Facial Expression Recognition Based on the Belief Theory: Comparison with Different Classifiers

Z. Hammal[1], L. Couvreur[2], A. Caplier[1], and M. Rombaut[1]

[1] Laboratory of images and signals LIS,
46 avenue Félix Viallet, F-38031 Grenoble, France
[2] Signal Processing Laboratory, Faculté Polytechnique de Mons,
1 Avenue Copernic, B-7000, Mons, Belgium

Abstract. This paper presents a system for classifying facial expressions based on a data fusion process relying on the Belief Theory (BeT). Four expressions are considered: *joy, surprise, disgust* as well as *neutral*. The proposed system is able to take into account intrinsic doubt about emotion in the recognition process and to handle the fact that each person has his/her own maximal intensity of displaying a particular facial expression. To demonstrate the suitability of our approach for facial expression classification, we compare it with two other standard approaches: the Bayesian Theory (BaT) and the Hidden Markov Models (HMM). The three classification systems use characteristic distances measuring the deformations of facial skeletons. These skeletons result from a contour segmentation of facial permanent features (mouth, eyes and eyebrows). The performances of the classification systems are tested on the Hammal-Caplier database [1] and it is shown that the BeT classifier outperforms both the BaT and HMM classifiers for the considered application.

1 Introduction

The human-machine interface (HMI) is definitively evolving to an intelligent multimodal interface, combining various human communication modes. Among others, facial expression is a very efficient mean for human beings to communicate their intention.

In this work, we propose a rule-based system for automatically classifying facial expressions. This system relies on the Belief Theory (BeT). Like other methods [2,3,4], our approach is based on facial deformation features (eyes, eyebrows and mouth). It allows to deal with uncertain data and recognize facial expressions in the presence of intrinsic doubt. Clearly, humans do not behave in a *binary* way: they do not produce *pure* expressions but rather combinations of them. Our system is able to identify either pure expressions as well as mixed ones. In order to demonstrate the efficiency of BeT system for the purpose of facial expression recognition, its performances are compared with those of more classical approaches, namely Bayesian Theory (BaT) and the Hidden Markov Models (HMMs).

Section 2 presents how video data are preliminary processed in order to recognize facial expression. In section 3 we describe the Belief Theory classifier, in section 4 the Bayesian classifier and in section 5 the HMM classifier. Section 6 describes the

F. Roli and S. Vitulano (Eds.): ICIAP 2005, LNCS 3617, pp. 743 – 752, 2005.

video database used in this work and presents a comparison between the performances of the three classifiers.

2 Facial Expression Analysis

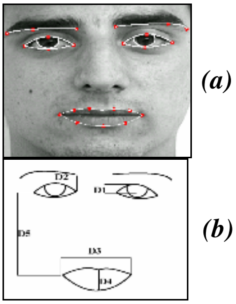

Fig. 1. a) facial features segmentation; b) facial skeleton and characteristic distances.

In this section, we describe how a video sequence of face images is analysed for recognition of facial expressions. First, the contours of facial features are automatically extracted in every frame using the algorithms described in [5,6] (Fig. 1.a).

Next, five characteristic distances are defined and estimated (Fig. 1.b): eye opening (D_1), distance between the inner corner of the eye and the corresponding corner of the eyebrow (D_2), mouth opening width (D_3), mouth opening height (D_4), distance between a mouth corner and the outer corner of the corresponding eye. These distances form together a characteristic vector associated to each facial expression and can be used for modeling and recognizing facial expressions.

The BeT recognition process involving all the distances D_i yields to a classification of every frame in the video sequence in terms of a single expression or a mixture of expressions. A state of doubt between two expressions can appear. In order to cope with it, a post processing based on the analysis of transient wrinkles in the nasal root and based on the analysis of the mouth shape is added.

The presence or absence of wrinkles in the nasal root (Fig. 2.a) is detected by using a Canny edge detector. If there is about twice more edges points in the nasal root of the current frame than in the nasal root of a frame with the *neutral* expression, the presence of transient wrinkles is validated, and discarded otherwise. The mouth shape is also used (Fig. 2.b, 2.c). According to the expression, the ratio between length and width of the mouth is larger or smaller than its corresponding value for the *neutral* expression.

