Ensemble of Hankel Matrices for Face Emotion Recognition

Liliana Lo Presti(✉) and Marco La Cascia

DICGIM, Universitá Degli Studi di Palermo,
V.le delle Scienze, Ed. 6, 90128 Palermo, Italy
liliana.lopresti@unipa.it

Abstract. In this paper, a face emotion is considered as the result of the composition of multiple concurrent signals, each corresponding to the movements of a specific facial muscle. These concurrent signals are represented by means of a set of multi-scale appearance features that might be correlated with one or more concurrent signals. The extraction of these appearance features from a sequence of face images yields to a set of time series. This paper proposes to use the dynamics regulating each appearance feature time series to recognize among different face emotions. To this purpose, an ensemble of Hankel matrices corresponding to the extracted time series is used for emotion classification within a framework that combines nearest neighbor and a majority vote schema. Experimental results on a public available dataset show that the adopted representation is promising and yields state-of-the-art accuracy in emotion classification.

Keywords: Emotion · Face processing · LTI systems · Hankel matrix · Classification

1 Introduction

Emotion recognition deals with the problem of inferring the emotion (i.e. fear, anger, surprise, etc.) given a sequence of face images. Due to strong inter-subject variations, especially in some kind of emotions (such as fear or sadness), and the difficulty to extract reliable feature representations because of illumination changes, biometric differences, and head pose changes, emotion recognition is a challenging problem. Nonetheless, recognition of face expressions and emotions is of great interest in many fields such as assistive technologies [21], [10], socially assistive robotics [23], computational behavioral science [25], [18], [35], and the emerging field of audience measurement [11].

A vast literature on affective computing [35], [27], [21], has shown that an emotion can be identified by a subset of detected action units. This suggests that face emotion results as combination of movements of various facial muscles. Therefore in this paper we assume that a composition of multiple concurrent signals yields to a face emotion. We use a restricted set of appearance features

V. Murino and E. Puppo (Eds.): ICIAP 2015, Part II, LNCS 9280, pp. 586–597, 2015.
DOI: 10.1007/978-3-319-23234-8_54

– computed on a frame-per-frame basis – that may be correlated with one or more of these concurrent signals. Given a sequence of face images corresponding to an emotion, the extraction of these appearance features yields to a set of time series, one for each appearance feature. Considering that face emotions are not instantaneous, we aim at using the dynamics regulating each sequence of appearance features to recognize among different emotions.

We propose to model a sequence of face appearance feature as the output of a Linear Time Invariant (LTI) system. Motivated by the success of works in action recognition [12], [17], that represent action-dynamics in terms of Hankel matrices, in this paper we explore the use of Hankel matrices to represent emotion-dynamics. We adopt a multi-scale Haar-like feature based appearance representation to obtain a set of time series (one for each spatial scale and Haar-like feature). Hence we represent a sequence of face images by means of an ensemble of Hankel matrices where each Hankel matrix embeds the dynamics of one of the extracted Haar-like feature time series. Nearest-Neighbor classifier combined with a majority vote schema is used for classification purposes.

We validated our work on the publicly available extended Cohn-Kanade dataset [20]. Our experiments show that there is a clear advantage in adopting a dynamics-based emotion representation over using the raw measurements. Furthermore, our experiments highlight that the dynamics of different appearance features contribute differently to the emotion recognition. Overall, our novel emotion representation permits to achieve state-of-the-art accuracy values in comparison to works that use accurate face landmarks.

The plan of the work is as follows. In Section 2, we present works that are related to our emotion-dynamics representation. In Section 3 we describe how we extract a multi-scale face appearance description; Section 4 introduces the Hankel matrix-based representation and describes how to build an ensemble of Hankel matrices to describe face appearance dynamics; Section 5 presents details about the adopted classification framework. Finally, in Sections 6 and 7, we present experimental results, and conclusions and future directions respectively.

2 Related Work

Face detection [31], face recognition [37], [14] and facial expression analysis [6] have been deeply studied in past years, resulting in a vast literature reviewed in [35], [27]. In this section, we focus on works that embed the temporal structure of the face image sequence in the feature representation or in the emotion model.

