ADVANCES
IN
INTELLIGENT COMPUTING
AND
MULTIMEDIA SYSTEMS

Edited by:

Mahbubur Rahman Syed
Orlando R. Baiocchi
Department of Electrical and Computer Engineering
North Dakota State University, Fargo ND, USA

and

George E. Lasker
School of Computer Science
University of Windsor, Canada

Published by

THE INTERNATIONAL INSTITUTE FOR ADVANCED STUDIES
IN SYSTEMS RESEARCH AND CYBERNETICS

ADVANCES
IN
INTELLIGENT COMPUTATION AND MULTIMEDIA SYSTEMS

Edited by

Syed Mahbubur Rahman
Orlando R. Baiocchi
North Dakota State University
USA

&

George E. Lasker
Editor - in - Chief
University of Windsor
Canada

Canadian Cataloguing in Publication Data

International Symposium on Intelligent Multimedia and Distance Education (1999 : Baden-Baden, Germany)
Advances in intelligent computation and multimedia systems

Symposium held as part of the 11th International Conference on Systems Research, Informatics and Cybernetics, held Aug. 2-7, 1999 in Baden-Baden, Germany.
Includes bibliographical references and index.
ISBN 0-921836-80-5

1. Information technology - Congresses. 2. Artificial intelligence - Congresses. 3. Electronic data processing - Congresses. 4. Multimedia systems - Congresses. I. Rahman, Syed Mahbubur. II. Baiocchi, Orlando R. III. Lasker, G.E. (George Eric), 1935- . IV. International Institute for Advanced Studies in Systems Research and Cybernetics. V. International Conference on Systems Research, Informatics and Cybernetics (11th : 1999 : Baden-Baden, Germany.) VI. Title.

QA75.5.I65 1999 004 C99-9314874

ISBN 0-921836-80-5

Publisher: The International Institute for Advanced Studies
in Systems Research and Cybernetics
c/o Dr. George E. Lasker
University of Windsor
Windsor, Ontario
Canada N9B 3P4

TABLE OF CONTENTS

Synthesis of the sign language of the deaf from the sign video clips

Slavko Krapež and Franc Solina

Computer Vision Laboratory
Faculty of Computer and Information Science
University of Ljubljana
Trzaska 25, SI-1001, Ljubljana, Slovenia
{ slavkok, franc }@razor.fer.uni-lj.si

Abstract

The aim of the article is to present the system for automatic synthesis and visualization of the Slovene sign language of the deaf (SSL) sentences. Synthesis occurs by assembling individual previously filmed video clips of sign demonstrations (1800 most frequent signs). Video clips consistency and optimization of passages between individual video clips enables smooth joining of video clips on a personal computer in real time offering high quality video. Existing systems for sign language synthesis are based on an artificial sign language demonstrator and are merely prototypes. Our system uses a live person for signing and has a large set of sign video clips. Therefor it enables a user to perform a translation of different texts into sign language.

Keywords: multimedia, digital video, sign language of the deaf, computer vision.

1 Introduction

Deaf people have, as a marginal community, big troubles in communicating with hearing people. Usually they have a lot of problems even with such simple tasks as understanding of the written language. Deaf people have a lot of problems understanding written language but they are very skilled in using sign language, which is their native language. Our long-term goal is to build a system able to translate written language (books, newspapers, e-mails, letters, HTML documents, ...) and in connection with a speech recognition also speech (conversations, radio and TV programs, phone calls, ...) to sign language. So far we have achieved this goal just partially, since the translator from text to a sequence of sign names has not been implemented yet. It would be quite straightforward to implement such translator in the case of English language, because English and sing language are grammatically very similar. In most cases it consists of only word to sign translation. For languages with more difficult grammar one possible solution would be to build such a translator using a translator from that specific language to English language and then to sign language.

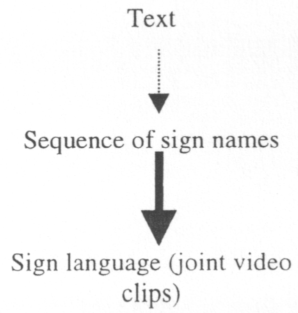


Figure 1: Translation from text to the sign language.

