

Sign Language Recognition: an Application of the Theory of Size Functions

Claudio Uras and Alessandro Verri
Dipartimento di Fisica dell'Università di Genova
Via Dodecaneso 33, 16146 Genova, Italy
E-mail: uras@ge.infn.it

Abstract

This paper discusses the use of certain integer valued functions of two real variables, named *size functions*, for shape representation and recognition. The recognition of the signing alphabet is described as a study case. A number of size functions are computed from the edge map of the viewed sign and a feature vector based on the obtained size functions is formed. A training set of feature vectors built from real images and the k -nearest-neighbor rule are employed for the classification of unpreviously seen signs. The proposed system performs recognition at about 2Hz with feature vectors of small dimension. The reported experiments indicate that size functions can be effectively used for the recognition of nonrigid shapes.

1 Introduction

Recognition is a fundamental aspect of vision. Most of the existing recognition systems (see [1, 2, 3, 4, 5, 6], for example) are based on rather strong geometric assumptions that might not be suitable for the analysis of natural shapes. Recently, a theory for shape representation and recognition based on the notion of *size function*, an integer valued function of two real variables, has been proposed [7, 8]. This theory seems to be adequate for the analysis of shapes for which a geometric model can be difficult to obtain [9].

In previous works [9, 10] an implementation of the theory based on the extraction of the outline of the viewed shapes was proposed, and the properties of size functions in the framework of computer vision were discussed. A main limitation of the proposed implementation was the fragility with respect to edge fragmentation. This paper presents a new interpretation of the theory in which no assumption about the topological connections of the extracted edge points is needed. This interpretation lead to an improved version of a system recently proposed for the recognition of the alphabet of the sign language [10].

The paper is organized as follows. First, a new interpretation of the theory of size functions is presented. Then, the representation of shape which can be obtained by means of families of size functions is described. Finally, the proposed recognition system and the experimental results are illustrated.

2 The theory of size functions revisited

The theory of size functions provides a geometric and topological framework in which the similarity between different functions can be assessed. Let us illustrate the main idea of the theory through the example of Fig.1.

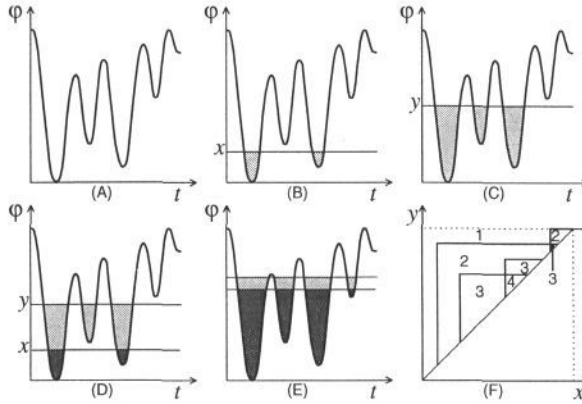


Figure 1: The concept of size function. (A). Graph of some measuring function φ . (B) and (C). The shaded regions identify the points with $\varphi \leq x$ and $\varphi \leq y$ respectively. (D). The darker shading identify the regions with both $\varphi \leq x$ and $\varphi \leq y$. (E). The same construction of (C) for a different value of x and y . (F). Graph of the size function $\ell_\varphi(x, y)$. Each label denotes the value of ℓ_φ within the underlying region. Throughout the paper the graph of a size function is always reproduced in the triangular region between the minimum and maximum value of the measuring function with $x < y$.

Let us assume that Fig.1(A) reproduces the graph of a function φ , named *measuring function*, obtained as the result of some measurement. The goal of the theory is to represent both the qualitative and quantitative aspects of the behavior of φ . This is done by means of an integer valued function of two real variables $\ell_\varphi = \ell_\varphi(x, y)$ named *size function*. Let us evaluate the size function ℓ_φ at a specific point (x, y) (with $x < y$) of the real plane. The shaded regions of Fig.1(B) and (C) identify the parts of the graph with $\varphi \leq x$ and $\varphi \leq y$ respectively. The superposition of Fig.1(B) and (C) is displayed in Fig.1(D). The darker regions of Fig.1(D) clearly correspond to the intersection between the parts of the graph with $\varphi \leq x$ and $\varphi \leq y$.

