Functional Layered Video Coding for Privacy Conscious Video Communication System

Suriyon TANSURIYAVONG[1], Takayuki SUZUKI[2], Somchart CHOKCHAITAM[3] and Masahiro IWAHASHI[2]

[1]Okinawa National College of Technology, Nago, Okinawa, 905-2192 Japan.(suriyon@okinawa-ct.ac.jp)
[2]Nagaoka University of Technology, Nagaoka, Niigata, 940-2188 Japan.
[3]Thammasat University, Rangsit, Pathum-Thani 12121,Thailand.

ABSTRACT

The present paper proposes a new video communication system that transmits awareness at very low bit rate and displays detailed video information by sending additional information on demand. The term awareness denotes information without details related to privacy and expresses the existence or motion of a person. The proposed system decomposes the region representing a person into (a) a basic layer, which contains awareness, and (b) an enhanced layer, which contains other detailed information. This processing is implemented by the basic constituent technology of JPEG2000, including wavelet transform and bit plane decomposition. The basic layer contains (1) the region of the person, (2) the bit plane component, and (3) the bandwidth image component. Therefore, the awareness information is transmitted at a very low bit rate. It is confirmed by quantitative measurement that the bit rate can be reduced to 1/10 that of the exiting method. The proposed system can be applied in the field of welfare, particularly, to remotely monitor patients and to provide remote nursing care for the elderly who lives alone.

Keywords: awareness, layered, video coding, JPEG2000, privacy

1. INTRODUCTION

Recently, always-connected video awareness has been proposed to support communication between remote sites [1,2].

However, always-connected video raises concerns regarding privacy. Several video communication systems have also been proposed in term of protecting the privacy of individuals in the scene, while providing his/her awareness information [3-7]. However, these systems are implemented, separately, using video recognition methods such as moving object detection, object contour extraction, and video compression software, such as MPEG, H.261, and H.264, which has caused systems to become larger and more complicated. Because of the focus on providing awareness, the data transfer bit rates of these systems are very high.

The present paper proposes a privacy conscious video communication system that balances awareness and privacy based on JPEG2000 technology. The person region is extracted into (a) a basic layer, which contains awareness, and (b) an enhanced layer, which contains other detailed information, by using the multi-resolution expression of wavelet transform and the bit-plane decomposition technique, both of which are constituent technologies of JPEG2000 [8-10]. The basic layer is focused to convey awareness information and contains (1) the person region, (2) the bit plane component, and (3) the bandwidth image component. Therefore, the awareness information is transmitted at a very low bit rate.

In the proposed method, the compression component contributes to object recognition in order to reduce the system complexity and shorten the development term of the system.

2. PRIVACY CONSCIOUS VIDEO COMMUNICATION SYSTEM

The privacy conscious video communication system is a system that provides awareness information, while protecting the privacy of individuals in the images. The system works as a conventional video communication tool and serves as an always-connected video awareness system to support communication between individuals located at remote sites. In the proposed system, an individual located far from the camera at the sending site is automatically displayed as awareness information (the image is semi-transparent, blurred, or contoured, for example) at the receiving site, in order to protect the privacy of the individual.

This system has the functions shown in Fig. 1. The cases in which the proposed system is particularly useful are described below.

1. Case in which an individual who wants to communicate with an individual at a remote site approaches the camera. In this case, at the receiving site, a detailed “video” image is displayed and the individuals can communicate with each other as in a conventional video communication system.

2. Case in which an individual who does not plan to talk but is ready to respond if someone calls. In this case the individual is located far from the camera, and only awareness information (existence or motion, for example) is sent to the receiving site, at a very low bit rate, without privacy-related details.

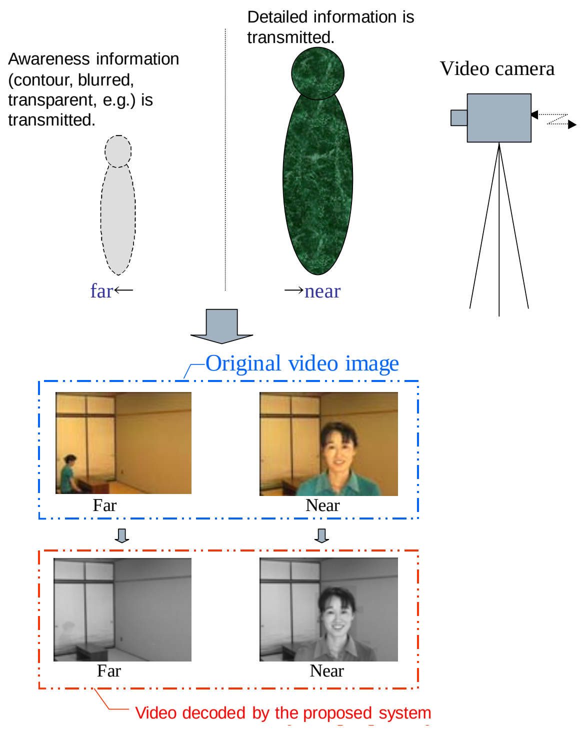


Fig.1: *Configuration of the privacy conscious video communication system.*

The display of awareness is referred to as awareness communication. Awareness communication is a new type of communication [11]. In case 2, the display of awareness has the advantage of protecting the privacy of the observed individual.

