Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 743–750
ISBN 978-83-60810-22-4

Automatic classification of gestures: a context-dependent approach

Mario Refice
IEEE Senior member
Department of Electrical Engineering and Electronics,
Polytechnical University of Bari, Italy Email: refice@poliba.it

Michelina Savino
Department of Psychology and Educational Sciences,
University of Bari, Italy
Email: m.savino@psico.uniba.it

Michele Adduci, Michele Caccia
Department of Electrical Engineering and Electronics,
Polytechnical Univ. of Bari, Italy
Email: {adduci.michele, michelecaccia84@gmail.com}

Abstract—**Gestures represent an important channel of human communication, and they are "co-expressive" with speech. For this reason, in human-machine interaction automatic gesture classification can be a valuable help in a number of tasks, like for example as a disambiguation aid in automatic speech recognition. Based on the hand gesture categorization proposed by D. McNeill in his reference works on gesture analysis, a new approach is here presented which classifies gestures using both their kinematic characteristics and their morphology stored as parameters of the templates pre-classified during the training phase of the procedure. In the experiment presented in this paper, an average of about 90% of correctly classified gesture types is obtained, by using as templates only about 3% of the total number of gestures produced by the subjects.**

I. Introduction

STUDIES on human gestures have been received considerable attention because they represent an important channel in human communication. The seminal work of Kendon [1] has set up a comprehensive scheme of classification and interpretation of human gestures in different languages and cultures. Based on these studies, McNeill [2] developed a detailed interpretation framework, according to which gestures and language are strictly interwined, as they are "co-expressive" in human communication.

As an almost natural consequence, studies on human gestures have received considerable attention also by applications developers, typically in technological fields like human computer interaction (HCI). Most of these research works have dealt with the problem of recognizing human gestures automatically by means of special equipments (gloves or similar pointing devices) or just taking advantage of the available Computer Vision technology, or even developing new methods for gestures recognition. Among the huge amount of work currently available on this matter, in this paper only a few cases will be recalled, which refer to the mentioned main approaches. In the Gesture Interpretation Module, developed within the project SMARTKOM [3], for example, the gesture channel is combined with other two input channels, i.e. face recording and speech. The project makes use of a very sophisticated equipment whose main purpose is to provide a useful laboratory environment for multimodal interaction studies. Yingen Xiong & Francis Quek [4] were able, using methods of computer vision, to analyze the hand motion of oscillating frequencies of gestures accompanying speech, and demonstrate that oscillatory gestures reveal portions of the multimodal discourse structure. Andrew D. Wilson & Aaron F. Bobick [5], on the other hand, proposed an extension of the standard HMM method of gesture recognition which shows a better performance in the representation, recognition and interpretation of pointing gestures.

In this paper, a different procedure is described, which aims at classifying hand gestures via a hybrid approach using both the spatial location of the movement and a morphological comparison of the movement with a set of reference templates obtained from the specific context. It makes use of standard equipments, i.e. standard 2D video recording, and is able to classify the main hand gestures of a human being while s/he is talking in a conversation environment. The application domain taken as a reference consists of automatic transcription systems where gestures capture can solve some interpretation ambiguities in the recognition of spoken sentences produced by a talker involved in a conversation with a single interlocutor or in front of an audience, such as in a conference environment.

Since, according to [2], gesture and speech are "co-espressive", automatic gesture analysis can help in assigning the correct semantic or pragmatic salience during the speech recognition process. In fact, it has been observed that gestural beats are normally syncronized with prosodically prominent syllables in speech (see for example [6], [7], [8]), and that iconic/metaphoric gestures are normally realised in relation to semantically salient words.

As it can be inferred from what discussed above, in our work we deal with communicative gestures only (i.e. speech accompanying gestures), whereas those unconsciously produced – also called "idiosyncratic" gestures – have been obviously not considered.

The methodology adopted in this system is inspired by McNeill's [2] classification of hand gestures into four main categories (iconics, metaphorics, deictics, beats). For gesture identification, instead, it assumes as discrimination factors both the kinematic of the gesture itself and its classification based on a template matching technique.