The size function $\ell_\varphi(x, y)$ (see Fig.1(D)) can be computed as the number of distinct regions in light gray which have a segment in common with at least one region in dark gray. In mathematical terms, $\ell_\varphi(x, y)$ equals the number of connected components of the set $\varphi^{-1}((-\infty, y])$ which contain at least a point of the set $\varphi^{-1}((-\infty, x])$. Hence, in the specific example of Fig.1(D), $\ell_\varphi = 2$ since only the leftmost and rightmost regions in light gray contribute to ℓ_φ .

Figure 1(E) shows the same construction of Fig.1(D) for a different choice of (x, y) . For (x, y) ranging over the real plane (with $x < y$) an integer valued function $\ell = \ell_\varphi(x, y)$ is obtained. The graph of ℓ_φ with x and y which range in the horizontal and vertical axis respectively is reproduced in Fig.1(F).

Let us briefly recall the main results of the theory. A first fundamental result of the theory ensures the finiteness of a size function within the triangular region

$$T_\varphi = \{(x, y) : \varphi^{\min} \leq x < y \leq \varphi^{\max}\},$$

with φ^{\min} and φ^{\max} the minimum and maximum of the measuring function φ respectively. Moreover, the normalization

$$\hat{\varphi} = \frac{\varphi - \varphi^{\min}}{\varphi^{\max} - \varphi^{\min}}$$

ensures that the size function $\ell_{\hat{\varphi}}$, computed within the triangular region

$$T_{\hat{\varphi}} = \{(x, y) : 0 \leq x < y \leq 1\},$$

is invariant for scale transformation.

Second, it can easily be proved that a size function $\ell_\varphi(x, y)$ is nondecreasing with x , nonincreasing with y , and continuous on the left in both x and y . Consequently, the restriction of a size function along a line of constant x or y is always monotonic (see Fig.1(F), for example) and the entire function piecewise constant. Furthermore, the pointwise computation of a size function in the region T_φ is usually redundant and unnecessary. All the information about a size function ℓ_φ resides in the location and value of the jumps of ℓ_φ along the horizontal and vertical direction of the real plane.

Third, let us describe the extension of the theory to the discrete case. Let us sample the measuring function φ of Fig.2(A) at a finite number of points $p^1 < \dots < p^i < \dots < p^N$. Furthermore, let G be the graph whose vertices are the points p^i and whose edges connect adjacent points (see Fig.2(B)), and $G_{\varphi \leq y}$ the subgraph of G naturally induced by the measuring function φ (see Fig.2(C)). In this discrete setting, the size function ℓ_φ induced by the measuring function φ is defined as the number of connected components of the graph $G_{\varphi \leq y}$ which contain at least a vertex with $\varphi \leq x$ (see Fig.2(C)).

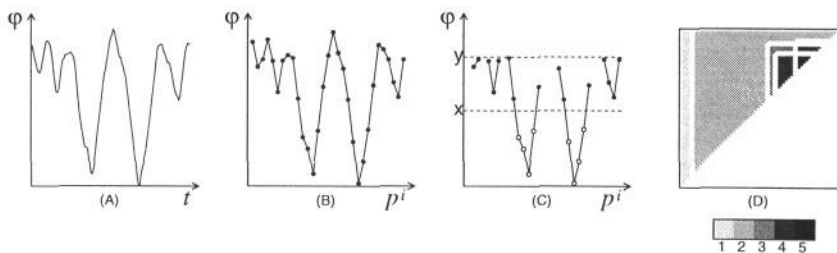


Figure 2: Discrete size function. (A). The graph of a measuring function φ in the continuous case. (B). The graph G discrete approximation of φ . For better visualization, only a few vertices of the graph are displayed. (C). The subgraph $G_{\varphi \leq y}$. The open circles mark the vertices with $\varphi \leq x$. (D). The gray value encoded size function ℓ_φ with the corresponding white bands (see text for details). The gray code is shown below.