Dynamics-based emotion recognition has been proposed in [5] where horizontal and vertical movements of tracked landmarks of different face parts such as eyebrows, eyelids, cheeks, and lip corners jointly with spatio-temporal appearance features are used to describe a sequence of face images. Temporal changes in the face appearance are described by means of the Complete Local Binary Patterns from Three Orthogonal Planes (LBP-TOP) [36] and classification is performed by SVM. While [5] attempts to embed information about the dynamics at a feature representation level, works such as [19], [28] account for the temporal structure of the sequences of descriptors in the emotion model. In [22], restricted Boltzmann machine with local interactions (LRBM) is used to capture the spatio-temporal patterns in the data. RBM is used as a generative model for data representation instead of feature learning, and data need to be pre-aligned. In [19] time-series kernel methods are used for emotional expression estimation using landmark data only. The work shows that emotion recognition may be done by adopting either the Dynamic Time Warping (DTW) kernel or the Global Alignment (GA) kernel [3], [4]. In [28], a Bayesian approach is used to model dynamic facial expression temporal transitions. Facial appearance representation is computed in terms of Local Binary Patterns (LBP), and an expression manifold is derived for multiple subjects. A Bayesian temporal model (similar to HMM with a non parametric observation model) of the manifold is used to represent facial expression dynamics.

Works such as [9], [32] use landmarks located on face parts such as eyes, eyebrows, nose and mouth to describe an emotion. In [9], a Constrained Local Model (CLM) is used to estimate facial landmarks and extract a sparse representation of corresponding image patches. Emotion classification is performed by least-square SVM. Wang et al. [32] propose to use Interval Temporal Bayesian Network (ITBN) to capture the spatial and temporal relations among the primitive facial events.

Hankel matrices have been already adopted for action recognition in [12], which adopts a Hankel matrix-based bag-of-words approach, and in [17], which models an action as a sequence of Hankel matrices and uses a set of HMM trained in a discriminative way to model the switching between LTI systems. In [16], we have showed how the dynamics of tracked facial landmarks can be modeled by means of Hankel matrices and can be used for facial expression analysis.

Whilst it is possible to obtain a reasonably accurate estimate of the face region [31], getting a reliable estimation of facial landmarks is still an open problem despite the remarkable progress described in [2], [38]. The adoption of appearance feature extracted from the detected face region to describe an emotion, as done indeed in [24], [28], [35], [27], might be a convenient choice. Therefore, in this paper we adopt appearance features to represent a face expression. In contrast to [16], we do not model landmark trajectories but we use an ensemble of Hankel matrices to describe the dynamics of sequences of appearance features computed at multiple spatial scales. We demonstrate that, without an accurate estimation of facial landmarks, our novel representation can achieve state-of-the-art accuracy in emotion recognition.

3 Multi-Scale Face Appearance Representation

Given a face image, we need to extract a proper appearance representation for the shown face expression. Considering the success of Haar-like features in face detection we adopt this kind of features to build our face appearance descriptor.

Haar-like features resemble Haar wavelets and have been developed by Viola and Jones for face detection [31]. A Haar-like feature is computed by considering adjacent rectangular regions in a detection window. The pixel intensities in each

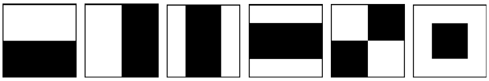

Fig. 1. The set of six Haar-like features used in this paper.

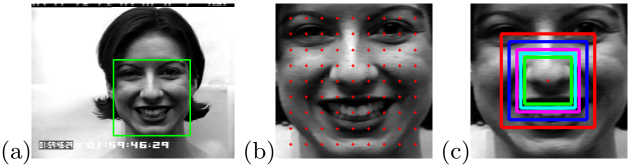

Fig. 2. Haar-like features are extracted from the face region at different spatial scales: (a) the face region is detected and cropped; (b) centers of the sliding window used to compute the Haar-like features; (c) multiple scales used to calculate Haar-like features.

region are summed up and the difference between these sums yields the Haar-like feature. In [31], Haar-like features are compared against a threshold and used to detect the face; therefore they are used as weak classifiers and a high number of features are considered in order to build a strong classifier. The key advantage of a Haar-like feature over most other features is that it can be calculated in constant time due to the use of integral images.

A number of Haar-like features have been used in literature [13], [8], and Haar-like features and/or simple variations have been formerly used in literature for emotion recognition [27],[33],[34] within boosting approaches.