Sign language is a set of signs (Slovene sign language (SSL) contains about 4000 different signs). The sign in a sign language equals to a word in written language. Similarly, a sentence in written language equals to the sign sentence in the sign language.

This article introduces a new method for joining sign video clips in video clips of complete sign sentences. The joining can be processed on a personal computer in real time. We achieved smooth transitions between sign video clips using the information about the demonstrator palm positions while signing. For that purpose a special program for palm position extraction from video clips was written based on the methods from computer vision.

The described system contains a large set of individual sign video clips (1800), which make possible the translation of various texts to sign language. Words that are not in the system can be shown using video clips containing the finger alphabet. In principle any text can be shown in sign language.

2 Related work

The prevalent idea in the development of a system for sign language synthesis is to use a synthetic person [5][6]. Such systems are usually teached using Datagloves, which contains sensors for defining palm and finger position and orientation.

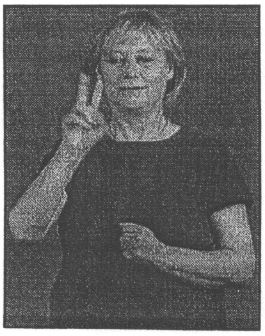
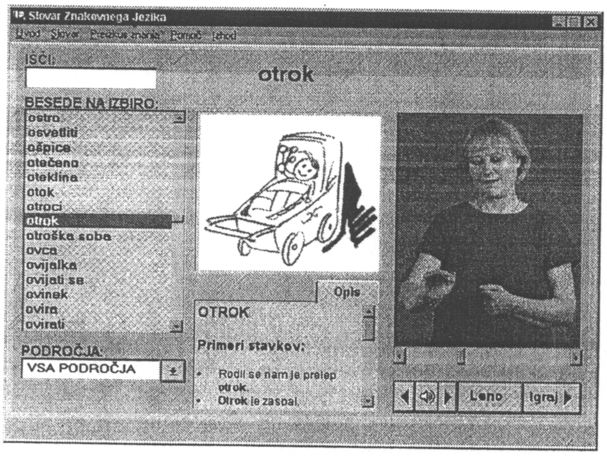


Figure 2: The appearance of real person demonstrator.

Since realistic animation of persons is still a time consuming task we decided to use video clips of a real person as a sign language demonstrator (figure 2). The synthetic demonstrator is no match for the real one in sense of appearance and consequently also of acceptance as a teaching tool.

3 Slovene sign language dictionary for the deaf

Slovene multimedia sign language dictionary [1][2] for the deaf on CD-ROM (figure 3) is composed of 1800 most frequent signs that are used by the deaf people in every day conversations.



otrok

- ostro
- osvetliti
- ošpice
- otečeno
- oteklina
- otok
- otroci
- otrok
- otroška soba
- ovca
- ovijalka
- ovijati se
- ovinek
- ovira
- ovirati

VSA PODROČJA

OTROK

Primeri stavkov:

- Rodil se nam je prelep otrok.
- Otrok je zaspal.

Figure 3: User interface of the Slovene sign language dictionary on CD-ROM.

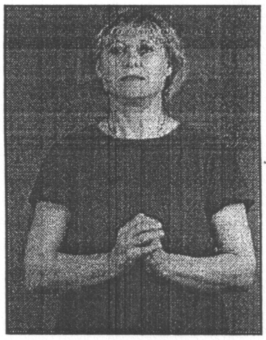


Figure 4: Example of one frame of a *sign video clip*. The demonstrator is in the start position.

Sign video clips from the dictionary are consistent (since all clips were filmed in one session using just one sign demonstrator, the same start and end positions of the arms on all video clips are used, see figure 4). Sign video clips from the CD-ROM can be therefore assembled into sign language sentences using a special process, which enables smooth transitions between individual sign video clips. The dictionary therefore represents basins for the sign language synthesizer.

4 Video clips joining

Joining of two digital video clips (each in his own file) should be done in such a way that the viewer perceives them as a single video sequence with smooth motion. The main idea is shown on figure 5.