It can be shown that the estimates of a size function ℓ_φ can be divided into two groups [8]. The estimates of the first group are known to be equal to the

exact values of the continuous case. For the size function l_φ induced by the measuring function of Fig.2(A), these estimates are reproduced in Fig.2(D). The monotonicity properties of a size function l_φ make it possible to interpolate the values of l_φ between the points of the mesh.

The white bands within the triangular region T_φ of Fig.2(D) identify the locations of the estimates of the second group. Intuitively, the white bands account for the possible difference between the size functions of two measuring functions which happen to have identical sampling. Interestingly, these differences are localized in the neighborhood of the lines of discontinuity of the size function computed from the sampling points only. Hence the white bands can be thought of as regions of uncertainty for the “true” size function. Due to the monotonicity properties the uncertainty on l_φ is more on the location of the discontinuities than on the actual values.

The actual extension of these regions of uncertainty can be determined in terms of the modulus of continuity of the measuring function $\epsilon_\varphi(\delta)$ [8]. If a size function is computed at a sufficiently high resolution, the white bands can be obtained by thickening each line of discontinuity in the diagram of the discrete size function of an amount equal to $\epsilon_\varphi(\delta)$, with δ equal to the maximum sampling distance. This is the procedure employed to generate the uncertainty bands of Fig.2(D). An algorithm for the computation of a size function was described in [11]. In essence the algorithm computes l_φ on a mesh of equally spaced points (x, y) within the triangle T_φ .

Finally, let us comment on the previous presentation of the theory. The theory of size functions was developed for the description of shape by means of topological tools [7]. In particular, in the original interpretation, the edges of the graph G were induced by the topological properties of the underlying continuous shape. In essence, the graph G was a discrete approximation of the viewed shape. Since size functions depend critically on the topological structure of the graph G , it is evident that the implementation of the theory proposed in previous papers [9, 10] was not stable with respect to changes of the topological connections of G due to edge fragmentation. In the next section a one parameter family of measuring functions which can be computed directly from the edge map of the viewed shape, with no assumption about the topological connections of the edge points, is described.

3 Shape representation

According to the proposed framework the representation of shape is based on the definition of certain functions, named measuring functions. Unfortunately, the theory does not provide formal tools to determine measuring functions, nor to compare the representations of the same shape that can be obtained by means of different measuring functions. Therefore, the search for measuring functions must be carried out heuristically. Let us illustrate the empirical principles that were employed in the search for measuring functions on a specific problem: the representation of the alphabet of the sign language.

Figure 3 illustrates the edge maps of the signs of the alphabet of the sign language performed by the same subject from “a” to “y” (the “z” sign requires the temporal dimension and has not been considered). Visual inspection of Fig.3

suggests that an adequate measuring function should not be invariant for rotation over the image plane (see the “k-v” and “d-p” sign pairs for example).

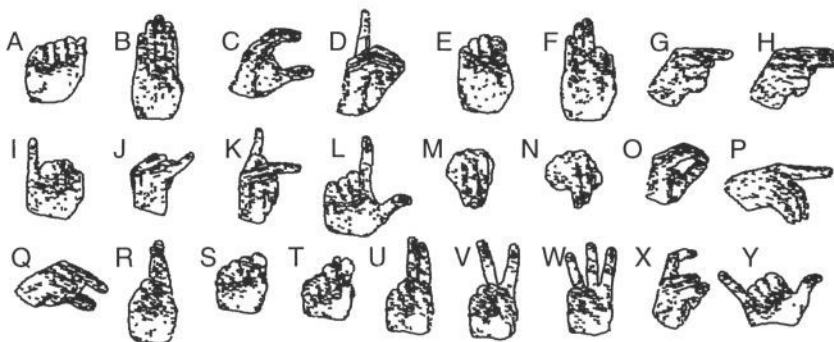


Figure 3: The alphabet of the sign language. The edge maps were extracted by means of standard edge detection techniques from real images. All the signs were performed by the same subject.

