XVII IMEKO World Congress
Metrology in the 3rd Millennium
June 22-27, 2003, Dubrovnik, Croatia

FUZZY GLOVE FOR GESTURE RECOGNITION

Thomas ALLEVARD, Eric BENOIT, Laurent FOULLOY

LISTIC-ESIA, Université de Savoie, Annecy, France

Abstract - This paper presents an application of a fuzzy rule-based aggregation to a dataglove for the recognition of gestures. The fuzzy glove is a dataglove that has fuzzy sensor functionalities. The approach used for the definition of numerical to linguistic conversion, and for the definition of the sets of rules is discussed.

Keywords: Fuzzy Sensors, Dataglove, Gesture recognition.

1. INTRODUCTION

Automatic human gestures recognition is a research subject which is studied since the seventies. The first goal is to create a friendly Human Computer Interface (HCI), a computer that understands a user expressing himself with speech and gestures (Put-That-There [1]). From then, this activity has grown to become a very active research field with a lot of different applications like sign-language recognition, robot remote control or musical creation.

There are two gesture recognition families. The sensor-based systems (usually glove-based system) [2] and the vision-based systems [3]. Both have advantages and drawbacks. Vision supporters blame glove-based methods for being cumbersome and constraining for the execution of natural gestures. On the other hand, vision-based systems constrain the user to stay in the restricted field of the camera and even sometimes to wear reflecting gloves in order to improve image segmentation. The advent of wireless datagloves makes it possible to imagine an embedded sign recognition system that could be used anywhere, in streets as well as in laboratories, an idea which seems inconceivable with a vision-based recognition method.

Human gesture recognition comes within the more general framework of pattern recognition. In this framework, systems consist of two processes: the representation process and the recognition process. Whatever the acquisition system, a representation model of the studied gestures must be chosen. It leads to wonder what is called a gesture. Of course, the answer depends on the application to be developed and the kind of gestures to be recognized. Different models can be used to recognize static postures of a small sign vocabulary or to recognize a sentence of French sign language for example. In the first problem, a very simple model is sufficient whereas in the latter, the very complex structure of French sign language, its temporal aspects, the co-articulation effect, must be taken into account.

Sign language are studied by linguists who try to understand their organization. But these studies are more of a phonological type as remarked by Braffort [4]. Their goal is to construct an alphabet which would help transcribing a sign language speech into a graphical representation, to code the sign language. Hence these studies do not take care of the realization of the signs, the articulations engaged in the hand motion. A joint-based study of the gesture realization is more general and more understandable for non-linguist people. Such a study has been performed by Braffort for the french sign language. It was shown that a gesture is described by four independent parameters : the hand configuration, its orientation, its location around the body and its movement. In the following, this four parameters will be used to characterised the gestures to be recognized.

This paper is concerned with the recognition of the hand configuration which is actually the hand shape and is often called the hand posture. Each hand posture corresponds to an order for a small robot to be controlled with gestures.

There are mainly three kind of gestures representation methods: the prototype based methods, the discrete spaces based ones and the kinematic and dynamic characteristics based ones. A statistic model is often added to one of these methods to deal with uncertainties occurring in gestures realisation (inter- or intra-user variations). The discrete space based methods consist of partitioning the raw data representation space into bounded areas which correspond for example to key positions of the fingers (like *bent* or *straight*). They have been criticized for their lack of robustness that comes from the coarse boundaries between the different areas (see Braffort [4]).

In this paper, it is proposed to revisit this representation method with a fuzzy partitioning of the data representation space. Instead of having coarse boundaries, the areas corresponding to the key positions will be considered as fuzzy subsets of the raw data spaces. The configuration of each finger and their relative positions are characterized with this method. Fuzzy aggregation rules are then used to extract the hand postures from the linguistic description of the fingers configurations. This approach differs from related works [5] by the fact that the physical to numerical interfaces and the numerical to linguistic interfaces are part of the same fuzzy sensor [6].

