Application of Sign Language Image Synthesis Technology for Free Mobility Project

Yusuke Suzuki Ichiro Miyamoto
Koichi Takeuchi

The number of people certified as having a physical disability in Japan exceeds 300,000, according to statistical data provided by the Ministry of Health, Labour and Welfare. Approximately 60,000 of these people use “Japanese Sign Language” as their mother tongue, which has a different linguistic system from the Japanese language.

A significant amount of information in daily life is provided only in audible form, such as voice announcements. The problem with this is that the information is not accessible to people with impaired hearing. To be able to provide such information in sign language has a great potential for benefiting persons with impaired hearing.

Considerable research into creating system expressing sign language is ongoing, with some products already being commercialized[1), 2), 3)]. Many such systems use images that are synthesized with computer graphics (hereafter referred to as “CG”) to display expressions in sign language. A benefit for using CG is its ability to synthesize sign language expressions for arbitrary sentences, however, other issues need to be resolved, such as the creation of expressions that involve non-manual signals (hereinafter referred to as “NMS”), which are extremely important for conversing in sign language, as they rely on expressions, the movements of a body and those of lips[4), 5), 6)].

The authors have been developing a sign language image display system that uses real captured images. Actual images have superior expressive capability with regards to NMS described above in comparison with CG and another advantage is that people who actually converse using sign language tend to prefer such images[7)]. However, issues remain to be resolved, such as the fact that in comparison with CG it is more difficult to synthesize sentences freely. This paper describes a system intended to resolve such issues by utilizing the techniques of image synthesis.

Sign language image synthesis system

The sign language image synthesis system uses the images of real living persons whose actions perform actual sign language motions have been captured. We placed our emphasis on the important role NMS expressions play, with the knowledge that conversations do not take place with the movement of hands alone. The analysis and research of NMS are still a work in progress and our understanding of it has not yet reached a level where it is possible to synthesize NMS with CG. Furthermore, when sign language movement data are added onto a system using CG, a technology known as motion capture is often employed to capture the movement of people. Capturing the accurate movements of expressions and handling such data are quite difficult, whereas the proofing of this kind of data tends to require large amounts of manual work.

Images that capture people actually conversing with sign language are much more expressive in terms of NMS in comparison with CG, therefore, the sign language is much easier to understand. Furthermore, since these people are used to see actual persons using sign language, there is the added advantage that actual people who use sign language prefer this mode of imagery. For this reason the captured images of people who are actually using sign language will be considered ideal for displaying information in sign language. However, captured images lack flexibility and require new captured images to be taken even when a slight change in the content occurs. The problem is that the synthesis of sentences is more difficult than CG, which is more flexible for composing sentences.

The authors considered all such issues and aimed to realize a sign language image synthesis that merges the advantages of the two methods well. This means the development of a sign language image synthesis technology with the flexibility of CG and expressive capability of actual captured images. Methods for preparing the aforementioned images and their benefits are summarized in **Table 1**. Checked items are features for which the particular method proved to be advantageous.

Table 1 Comparison of sign language imaging systems

	Methods		
	Captured motion picture	CG	Our system
Dynamic response to changing conditions		√	√
Scalability		√	√
Expressive capabilities of NMS	√		√
Familiarity for persons with hearing impairment	√		√

Our system synthesizes a series of sign language sentences by selecting and linking up all the required images that have been fragmented into units of words. These image fragments are captured from the images of a real persons while conversing in sign language. Furthermore, this system also incorporates highly scalable data, which allows data to be added to the system easily once the images are captured and clipped into fragments. In addition, the system features compensation with the insertion of blank frames between the connected images using an image synthesis technique.

A conceptual diagram of the system configuration on trial is shown in **Fig. 1**.

This system performs a morpheme analysis of sentences that are entered sentences and divides the sentences into fragments comprised of individual Japanese words. The system then selects from the database the sign language image files that correspond to these individual words, links them up and outputs them as a series of sentences. Individual components and the actual operation of the system are described below, based in **Fig. 1**.

