Video-Based Recognition of Fingerspelling in Real-Time

Kirsti Grobel and Hermann Hienz

Lehrstuhl für Technische Informatik, RWTH Aachen
Ahornstraße 55, D - 52074 Aachen, Germany
e-mail: grobel@techinfo.rwth-aachen.de

This paper presents a visual prototype for the recognition of 31 different letters and 7 additional handshapes used in German sign language. Fingerspelling is used in sign languages to spell names and words. A letter is defined by one handshape, hand orientation and partly motion within the German fingeralphabet. Using the presented prototype a signer performs fingerspelling in front of a single video camera. The signer has to wear a coloured glove, but is nearly free in choosing the clothes. The letters are recognised with an accuracy of 93% by the system.

Keywords: Image Processing, Hand-Gestures, Human-Computer Interface

1 Introduction

Gestures are part of everyday natural human communication, where they e.g. emphasize the spoken language spontaneously and instinctively. Deliberate and controlled gestures are the primary medium of visual-spatial languages, such as sign language, which is the native language of the deaf.

Video-based recognition of gestures can be used in a wide field of applications. In medicine motion analysis systems are used to acquire motion data and to diagnose motor distortions. Controlling robots or machines, through hand-gestures offers a new step in human-computer interaction. Such an gesture-based input device provides an efficient interaction technique which could be for great benefit of those with physical impairments. Manipulating three dimensional objects in virtual reality leads to a further interesting application area. Huang [4] gives an overview over current developments in hand gesture recognition. [2, 3, 9] present some examples for recognition of fingerspelling in particular.

Our long term goal is the development of a video-based sign language recognition system. Such a system will consist of several components to recognise handshape, motion, position and orientation of the hand.

Fingerspelling is used within sign languages, e.g. for spelling names and words [1]. A letter consists of one handshape, handorientation, and partly motion. Figure 1 shows some examples of different letters. For the fingeralphabet, the most distinguishing parameter is handshape. The orientation is mainly holding up the hand and partly turning the hand into one direction, e.g. H and X (figure 1). Motion is for example included in the letter Z, where the handshape of letter D is used and the outline of the letter Z is copied.

The developed system has to fulfil several requirements. To offer a user friendly system, real-time recognition is required and performed letters have to be recognised with high accuracy. The whole signing space must be observed in order to develop a sign language recognition system in the near future.

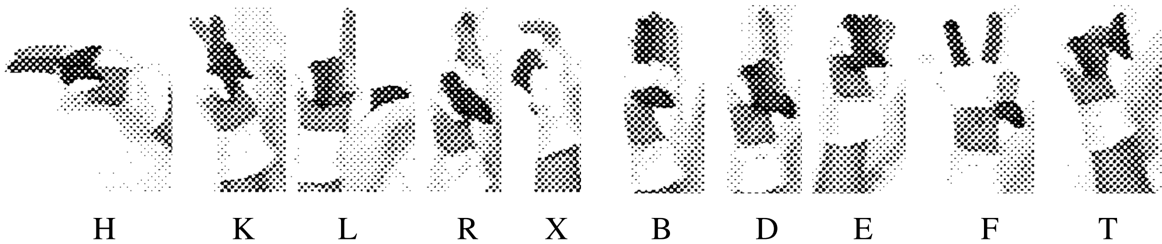

Figure 1: Examples of letters

To meet these demands, additional hardware, such as a modular image processing system, has been added to the PC. The signer has to wear a coloured glove to make the segmentation of the image easier and with that to speed up the recognition process. Only a single video camera observes the whole signing space and allows a simple setup of the system and minimises the required computing capacity. The lack of information due to the two dimensional view of the three dimensional scene is mainly solved by using relative features of handshapes instead of e.g. reconstructing the exact hand configuration [3, 8].

In this paper a prototype for real-time video-based recognition of 31 letters (including the German Ä, Ö, Ü, and ß), partly with motion and variations in handorientation and 7 additional gestures used for fingerspelling and controlling the input of letters is presented.

2 Recognition of Fingerspelling

The following chapter describes the structure of the whole system including distribution of modules through the used hardware. The image processing parts are presented in more detail.