2. RECOGNITION PROCESS

Fig. 1. illustrates the general structure of the recognition process. The physical to numerical interface is performed by the Cyberglove® from Immersion. It is integrated with a numerical to linguistic interface: the fuzzy glove.

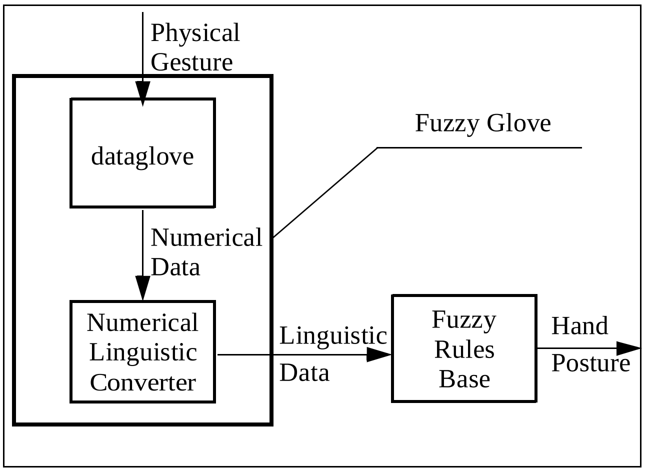

Fig. 1. Structure of the recognition process

The Cyberglove® has 15 bending sensors (Fig. 2.).

This glove was chosen because it gives an accurate and independent measurement of the different joint angles.

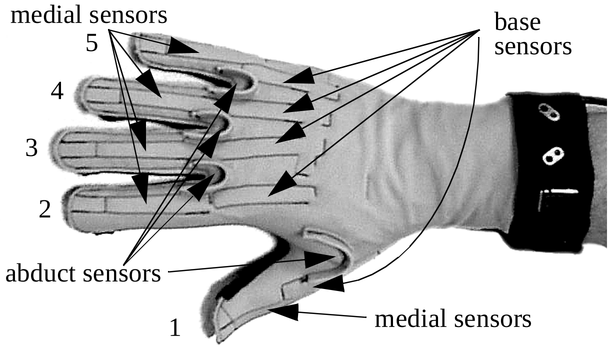

Fig. 2. Picture of the cyberglove, name of the sensors and fingers numbering

Representational theory of measurement [7] formally defines measurement process as an homomorphism M from an empirical relational system <Q,R> to a symbolic relational system <L,T>. According to this definition, numerical measurements are only particular cases of measurements where symbols are numbers. But there are many applications where this restriction of symbols to number is very limiting. Fuzzy symbolic sensors [6] have been introduced to fill this gap.

Whereas the Cyberglove® provides a numerical description of a hand posture, the task of the fuzzy glove is to give a linguistic description of the same hand posture. Hence, the fuzzy glove is typically a fuzzy symbolic sensor.

3. FUZZY SENSORS

A symbolic or linguistic sensor is compound of one or more numerical sensors and one or more numerical to linguistic converters. A converter receives numerical measures from d sensors which take their values on generally continuous spaces N_k, k= 1…d. The Cartesian product of those d spaces is called the universe of discourse X of the converter. So d is the dimension of this universe of discourse. For each numerical value x of X, the converter returns a set of linguistic terms $d(x)$, called the description of x, taken from a discrete lexical set L (Zadeh [8]) .

Dually, given a lexical term *word*, the set of the numerical values the descriptions of which contain *word* is called the meaning of *word*: $m(word)$. So we have:

$$\forall x \in X, \forall word \in L, x \in m(word) \Leftrightarrow word \in d(x) \qquad (1)$$

Fuzzy sensors are symbolic sensors where the meanings and the descriptions are fuzzy subsets respectively of X and L. Hence, we have:

$$\forall x \in X, \forall word \in L, (\mu_{m(word)}(x) = \mu_{d(x)}(word)) \qquad (2)$$

The image of L by m, $m(L)$, is a set of fuzzy subsets of X. It is imposed that $m(L)$ is a strict fuzzy partition of X that is :

$$\forall x \in X, \sum_{w \in L} \mu_{m(w)}(x) = 1 . \qquad (3)$$

The definition of this fuzzy partition defines completely the converter.