The system here presented makes use of the OCV [9] package which is an open source set of software modules covering most of the functionalities involved in state of the art video processing techniques.

The recognition phase performs its function in real time and produces an output, which is subsequently analyzed to give the classification of the gestures produced by the subject.

In section II the main outcomes of the McNeill's experimental work, which is the background knowledge of the present application, is recalled. In section III the recognition procedure adopted in this work is presented, along with the classification procedure. Finally, in section IV some results of the developed system are shown and discussed with reference to a specific experiment.

II. McNeill's classification scheme

As mentioned in section I, some basic assumptions can be made about gestures. First of all, they imply a movement, either of hands or head or some other part of the human body. With this respect the foundations of movements, namely of the hands, has to be acknowledged to McNeill work [2].

According to McNeill, spontaneous movements produced by humans while talking can be classified as:

Iconics: *"they bear a close formal relationship to the semantic content of speech [...] hand appears to grip something and pull it from the upper front space back and down near to the shoulder."* (McNeill, 1992: 12)

Metaphorics: *"The gesture present an image of the invisible-an image of an abstraction. The gesture depicts a concrete metaphor for a concept [...] Hands rise up and offer listener an object."* (McNeill, 1992:14)

Beats: *"The hand moves along with the rhythmical pulsation of speech [...].The typical beat is a simple flick of the hand or fingers up and down, or back and forth; the movement is short and quick and the space may be the periphery of the gesture space (the lap, an armrest of the chair, etc.)"* (McNeill, 1992:15)

Deictis: *"[...] is the familiar pointing [...] Points to space between self and interlocutor"* (McNeill,1992:18)

Using data coming from his experiments, McNeill shows also a diagram of the final position of hand gestures; such diagrams are shown in Fig. 1 a), b), c), d) respectively. Here, dots represent the density of spatial usage for each gesture category.

Even though the McNeill scheme is a descriptive one, mainly having the purpose of describing the psychological background of the producer, the experiment described in the present paper assumes the mentioned classification scheme as one basic assumption in order to classify gestures accordingly to their spatial location; a further discrimination is achieved by trying to match the image of the gesture with some pre-classified examples, which are context, and individual, dependent.

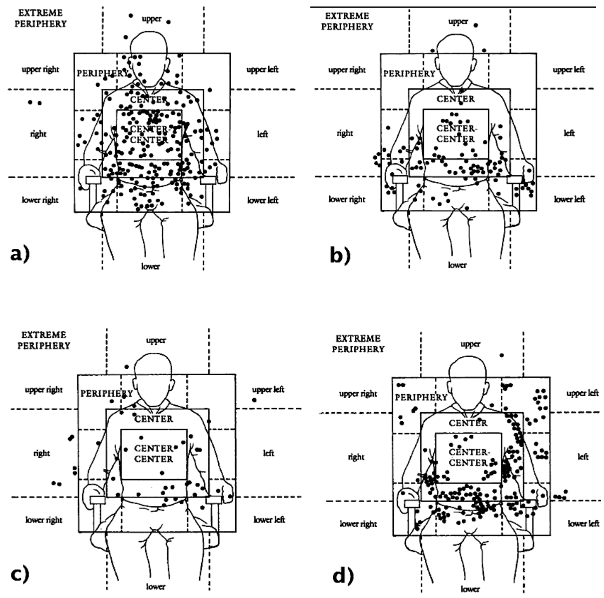

Fig 1. Spatial location and spatial density (dots) of gestures, according to McNeill's schematization, in four gesture types: a) iconics, b) metaphorics, c) deictics, d) beats (reproduced from McNeill, 1992:: 90-91).

III. Method and Phases of the System

Our application system consists of two different and separate phases: the training phase and the identification phase. In the training phase, the database containing the examples – here called templates – is built up through some basic modules, i.e. image acquisition, image preprocessing, features extraction and classification. These templates are used in the second phase for the automatic identification procedure which is based on template matching criteria. The basic features used in this work consist of a number of kinematic parameters, like the position of the centre of gravity of both hands, the speed of their movement, the angle of their movements, and the classification assigned to the template of that movement during the training phase.

