Chapter in *Design and Management of Multimedia Information Systems*
Mahbubur Rahman Syed, Editor
Idea Group Publishing, Hershey PA, 2001, pages 268–281

Chapter XIII

Multimedia Dictionary and Synthesis of Sign Language

Franc Solina, Slavko Krapež, Aleš Jaklič
University of Ljubljana, Slovenia

Vito Komac
Zoom promotion, Ljubljana, Slovenia

1 Introduction

Deaf people, as a marginal community, may have severe problems in communicating with hearing people. Usually, they have a lot of problems even with such, for hearing people simple tasks, as understanding the written language. However, deaf people are very skilled in using the sign language, which is their native language. A sign language is a set of signs or hand gestures. A gesture in a sign language equals to a word in a written language. Similarly, a sentence in a written language equals to a sequence of gestures in a sign language.

In the distant past deaf people were discriminated and believed to be incapable to learn and think independently. Only after year 1500 first attempts were made to educate deaf children. An important breakthrough was the realization that hearing is not a prerequisite for understanding ideas. One of the most important early educators of the deaf and the first promoter of sign language was Charles Michel De L'Epée (1712–1789) in France. He founded the fist public school for deaf people. His teachings about sign language quickly spread all over the world.

Like spoken languages different sign languages and dialects evolved around the world. According to the National Association of the Deaf the American Sign Language (ASL) is the third most frequently used language in the United States, after English and Spanish. ASL has more than 4,400 distinct signs. The Slovenian sign language (SSL), which is used in Slovenia and also serves as a case study sign language in this chapter, contains approximately 4000 different gestures for common words. Signs require one or both hands for signing. Facial expressions which accompany signing are also important since they can modify the basic meaning of a hand gesture. To communicate proper nouns and obscure words sign languages employ finger spelling. Since the majority of signing is with full words, signed conversation can proceed with the same pace as spoken conversation.

So far, sign language dictionaries, textbooks and manuals relied on illustrations using drawings or photographs which were augmented by text descriptions. Multimedia technology seems an ideal medium for presentation, reference and learning of such gestural knowledge since it can also incorporate video material. An overview of computer-aided learning systems for education of hearing-impaired people can be found in [Alonso et al., 1995].

The most basic version of a multimedia dictionary of a sign language consists of words accompanied by video clips showing the corresponding gesture. Sound recordings of spoken words can be added, as well as illustrations and examples of sentences demonstrating the usage of a word.

Such a multimedia sign language dictionary can serve multiple functions. Primarily, it supports the learning process of sign language in special educational institutions for the deaf as well as enables normal hearing people who are in a daily contact with deaf people to learn to communicate with them in the sign language. Another goal of a sign language dictionary is to standardize a given sign language. Since communities of deaf people are often isolated from each other there is a great tendency to develop local dialects which are then not easily understood by other communities of deaf people. A dictionary can unify the meaning of signs and at the same time define a standard way of performing a sign. A dictionary can also somewhat fill the gap in the number of qualified sign language instructors.

We developed a multimedia dictionary of the Slovenian Sign Language (SSL) to address the needs in Slovenia. Using the same methods and the same framework we could easily produce dictionaries for other sign languages.

In the first half of the chapter, we give some background on how our sign language dictionary evolved, describe its structure, and give examples of its user interface. Based on our sign language dictionary, we developed a method of synthesizing the sign language, which makes possible a translation of written texts or, in connection with a speech recognition system, of spoken words to the sign language. This sign language synthesis method is described in the second half of the chapter.

2 Background

We presented our first concept of a multimedia sign language dictionary for the deaf on CD-ROM in 1995 [Jaklič et al., 1995a, Jaklič et al., 1995b]. This was to our knowledge one of the first concepts for a multimedia dictionary of sign language which we demonstrated at the New Talent Pavilion, MILIA'95, in Cannes, France. A similar approach was used for the American Sign Language Dictionaries on CD-ROM [Sternberg, 1994, PC, 1995].

A pilot application of our sign language dictionary consisting of less than 100 words was made in 1996 [Krapež, 1996] and the final application in 1999 which includes also synthesis of sign language sentences [Krapež and Solina, 1999]. The final version of the Slovenian Sign Language dictionary includes 2504 most frequent words that are used by the deaf people in every day conversations. The CD ROM-based Dictionary of the Slovenian Sign Language of the deaf was selected among the top 15 products in the Student Europrix'99 MultiMediaArt Competition [EuroPrix, 1999].