As a numerical sensor provides the values of some numerical variables, a fuzzy sensor provides the values of some linguistic variables. Hence, a linguistic variable takes values which are some fuzzy descriptions.

The general organization of a fuzzy sensor with n sensors and p converters is given in Fig. 3.

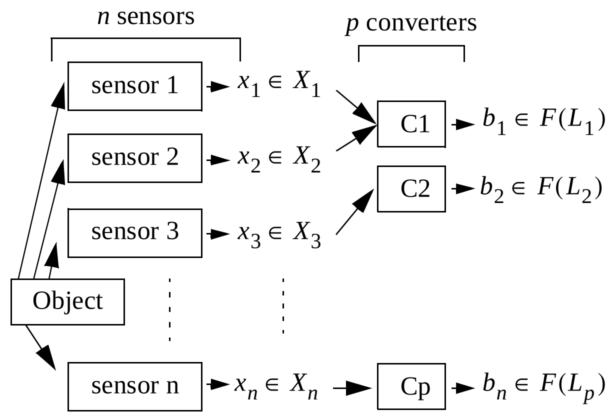

Fig. 3. Structure of a fuzzy sensor

As the number of numerical and linguistic variables is large, the following notations are defined. If an entity is named *sensor* its universe of discourse is written X_{Sensor}, its lexical set is noted L_{Sensor}, its numerical value is noted x_{Sensor} and its linguistic value is simply noted *Sensor*.

When the universe of discourse X of a converter C is mono-dimensional (d=1), a fuzzy partition can be easily defined on X. In this paper, the fuzzy meanings are extrapolated from a set of characteristic values. The numerical to linguistic conversion in the mono-dimensional case is illustrated by the example of the index medial sensor in Fig. 4. In this example, the fuzzy description of the numerical value 1 is given under the so-called additive form by: $d(1) = 0.2/half + 0.8/straight$.

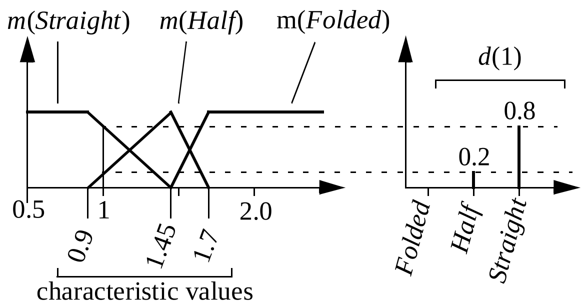

Fig. 4. Numerical to linguistic conversion

When $d > 1$, the solution which is generally adopted is to construct d mono-dimensional fuzzy partitions, one for each sensor. These partitions provide linguistic values taken from lexical sets $L_k=\{w_{k,1}, w_{k,2}, w_{k,3},\ldots\}$. $k=1,\ldots,d$. Basically, the d-dimensional numerical space is transformed into a d-dimensional linguistic space. This linguistic information can then be fused using a set of fuzzy rules (Fig. 5.).

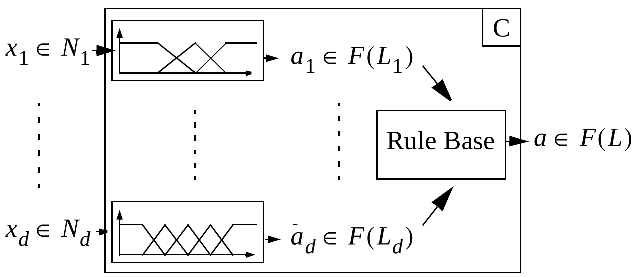

Fig. 5. Structure of a converter.