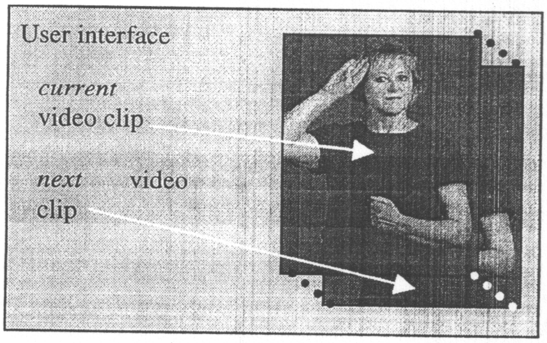


Figure 5: The joining of video clips.

The figure shows a form, containing two controls (for example *picture box*) in which video clips can be shown. The controls are laying one over another. In the upper control current video clip is played, when this ends, the lower video clip becomes the upper and it plays. That gives the impression of just one video clip being played and actual merging of files is not necessary.

4.1 Definition of the video clip joining problem

The problem of video clip joining will be discussed in the case of two video clips. This can be easily generalized to any number of video clips. Let us represent the first video clip we would like to show as a n_1 dimensional vector v_1

$$v_1 = \langle p_1, p_2, \ldots, p_{n_1} \rangle$$

where the components of the vector correspond to the pictures that form the video clip. n_1 is the number of pictures in the video clip. The second video clip, is represented as:

$$v_2 = \langle r_1, r_2, \ldots, r_{n_2} \rangle$$

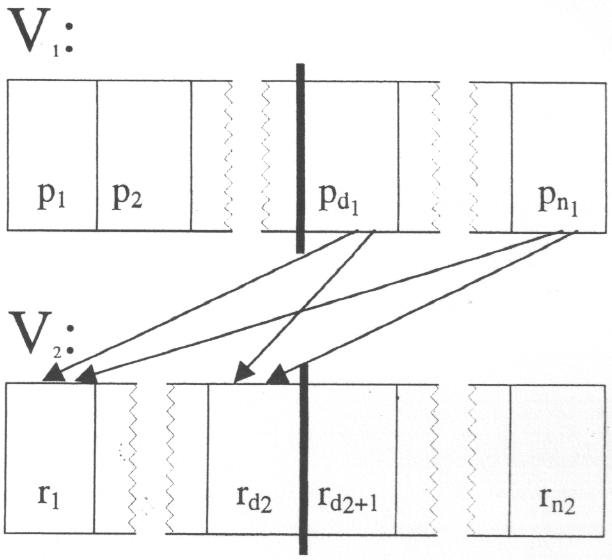
The joining can occur on any pair of pictures from v_1 and v_2. Let us indicate the point in which joining should occur with t_1 and t_2. They represent the indexes of pictures, which are, in the case of the first video clip, shown as the last and as the first in the case of the second video clip.

$$t_1 \in [\, 1 \ldots n_1 \,]$$
$$t_2 \in [\, 1 \ldots n_2 \,]$$

Let us introduce another two variables d_1 and d_2. Let d_1 be the index of the picture in the first video clip, from which on, the joining is allowed, d_2 is the index of the picture in the second video clip, till which the joining is allowed. With this constraints t_1 and t_2 can occupy the following values.

$$t_1 \in [\, d_1 \ldots n_1 \,]$$
$$t_2 \in [\, 1 \ldots d_2 \,]$$

The joining of video clips should be performed in such a way that the transition from one video clip to another wouldn't be noticed. Let us introduce a criteria function *dif(i,j)*. This function calculates the difference between the picture p_i and r_j depending on given criteria. The smaller is the function value the more both pictures resemble each other. The calculation of t_1 and t_2 is performed by finding the values of arguments for which the function *dif* has a minimum (figure 6).



$$\mathbf{arg\ min}\ (\mathit{dif}\ (\mathbf{t_1, t_2}))$$
$$\mathbf{t_1 \in [\, d_1..n_1 \,]}$$
$$\mathbf{t_2 \in [\, 1..d_2 \,]}$$

Figure 6: Determine the t_1 and t_2 values.