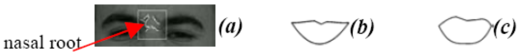

Fig. 2. (a) wrinkles in the nasal root, examples of mouth shapes in case of : (b) *joy*, (c) *disgust*

3 Classification by the Belief Theory

3.1 Definition of the Symbolic States

As shown from our expertise database [1], every characteristic distance D_i can be higher, lower or roughly equal to its value $D_{i,neutral}$ defined for the *neutral* expression, whatever the actual face expression. This comes naturally to the definition of three symbolic states:

- the *higher* state C^+ if D_i is significantly higher than $D_{i,neutral}$;
- the *lower* state C^- if D_i is significantly lower than $D_{i,neutral}$;
- the *neutral* state S if D_i is close to $D_{i,neutral}$.

Hence, one can identify the symbolic state of every characteristic distance for a given face image and analyse its time evolution along a video sequence. Fig. 3 presents the evolution of D_2 (distance between the interior corner of the eye and the interior corner of the eyebrow) and D_5 (distance between one mouth corner and the external corner of the corresponding eye) for several persons and for a given expression. In each case, the video sequence starts with a *neutral* expression, goes towards the actual expression and returns to the neutral expression. Clearly, we observe similar time evolutions of the characteristic distance, thereof the corresponding symbolic state, whatever the subject. This observation also holds for the other characteristic distances and facial expressions.

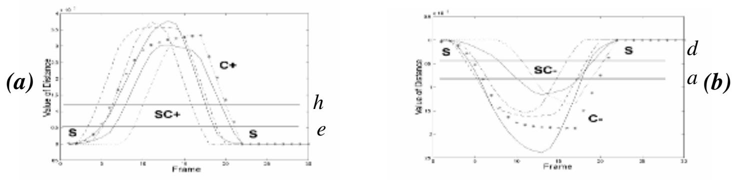

Fig. 3. Examples of time evolution of characteristic distances (a) D_2 and (b) D_5 in case of *surprise* and *joy* face expression, respectively. Thresholds *h*, *e*, *d* et *a* are defined in section 3.3.

3.2 The Belief Theory

Originally introduced by Dempster and Shafer and revisited by Smets [7], the Belief Theory (BeT) can be seen as a generalization of the probability theory. It requires the definition of a set $\Omega = \{E_1, E_2, \ldots, E_N\}$ of N exclusive and exhaustive assumptions. We also consider the power set 2^{Ω} that denotes the set of all subsets of Ω. To each element A of 2^{Ω} is associated an elementary piece of evidence *m(A)* which indicates all confidence that one can have in this proposal. The function *m* is defined as:

$$m: \quad 2^{\Omega} \rightarrow [0,1] \qquad (1)$$

$$A \mapsto m(A) \quad \text{with} \quad \sum_{A \subseteq \Omega} m(A) = 1$$

In our application, the assumptions E_i correspond to the four facial expressions : *joy* (E_1), *surprise* (E_2), *disgust* (E_3) and *neutral* (E_4); 2^{Ω} corresponds to single expressions or combinations of expressions, that is, $2^{\Omega} = \{E_1, E_2, E_3, \ldots, E_1 E_2, E_2 E_3, \ldots\}$, and A is one of its elements.

3.3 Modelling Process

The modelling process aims at computing the state of every distance D_i and at associating a piece of evidence. Let define the basic belief assignment (BBA) m_{Di} as:

$$m_{Di} : \quad 2^{\Omega'} \rightarrow [0,1] \qquad (2)$$
$$B \mapsto m_{Di}(B)$$

With $\Omega' = \{C^+, C^-, S\}$, $2^{\Omega'} = \{C^+, C^-, S, SC^+, SC^-, C^-SC^+\}$, where $S \cup C^+$ (noted SC^+) states the doubt between S and $C+$, $S \cup C^-$ (noted SC^-) states the doubt between S and C^- and $m_{Di}(B)$ is the piece of evidence (PE) of each state B.

A numerical/symbolic conversion is carried out, which associates to each value of *Di* one of the symbols of $2^{\Omega'}$. To carry out this conversion, we defined a model for each distance using the states of $2^{\Omega'}$ (Fig. 4). We assume that the symbol C^-SC^+ is impossible.