3 The structure and user interface of the sign language dictionary

The SSL multimedia dictionary contains 1800 individual sign video clips which represent 2500 words. This is possible because, like some words, a given sign can have several meanings, depending on the context. A good user interface for the dictionary is very important since it is not used only as a reference but also for learning (Figs. 1, 2).

Figure 1: Graphical user interface of a sign language dictionary. The examples in this work are from the *Slovenian sign language dictionary*. On the left side of the screen the user selected the word “baby” (“otrok” in Slovenian) which is illustrated with a drawing in the middle, some examples of using the word in sentences bellow the illustration, and the video clip of the corresponding sign on the right.

The left side of the screen enables the selection of words. The top left entry field is for searching words in the dictionary. As the sign language can be best learned in topic groups (i.e. family, work, nature etc.), the user can select a topic area in the left bottom corner for more effective search and navigation through the dictionary. Deaf people have in general difficulties with spelling. Therefore, when a word is entered, a list of possible matches is displayed on the screen to give a better overview over the database content and to help in the correct word selection.

After a word is selected, the user is presented with some written examples of sentences that demonstrate the use of the selected word, and if possible, also with a drawing illustrating the concept. A professional cartoonist was engaged for illustrating the words in the dictionary. However, abstract concepts are in general difficult to visualize and therefore not all words in the dictionary have corresponding illustrations.

The most important feature of the sign language dictionary is the video

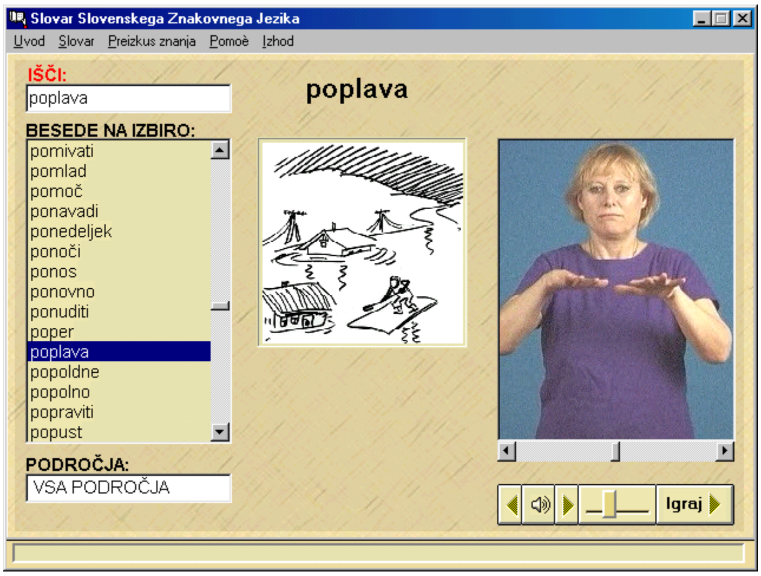

Figure 2: Another example of the graphical user interface where the word "flood" ("poplava" in Slovenian) is selected.

player on the right side of the graphical user interface where the corresponding sign demonstration can be played. The pronunciation of the selected words can be heard at the normal video playback speed. The sound volume can be readjusted by the user. The video clips can be played also in slow motion for easier visual tracking of the movements or even examined frame by frame by moving the slider along the time axis.

To fit the entire dictionary on a single CD ROM we had to optimize the video frame rate. We settled on 15 frames per second which is sufficient for smooth reproduction of the sign video clips. This is in accordance with the study of encoding ASL image sequences at extremely low information rates which concluded that 10 frames per second are sufficient [Sperling et al., 1985]. The video clips were compressed using the Intel Indeo Video 5 Codec [Codec, 2000] which is based on JPEG compression method.

Words which are not in the dictionary or proper names can be shown by spelling and using the video clips which contain the signs for individual letters. These signs are referred to as the finger alphabet (Fig. 3). In the SSL one hand is used for each letter in the finger alphabet. Therefore, any text can be shown finger spelled.

The complete finger alphabet forms also a special part of our sign language dictionary showing only still images of gestures for individual letter signs (Fig. 4). In this part of the dictionary the user can quickly learn the finger alphabet and also test his knowledge of finger spelling.

Using the same framework one can produce multimedia dictionaries of other sign languages.