As a preliminary stage, we used two videoclips showing two popular moto racers, Jorge Lorenzo and Valentino Rossi, recorded during an interview. These videos are freely available [10] [11]. The two excerpts last about 10 minutes and appear to be recorded under the same conditions. Moreover, in the videos the two subjects wear the same type of dress. In the paper, these two videos will be used as a reference for describing the system in some details.

A. *Image preprocessing and movements classification*

The steps which have been considered in this system consist of:

1. skin detection by means of color transformation
2. motion detection by means of background subtraction
3. cleaning up of the resulting image
4. contours extraction by means of edge detection.

The color transformation step aims to represent the color image according to the HSV (Hue, Saturation, Value) scale instead of the standard RGB (Red, Green, Blue). The reason for this conversion is due to the possibility of tuning the computer vision algorithm for each color channel of the image, thus avoiding the correlation effect of the components (Red, Green, Blue) with respect of the light intensity captured by the image. In the HSV representation, these components are uncorrelated each other. Here the following values have been experimentally determined for the three mentioned components: $0 > H < 20 \quad 30 > S < 150 \quad 80 > V < 255$. These values have demonstrated to better represent the skin characteristics of the two subjects. Result of the skin color detection algorithm is shown in Fig. 2, where the step for background subtraction has also been applied.

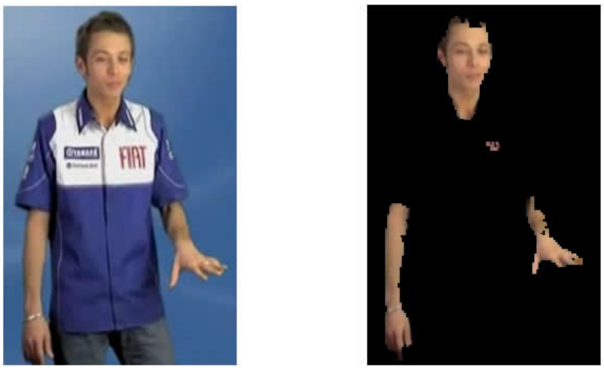

Fig. 2 Skin color detection and background subtraction

This procedure, of course, is not able alone to isolate the subject's hands from the rest, such as face and/or other possible noise of the image, but it eliminates all the static background of it, which is of no interest for the present application. On the other hand, this step allows considering also the head movements, which might be used in further applications.

It is worth noting that the isolation of hand from head movements can be obtained in a quite straightforward way by considering the coordinates of the contours extracted in the step 4 of this procedure, or by using some useful parameters provided by the OpenCv package, selecting only those contours of interest.

Since in this case we are looking for hand movements, a background subtraction allows isolating the parts of the body which have changed position with respect to the previous frame. Since people tend to move their hands more than their heads, in most of the cases such background subtraction will isolate only the hands from the rest of the image.

At this stage, the image has to be cleaned up for eliminating the noise still present in it, as it appears in Fig. 2. For this purpose, the standard algorithms of "blurring" and "smoothing" are applied. Fig. 3 shows the result of such filtering procedure, where a Gaussian filter and a subsequent threshold operation with a threshold value of 100 (experimentally determined) have been used.

Fig 3. Result from blurring and a threshold filtering of the image

The contours of the image portions of interest are obtained by using the edge detector algorithm proposed by Canny [12]. In our case, a rectangular contour has been considered. Such a contour shape has also the advantage of identifying the Centre Of Gravity (COG) of each hand as the centre of the rectangle. On the basis of COG identification, the relating kinematic parameters, such as speed and angle, can be easily computed.

The result of the edge detection is shown in Fig. 4. It is worth noting that the rectangular shape has the disadvantage of not being sensitive to the hand shape (position/orientation of the fingers), it has nevertheless the advantage of being independent from the hand shape. As a consequence, the information derived so far can be used to identify all gesture types realised by moving hands.

