hatice Çınar Akakın[1] and Bülent Sankur[1]

Bogazici University, Electrical & Electronics Engineering Department, Bebek, Istanbul

Abstract. Automatic analysis of head gestures and facial expressions is a challenging research area and it has significant applications in human-computer interfaces. In this study, facial landmark points are detected and tracked over successive video frames using a robust method based on subspace regularization, Kalman prediction and refinement. The trajectories (time series) of facial landmark positions during the course of the head gesture or facial expression are organized in a spatiotemporal matrix and discriminative features are extracted from the trajectory matrix. Alternatively, appearance based features are extracted from DCT coefficients of several face patches. Finally Adaboost algorithm is performed to learn a set of discriminating spatiotemporal DCT features for face and head gesture (FHG) classification. We report the classification results obtained by using the Support Vector Machines (SVM) on the outputs of the features learned by Adaboost. We achieve 94.04% subject independent classification performance over seven FHG. [1] [2]

1 Introduction

Human face is a rich source of nonverbal information. Indeed, not only it is the source of identity information but it also provides clues to understand social feelings and can be instrumental in revealing mental states via social signals. Facial expressions form a significant part of human social interaction [1, 2]. While communicating, we express ideas that are visualized in our minds by using words integrated with nonverbal behaviors. Therefore when the body language and verbal messages are used in complementary roles, our messages can be more clear and can be conveyed more accurately. Face then functions as a channel in communicating the emotional content of our messages. Gestures, eye and head movements, body movements, facial expressions and touch constitute the nonverbal message types of our body language. Therefore, empowering computers with the capability to recognize and to respond to nonverbal communication clues is important [3–6].

In this study, we consider two data representation types, namely facial landmark trajectories and intensity image patches on expressive regions of the face

[1] This work was supported by TUBITAK project with number 107E001 and by Bogazici University project with number 09HA202D.

[2] Thanks to Arman Savran for providing his Adaboost code.

extracted throughout the video sequences. DCT features are extracted from those face representations for the automatic analysis of facial expressions and head gestures. Adaboost algorithm is exploited in order to reduce the dimensionality of the total features and to obtain more discriminative DCT coefficients for the classification.

The paper is organized as follows. In the next section we briefly review related works. Section 3 describes the data representation types and extracted features. Classification method is explained in Sect. 4. Section 5 gives details of the used dataset and presents implemented classifiers and the experimental results. Finally, conclusions are drawn in Sect. 6.

2 Related Work

Most of the work in the literature on facial expression analysis is focused on the six basic emotions, i.e., happiness, surprise, sadness, fear, anger and disgust [7–10, 5, 11, 12]. The majority of facial expression recognition systems attempt to identify Facial Action units (FAUs) [13, 7, 8, 14–16, 12] based on Facial Action Coding System (FACS) [17]. In FACS, the facial behavior is decomposed into 46 action units (AUs), each of which is anatomically related to the individual facial muscles. Although they only define a small number of distinctive AUs, different combinations of AUs can be sufficient for accurately detecting and measuring a large number of facial expressions.

Head displays, sometimes called as emblems [14, 18] fulfill a semantic function and provide conversational feedback. Examples of emblems are head nodding (head up and down) and head shaking (head swinging left and right) with or without accompanying facial expressions. In social interactions head and facial displays may convey a message, provide conversational feedback, and form a communicative tool [2, 1]. For example, head nod is an affirmative cue, frequently used throughout the world to indicate understanding, approval and agreement [2, 1, 19–21]. On the other hand, head shake is almost a universal sign of disapproval, disbelief, and negation [2, 1, 19–21]. Prediction of frustration and human fatigue detection problems were analyzed by integrating information from various sensory information [22–24].

Bartlett et al. [25] used Gabor filters for appearance based feature extraction from the still images. They obtained their best recognition results by selecting a subset of Gabor filters using AdaBoost and then training Support Vector Machines on the outputs of the filters selected by AdaBoost. Shan [11] studied facial representation based on LBP features for facial expression recognition. They examined different machine learning methods, including template matching, SVM, LDA, and the linear programming technique on LBP features. They obtained their best results with Boosted-LBP by learning the most discriminative LBP features with AdaBoost, and the recognition performance of different classifiers were improved by using the Boosted-LBP features.