Let us describe a measuring function not invariant for rotation which was found to be adequate. The edge map of the sign of Fig.4(A) is extracted and enclosed with the rectangular box shown in Fig.4(B). For each point p of the horizontal segment through the center of the rectangular box, let $h(p)$ be the distance between p and the furthest edge point which lies in the vertical strip centered in p shown in Fig.4(B). Then, the function

$$H_0(p) = \begin{cases} h(p) & \text{if the strip is nonempty} \\ 0 & \text{otherwise} \end{cases}$$

can be thought of as a measuring function. The graph of $H_0 = H_0(p)$ for the edge map of Fig.4(B) is displayed in Fig.4(C). The induced size function ℓ_{H_0} is reproduced in Fig.4(D). Notice that the *ad hoc* definition of H_0 at the points below the horizontal segment ensures that ℓ_{H_0} looks only at the portion of shape lying above the horizontal segment through the center of the rectangular box.

The measuring function H_0 was used to generate the size functions of the different signs. It was found that the size function ℓ_{H_0} is (a) usually tolerant to small changes of the apparent orientation of the viewed signs, and (b) largely insufficient to distinguish between all the different signs.

On the other hand, it is clear that the size function ℓ_{H_0} is tuned to a particular direction of the image plane. In order to produce size functions able to distinguish between the different signs, it was found convenient to regard H_0 as a particular case (the case $\theta = 0$) of a family of measuring functions indexed by the angle θ with $0 \leq \theta < 360^\circ$. A generic measuring function of the family, H_θ , is tuned to the direction θ and was computed in two steps. In the first step, the edge points were rotated clockwise of an angle θ about the center of the rectangular box enclosing the sign. In the second step, the function H_0 was computed on the rotated edge points. An example of ℓ_θ computed for $\theta = 0^\circ, 30^\circ$ and 120° for an “l” sign is shown in Fig.5.

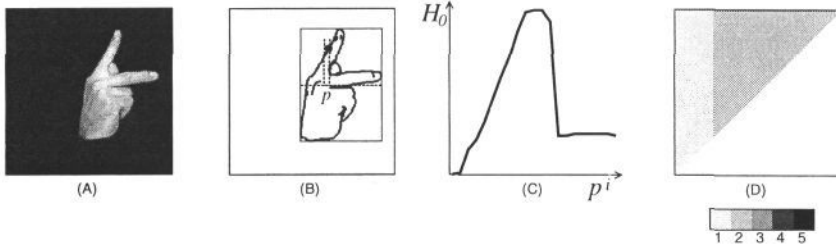


Figure 4: A measuring function not invariant to rotation. (A). An image of a “k” sign. (B). The edge map of the sign of (A) obtained by means of standard edge detection techniques and the geometric construction for the computation of the measuring function H_0 . (C). The graph of the measuring function H_0 relative to the edge map of (B). (D). The induced size function ℓ_{H_0} .

Since for each choice of θ the measuring function H_θ conveys little information about the viewed sign, it is legitimate to ask whether the shape information encoded by the collection of measuring functions H_θ is sufficient to discriminate between the different signs. Empirical evidence of the “completeness” of the family H_θ for a particular sign is shown in Fig.6. Figure 6(A) displays the original edge map of a “p” sign. Figure 6(B) reproduces the set of “furthest edge points” for $\theta = 0$ and associated strips relative to the edge map of Fig.6(A). Figure 6(C) shows the set of “furthest edge points” for all the used angles. The fact that the edge map of Fig.6(C) appears to be a good approximation of the outer contour of the edge map of Fig.6(A) indicates that the family of measuring functions H_θ takes into account the relevant shape information contained in the viewed sign.

Let us now discuss the construction of the devised recognition scheme and the obtained experimental results.

4 Shape recognition

The devised recognition system, similar to the system described in [10], consists of four stages. In the first stage, the image is acquired and an edge map of the viewed sign is produced. Since size functions can be computed independently of the topological structure of the obtained edge map, this step is considerably less critical than in the previously proposed implementation [10].

In the second stage the obtained edge map is given as input to the algorithm for the computation of the size functions ℓ_{H_θ} with $\theta = 0^\circ, 15^\circ, \dots, 330^\circ$. At the adopted resolution, each of the 24 obtained size functions ℓ_{H_θ} consists of 120 integer values computed on a mesh of equally spaced points within the triangular region T_{H_θ} .