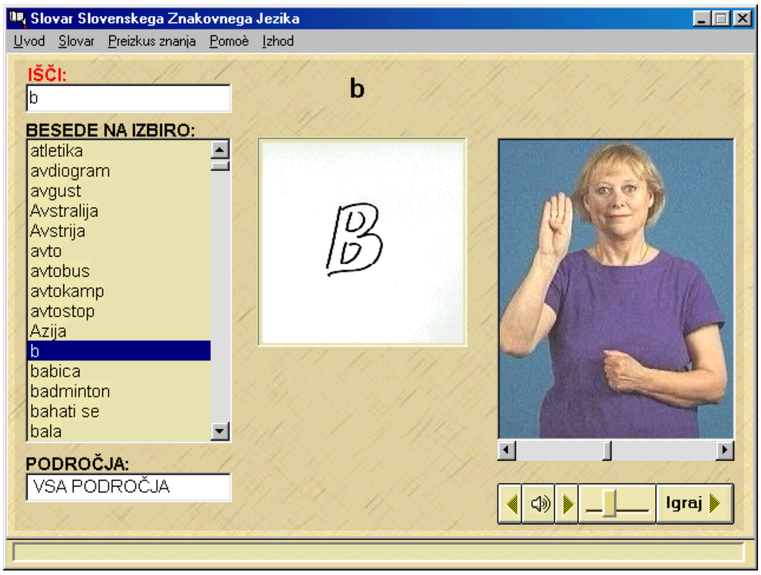

Figure 3: The demonstrator shows the finger alphabet sign for the letter “B”.

Figure 4: The finger alphabet used in Slovenian Sign Language. Each letter has a corresponding one hand sign. By typing a letter, the corresponding hand sign is displayed (for letter “V” in this example). For words in the entry field (“voda” is “water” in Slovenian) the corresponding signs are displayed bellow.

4 Synthesis of sign language

For translation of spoken or written language to sign language a well trained human sign interpreter is needed. In many situations where deaf people are involved exists the need for translation into the sign language. Sign language interpreters are not always available and beside that they are an expensive solution. With modern information technology it is possible to replace a human sign language interpreter by synthesizing the sign language.

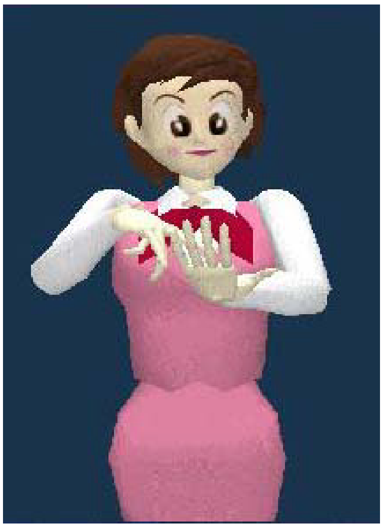

Figure 5: A synthetic sign language demonstrator made possible using computer graphics techniques [CRL, 1998]

The prevalent idea in the development of a system for sign language synthesis is to use a synthetic person [CRL, 1998]. Such systems are usually taught by a human sign language demonstrator using datagloves which contain sensors for capturing the palm and finger position and orientation which in turn animates the synthetic person (Fig. 5). A system for 3D hand animation for finger spelling is described in [Alonso et al., 1995]. Besides gestures performed with hands the accompanying facial expressions are also very important for sign language comprehension. Realistic animation of the human body motions and facial expressions is still a time consuming task. A synthetic demonstrator is at present no match for a real one in sense of appearance and consequently also of acceptance as a teaching tool. We decided therefore to solve the problem of sign language synthesis by using video clips of a real sign language demonstrator.

At this point, something must be said about the grammar of the sign languages. Sign languages in general allow much more flexibility in word order than their spoken equivalents. Some sign languages are grammatically very similar to their spoken languages, such as the Signed Exact English (SEE), a variant of the American Sign Language (ASL). The grammar of the Slovenian Sign Language is on the other hand quite apart from the grammar of spoken and written Slovenian language where the endings of words change according to the case, number, gender, conjugation and tense. The grammar of SSL is in fact closer to the English grammar than to the Slovenian grammar.

To perform a translation of a written language to its sign language equivalent, a syntactic text analyzer is needed first, to translate the written text into sequences of sign name tags. This problem is specific for each language. Next, for each sequence of these sign name tags, which forms a sentence, the appropriate sign video clips must be assembled. This gluing of video clips must be done in such a way that the observer perceives just one smooth signing motion. In our work that we describe here, we have concentrated only on the second part of the translation, namely, in producing signed sentences from a sequence of individual sign video clips. This problem is independent of any particular sign language.

4.1 Sign language synthesis using sign video clips

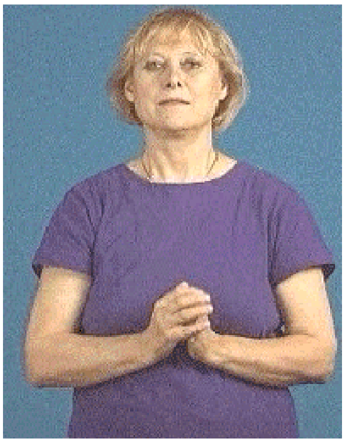

Figure 6: Example of the first frame of a sign video clip. The demonstrator has its arms in the start position.

