

Digital Encoding Applied to Sign Language Video

Kaoru NAKAZONO^{†a)}, Yuji NAGASHIMA^{††}, *Members*, and Akira ICHIKAWA^{†††}, *Fellow*

SUMMARY We report a specially designed encoding technique for sign language video sequences supposing that the technique is for sign telecommunication such as that using mobile videophones with a low bitrate. The technique is composed of three methods: gradient coding, precedence macroblock coding, and not-coded coding. These methods are based on the idea to distribute a certain number of bits for each macroblock according to the evaluation of importance of parts of the picture. They were implemented on a computer and encoded data of a short clip of sign language dialogue was evaluated by deaf subjects. As a result, the efficiency of the technique was confirmed.

Key words: sign language, quality evaluation, picture encoding, assistive technology

1. Introduction

We discuss a technique for visual communication using sign language. It is very helpful for deaf people if telecommunication of sign language becomes available over mobile videophones. However, in the present circumstances, the communication speed (bitrate) is limited to 64 kbps in the case of Japanese mobile videophones, which cannot offer a useful service.

The authors designed an improved video encoding technique that was specially designed for improving the intelligibility of sign language video sequences. The technique uses characteristics of sign language based on the idea of distributing number of bits according to the evaluation of importance of parts of the picture. The technique was implemented in a test encoding program in accordance with the H.263 standard.

In the authors' previous paper, the evaluation test comparing general and specially designed encoding techniques was carried out by deaf subjects. However, results of the evaluation were nearly equal and the efficiency was not sufficiently verified [1]. In that experiment, the tested pair of data was encoded in the same size (CIF or QCIF, which is defined in Sect. 3), the same target bitrate and the same target framerate. Whereas the framerate is most conclusive for determining the intelligibility of sign video sequences, it is understandable that the results of the evaluation were nearly

equal.

In this study, the implementation of the encoder was refined and encoding parameters were adjusted to attain a framerate as high as possible for the given target bitrate. The tested pair of data was encoded in the same size and the same target bitrate but obtained framerates were different. Then, evaluations of the renewed data were carried out. The results indicated that the specially designed encoding technique is effective in particular in the case of a very low bitrate. These results assure us that the technique is suitable for the low bitrate videophone such as the mobile phone.

In the rest of the paper, we first discuss characteristics of sign language video sequences, introduce the basic idea of encoding techniques, and review related work in Sect. 2. In Sect. 3, details of encoding techniques are shown. In Sect. 4, the process of making the test data, steps of the evaluation tests, and results are presented. Section 5 concludes the paper by reviewing these discussions and considering future work.

2. Outline of Encoding Techniques for Sign Language Video

2.1 Basic Idea of the Encoding

Information around the face of a signer is known to be important for sign language communication. For example, facial expression and line of sight conveys important information. Furthermore, the expression of precise meaning or nuance is also shown in the same region.

On the other hand, human vision has the following characteristics: 1) the center of the visual field is good at recognizing precise but motionless objects or patterns, 2) the corner of the visual field is weak in precise recognition, but good at recognizing swift motion. In fact, people who use sign language daily do not follow the motion of the hands of their companion, but gaze at the facial region. This finding was confirmed by an experiment using an eye camera that can record the line of sight [2], [3].

From these findings, we have arrived at the following prospect, i.e., an encoding technique that attaches importance to the facial area of the picture will be effective for improving the intelligibility of sign language communication. Using this policy, two effects are expected. First, a more detailed picture can be sent at the same bitrate as not using this policy. Second, a higher framerate movie can be sent without degrading the preciseness of the important re-

Manuscript received September 25, 2005.

Manuscript revised January 10, 2006.

[†]The author is with NTT Network Innovation Laboratories, NTT Corporation, Musashino-shi, 180-0012 Japan.

^{††}The author is with Kogakuin University, Hachioji-shi, 192-0015 Japan.

^{†††}The author is with Chiba University, Chiba-shi, 263-8522 Japan.

a) E-mail: nakazono@core.ecl.net

DOI: 10.1093/ietisy/e89-d.6.1893

gion.

Then, how can we detect the facial area? The facial area can be detected by using an image processing technique. However, in thinking about the usage of mobile videophones, the position of the signer can be assumed to be settled in the center of the picture.

2.2 Three Components of Encoding Technique

Based on the idea stated in the previous section, the following three methods are presented. The full effect of these methods can be obtained when they are combined.