In the third stage the information contained in the computed size functions is used to produce a feature vector v . For the sake of simplicity and in the attempt to reduce the dimensionality of the computed data, it has been decided to encode the information contained in the size function ℓ_{H_θ} , with $\theta = j \cdot 15^\circ$, in the j -th component v_j of a 24-dimensional feature vector, with $j = 0, 1, \dots, 23$. Empirical evidence suggested that the number of 1s in each of the computed size functions was

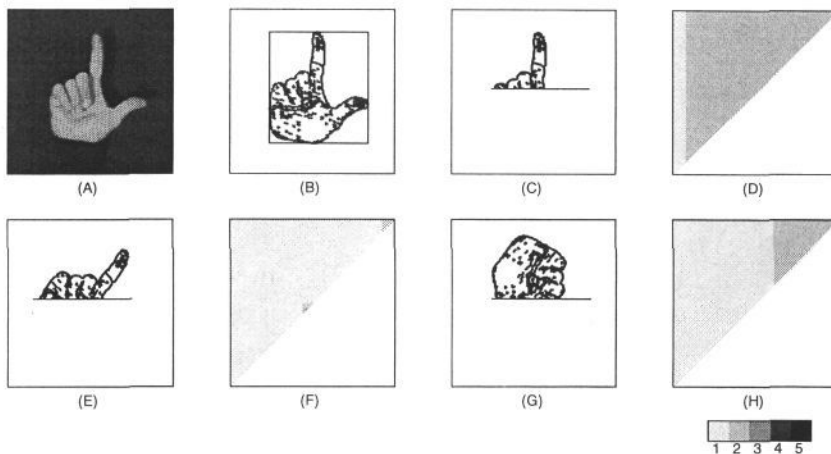


Figure 5: A family of size functions for sign language recognition. (A). An image of an “I” sign. (B). The edge map of the sign of (A) obtained by means of standard edge detection techniques and the geometric construction for the computation of the measuring function H_θ . (C) and (D). The subset of the edge points used for the computation of H_θ with $\theta = 0^\circ$ and the induced size function. (E) and (F). Same as (C) and (D) with $\theta = 30^\circ$. (G) and (H). Same as (C) and (D) with $\theta = 120^\circ$.

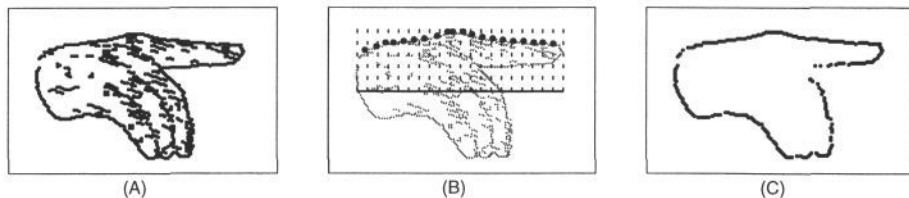


Figure 6: Empirical evidence of the “completeness” of the family of measuring functions H_θ . (A) Edge map of a “p” sign. (B) and (C) See text.

sufficient to characterize each component. The effectiveness of this considerable data reduction is mainly due to the redundancy of the produced representations. Clearly other choices are possible and no claim of optimality is made about this particular choice. Ultimately, the validity of this choice was assessed *a posteriori*.

In order to smooth out the effects of the coarse angular sampling, it was found convenient to compute the family of size functions ℓ_{H_θ} at a finer angular resolution, $\theta = 0^\circ, 5^\circ, \dots, 355^\circ$. Consequently, the j -th vector component, $j = 0, 1, \dots, 23$, was actually obtained as the average number of 1s in ℓ_{H_θ} with $\theta = j \cdot 15^\circ, j \cdot 15^\circ \pm 5^\circ, j \cdot 15^\circ \pm 10^\circ$.

In the fourth and last stage the feature vectors are used to classify the viewed signs. A number of images were processed offline and the obtained feature vectors grouped to form the training set. The k -nearest-neighbor rule was employed for the classification stage. Let us now discuss the results of the performed experiments.