Using video clips of a real person for sign language synthesis requires a consistent set of sign video clips so that when several clips are glued together, the viewer gets the impression that he is observing continuous signing of the demonstrator. This means that all video clips in the set must be captured using the same sign demonstrator and under the same conditions (i.e. appearance of the demonstrator, lightning, camera view) so that the basic prerequisites for consistency among individual video clips are met. In producing our sign language dictionary we achieved the highest possible consistency by videotaping the entire sign vocabulary in one session with just one sign demonstrator. This requires good planning since any later additions are very difficult. The ASL multimedia dictionary [Sternberg, 1994], for example, uses several different sign demonstrators which prevents its use for sign language synthesis.

It is important that the start and end positions of the arms for all gestures are the same on all video clips (Fig. 6). Given these prerequisites, the sign video clips from a sign language dictionary can be glued into sign language sentences, provided that the gluing of video clips is optimized. Optimization means that redundant moves of the hands between consecutive gestures must be eliminated. The SSL dictionary which consists of such a consistent set of sign video clips therefore represents the basis for the sign language synthesizer.

We introduced a new method for joining sign video clips into video clips of complete sign sentences. We eliminated redundant moves of the hands and produced smooth transitions between sign video clips by tracking the position and orientation of the demonstrator's palms while signing. For that purpose we wrote a special program for palm position extraction from video clips using computer vision methods [Klette and Zamperoni, 1994, Marshal and Martin, 1993]. Palm position extraction is done by preprocessing all video clips and saving the palm positions in special files related to each video clip. The actual sign language synthesis from a sequence of sign name tags can be achieved in real time on a personal computer.

4.2 Definition of the video clip joining problem

Joining two digital video clips (each contained in its own file and consisting of several individual frames) should be done in such a way that the viewer perceives them as a single video sequence with smooth motion. This can be done by eliminating some frames at the end of the first video clip and some frames at the beginning of the second video clip to form the final joined video sequence.

The problem of video clip joining will be discussed for the case of two video clips. This can be easily generalized to any number of consecutive video clips. Let us represent the first video clip that we would like to show as a n dimensional vector $\mathbf{v}_1$

$$\mathbf{v}_1 = \langle p_1, p_2, \ldots, p_n \rangle, \tag{1}$$

where the components of the vector $\mathbf{v}_1$ correspond to the frames that form the video clip and n is the number of frames in the video clip. The second video clip is in the same way represented as

$$\mathbf{v}_2 = \langle r_1, r_2, \ldots, r_m \rangle. \tag{2}$$

Joining of video clips means that after the video frame p_u in $\mathbf{v}_1$ follows the video frame r_v in $\mathbf{v}_2$ (Fig. 7) where index u represents tha last video frame shown in the first video clip $\mathbf{v}_1$ and index v the first video frame shown in the second video clip $\mathbf{v}_2$. In general holds

$$u \in [1 \ldots n] \tag{3}$$

$$v \in [1 \ldots m]. \tag{4}$$

But the central part of both video clips must be preserved. Therefore the values of u and v must be more restricted. Let e be the index of the frame in the first video clip, from which on the joining is allowed, and f the index of the frame in the second video clip, till which the joining is allowed. With these constraints u and v can occupy the following values

$$u \in [e \ldots n] \tag{5}$$

$$v \in [1 \ldots f]. \tag{6}$$

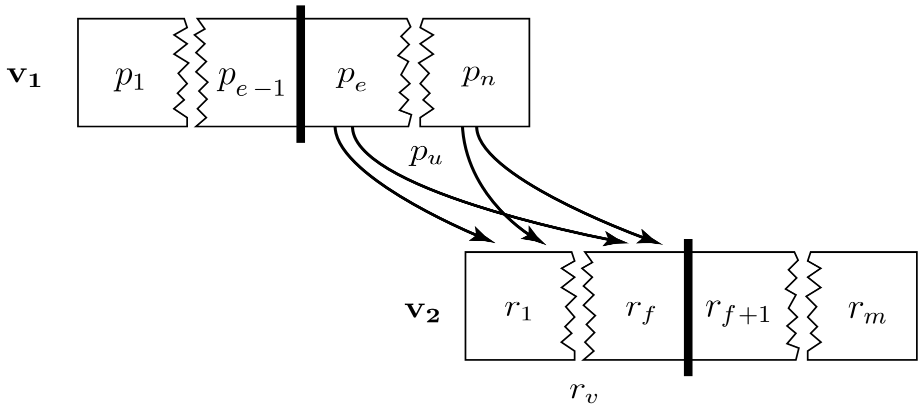

Figure 7: Two video clips are joined at frames p_u from the first video clip $\mathbf{v_1}$ (above) and frame r_v from the second video clip $\mathbf{v_2}$ (bellow). Frame p_u is selected among frames between p_e and p_n in the first video clip. Frame r_v from the second video clip is selected among the frames between r_1 and r_f.