In this paper we only use the six most common features depicted in Figure 1. Intuitively, a multi-scale approach might account for different intensity of the emotion, which may change from subject-to-subject. Therefore we extract Haar-like features at different spatial scales. In this preliminary work, we do not model the weights of each extracted feature; modeling these weights/performing feature selection remains a topic of future investigations. The main steps we perform to extract our face appearance representation are:

- we detect the face region (as shown in Fig. 2 (a));
- within the face region, we consider a set of uniformly sampled points (red dots in Fig. 2 (b));
- we center windows of varying spatial scales at each of these sampled points (Fig. 2 (c) shows the windows centered at a representative point on the subject's nose. Each color indicates a different scale.);
- we extract our Haar-like features from each of the selected windows. Whenever the sliding window exceeds the size of the face region (especially along the boundary), the window is cropped so to consider only the pixels within the face area. In our implementation, the white and black rectangular regions of each Haar-like feature are computed in proportion to the window size, therefore the cropping does not affect the computation of the Haar-like features.

4 Ensemble of Hankel Matrices for Emotion-Dynamics

In this section, first we briefly review LTI systems and Hankel matrix, then we describe our ensemble of Hankel matrices for emotion-dynamics representation.

4.1 Hankel Matrix-Based Dynamics Representation

In a LTI system, two linear equations regulate the behavior of the system:

$$\begin{aligned} x_{k+1} &= A \cdot x_k + w_k; \\ y_k &= C \cdot x_k. \end{aligned} \tag{1}$$

The first equation is known as the *state equation* and involves the variable $x_k \in R^u$, which represents the u-dimensional internal state of the LTI system. The second equation is known as the *measurement equation* and provides a link between the state of the system x_k and the v-dimensional observable measurement y_k. In such equations the matrices A and C are constant over time, and $w_k \sim N(0, Q)$ is uncorrelated zero mean Gaussian measurement noise.

It is well known [30] that, given a sequence of output measurements $[y_o, \ldots, y_\tau]$ from Eq. 1, its associated truncated block-Hankel matrix is

$$\widetilde{H} = \begin{bmatrix} y_0, & y_1, & y_2, & \ldots, & y_m \\ y_1, & y_2, & y_3, & \ldots, & y_{m+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ y_n, & y_{n+1}, & y_{n+2}, & \ldots, & y_\tau \end{bmatrix}, \tag{2}$$

where n is the maximal order of the system, τ is the temporal length of the sequence, and it holds that $\tau = n + m - 1$.

The Hankel matrix embeds the observability matrix Γ of the system, since $\widetilde{H} = \Gamma \cdot X$, where $X = [x_0, x_1, \cdots, x_\tau]$ is a matrix formed by the sequence of internal states of the LTI system.

As previously done in [12], [17], we normalize the Hankel matrix $\widetilde{H}$ as follows:

$$H = \frac{\widetilde{H}}{\sqrt{||\widetilde{H} \cdot \widetilde{H}^T||_F}}. \tag{3}$$

and compare two Hankel matrices H_p and H_q by the following similarity score:

$$s(H_p, H_q) = ||H_p^T \cdot H_q||_F, \tag{4}$$

which can be easily derived from the dissimilarity score in [12]. We have experimentally found that our similarity score is numerically more stable and fast to compute than the dissimilarity score. Such score can be regarded as an approximation of the cosine of the subspace angle between the spaces spanned by the columns of the Hankel matrices. As such, it can convey the degree to which two Hankel matrices may correspond to the same dynamical system.

4.2 Emotion-Dynamics Representation

The simple and fast appearance feature extraction described in Section 3 yields to a set of time series $Y = \{y^{i,j}\}_{i=1,j=1}^{i=N,j=S}$ where $y^{i,j} = \{y_1^{i,j}, \cdots y_\tau^{i,j}\}$ is the time series corresponding to the i-th Haar-like feature at the j-th spatial scale (N is the number of Haar-like features, and S is the number of scales). Each element $y_t^{i,j}$ of this time series is a vector of features computed at the uniformly sampled points and representing the t-th face in the face image sequence.