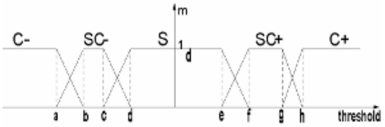

Fig. 4. Model of distances states

In Fig.4., *m* is the PE associated to each possible state in $2^{\Omega'}$ and the thresholds *(a... h)* are the limit values of D_i corresponding to each state or subset of states. For each distance D_i, the threshold ***h*** (resp. ***a***) of the state C^+ (resp. C^-) corresponds to the average of the maximum (resp. minimal) values of D_i for all the subjects and all the expressions of the expertise database. The thresholds ***d*** and ***e*** of the state *S* are defined in the same way.

The median of the maximum values of each distance for all the subjects and all the expressions of the expertise database is computed. The thresholds ***f***, ***b*** *(*resp. ***c, g***) of the intermediate states are defined by mean+median (resp. mean-median) of each state *(*C^+, C^-, *S).*

3.4 Recognition Process

Analysis. The analysis of the states for the five distances associated to each of the four expressions (*joy, surprise, disgust* and *neutral*) allows us to exhibit for each expression a specific combination of these states. Table 1 shows the resulting states combinations.

For example, in case of *joy* (E_1), the mouth is opening (C^+ state for D_3 and D_4), the corners of the mouth are going back toward the tears (C^- state for D_5) and the eyebrows are slackened (*S* state for D_2). The distance between the interior corner of the eye and the interior corner of the eyebrow decreases (C^- state for D_2) and the eyes become slightly closed (C^- state for D_1).

Table 1. Theoretical table of D_i states for each expression

	D_1	D_2	D_3	D_4	D_5
Joy E_1	C^-	S / C^-	C^+	C^+	C^-
Surprise	C^+	C^+	C^-	C^+	C^+
Disgust	C^-	C^-	S / C^+	C^+	S / C^-
Neutral	S	S	S	S	S

The proposed combinations of symbolic states are similar to the MPEG-4 [8] description of the deformations undergone by facial features for such expressions , yet they give some extensions.

Note that in some cases, two different states are possible for a given distance (for example, see D_2 for *joy*, D_3 for *disgust*). This can lead to doubt between two expressions. For example, the classifier is not always able to distinguish *disgust* and *joy* because both expressions can be described by the same combination of states.

The expression E_5 is added as the *unknown* expression or rejection class. It represents all the expressions that do not correspond to any of the state combination in Table 1.

Combination and Decision. We have several sources of information ,namely the distances D_i, to which we associate PEs. Our goal is to obtain a PE which takes into account all the available information. The BBA is obtained using the rule of conjunctive combination or orthogonal sum. In the case of two distances D_1 and D_2, the orthogonal sum is defined in the following way:

$$m = m_{D_1} \oplus m_{D_2} \tag{3}$$

$$m(A)=\sum_{B\cap C=A} m_{D1}(B) m_{D2}(C) \tag{4}$$

A, B and C are expressions or subsets of expression.

This allows to obtain propositions whose number of elements is lower than the initial ones and to associate them a piece of evidence. The final PE is thus more accurate. More explicitly, if one takes two basic belief assignments: $m_{D1}(E_1 \cup E_3)$ $m_{D1}(E_1)$ $m_{D1}(E_2)$

$m_{D2}(E_1)$ $m_{D2}(E_2)$ $m_{D2}(E_1 \cup E_2)$

their combination gives the results of Table 2. The piece of evidence of each expression by the combination of results of the two distances is calculated by:

Table 2. Example of combination of PEs of two distances. ∅ is the empty set

D_1 / D_2	E_1	E_2	$E_1 \cup E_3$
$E_2 \cup E_3$	∅	E_2	E_3
E_1	E_1	∅	E_1
E_2	∅	E_2	∅

$m_{D12}(E_1) = m_{D1}(E_1) .\ m_{D2}(E_1) + m_{D1}(E_1)\ m_{D2}(E_1 \cup E_3)$,
$m_{D12}(E_2) = m_{D1}(E_2 \cup E_3) .\ m_{D2}(E_2) + m_{D1}(E_2) . m_{D2}(E_2)$,
$m_{D12}(E_3) = m_{D1}(E_2 \cup E_3) . m_{D2}(E_1 \cup E_3)$,
$m_{D12}(\varnothing) = m_{D1}(E_2 \cup E_3) . m_{D2}(E_1) + m_{D1}(E_1) . m_{D2}(E_2) + m_{D1}(E_2) . m_{D2}(E_1) + m_{D1}(E_2) .\ m_{D2}(E_1 \cup E_3)$.