Using fuzzy rules to classify the numerical data is not always possible. Rule based classifiers have a parallel axis bias, that means the decision boundaries are parallel to the coordinate axis. If the different classes cannot be separated by such boundaries the data shall be projected on more discriminant axis, using the Discriminant Components Analysis (DCA). If it is still impossible to separate the classes, more rules shall be used which increase the complexity of the fuzzy sensor and hence its intuitiveness.

Many algorithms allow to infer rules from a set of data (rule inducing, [9]), but the rules created by such algorithms have no semantic. These algorithms are useful mainly when rules have to be modified regularly, which is not the case here. The rules can also be given by an expert. When there is no expert knowledge, an analysis of the data can help to acquire this missing knowledge.

4. THE FUZZY GLOVE

As it has been said in section 2, the task of the fuzzy glove is to give a linguistic description of the hand posture. This description must be meaningful. This means that given this description, one should be able to understand the shape of the gesture or to realize the gesture without particular knowledge about sign languages.

A first knowledge is introduced by assuming that a hand posture is a special combination of fingers configurations. Hence, a hand posture is fully depicted by the description of each finger configuration and of their relative positions.

It is also admitted that the configurations of fingers (index to pinkie) must be described by one linguistic variable each, as well as the relative position of finger 2/3, 4/3 and 5/4. The description of the thumb position and configuration is less intuitive: two linguistic variables are devoted to this task. Hence, the thumb is treated differently from other fingers. In what follows, the term “fingers” will signify: “all fingers excepted the thumb”.

Pictures of the five configurations of fingers used in this paper are given in Fig. 6. It is pointed out that though they have been found by an independent analysis of data, these five configurations of the fingers are the same as the one used in the Hamnosys sign language notation system [10].

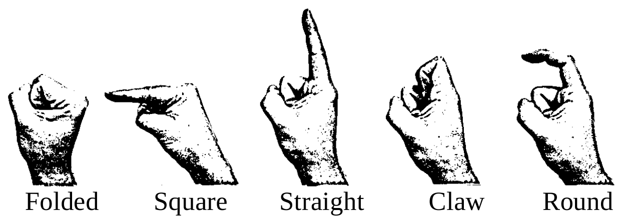

Fig. 6. Picture of the five key configurations of a finger

It must be decided which sensors will contribute to the evaluation of the different linguistic variables. It seems obvious that given a finger, its configuration shall be evaluated from the values of the two sensors measuring its bending angles: the medial and base sensors. Hence the universe of discourse of the corresponding numerical to linguistic converter is two-dimensional and fuzzy rules have to be defined.

Each of the three relative positions (2/3, 4/3 and 5/4) shall also be evaluated from the values of the corresponding abduct sensor. The corresponding converters are thus mono-dimensional and a simple fuzzy partition has to be defined.

The last four bending sensors are used to evaluate the thumb position and configuration. A fuzzy rule base must also be defined.

5. LINGUISTIC DESCRIPTION OF FINGERS CONFIGURATIONS

For a better understanding of the process, the linguistic description of the index finger is described in details. The rule base presented here, as well as the mono-dimensional fuzzy partitions and the different lexical sets, have been defined by an analysis of the data which is described in next section.

To qualify linguistically the configuration of a finger, two sensors are used: the finger base sensor and finger medial sensor. Two mono-dimensional partitions have then to be defined on the numerical spaces of those two sensors. The corresponding lexical sets are respectively: $L_{BaseIndex}$ = {*Folded*, *Straight*} and $L_{MedialIndex}$ = {*Folded*, *Half*, *Straight*}. The final lexical set used to describe the whole index finger is: L_{Index} = {*Folded*, *Straight*, *Square*, *Round*, *Claw*}. The set of rules is given in the table of Fig. 7.