At the sending site, we use a fixed camera. Therefore, the background image is approximately fixed, so that it is not necessary to send the background image in real time. First, we send the background image and store it in memory at the receiving site.

We then create a bit stream of the data corresponding to person region, and embed the awareness information into the tile's header (tile part header). Next, we encode and send only the data corresponding to the person region in order to reduce the transfer bit rate.

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

3. SIGNAL PROCESSING BY PROPOSED METHOD

3.1 Person Region Extraction Using Multi-resolution Expression in JPEG2000

Generally, using background image subtraction to extract a foreground image creates a hole in the region; especially when the background image is similar to the foreground image. Here, we examine how to extract the foreground image with the minimum implementation cost. The proposed method employs multi-resolution expression by wavelet transform, which is included in the technology of JPEG2000.

Figure 2 shows subband decomposition using wavelet transform. We execute wavelet transform for the input image and the background image. Once wavelet transform has been executed, low-pass and high-pass images for the vertical and horizontal directions (a total of four images) are created. In one stage of wavelet transform, the resolution of each image is scaled to half size. We select the low-pass image, and repeat the processing stage of wavelet transform on the low-pass image. Therefore, according to the number of processing stages, N, low-pass images of different sizes are created (in the present paper, N=3).

Then, we execute image subtraction between the current input image and the background image of each stage. The results of background subtraction images are shown in Fig. 3(c).

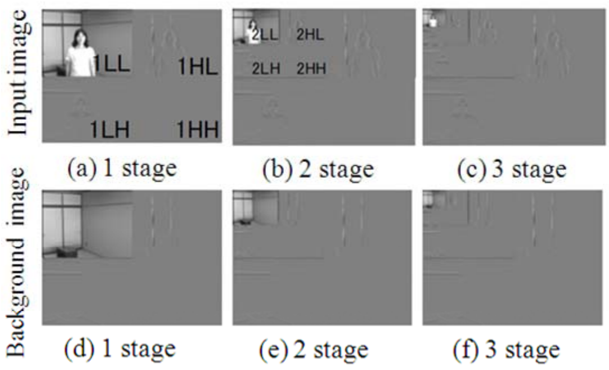


Fig.2: *Subband decomposition using wavelet transform.*

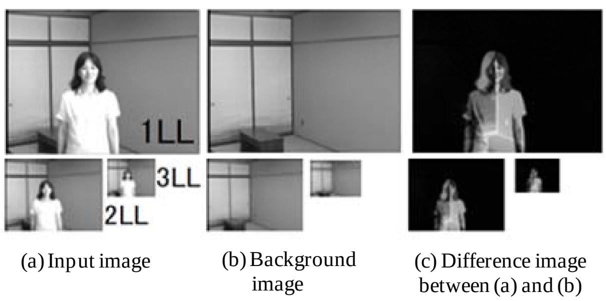


Fig.3: *Image subtraction between the current input image and the background image.*

Next, we execute binarize and enlargement processing for each image of Fig. 3(c). We also replace zero value (zero padding) to the non-person region. Finally, each of the images is restored to its original size, and the images are then combined into one image by the logical OR operation as shown in Fig. 4.

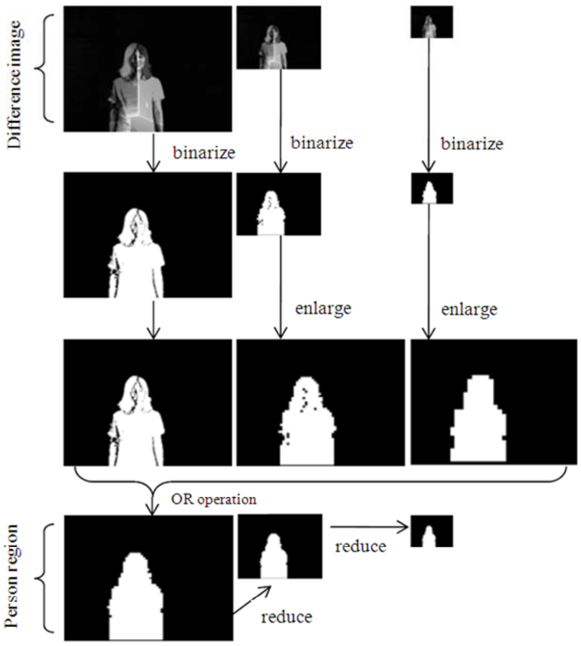


Fig.4: *Moving object extraction using wavelet transform.*

As shown in Fig. 4, moving person extraction using multi-resolution expression can reduce undesired holes in the obtained region. The effectiveness of this method is thus confirmed. The labeling process is then performed, and the number of pixels in the object region is calculated as the total number of pixels in the person region. Based on this value, the transparency rate (T) is set, as shown in Table 1.