Conflicts, noted ∅, can appear in case of incoherent sources. In the scope of the presented application, the conflict corresponds to a configuration of distance states which does not appear in Table 1. It comes from the fact that Ω is not exhaustive. The additional *unknown* expression or class of reject E_5 represents all these conflicts states (Table 2).

The decision is the ultimate step of the classification process. It consists in making a choice between various assumptions E_i and their possible combinations. Making a choice means taking a risk, except if the result of the combination is perfectly reliable: $m(E_i)=1$. Here, the accepted proposal is the one with maximum value of PE.

4 Bayesian Classifier

In this work, the data and the classes of the Bayesian classifier consist in the distance vectors and the facial expressions, respectively. The statistical models aim at modelling the probability density functions of the observation data for every class.

Here, the probability density functions are defined as mixtures of 3 Gaussian components $N(\mu_{k,},\Sigma_k)$:

$$p(x|y)=\sum_{k=1}^{3} w_k N(\mu_k,\Sigma_k) \tag{5}$$

where x and y denote the distance vector and the facial expression class, respectively. The parameters of these models, *i.e.* the mean vectors μ_k, the covariance matrices Σ_k and the mixing weights w_k, are estimated in a Maximum Likelihood (ML) sense independently for every class. Since this problem is a missing data problem, it can be addressed by the Expectation-Maximization (EM) algorithm [9] .

During recognition experiments, a distance vector is derived for every frame. Consecutive distance vectors are assumed to be statistically independent as well as the underlying class sequences. This vector is presented to each mixture of Gaussians and its likelihood score is computed. The vector is eventually assigned to the class corresponding to the mixtures of Gaussians with the highest likelihood score:

$$\tilde{y}=\underset{y}{\operatorname{argmax}}\, p(x|y) \quad \mathrm{y} \in \{\mathrm{E}_1, \mathrm{E}_2, \mathrm{E}_3, \mathrm{E}_4, \mathrm{E}_5\} \tag{6}$$

5 HMM Classifier

Hidden Markov Models (HMM) are widely used in many fields (*e.g.*, automatic speech recognition) [10] where temporal (or spatial) dependencies are present in the data.

For the recognition of facial expressions, we adopt a 5-state HMM, one state per expression, whose topology is depicted in Fig. 5. As can be seen, the HMM forces the state sequence to start in the *neutral* state (E_4), then can stay some times in either the *joy* state (E_1), the *surprise* state (E_2) or the *disgust* state (E_3). An *unknown* state (E_5) is also considered for representing any other expression. Such topology is practically realized by forcing some transition probabilities to zero.

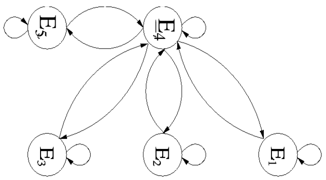

Fig. 5. HMM classifier topology. Branches between the first E_4 and the last E_4 should not be bi-directional. Besides, there should be only one E_4 state to which you loop back after leaving an expression state.

The state probability density functions are defined as mixtures of 3 Gaussian components. All the HMM parameters are estimated by an EM-style algorithm. Given some observation data, *i.e.* sequences of vectors of characteristics distances, and an initial estimate, the HMM parameters are refined using the Baum-Welch algorithm [10]. Note that its Viterbi approximation can be used as well.

During the recognition experiments, a sequence of distance vectors is presented to the HMM. The most likely state sequence is searched by a Viterbi algorithm [10]. Hence, a state is assigned to every distance vector, equivalently an expression is assigned to every frame. Unlike for the Bayesian classifier, the state choice is not taken independently at each time instant but rather globally it is assumed that there exists some dependency between consecutive face classes. Such criterion allows to recover many errors of the Bayesian classifier.

6 Results and Discussion

6.1 Database

The Hammal-Caplier database is used for our experiments (21 subjects and 4 expressions) [1]. Each video recording starts and ends with a *neutral* state (for example Fig. 6). The sequences have been acquired during 5 seconds at 25 images/second.