There are relatively few papers in the literature addressing the FHG detection issue. In Kang et al. [20], location of eyes is detected and tracked in video sequence, and the resulting trajectory is used to recognize head shake and head nod gestures using HMMs. Somewhat similarly, Kapoor and Picard [19] used an active camera with infrared LEDs to track pupils. The position of pupils are used as observations by a discrete HMM pattern analyzer to detect head nods/shakes. Morency et al. [21] investigated how dialog context from an embodied conversational agent can improve visual recognition of user gestures such as head nod and head shakes. For recognizing these gestures, they tracked head position and rotation, then computed head velocity vector and used SVM classifiers. In Aran's study [26] a multi-class classification strategy for Fisher scores was proposed and tested on a hand gesture dataset and a sign language expression dataset [27].

3 Data representations and spatiotemporal features for FHG

Once the facial landmark points are tracked in each FHG video frame, two different data types are extracted: *i*) Landmark trajectories; *ii*) Intensity face image patches. Even though, head and facial gestures may differ in total duration they mostly follow a fixed pattern of temporal order. Therefore, in order to process extracted data from face videos, we used both spatial normalization and temporal normalization. The details of the feature extraction process is given in the following subsections.

3.1 Facial landmark trajectories

A number (l) of facial landmark points are detected and tracked over successive video frames using an automatic landmark detection and algorithm [28] (Figure 1). The algorithm detects facial landmarks in the initial frame using DCT-features trained with SVM classifiers, and then applies a multi-step tracking method based on adaptive templates, Kalman predictor and subspace regularization for the subsequent frames.

Once the landmark coordinates are detected over the successive frames, the landmark coordinate data of the face video is reduced to a FxT matrix P. Here, each row of the P matrix represents the time sequence of the x or y coordinates of one of the 17 landmarks. In order to obtain landmarks independent from the initial position of the head the first column is subtracted from all columns of P, so that we only consider the relative landmark displacements with respect to the first frame. This presupposes that the landmark estimates in the first frame of the sequence are reliable.

In our work we used 17 ($l = 17$) landmarks as illustrated in Figure 1 which resulted in $F = 34$ coordinates.

3.2 Feature extraction from face image patches

Deformations occurring on the face during an expression involves changes over whole regions, such as mouth and eye regions. The estimated landmarks enable

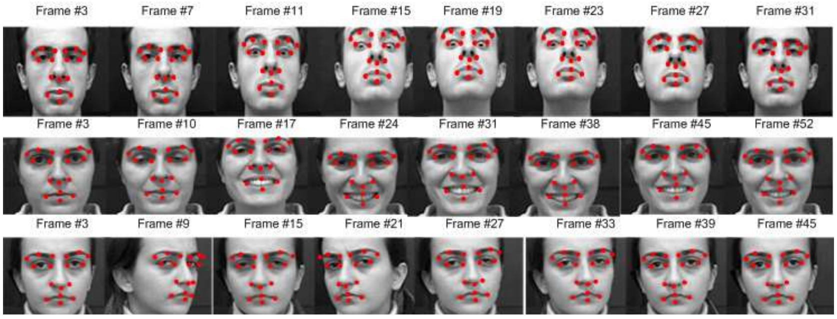

Fig. 1: Illustration of the 17 tracked facial landmarks on sample image sequences

us to parse the face into regions of interest. We have heuristically selected four patches covering the most expressive facial parts, as shown in Figure 3. Patch sizes are chosen large enough to cover a whole expressive face region. Furthermore patches are positioned using the tracked facial landmark locations. In that respect, sizes and semantic positions of the patches do not vary with changes in head orientation.

Extracted patches are scaled into fixed block size as in Table 1. The discriminative features from patches consist of DCT coefficients, not from the whole patch but from the 16x16 non-overlapping blocks tessellating the patch. Since the expressive eye region is critical, it is doubly covered. Beside the eye and eyebrow patches, one larger patch that jointly covers them and that overlaps with the other four (16x16 DCT block) is used in order to better interpret the appearance changes between the eyebrows. We selected the first 20 DCT coefficients (after skipping the DC value) from the zigzag order. All DCT block patterns are then concatenated into a single vector to form the feature vector 20 x (total block number = 15). Since the patch-based, 300-coefficient long appearance feature is extracted from each of the T frames, the gesture video thus generates 300xT dimensioned feature matrix S. As can be seen from Figure 2, the rows of the S matrix represents the temporal changes of the selected DCT coefficients and the columns represents the selected 300 DCT coefficients (spatial features extracted at time k).