The morphology of the hand (closed vs. open fingers, for example), which can be discriminative for some types of gestures, will be taken into account in a subsequent template matching step described in subsection III.B. It is also important noting that, as shown in Fig. 4, the region detected by the mentioned algorithm does not select the hand but also the arm. This is due to the skin detection method previously described, which is of course not able to distinguish the hand from the arm since they are both characterised by the same skin color. However, this feature does not affect the hand movement classification scheme adopted here.

Using the kinematic information, the simple algorithm shown in Table I allows determining the position of each movement, according to the spatial location of hand gestures proposed by David McNeill and illustrated in section II.

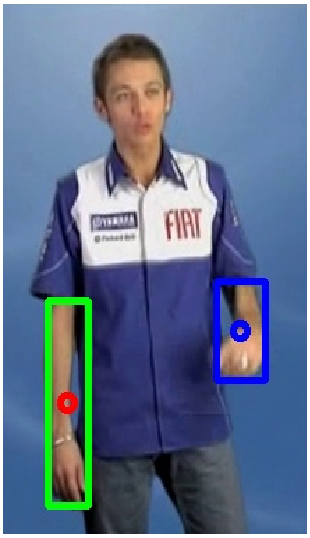

Fig 4. Extraction of contour and Centre Of Gravity of both hands

It is worth noting that hand movements are measured with respect to the rest position. In our case, this position corresponds to the subjects' hands in their pockets.

I. *Templates*

As previously mentioned, in this paper a novel aspect of the proposed system consists of assuming that the correct classification of the captured gestures depends also on the predetermined assignment of a class to a gesture prototype. This assignment task is performed through a training phase, and such a task allows the classification to be context- and subject-dependent.

In the training phase, a set of prototypical gestures are selected, classified and stored in a reference database. In order to save the needed computational time, a limited amount of features are extracted by each template, and they are used for the final automatic classification phase.

The features here adopted are the Hu moments [13] of the image.

Generally speaking, for each image a set of moments can be computed using the following definitions:

$$M_{ij}=\sum_x \sum_y x^i y^j I(x,y) \qquad (1)$$

and the normalised ones as:

$$M_{ij}^{norm}=\frac{M_{ij}}{\sum_x \sum_y I(x,y)} \qquad (2)$$

being *I (x,y)* the intensity of the pixel (x,y).

The moment M_{00} represents the area of the image, while the centroid of it corresponds to

$$(x_c, y_c) = (M_{10}/M_{00}, M_{01}/M_{00}) \qquad (3)$$

and may be assumed also as the Center Of Gravity (COG).

TABLE I.
ALGORITHM FOR DETERMINING THE POSITION OF THE MOVEMENTS

```
if abscissas _absolute_difference < 30
{
// if the present COG has the same abscissa
// of the previous one

        if ordinal_absolute_ difference < 30
        // if the present COG has the same ordinal
        // of the previous one

// the COG has not significativelly moved
direction = centre

        else if ordinal_difference > 0
        // the present COG is lower
        direction = to lower
        else if ordinal_difference < 0
        // the present COG is upper
        direction = to upper
}
else if abscisses_difference > 0
{
// if the present COG absisa is towards right

        if ordinal_absolute_difference < 30
        // if the present COG has the same ordinal
        direction = to right
        else direction = unidentified
        // non implemented direction
}
else if abscissa_difference < 0
{
// if the present COG abscissa is towards left

        if ordinal_absolute_difference < 30
        // if the present COG has the same ordinal
        direction = to left
        else direction = unidentified
}
else direction = unidentified
```

From these coordinates, the relative moments – also known as central moments – can be obtained, which are translation invariants:

$$\mu_{ij}=\sum_x \sum_y (x-y_c)^i (y-y_c)^j I(x,y) \qquad (4)$$

The Hu moments can be derived from the normalized central moments of the image, and the first seven of them have been demonstrated to be able to represent the features of an image, being invariants under different geometrical variations [14].

According to this set of features, four prototypical templates have been chosen for the examples presented in this paper. These templates represent all the prototypical gestures produced by the subjects under examination. For coding advantage, the four templates are numbered from 5 to 8, as shown in Fig. 5.