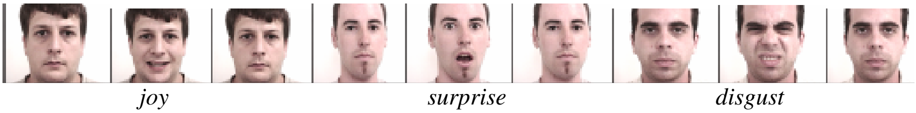

Fig. 6. Examples of expressions. Each record starts and finishes with a neutral state

For the expertise step of the BeT, 1170 frames (13 subjects and 4 expressions) of the Hammal-Caplier expertise database have been considered. All the frames of the expertise database are segmented and the five distances defined on Fig. 1 are computed and used in order to define the thresholds of section 3.3 and to build Table 1.

In order to evaluate the robustness to different variations (gender, ethnicity, difference of expressions ,etc), the BeT system is tested on the Hammal-Caplier test database (630 frames for 8 subjects and 4 expressions).

For the HMM and Bayesian classifiers all the data of Hammal-Caplier database are used for the training step and the test is carried out by a 21-fold cross validation. It consists in taking the data from 20 out of 21 subjects for the training step and in using the data of the remaining subject for the test step. This process is repeated 21 times, considering a different test subject each time. The classification rate is the average over 21 results (Table 4).

6.2 Results

Results of Belief Theory Classification. Table 3 presents the classification rates for the frames of the Hammal-Caplier test database. The correct expressions and the recognized expressions are given in the first column and the first row, respectively.

Expressions E_1 (*joy*), E_2 (*surprise*) and E_4 (*neutral*) yield to good classification rates. On the contrary, the classification rate E_3 *(disgust*) is lower. This is due to individual variability (Fig. 7.a) and to the difficulty for a non actor people to simulate this expression (Fig. 7.b).

For E_1, there is a high rate of total doubt between E_1 and E_3 : the system is sure that it is one of the two expressions but is not able to know which one. This has to be related to the definition of Table 1 with two possible different states for a given distance. In the Hammal-Caplier database, the *unknown* state E_5 often appears for intermediate frames where the person is neither in a *neutral* state, nor in a particular expression (Fig.7.c).

In order to choose between *joy* and *disgust* in case of doubt, we add a post-processing state which takes into account information about transient features and mouth shape (Sect 2). Nasal root wrinkles (Fig. 2.a) are characteristic for *disgust*. This is used to solve the problem of doubt between *joy* and *disgust*. In the case of absence of transient features, we use the ratio between length and width of the mouth (Fig. 2.b, 2.c). Our analysis shows that this ratio is larger than its value for the *neutral* expression in the case of *joy* and lower in the case of *disgust*. With the proposed post-processing step the recognition rate for E_1 (*joy*) increases by 15% and $E_1 \cup E_3$ (*joy-disgust*) decreases by 17% (2% of false detection of *disgust*). We increase by 19% for E_3 (*disgust*) and $E_1 \cup E_3$ (*joy-disgust*) decreases by 11% (5% of false detection of *joy*).

Table 3. Classification rates on the Hammal-Caplier database

Syst\Exp	E_1	E_2	E_3	E_4
E_1 joy	76.36	0	9.48	3
E_2 surprise	0	84.44	0	0
E_3 disgust	0	0	43.10	2
$E_1 \cup E_3$	10.90	0	8,62	0
E_4 neutral	6.66	0.78	15.51	88
E_5 unknown	6.06	11.8	12.06	0
Total	**87.26**	**84.44**	**51.2**	**88**

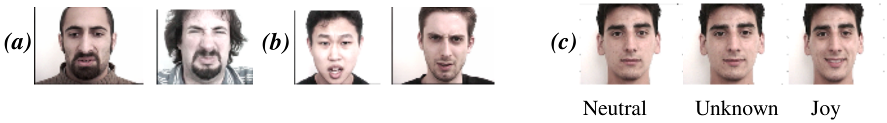

Fig. 7. Examples of *disgust* expressions : (a) individual variability; (b) poor simulations. (c): Example of *unknown* state: 3 consecutive frames from *neutral* to *joy*.

Given the fact that the state of doubt *joy-disgust* is related to the rules defined in the Table 1, it is not due to classification errors of the proposed system. It is thus possible to consider it as a good classification and to associate it to the corresponding expression which allows us to add their respecting rates leading to the results of the last row of Table 3.

Results of Bayesian Theory and HMM. Classification rates of Bayesian classifier are lower than those of belief theory classifier (Table 4 left). The best results are those of the *neutral* expression. A very low rate of classification is noted on the whole set of expressions. This is due on the one hand to the fact that the Bayesian classifier assumes specific form of the statistical distributions of the classes, which may be a wrong assumption for our dataset and on the other hand to the lack of training data.