$$S = \begin{bmatrix} DCT_{1,1} & \cdots & DCT_{1,60} \\ \vdots & \ddots & \vdots \\ DCT_{(kx15),1} & \cdots & DCT_{(kx15),60} \end{bmatrix} \tag{1}$$

$$P = \begin{bmatrix} P_{1,1} & \cdots & P_{1,60} \\ \vdots & \ddots & \vdots \\ P_{34,1} & \cdots & P_{34,60} \end{bmatrix} \tag{2}$$

$$\begin{array}{rcl} & \xrightarrow{\text{Temporal Domain}} & \\ S = & \begin{bmatrix} DCT_1^1 & \cdots & DCT_T^1 \\ \vdots & \ddots & \vdots \\ DCT_1^{300} & \cdots & DCT_T^{300} \end{bmatrix} & \Big\downarrow \text{Spatial Domain} \end{array}$$

Fig. 2: Representation of S matrix which is composed of DCT features of image patches

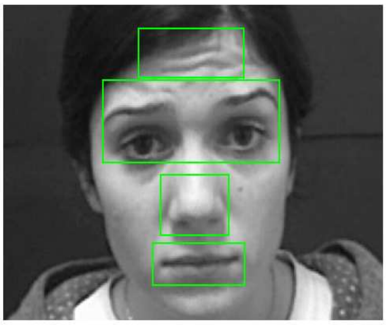

Fig. 3: Facial patches defined on a sample image

$$P = \begin{bmatrix} P_{1,1} & \cdots & P_{1,T} \\ \vdots & \ddots & \vdots \\ P_{34,1} & \cdots & P_{34,T} \end{bmatrix} \tag{3}$$

Since duration of FHG videos is variable depending on the gesture type and upon the actor, we normalized length of the landmark trajectories and appearance features by using the “resample” function of the Matlab so that all gestures spatiotemporal consisted had length T. Note that “resample” function basically changes the sampling rate of a given sequence to a desired one using a polyphase implementation. The resulting spatiotemporal trajectory matrix P has rows corresponding to landmark coordinates and columns corresponding to normalized time index; similarly appearance feature matrix S has rows corresponding to spatial features (300 DCT coefficients) and columns corresponding to normalized time index. In our study we chose T as 60 which is also the average length of the gesture sequences with 11 frames standard deviation.

Table 1: Facial patch dimensions and sub-block sizes on original intensity frames

Patch Label	Region on the Face	Patch Size	number of 16x16 blocks
1	eyes and eyebrows	16x64	5
2	nose	16x32	2
3	mouth	16x64	4
4	forehead	16x64	4

In order to decrease the dimensionality of P and S matrices, in our case 34x60=2040 for landmark trajectory data and 300x60=18000 for appearance data, respectively, we apply DCT to each row of the data matrix P and S to extract the temporal information of the data matrices. Here rows of the P and S matrix correspond to normalized time domain of the data. We select the first 20 DCT coefficients, by excluding the DC term, and normalize the resulting DCT feature vector to unit norm. DCT is chosen because it is known to have good energy compaction property for highly correlated data and can serve the purpose of summarizing and capturing the data content. Hence we get 680x1 (34x20) dimensioned DCT coefficients for trajectory matrix P and 6000x1 (300x20) dimensioned DCT coefficients for appearance features.

4 Classification of FHGs

A set of discriminative and effective features should be selected from the DCT coefficients to construct the FHG classifier. It is known that the motion of certain landmarks are more expressive and hence contain more discriminative information, and this selection depends on the face and head gesture types. Therefore it would pay to pinpoint these more discriminative and effective features per gesture. The Adaboost [29] algorithm seems to be the right tool for optimal feature selection [30, 11, 31] from the high dimension data. In this paper, we use the Adaboost [29] learning to select 7 to 30 percent of the features from the initial set extracted features.

4.1 Boosting the DCT features

A sequence of weighted weak classifiers are boosted to form a final strong classifier. A weak classifier is designed by selecting a single feature performance and by setting optimally a threshold such that the best classification performance is achieved. In this study, weak classifiers are chosen as nearest mean classifiers.