(1) Gradient Coding

In the usual encoding process, target data amounts are distributed evenly to all MBs (macroblocks[†]). For the sign video sequence in which the signer's position is fixed, we can distinguish between important regions and unimportant regions in advance. Consequently, if we assign a large number of bits around the signer's face and reduce the number of bits at the corners of the picture, better encoding of sign videos can be expected at the same bitrate. In the case of H.26X encoding, picture quality is controlled by the QP (Quantization Step Size.^{††}) We call the method "gradient coding."

(2) Precedence Macroblock Coding

Thinking of the usage of mobile videophones, it is possible that the background of pictures include more complex movements like a crowded street, as shown in Fig. 1. In that case, a lot of data would be exhausted for encoding of the background and sufficiently good image quality or a sufficient framerate might not be obtained.

Generally, MBs are processed in a fixed raster scan order. However, when processing in this order, a problem may occur when the bitrate is very low. If a large number of bits has been exhausted in the upper part of the picture, for example, very fine background or drastic change were included in that region, a sufficient number of bits might not be allocatable even if a very important region emerges afterwards.

Thus, when an important region, where a large number of bits is required, is known beforehand, the following methods are effective. We start the encoding process from the important region leaving the unimportant region for later. This method is called "precedence macroblock coding."

(3) Not-coded Coding

Even if the two former methods are implemented, a failure of encoding is sometimes caused by the following factors.

During the process of image encoding, QP is always adjusted. QP is the parameter that controls the fineness and data amount of a MB.

However, even if the value of QP is 31, which is the upper limit, a number of bits are generated. Indeed, inspecting the actual coded (by $QP = 31$) data of the sign video with a crowded background, 10 to 100 bits were generated for one



Fig. 1 Sign picture composed with background movie.

MB of background region.

Therefore, when the encoded bits exceed the target number of bits, even if the QP is set to 31, maintaining the target bitrate becomes impossible. This phenomenon leads to output buffer overflow or failure of maintaining the target framerate.

Here, a powerful bitrate control method using the COD bit is presented. Originally, the COD bit is decided in the H.263 data structure and set ON when there was no change from the previous frame. On the contrary, by the new method, the COD is set ON mandatorily when the calculated value of QP is bigger than 31. Thus, the MB is not coded and the number of bits for the MB is reduced to only one. This method is called "not-coded coding."

2.3 Related Work

In video encoding, an idea such as adding weight to the important region is not unique. Here, we will call this "priority encoding." There are some works concerning priority encoding for sign language videos.

Schumeyer [4], [5] proposed an encoding idea of distributing a lot of bits to the face or hand region of the signer and carried out a very simple evaluation. For the designation of face and hand regions, two methods were tested. One is simply designating a fixed central region of the picture. The other is automatically detecting the region using color information. The encoding format is H.263 and QP values are limited to two levels.

Muir et al. [2] analyzed the line of sight of people who are watching a signer with an eye camera, and proposed an encoding technique that concentrates heavily on the face and hand regions. Furthermore, they pointed out that for the study of encoding a sign video, establishing subjective evaluation criteria is important instead of conventional objective evaluation methods like the S-N ratio. However, these propositions are no more than just ideas.

[†]In H.26X encoding, pictures are processed by a unit of a small bunch of pixels called macroblocks.

^{††}When the value of QP becomes large, preciseness of the picture is degraded but the number of bits is decreased. When the value of QP becomes small, preciseness is improved but the number of bits is increased.

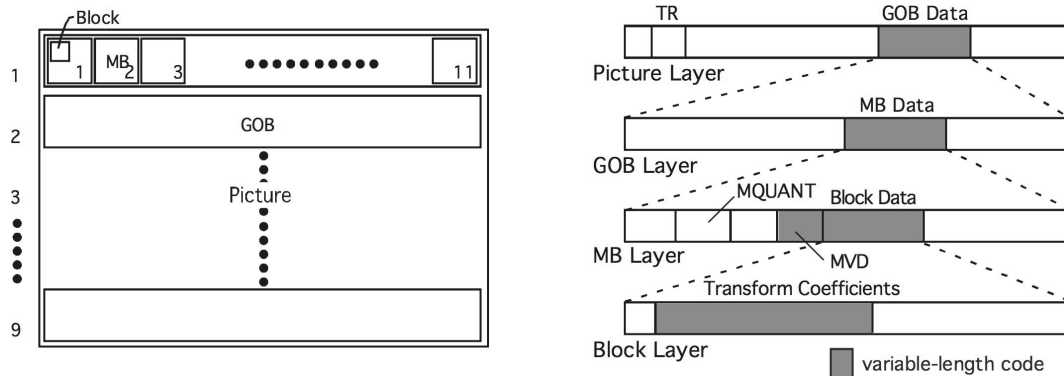


Fig. 2 Geographic structure (of QCIF) and data structure of H.263 picture.