We use the set of time series Y to build an ensemble of Hankel matrices $H = \{H^{i,j}\}_{i=1,j=1}^{i=N,j=S}$ where each Hankel matrix $H^{i,j}$ is built upon the time series $y^{i,j}$ and, therefore, is associated with the i-th Haar-like feature and the j-th spatial scale. Before calculating the Hankel matrix, the sequence $y^{i,j}$ is made zero mean. We note the following:

- each vector $y_t^{i,j}$ is an ordered set of appearance features extracted from different parts of the face region. The set of Hankel matrices $H^{i,j}$ captures the dynamics of the Haar-like features over the whole face;
- each Hankel matrix is built upon a single Haar-like feature;
- each Hankel matrix is built upon a single scale;
- modeling separately Haar-like features at different spatial scales has computational advantages in terms of memory and time complexity;
- Hankel matrices can be obtained by a simple and fast reordering of the elements in the vector $y_t^{i,j}$. Therefore, from a computational point of view, the adoption of Hankel matrices over other time series representation is particularly appealing.

5 Emotion Classification

To test the effectiveness of our novel representation we have adopted the simple and widely used nearest-neighbor classifier (NN). We compare Hankel matrices by using the similarity score in Eq. 4. Given an ensemble of Hankel matrices, each Hankel matrix contributes to the emotion classification by voting for a class (predicted by NN). Comparison of Hankel matrices is done on equal terms of Haar-like feature and scale (we compare only Hankel matrices that share the same scale and Haar-like feature). Decision on the predicted class is performed considering a majority vote schema.

Other classification frameworks might be used, such as an LTI system codebook based representation similar to that proposed in [12], or a state-based approach similar to that in [17]. Alternatively, system identification techniques such as the ones applied in [29], [26] can be adopted at the cost of an increased overall time complexity. Even if stronger classification frameworks might be adopted as well, NN allows us to study the effectiveness of our representation without introducing further classifier-dependent parameters.

Table 1. Accuracy in Emotion Classification on the CK+ dataset. Red font indicates the best accuracy value per emotion, while bold font highlights the second best performance. **Different validation protocol (10-fold cross validation)

Features	Method	An.	Con.	Disg.	Fear	Hap.	Sad	Surp.	Avg
▄	DTW + NN	37.8	55.6	55.9	16	73.9	21.4	73.5	47.7
▐	DTW + NN	40	38.9	32.2	20	69.6	10.7	54.2	37.9
▮	DTW + NN	40	44.4	22	20	63.8	14.3	50.6	36.4
▬	DTW + NN	42.2	66.7	62.7	12	78.3	10.1	73.5	49.4
▞	DTW + NN	35.6	38.9	54.2	12	65.2	10.7	66.3	40.4
▣	DTW + NN	57.8	61.1	59.3	16	68.1	14.3	72.3	49.8
▄ + ▣	DTW + NN	53.3	55.6	62.7	16	72.5	10.7	81.9	50.4
▐ + ▣	DTW + NN	46.7	72.2	52.5	20	79.7	7.1	67.5	49.4
▮ + ▣	DTW + NN	48.6	55.6	50.8	24	78.3	17.9	65.1	48.6
▬ + ▣	DTW + NN	60	55.6	59.3	16	76.8	14.3	80.7	51.8
▞ + ▣	DTW + NN	53.3	66.7	57.6	20	78.3	7.1	79.5	51.8
▮ + ▬	DTW + NN	44.4	61.1	50.8	24	84.1	10.7	73.5	49.8
all	DTW + NN	42.2	72.2	59.3	20	87	14.3	83.1	54
▄	Hankel + NN	62.2	72.2	88.1	40	100	42.9	92.8	71.2
▐	Hankel + NN	71.1	61.1	81.4	44	94.2	64.3	87.9	72
▮	Hankel + NN	57.8	61.1	81.4	44	97.1	53.6	84.3	68.5
▬	Hankel + NN	44.4	66.7	84.7	40	**98.5**	21.4	94	64.3
▞	Hankel + NN	77.8	83.3	83	48	97.1	42.9	90.4	74.6
▣	Hankel + NN	71.1	**77.8**	91.5	48	100	60.7	96.4	77.9
▄ + ▣	Hankel + NN	68.9	**77.8**	93.2	44	100	57.1	96.4	76.8
▐ + ▣	Hankel + NN	82.2	83.3	91.5	44	100	**78.6**	91.6	81.6
▮ + ▣	Hankel + NN	75.6	83.3	89.8	48	100	71.4	92.8	80.1
▬ + ▣	Hankel + NN	60	72.2	89.8	40	100	53.6	94	72.8
▞ + ▣	Hankel + NN	**84.4**	**77.8**	89.8	**56**	100	64.3	95.2	81.1
▮ + ▬	Hankel + NN	62.2	**77.8**	89.8	44	100	57.1	91.6	74.6
all	Hankel + NN	86.7	83.3	96.6	52	100	71.4	**97.6**	**83.9**
CAPP	SVM [20]	70	21.9	**94.7**	21.7	100	60	98.7	66.7
LDN	RBF-SVM [24]**	71.7	73.7	93.4	90.5	95.8	78.9	**97.6**	85.9
Shape (SPTS)	SVM [20]	35	25	68.4	21.7	98.4	4	100	50.4
Shape+CAPP	SVM [1]	70.1	52.4	92.5	72.1	94.2	45.9	93.6	74.4
Shape	ITBN [32]	91.1	78.6	94	83.3	89.8	76	91.3	86.3
Shape	LRBM [22]	97.8	72.2	89.8	84	100	78.6	97.6	88.6
Shape + Hankel	NN [16]	91.1	83.3	94.9	84	100	71.4	98.8	89.1

