Sign Language in the Intelligent Sensory Environment

Ákos Lisztes*, Ákos Antal, Andor Gaudia*, Péter Korondi***

Budapest University of Technology and Economics
*Department of Automation and Applied Informatics, Budapest, Hungary
** Department of Mechatronics, Optics and Instrumentation Technology

Abstract: It is difficult for most of us to imagine, but many who are deaf rely on sign language as their primary means of communication. They, in essence, hear and talk through their hands. This paper proposes a system which is able to recognize the signs using a video camera system. The recognized signs are reconstructed by the 3D visualization system as well. To accomplish this task a standard personal computer, a video camera and a special software system was used. At the moment the software is able to recognize several letters from the sign language alphabet with the help of color marks. The sign language recognition is a function of an Intelligent Space, which has ubiquitous sensory intelligence including various sensors, such as cameras, microphones, haptic devices (for physical contact) and actuators with ubiquitous computing background.

Keywords: sign language, deaf, image processing, marker, CCD, camera, OpenGL, WISDOM, recognition, reconstruction, intelligent space

1 Introduction

Sign languages are visual languages. They are natural languages which are used by many deaf people all over the world, e.g. HSL (Hungarian Sign Language) or ASL (American Sign Language). In 1988, the European Parliament passed a resolution stating that the sign languages of the European Community should be officially recognized by the Member States. To date only Sweden, Greece, Portugal and Finland have ratified this resolution into legislation. Nevertheless, the sign languages of Europe are living and developing languages. They can be characterized by manual (hand shape, hand-orientation, location, motion) and non-manual (trunk, head, gaze, facial expression, mouth) parameters. Mostly one-handed and two handed signs are used. Sign languages occupy a 3D signing space usually considered to be within the arms reach horizontally and from the top of the head to the waist [1].

In sign language the hands convey most of the information. Hence, vision-based automatic sign language recognition systems have to extract relevant hand features from real life image sequences to allow correct and stable gesture classification. Using only the position information of the hands [2] or additionally their 2D orientation and some simple 2D shape properties [3] [4], recognition results of up to 97.6% for 152 different single gestures have been achieved so far. Nevertheless, more detailed information about the hand shape becomes necessary if a general recognition system is supposed to be constructed with a significantly larger gesture vocabulary and with the aim of distinguishing between so-called minimal pairs of signs, which merely differ in hand shape.

Most recent gesture recognition applications use either low level geometrical information [5], 2D models of the contour [6] or appearance models [7] [8] to describe the 2D hand image. The resulting parameters are appropriate to distinguish between a few clearly separable hand shapes from a fixed viewing angle. Although there are only a small number of different hand shapes employed during the performance of sign language, these can appear at any posture, creating a large amount of different hand images. The problem is that the degree of similarity between these 2D appearances does not correspond well to the degree of similarity between the corresponding constellations of the real hand. Because of this, an improvement of recognition results through direct usage of any extracted 2D features cannot be expected.

The most suitable information for gesture classification are the real hand parameters which include the finger constellation and the 3D hand posture. Several approaches have been published to extract these from images without the usage of any aids like marked gloves. In [9] the given image is compared with a large database of rendered hand images using a set of similarity criteria. The natural hand parameters of each image are included in the database. Another approach is to detect relevant points like the finger tips in the image and to adapt a simple 3D model to the positions found [10]. Furthermore, different methods have been developed to construct a deformable 3D model of the hand and to adapt it to the image content.

The main goals of the proposed system are the following:

- To recognize some very basic elements of manual sign language, based on 2D visual information.
- To visualize the positions of the fingers with a 3D graphical Hand Simulation.
- To make connection with other users in order to exchange sign language data over the Internet.

This paper is structured as follows: Section 2 gives a general description about our system. Section 3 introduces the details of the system. Finaly in Section 4 the results are summarized and further plans are explained.

2 General Description

2.1 Intelligent Space

The main problem is the bandwith of the existing LAN applications. We can acquire bigger amout of information than we can transfer through the fastes computer network line. To reduce the data transfer the intelligence of the system must be distributed. A conceptual figure of the Intelligent Space with ubiquitous sensory intelligence is shown in Figure 1.

