Fuzzy Rule-Based Hand Gesture Recognition*

Benjamín C. Bedregal[1], Antônio C. R. Costa[2] and Graçaliz P. Dimuro[2]

[1] Department of Informatics and Applied Mathematics
Federal University of Rio Grande do Norte, Brazil
[2] Informatics School, Graduate Programme in Computer Science
Catholic University of Pelotas, Brazil
bedregal@dimap.ufrn.br, {rocha,liz}@ucpel.tche.br

Abstract. This paper introduces a fuzzy rule-based method for the recognition of hand gestures acquired from a data glove, with an application to the recognition of some sample hand gestures of LIBRAS, the Brazilian Sign Language. The method uses the set of angles of finger joints for the classification of hand configurations, and classifications of segments of hand gestures for recognizing gestures. The segmentation of gestures is based on the concept of monotonic gesture segment, sequences of hand configurations in which the variations of the angles of the finger joints have the same sign (non-increasing or non-decreasing). Each gesture is characterized by its list of monotonic segments. The set of all lists of segments of a given set of gestures determine a set of finite automata, which are able to recognize every such gesture.

1 Introduction

Fuzzy set theory [1] is the oldest and most widely reported component of present-day soft computing (or computational intelligence), which deals with the design of flexible information processing systems [2], with applications in control systems [3], decision making [4], expert systems [5] etc. The significance of fuzzy set theory in the realm of pattern recognition was justified in [2].

A fuzzy system encompasses the implementation of a (usually nonlinear) function, defined by a linguistic description of the relationship between its input variables. Standard fuzzy systems presents an architecture such as the one depitecd in Fig. 1. The *fuzzificator* is the component that computes the membership degrees of the crisp input values to the linguistic terms (fuzzy sets) associated to each input linguistic variable. The *rule base* is composed by inference rules associating linguistic terms of input linguistic variables to linguistic terms of output linguistic values. The *information manager* is the component for searching in the rule base the adequate rules to be applied for the current input. The *inference machine* gives the membership degrees of the output values in the output sets, by the application of the rules selected in the rule base.

* This work is partially supported by FAPERGS and CNPq (Proc. 470871/2004-0, Proc. 470556/2004-8).

Finally, the *defuzzificator* determines a single output value as a function of the output values and their membership degrees to the output sets.

We remark, however, that there are many approximate methods (e.g., classification or pattern recognition procedures) that do not produce a single final result. On the contrary, they may give several alternative solutions to a single problem (e.g., the different classes to which a given input may belong). Examples of such methods are several fuzzy methods for pattern recognition [2], such as fuzzy relations, fuzzy clustering, fuzzy neural systems etc. [6], with applications to signature verification [7], and face recognition [8], for example. Thus, for some specific applications, it is reasonable to consider a fuzzy rule based method, which determines a system architecture as shown in Fig. 2.

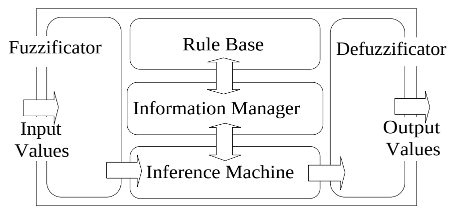

Fig. 1. Architecture of standard fuzzy systems.

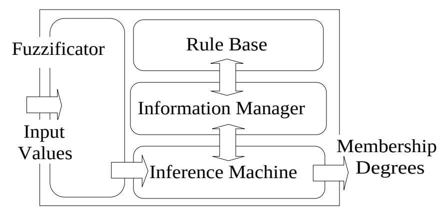

Fig. 2. Architecture of a fuzzy rule based system.