4.2 Classification with SVM

We combine Adaboost selected feature with SVM [32] classification. Note that, we run the Adaboost-based feature selection separately for each experiment

Table 2: Characteristics of the BUHMAP FHG videos

Head and Facial Gesture Classes in BUHMAP DB [27]	
Head shaking (**G1**):	Rotating head left and right sides repetitively
Head up (**G2**):	Tilting the head back while at the same time raising the eyebrows
Head forward (**G3**):	Moving head forward and raising eyebrows
Sadness (**G4**):	Lips turned down, eyebrows down
Head up-down (**G5**):	Nodding head repetitively
Happiness (**G6**):	Smile and expression of joy
Happy up-down (**G7**):	Nodding with smile

Table 3: Test set and experiment setup

Test	Subjects (S)	Class (C)	Repetitions (R)
	11	7 (**G1**, **G2**, **G3**, **G4**, **G5**, **G6**, **G7**)	5
Experiment	**Training**	**Testing**	**Method**
11 fold	10 S, 5 R, 350 videos	1 S, 5 R, 35 videos	Leave-one-S-out cross validation

setup and then SVM is trained for the two-class classification problem. Therefore we formulate C (number of classes) two-class problems, and in each one we separate one of the classes from the ensemble of all other classes. This result in C different SVMs. When a test feature vector arrives, we calculate the output of each SVM classifier for this test data, where the C outputs give the class likelihoods. Then the classifier with maximum probability is declared as the gesture class of the test data. In order to find the parameter setting, we carried out grid-search on the hyper-parameters in the 11-fold cross-validation and selected the parameters with maximum recognition accuracy. Note that, radial basis function is used as an SVM kernel in the implementation of SVM classifier.

5 Experimental results

5.1 Video database (BUHMAP)

We tested our FHG recognition algorithm on the BUHMAP video data-base [27] (http://www.cmpe.boun.edu.tr/pilab/pilabfiles/databases/buhmap/). BUHMAP includes seven non-manual gesture classes (but not including neutral state) selected from Turkish Sign Language (TSL). The details of the gesture classes are given in Table 2. Our test set includes seven gesture types acted by eleven subjects, with five repetitions each, hence overall 385 video shots.

The videos are recorded at 30 fps at the resolution of 640x480. Each video starts and ends in the neutral state of the face.

As presented in Table 3 an 11-fold cross-validation scheme is carried out for training and testing any one feature set and classifier combination. For each fold,

Table 4: Proposed classifiers FHG classification

Classifier	Data / Size	Feature selection / Size/Classification
$P^{200}_{DCT\_ADA}$:	DCT of Trajectory matrix P / 680x1	Adaboost/200/SVM
P^{680}_{DCT} :	DCT of Trajectory matrix P / 680x1	$-$/680/SVM
$S^{400}_{DCT\_ADA}$	DCT of intensity image patches S / 6000	Adaboost/400/SVM
S^{6000}_{DCT}	DCT of intensity image patches S / 6000	$-$/6000/SVM
$(S+P)^{600}_{DCT\_ADA}$	DCT of intensity image patches S + DCT of Trajectory matrix P/ 6680	Adaboost/600/SVM
$(S+P)^{6680}_{DCT}$	DCT of intensity image patches S + DCT of Trajectory matrix P/ 6680	$-$/6680/SVM

one subject's gesture samples (7x5=35 gesture samples) are left out as test set and the 350 gesture samples of the remaining subjects are used for training. Thus for each fold, each gesture class has 5 test samples and 50 positive training samples. Notice that, recognition results reported in this study are computed as the average of 11-fold testing.

5.2 Results

The classifiers chosen for the face and head gesture classification problem are given in Table 4. The classification performance of DCT features and boosted-DCT features are given in order to compare the performances of these two feature extraction methods. Table 5 represents the recognition results of the each individual classifier over seven gesture classes. We give the following clarifications for the Tables 4 and 5: Set P denotes the landmark trajectory features and S denotes the sequence of image patch features. Furthermore the superscript indicates the number of features used and the subscript indicates the selection method. Thus for example:

- $P^{200}_{DCT\_ADA}$: 200 DCT coefficients out of 680x1 available DCT coefficients of trajectory matrix P have been selected via Adaboost.
- $(S+P)^{600}_{DCT\_ADA}$: DCT features from the trajectory matrix and image patch sequence have been pooled, and then 600 of them have been selected via Adaboost.

The results show that:

(i) Feature selection by Adaboost algorithm improves the classification performance about 3 to 4 percentage points.

(ii) 92.22 % best individual classification performance is obtained with boosted-DCT features extracted from face intensity image patches ($S^{400}_{DCT\_ADA}$).

(iii) Feature-based fusion of boosted-DCT features of trajectory matrix and boosted-DCT features of intensity image patches ($(S+P)^{600}_{DCT\_ADA}$) surpass the classification performance of boosted-DCT features of intensity image patches.