In another paper, Muir et al. [6] proposed three implementation ideas for the priority encoding of sign language video: 1)using object-based coding like MPEG-4 or H.264, 2)controlling each quantization parameter by a MB unit and using MB skip mode, 3)making the unimportant region hazy by preprocessing. The second idea resembles our technique but these ideas have not been implemented.

Tan et al. [7] implemented an improved encoding technique using automatic detection of face and hand regions. Their encoding object is a video of a lecture by teachers. Furthermore, they carried out a simple evaluation. Their automatic detection technique was advanced. However, in thinking of applying this technique to sign language communication using mobile videophones, problems will occur because distinguishing the hands and face of one’s companion from those of other people (in the background) is difficult.

Agrafiotis et al. [8] also implemented automatic detection of a signer’s face and priority encoding. Their technique was implemented using H.264 but object-based coding was not used. An intelligibility test and an opinion test were carried out although these tests were elementary.

Differences between these related works and the authors’ stand points are as follows:

1. We think the combination of three encoding techniques is important and partial adoption of these techniques cannot achieve the full effect.
2. Most of these studies did not pay attention to processing of background area which requires an unexpectedly large number of bits.
3. Most of these studies did not carry out credible evaluations.

3. Detailed Encoding Methods for Sign Language Video

In this section, details of three kinds of specially designed encoding methods for sign language videos will be presented. All of these enhancements can be implemented within H.263 specifications and encoded data can be decoded by a general H.263 decoder.

At first, the data structure of H.263 will be presented. One picture is comprised of four layers, as shown in Fig. 2. A CIF (Common Intermediate Format) picture is comprised of 352×288 pixels. The QCIF (Quarter Common Intermediate Format) has 176×144 pixels. One picture is divided into groups of blocks (GOBs.) A GOB is comprised of 16 lines of pixels. There are 9 GOBs in a picture for QCIF, and 18 for CIF. Each GOB is divided into MBs. An MB is comprised of 16×16 pixels. Thus, a GOB is comprised of 11 MBs for QCIF, and 22 MBs for CIF. One MB is divided into four blocks. A block is comprised of 8×8 pixels. Symbols included in right half of Fig. 2 means as follows: TR is short for Temporal Reference, MQQUANT means quantization parameter of MB, and MVD is short for Motion Vector Data.

3.1 Gradient Coding

Normally, during the encoding process of a frame, the QP of the current MB is calculated on the basis of the index value, which is called the Error. The Error is the difference between the accumulated target number of bits and accumulated actual number of bits generated by the encoding process. The Error takes a positive value when more bits are generated than the target number of bits and a negative value in the opposite case. QP takes an integer value between 1 and 31. Roughly, when the Error takes a large value, QP is set to a large value, and when the Error takes a small value, QP is set to a small value.

In the case of general purpose encoding, when the target numbers of bits are calculated, an equal number of bits are distributed to each MB. On the other hand, gradient coding distributes a different number of bits to each MB. The ratio of the number of bits distributed to every MB is called “weight.” The set of weights indicates the importance of each MB and must be determined prior to the encoding process. If a MB was judged as important and given a heavy weight, the QP of the MB is controlled to take a small value, and then, a detailed picture is obtained in the MB.

3.2 Precedence Macroblock Coding

We can think of various patterns for the encoding order of MBs, for example, starting from a MB in the center of the picture, proceeding like a spiral, and then finishing at the edge. However, we will simplify the order to a two-tier order. The encoding order of GOBs are specified primarily, and the order of MBs are specified subsidiarily. In the first GOB, MBs are processed in a designated order. After all MBs in the first GOB are processed, the second GOB is processed and the process is continued.

Note that the data format of H.263 does not allow an arbitrary MB order. Consequently, the encoded data of each MB is stored once in the buffer that has the same structure as that of the H.263 data stream. Then, after all encoding of MBs in a frame is complete, the data is sent out forming a regular data structure.