Medial Sensor of index \ Base sensor of index	*Straight*	*Folded*
Folded	*Claw*	*Folded*
Half	*Round*	*Round*
Straight	*Straight*	*Square*

Fig. 7. Set of rules for Index

IF *BaseIndex* is *Straight* AND *MedialIndex* is *Folded* THEN *Index* is *Claw* (4)

The application of these rules produces a fuzzy subset of the lexical set L_{Index}.

In the case of the index finger, this fuzzy subset is the value of the linguistic variable *Index*, that describes the index posture. Each rule is equivalent to an equation that links this linguistic variable with the intermediary linguistic variables *MedialIndex* and *BaseIndex*.

In this paper, the rules are interpreted conjunctively and, for example, the degree of membership of *Claw* to *Index* is given by (5):

$$\mu_{\text{Index}}(Claw)=T(\mu_{\text{BaseIndex}}(Straight), \mu_{\text{MedialIndex}}(Folded)) \quad (5)$$

Where T is a triangular norm.

The same conclusion can be present in more than one rule: it is the case for *Round*. Then, the rules are combined disjunctively:

$$\mu_{\text{Index}}(Round)=\perp(T(\mu_{\text{BaseIndex}}(Straight),\mu_{\text{MedialIndex}}(Half)), T(\mu_{\text{BaseIndex}}(Folded),\mu_{\text{MedialIndex}}(Half))) \quad (6)$$

Where T and $\perp$ are respectively a triangular norm and a triangular co-norm.

Finally, to a numerical value x_{Index} = ($x_{BaseIndex}$, $x_{MedialIndex}$), corresponds a linguistic value *Index*. Dually, to a lexical value $w \in L_{Index}$, corresponds a fuzzy meaning $m(w) \in F(X_{Index})$. It has been shown in [11] that if the T-norm and T-conorm in (5) and (6) are respectively the product and the bounded sum, and if the rule base is complete, then $m(L_{Index})$ is also a fuzzy partition of X_{Index}.

The same lexical sets and rule base are used for the fuzzy description of the middle, the ring and the pinkie fingers. The meaning of basic linguistic terms differs for each finger.

6. DATA ANALYSIS

It is considered as common sense knowledge that relative positions of 2/3, 4/3 and 5/4 can either be *Separated* or *Together*. Hence, those linguistic variables will take values on the same lexical set: L_{relative} = {*Separated*, *Together*}. A fuzzy partition of the corresponding abduct sensor is easily defined by taking two characteristic values respectively of the classes *m*(*Separated)* and *m*(*Together)*.

But there is not common sense knowledge nor about how a finger configuration shall be described neither how the thumb configuration and position shall be described. To acquire this knowledge, an analysis of the data must be conducted.

6.1. Analysis of finger description

First, the analysis of fingers description is presented. A set of natural hand postures have been measured. These measures are then represented in the Universe of discourse of the studied finger (Fig. 8.). The results and treatments are the same for all fingers.

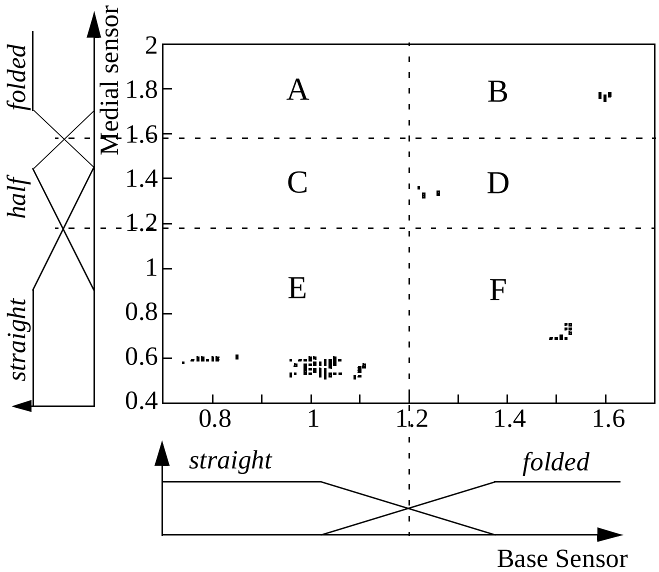

Fig. 8. Data analysis of finger postures

Four clusters can be discerned. They correspond to four different configurations the finger can take in a natural hand posture. It is possible to separate them with axis parallel boundaries. The edges that have been chosen are represented in Fig. 8. Other choices could be made but this one uses only three borders and so limits the number of rules.