It is possible to find an extensive literature about methods and systems for gesture recognition in general, and hand gesture recognition in particular. There are systems for the recognition of 3-D and 2-D gestures captured by different devices (data gloves, cameras etc.) [9], systems for the graphical recognition of traces left on tablet devices [10] etc. Among several methods for gesture recognition, there are methods based on fuzzy logic and fuzzy sets, methods based on neural networks, hybrid neuro-fuzzy methods [11], fuzzy rule [12] and finite state machine [13] based methods, methods based on hidden Markov models [14] etc. In particular, considering methods for sign language recognition, some literature can be found related to fuzzy methods, such as, for example, fuzzy decision trees [15] and neuro-fuzzy systems [16].

In this paper, we propose a fuzzy rule-based method for the recognition of hand configurations and hand gestures acquired from a data glove, with an application to the recognition of some sample hand gestures of LIBRAS, the Brazilian Sign Language [17]. The method uses the set of angles of finger joints for the classification of hand configurations, and classifications of sequences of hand configurations for recognizing gestures. The segmentation of gestures is based on the concept of *monotonic gesture segment*, sequences of gestures in which the variations of the angles of the finger joints have the same sign (non-increasing or non-decreasing).

Any monotonic gesture segment is characterized by an initial hand configuration, a terminal hand configuration and a list of relevant intermediate configurations. Each gesture is characterized by a list of monotonic segments. That set of lists of segments determine a set of finite automata, which are able to recognize the gestures being considered.

The paper is organized as follows. In Sect. 2, we introduce our fuzzy rule-based method for hand gesture recognition. A case study is discussed in Sect. 3, with the recognition of LIBRAS hand gestures. Section 4 is the Conclusion.

2 The Fuzzy Rule Based for Hand Gesture Recognition

The objective is to recognize some hand gestures with data obtained from a data glove. Consider a hypothetical data glove with 15 sensors, as shown in Fig. 3. The fingers are labelled as: F1 (little finger), F2 (ring finger), F3 (middle finger), F4 (index finger) and F5 (thumb). The joints in the fingers are labelled as J1 (the knuckle), J2 and J3, for each finger. A separation between two fingers is labelled as Sij to indicate that it is a separation between the fingers Fi and Fj.

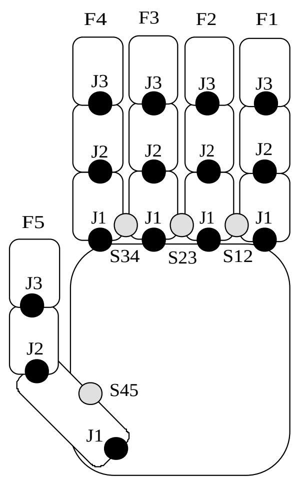

Fig. 3. Localization of sensors in the data glove.

Since any movement can be represented as a sequence of frames, a hand movement using a data glove is represented as a sequence of hand configurations, one for each discrete time instant. That is, at each time instant, the data glove sensors should provide the set of angles of joints and finger separation that characterizes a hand configuration.

In order to simulate this data transfer, a random generator of hand configurations was implemented, generating at each instant one hand configuration represented by a tuple of angles corresponding to each sensor shown in Fig. 3:

```
( (F1J1,F1J2,F1J3), S12, (F2J1,F2J2,F2J3), S23, (F3J1,F3J2,F3J3), S34,
  (F4J1,F4J2,F4J3), S45, (F5J1,F5J2,F5J3) )
```

Given a hand configuration c and a sensor s, denote the value of each sensor angle by $s(c)$, e.g., F1J1(c), S45(c) etc.

2.1 Fuzzification

To each sensor corresponds a linguistic variable, whose values are linguistic terms representing typical angles of joints and separations. For the joints in the fingers (linguistic variables F1J1, F1J2, F1J3 etc.) the linguistic terms are: STRAIGHT, CURVED and BENT. For the separations between fingers F1 and F2, F2 and F3, F4 and F5 (linguistic variable S12, S23, S45), the linguistic terms are: CLOSED, SEMI-OPEN and OPEN. For the separations between fingers F3 and F4 (linguistic variable S34), the linguistic terms are: CROSSED, CLOSED, SEMI-OPEN and OPEN. Tables 1 and 2 present the notations used for linguistic terms of linguistic variables representing joints and separations, respectively. Figures 4, 5, 6, 7 and 8 show the fuzzification adopted for those variables.