3.3 Not-coded Encoding

With this method, when the calculated value of QP is bigger than 31, the COD is set to 1 and the MB is not coded. Thus, only one bit is spent for the MB and powerful information compression is achieved. In the internal encoding processing, the MB is treated so that it does not change from the previous frame.

When the movie, which is encoded in the above way, is played back, the not-coded MB is displayed as stopped. Thereafter, when the MB returns to normal encoding mode, it starts moving again. Consequently, looking at the whole screen, the stopped MB appears here and there.

However, if the MB near the viewer's point of gaze is stopped, a strange feeling could be given to the viewer. To prevent this phenomenon, not-coded coding can be applied only at designated MBs, normally the corners of the picture.

4. Confirmation of Effectiveness by Experiments

4.1 Evaluation Experiments of Sign Language Video

Sign is a kind of natural language that transmits linguistic information using human vision. When we evaluate the quality of sign video, we should be careful not to be swayed by the naive impression of the appearance of the video. We must evaluate how well the linguistic information was transmitted. Thus, the voice evaluation technique was treated as a model of study instead of the image evaluation technique. Details of evaluation techniques are reported in our paper [9].

Two kinds of evaluations, the intelligibility test and the opinion test, were accomplished.

The processes of the intelligibility test are as follows: 1)a short video sequence of sign language was presented to subjects, 2)subjects were instructed to write down the content of the sentences, 3)dictated sentences are evaluated and

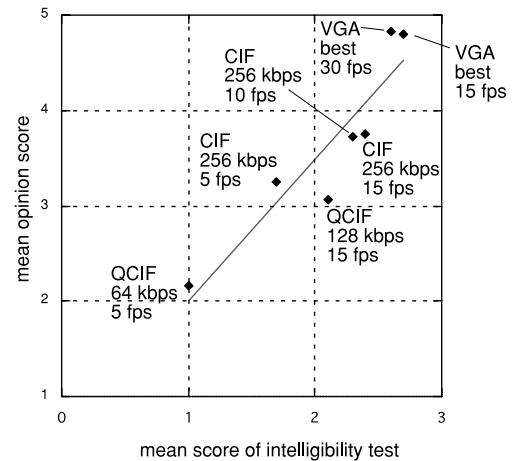


Fig. 3 Comparison of score between intelligibility test and opinion test.

scored at four levels, from 0 to 3. For the scoring of sentences, we must be careful not to be affected by the difference in subjects' ability in written Japanese.

In this test method, the intelligibility of video can be evaluated straightforwardly from the dictated sentences. On the contrary, the method has some problematic points: 1)a sign sentence cannot be presented more than twice to the same subject, 2)the rating process requires a long time, 3)maintaining adequacy and consistency in rating is difficult. For these reasons, carrying out grand-scale experiments across many conditions is difficult.

The processes of opinion test are as follows: 1)a short video sequence of sign language was presented to subjects, 2)subjects were asked to evaluate the intelligibility of the sign data at five levels, from 1 to 5, 3)the mean value of the scores is called the MOS (Mean Opinion Score) and is used for the evaluated value of the data.

In the opinion test, one sign sentence can be presented more than once to the same subject. Furthermore, rating process is not required. Relatively grand-scale experiments can be accomplished by this method because of these reasons. However, the reliability of the results as a measure of sign video quality is not obvious because the source of measurement is a subject's impression.

In our paper [9], sign video sequences were evaluated by both the intelligibility test and the opinion test. The results indicate a strong correlation between these two evaluation methods, as shown in Fig. 3. From these observations, we concluded that the opinion test can be used instead of the intelligibility test for the evaluation of the quality of digitally encoded sign language video.

4.2 Experiment System

The authors are developing platform software to study digital encoding of video on a Macintosh computer with Mac OS X. Using this program, various parameters and encoding techniques can be customized. It is called "AdapSync." Actually, the encoding engine of this program was developed

based on a “TMN,” Test Model Near-term, which is an experimental program package for H.263 encoding that was developed by Telenor Research & Development.

The encoding technique that was implemented in the TMN can be considered as general purpose. From here on, this encoding technique will be called “TMN” and the specially designed encoding technique for sign language will be called “AdapSync.”

Evaluation experiments were carried out with a notebook personal computer (iBook, Apple Computer.) Test data were stored in the computer hard disk in QuickTime movie format. The presentation of movies is controlled by an experimental program that was developed by SuperCard, an application developing software. Furthermore, evaluated scores were input by subjects clicking icons directly. The collection and processing of result data were also executed by the computer.

