Recognising Human Emotions from Body Movement and Gesture Dynamics

Ginevra Castellano[1], Santiago D. Villalba[2], and Antonio Camurri[1]

[1] Infomus Lab, DIST, University of Genoa*
[2] MLG, School of Computer Science and Informatics, University College Dublin
ginevra.castellano@unige.it santiago.villalba@ucd.ie

Abstract. We present an approach for the recognition of acted emotional states based on the analysis of body movement and gesture expressivity. According to research showing that distinct emotions are often associated with different qualities of body movement, we use non-propositional movement qualities (e.g. amplitude, speed and fluidity of movement) to infer emotions, rather than trying to recognise different gesture shapes expressing specific emotions. We propose a method for the analysis of emotional behaviour based on both direct classification of time series and a model that provides indicators describing the dynamics of expressive motion cues. Finally we show and interpret the recognition rates for both proposals using different classification algorithms.

1 Introduction

One critical aspect of human-computer interfaces is the ability to communicate with users in an expressive way [1]. Computers should be able to recognise and interpret users' emotional states and to communicate expressive-emotional information to them. Recently, there has been an increased interest in designing automated video analysis algorithms aiming to extract, describe and classify information related to the emotional state of individuals. In this paper we focus on video analysis of movement and gesture as indicators of an underlying emotional process.

Our research aims to investigate which are the motion cues indicating differences between emotions and to define a model to recognise emotions from video analysis of body movement and gesture dynamics. The possibility to use movement and gesture as indicators of the state of individuals provides a novel approach to quantitatively observe and evaluate the users in an ecological environment and to respond adaptively to them. Our research is grounded in the Component Process Model of emotion (CPM) proposed by Scherer [2], and explores the component of the model representing motor activation in emotion. Specifically, the CPM describes a relationship between the results of different event appraisals (such as novelty, intrinsic pleasantness, goal conduciveness etc.)

* Part of this work was carried out in the context of the EU Project HUMAINE (Human-Machine Interaction Network on Emotion)(IST-2002-507422).

and the response patterning with respect to physiological arousal, action tendency, subjective feeling and, in particular, motor expression.

In this paper we present an approach for the recognition of four acted emotional states (anger, joy, pleasure, sadness) based on the analysis of body movement and gesture expressivity. According to research showing that distinct emotions are often associated with different qualities of body movement (see for example [3]), we use non-propositional movement qualities (e.g. amplitude, speed and fluidity of movement) to infer emotions, rather than trying to recognise different gesture shapes expressing specific emotions, to investigate the role of movement expressivity versus shape in gesture. We propose a method for the analysis of emotional behaviour based on direct classification of time series and on a model that provides indicators describing the dynamics of expressive motion cues. Finally we show and interpret the recognition rates for both proposals using different classification algorithms.

2 Motivations and Related Work

Research aiming to endow computers with the ability to mimic human emotional intelligence and to recognise emotions communicated by others through body movement needs to be grounded in neuroscience and psychology. Psychological studies on visual analysis of body movement show that human movement differs from other movements because it is the only visual stimulus we have experience of both perceiving and producing [4]. Understanding the human processing mechanism of the stimuli coming from the so called "biological movement"[5], the one produced by animals in general and humans in particular, would lead to better designed computational models of emotion based on visual analysis of affective body language. Further, theories linking the response to motion visual stimuli to mirror motion-visual neurons would allow a deeper understanding of how we comprehend others' emotions and develop empathy [6]. Finally, several studies from psychology focus on the relationships between emotion and movement qualities, and investigate expressive body movements [7–9].

In human-computer interaction a central role is played by automated video analysis techniques aiming to extract physical characteristics of humans and use them to infer information related to the emotional state of individuals. Modelling emotional behaviour starting from automatic analysis of movement and gesture is currently still a poorly explored field though. Most of the previous studies, in fact, focus on emotion recognition based on audio and facial expression data. Nevertheless, some attempts were made in the direction of designing systems able to analyse expressive body movements and use this information to recognise emotions.

