PRIVACY CONSCIOUS VIDEO COMMUNICATION SYSTEM BASED ON JPEG2000

Suriyon TANSURIYAVONG[1], *Takayuki SUZUKI*[1], *Somchart CHOKCHAITAM*[2] *and Masahiro IWAHASHI*[1]

[1]Nagaoka University of Technology, Nagaoka, Niigata, 940-2188 Japan
[2]Thammasat University, Rangsit, Pathum-Thani 12121,Thailand
E-mail: suriyon@vos.nagaokaut.ac.jp

ABSTRACT

This paper proposes a “privacy conscious” video communication system utilizing core technology of the JPEG2000 international standard for digital image data compression. The system works not only as a conventional video communication tool but also serves as an “awareness communication” tool. In addition, bit rate for transmitting video data and hardware complexity of the system are also reduced since it utilizes “bit-plane decomposition” and “multi-resolution expression” in the JPEG2000. It is confirmed by quantitative measurement that total amount of video data can be reduce to 1/5. The system can be used for monitoring an elderly or a patient in a hospital for 24 hours a day.

1. INTRODUCTION

Recently, there are an always-connected video awareness which have been proposed to support informal-communication between cooperative groups works[1,2].

According to those proposal, several video communication system have been also proposed in the term of protecting a "privacy" of the person in the scene while providing their "awareness" information[3-7]. However, these systems are implemented a video "recognition" such as moving object detection, object's contour extraction, and video "compression" such as the MPEG or JPEG separately, which made a system become large and complicated.

Our idea to reduce the scale of the system is to construct a video recognition and video compression together in one system. Furthermore, it would be better if it can use a part of compression-utility for recognition or vice versa. It might bring an sub-effectiveness for reducing scale and shorting development term of the system.

This paper proposes a “privacy conscious video communication system”; an intermediate system balancing awareness and privacy base on the idea of using technology of JPEG2000. Person region is extracted based on the “multi-resolution expression” of Wavelet transform technique, and transparent image of person is created based on “bit-plane decomposition” technique, which these techniques are a constituent technology of JPEG2000[8-10]. A part of the compression is utilized for “recognition” of an object contributing to reducing hardware complexity and shorting development term of the system..

In this system, the persons who stay far from camera at the sending site are automatically displayed as a high-transparent person at the receiving site, in order to protect their privacy. The system works not only as a conventional video communication tool but also serves as an always-connected video awareness system to support informal-communication of people between remote sites. Also, the system can be used for monitoring an elderly or a patient in a hospital for 24 hours a day.

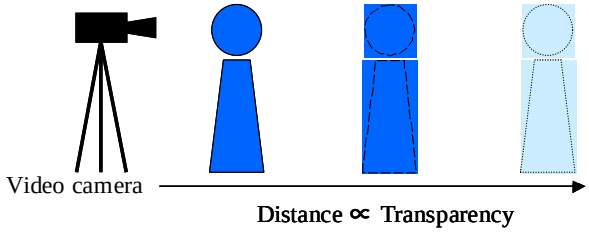


Fig.1: The transparency of object is variable corresponding to the distance from camera.

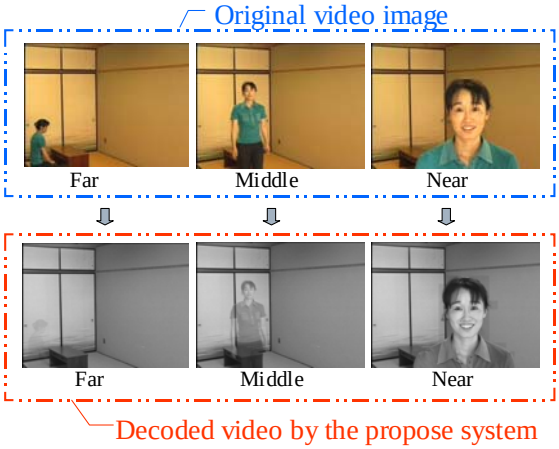


Fig.2: Illustrated of system’s specification.

2. PRIVACY CONSCIOUS VIDEO COMMUNICATION SYSTEM

The privacy conscious video communication system is a system supports informal communication between the networked remote sites. It has functions as shown in Fig.1. In this system, as shown in Fig.2, the transparency of object at receiving site is proportionally depending on the distance between object and camera at the sending site. The

useful cases of using variable-transparent object are described as below.