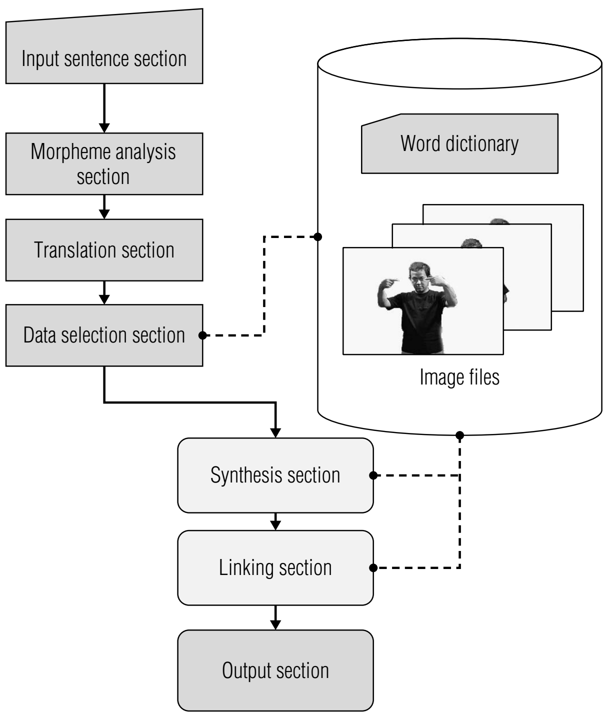


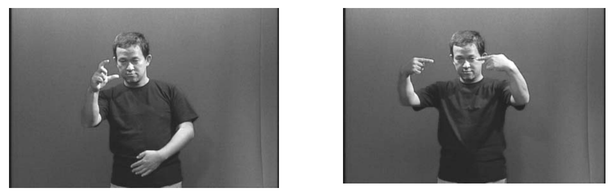
Fig. 1 System configuration

The system is comprised of data, a input sentence section, morpheme analysis section, translation section, data selection section, synthesis section, linking section and output section. The data used are primarily comprised of word dictionaries and image files. Word dictionaries indicate a collation between Japanese language and sign language image files. Image files are the group of sign language components necessary to express sentences that have been divided into units of fragments, which are almost individual words. The morpheme analysis section performs a morpheme analysis on the Japanese language sentences entered into the system.

The translation section starts the translation process with an analysis on the sequence of individual words, then adjusts their sequential relationships according to grammar rules for sign language and attaches sign language word labels to the sentence. The sign language word label is the assigned Japanese language label used primarily to identify individual words of sign language.

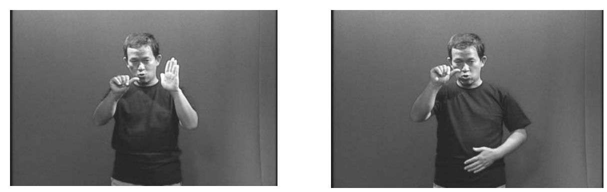
Because Japanese sign language is a different language from spoken Japanese, a meaning that can be expressed with a single word in Japanese may be expressed in more than one way in sign language.

For example, words used to describe weather conditions, like the morpheme taken for the expression “moving north”, as in a “typhoon is moving north” and a “cold front is moving north” are quite different, as shown in **Fig. 2**. Furthermore, the position of the left hand varies greatly for the words for numbers depending on whether it is used in relation to “typhoon” or “temperature”, as shown in **Fig. 3**. Individual image files for each situation vary depending on the relationship of the word stored and the image files most suitable as well as the data selection that takes place following the translation.



“moving north” (typhoon)　　“moving north” (cold front)

Fig. 2 Variance of expression depends on word relationship



Five (degrees temperature)　　Five (number of typhoon)

Fig. 3 Variance of numerical morpheme depends on word relationship

Connecting image issue

Image files are fragmented into units of words and there are no connecting images to provide a link between the end of one word and the beginning of the next. Therefore, by simply linking these components and displaying them does not result in a final completed series of movements of a person but a series of images with blank segments (gaps) in the movements of the person, which makes display of the final image file appear odd.

To resolve this problem synthesized images are created to complement these blank sections of movement between image files with the use of an image synthesis technique. Such synthesis of the image files include a series of sign language images output as synthesized images.

Image synthesis

The flow of an intermediate image synthesis is described according to **Fig. 4** (on the next page). Two files are extracted from the image files. The file that will be displayed first is referred to as the leading file, whereas the file displayed next is referred to as the trailing file.

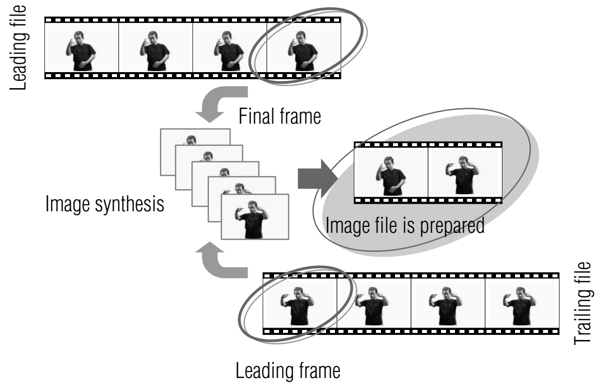


Fig. 4 Flow of image synthesis

(1) Two image files are prepared.
(2) The last frame at the end (final frame) of the motion in the leading file is extracted and the first frame at the beginning (leading frame) of the motion in the trailing file is also extracted.
(3) The morphing process, a technique used in image synthesis, is applied on the two extracted frames. Multiple intermediate image frames are generated to bridge the change from the final frame to the leading frame. Details of this process are described in the next section.
(4) The generated image frames are linked together to prepare the image file.
(5) The prepared image file is inserted between the leading file and trailing file.

This series of processes provide an effect to make the motion of the person conversing in sign language appear seamless.

Details of image synthesis

The details of the method used for image synthesis is described in **Fig. 5**. This method evolved from improving basic experiments repeatedly, based on Warping[8].