Table 2. Confusion Matrix on the CK+ dataset when all the six Haar-like features are used. True labels are on rows, and predicted labels are on columns.

Tr. vs Pr.	**Angry**	**Contempt**	**Disgust**	**Fear**	**Happy**	**Sadness**	**Surprise**
Angry	**86.67**	0	2.22	2.22	0	6.67	2.22
Contempt	0	**83.33**	0	0	5.56	5.56	5.56
Disgust	0	0	**96.61**	0	1.69	0	1.69
Fear	12	4	0	**52**	24	4	4
Happy	0	0	0	0	**100**	0	0
Sadness	7.14	3.57	0	0	3.57	**71.43**	14.29
Surprise	0	1.20	0	0	1.20	0	**97.59**

6 Experimental Results

We have performed experiments in emotion recognition on the widely adopted Extended Cohn-Kanade dataset (CK+) [20]. This dataset provides facial expressions of 210 adults. Participants were instructed to perform several facial displays representing either single or combinations of action units. Based on the coded action units and by means of a validation procedure of the assigned label, the segmented recording of the participants' emotions were classified into 7 categories (in brackets the number of available samples): *angry (45), contempt (18), disgust (59), fear (25), happy (69), sadness (28), surprise (83).* In total there are 327 sequences of the 7 annotated emotions, performed by 118 different individuals. The number of frames of these sequences ranges in $[6, 71]$ with an average value of about 18 ± 8.6. The dataset provides landmark tracking results obtained by active appearance model, which we use in our experiments only to detect the face region. We adopted the validation protocol suggested in [20], which is leave-one-subject-out cross validation.

When extracting the Haar-like features, we sample the center location uniformly with a step equals to 10% of the size of the detected face region, yielding a 81 dimensional vector for each Haar-like feature. The spatial scales (size of the window used to calculate the Haar-like feature) are also computed in proportion to the face region size and the percentage ranges in $\{30, 35, 40, 50, 60\}$. The order of each Hankel matrix has been empirically set to 2. To extract the Haar-like features we have modified the implementation used in [7], [15].

6.1 Results

We have performed an extensive validation of the dynamics-based emotion representations whose results are reported in Table 1. The table reports the per-class classification accuracy values for each of the emotion classes, and the average accuracy. The table is divided in **4** parts. The first part presents accuracy values in classification when the raw features are adopted (namely Haar-like features). In this case, as the face image sequences have different lengths, dynamic time warping (DTW) is used to align the sequences and nearest-neighbor classifier is used over the aligned sequences. For a fair comparison, also when adopting the raw features, different Haar-like features and spatial scales are compared separately and a majority vote schema is used to predict the final class.

The second part of the table presents results when an ensemble of Hankel matrices is used. Both the first and second part of the table report performance when a single Haar-like feature is used, when a pair of Haar-like features is used and, finally, when all the six Haar-like features are used.

By comparing the first and second part of the table, there is a clear advantage in using an ensemble of Hankel matrices to represent the emotions over using directly the Haar-like features. On average, the increase of performance in using the dynamics-based representation with respect to the raw measurements is of about 60.3%.