The joining of video clips should be performed in such a way that the transition from one video clip to the next one would be as smooth as possible. Let us introduce a criteria function $\mathcal{DIFF}(i, j)$. This function evaluates the difference between frames p_i and r_j depending on a given joining criterion. The smaller is the function value, the more both frames resemble each other. The calculation of u and v is performed by finding the values of arguments for which the function $\mathcal{DIFF}$ has a minimum value.

4.3 Automatic sign language video clips joining

Note that a very smooth transition between two sign video clips can be achieved simply by joining complete sign video clips since in each clip the sign demonstrator begins and ends in the same position with the hands at the waist level, which we refer to as the start position (Fig. 6). But this would be a very unnatural way of signing, similar in its effect to trying to synthesize speech by playing in sequence prerecorded words. Redundant motions of the arms (from the start position of the arms into the demonstration of a sign) between subsequent signs must be eliminated. In some instances, however, such as between sentences or paragraphs, such pauses are in fact required. Therefore, besides a smooth transition between two video clips, an additional requirement is necessary.

In order to be able to perform optimal joining of sign video clips some extra data is needed. We need to know the position of palms in each video frame. Each video clip has a related file containing data about the position of demonstrator's palms performing a particular sign. We obtained the positions of palms by preprocessing all video clips by a program based on standard computer vision algorithms, which we wrote in C++. The input to that program are the sign video clips. The output are files containing data about palm positions for every frame forming the video clip. The positions of palms are used for evaluation of the $\mathcal{DIFF}$ function. Fig. 8 shows the arm vectors which are computed in each video frame of a sign video clip. The palm position is at the end point of the arm vector.

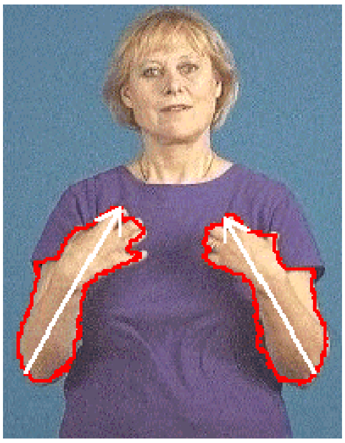

Figure 8: The arm vectors are computed in each video frame and the positions of palms are stored.

Something more about the use of arms while signing must be explained. Most signs are performed at the chest and head level. Signs performed at the waist level or lower are rare but still to be considered. Approximately two thirds of signs are performed with one hand. For one hand signs one can use either the right or the left hand. That does not affect the meaning of the sign. In our case, all one handed signs are performed with the right hand. We call these signs *one-hand signs*. Approximately one third of signs are performed with both hands. We call these *two-hand signs*.

4.3.1 Joining criteria and their selection

Conditions that enable a smooth transition between video clips in the sense of similar palm positions are referred to as the *joining criteria*. After analyzing the position of palms during the performance of a sign gesture in all types of signs in our sign language dictionary, we proposed four different joining criteria and their corresponding $\mathcal{DIFF}$ functions to determine the optimal transition point between two sign video clips. The four joining criteria are:

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

Start position criterion Using this criterion, the transition should occur at points where arms are in the start position.

The function $\mathcal{DIFF}(i,j)$ for a frame pair (p_i, r_j) should therefore have a minimum value when the last frame of the first video clip ($i = n_1$) is glued to the first frame of the second video clip ($j = 1$) (see Fig. 9a).

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 a certain value.

Function $\mathcal{DIFF}(i,j)$ for a frame pair (p_i, r_j) has in this case its minimum when palms on frame p_i are outside the start position for the last time and outside the start position for the first time in frame r_j (see Fig. 9b).

Palms over the chest criterion According to this criterion the transition between video clips is performed when palms are over the chest.

Function $\mathcal{DIFF}(i,j)$ for a frame pair (p_i, r_j) has its minimum in this case when the palms on frame p_i are for the last time over the chest and palms on frame r_j are for the first time over the chest region (see Fig. 9c).

Palms close to each other criterion Using this criterion, the transition occurs at points where the palms from the first and the second video clip are close to each other.