Burgoon et al. [10] proposed an approach for analysing cues from multiple body regions for the automated identification of emotions displayed in videos. Camurri et al. [11] classified expressive gestures in human full-body movement (dance performances). They identified cues deemed important for emotion recognition, like quantity of motion and contraction/expansion of the body, and showed how these cues could be tracked by automated recognition techniques. Kapur et al. [12] used full-body skeletal movements data obtained with a technology based on the VICON motion capturing system to classify four emotional states. Bianchi-Berthouze and Kleinsmith [13] proposed a model that can self-organise postural features into affective categories to give robots the ability to incrementally learn to recognise affective human postures through interaction with human partners. Other studies show that expressive gesture analysis and classification can be obtained by means of automatic image processing [14] and that the integration of multiple modalities (facial expressions and body movements) is successful for multimodal emotion recognition. Gunes and Piccardi [15] for example fused facial expression and body gesture information at different levels for bimodal emotion recognition. Further, el Kaliouby and Robinson [16] proposed a vision-based computational model to infer acted mental states from head movements and facial expressions.

3 Experimental Setup and Data Collection

Our work is based on a corpus of 240 gestures collected during the Third Summer School of the HUMAINE (Human-Machine Interaction Network on Emotion) EU-IST project, held in Genoa in September 2006. Ten participants (six male and four female) were asked to act eight emotional states (anger, despair, interest, pleasure, sadness, irritation, joy and pride) equally distributed in the valence-arousal space. In this study we focus mainly on four emotions: anger, joy, pleasure and sadness (see Table 1).

	positive-valence	*negative-valence*
high-arousal	joy	anger
low-arousal	pleasure	sadness

Table 1. The emotions in the space valence-arousal.

We asked the participants to perform the same gesture (raising and lowering the arms in the coronal plane, starting with the arms down by the side while standing in the rest position) trying to express the different emotional conditions. We chose this experimental setup because we are interested in investigating the role of movement expressivity versus shape in gesture. For this reason, we decided to use the same gesture, whose shape does not appear to convey any obvious emotional expression or meaning, while expressing different emotions. In this way we evaluate whether it is possible to recognise the emotions based only on expressivity. The gesture was repeated three times for each emotion, so that we collected 240 posed gestures.

During the experiment, subjects' full-bodies were recorded by a DV camera (25 fps) viewed from the front. A uniform dark background was used in order to make the silhouette extraction process easier. Further, a long sleeved shirt was worn by all the participants in order to make the hand tracking feasible with our available technical resources.

3.1 Feature Extraction

Our approach to video analysis of gesture and body movement allows the silhouette and body parts to be tracked without the need for markers. We used the EyesWeb platform [17] to extract the whole silhouette and the hands of the subjects from the background. The EyesWeb Expressive Gesture Processing Library [18] was used to compute five different expressive motion cues: quantity of motion and contraction index of the body, velocity, acceleration and fluidity of the hand's barycentre.

The quantity of motion (QoM) is a measure of the amount of detected motion, computed with a technique based on silhouette motion images (SMIs). These are images carrying information about variations of the silhouette shape and position in the last few frames (see Figure 1).

$$SMI[t] = \sum_{i=0}^{n} Silhouette[t-i]\} - Silhouette[t] \qquad (1)$$

The SMI at frame t is generated by adding together the silhouettes extracted in the previous n frames and then subtracting the silhouette at frame t. The resulting image contains just the variations that happened in the previous frames.

Fig. 1. A measure of QoM using SMIs (the shadow along the arms and the body) and the tracking of the hand.

QoM is computed as the area (i.e., number of pixels) of a SMI, normalised in order to obtain a value usually ranging from 0 to 1. That can be considered as an overall measure of the amount of detected motion, involving velocity and force.

$$QoM = Area(SMI[t,n])/Area(Silhouette[t]) \qquad (2)$$

The contraction index (CI) is a measure, ranging from 0 to 1, of the degree of contraction and expansion of the body. CI can be calculated using a technique related to the bounding region, i.e., the minimum rectangle surrounding the body: the algorithm compares the area covered by this rectangle with the area currently covered by the silhouette.