4.3 Making of Material Data

Japanese Sign Language (JSL) conversations by deaf people were recorded as the material of experimental data. Scenes in which two people were conversing facing each other were recorded using prompters instead of standing alone in front of the video camera. The communication partner’s picture was displayed on each person’s prompter. For convenience, the two people are called “the Signer” and “the Dialoguer.” Furthermore, video of a sign word was displayed on the signer’s prompter as a seed of sign conversation[†]. The Signer was instructed to start conversation associated with the word. The Dialoguer was instructed to give responses or to ask the Signer questions. For evaluation experiments, video data of “the Signer” were adopted.

Some short video sequences were selected from the long sign videos based on the following criteria: 1) pure JSL is shown^{††}, 2) very common sign expressions are used, 3) the video sequence has a length that is suitable for experiments, about 7 or 8 seconds. The first and second criteria were evaluated by CODAs (Child Of Deaf Adults.)

Although the original videos were recorded in the studio with a blue screen in the background, a recorded scene of many people passing through a crowded street was composed to simulate a conversation by videophone on the street, as shown in Fig. 1.

4.4 Encoding of Test Data

In this section, the method of encoding test data from the material data will be presented. Various innovations were introduced in the encoding of new test data because significant differences between two techniques could not be found in our former evaluations [1].

Advantages of AdapSync can be summarized as follows: 1) the preciseness of the important region can be improved while a high framerate is maintained, 2) the framerate can be raised while the preciseness of the important region is maintained. There are trade-offs between these two options.

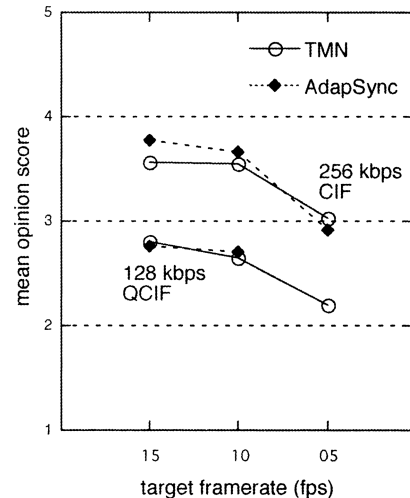


Fig. 4 Comparison of MOS between TMN and AdapSync (results of the previous experiments [1]).

In our paper [9], a lot of sign videos with different encoding parameters were evaluated. The results showed the following: comparing the video with low preciseness and high framerate to the video with high preciseness and low framerate, under the condition of same target framerate, the former was more intelligible.

In our other paper [10], sign video sequences in which the QP were fixed throughout the sequences were evaluated with a fixed framerate of 30 fps^{†††}. The results indicated that the intelligibility was not affected by the degradation of QP.

These consequences imply that it is better to maintain a higher framerate even if the preciseness is degraded, especially in the case of very low bitrate.

In former experiments [1], a pair of video sequences encoded by different techniques (TMN and AdapSync) were built with the same size, the same target bitrate and the same target framerate. The results are shown in Fig. 4. It was matter of course that the efficiency of AdapSync was not verified because both data had almost identical framerates.

In order to verify the efficiency, the best quality data for each technique must be presented under conditions of equal encoding size and target bitrate.

For evaluation experiments described in this paper, test data were newly encoded with retuned parameters based on the knowledge of former experiments. The method of making the data in the case of QCIF size will be described below.

First, a video sequence of sign data with a blue screen in the background was decoded by a fixed QP (31) and fixed framerate (30). The picture was divided into grids by the size of MBs and the correspondence between each MB and

[†]These words were selected from the JSL word database, KOSIGN Ver. 2.

^{††}In Japan, Signed Japanese, a kind of simultaneous communication, is also widely used besides JSL.

^{†††}QPs, which are usually controlled step-by-step, dominate the preciseness of the video.