1. In the case when the members who want to communicate each other, they come close to the camera. In this case, at the receiving site, [non-transparent object] is displayed and members can talk each other like as a conventional video communication system.

2. In the case when the member who has intention not to be talked, or not to be observed, he/she stays far from camera, then [transparent object] is displayed at the receiving site. Displaying the transparent-object has a merit to protect a privacy of person who intents not participate in the talk.

3. In the case that the member who has no intention to talk right that time, but he/she is ready to response if someone calls. In this case [semi-transparent object] is displayed at the receiving site. The displaying of semi-transparent object is the so-called "an awareness communication". It is a new type of communication means [11].

This system is a one of sample system that can cope with our proposal. At the sending site, we use the fixed camera. Therefore, the background image is almost fixed, so, it is not necessary to send background image in real-time. Firstly, we send the background image and store it in the memory at the receiving site.

We use the distance information between object and camera, which is decided from the size of person region to adjust the degree of the transparency. We create a bit stream by making the person region corresponding to the JPEG2000 tile (the small rectangular region), and embedding the transparency information into tile's header (tile part header). We encode and send only the tiles that corresponding to person region, avoid sending non-important data in order to reduce transfer bit rate.

At the receiving site, the tile part header of bit stream from the sending site is analyzed, and then the tile information and the transparent information are extracted. By combining these information with the background image, which once has been stored in the memory, the reconstruction of transparent person with background image is carried out.

Next, we describe the technique of using JPEG2000 technology in detail for signal processing.

3. SIGNAL PROCESSING BY PROPOSED METHOD

3.1 Object Region Extraction Using "Multi-Resolution Expression" in JPEG2000

Generally, using the background image subtraction to extract the foreground image usually raises a hole in the region, in particular, when the background image is similar to the foreground image. The problem discussed here is how to extract the foreground image with the minimum implementation cost. In order to cope with this issue, propose method employed the multi-resolution expression by wavelet transform, which already included in technology of JPEG2000.

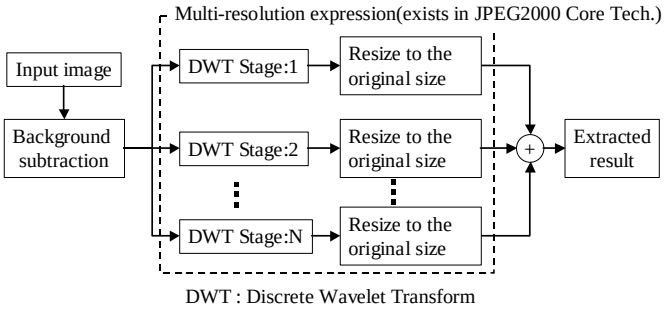


Fig.3: Moving person extraction by using multi-resolution expression(proposed method).

Figure 3 shows the using multi-resolution expression to extract moving object. We execute wavelet transform for the result image that obtained from the background subtraction. Once the wavelet transforms has been executed, the vertical and horizontal direction of the low-pass and high-pass, total 4 images are created. For one stage of wavelet transform, the resolution of each image is scaled to a half size. We pick up the low-pass image, and repeat the processing stage of wavelet transforms to the low-pass image. Therefore, according to the number of processing stage:N, the low-pass images in difference size are created. Finally, restore each of them to the original size, and then reconstruct them into one image as an original by the logical OR operation.

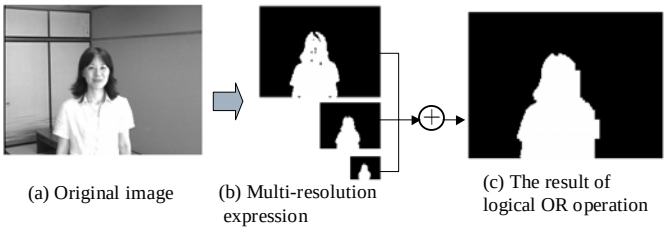


Fig.4: Moving Object extraction results. Undesired holes in the object region are reduced by proposed method.

As shown in Fig.4, the result of moving object/person extraction using multi-resolution expression is able to reduce an undesired hole in the obtained region; the effectiveness of this method is confirmed. After that, the labeling process is performed, and the amount pixel in the object region is calculated, as the "total pixel in the object region". We use this value to set the transparency rate, as shown in Table1.

Table 1: Transparency rate of the object is depending on distance between object and camera.