Velocity (Vel) and acceleration (Acc) are related to the trajectory followed by the hand's barycentre in a 2D plane. Fluidity gives a measure of the uniformity of motion, so that fluidity is considered maximum when, in the movement between two specific points of the space, the acceleration is equal to zero. It is computed as the Directness Index [18] of the trajectory followed by the velocity of hand's barycentre in the 2D plane.

Due to technical and time constraints during the recordings our data became noisy somehow. Since we didn't use markers the hand-tracking was difficult, even in our controlled environment. We lost 14 full videos and some frames at the end of a small number of other videos. This is only a reflection of the current limitations of technology that a deployed recognition system should be able to handle. Missing and noisy data, though, are expected as inputs to a production system.

The gestures are then described by the profiles over time of expressive motion cues. In order to compare the gestures from all the subjects, the data were normalised considering the maximum and the minimum values of each motion cue in each actor. We used these normalised time series as inputs for the recognition experiment described in section 4.2.

A further processing step converted these temporal series into a fixed set of indicators or meta-features (see Figure 2) conveying information about the dynamics of the gesture expressivity over time.

- **Initial and Final Slope**: slope of the line joining the first value and the first relative extremum, slope of the line joining the last value and the last relative extremum.
- **Initial (Final) Slope of the Main Peak**: slope of the line joining the absolute maximum and the preceding (following) minimum.
- **Maximum, Mean, Mean / Max, Mean / Following Max**: the maximum and mean values and their ratio, ratio between the two first biggest values.
- **Maximum / Main Peak Duration, Main Peak Duration / Duration**: ratio between the maximum and the main peak duration, ratio between the peak containing the absolute maximum and the total gesture duration.
- **Centroid of Energy, Distance between Max and Centroid**: location of the barycentre of energy, distance between the maximum and the barycentre of energy.
- **Shift Index of the Maximum, Symmetry Index**: position of the maximum with respect to the centre of the curve, symmetry of the curve relative to the maximum value position.
- **Number of Maxima, Number of Maxima preceding the Main One**: number of relative maxima, number of relative maxima preceding the absolute one.

Fig. 2. The features computed from the expressive motion cues.

Starting from the temporal profiles of each cue, information about its shape was calculated. Automatic extraction of the selected features was made using new software modules developed in EyesWeb. This process was made for each motion cue, so that each gesture is characterised by a set of 80 (5x16) meta-features. These are the inputs for the classifiers used in the experiment described at section 4.3.

4 Emotion Recognition and Experimental Results

Our emotion recognition system is based on gesture dynamics captured from video via computer-vision techniques (see figure 3). We face a time-series classification problem, which can be solved in several ways. Two of the possible approaches are [19]:

- Use of techniques that specifically deal with temporal classification, working directly over the series data. Hidden Markov Models or Dynamic Time Warping are examples of this. We follow this approach in the recognition experiment described in section 4.2.
- Representing the problem in such a way as to allow the application of propositional concept learners (the more widespread and best understood classification techniques). This approach constructs a set of features trying to best describe the series. This is the direction we follow in the recognition experiment described in section 4.3.