These edges define 6 fuzzy areas: A,B,C,D,E and F. Though there is no cluster within square A, it still represents a configuration of the finger which is simply not present in the studied set of postures but could be present in some other gestures and thus must be taken into account. It has also been found that finger configurations having the same appearence can be present in both areas C and D. Thus, the union of those two areas represents one configuration.

Hence, only five fuzzy classes are defined: A,B,C $\cup$ D , E and F. These five classes are the fuzzy meanings of some linguistic terms that are to be defined. Five linguistic values are chosen that evoke the corresponding configuration: {*Folded* (B), *Square* (F), *Straight* (E), *Claw* (A), *Round* ($C \cup D$)}. The two fuzzy partitions of the axis are deduced from those boundaries and symbols corresponding to these partition can then be defined.

The medial sensor axis is partitioned by 3 fuzzy subsets. The corresponding symbols are: {*folded*, *half*, *straight*}. The base sensor is partitioned by 2 fuzzy subsets and their corresponding meanings are {*folded*, *straight*}.

Finally, the rule base is the one presented in Fig. 7.

Such an analysis without previous knowledge has been possible because a finger has only 2 degrees of freedom and that the base and medial sensors are independent. But those 2 conditions are not true for the thumb. Moreover, the numerical values corresponding to two similar thumb configurations can be found very far one from another in the universe of discourse if the configurations of other fingers is different.

6.2. Analysis of the thumb description

For the fingers, the analysis of data alone has taught which configurations had to be discriminated and how to discriminate them. For the thumb, key configurations have to be defined first. This can be considered as introducing expert knowledge.

It is decided that two linguistic variables describe the thumb: one describes its configuration (*Bent* or *Straight*) and another its orientation (*Aside*, *External*, *Ahead* or *Internal*).

A set of gestures where all these key positions were represented, with different positions for the other fingers have been measured. A first DCA is made to evaluate the axis that discriminate the most the gestures with a *Bent* thumb from the gestures with a *Straight* one. This axis is found to be very close from the axis of the tip sensor. Hence, this sensor is devoted to the evaluation of the thumb configuration while the three others (medial, base and abduct sensors) are utilized to evaluate its orientation.

The second DCA was thus made on 3-dimensional data. Two discriminant axis have been extracted. Once projected on these axis, the four clouds corresponding to each key configuration are well separated. The boundaries that have been chosen are represented in Fig. 9.

These lines are not parallel to the axis and so they do not define directly a set of rules and it seems impossible to use a rule base for the classification.

Actually, to each border line corresponds a perpendicular axis, which is the most discriminant axis of the 2 sets of points separated by the border. This axis goes through the centres of gravity of these two sets. Three mono-dimensional partitions are defined on the three discriminant axis: X_{axis1}, X_{axis2}, X_{axis3}. They give a fuzzy description on the lexical sets $L_{axis1} = L_{axis2} = L_{axis3} = \{Neg, Pos\}$. So three discriminant linguistic variables are defined : *Axis1*, *Axis2* and *Axis3*.