Distance	Total pixel in the object region: A[pixel]	Transparency rate: T
Near	A >15000	$T=2^{0}$
Quite near	15000≧ A > 10000	$T=2^{-1}$
Middle	10000≧ A > 6000	$T=2^{-2}$
Quite far	6000 ≧ A > 3000	$T=2^{-3}$
Far	A ≦ 3000	$T=0$

There are several idea in order to control the transparency rate. For instance, using multi camera as a stereo vision, to calculate the position of the person, and

then control the transparency rate, and so forth. However, in this paper we use single camera for computationally feasible and relatively inexpensive.

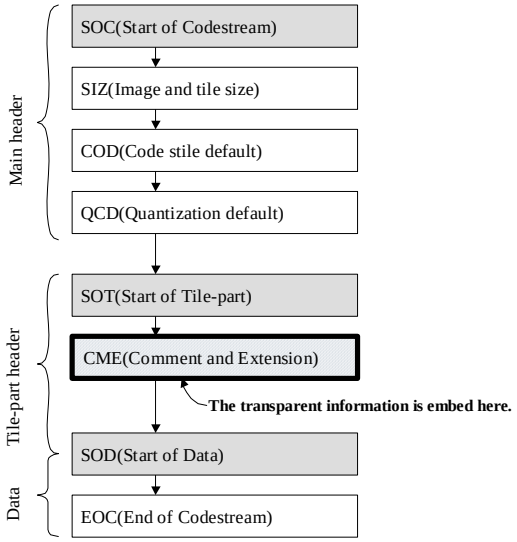


Fig.5: The bit stream composition of the system.

3.2 Transparency Rate Control using "Bit-plane decomposition" in JPEG2000

In the JPEG 2000 compression algorithm, each frame is decomposed into adjacent small rectangular called "tile". The proposed system classifies each tile into "object" or "background". In this paper, all the tile which contains a part of (or all) the person region are categorized into "object". Identifier of "object or not" is embedded into CME (comment and extension) header at tile-part header in the JPEG2000 bit-stream composition as shown in Fig5. This system encodes and sends only the tiles that corresponding to object region, in order to reduce transfer bit rate of video data.

Transparency rate(T) of an object is controlled according to table1 under the assumption that the total of pixels in the object region is related to the distance between object and camera. At the receiving site, the tile part header of bit stream from the sending site is analyzed, and then the tile information and the transparent information are extracted. A receiver decodes the object tile and adds it to previously received background by

$$\begin{aligned} Transparent\,output\,signal &= Object\,signal \times T \\ &+ Background\,signal \times (1-T) \end{aligned} \quad (1)$$

This procedure brings an advantage that the system does not directly extract contour of an object. Just a group of tiles, which totally covers an object, is recognized and reconstructed as an original object. A viewer can see a transparent object with its exact contour. Namely, the system does not require any complicated contour extraction procedure such as Ref.[12-14].

4. EXPERIMENTAL RESULTS

Experiment results in this paper were given under the environment listed in table2.

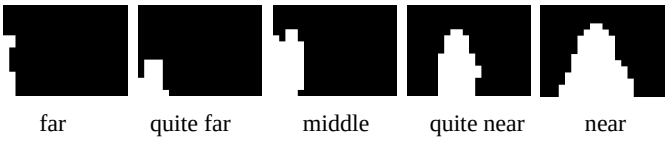
Table2: System Environment

OS	WindowsXP
CPU	Pentium4 2.0Ghz
RAM	512MB
Software	Visual C++ 6.0
Movie	Video image 1,110 frame(320×240pixel, 15frame/sec)
Tile size	16×16pixel
Coding	Lossless coding with 5/3 wavelet filter

4.1 Person-region recognition

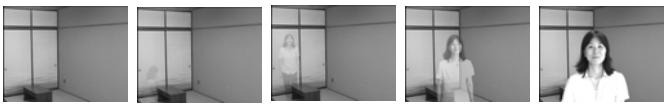
First, we evaluate the person-region recognition process. The way to evaluate is to judge whether person tiles cover the whole person region or not. If the answer is yes, it means that the recognition is successful.

The result of person tiles in recognition process is shown in Fig.6(a). These tiles are successfully extracted. There is no undesired holes in the region. The transparent persons are shown in Fig.6(b) as the final results. We can confirm that the transparency of the person is depending on the distance from the camera. In the total 1110 frames of the video data, the person recognition rate is 92.9%.



far quite far middle quite near near

(a) Tiles recognized as object region.



far quite far middle quite near near

(b) Video signal decoded by the proposed method.