Table 1. Linguistic terms of linguistic variables representing finger joints.

Linguistic Term	Notation
STRAIGHT	St
CURVED	Cv
BENT	Bt

Table 2. Linguistic terms of linguistic variables representing finger separations.

Linguistic Term	Notation
CROSSED	Cr
CLOSED	Cd
SEMI-OPEN	SOp
OPEN	Op

The hand configuration is the main linguistic variable of the system, denoted by HC, whose linguistic terms are names of hand configurations, which names are application dependent. For instance, in Sect. 3, names of Brazilian Sign Language (LIBRAS) hand configurations (see Fig. 10) were used for such linguistic terms.

2.2 The Recognition Process

The hand gesture recognition process is divided into four steps: (1) recognition of finger configurations, (2) recognition of hand configurations, (3) segmentation of the gesture in monotonic hand segments and (4) recognition of the sequence of monotonic hand segments.

For the step 1 (*recognition of finger configurations*), 27 possible finger configurations are considered. These configurations are codified in the following format: XYZ, where X is the value of the linguistic variable corresponding to the first joint J1, Y is the value of the linguistic variable corresponding to the second joint J2 and Z is the value of the linguistic variable corresponding to the third joint J3. For example, StStSt is used to indicate that the three joints are

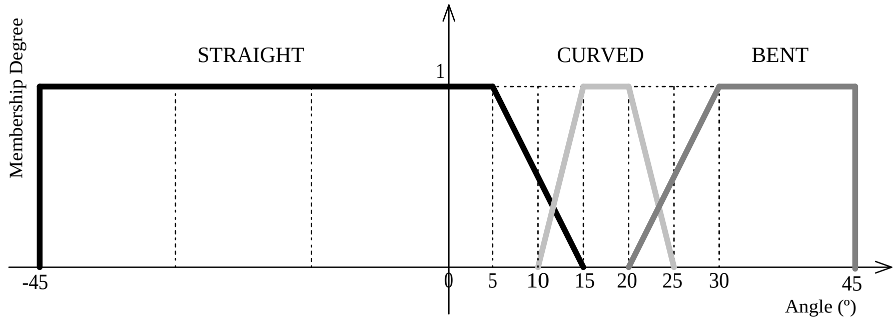

Fig. 4. Fuzzification of the linguistic variable of the joint F5J2 in the thumb finger F5.

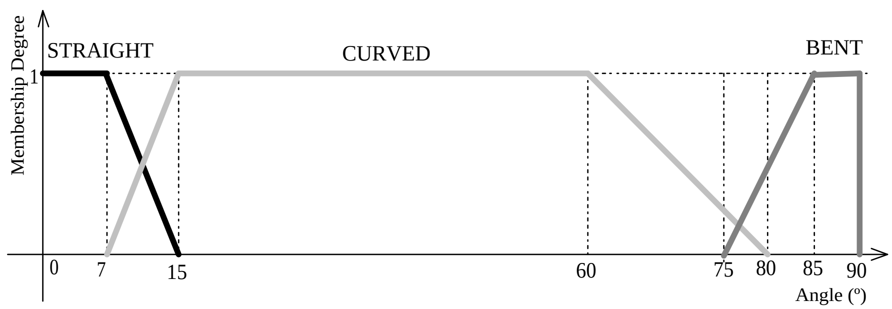

Fig. 5. Fuzzification of the linguistic variables of remaining finger joints.

STRAIGHT, StCdCd indicates that the first joint is STRAIGHT whereas the others are CURVED etc.