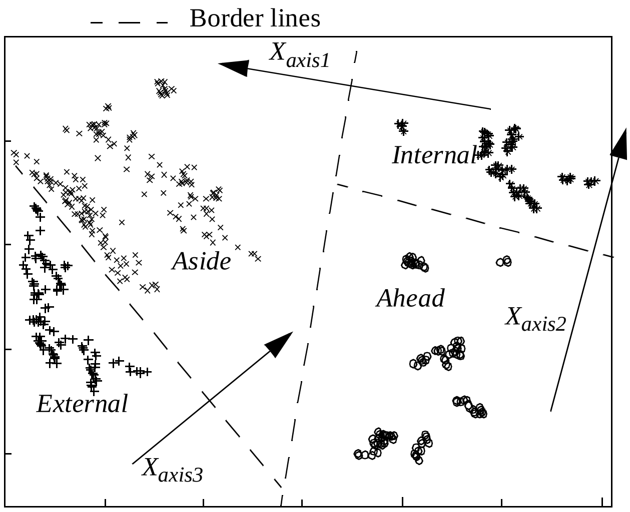

Fig. 9. Data Analysis of thumb postures

The linguistic variable *ThumbOr* is a fuzzy description on the Lexical set $L_{ThumbOr} = \{Ahead, Internal, Aside, External\}$. Its value can then be deduced from the values of the three discriminant linguistic variables by a set of four rules:

if *Axis1* is *Neg* and *Axis2* is *Neg* then *ThumbOr* is *Ahead*
if *Axis1* is *Neg* and *Axis2* is *Pos* then *ThumbOr* is *Internal*
if *Axis1* is *Pos* and *Axis3* is *Neg* then *ThumbOr* is *External*
if *Axis1* is *Pos* and *Axis3* is *Pos* then *ThumbOr* is *Aside* (7)

This base of rules is equivalent to a base made of 8 rules with 3 entries as shown in Fig. 10.

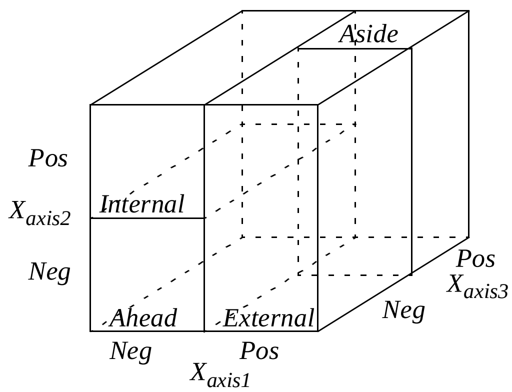

Fig. 10. Rule base for ThumbOrientation

Other supervised classification methods could have been used to construct a fuzzy partition of the universe of discourse. A Delaunay triangulation of the centres of gravity of the four clouds could be made, but some points could be classified in the wrong category by this method. The same problem would happen with a Nearest neighbour method using the centres of gravity as prototypes of the class. Moreover, the rule base method is preferred for its intuitiveness and simplicity.

It is difficult to avoid the step of choosing heuristically the different categories. FCM methods cannot be employed here because it will bring together points which correspond to different thumb configurations.

7. HAND POSTURES RECOGNITION

The fuzzy glove provides a linguistic description of the current hand posture given by nine linguistic variables. Each linguistic variable takes values which are fuzzy subsets of a lexical set. From this linguistic description, it must be evaluated if the current posture is one of the hand postures to be recognized.

A posture to be recognized is called a morphem. A morphem is defined linguistically by giving the position of each finger. Actually, the description of a morphem can be seen as a rule:

IF [(*Index* is *straight)* AND (*Middle* is *folded)*
AND ... AND (*Thumb_config* is *folded)*
AND (*Thumb_position* is *(ahead* OR *internal*))]
THEN (Current_gesture is pointing). (8)

The IF...THEN... statement is understood conjunctively. This means that the degree of truth of the conclusion is equal to the degree of truth of the premise. The premise is a conjunction of nine independent conditions. The basic conditions are of «*A* is *b*» kind where *A* is a fuzzy description (current value of a linguistic variable) and *b* is an element of the same lexical set. This way of thinking allows to compare A with subsets which are not singletons. In the last condition of (8), the fuzzy subset *Thumb_position* is compared to the subset {*ahead,internal*}.