Looking at the performance of each single Haar-like feature, the most informative one is the concentric squared regions (the last of the six features). Therefore we have performed experiments to study the performance of this feature when coupled with another Haar-like feature. As the table shows, there is an improvement with three of the five Haar-like features. There is no improvement when the Haar-like feature is coupled with the first Haar-like feature and a degradation of the performance when coupled with the vertical bands Haar-like feature. What is striking is that in all the experiments, the emotion Happy is always correctly recognized 100% of times. This suggests that our ensemble of Hankel matrices can be appropriate for smile detection. A further improvement of the performance is obtained when all the Haar-like features are used together, at the cost of an higher computational complexity. We suspect that not all the features are actually contributing to the recognition of the emotion, and feature and scale selection techniques may help to achieve more accurate results.

The third part of the table reports accuracy values of state-of-the-art methods adopting only appearance features. The class for which our method seems to fail the most is the emotion *Fear*. If we ignore this class, our method achieves even better accuracy values of the most competitive method in [24]. For completeness, the fourth part of the table reports the performance of techniques adopting accurate estimation of facial landmarks (provided together with the dataset). Even if these methods are not directly comparable with the ones that use only appearance information, we note that our appearance-based representation competes already very well against these techniques.

Finally, Table 2 reports the confusion matrix of our method. The class *Fear* is confused mostly with the class *Happy*. Some confusion is also present between the classes *Sadness* and *Surprise*. We believe that these ambiguities might be probably solved with fine-grained appearance descriptors, such as the Local Directional Number (LDN) pattern introduced in [24].

7 Conclusions and Future Work

In this paper we have proposed to use an ensemble of Hankel matrices to represent the dynamics of face appearance features, where each Hankel matrix embeds the dynamics of a single appearance feature at a given spatial scale.

We have tested our novel emotion representation on a widely used publicly available benchmark (CK+). Our experiments demonstrate that, on equal terms of classification framework and feature representations, the dynamics-based emotion representation achieves about 60.3% of increase in the accuracy values with respect of using directly the raw measurements. Overall, our approach achieves competitive performance with respect to more sophisticated machinery or methods that use accurate shape information.

Our formulation is general and it is not limited to the adopted face appearance representation. We therefore aim at extending our work by considering other appearance features. Moreover, we believe that feature and scale selection techniques (i.e. boosting) might led to an increase of the accuracy of our approach. In this paper, we have focused on the problem of classifying segmented emotion sequences. In future works we aim at tackling with the problem of emotion intensity estimation and emotion detection in face image sequences. In this sense, we will explore how face appearance feature dynamics correlate with the intensity of face emotions and if they can help in detecting subtle changes in face expressions.

Acknowledgments. This work was partially supported by Italian MIUR grant PON0101687, SINTESYS - Security and INTElligence SYStem.

References

1. Chew, S.W., Lucey, P., Lucey, S., Saragih, J., Cohn, J.F., Sridharan, S.: Person-independent facial expression detection using constrained local models. In: Conf. and Workshop on Automatic Face & Gesture Recognition (FG), pp. 915–920. IEEE (2011)
2. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI) **23**(6), 681–685 (2001)
3. Cuturi, M.: Fast global alignment kernels. In: Int. Conf. on Machine Learning (ICML), pp. 929–936 (2011)
4. Cuturi, M., Vert, J., Birkenes, O., Matsui, T.: A kernel for time series based on global alignments. In: Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), vol. 2, pp. II-413. IEEE (2007)
5. Dibeklioğlu, H.: Enabling dynamics in face analysis. Ph.D. thesis, University of Amsterdam (2014)
6. Fasel, B., Luettin, J.: Automatic facial expression analysis: a survey. Pattern Recognition **36**(1), 259–275 (2003)
7. Hare, S., Saffari, A., Torr, P.H.: Struck: Structured output tracking with kernels. In: Int. Conf. on Computer Vision (ICCV), pp. 263–270. IEEE (2011)
8. Huang, C., Ai, H., Li, Y., Lao, S.: High-performance rotation invariant multiview face detection. IEEE Trans. on Pattern Analysis and Machine Intelligence **29**(4), 671–686 (2007)
9. Jeni, L.A., Girard, J.M., Cohn, J.F., De La Torre, F.: Continuous AU intensity estimation using localized, sparse facial feature space. In: Conf. on Automatic Face & Gesture Recognition (FG), pp. 1–7. IEEE (2013)