5 Automatic sign language video clips joining

For sign video clip joining we would like to achieve the following two goals:

- as smooth transitions as possible between video clips and
- minimal additional motion of the arms from the start position of the arms into the demonstration of the sign.

Sign video clips start and end with the demonstrator's arms in the start position (figure 4). In the start position the sign demonstrator joins the hands in the belt height.

5.1 Sign video clips

Sign sentences are generated from the sign video clips contained on the CD-ROM of the *Slovene sign language dictionary* (see section 3).

In order to perform optimal sign video clips joining some extra data is needed. Each video clip has a related file containing data about the position of demonstrator's palms performing a particular sign. Positions of palms are computed by a program *Tracking palm movements in video clips*, written in C++[4][3]. The input to that program is the sign video clip. The output is a file containing data about palm positions for every picture forming the video clip.

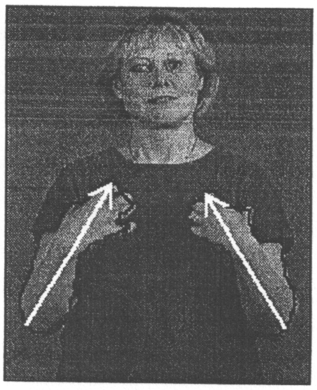


Figure 7: Arm vectors.

Figure 7 shows the arm vectors computed from one picture of a sign video clip. The palm position is the end point of the arm vector.

5.2 Suggested joining criteria

Conditions that enable smooth transition between video clips in the sense of similar palm positions are refereed to as joining criteria. We propose four possible criteria for determine the transition point between two sign video clips.

1. Palms in start position.
2. Palms outside the start position
3. Palms over the chest.
4. Palms close to each other.

1. Start position criterion

Using this criterion, transition should occur in points where arms are in the start position. Function *dif(i,j)* should therefore have a minimum in case when $i = n_1$ and $j = 1$.

2. Palms outside the start position criterion

Palms are outside the start position when they are near the start position but not joined or when their distance from the start position exceeds certain value. Function *dif(i,j)* has its minimum in the case of a picture pair (p_i, r_j), where palms on picture p_i are outside the start position for the last time and for the first time outside the start position on picture r_j.

3. Palms over the chest criterion

According to this criterion the transition between video clips is performed when palms are over the chest. Function *dif(i,j)* has its minimum for a picture pair (p_i,r_j) where the palms in the picture p_i are for the last time over the chest and where palms in picture r_j are for the first time over the chest region.

4. Palms close to each other criterion

Using this criterion the transition occurs in a point where the palms from the first and the second video clip are close to each other. Function *dif(i,j)* is defined as: *dif(i,j)* = d(i,j) + w(i,j) where d(i,j) represents the distance between palms from pictures p_i in r_j. $w(i,j) = ((i - d_1) + (d_2 - j))*K$, K is an empirical constant value (in our case $K = 5$).

5.3 Implementation of automatic sign video clips joining

At this point we have to say something about the use of arms in signing. Most signs are performed in the height of the chest and head. Signs performed in the height of the belt or lower are rare but still to consider. Approximately two thirds of signs are performed with one hand. The hand can either be the right or the left, that doesn't affect the meaning of the sign. In our case they are all performed with the right hand. We will call that signs *one-hand signs*. Approximately one third of signs are performed with both hands. We will call them *two-hand signs*.