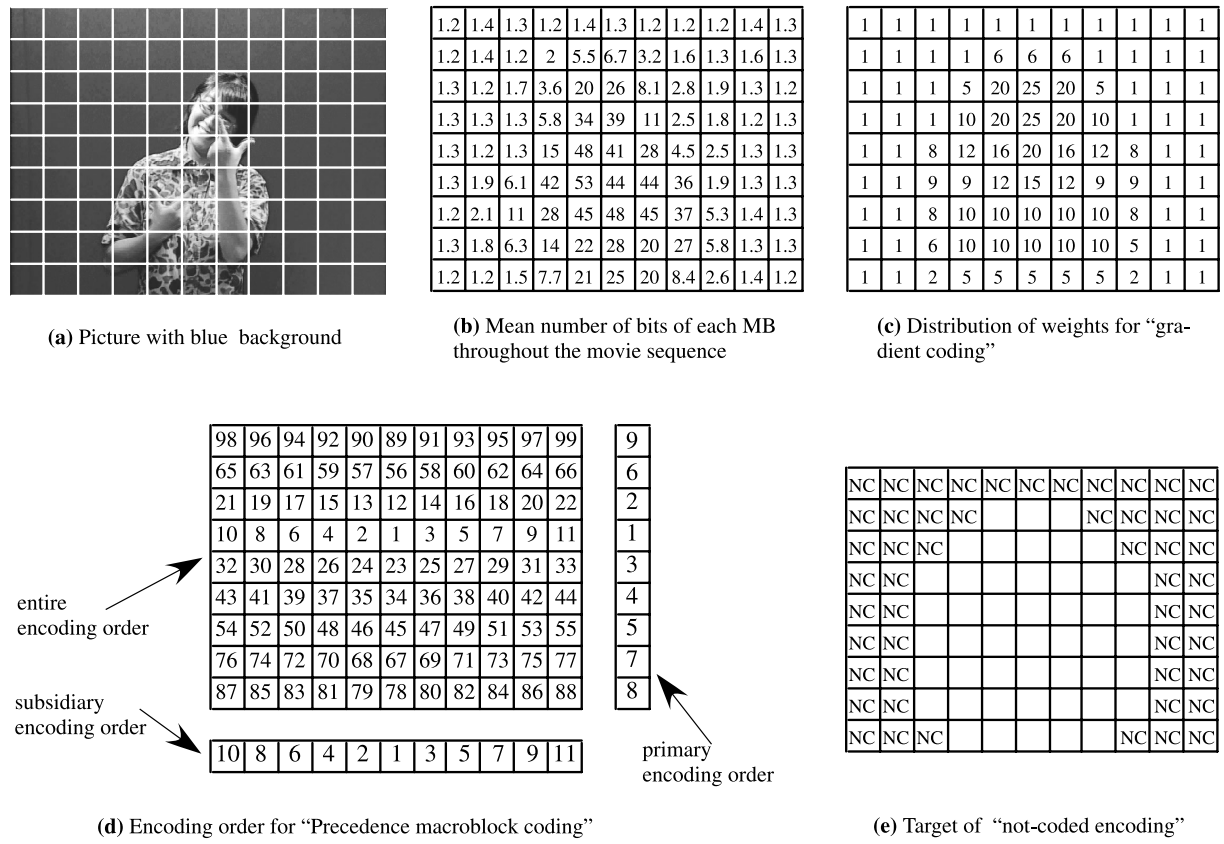


Fig. 5 Sample of the procedure of making up encoding parameters.

each picture segment was made clear, as shown in Fig. 5 (a).

The mean number of bits of each MB throughout the sequence was calculated. Samples of the mean number of bits are shown in Fig. 5 (b). MBs can be classified into two groups: a group of MBs whose mean number of bits is very small, for example, less than two, and the group of MBs whose mean number of bits is relatively big, for example, greater than two. The former corresponds to the background region and the latter corresponds to the region of the picture of the signer. We call the former "background MBs" and the latter "foreground MBs."

Even if people and vehicles move in background MBs, a large number of bits should not be wasted. Here, the ratio of number of bits, weight, of most important MB to background MB was set to 25:1, respectively. Consequently, the weights of the most important MBs around the face of the signer were set to 25. The weights of the other MBs in the foreground were set to some value less than 25, and they gradually declined as the distance from the facial area increased. Finally, weights of the background MBs were set to 1.

In this way, the set of weights that was used in "gradient coding" was determined. A sample distribution of weights is shown in Fig. 5 (c).

Second, the encoding order that was used for "precedence macroblock coding" was determined. The order was determined according to the weight. In this regard, the order

was determined by a two-tier order as stated in Sect. 3.2. A sample of the encoding order is shown in Fig. 5 (d).

Third, MBs that were targets of not-coded encoding were specified. Specified MBs were basically background MBs at the edge of the picture. A sample of targets of not-coded encoding is shown in Fig. 5 (e). MBs marked by "NC" are targets.

A provisional suite of parameters was set up.