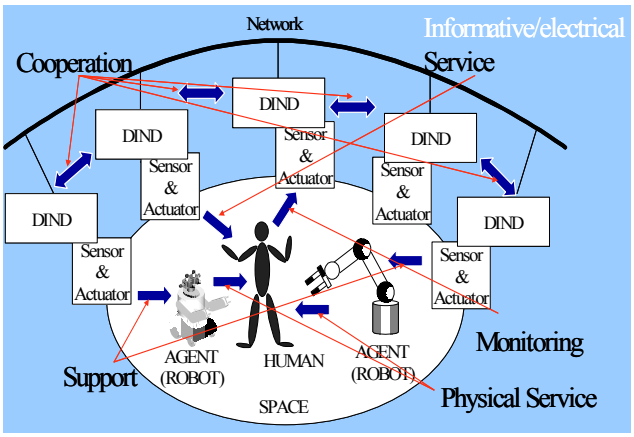
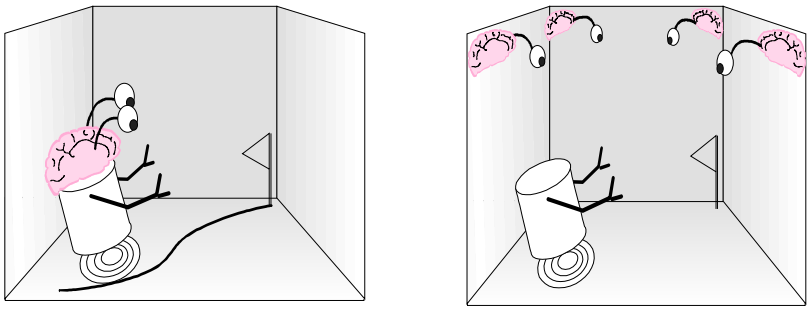


Figure 1
Intelligent Space Concept

The Ubiquitous Sensory Intelligence is realised by Distributed Intelligent Networked Devices [11], robots, which are physical agents of the Intelligent Space, and Human. In the Intelligent Space, DINDs monitor the space, and achieved data are shared through the network. Since robots in the Intelligent Space are equipped with wireless network devices, DINDs and robots organize a network. The Intelligent Space based on Ubiquitous Sensory Intelligence supply information to the Human beings and it can help them physically by using robot agents. Conventionally, there is a trend to increase the intelligence of a robot (agent) operating in a limited area. The Ubiquitous Sensory Intelligence concept is the opposite of this trend. The surrounding space has sensors and intelligence instead of the robot (agent). A robot without any sensor or own intelligence can operate in an Intelligent Space. The difference of the conventional and Intelligent Space concept is shown in Figure 2. There is an intelligent space, which can sense and track the path of moving objects (human beings) in a limited area. There are some mobile robots controlled by the intelligent space, which can guide blind persons in this limited area. The Intelligent Space tries to identify the behaviour of moving objects (human beings) and tries to predict their movement in the near future.



Conventional concept

Ubiquitous Sensory Intelligence concept

Figure 2
Comparison of conventional and Ubiquitous Sensory Intelligence concept

Using this knowledge, the intelligent space can help avoiding the fixed objects and moving ones (human beings) in the Intelligent Space. A mobile robot with extended functions is introduced as a mobile haptic interface, which is assisted by the Intelligent Space. The mobile haptic interface can guide and protect a blind person in a crowded environment with the help of the Intelligent Space. The Intelligent Space learns the obstacle avoidance method (walking habit) of dynamic objects (human beings) by tracing their movements and helps to the blind person to avoid the collision. The blind person communicates (sends and receives commands) by a tactile sensor. The prototype of the mobile haptic interface and simulations of some basic types of obstacle avoidance method (walking habit) are presented. Some other Intelligent Space projects can be found in the Internet [12, 13, 14, 15].

2.2 Sign Language Recognition as a Type of DINDs

We can use as a definition: A space becomes intelligent, when Distributed Intelligent Network Devices (DINDs) are installed in it [11]. DIND is very fundamental element of the Intelligent Space. It consists of three basic elements. The elements are sensor (camera with microphone), processor (computer) and communication device (LAN). DIND uses these elements to achieve three functions. First, the sensor monitors the dynamic environment, which contains people and robots. Second, the processor deals with sensed data and makes decisions Third, the DIND communicates with other DINDs or robots through the network. Figure 3 hows the basic structure of human decision and DIND.