The 27 possible finger configurations determine 27 inference rules that calculate membership degree of each finger to each configuration. For example:

```
If F4J1 is STRAIGHT and F4J2 is CURVED and F4J3 is CURVED
Then F4 is StCdCd
```

The next step is 2 (*recognition of hand configurations*), where the hand configuration is determined, considering each finger configuration and each separation between fingers. For example, the rule for the hand configuration [G] of LIBRAS (see Fig. 10) is described below:

```
If F1 is BtBtSt and S12 is Cd and F2 is BtBtSt and S23 is Cd and
   F3 is BtBtSt and S34 is Cd and F4 is StStSt and S45 is Cd and
   F5 is StStSt
Then HC is [G]
```

In 3 (*segmentation of the gesture in monotonic hand segments*), we divide each gesture in a sequence of k limit hand configurations $l_1, \ldots, l_k$, where l_1 is the initial gesture configuration and l_k is the terminal gesture configuration.

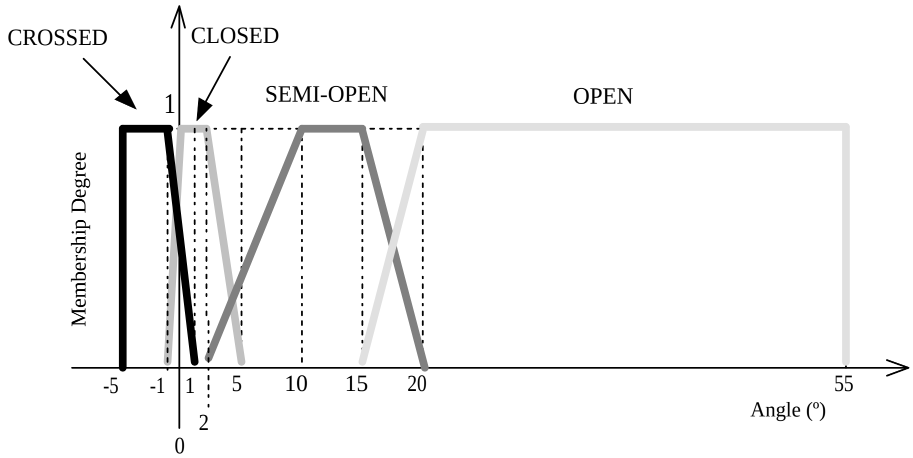

Fig. 6. Fuzzification of the linguistic variable of the separation S34 between the middle finger F3 and the index finger F4.

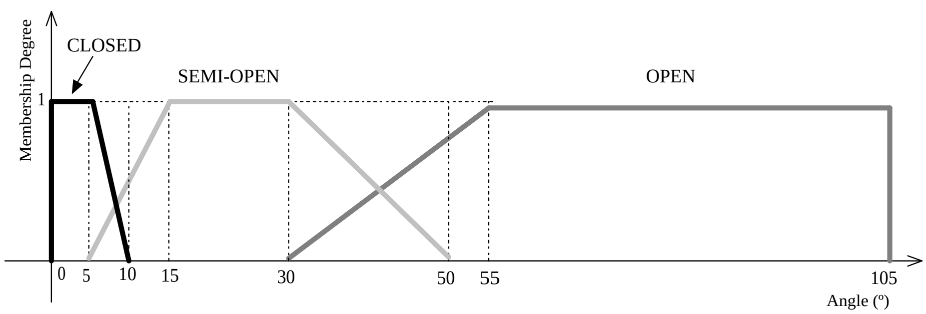

Fig. 7. Fuzzification of the linguistic variable of the separation S45 between the index finger F4 and the thumb finger F5.

The limit configurations are such that, for each c between l_i and l_{i+1}, and for each sensor s, $s(c) - s(l_i)$ has the same sign of $s(l_{i+1}) - s(l_i)$, for $i = 1, \ldots, k-1$ (a difference equal to 0 is compatible with both negative and positive signs).

The limit hand configurations are the points that divide the gesture into monotonic segments, that is, segments in which each sensor produces angle variations with constant (or null) sign. For each monotonic segment $l_i l_{i+1}$, l_i and l_{i+1} are its initial and terminal hand configurations, respectively.