Table 5: Proposed classifiers FHG classification (C1:)

Classifier	**G1**	**G2**	**G3**	**G4**	**G5**	**G6**	**G7**	**Total**
$P^{200}_{DCT\_ADA}$	96.4	100	89.1	70.9	80	89.1	78.2	86.2
P^{680}_{DCT}	98.2	98.2	89.1	67.3	70.9	87.3	72.7	83.4
$S^{400}_{DCT\_ADA}$	100	92.7	87.3	89.1	100	81.8	94.6	92.2
S^{6000}_{DCT}	96.4	94.6	87.3	90.9	90.9	76.4	85.5	88.8
$(S+P)^{600}_{DCT\_ADA}$	96.4	98.2	89.1	90.9	90.9	80	89.1	93.77
$(S+P)^{6680}_{DCT}$	100	100	87.3	98.2	98.2	83.6	89.1	90.65
$P^{200}_{DCT\_ADA} + S^{400}_{DCT\_ADA}$	100	100	92.7	92.7	92.7	89.1	90.9	94.03

Table 6: Decision fusion of $P^{200}_{DCT\_ADA}$ and $S^{400}_{DCT\_ADA}$

	G1	**G2**	**G3**	**G4**	**G5**	**G6**	**G7**
G1	**100**	0	0	0	0	0	0
G2	0	**100**	0	0	0	0	0
G3	0	7.3	**92.7**	0	0	0	0
G4	0	0	1.8	**92.7**	1.8	1.8	1.8
G5	0	0	1.8	5.4	**92.7**	0	0
G6	0	0	0	0	0	**89.1**	10.9
G7	0	0	1.8182	0	1.8	5.4	**90.9**

(iv) Best overall classification performance (94.03 % Table 6) is achieved by decision fusion of boosted-DCT features of trajectory matrix ($P^{200}_{DCT\_ADA}$) and and boosted-DCT features of face intensity image patches ($S^{400}_{DCT\_ADA}$).

Note that, decision combination is implemented by summing the scores of the classifiers.

6 Conclusion

In this study we have analyzed spatiotemporal feature extraction methods based on accurate tracking of facial landmarks on facial expressions and head gesture sequences. Two types of data representations are investigated, namely facial landmark trajectories and patches of face intensity images. Both modalities have been subjected to DCT transformation for feature extraction. Selection of DCT features is implemented both heuristically using only the low-pass coefficients and algorithmically, using the Adaboost algorithm. The first conclusion is that the proposed classifiers perform satisfactory FHG identification as they achieve scores well above 90 %. In fact, our method surpass significantly the average classification performances reported recently, i.e., 77 % in [26] and 86.4% [28] using the subset of BUHMAP dataset (210 videos of four subjects).

An interesting observation is that sequence and subspace classifiers have very similar performances. While sequence classifiers (e.g. HMM) are designed to compensate for time variations between sequences, the fact that, subspace classifiers with spatiotemporal features have on a par performance can be attributed to the mitigation of this variability by linear time normalization. The best classification result of an individual classifier (without any decision fusion) is achieved for a database with seven gesture classes is 92.2% is state-of-the-art performance. Work is progressing on FHG classification involving a larger set of gestures and mental states.