Last, some sign video sequences with people and vehicles moving around in the background were encoded for the investigation of the parameters. Video appearance, bitrates, and framerates of obtained movies were examined and the parameters were tuned again. This procedure was repeated several times to improve subjective impression of video quality.

The suite of parameters for QCIF size video for the specially designed encoding technique was obtained by the above procedure. Parameters for CIF size video are also determined by a similar method.

Material data were encoded into the experimental data of AdapSync using these parameters. Predetermined encoding conditions of both techniques are shown in Table 1. Conditions that were actually used for evaluation are marked with a \surd in the table.

Table 1 Table of encoding conditions used for evaluation tests.

bitrate	TMN		AdapSync	
	CIF	QCIF	CIF	QCIF
256 kbps	√		√	
128 kbps	√	√	√	√
64 kbps		√	√	√

Table 2 Evaluation category of opinion test.

Measure	Definition of Categories
5	Very easy to read
4	Somewhat easy to read
3	Somewhat hard to read
2	Hard to read
1	Almost impossible to read

4.5 Evaluation Method

Seventeen native JSL speakers, two in their twenties, five in their thirties, three in their forties, six in their fifties and one in his sixties, were subjects. Sixteen subjects lost their hearing congenitally or younger than three years old. The remaining person lost her hearing between five and seven years old.

Sixteen kinds of sign video were encoded by sixteen kinds of encoding variation[†]. Sixteen kinds of distinct data were selected and presented to subjects. Note that sign data with the same content were presented to subjects only once. Subjects were asked to evaluate the intelligibility of the sign video, not to evaluate the preference of picture quality. Evaluation categories are shown in Table 2.

4.6 Results and Discussion

The comparison of the MOS of two kinds of data, using the usual encoding technique, TMN, and the specialized encoding technique for sign video, AdapSync, are shown in Fig. 6. For ease of observation, the graph is divided into two parts which correspond to each sampling image size. Furthermore, framerates obtained by practical encoding were also plotted.

At first, look at the CIF size case. At 256 kbps, both encodings have almost full marks of 5 points. At 128 kbps, the score for AdapSync degraded slightly, but the score for TMN degraded significantly. As TMN cannot encode the video at a practical framerate at 64 kbps, the data of TMN is not plotted at 64 kbps. The average framerate exhibits the same pattern as that of the MOS.

Next, we look at QCIF size results. At 128 kbps, almost no difference was seen between the two kinds of encoding. At 64 kbps, the evaluation for AdapSync is much better than that of TMN. The average framerate exhibits the

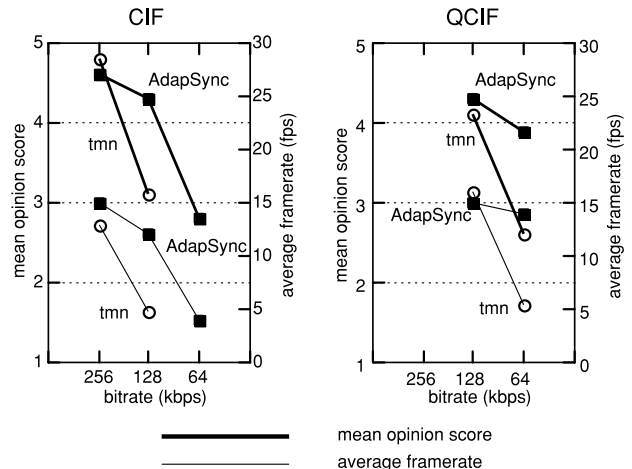


Fig. 6 Comparison of MOS and obtained framerate between TMN and AdapSync.

same pattern as that of the MOS.

These results indicate that AdapSync is effective for the encoding of sign language video. In particular, the effectiveness is noticeable in the case where a sufficient framerate cannot be obtained by the usual encoding technique.

5. Conclusion

Considering telecommunication for deaf people, a video encoding technique specially designed for sign language was proposed. The technique is composed of three kinds of extension: gradient coding, precedence MB coding, and not-coded coding. These methods were designed to meet H.263 standards, and encoded data can be decoded by a general H.263 decoder.

The encoding methods were implemented on a computer and encoded video sequences of sign language were evaluated by deaf subjects. As a result, the efficiency of the technique was confirmed.

These experiments were accomplished in an offline environment. Evaluations were performed by watching the one-way signing. These circumstances are far different from actual conversations that occurs over a videophone. In the future, we will construct a real-time videophone system and carry out evaluations of sign language conversations.