For tuning purpose of the classification algorithm, a number of counter-examples have also been considered, which correspond to prototypes of the unconscious gestures produced by the subject. In this way, an optimal value of false acceptance vs false rejection (error rate) can be obtained.

For the reported examples, these counter-examples are coded by numbering them from 0 to 4, as shown in Fig. 6. Here a typical case of counter-example is represented by the template 4, where the two hands are connected together, i.e. a hand movement which cannot be certainly considered as intentionally communicative.

The features (Hu moments) of all the coded gestures are included in the database and are used in the classification algorithm.

We have also been developing a user-friendly interactive procedure which allows the system user selecting the most representative frames, computing the feature, storing them in the database, and finding the most useful set of empirical parameters to be used during the recognition phase. The details of such procedure are beyond the scope of this paper and will not be discussed here any further.

B. The recognition phase

Each frame of the video under examination undergoes the image preprocessing steps described in Section III. A, the kinematic features are computed and the gesture spatial position is determined.

A moving window detects the regions of interest (in this case, hands and/or arms), computes the Hu moments of the image, and compares them with the stored templates information. Among the possible successful comparisons, the one having the minimum value of the Mahalanobis distance [15] is selected

The matching between any template and the examined gesture is checked by applying the Mahalanobis distance:

$$D_M(x)=\sqrt{(x-y)^T S^{-1}(x-y)} \qquad (5)$$

where S is the covariance matrix.

This distance is also known as generalized squared interpoint distance, because it is scale invariant and takes into account the correlations within the data. If it is close to zero, the two vectors are considered similar (or coincident), while they are not if the distance is greater than 1.

Table II shows the classification algorithm of the proposed system. It makes use of both the kinematic parameters computed in this phase (including the gesture spatial position) and the image features stored in the database. It this way, both the spatial classification proposed by McNeill and the template matching approach are taken into account, where the latter provides the needed context and individual variability for the classification.

Fig 5. Templates used for intentional gestures (numbered from 5 to 8)

Fig 6. Templates referring to unconscious gestures (numbered from 0 to 4)

TABLE II

THE CLASSIFICATION ALGORITHM

```
if ordinal_absolute_ difference < 10
      && abscissae_absolute_difference < 10
     // small movements are neglected
return "No hands movement ";


if num_template >= −1 && num_template < 5
     && position == undefined
return "Unconscious " ;

if num_template == 4 && position == center
   // templates numbered up to 5 represent
   //unconscious gestures
   // such as hand in a pocket or crossed arms
return " Unconscious " ;

if num_template == 5 && speed_absolute_difference < 10
    && position == previous_position
   // template related with united hands
   // speed lower than 10 pixel / frame
return "Iconic" ;


if num_template == 6
     && ( position == lower | | position == center )
return "Metaphoric" ;

if num_template == 7 && angle_absolute_difference <0.3
   // angles are measured in radiants
   // for small variations (0.3 radiants = 10 degrees )
   // the direction is the same
return "Deictics" ;

if num_template == 8
   && ( position == lower | | position == to right | | position == to left )
   && speed_absloute_diffrence > 10
   && angle_absolute_difference > 3 )
   // movements larger than 10 pixels / frame
   // 3 r adiants = about 180 degrees , means opposite
   // direction
return "Beat" ;

if no_previous_rule_valid
return "Gesture non recognised-non valid " ;
```

IV. RESULTS AND DISCUSSION

In order to test the performance of the system, both in terms of its robustness in classifying gestures and its generality with respect to the used templates, two test trials are here presented.

The first trial classifies the hand gestures produced by Jorge Lorenzo by making use of the templates related to the same subject.

The other one classifies the hand gestures produced by Valentino Rossi, by making use of the templates extracted from Jorge Lorenzo's video instead. As mentioned above, for each classification session a file is automatically produced which can be inspected and statistically analyzed for both evaluation of results and a possible further tuning of the system. The two videos last for about 10 minutes each, and a total of 543 gestures and 354 gestures were produced, respectively.