The procedure for step 3 is the following. To find any monotonic segment $l_i l_{i+1}$, the next n configurations sent by the data glove after l_i are discarded, until a configuration c_{n+1}, such that the signs of $s(c_{n+1}) - s(c_n)$ and $s(c_n) - s(l_1)$ are not the same (or, c_{n+1} is the last configuration of the gesture). Then, c_n (resp., c_{n+1}) is the terminal hand configuration l_{i+1} of the considered monotonic segment, and also coincides with the initial configuration of the next segment

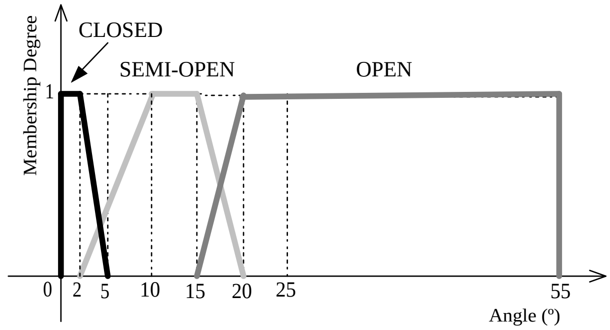

Fig. 8. Fuzzification of the linguistic variables of the separations between remaining fingers.

$l_{i+1}l_{i+2}$ (if there is one). The process starts with $l_i = l_1$, which is the initial gesture configuration, and is repeated until the end of the gesture, generating the list of k limit hand configurations.

In 4 (*recognition of the sequence of monotonic hand segments*), the recognition of each monotonic segment l_il_{i+1} is performed using a list of reference hand configurations $r_1, r_2, \ldots, r_m$ that characterizes the segment, where r_1 and r_m are the initial and terminal hand configurations of the segment, respectively. A monotonic segment is recognized by checking that it contains its list of reference hand configurations. The process is equivalent to a recognition based on a linear finite automaton (shown in Fig. 9), where $l_i = r_1$ and $l_{i+1} = r_m$.

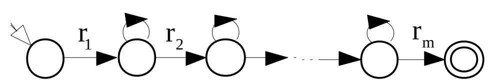

Fig. 9. Automaton for the recognition of monotonic segments.

3 Case Study: Hand Gestures of LIBRAS

As any other sign language (e.g., ASL – American Sign Language, used in the USA), LIBRAS (Língua Brasileira de Sinais – Brazilian Sign Language) is a natural language endowed with all the complexity normally found in the oral-auditive languages. Thus, it can be analyzed at all the various linguistic levels encountered in such languages, such as the "phonetic-phonological" level (also called "cheremic" level, for its relationship with the movement of the hands), the syntactic level, and the semantic and pragmatic levels [17].

As a language of the specific modality called visual-gestural, however, the elements that constitute many of those linguistic levels are of a specific nature.

For instance, the main parameters that characterize the “phonological” units of sign languages are: the configurations of the hands used in the gestures, the main spatial location (relative to the persons who is signing) where the movements of the gestures are performed, the different movements (of the fingers in the hand, of the hands and arms in the space, of the whole body) that constitute the gesture, the facial expressions that express different syntactic, semantic and pragmatic marks during the production of the signs etc.

In the various works on automatic recognition of sign languages that have been developed along the years (see Sect. 1) the recognition of hand gestures has occupied a prominent place. Using capture devices like data gloves and cameras, hand gestures have been analyzed and recognized in order to allow the computer understanding of such basic component of sign languages.

To support that recognition process, a reference set of hand configurations is usually adopted, driven either from the linguistic literature on sign languages, or dynamically developed by the experimenters with an ad hoc purpose. For our purposes, we have chosen a standard set of hand configurations (some of them shown in Fig. 10), taken from the linguistic literature on LIBRAS [17].

Since we take the set of hand configurations from the literature, our method requires that each sign be thoroughly characterized in terms of its monotonic segments and the sequences of hand configurations that constitute such segments, and that the identification of the monotonic segments and hand configurations be manually provided to the system. Of course, a capture device such as a data glove can be used to help to identify the typical values of the angles of the finger joints, but the final decision about the form of the membership functions that characterize the linguistic terms used in the system has to be explicitly taken and manually transferred to the system.