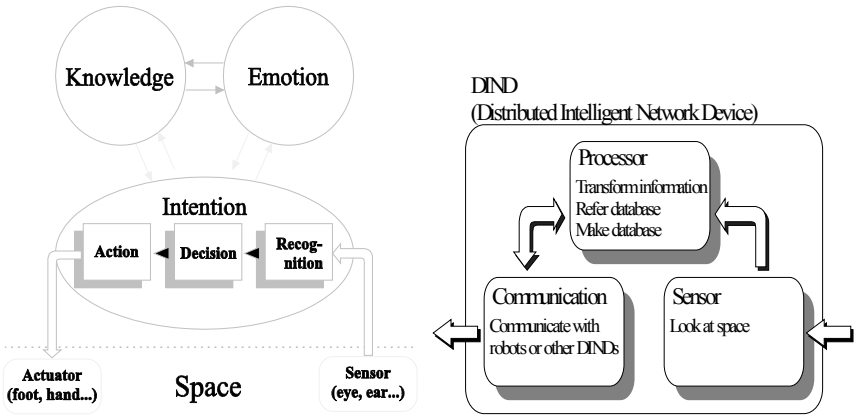


Figure 3
Fundamental structures of human decision and DIND

2.3 The Alphabet of the Sign Language

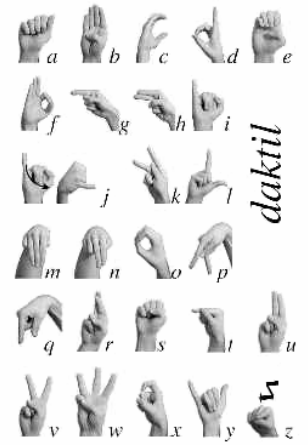


Figure 4
Alphabet of the Sign Language

A, B, E, I, U, V and W letters can be recognized without examining the position of the thumb (see in Fig. 5-11).

Figure 5
Camera picture of letter A

Figure 6
Camera picture of letter B

Figure 7
Camera picture of letter C

Figure 8.
Camera picture of letter I

Figure 9
Camera picture of letter U

Figure 10
Camera picture of letter V

Figure 11
Camera picture of letter W

To recognize all of the letters two or more cameras and a rather complicated marking system is needed so we limited our system's facilities to recognize those letters which neglect the use of thumb.

2.4 Marker Points

For finding the joints of the fingers on the picture of the camera we had to sign them. First we mark the joints with red points but in this case two problems appeared: which point belongs to which finger and this point an inner point or an outer point. To solve this problem, different colors would have to be used as it shown in Figure 12. and 13., but in that case the finding of the joints would be more difficult, because there are more colors.

Figure 12
The inner color points of the prepared glove

Figure 13
The outer color points of the prepared glove

2.5 System Elements

We designed a system which uses a CCD video camera to recognize the finger positions(see in Fig. 14) To help our system we used specially marked gloves as described in section 2.4. The image of the observed hand is transferred to a standard desktop computer - using a video digitalisator card - where an image is analyzed by our image recognition program. If the analyzed hand is recognized as a sign for deaf people the corresponding letter is displayed on the screen. While displaying a recognized letter the program is able to display the signs using a 3D hand visualization software (see in Fig. 15). Multiple hand visualization programs can be connected to the recognition software through the Internet.