The prototype gestures used as templates were randomly selected among all hand gestures produced by the subject during the interview. Of course, the frames corresponding to these prototypes have been eliminated from the total amount of gestures analysed in test the procedure. Results of automatic classification are shown in Table III.

TABLE III

GESTURES CORRECTLY CLASSIFIED

	Deictic	**Iconic**	**Metaphoric**	**Beat**
Jorge Lorenzo (calibration)	99%	None	100%	85%
Valentino Rossi (test)	83%	None	100%	78%

Note that results for iconic gestures are due to the fact that, for this category, the Jorge Lorenzo had realised only one gesture (used as template), whereas Valentino Rossi had never produced iconic gestures during his interview.

As it was expected, the test in autocorrelation gives a better performance but results of the crossed test appears to be also encouraging.

In order to test the accuracy of the system in classifying gestures, we submitted the set of parameters measured in the recognition phase to a Neural Network, and to a clustering algorithm, namely a RBFN (Radial Basis Function Network) and K-mean clustering analysis.

The results of the RBFN model on the Jorge Lorenzo video are reported in Table IV, whereas Table V shows the results of the same process for the Valentino Rossi video.

Unfortunately the well-known K-means algorithm does not provide any figure which is able to give an estimation of the clustering goodness, since the number of clusters is an input parameter for the algorithm. We have performed several runs on the data, adopting a number of clusters spanning from 2 to 12, and found empirically that 5 clusters show the best compromise between the number of clusters and the population of each cluster. This result is confirmed also by the model analyzed by the RBF Network, as previously shown in Tables IV and V. Of course, in the latter case, a considerable amount of computational time is required.

We conclude therefore that the accuracy of the gesture classification produced by our proposed system is compatible with that coming from a Neural Network and a clustering algorithm

However, some considerations need to be pointed out.

First of all, the considered scenes. In the examples presented, the two subjects belong to a scene that never changes, they wear the same dresses and the illumination of the scene does not change during the video recording. This particular situation helps in solving most of the problems which are usually encountered in the automatic tracking of objects. This might appear as a limitation of the proposed approach. On the other hand, these ideal environmental characteristics can be commonly found in the video recordings of a conference speaker, i.e. the kind of application domain we are looking at.

TABLE IV

GESTURES CLASSIFIED BY RBFN (JORGE LORENZO)

```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances         443          81.5838 %
Incorrectly Classified Instances       100          18.4162 %
Kappa statistic                          0.6924
Mean absolute error                      0.1041
Root mean squared error                  0.2261
Relative absolute error                 41.7695 %
Root relative squared error             64.1395 %
Total number of Instances              543

=== Detailed Accuracy By Class ===

TP Rate    FP Rate   Precision   Recall   F-Measure   Class
  0.559      0.055      0.755     0.559      0.643     Deictic
  0.903      0.272      0.804     0.903      0.851     Unconscious
  0.785      0.021      0.836     0.785      0.81      Beat
  0.958      0.002      0.958     0.958      0.958     Unrecognized/Unconscious
  1          0          1         1          1         Metaphoric

=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
  71  53   2   1   0 |   a =   Deictic
  22 271   7   0   0 |   b =   Unconscious
   1  13  51   0   0 |   c =   Beat
   0   0   1  23   0 |   d =   Unrecognized/Unconscious
   0   0   0   0  27 |   e =   Metaphoric
```

TABLE V

GESTURES CLASSIFIED BY RBFN (VALENTINO ROSSI)

```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances         348          98.3051 %
Incorrectly Classified Instances         6           1.6949 %
Kappa statistic                          0.949
Mean absolute error                      0.0076
Root mean squared error                  0.0588
Relative absolute error                  5.8156 %
Root relative squared error             23.2732 %
Total number of Instances              354

=== Detailed Accuracy By Class ===

TP Rate    FP Rate   Precision   Recall   F-Measure   Class
  0.979      0          1         0.979      0.99      Unconscious
  1          0          1         1          1         Unrecognized/Unconscious
  1          0          1         1          1         Beat
  1          0.018      0.76      1          0.864     Deictic
  1          0          1         1          1         Metaphoric

=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
 284   0   0   6   0 |   a =   Unconscious
   0  14   0   0   0 |   b =   Unrecognized/Unconscious
   0   0  25   0   0 |   c =   Beat
   0   0   0  19   0 |   d =   Deictic
   0   0   0   0   6 |   e =   Metaphoric
```