We illustrate here the application of the method by the definition of the necessary parameters for the recognition of the hand gestures that constitute the sign CURIOUS, in LIBRAS. CURIOUS is a sign performed with a single hand placed right in front of the dominant eye of the signer, with the palm up and hand pointing forward. The initial hand configuration is the one named [G1] in Fig. 10. The gesture consists of the monotonic movement necessary to perform the transition from [G1] to [X] and back to [G1] again, such movements been repeated a few times (usually two or three). Thus, a possible analysis of the hand gestures that constitute the sign CURIOUS in LIBRAS is:

Initial configuration: [G1]
Monotonic segment S1: [G1]-[G1X]-[X]
Monotonic segment S2: [X]-[G1X]-[G1]
State transition function for the recognition automaton: see Fig. 11.

To support the recognition of the monotonic segments of CURIOUS, we have chosen to use one single intermediate hand configuration, [G1X]. It is an intermediate configuration that does not belong to the reference set (Fig. 10) and whose characterization in terms of the set of membership functions for linguistic terms was defined in an ad hoc fashion, for the purpose of the recognition of CURIOUS. Together with [G1] and [X], it should be added to the list of hand configurations used by the recognition system.

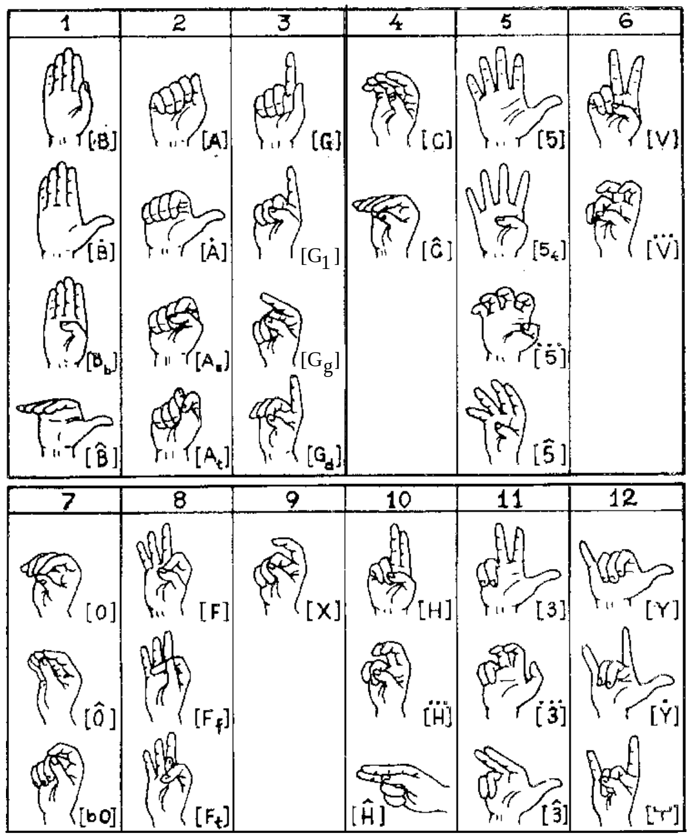

Fig. 10. Some LIBRAS hand configurations.

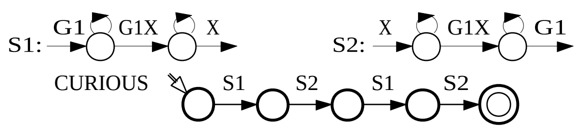

Fig. 11. Automaton for the recognition of hand gestures of the sign CURIOUS.

4 Conclusion and Final Remarks

We presented a fuzzy rule-based for the recognition of hand gestures. The method is highly dependent on a detailed previous analysis of the features of the gestures to be recognized, and on the manual transfer of the results of that analysis to the recognition system. This makes it suitable for the application to the recognition of hand gestures of sign languages, because of the extensive analysis that linguists that have already done of those languages. Prototypes of a random gesture generator and of the gesture recognizer were implemented in the programming language Python. In the fuzzification process, we considered only trapezoidal fuzzy sets and the minimum (or Gödel) t-norm, motivated by simplicity. Initial experimentation indicated promising results. Future work is concerned with the recognition of arm gestures, by including the analysis of the angles of arm joints.