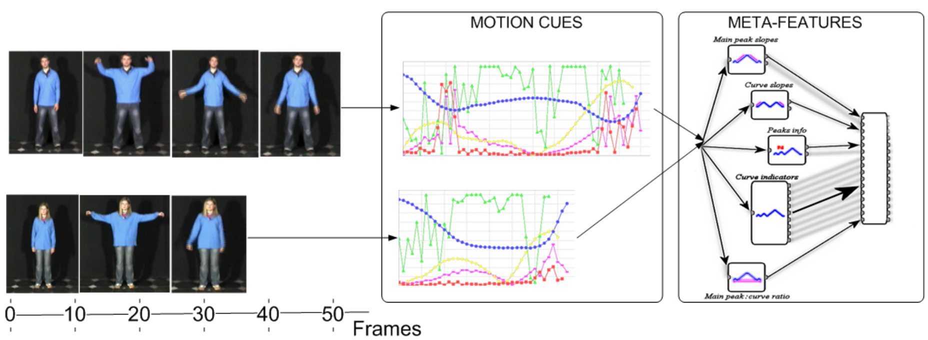

Fig. 3. Representing gesture videos as cue-time-series and meta-features. Several of the usual problems in series classification, such as scaling, misalignment and different duration are present.

4.1 Performance Estimation Procedure

Credibility is the main concern when evaluating the generalisation capability of a learning system [20]. In our case the small size of our dataset poses a challenge to both error estimation and overfitting (over-representation of the system performance) avoidance.

All the quantitative results presented here were obtained using leave one out cross validation (LOOCV). LOOCV allows us to train with the greatest amount of data each time, which increases the chance that the classifier is accurate and that both train and test data are representative. Since no random sampling is involved its results are deterministic. Standard 10-fold cross validation estimation was exhibiting a big, undesirable variance.

We also followed the sometimes overlooked principle of not allowing the learning machines to see in any way the testing data during the training phase. We found extremely easy to memorise our data and introduce overfitting artifacts by full-dataset pre-discretisation and pre-feature-selection, leading to fallacious error estimations. Supervised discretisation and feature selection were performed as an integral part of the whole classification system, that is, for each iteration of the LOOCV.

4.2 A Lazy Classifier based on Dynamic Time Warping

Dynamic Time Warping (DTW)[21] is a well known algorithm for measuring similarity between two sequences which may vary in phase or speed. It computes both a distance between series and a *warping path* or correspondence between pairs of points from the two series. In the last decade DTW has emerged as a common solution for time series alignment and comparison. It has already been applied to gesture temporal alignment for the same gesture performed with different styles [22].

Once we have a distance measure between time series, a world of possibilities is presented. For example, we can apply clustering algorithms, or use the distance matrix as input for Kernel-based algorithms (e.g. SVMs) or other dissimilarity based classifiers. In this study we used simple a nearest neighbour [20] based on DTW distance (1NN-DTW) as the classifier for our system. A new gesture is assigned to the emotion of the gesture in the training set to a minimum distance, its nearest neighbour. Different from the other alternatives, this simple approach allows us to interpret the results in a more straightforward way and, furthermore, it has proven effective.

There are several perspectives from which the recognition system should be evaluated. Table 2 shows the results for the dataset containing the eight original emotions and those with a reduced set containing only four emotions (anger, joy, pleasure, sadness). We evaluated the performance of the system trained using only the personal set of gestures (i.e. training one classifier per subject, "personal space") and the universal set (i.e. an inter-subject enabled classifier using the universal set of gestures, "universal space"). From the results presented in table 2 we draw several conclusions.

– Our system is not able to discriminate successfully between the 8 emotions. We think that this is due not only to the increased complexity for the learning task, but also to inherent lacks in the task definition. Maybe it's just that

Cue ↓	4 Emotions (PS)	4 Emotions (US)	8 Emotions (PS)	8 Emotions (US)
Acceleration	0.54	0.57	0.73	0.77
CI	0.41	0.42	**0.53**	**0.59**
Fluidity	0.56	0.56	0.80	0.82
QoM	**0.34**	**0.37**	0.57	0.63
Velocity	0.40	0.43	0.64	0.65

Table 2. LOOCV errors for the 1NN-DTW classifier. Cues per rows, in columns experiment for 4 / 8 emotions, personal (PS) versus universal (US) classifiers. In the case of the personal space the error rate is averaged over all subjects.