Fig.6: Simulation results for a video signal. When the object comes to the camera, it gradually displays as a non-transparent object.

4.2 Transfer bit rate for communication

Next, we evaluated total bit rate to be transferred to a receiver. For encoding in both existing method and propose method, we used lossless compression of JPEG2000 with 5/3 wavelet filter.

The reason that we used lossless compression in this paper is to confirm if proposed algorithm is work properly or not. Since, some results can not be confirmed in the case of using lossy compression, such as the tile noises that occurred around the person region, and so on.

We evaluated transfer bit rate of image as below.

4.2.1 Bit rate for sending only person's tile

Figure 7 shows transfer bit rate in the case of sending only person's region tile. It is confirm that sending only person's tile is able to reduce the transfer bit rate because amount of data is depending on the distance between the person and camera.

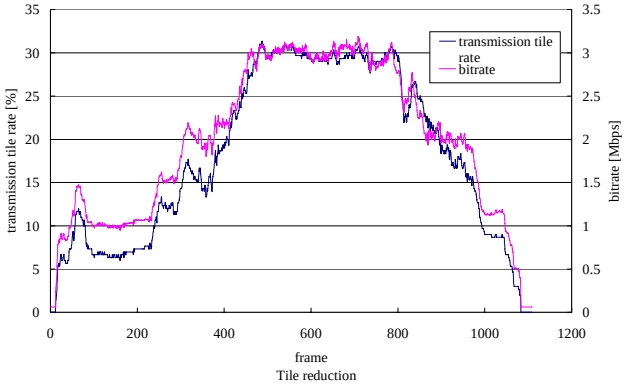


Fig.7: Bit rate for sending only person's tile

4.2.2 Bit rate for sending only the upper bit plane(semi-transparent image)

Figure8 shows the bit rate of sending only upper bit plane. It is confirmed that sending only upper bit plane (transparent image) is also able to reduce bit rate of video data.

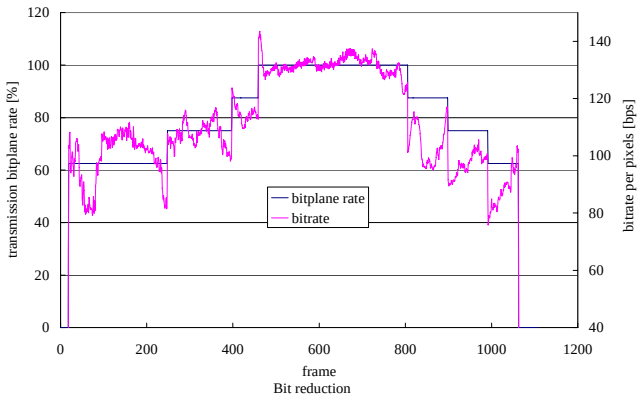


Fig.8: Bit rate for sending only the upper bit plane(semi-transparent image)

4.2.3 Bit rate for sending only a part of bandwidth

Figure 9 shows bit rate in the case of only sending a part of bandwidth as an awareness information. When object come close to camera, the LL part of 3rd , 2nd and 1st state is send.. The total amount of pixel is (0.25)*2, (0.25)*1, (0.25)*0. Therefore, it is confirmed that ,sending only a part of bandwidth is also able to reduce bit rate of data.

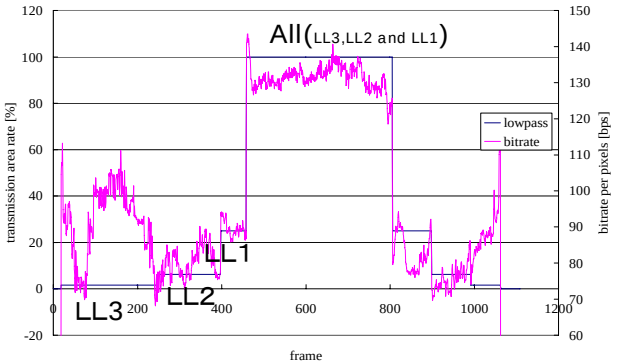


Fig.9: Bit rate for sending only a part of bandwidth(low pass image)

4.2.4 Compressibility

Table3 shows data size and compress rate of 1110 frames video data. Proposed method performs 1/5 compress rate compared to the existing method. The effectiveness of the proposed method has confirmed by quantitative measurement.