References

1. L.A. Zadeh, Fuzzt Sets, *Information Control* **8**, 338–353 (1965).
2. S. Mitra, S.K. Pal, Fuzzy Sets in Pattern Recognition and Machine Intelligence, *Fuzzy Sets and Systems* **156**, 381–386 (2005).
3. G. Chen, T.T. Pham, *Fuzzy Sets, Fuzzy Logic, and Fuzzy Control Systems* (CRC Press, Boca Raton, 2001).
4. C. Carlsson, R. Fuller, *Fuzzy Reasoning in Decision Making and Optimization* (Physiva-Verlag Springer, Heidelberg, 2002).
5. W. Siler, J.J. Buckley, *Fuzzy Expert Systems and Fuzzy Reasoning* (John Wiley & Sons, Inc., New York, 2004).
6. C.T. Lin, C.S.G. Lee, *Neural Fuzzy Systems: A Neuro-fuzzy Synergism to Intelligent Systems* (Prentice Hall, Upper Saddle River, 1996).
7. M. Hanmandlu, M.H.M. Yusof, V.K. Madasu, Off-line Signature Verification and Forgery Detection Using Fuzzy Modeling, *Pattern Recognition* **38**(3), 341–356 (2005).
8. K. Kwak, W. Pedrycz, Face Recognition Using a Fuzzy Fisherface Classifier, *Pattern Recognition* **38**(10), 1717–1732 (2005).
9. O. Bimber, Continuous 6DOF Gesture Recognition: a Fuzzy-logic Approach, in: Proc. of VII Intl. Conf. in Central Europe on Computer Graphics, Visualization and Interactive Digital Media, V. 1. (1999), pp. 24–30
10. J. Ou, X. Chen, J. Yang, Gesture Recognition for Remote Collaborative Physical Tasks Using Tablet PCs, in: Proc. of IX IEEE Intl. Conf. on Computer Vision, Work. on Multimedia Technologies in E-Learning and Collaboration (Nice, 2003).
11. N.D. Binh, T. Ejima, Hand Gesture Recognition Using Fuzzy Neural Network, in: Proc. ICGST Conf. Graphics, Vision and Image Proces (Cairo, 2005), pp. 1–6.
12. M. Su, A Fuzzy Rule-based Approach to Spatio-temporal Hand Gesture Recognition, *IEEE Transactions on Systems, Man and Cybernetics, Part C* **30**(2). 276–281 (2000).
13. P. Hong, M. Turk, T.S. Huang, Gesture Modeling and Recognition Using FSM, in: Proc. of IEEE Conf. Face and Gesture Recognition (Grenoble, 2000), pp. 410–415.
14. G. Rigoll, A. Kosmala, S. Eickeler, High Perfomance Real-time Gesture Recognition Using Hidden Markov Models, in: Gesture and Sign Language in Human-Computer Interaction, Proc. of International Gesture Workshop, Bielefeld, 1997, edited by I. Wachsmuth, M. Frölich (n. 1371 in LNAI, Springer, 1998), pp. 69–80.
15. G. Fang, W. Gao, D. Zhao, Large Vocabulary Sign Language Recognition Based on Fuzzy Decision Trees, *IEEE Transactions on Systems, Man and Cybernertics* **34**(3), 305–314 (2004).
16. O. Al-Jarrah, A. Halawani, Recognition of Gestures in Arabic Sign Language Using Neuro-fuzzy Systems, *Artificial Intelligence* **133**(1–2), 117–138 (2001).
17. L.F. Brito, *Por uma Gramática de Línguas de Sinais* (Tempo Brasileiro, Rio de Janeiro, 1995) (in Portuguese).