those emotions are indistinguishable from the gesture at hand or that our features are not relevant.
- For the gesture considered QoM is the clear winner in the 4-emotions case.
- As it is the case in the real world, it's desirable for the classifier to get the knowledge of the actual subject nuances. No "universal" classifier can be flexible enough as to adapt to the modes of all subjects. By inspecting the nearest neighbour for each gesture, the 1NN classifier allows to measure an indicator in the universal space that we call the doubleHit / hit ratio. It is the percentage of correctly classified gestures (*hits*) that, furthermore, are correctly classified in the subject space (*doubleHits*). These values are of 69% and 77% for 4 and 8 emotions, that is, most of the emotion hits correspond to a gesture from the same subject. These results seem not coming by pure chance since, in fact, the more a universal system becomes a good predictor of the subject (41% and 35% respectively) the less the doubleHit / hit ratio. What we understand from those numbers is that the personalisation aspect is crucial. An hypothetic scenario would be starting from an universal recognitor and then to do a progressive fine tuning towards each specific subject.

4.3 A Simple Meta-Features Approach

Design of a good set of features to describe the considered problem domain is one of the key steps, if not the key, towards successful generalisation. A paradigmatic example comes from speech recognition, a mature biometric-based tough recognition problem. Features constructed to capture several aspects of speech, relying on models of human vocal tract and human sound perception, enable speaker-independent systems with a high word recognition rate.

A meta-feature is an abstraction of some substructure that is observed within the data [19]. For time series several meta-features have been proposed [19, 23]. By means of meta-features calculation we intend to present our data in a format that is appropriate for feature-vector-based classifiers.

We decided to use several simple holistic meta-features (see Figure 2), in the hope that they were neither too naive nor myopic as to extract useful information from our movement cues time series. If useful those meta-features (from now on simply "features") would provide interpretable results, allowing for a more comprehensible classification system. Further, since their calculation requires few computational efforts, they would be easily used to enable a real time recognition system.

We used Weka [20], an open source suite of machine learning tools for data mining, as the basis for classification. We adopted a comparative approach between three different classifiers.

- A simple 1-nearest-neighbor (1NN). It's the natural counterpart to the DTW-1NN approach presented in section 4.2.
- A decision-tree. We use J48, Quinlan's C4.5 implementation in Weka.
- A Bayesian Network. One of the state of the art techniques for emotion recognition [24]. After a first exploration of the different implementations we chose Hidden Naive Bayes (HNB) [25]. The rationale behind its better performance, compared to other Bayesian network architectures, is its more suitable handling of the independence assumption, given that our features are obviously non independent.

For HNB it is necessary to discretise the features. Using weka supervised discretisation module with the Kononenko's MDL criterion to evaluate the quality of the intervals exhibited a great improvement over using Weka's default setting. To be fair, we also tried the other two classifiers with the features discretised in this way. Our feature set is of high dimensionality for our training set size, so we performed two simple minded feature selection processes with a shallow forward search using both, a correlation based and a wrapper approach, to measure the merit of each features subset [20]. In all cases we used the default values for the classifiers parameters. The results are shown in table 3.

Dataset →	All	CB-FS	Wr-FS	D-ALL	D-CB-FS	D-Wr-FS
HNB	-	-	-	47.22	37.92	51.85
J48	**52.78**	47.22	48.15	56.48	48.15	47.22
1NN	53.70	**43.52**	**37.04**	**44.44**	**32.41**	**44.44**

Table 3. LOOCV error rates for our meta-features experiment applied to the 4-emotions dataset. In columns the different preprocessing steps (All= no preprocessing, CB-FS= Correlation Based Feature Selection, Wr-FS = Wrapper Feature Selection, D = Discretisation).

The three chosen classifiers produce in different levels interpretable results which allowed us to confirm some of the hypotheses coming from the previous analysis. One of them is that the same cues remain the winners with our metafeatures, since the constructed trees and the selected features show a strong bias towards the choice of QoM based features, with a minor role played by CI. In particular, the maximum of QoM seems to be one of the most significant features: it discriminates between "high arousal" emotions (anger and joy) and "low arousal" emotions (pleasure and sadness), with the first ones showing higher values for the maximum of QoM. Further, the mean of CI discriminates between "positive" and "negative" emotions: pleasure and joy show low values for mean of CI, whereas anger and sadness show high values for this feature. Another obvious conclusion is that we are generating irrelevant and redundant features. A classifier such as 1NN, which is very sensitive to the presence of noise, gets a strong improvement when the dimensionality is reduced.