5 Experiments

Five series of experiments were performed. In each series a different training set was tested by means of 20 sequences of unpreviously seen signs performed by two different subjects $S1$ and $S2$ (10 sequences each). The first three training sets, $T1$, $T2$, and $T3$, consisted of 10 sequences of signs from "a" to "y" (i.e. 250 signs for each training set). The training set $T1$ consisted of 10 sequences performed by the subject $S1$, $T2$ of 10 sequences performed by the subject $S2$, and $T3$ of 5 sequences from $T1$ and 5 from $T2$ respectively. Finally, the training sets $T4$ and $T5$ consisted of 5 sequences from $T1$ and $T2$ respectively. The k -nearest neighbor classification rule was applied with $k = 3$.

The recognition errors relative to each of the five training sets are displayed from left to right in Tab.1. For each of the signs displayed in the leftmost column, the number of errors relative to $T1, \dots, T5$, each tested by means of the sequences performed by $S1$ and $S2$ respectively, is reported. The letter in parentheses refers to the most common erroneous classification. It can easily be seen that if the training and test set consist of signs performed by the same subject ($T1 - S1$, $T2 - S2$) the average recognition rate is 98% for both $S1$ and $S2$. The percentage drops to about 82%, if the system must recognize signs performed by unpreviously seen subjects ($T1 - S2$ and $T2 - S1$), but is again well above 90% for both $S1$ and $S2$ if the training set is $T3$ (that is, a mixture of signs performed by $S1$ and $S2$).

Notice the overall graceful degradation of the recognition rates when smaller training sets are used, namely $T4$ and $T5$, (about 90% and 70% if the training and test sets are from the same subject and different subjects respectively).

A few comments are in order. First, it is interesting to note that, when larger training sets were used in the experiments, namely $T1$, $T2$, and $T3$, 9 signs ("c", "f", "j", "k", "o", "p", "v", "w", "y") were always classified correctly in all the performed experiments. If the training and test set refer to the same subject, 20 of the 25 signs were always classified correctly in all the performed experiments. Second, the vast majority of the errors involves signs for which the encoding of information about the internal edges appears to be important. Third, the style of signing seemed to be critical for a few signs only (the "l" and "m" for $S1$ and the "d" and "l" for $S2$).

The entire process (image acquisition, size function computation, and classification) takes less than 1 second on a SPARC Station 10 when a training set made of 250 feature vectors is used. The image acquisition requires more than 50% of the total time.

6 Conclusions

Finally, let us summarize the obtained results. This paper presented a scheme for the representation and recognition of static hand gestures. The scheme is based on the concept of size function, an integer valued function of two real variables which encodes information about the qualitative and quantitative properties of the viewed shape. Unlike a previously proposed version, the presented scheme can handle shapes described by fragmented edge maps. The scheme has been tested on the problem of the recognition of the signing alphabet.

Table 1: Sign recognition from the training sets $T1$, $T2$, $T3$, $T4$, and $T5$ (see text).