2.1 System Overview

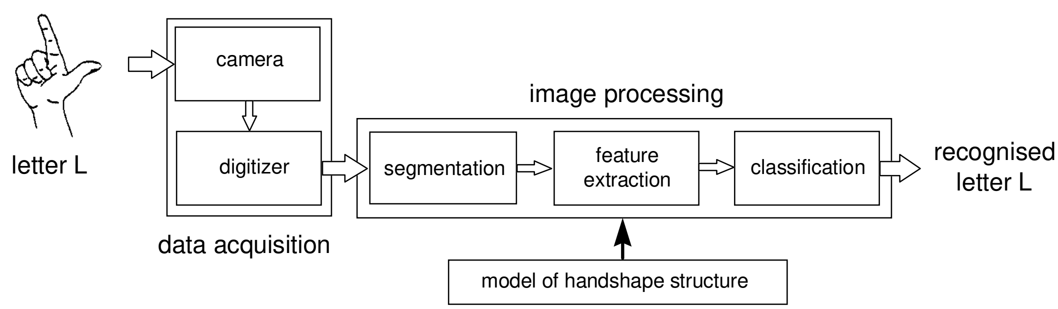

Figure 2: Components of the recognition system

The system consists of three functional units, such as data acquisition, image processing, and model of handshape structure (see figure 2).

The first unit acquires image data by using a CCD-video camera. After grabbing and digitizing, the image is computed by the second unit which provides image processing routines for segmentation, feature extraction, and classification. The image processing [5,6] routines depend on the chosen underlying handshape structure model, which contains three components: the coloured areas, the features of these areas, and the relations between these areas. The coloured areas are put onto a cotton glove and support the segmentation. The features of these areas and their relations are computed by the feature extraction module. These features form the input for the classifier, which calculates the output, the recognised letter of the system.

2.2 Image Segmentation

The segmentation separates the interesting coloured areas within an image from the background. Each finger of the glove is fully coloured and has its own colour (figure 3). Coloured areas are chosen because they allow an easy and fast segmentation process, which is important with respect to computing time. Using different colours for each area provides an easy matching between the colour and the finger, respectively the ball of the hand. A simple and therefore fast segmentation method, a thresholding algorithm has been implemented. Within the thresholding algorithm, the ideal colour is assigned to all pixels with the value according the definite rectangular solid in the three-dimensional colour space.

2.3 Feature Extraction

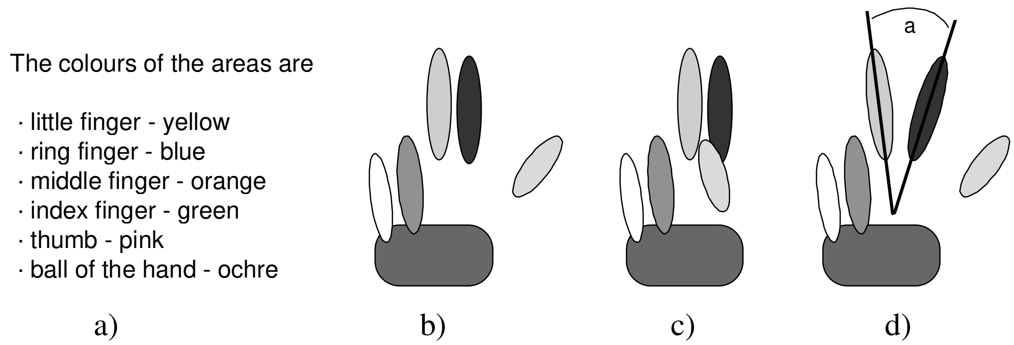

Figure 3: a) Colour coding and b)-d) relations between the coloured areas

The feature extraction module calculates the interesting features of the areas. Therefore the segmented image is scanned pixel by pixel, and the In-Out-Code (I/O-Code) for these areas are stored. The further computation bases on this data.

The most important features for the recognition of letters are colour, visibility, size, centre of gravity, curvature, circumference, and orientation. Using those features, relations between areas respectively fingers are calculated like: Contact of two areas? yes / no, manner of contact of two fingers, relative size of one area compared to one another, distance of centres of gravity (COG), angle between two fingers, and relative length of one finger.

An important characteristic is the *contact of two areas*. If, e.g., the ring finger is bent, contact occurs between the finger and the ball of the hand (figure 3b). Furthermore the *manner of contact* between two fingers is used for the description of handshapes. In figure 3b the index finger and the middle finger have contact on the whole length of the fingers. In figure 3c the thumb touches the index finger only with its tip. The computation of the *angle between two areas* (figure 3d) is based on the orientation of themselves.

For recognition of motion an additional feature is necessary: the motion of the whole hand. Motion vectors for a sequence of images are determined on the basis of the COGs of the whole hand.

Figures 4a-b give an example for the image processing steps until the feature extraction. Figure 4a shows an input image, where the AP-handshape is performed. Figure 4b shows a segmented part of the input image with homogeneous regions, additionally the extracted areas and the centres of gravity are drawn in.