Finally we report one of the confusion matrices in table 4.

Ang	**Joy**	**Ple**	**Sad**	**←true class**
0.9	0.1	0	0	Anger
0.2	**0.44**	0.28	0.08	Joy
0.0	0.21	**0.62**	0.17	Pleasure
0.08	0.08	0.36	**0.48**	Sadness

Table 4. Confusion matrix of the HNB classifier used after a Correlation based Feature Selection Process (D-CB-FS).

This table highlights that "negative" emotions (anger and sadness) were confused with the correspondent "positive" emotion with the same arousal characteristics. That suggests that arousal plays an important role in the recognition of emotion by movement and gesture and this is confirmed by the decision trees structure, that shows that QoM discriminates between "high" and "low arousal" emotions. Nevertheless, "positive" emotions (joy and pleasure) were misclassified with the correspondent "positive" emotion with opposite arousal characteristics. It seems that also valence plays a role in the recognition process. This is also confirmed from the decision trees structure, where it is evident that joy and pleasure have lower values of CI, i.e., the generated movements are more expanded than those of anger and sadness.

5 Conclusions and Future Work

We have presented an approach for automated video analysis of human gesture dynamics for emotion recognition. Since we do not use markers, this approach allows us to analyse human emotional behaviour in ecological environments in a non-intrusive way. We used movement expressivity to infer emotions and, in particular, we proposed a method for the analysis of emotional behaviour based on both direct classification of time series and a model that provides descriptors of the dynamics of expressive motion cues. We presented and analysed the classification results of our approach, finding that QoM is the most significant cue in differentiating between the emotions, with a minor role played by CI.

With this study we investigated the role of movement expressivity versus shape in gesture. To the best of our knowledge our experiments are novel.

No other publication have addressed the same problem of out-of-context intra-gesture emotion classification using dynamic movement information. Results showed how expressive motion cues allow to discriminate between "high" and "low arousal" emotions and between "positive" and "negative" emotions. This result appears interesting because, since we considered gestures with the same shape, we can conclude that movement expressivity allows to differentiate the four emotions considered.

Our work is applicable to several different scenarios and the final scope is to complement other modalities in a multimodal fusion recognition system. For this purpose though, it's necessary to collect more data to feed these kind of systems. Since with the small datasets available no definitive conclusions can be drawn nor a robust system can be constructed; we plan further recordings with a larger and more representative set of subjects and gestures.

We plan to extend the meta-features based system to a broader feature space considering local features of the time series, as proposed in [23]. More classification schemes and better model selection are to be explored, which constitutes a work in progress. A cleverer cue fusion is also in study.

Last but not least, the base assumption behind our system is that the subjects were actually expressing the requested emotion. Future work includes perceptive tests to verify how the acted emotions are perceived by humans. Neglecting further discussion about this, how to construct a pervasive recognition system based on "real-life" emotion data is still a challenging open question.

6 Acknowledgments

We would like to thank Gualtiero Volpe for his advice and useful discussion.

References

1. Picard, R.W.: Affective Computing. The MIT Press (September 1997)
2. Scherer, K.R.: On the nature and function of emotion: a component process approach. In Scherer, K.R., Ekman, P., eds.: Approaches to emotion. Hillsdale, NJ: Erlbaum (1984) 293–317
3. Pollick, F., Paterson, H., Bruderlin, A., Sanford, A.: Perceiving affect from arm movement. Cognition **82** (2001) B51–B61
4. Shiffrar, M., Pinto, J.: The visual analysis of bodily motion. In Prinz, W., Hommel, B., eds.: Common mechanisms in perception and action: Attention and Performance. Oxford University Press (2002) 381–399
5. Giese, M.A., Poggio, T.: Neural mechanisms for the recognition of biological movements. Nature Reviews Neuroscience **4**(3) (March 2003) 179–192
6. Rizzolatti, G., Fogassi, L., Gallese, V.: Mirrors in the mind. Scientific American **295**(5) (November 2006) 54–61
7. Boone, R.T., Cunningham, J.G.: Children's decoding of emotion in expressive body movement: the development of cue attunement. Developmental psychology **34**(5) (September 1998) 1007–1016