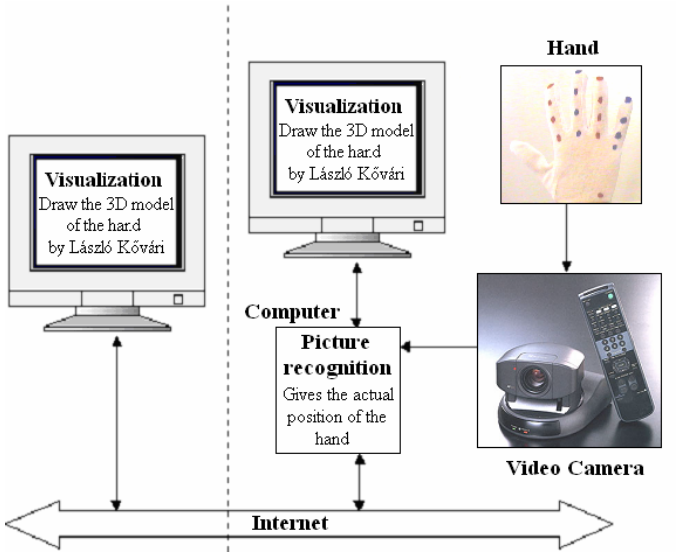


Figure 14
Block diagram of the sign recognition and sign reconstruction system

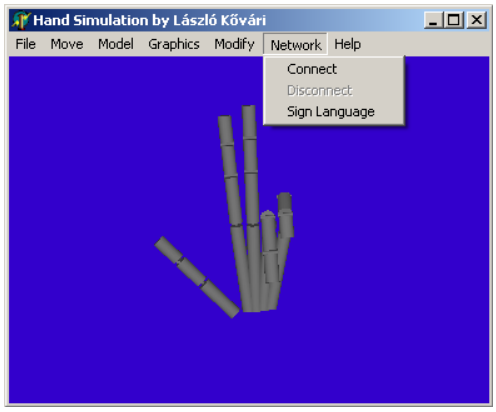


Figure 15
The Hand Simulation software

3 Detailed Description

The video camera system produces a standard composite video signal which is recognized by a video digitalisator card. The captured image is transferred to the image recognition software 10 times per second.

The image recognition software scans the image pixel by pixel using the algorithm described as follows:

3.1 Algorithm for Finding the Marker Points

The problem at finding the marker point is that they are bigger than one pixel. That is why after finding for example a red pixel it must be decided whether this pixel belongs to an earlier found joint or it belongs to a new one. This problem is well known, there exist many so-called edge detection algorithms. But these algorithms are quite sophisticated, usually not too fast, and they know much more than what we need. This is why we decided to use a self-created algorithm, which is worse than any of these edge detection algorithms, but enough for our purposes. We are interested only about the en frame rectangle of the joints. So the algorithm works in the following way:

The computer starts to scan the picture, and after it finds a red point the following two things can happen:

- If this point is not further to the middle of a previously found rectangle than a given limit, then we suppose, the point belongs to that rectangle. In this case we increase the size of that rectangle, so that it encloses this new point too.
- If it is not enough close to any other rectangles, then we suppose that it is a new joint, and we make a new rectangle for it.

To find the color marks the program gets RGB component of each pixel. Instead of looking for a concrete color we are examining a range of colors. We can set this range with the help of the scroll bars or edit boxes which can be seen in the left bottom quarter of the program screen. (see in Fig. 16) It is important to the lighting conditions are to be nearly constant so the RGB components of the color points change in narrow range.

3.2 Evaluation of the Positions of the Points

First of all a red dot at the lower right corner of the palm of the hand is found; this will be the base point of the further analysis. Four different groups of the points are created, one for each finger. The points are separated into those groups by reading their color code and position according to the base point. Each group should contain 4 points. If some points are missing, they are automatically added by the program.

Each group is examined by the algorithm and the angles of the joints' are calculated. These calculated angles are compared to the stored values – angle values for letters A, B, E, I, U, V and W – and the result of the comparison will be the recognized letter.