Figure 4a: Input image

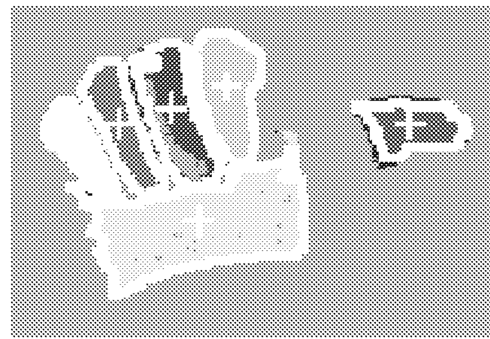

Figure 4b: Part of the segmented image: Extracted areas and centres of gravity

2.4 Classification of Letters

A rule-based module is finally used for classification. The principle order of the analysis is firstly the analysis of visibility, which is an easy to calculate feature, and secondly, the analysis of contact. It is preferred to analysis contact on the whole length or no contact, because this is more reliable to calculate than the contact of tips. The analysis of these two parts is often sufficient for discriminating the letters. In case of variations in orientation, the relative location of COGs one to another is analysed. If it is necessary to look at motion, an additional calculation of motion vectors takes place.

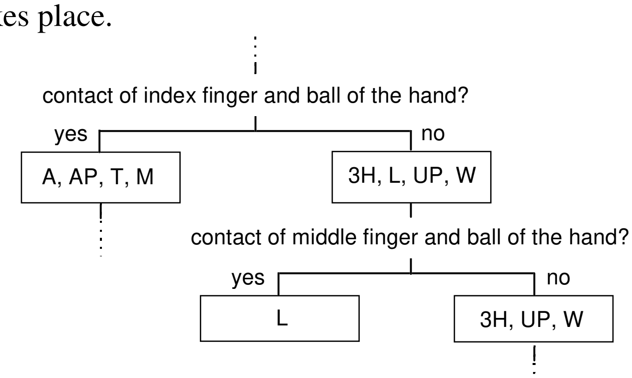

Figure 5: Part of the classification rules

During classification the whole pool of letters is subdivided into smaller sets of letters. Figure 5 presents a small part of the whole rule-set with the final leaf for the letter L. After decisions concerning the visibility of the ball of the hand, contact of the little finger with the ball of the hand, and contact of the thumb with either ring or middle finger, it is asked whether the index finger has contact to the ball of the hand. The pool of remaining letters is divided into two sets A, AP, T, M and 3H, L, UP, W. The next question ‘contact of middle finger and the ball of the hand?’ separates finally letter L from the rest.

2.5 Image Processing Hardware

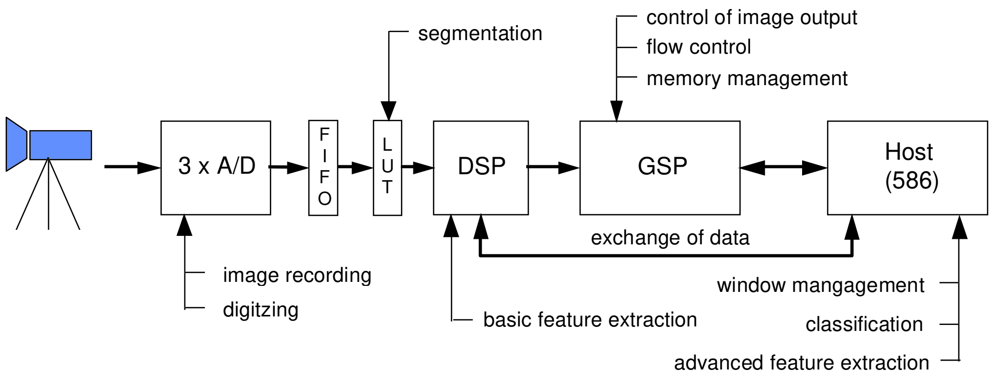

Figure 6: Structure of the used hardware and the distribution of tasks among the different units

The used hardware (figure 6) consists of the host, which is a PC equipped with a Pentium 90 processor and a modular image processing system from Matrix Vision connected over the ISA bus. The image processing system contains a basic board with a graphical signal processor (GSP), a framegrabber, and a digital signal processor (DSP), which is a pipelining module mounted right after the framegrabber. The framegrabber includes three analog to digital converters, a first-in-first-out buffer(FIFO) for buffering image data, and a Look-Up-Table (LUT).