Function $\mathcal{DIFF}(i,j)$ is defined in this case as:

$$\mathcal{DIFF}(i,j) = d(i,j) + w(i,j) \tag{7}$$

where $d(i,j)$ represents the distance between palms in frames p_i and r_j. $w(i,j)$ is a weight function which penalizes the joining of video clips at frames distant from frames d_1 and d_2 which represent the first and the last video frame allowed for joining in the first and the second video clip, respectively

$$w(i,j) = ((i - d_1) + (d_2 - j)) * K, \tag{8}$$

where K is an empirical constant value (in our case $K = 5$) (see Fig. 9d).

Which of the four proposed joining criteria is used depends on the type of the two signs we would like to join. The criterion for joining two sign video clips is selected according to the following four situations:

1. a delay between signs is needed $\Rightarrow$ *start position criterion*,
2. one of the signs is performed at the waist level $\Rightarrow$ *palms outside the start position criterion*,
3. exactly one of the signs is a two-hand sign $\Rightarrow$ *palms over the chest criterion*,
4. both signs are one-hand signs or two-hand signs $\Rightarrow$ *palms close to each other criterion.*

If conditions enable the use of more than one criterion, we use the first one that matches a situation in the above given ordering.

4.4 Results of automatic joining of sign video clips

Let us have a look at some transitions between two video clips obtained with this method of automatic video clip joining. Fig. 9 shows examples for each of the four joining criteria that we proposed. The left side of each example shows the last shown video frame of the first sign video clip and the first shown video frame of the second sign video clip. Fig. 9a shows an example using the *start*

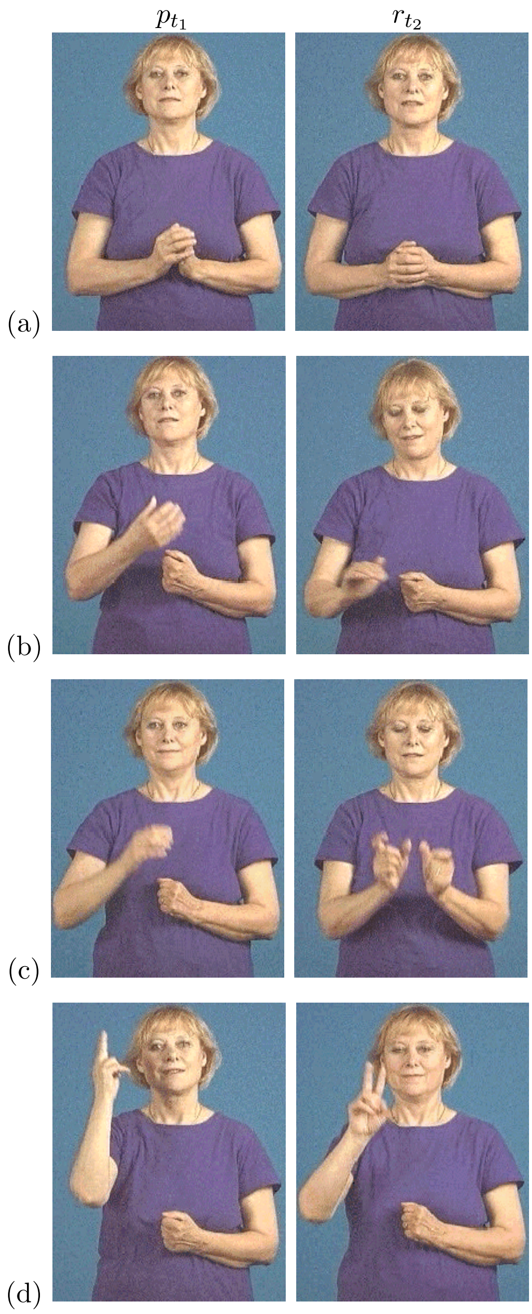

Figure 9: Four examples of automatic joining of video clips. For each example are shown on the left the last frame of the first video clip (p_u) and on the right the first frame of the second video clip (r_v), according to (a) *start position criterion*, (b) *palms outside the start position criterion*, (c) *palms over the chest criterion* and (d) *palms close to each other criterion* for optimal joining of two video clips.

position criterion. This kind of transition is useful when we want to emphasize the end of a sentence or the end of a paragraph. It is useful also in the case where the values of the palm positions are not available.

Fig. 9b shows the transition between two signs where either one of them is performed at the waist level. *Palms outside the start position criterion* is used in this case to prevent arms from being joined in the start position. The results obtained on one-hand and two-hand signs using the *palms over the chest criterion* are shown in Fig. 9c. Fig. 9d 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. The synthesis method 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 due 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 of making an impression of continuous transition between two frames that are similar enough and shown quickly one after another.