Secondly, the approach here proposed strongly relies on the availability of suitable templates which describe the morphology of the gestures. However, the reported examples demonstrate that only a limited amount of templates is needed: 9 templates (4 positive and 5 negative), meaning about 3% of the total amount of gestures to be examined. Moreover, this methodology guarantees the adherence of the classification to the individual subject, and the few templates selected during the calibration phase perform well also for similar environmental conditions. Even the possibility of having a larger set of templates does not affect the performance of the system, since the comparison is made among few parameters (the Hu moments) and does not require a heavy computation time.

We are aware that our system does not solve the general problem of automatically classifying human gestures from a video. On the other hand, we are confident that our simple system can be useful for some specific applications, where its limited computational time can be more effective than more sophisticated, resource-consuming and, in any case, not always highly perfoming systems

REFERENCES

[1] A. Kendon, “Some relationships between body motion and speech. An analysis of an example”, in A.Siegman & B.Pope (Eds.) Studies in Dyadic Communication, Elmsford, Pergamon Press, New York, 1972, pp. 172-210.

[2] D. McNeill, “Hand and mind: what gestures reveal about thought”, University of Chicago Press Chicago, 1992

[3] Rui Ping Shi, Johann Adelhardt, Anton Batliner, Carmen Frank, Elmar Noeth, Viktor Zeissler, Heinrich Niemann, “The Gesture Interpretation Module”, SMARTKOM: Foundations of Multimodal dialog systems, W. Wahlster Ed., Springer, Berlin 2006, pp.209-219

[4] Yingen Xiong and Francis Quek, “Hand Motion Gesture Frequency Properties and Multimodal Discourse Analysis”, International Journal of Computer Vision 69(3), 353–371, 2006.

[5] Andrew D, Wilson & Aaron F. Robick, “Parametric Hidden Markov Models for Gesture Recognition”, IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 21, no. 9, September 1999, pp.884-900

[6] Y. Yasinnik, M. Renwick, S. Shattuck-Hufnagel, “The timing of speech-accompanying gestures with respect to prosody”, in Proc. of the International Conference “From Sound to Sense”, MIT, Cambridge, Mass, June 10-13, C97-C102, 2004.

[7] M. Savino, L. Scivetti, M. Refice, “Integrating Audio and Visual Information for Modelling Communicative Behaviours Perceived as Different”, in Proceedings of Language Resources and Evaluation Conference (LREC 2008), Marrakesh 28-30 May 2008 (on CD-ROM).

[8] A. Esposito, D. Esposito, M. Refice, M. Savino, S. Shattuck-Hufnagel, “A Preliminary Investigation of the Relationship Between Gesture and Prosody in Italian”, in A. Esposito, M. Bratanic, E- Keller, M. Marinaro (eds), Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue, IOS Press, NATO Security through Science Series, Amsterdam, 2007, pp. 65-74.

[9] OpenCV : http//opencv.willowgarage.com/wiki/

[10] Jorge Lorenzo video clip : http//www.youtube.com/watch?v=DJLDn4MOF2Y

[11] Valentino Rossi video clip: http//www.youtube.com/watch?v=d3C8VWrmkcc

[12] J. Canny, “A computational approach to edge detection”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, pp. 679-714, 1986.

[13] M.K. Hu, “Visual Pattern Recognition by Moments Invariants” – RE Transaction in Information Theory, vol. IT-8, pp.179-187, 1962.

[14] A. Papoulis, *Probability, random variables, and stochastic processes*, McGraw-Hill, 1991.

[15] P.C. Mahalanobis, “On the generalized distance in statistics”, *Proceedings of the National Institute of Sciences of India* **2** (1): 49–55, 1936.