8. CONCLUSION

Results have shown that this method is efficient for the recognition of hand postures. Independently of its efficiency to well recognize hand postures, this method has several advantages compared to other gesture recognition systems. Its main quality is to be easily understandable and then easily usable. The recognition principle is very intuitive and the system can explain the user how his gesture is classified. The fuzzy subset theory has been created to model human knowledge and perception. According to Dubois et Al. [12], fuzzy logic specificity is its capability of bridging the gap between articulated linguistic descriptions and numerical models of systems. Fuzzy logic is here used for what it has been created and where it is efficient.

This fuzzy posture model can be associated with a syntactic recognition method using fuzzy grammars [13]. Gesture recognition process using formal grammars have been criticized for there rigidity. Introducing fuzziness in such models makes them more flexible and hence adapted to the processing of varying data as the ones coming from gestures. They can be used to represent the relations between the different semantic levels of a gesture recognition process.

In future works, this method will be applied to the representation of more complex gestures. Temporal aspects of gestures will be taken into account.

REFERENCES

[1] R.A. Bolt, "Put-That-There: voice and gesture at the graphic interface", *Computer Graphics*, vol. 14, no. 3, pp. 262-270, July 1980.

[2] D.J.Sturman , D. Zeltzer, "A survey of glove-based input", *IEEE Computer Graphics and Applications*, vol. 14, no. 1, pp. 30-39, January 1994.

[3] V.I. Pavlovic, R. Sharma, T.S. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7 , pp. 677-695, July 1997.

[4] A.Braffort, "Gesture recognition and comprehension, implementation for sign language" ("Reconnaissance et compréhension de gestes, application à la langue des signes"), PhD Thesis, LIMSI, Paris-XI University, June 1996.

[5] E.J. Holden, G.G. Roy, R. Owens, "Adaptive classification of hand movement", *Proceedings of the International Conference on Neural Networks*, vol. 3, pp. 1373-1378, 1995.

[6] G. Mauris, E. Benoit, L. Foulloy, "fuzzy symbolic sensors : from concept to applications", Measurement, vol. 12, no. 4, pp. 357-384, Feb 1994.

[7] L. Finkelstein, M.S. Leaning, "A review of the fundamental concepts of measurement", *Measurement*, vol. 2, no. 1, pp. 25-34, Jan-Mar 1984.

[8] L.A. Zadeh, "Quantitative fuzzy semantics", *Information Sciences*, vol. 3, pp.159-176, 1971.

[9] J.C. Bezdek, J.Keller, R. Krisnapuram, N.R. Pal, "Fuzzy models and algorithms for pattern recognition and image processing", kluwer Academics publishers, 1999.

[10]S.Prillwitz, R.Leven, H.Zienert, T.Hamke, and J. Henning. "HamNoSys Version 2.0: Hamburg Notation System for Sign Languages: An Introductory Guide", *Signum Press*, 1989.

[11]E. Benoit, G.Mauris, L. Foulloy, "Fuzzy sensor aggregation: application to comfort measurement", proc of Int. Conf. IPMU, pp 721-726, July 1994

[12]D. Dubois, H.T. Nguyen, H. Prade, M. Sugeno, "Introduction : The real contribution of fuzzy systems", in: *Fuzzy systems: Modelling and control* (Eds. H.T. Nguyen, M. Sugeno), Kluwer, Boston, pp. 8-10, 1998.

[13]H. Bunke, "syntactic and structural pattern recognition : theory and applications", World scientific, Singapore, 1990.

AUTHORS: Thomas ALLEVARD, thomas.allevard@univ-savoie.fr. Eric BENOIT, eric.benoit@univ-savoie.fr. Laurent FOULLOY, laurent.foulloy@univ-savoie.fr.
LISTIC-ESIA, Université de Savoie, B.P. 806, 74016 Annecy cedex, France