The system for sign language synthesis can show video clips with normal speed or with lower or faster speed than normal. In this way, it can be adapted to the user's ability of recognizing sign language. An example of a synthesized sign language sentence is shown in Fig. 10.

5 Future trends

We can expect to see similar multimedia dictionaries for most sign languages in the near future. Using the same technology and methods described in this text we could easily produce dictionaries for other sign languages. One can also envision multilingual dictionaries for several different sign languages as well as their corresponding spoken languages. An open problem for such multimedia sign language dictionaries is how to search through the visual material. For example, if we see a gesture that we do not understand, how could we easily find its interpretation? A "natural" organization of the visual material (i.e. one-hand sign versus two-hand signs etc.) in combination with some simple visual search methods could address such needs.

The long-term goal is to build systems able to translate the written language (books, newspapers, e-mails, letters, HTML documents, ...) and in connection with speech recognition also spoken language (conversations, radio and TV programs, phone calls, ...) to the sign language.

The natural complement to sign language synthesis is sign language recognition. Similar to early speech recognition, first attempts at sign language recognition concentrated on isolated signs or fingerspelling. For caption of gestures instrumented gloves can be used but passive, vision based methods are of course much preferred. Current advanced efforts at sign language recognition concentrate on real-time continuous gesture recognition from image sequences which is based on hidden Markov models [Starner et al., 1998, Vogler and Metaxas, 1998]. These research systems are typically capable of recognizing up to 100 different

Figure 10: Example of a synthesized sign language sentence "I am walking home". The sentence is assembled out of three video clips ("I" + "walking" + "home"), each shown in its own row. Only every fifth video frame is shown in the figure. The actual video frame rate is 15 frames per second.

words or signs with an 80–90% reliability. Beside from the usual "second-person" viewpoint for observation of gestures, it is possible also to recognize signs from the "first-person" viewpoint, which is obtainable with a camera mounted on the head of the signing person and pointed downwards towards his or her hands. In this way, a wearable computer based system capable of real-time translation from the sign language to spoken language is envisable [Starner et al., 1998].

6 Conclusion

We developed a multimedia dictionary of sign language which supports sign language synthesis. Each entry for the selected word consists of the video clip showing a demonstration of the corresponding sign. In addition to the video clip, a drawing illustrating the selected word is given, where visualization is possible. Examples of sentences using the selected word are also included. Words can be searched in the entire dictionary or only in selected topic areas. For uncommon words and proper names the finger alphabet can be used for spelling. In our implementation we used the Slovenian Sign Language. The dictionary consists of more than 2500 entries and fits on a single CD-ROM.

Since the video clips showing sign demonstrations were taped under uniform conditions (same sign demonstrator, camera position, illumination etc.) sign language synthesis became possible. Given a sequence of name tags for corresponding sign video clips, we can join these individual video clips so that all the cuts appear smooth and redundant movements of hands are eliminated. Sign language synthesis runs in real-time on a personal computer. The system is especially suitable for prerecording known sentences used in education of deaf people through multimedia. With the help of a suitable syntactic analyzer, the translation of written texts into sign language is also possible.

The dictionary is Web ready and a short demonstration using a Java user interface and MPEG encoded video sequences is already available [SSL on Web, 1999]. On the World Wide Web are available several sign language dictionaries. Some good examples are the American Sign Language [ASL, 2000, HandSpeak, 2000] and the German Sign Language Dictionary on Psychological Terms [GSL, 2000]. Sign language synthesis on the Web is, however, technologically much more demanding. The main problem is to assure that joined video clips will always play without a visible delay. We are currently investigating a pure JAVA implementation of the dictionary and of sign language synthesis, to make it platform independent.

7 Acknowledgment

This work was supported by the Ministry of Science and Technology of Republic of Slovenia (Project no. L2-0521) and Zoom Promotion, Ljubljana. We would like to thank also Ljubica Podboršek, who acted as the sign demonstrator, and Dora Vodopivec for their enormous help in the development of the Slovenian sign language dictionary on CD-ROM, Matej Gašperič for helping with the design of the graphical user interface, and Marjan Bregar for his drawings.

References

[Alonso et al., 1995] Alonso, F., de Antonio, A., Fuertes, J. L., and Montes, C. (1995). Teaching communications skills to hearing-impaired children. *IEEE Multimedia*, 2(4):55–67.

[ASL, 2000] ASL (2000). *American Sign Language Dictionary.* The Communication Technology Laboratory at Michigan State University. http://commtechlab.msu.edu/sites/aslweb/.