References

1. Knapp, M.L., Hall, J.A.: Nonverbal communication in human interaction. 6th edn. Belmont, CA : Wadsworth/Thomson Learning (2006)
2. A.Mehrabian, S.R.Ferris: Inference of attitude from nonverbal communication in two channels. Journal of Counseling Psychology **31**(3) (June 1967) 248–252
3. Vinciarelli, A., Pantic, M., Bourlard, H.: Social signal processing: Survey of an emerging domain. Image and Vision Computing **27**(12) (2009) 1743 – 1759
4. Gatica-Perez, D.: Automatic nonverbal analysis of social interaction in small groups: A review. Image and Vision Computing **27**(12) (2009) 1775 – 1787
5. Sebe, N., Lew, M., Sun, Y., Cohen, I., Gevers, T., Huang, T.: Authentic facial expression analysis. Image and Vision Computing **25**(12) (December 2007) 1856–1863
6. Zeng, Z., Pantic, M., Roisman, G., Huang, T.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(1) (January 2009) 39–58
7. Bailenson, J.N., Pontikakis, E.D., Mauss, I.B., Gross, J.J., Jabon, M.E., Hutcherson, C.A.C., Nass, C., John, O.: Real-time classification of evoked emotions using facial feature tracking and physiological responses. Int. J. Hum.-Comput. Stud. **66**(5) (2008) 303–317
8. Zhang, Y., Ji, Q.: Active and dynamic information fusion for facial expression understanding from image sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(5) (May 2005) 699–714
9. Hupont, I., Cerezo, E., Baldassarri, S.: Facial emotional classifier for natural interaction. **7**(4) (2008) 1–12
10. Dornaika, F., Davoine, F.: Simultaneous facial action tracking and expression recognition in the presence of head motion. International Journal of Computer Vision **76**(3) (March 2008) 257–281
11. Shan, C., Gong, S., McOwan, P.: Facial expression recognition based on local binary patterns a comprehensive study. Image and Vision Computing **27**(6) (May 2009) 803–816
12. Tsalakanidou, F., Malassiotis, S.: Real-time 2d+3d facial action and expression recognition. Pattern Recognition **43**(5) (2010) 1763 – 1775
13. Wang, T., James Lien, J.J.: Facial expression recognition system based on rigid and non-rigid motion separation and 3d pose estimation. Pattern Recognition **42** (2009) 962–977
14. Kaliouby, R.A.: Mind-reading machines: automated inference of complex mental states. Technical report, UCAM-CL-TR-636 (2005)
15. Tong, Y., Wang, Y., Zhu, Z., Ji, Q.: Robust facial feature tracking under varying face pose. Pattern Recognition **40** (2007) 3195–3208
16. Pantic, M., Rothkrantz, L.: Facial action recognition for facial expression analysis from static face images. IEEE Transactions on Systems, Man, and CyberneticsPart B: Cybernetics **34**(3) (June. 2004) 1449–1461

17. Ekman, P., Friesen, W.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)
18. Kanaujia, A., Huang, Y., Metaxas, D.: Emblem detections by tracking facial features. In: Conference on Computer Vision and Pattern Recognition Workshop. (2006) 108
19. Kapoor, A., Picard, R.W.: A real-time head nod and shake detector. In: Proceedings from the Workshop on Perspective User Interfaces. (2001)
20. Kang, Y.G., Joo, H.J., Rhee, P.K.: Real time head nod and shake detection using hmms. In: Knowledge-Based Intelligent Information and Engineering Systems. Volume 4253., Springer Berlin / Heidelberg (2006) 707–714
21. Morency, L.P., Sidner, C., Lee, C., Lee, C., Darrell, T.: Contextual recognition of head gestures. In: Proc. of the 7th Int. Conf. on Multimodal Interfaces. (2005) 18–24
22. Kapoor, A., Burleson, W., Picard, R.W.: Automatic prediction of frustration. International Journal of Human-Computer Studies **65**(8) (2007) 724 – 736
23. Ji, Q., Lan, P., Looney, C.: A probabilistic framework for modeling and real-time monitoring human fatigue. IEEE Transactions on Systems, Man, and Cybernetics, Part A **36**(5) (2006) 862–875
24. Yang, J.H., Mao, Z.H., Tijerina, L., Pilutti, T., Coughlin, J.F., Feron, E.: Detection of driver fatigue caused by sleep deprivation. Trans. Sys. Man Cyber. Part A **39**(4) (2009) 694–705
25. Bartlett, M.S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J.: Recognizing facial expression: Machine learning and application to spontaneous behavior. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 2., Los Alamitos, CA, USA, IEEE Computer Society (2005) 568–573
26. Aran, O., Akarun, L.: A multi-class classification strategy for fisher scores: Application to signer independent sign language recognition. Pattern Recognition **43**(5) (2010) 1776 – 1788
27. Aran, O., Ari, I., Guvensan, M.A., Haberdar, H., Kurt, Z., Turkmen, H.I., Uyar, A., Akarun, L.: A database of non-manual signs in turkish sign language. In: IEEE 15th Signal Processing and Communications Applications Conference (SIU '07). (June 2007)
28. Akakin, H.C., Sankur, B.: Analysis of head and facial gestures using facial landmark trajectories. In: COST 2101/2102 Conference. (2009) 105–113
29. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences **55**(1) (1997) 119 – 139
30. Littlewort, G., Bartlett, M.S., Fasel, I., Susskind, J., Movellan, J.: Dynamics of facial expression extracted automatically from video. In: Journal of Image and Vision Computing. (2004) 615–625
31. Yang, P., Liu, Q., Metaxas, D.N.: Boosting encoded dynamic features for facial expression recognition. Pattern Recognition Letters **30**(2) (2009) 132 – 139 Video-based Object and Event Analysis.
32. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. (2001) Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.