10. Lacava, P.G., Golan, O., Baron-Cohen, S., Myles, B.S.: Using assistive technology to teach emotion recognition to students with asperger syndrome a pilot study. Remedial and Special Education **28**(3), 174–181 (2007)
11. Lee, H.Y., Lee, W.H.: A study on interactive media art to apply emotion recognition. International Journal of Multimedia & Ubiquitous Engineering 9(12) (2014)
12. Li, B., Camps, O.I., Sznaier, M.: Cross-view activity recognition using Hankelets. In: Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 1362–1369. IEEE (2012)
13. Lienhart, R., Maydt, J.: An extended set of haar-like features for rapid object detection. In: Int. Conf. on Image Processing, vol. 1, pp. I-900. IEEE (2002)
14. Lo Presti, L., La Cascia, M.: An on-line learning method for face association in personal photo collection. Image and Vision Computing (2012)
15. Lo Presti, L., La Cascia, M.: Tracking your detector performance: how to grow an effective training set in tracking-by-detection methods. In: Int. Conf. on Computer Vision Theory and Applications (VISAPP), pp. 1–8 (2015)
16. Lo Presti, L., La Cascia, M.: Using Hankel matrices for Dynamics-based Facial Emotion Recognition and Pain Detection. In: Int. Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1–8 (2015)
17. Lo Presti, L., La Cascia, M., Sclaroff, S., Camps, O.: Gesture modeling by hanklet-based hidden markov model. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9005, pp. 529–546. Springer, Heidelberg (2015)
18. Lo Presti, L., Sclaroff, S., Rozga, A.: Joint alignment and modeling of correlated behavior streams. In: Int. Conf. on Computer Vision-Workshops (ICCVW), pp. 730–737 (2013)
19. Lorincz, A., Jeni, L.A., Szabó, Z., Cohn, J.F., Kanade, T.: Emotional expression classification using time-series kernels. In: Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 889–895. IEEE (2013)
20. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The Extended cohn-kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression. In: Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 94–101. IEEE (2010)
21. Lucey, P., Cohn, J.F., Prkachin, K.M., Solomon, P.E., Matthews, I.: Painful data: the UNBC-McMaster shoulder pain expression archive database. In: Conf. and W. on Automatic Face & Gesture Recognition (FG), pp. 57–64. IEEE (2011)
22. Nie, S., Wang, Z., Ji, Q.: A generative restricted Boltzmann machine based method for high-dimensional motion data modeling. Computer Vision and Image Understanding (2015)
23. Rabbitt, S.M., Kazdin, A.E., Scassellati, B.: Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. Clinical Psychology Review **35**, 35–46 (2015)
24. Ramirez Rivera, A., Castillo, R., Chae, O.: Local directional number pattern for face analysis: Face and expression recognition. IEEE Transactions on Image Processing (TIP) **22**(5), 1740–1752 (2013)
25. Rehg, J.M., et al.: Decoding children's social behavior. In: Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 3414–3421. IEEE (2013)
26. Sankaranarayanan, A.C., Turaga, P.K., Baraniuk, R.G., Chellappa, R.: Compressive acquisition of dynamic scenes. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 129–142. Springer, Heidelberg (2010)

27. Sariyanidi, E., Gunes, H., Cavallaro, A.: Automatic analysis of facial affect: A survey of registration, representation and recognition. EEE Trans. on Pattern Analysis and Machine Intelligence (PAMI) (2014)
28. Shan, C., Gong, S., McOwan, P.W.: Dynamic facial expression recognition using a Bayesian temporal manifold model. In: BMVC, pp. 297–306 (2006)
29. Slama, R., Wannous, H., Daoudi, M., Srivastava, A.: Accurate 3D action recognition using learning on the Grassmann manifold. Pattern Recognition (PR) **48**(2), 556–567 (2015)
30. Viberg, M.: Subspace-based methods for the identification of linear time-invariant systems. Automatica **31**(12), 1835–1851 (1995)
31. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision **57**(2), 137–154 (2004)
32. Wang, Z., Wang, S., Ji, Q.: Capturing complex spatio-temporal relations among facial muscles for facial expression recognition. In: Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 3422–3429. IEEE (2013)
33. Yang, P., Liu, Q., Metaxas, D.: Similarity features for facial event analysis. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 685–696. Springer, Heidelberg (2008)
34. Yang, P., Liu, Q., Metaxas, D.N.: Boosting coded dynamic features for facial action units and facial expression recognition. In: Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 1–6. IEEE (2007)
35. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI) **31**(1), 39–58 (2009)
36. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI) **29**(6), 915–928 (2007)
37. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys (CSUR) **35**(4), 399–458 (2003)
38. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 2879–2886. IEEE (2012)