The CCD-camera is connected to the A/D-converters, which digitize the analog RGB-signal. The FIFO buffers the incoming stream of image data. By passing the LUT, the image is segmented. A new value is assigned to each pixel regarding the coloured areas and the background. The DSP calculates and stores the I/O-Code of the interesting areas on local memory. Basic features of these areas, as size and COG, are calculated and stored as well on the DSP.

The GSP is responsible for internal memory management, initialisation and for functioning of the framegrabber. It controls the flow of image data among the unit and controls the output of image data on monitors. The host directly collects data from the local memory on the DSP-board through the ISA bus and calculates advanced features and classifies the letter. [7]

3 Evaluation of the System

The analysed image size is 768 pixel *512 pixel interlaced in true colour. The used equipment (besides computer hardware) consists of three halogen projectors (500 W), and one RGB-camera. The three projectors and the camera are arranged as follows: Two projectors are located on the left and right hand side of the subject and an additional projector is put on the floor. The camera is located a bit underneath the eye-level of the subject. During evaluation the subject has to wear the coloured glove. As a rule subjects can wear their normal clothes.

The system has been tested with five different subjects performing two series. The system calculates fourteen frames per second. The required computing time is real-time for the subjects. The recognition results are presented in table 1. An average recognition rate of 93% was achieved. It is interesting that the results reached by the subjects 1 and 2, who are very familiar with the fingeralphabet, are better than the others.

The reached recognition rate is good. Looking at the series in detail, it is noticeable that mainly the letters P, Q, BP, and BH are recognised wrongly. Partly the letters P and Q, where thumb, indexfinger, and middlefinger are visible, are recognised as letter C, where the middlefinger is not visible. The position of the thumb is the only difference between the B (thumb touching the palm of the hand), BH (thumb beside the palm), and BP (thumb stretched). For both parts of the rule-based module, a readjustment of firstly the analysis of visibility feature of the middlefinger and secondly a readjustment of the contact feature of the thumb should eliminate the misclassification.

Table 1: Recognition result

	Subject1 series 1/2	Subject2 series 1/2	Subject3 series 1/2	Subject4 series 1/2	Subject5 series 1/2
Rate in %	94,7 94,7	97,4 94,7	94,7 89,5	86,8 89,5	92,1 92,1
aver.p.pers.	94,7 %	96,1 %	92,1 %	88,2 %	92,1 %
aver.overall	93 %				

4 Summary and Future Work

In this paper an image processing system for recognition of fingerspelling has been presented. The system operates in real-time and uses a single video camera. A recognition rate of about 93% is reached. The use of a coloured glove has been accepted by the users. A higher recognition rate will be reached by modifying the feature extraction module.

Since the goal is recognition of sign parameters of isolated signs used in sign languages, the next steps are to expand the system for full motion analysis used within sign languages and to look closer at recognition of handshapes in motion.

5 References

[1] Boyes-Braem, P. (1992). *Einführung in die Gebärdensprache und ihre Erforschung*. Hamburg: Signum-Verlag, 1992.

[2] Cui, Y. and J.J. Weng (1996). *View-Based Hand Segmentation and Hand-Sequence Recognition with Complex Backgrounds.* In 13th Int. Conf. on Pattern Recognition, Vol. III, pp. 617-621, Vienna.

[3] Grobel, K. (1994). *Recognition of Fingerspelling from Video*, 6th Biennial Conf. of the Int. Soc. for Augmentative and Alternative Communication ISAAC, pp. 282-284, Maastricht.

[4] Huang, T.S. and V.I. Pavlovic (1995). *Hand Gesture Modeling, Analysis, and Synthesis*. Int. Workshop on Automatic Face- and Gesture Recognition, pp. 73-79, Zürich.

[5] Jähne, B. (1991). *Digitale Bildverarbeitung*. Springer Verlag.

[6] Jain, A.K. (1989). *Fundamentals of Digital Image Processing*, Prentice Hall, New Jersey.

[7] Offner, G. (1996). *Entwicklung eines video-basierten Systems zur Charakterisierung von Hand-Arm-Bewegungen der deutschen Gebärdensprache in Echtzeit*, Diploma Thesis, Lehrstuhl für Technische Informatik, RWTH Aachen.

[8] Rehg, J.M. and T. Kanade (1994). *Visual Tracking of High DOF Articulated Structures: an Application to Human Hand Tracking*, Lecture Notes in Computer Science, Vol. 801(Computer Vision - ECCV 1994), Springer Verlag.

[9] Uras, C. and A. Verri (1995). *Hand Gesture Recognition from Edge Maps*, International Workshop on Automatic Face- and Gesture Recognition, pp. 116-121, Zürich.