8. De Meijer, M.: The contribution of general features of body movement to the attribution of emotions. Journal of Nonverbal Behavior **13**(4) (December 1989) 247–268
9. Wallbott, H.G.: Bodily expression of emotion. European Journal of Social Psychology **28**(6) (December 1998) 879–896
10. Burgoon, J.K., Jensen, M.L., Meservy, T.O., Kruse, J., Nunamaker, J.F.: Augmenting human identification of emotional states in video. In: Intelligence Analysis Conference, McClean, VA. (2005)
11. Camurri, A., Lagerlof, I., Volpe, G.: Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. International Journal of Human-Computer Studies **59**(1-2) (July 2003) 213–225
12. Kapur, A., Kapur, A., Babul, N.V., Tzanetakis, G., Driessen, P.F.: Gesture-based affective computing on motion capture data. In: ACII. (2005) 1–7
13. Bianchi-Berthouze, N., Kleinsmith, A.: A categorical approach to affective gesture recognition. Connection Science **15**(4) (December 2003) 259–269
14. Balomenos, T., Raouzaiou, A., Ioannou, S., Drosopoulos, A.I., Karpouzis, K., Kollias, S.D.: Emotion analysis in man-machine interaction systems. In: Machine Learning for Multimodal Interaction. (2004) 318–328
15. Gunes, H., Piccardi, M.: Bi-modal emotion recognition from expressive face and body gestures. Journal of Network and Computer Applications **In Press, Corrected Proof**
16. el Kaliouby, R., Robinson, P.: Generalization of a vision-based computational model of mind-reading. In: ACII. (2005) 582–589
17. Camurri, A., Coletta, P., Massari, A., Mazzarino, B., Peri, M., Ricchetti, M., Ricci, A., Volpe, G.: Toward real-time multimodal processing: Eyesweb 4.0. In: AISB 2004 Convention: Motion, Emotion and Cognition. (March 2004)
18. Camurri, A., Mazzarino, B., Volpe, G.: Analysis of expressive gesture: The Eyesweb Expressive Gesture processing library. In Camurri, A., Volpe, G., eds.: Gesture Workshop. Volume 2915 of Lecture Notes in Computer Science., Springer (2003) 460–467
19. Kadous, M.W.: Temporal Classification: Extending the Classification Paradigm to Multivariate Time Series. PhD thesis, School of Computer Science & Engineering, University of New South Wales (2002)
20. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems). Morgan Kaufmann (June 2005)
21. Keogh, E., Ratanamahatana, C.A.: Exact indexing of dynamic time warping. Knowledge and Information Systems **7**(3) (March 2005) 358–386
22. Heloir, A., Courty, N., Gibet, S., Multon, F.: Temporal alignment of communicative gesture sequences. Computer Animation and Virtual Worlds **17**(3-4) (July 2006) 347–357
23. Rodríguez, J.J., Alonso, C.J., Maestro, J.A.: Support vector machines of interval-based features for time series classification. Knowledge-Based Systems **18**(4–5) (2005) 171–178
24. Sebe, N., Cohen, I., Cozman, F.G., Gevers, T., Huang, T.S.: Learning probabilistic classifiers for human-computer interaction applications. Multimedia Systems **V10**(6) (October 2005) 484–498
25. Zhang, H., Jiang, L., Su, J.: Hidden naive bayes. In: AAAI 2005, The Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence. (2005) 919–924