[Codec, 2000] Codec (2000). *Indeo Video 5.* Ligos. http://www.ligos.com/indeo.

[CRL, 1998] CRL (1998). *Sign Language Recognition and Synthesis System.* Communications Research Laboratory, Japan. http://www.crl.go.jp./st/st821/research/sl-e.htm.

[EuroPrix, 1999] EuroPrix (1999). *EuroPrix99 Multimedia Art, Europe's Best in Multimedia.* http://www.europrix.org.

[GSL, 2000] GSL (2000). *German Sign Language Dictionary on Psychological Terms*. University of Hamburg. http://www.sign-lang.uni-hamburg.de/projects/SLDicts.html.

[HandSpeak, 2000] HandSpeak (2000). *HandSpeak*. http://www.handspeak.com.

[Jaklič et al., 1995a] Jaklič, A., Vodopivec, D., and Komac, V. (1995a). Learning sign language through multimedia. In *Proceedings of International Conference on Multimedia Computing and Systems*, pages 282–285, Washington.

[Jaklič et al., 1995b] Jaklič, A., Vodopivec, D., Komac, V., and Gašperič, M. (1995b). Multimedia learning tools for the hearing impaired. In *Proceedings of the World Conference on Educational Multimedia and Hypermedia ED-MEDIA 95*, pages 354–359, Graz.

[Klette and Zamperoni, 1994] Klette, R. and Zamperoni, P. (1994). *Handbook of Image Processing Operators*. Wiley.

[Krapež, 1996] Krapež, S. (1996). *Slovar znakovnega jezika za gluhe/Sign language dictionary for the hearing impaired*. Dipl. ing. thesis, University of Ljubljana, Faculty for Computer and Information Science.

[Krapež and Solina, 1999] Krapež, S. and Solina, F. (1999). Synthesis of the sign language of the deaf from the sign video clips. *Electrotechnical Review*, 66(4–5):260–265.

[Marshal and Martin, 1993] Marshal, A. D. and Martin, R. R. (1993). *Computer vision, Models and inspection*. World Scientific, Singapore.

[PC, 1995] PC (1995). *Personal Communicator CD-ROM*. The Communication Technology Laboratory at Michigan State University.

[Sperling et al., 1985] Sperling, G., Landy, M., Cohen, Y., and Pavel, M. (1985). Intelligible encoding of ASL image sequences at extremely low information rates. *Computer Vision, Graphics, and Image Processing*, 31:335–391.

[SSL on Web, 1999] SSL on Web (1999). *Slovenian Sign Language on the Web*. University of Ljubljana, Faculty of Computer and Information Science. http://lrv.fri.uni-lj.si/slavko.html.

[Starner et al., 1998] Starner, T., Weaver, J., and Pentland, A. (1998). Real-time American Sign Language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1371–1375.

[Sternberg, 1994] Sternberg, M. L. A. (1994). *The American Sign Language Dictionary on CD-ROM*. Harper Collins, New York.

[Vogler and Metaxas, 1998] Vogler, C. and Metaxas, D. (1998). ASL recognition based on a coupling between HMMs and 3D motion analysis. In *Proceedings International Conference on Computer Vision, Bombay, India*. IEEE Computer Society.

Franc Solina received his Dipl.Ing (1979) and M.Sc. degrees (1982) in electrical engineering from the University of Ljubljana and a Ph.D. degree (1987) in computer science from the University of Pennsylvania. He is a professor of computer and information science at University of Ljubljana and the head of the Computer Vision Laboratory. His research interests are segmentation and part-level object representation in computer vision, as well as Internet based applications of computer vision.

Slavko Krapež received his Dipl.Ing (1996) and M.Sc. degrees (1999) in computer and information science from the University of Ljubljana. His research interests are multimedia, especially multimedia support for learning of sign language and Internet based applications. Slavko Krapež is now with Hermes SoftLab in Nova Gorica, Slovenia.

Vito Komac is the C.E.O. of the marketing & research company Zoom Promotion in Ljubljana, Slovenia. He received his B.S. degree in Economics in 1977 from University of Ljubljana, Faculty of Economics. He is engaged in creating, planning and performing international marketing communication and public relations activities for several important clients. He writes articles for business magazines on marketing and opinion research.

Aleš Jaklič received his Dipl. Ing (1989), M. Sc. (1992) degrees in electrical engineering and a Ph.D. (1997) degree in computer science from the University of Ljubljana. He is a teaching assistant at the University of Ljubljana. His research interests span computer vision, multimedia and Internet technologies.