	T1		T2		T3		T4		T5	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
a	1(s)	9(s)	-	2(t)	-	2(s)	-	9(t)	-	2(s)
b	-	1(f)	-	-	-	1(f)	-	1(f)	-	1(f)
c	-	-	-	-	-	-	-	3(l)	-	-
d	-	9(f)	4(i)	-	-	2(r)	-	9(f)	5(i)	1(x)
e	-	-	1(j)	-	2(j)	-	2(m)	-	2(j)	-
f	-	-	-	-	-	-	3(b)	-	1(q)	-
g	-	4(h)	-	-	-	1(h)	1(h)	2(h)	-	-
h	-	-	1(g)	-	2(g)	-	1(g)	-	3(g)	-
i	-	-	3(j)	-	3(j)	-	3(d)	1(j)	3(j)	-
j	-	-	-	-	-	-	7(a)	5(m)	-	-
k	-	-	-	-	-	-	1(d)	-	-	-
l	-	8(k)	9(c)	-	-	-	-	2(k)	5(k)	-
m	1(n)	5(n)	9(n)	-	1(n)	-	1(n)	7(n)	9(n)	-
n	1(m)	-	1(m)	2(m)	1(m)	5(m)	1(m)	5(m)	2(m)	5(m)
o	-	-	-	-	-	-	-	3(j)	-	2(h)
p	-	-	-	-	-	-	-	-	-	-
q	-	1(h)	2(p)	-	1(p)	-	2(g)	1(h)	1(x)	-
r	1(u)	4(u)	1(u)	-	1(u)	4(u)	1(u)	5(u)	1(u)	2(u)
s	-	-	3(a)	-	1(a)	2(a)	-	-	6(a)	1(a)
t	-	-	3(s)	-	4(s)	2(e)	3(s)	-	4(s)	1(e)
u	1(r)	4(b)	3(r)	1(r)	1(r)	1(b)	-	3(b)	7(r)	2(r)
v	-	-	-	-	-	-	-	4(w)	-	2(w)
w	-	-	-	-	-	-	1(v)	-	2(v)	-
x	-	-	3(k)	-	-	-	-	3(u)	2(k)	-
y	-	-	-	-	-	-	-	1(i)	-	-

The obtained experimental results indicate that the system is able to perform recognition of the viewed sign from single images with very good percentages of success and reliability at the rate of $2Hz$, and is almost insensitive to the apparent shape changes induced by edge fragmentation and differences in edge localization, hand pose, and style of signing. Preliminary results indicate that the choice of the family of measuring functions is not critical [12].

It is concluded that size functions can be very useful for the analysis of natural shapes for which a geometric model can be difficult to obtain.

This research has been partially supported by the B.R.A. Project VIVA.

Laura Giovanelli checked the English.

References

- [1] Grimson, W.E.L. and Lozano-Perez, T., "Model-based recognition and localization from sparse range and tactile data", *Int. J. Robotics Res.* Vol. 3, pp. 3-35, 1984.
- [2] Grimson, W.E.L. and Lozano-Perez, T., "Localizing overlapping parts by searching the interpretation tree", *IEEE Trans. Patt. Anal. Mach. Intell.* Vol. 9, pp. 469-482, 1987.
- [3] Lowe, D.G., "Three-dimensional object recognition from single two-dimensional images", *Artif. Intell.* Vol. 31, pp. 355-395, 1987.
- [4] Ullman, S., "An approach to object recognition: Aligning pictorial descriptions", MIT AI Lab Memo 931, 1987.
- [5] Huttenlocher, D.P., "Three-dimensional recognition of solid objects from a two-dimensional image", MIT AI Lab Memo 1045, 1989.
- [6] Forsyth, D.A., Mundy, J.L., Zisserman, A., Coelho, C., Heller, A., and Rothwell, C.A., "Invariant descriptors for 3D object recognition and pose", *IEEE Trans. Patt. Anal. Mach. Intell.* Vol. 13, pp. 971-991, 1991.
- [7] Frosini, P., "Measuring shapes by size functions", *Proc. SPIE on Intelligent Robotic Systems*, Boston (1991).
- [8] Frosini, P., "Discrete computation of size functions", *J. Combin. Inf. Syst. Sci.* Vol. 17, No. 3-4, pp. 232-250, 1992.
- [9] Verri, A., Uras, C., Frosini, P., and Ferri, M., "On the use of size functions for shape analysis", *Biol. Cybern.* Vol. 70, pp. 99-107, 1993.
- [10] Uras, C. and Verri, A., "On the Recognition of the Alphabet of the Sign Language through Size Functions", *Proc. XII Int. Conf. Patt. Recog.*, Jerusalem (1994), Vol. II, pp. 334-338, IEEE Computer Society Press.
- [11] Uras, C. and Verri, A., "Studying shape through size functions", In: O, Y., Toet, A., Foster, D., Heijmans, H., Meer, P. (eds.), *Shape in Picture*, NATO ASI Series F Vol. 126, pp. 81-90, Springer-Verlag, Berlin Heidelberg, 1994.
- [12] Uras, C., Verri, A., "Computing Size Functions from Edge Maps", Accepted for publication: *Int. J. Computer Vision*.