Acknowledgments

We thank the native signers who supported us in recording and analysis.

This report is supported by Grants-in-Aid for Scientific Research (A)(1) #14208030 and Grant-in-Aid for Scientific Research on Priority Areas #16091209.

References

[1] K. Nakazono, S. Yanagihashi, and Y. Nagashima, "Encoding for moving picture of sign language and its evaluation," *Human Interface*, vol.7, pp.131-140, 2005.

[†]Other than the evaluation written in this paper, some encoding conditions were tested in the same experiment. This is the reason why the number of encoding variations is different from the number of checks in the Table 1.

- [2] L. Muir and I. Richardson, "Video telephony for the deaf: Analysis and development of an optimized video compression product," Proc. Tenth ACM International Conference on Multimedia, pp.650–652, 2000.
- [3] Y. Nagashima and K. Kanda, "Present stage and issues of sign linguistic engineering," HCI, pp.387–391, 2001.
- [4] R. Schumeyer, E. Heredia, and K. Barner, "Region of interest priority coding for sign language videoconferencing," First Workshop on Multimedia Signal Processing, pp.531–536, IEEE, 1997.
- [5] R. Schumeyer and K. Barner, "A color-based classifier for region identification in video," Visual Communications and Image Processing 1998, pp.189–200, SPIE, 1998.
- [6] L. Muir, I. Richardson, and S. Leaper, "Gaze tracking and its application to video coding for sign language," Proc. 23rd International Picture Coding Symposium, pp.321–325, 2003.
- [7] R. Tan and J. Davis, "Differential video coding of face and gesture events in presentation videos," Computer Vision and Image Understanding, vol.96, pp.200–215, 2004.
- [8] D. Agrafiotis, N. Canagarajah, D.R. Bull, and M. Dye, "Perceptually optimised sign language video coding based on eye tracking analysis," Electron. Lett., vol.39, no.24, pp.1703–1705, Nov. 2003.
- [9] K. Nakazono, S. Yanagihashi, Y. Nagashima, and A. Ichikawa, "Evaluations of digitally encoded motion picture of sign language," IEICE Trans. Inf. & Syst. (Japanese Edition), vol.J89-D, no.3, pp.541–551, 2006.
- [10] K. Nakazono, H. Yonehara, Y. Nagashima, and A. Ichikawa, "Evaluation of motion picture of sign language—effect of screen size, framerate and quantization step width," IEICE Technical Report, WIT2005-15, 2005.



Akira Ichikawa received the B.E. and Ph.D. degrees in Electronic Engineering from Keio University, Japan, in 1964 and 1981, respectively. In 1964 he joined the research staff of the Hitachi Central Research Laboratory, Japan, where he engaged in all areas of speech processing research as the head of the speech processing and human interface research group. From 1992, he was a professor of the Department of Information and Computer Sciences, Chiba University, Japan. Since 1998, he is a professor of

the Graduate School of Science and Technology, Chiba University. His current research interests include Spoken Language Processing and Assistive Information Technology for Disabled People. He received the IEICE Paper Award in 1989. He is a member of the IEEE, ESCA, the Acoustical Society of Japan, the Information Processing Society of Japan, the Japanese Society for Artificial Intelligence, the Association of Natural Language Processing, the Japanese Association of Sign Linguistics, and the Society of Logopedics and Phoniatrics.



Kaoru Nakazono received his B.E. and M.E. degrees in Biological Engineering from Osaka University in 1980 and 1982, respectively, and Ph.D. degree in Engineering from Chiba University in 2006. He joined the NTT Electrical Communication Laboratories in 1982. He worked for the Advanced Telecommunications Research Institute International (ATR) from 1990 to 1993. He is currently a Senior Research Engineer of NTT Network Innovation Laboratories. His current research interests include

Assistive Technology, Human Cognition and Communication, and Basic Computer Science. He is a member of the Information Processing Society of Japan, and the Human Interface Society of Japan.



Yuji Nagashima received the B.E., M.E., and Ph.D. degrees in Electronic Engineering in 1978, 1980, and 1993, respectively, from Kogakuin University. He joined the Faculty of Engineering at Kogakuin University as an assistant, became a lecturer, then an assistant professor. Since 2003, he has been a professor at the same university. His research interests include Human Interface Technology, Sign Linguistics Engineering, Assistive and Rehabilitation Engineering, and Bioinstrumentation. He is a member

of the Human Interface Society of Japan, IEEE, and ACM.