The classification rates of the HMM are comparable with those of the belief theory (Table 4 right). Similarly the classification rates of *disgust* are better than those of *joy*, *surprise* and *neutral*.

To model the *unknown* expression used in the BeT for the HMM and the Bayesian classifiers, we introduce an "*unknown state*" which gathers all the expressions that correspond to a set of configurations of distance states learned by the two systems as being *unknown,* contrary to the belief theory where the *unknown* expression corresponds to all the configuration of distances states unknown to the system. In other terms this is another finite set of facial expressions added to the four already defined ones and so does not contain all the possible facial configurations which can lead to classification errors. This is not the case for the belief theory which directly affects new configurations at *unknown* expression.

Table 4. Classification rates on Hammal-Caplier database. left: Bayesian classifier; right HMM classifier.

	Bayesian					HMM				
Syst\Exp	Joy	Surprise	disgust	neutral	unknown	joy	Surprise	disgust	neutral	unknown
E_1 joy	37.71	3.80	21.86	5.46	23.96	78.87	0	2.02	3.28	14.45
E_2 surprise	22.27	50.43	3.79	4.94	19.17	0	79.81	0	6.36	21.31
E_3 disgust	4.33	10.16	25.43	5.20	12.79	6.82	0	49.39	3.60	28.66
E_4neutral	7.62	20.47	2.21	79.85	24.56	4.48	10.75	6.37	75.25	25.84
E_5unknown	28.06	15.12	46.69	4.53	19.50	9.82	9.43	42.21	11.49	9.72
Total	**37.71**	**50.43**	**25.43**	**79.85**	**19.50**	**78.87**	**79.81**	**49.39**	**75.25**	**9.72**

The classification results of the three classifiers on the same data (characteristic distances) shows that the better results are obtained with the classifier based on BeT. In addition to its capacity of generalization, the use of the BeT emphasizes the fact that some expressions are not always dissociable (*joy* and *disgust*) and allows to recognize a mixture of facial expressions contrary to the HMM or Bayesian classifiers. For all these reasons we conclude that the BeT is better adapted to the problem of facial expressions recognition.

7 Conclusion

We present a method for classification of facial expressions based on the analysis of characteristic distances computed on skeletons of expression. The results of comparison with Bayesian Theory and HMM show that the best classification rates are those of the Belief Theory. To improve the results, we can increase the number and the quality of measurements, by taking into account the explicit information about the forms of contours of the skeletons of expression in addition to the characteristic distances and by taking into account the temporal evolution of measurements.

References

1. http://www.lis.inpg.fr/pages_perso/hammal/index.htm
2. Cohn, J., Zlochower, A.J., Lien, J.J., Kanade, T.: Feature-point tracking by optical flow discriminates subtle differences in facial expression. IEEE ICFGR, N°3, (1998) 396–401.
3. Pantic, M., Rothkrantz, L.J.M.: Expert system for automatic analysis of facial expressions. IVC Vol.118. (2000) 881–2000.

4. Cohen, I., Cozman, F.G., Sebe, N., Cirelo, M.C., Huang, T.S.: Semi supervised Learning of Classifiers : Theory, Algorithms and their Application to Human-Computer Interaction“, IEEE Trans. On PAMI, Vol.26, (2004) 1553–1567.
5. Eveno, N., Caplier, A., Coulon, P.Y.: Automatic and Accurate Lip Tracking. IEEE Trans. On CSVT, Vol. 14. (2004) 706–715.
6. Hammal, Z., Caplier, A. : Eye and Eyebrow Parametric Models for Automatic Segmentation. IEEE SSIAI, Lake Tahoe, Nevada (2004).
7. Smets, PH.: Data Fusion in the Transferable Belief Model. Proc. ISIF, France (2000) 21–33.
8. Malciu, M., Preteux, F.: MPEG-4 Compliant Tracking of Facial Features in Video Sequences. Proc. EUROIMAGE, ICAV3D, Greece (2001) 108–111.
9. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE, Vol.77. no.2, February (1989) 257–286.
10. Moon, T.K., Stirling, W.C.: Mathematical Methods and Algorithms for Signal Processing, Prentice-Hall, (2000).