Table 3: Comparison of compressibility

Encoding method	Data size [MB]	Compress
Original video image (non-compress)	86.4	1.0
Existing method	37.1	0.4
Proposed method 1	7.5	0.08
Proposed method 2	6.7	0.07

5. CONCLUSION

In this paper, a new communication system which works not only as a conventional video communication but also as an "awareness communication" was proposed. The system is "privacy conscious" and it was constructed with core technology of the JPEG2000 standard utilizing "bit-plane decomposition" and "multi-resolution expression" in the JPEG2000. It is confirmed that total amount of video data can be reduce to 1/5. The system can be applied to use in the filed of welfares, for example, monitoring and nursing-care for an elderly or a patient in a hospital, etc. It is our future work to adapt quantization and bit truncation, namely lossy compression, to the system for more bit rate reduction.

In addition, the variation of displaying awareness image are shown in Fig.10.

6. ACKNOWLEDGEMENTS

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (C), 17560334, 2005-2006.

7. REFERENCES

[1] Robert Fish, et al: "The Video Window System in Informal Communication", ACM Conf. Computer Supported Cooperative Work, CCSCW '90, ACM, pp.1-11, (1990).

[2] Paul Dourish: "Extending Awareness Beyond Synchronous Collaboration", Workshop on Awareness in Collaborative Systems CHI'97,(1997).

[3] Michael Boyle, Christopher Edwards and Saul Greenberg: " The Effects of Filtered Video on Awareness and Privacy", Proceedings of the CSCW'00, ACM,(2000).

[4] Suriyon Tansuriyavong and Shin-ichi Hanaki: "Privacy protection by concealing persons in circumstantial video image", Workshop on Perceptive User Interfaces (PUI'01),(2001).

[5] S. Hanaki, S. Tansuriyavong, S.Tsubaki: "Human Image Concealment Depending on Position for Video Awareness Systems", Supplement Proceedings of the Eighth European Conference on Computer Supported Cooperative Work, pp.31-33, (2003).

[6] Stefan Junestrand, et al, "Private and Public Spaces -

The Use of Video Mediated Communication in A Future Home Environment", Proc. CHI 2000, HAAG, ACM press, 2000.
[7] K.Yabuta, H.Kitazawa, S.Tanaka, "A Fixed Monitoring Camera Image Processing Method Satisfying both Privacy Protection and Object Recognition", Tech. Report of IEICE,SIP 2005-3, pp.13-18April, 2005.
[8] D.S.Taubman, M.W.Marcellin, JPEG 2000 - Image Compression fundamentals, standards and practice, Kluwer Academic Publishers, 2002.
[9] ISO/IEC FCD15444-1,"JPEG2000 IMAGE CODING SYSTEM,"March.2000.
[10] A. N. Skodras, C. A. Christopoulos and T. Ebrahimi "JPEG2000: The upcoming still image compression standard", Pattern Recognition Letters, Volume 22, Issue 12, pp. 337-1345, October 2001.
[11] Uwe M. Borghoff, J.H. Schlichter, U.M. Borghoff, Johann H. Schlichter ,Computer-Supported Cooperative Work: Introduction to Distributed Applications. Berlin; Tokyo: Springer, 2000.
[12] M. Kass, A. Witkin and D. Terzopoulos , "Snakes: active contour models", Int. J. Comput. Vision 1 4, pp. 321-331, (1987).
[13] S.Araki, T,Matsuoka, N.Yokoya, H.Tamemura, "Real-Time Tracking of Multiple Moving Object Contours in a Moving Camera Image Sequence", IEICE Trans. Inf. & Systems,vol.E83-D, no.7, pp.1583-1591, July 2000.
[14] A.E. Hassanien, M.Nakajima, "Feature-Specification Algorithm based on Snake Model for Facial Image Morphing", IEICE Trans. Inf. & Syst., vol E82-D, no.2,pp.439-446, Feb. 1999.

(a) Example 1: Contoured display（only HH1*）

(b) Example2： Blurred and semi-transparent display （LL3, LL2**, only LL1 and semi-transparent）

(c) Example3: Contoured, blurred and semi-transparent（HH1, HL1, LH1, LL2）

Fig.10: The variation of displaying awareness image base on JPEG2000.

*HH1 means HH(high pass) bandwidth of 1st state wavelet transform.
**LL2 means LL(low pass) bandwidth of 2nd state wavelet transform.