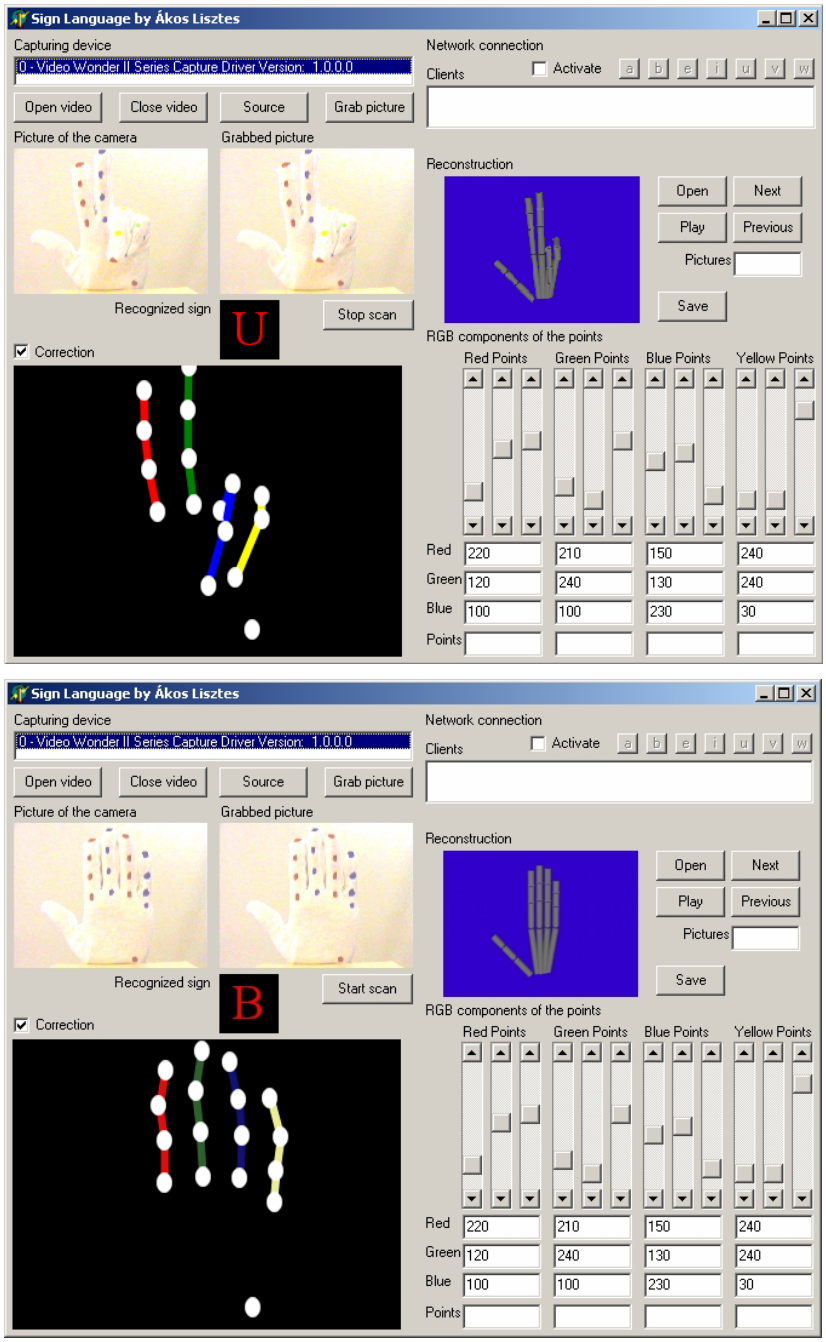


Figure 16
Sign Recognition of letter U and B

3.3 The Sign Reconstruction Software

A 3D hand animation software was developed to give a visual feedback of signs [16] (see in Fig. 15). Visual feedback has the aim of giving a quite real representation of the environment, and although at the moment this environment is far from complete, the program enables the user to wonder in full three-dimensional space.

4 Results and Further Plans

4.1 Performed Tasks

A Sign Recognition program was written in Delphi 6 which can detect the positions of the fingers' expect the thumb and can recognize the A, B, E, I, U, V and W letters. A Hand Simulation Program correctly reconstructs the recognized signs.

A network connection between the sign recognition software and the hand simulation software has been established and tested; It was able to transfer the recognized signs through the Internet.

Using this technique the network load was significantly lower than sending the whole video stream through the network.

4.1.1 Performance Test Results

We used the following system components for testing:

- Sony EVI D30/D31 Pan/Tilt/Zoom CCD camera
- Genius Video Wonder Pro III video digitalisator card
- GeForce4 MX440 video card
- Intel Pentium II 366 MHz processor
- 128 Mb RAM
- Windows 95 OSR2

The signal produced by the CCD camera is a standard composite video signal; This can be digitalised using a video digitalisator card. The video signal contains 25 frames per second but the digitalisator card was able to digitalise 16-25 frames per second. Our algorithm was able to process 10 frames per second (average) at 100% CPU utilization.