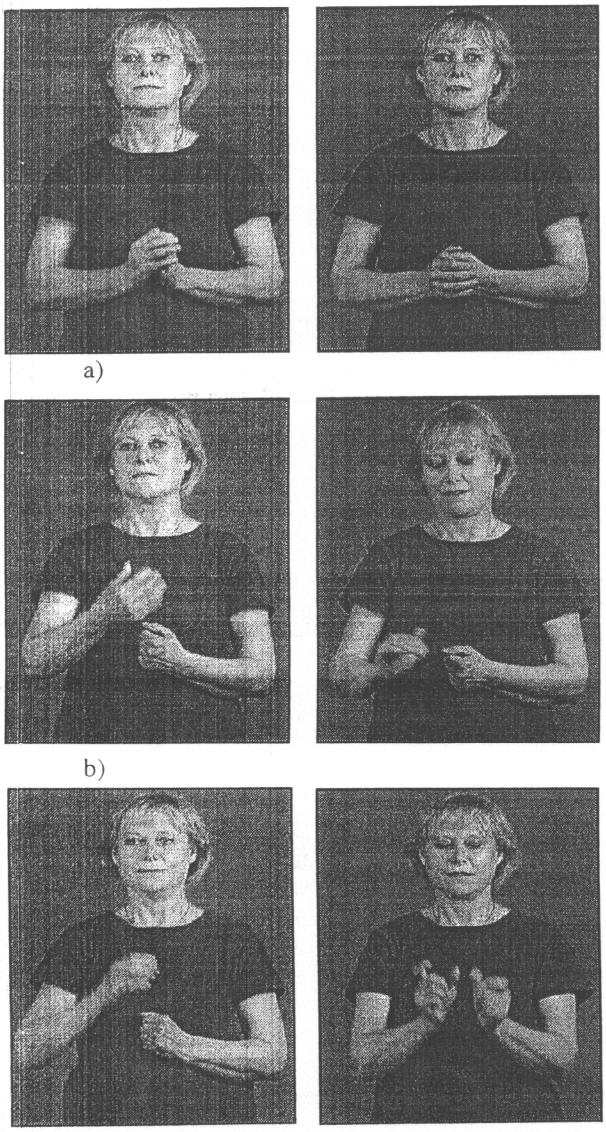
According to that, we propose automatic joining of video clips using all four criteria for joining. The use of a specific criterion depends on the type of the two signs we would like to join. The criterion for joining two sign video clips is chosen according to the following four possibilities. If the conditions enable us to use more criteria, we use the first that matches in order they are listed:

- signs define the end of the current and the beginning of the next sentence or a delay is needed => *start position criterion*
- one of the signs is performed in the belt height => *palms outside the start position criterion*
- exactly one of the signs is a two-hand sign => *palms over the chest criterion*

- both signs are one-hand signs or two-hand signs => *palms close to each other criterion*

6 Results of automatic sign video clip joining

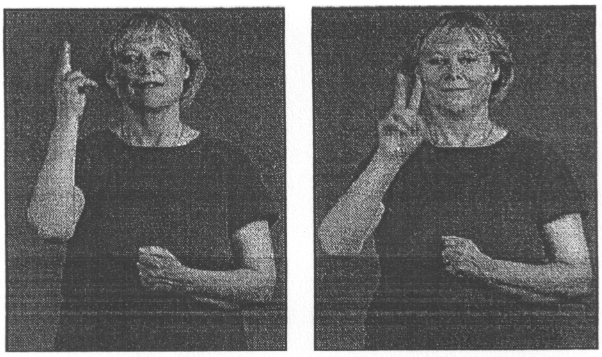
Let us have a look at some results of the transitions between two video clips obtained with the method of automatic video clip joining.



a)

b)

c)



d)

Figure 8: Shown are the last picture of the first and the first picture of the second video clip according to the a) *start position criterion*, b) *palms outside the start position criterion,* c) *palms over the chest criterion* and d) *palms close to each other criterion.*

Figure 8a shows the last shown picture of the current and the first shown picture of the next sign video clip using the *start position criterion*. That kind of transition is useful when we want to emphasize the end of the sentence or the end of the paragraph. It is useful also in the case where the values of the palm positions are not available.

Figure 8b shows the transition between two signs one of which is performed in the belt height. *Palms outside the start position criterion* is used in this case to prevent arms from being joint in the start position.

The results obtained on one-hand and two-hand signs using the *palms over the chest criterion* are shown on figure 8c.

Figure 8d shows the suggested transition between one-hand signs using the *palms close to each other criterion.*

The results of automatic sign video clip joining are very encouraging. This system allows us to join different kinds of sign video clips into sign sentences with smooth transitions between signs. The results of video clip joining are very good thanks also to the following two reasons:

- there are no visible delays in video clip playing between the transitions from one video clip to another and
- the human ability to form an impression of continuous transition between two pictures that are similar enough and shown quickly one after another.