First, a contour surrounding the arm region of a person from their palms to their elbows is obtained as a group of multiple points on a static frame, which has been clipped out of a moving image. The triangular meshes covering the aforementioned region are prepared next using the multiple points that comprise the contour. The triangular meshes are prepared in this manner for both the last frame of the leading file and the first frame of the trailing file.

The transitional velocity of the mesh region between the last frame of the leading file and the first frame of the trailing file is determined, based on their positional differences. The mesh from the last frame is then transitioned to the mesh of the first frame with a determined velocity, while the pixels inside the mesh region are output to a screen using their weighted average values. In this manner we obtain an image that shows a motion in which the mesh region or the region of the arm moves while changing its shape.

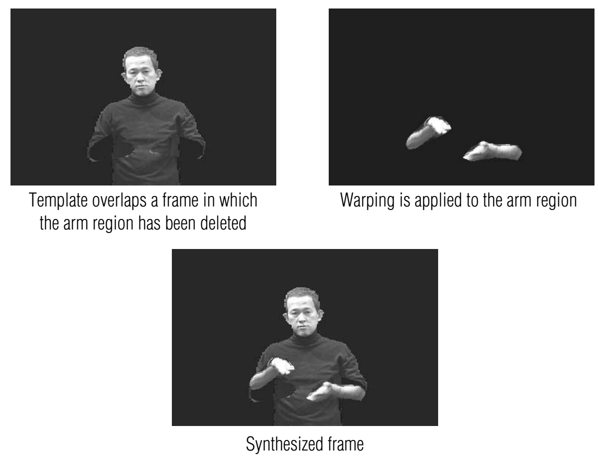


Fig. 5 Image synthesis diagram

This image is output, overlapping with the image Synthesized prepared previously that shows the entire body region, but with the arm region of the person hidden, as well as an image that has the arm region smeared to ensure that the background does not appear in the foreground. This results in an image that makes the arms appear to be moving in a natural manner.

Case example of experimental application

An information delivery system using the sign language image synthesis engine, described in the previous section, was developed and an experiment was conducted at Kobe Airport. This experiment was a part of the Free Mobility Project being promoted by the Ministry of Land, Infrastructure and Transport. The Free Mobility Project has a tenet for using information technology to provide support for persons with impaired mobility or a disability, to obtain information relating to movement and allow autonomy over their movements, whether they are elderly persons, physically disabled or persons with temporary disabilities brought on by injury.

The feasibility of the sign language image synthesis system was evaluated with a verification experiment for the project held at Kobe Airport with impaired hearing persons. A sign language image synthesized by the system was displayed on handheld devices carried by the experiment participants. The display of the sign language image on a handheld device is shown in **Fig. 6**. The results of a interview taken following the completion of the evaluation experiment showed many experiment participants supported the provision of information and they indicated that information using sign language was required. Another article in this issue, “Aiming to Realize Ubiquitous Services: Participation in Verification Experiment of Ubiquitous Services for Kobe Airport” (pp. 4 to 9), provides a detailed report on this experiment.

Fig. 6 Display of sign language image on handheld device

Conclusion

This paper described the sign language image synthesis system being developed. Many aspects of sign language are still unknown and then continued research is necessary in cooperative relationships with universities and research organizations of other business enterprises, as well as sign language users, who, after all, will become the actual users of the system.

References

1) Takao Kurokawa, “Trial for Bidirectional Translation between Japanese Sign Language and Japanese Language”, Collected Papers of the Fifth Conference of the Kansai Federation of Information Related Academic Societies, pp. 25 to 34, 2000.
2) Tomohiro Kuroda, et. al., “S-TEL Sign Language Transmission System”, Technical Report of the Institute of Electronics, Information and Communication Engineers, ET96-85, pp. 65 to 71, 1996.
3) Hirohiko Sagawa, “Sign Language Animation Software Mimehand and Its Applications”, MedCom., Vol. 13, No. 8, 2002.
4) Harumi Kimura and Yasuhiro Ichida, “Introduction to Sign Language”, Nihonbungeisha, 1995.
5) Hiroki Yonehara and Yuji Nagashima, “Experimental Study on Changes in Observing Visual Point According to Degree of Mastery in Sign Language”, Collected Papers for Human Interface Symposium 2002, pp. 233 to 236, 2002.
6) Tetsuhiko Kodama and Michiaki Yasumura, “Trial of Sign Language Animation Including Representation of Expressions”, Research Report for the Information Processing Society of Japan, 2003-HI-103, pp. 23 to 29, 2003.
7) Takashi Ichikawa et. al., “A Study on Ease of Viewing Sign Language Images on Screens of Handheld Devices”, Collected Papers for Human Interface Symposium 2003, pp. 309 to 312, 2003.
8) Thaddeus Beier and Shawn Neely, “Feature-based Image Metamorphosis”, Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques, pp. 35 to 42, 1992.

Authors

Yusuke Suzuki: Corporate Research and Development Center, Human Interface Laboratory, Interactive Communication Team.

Ichiro Miyamoto: Corporate Research and Development Center, Human Interface Laboratory, Interactive Communication Team.

Koichi Takeuchi: Corporate Research and Development Center, Human Interface Laboratory, Interactive Communication Team.