4.1.2 Functionality Test Results

The efficiency of the algorithm was around 90%. We tested the algorithm in ideal light conditions and the signs were shown for at least 5 seconds. No mistakes were made during the test only few times the program wasn't able to recognize the sign.

4.2 Further Plans

A video recognition system should be extended to handle the thumb and to recognize every used sign. Rendundancy of the recognition system should be improved to tolerate different light conditions as well.

The marker system should be simplified or neglected and the recognition system should work without any additional help. Later this system can be integrated to Intelligent Space systems as a DIND (Distributed Intelligent Network Device).

Conclusion

As we know there is a growing need for space monitoring systems all over the World but these systems' need larger and larger operating staff as well; The tendency in the World is that salaries are growing and to employ a security staff to operate these monitoring systems becomes uneconomical. Contrarily computer systems are getting cheaper and cheaper so they might substitute humans in this field.

We think that there is a growing need for cheap intelligent space systems which can be easily taught to make difference between usual and unusual situations. Intelligent space systems are able to communicate with humans in the monitored area, but we have to make them as flexible as possible. The proposed system fits into the Intelliget Space because it can be used as the interface between deaf people and computers.

Acknowledgement

The authors wish to thank the JSPS Fellowship program, National Science Research Fund (OTKA T034654, T046240), Hungarian-Japanese Intergovernmental S & T Cooperation Programme and Control Research Group of Hungarian Academy of Science for their financial support.

References

[1] Liddel, S. K. and R. E. Johnson. American Sign Language The phonological base. In: Sign Language Studies, 64: 195-277, 1989

[2] Yang, M.-H., N. Ahuja, and M. Tabb, "Extraction of 2D Motion Trajectories and its application to Hand Gesture Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(8), 2002, pp. 1061-1074

[3] Starner, T., J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer based video", IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(12), 1998, pp. 1371-1375

[4] Akyol, S., "Non-intrusive recognition of isolated gestures and signs", Ph.D. thesis, Achen University (RWTH), Germany, 2003

[5] Hardenberg, C. v. and F. Bérard, "Bare-Hand Human-Computer Interaction", ACM Workshop on Perceptive User Interfaces, Orlando, Florida, Nov. 2001

[6] Bowden, R. and M. Sarhadi, "Building temporal models for gesture recognition", BMVC, 2000, pp. 32-41

[7] Martin, J. and J. Crowley, "An Appearance-Based Approach to Gesture recognition", ICIAP (2), 1997, pp. 340-347

[8] Triesch, J. and C. v. d. Malsburg, "Robust Classification of Hand Postures against Complex Backgrounds", IEEE International Conference on Automatic Face and Gesture Recognition, 1996, pp. 170-175

[9] Athitsos, V. and S. Sclaroff, "An Appearance-Based Framework for 3D Hand Shape Classification and Camera Viewpoint Estimation", Technical Report 2001-22, Boston University, 2001

[10] Nölker, C. and H. Ritter, "GREFIT: Visual Recognition of Hand Postures", Gesture-Based Communication in Human-Computer Interaction, Springer, 1999, pp. 61-72

[11] J.-H. Lee and H. Hashimoto, "Intelligent Space - Its concept and contents – ", Advanced Robotics Journal, Vol. 16, No. 4, 2002

[12] Hashimoto Laboratory at The University of Tokyo http://dfs.iis.u-tokyo.ac.jp/~leejooho/ispace/

[13] MIT Project Oxygen MIT Laboratory for Computer Science MIT Artificial Intelligence Laboratory http://oxygen.lcs.mit.edu/E21.html

[14] Field Robotics Center, Carnegie Mellon University Pittsburgh, PA 15213 USA http://www.frc.ri.cmu.edu/projects/spacerobotics/publications/intellSpaceRobot.pdf

[15] [Institut of Neuroinformatics Universität/ETH Zürich Winterthurerstrasse 190 CH-8057 Zürich http://www.ini.unizh.ch/~expo/2_0_0_0.html

[16] Kővári, L.: Visual Interface for Telemanipulation, Final Project, Budapest University of Technology and Economics, 2000

[17] T. Akiyama, J.-H. Lee, and H. Hashimoto, "Evaluation of CCD Camera Arrangement for Positioning System in Intelligent Space", International Symposium on Artificial Life and Robotics, 2001