Several methods, such as the use of multiple cameras to achieve stereo vision and calculation of the position of the person, have been suggested for controlling the transparency rate. However, in the present paper, we use a single camera for the sake of remaining computationally feasible and relatively inexpensive.

Even if there are two or more people in the scene, the system will work properly to protect the privacy in the case that total number of pixels(A) in each blob (person region) is less than 15,000[pixel]. However, if those blobs are merged into a huge blob (A>15,000 [pixel]), the system will display detailed information even though they stay far from camera. This issue can be addressed by extra processing using the template of upper body to detect if the person is really in front of the camera.

Table 1: *Transparency of the person depends on the total number of pixels in the person region.*

Distance	Total number of pixels in the person region: A [pixel]	Transparency rate: T
Near (non-transparent)	A >15,000	$T = 2^0$
Far (transparent)	A≦15,000	$T = 2^{-n}$ where n = 1, 2, …, 8 ("n" is set by user)

3.2 Transparency Rate Control using Bit-plane Decomposition in JPEG2000

In the JPEG2000 compression algorithm, each frame is decomposed into small, adjacent rectangles called tiles. The proposed system classifies each tile as belonging to either the object category or the background category. In the present paper, all of the tiles that contain part of (or all of) the person region are categorized as belonging to the object category. The object identifier is embedded in the comment and extension (CME) header of the tile header in the JPEG2000 bit-stream composition, as shown in Fig. 5. In order to reduce the transfer bit rate of the video data, this system encodes and sends only the tiles that correspond to the object region.

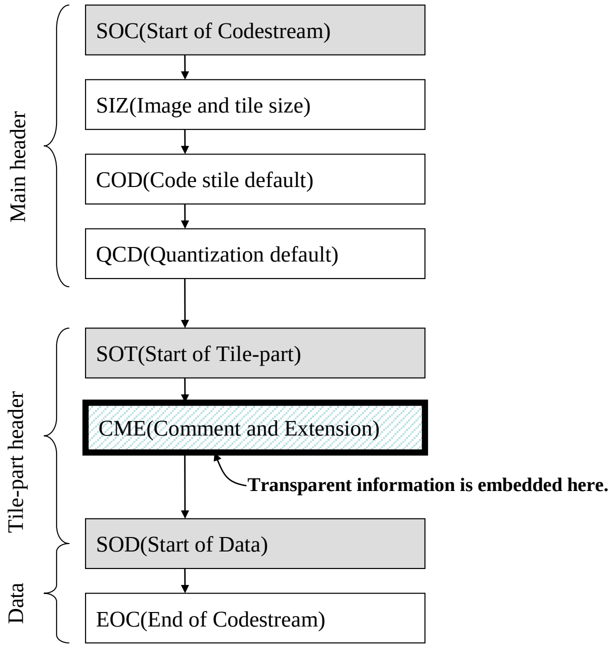


Fig.5: *Bit stream composition of the system.*

Equation (1) shows the transparency rate(T) of an object that is controlled according to Table 1. At the receiving site, the tile header of the bit stream from the sending site is analyzed, and the tile information is then extracted. A receiver decodes the object (person) region and adds it to the previously received background as follows:

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

This procedure has an advantage in that the system does not directly extract the contour of an object. A viewer can see a transparent object by its contour. In other words, the system does not require any complicated contour extraction procedure, such as those described in Refs. [12-14].

4. EXPERIMENTAL RESULTS

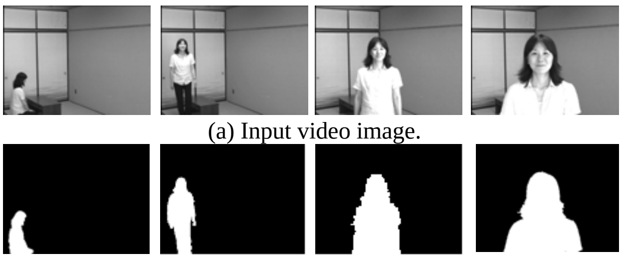
The experimental results of the present paper were obtained under the environment listed in Table 2.

Table 2: *System Environment*

OS	WindowsXP SP2
CPU	Pentium4 2.8 GHz
RAM	512 MB
Software	Visual C++ 6.0
Movie	Video image 1,110 frames (320*240 pixels, 30 frames/sec)
DWT	5/3 wavelet filter

4.1 Person region extraction

First, we evaluate the person region extraction process. Figure 6(a) shows examples of input images for 100, 320, 400, and 490 image frames. The person regions are successfully extracted as shown in Fig. 6(b). There are no undesired holes in the region. It is confirmed that proposed method is effective in reducing holes in extracted objects.



(a) Input video image.

(b) Person region extracted by the proposed method.

Fig.6: *Moving person extraction results. Undesired holes in the object region are reduced by the proposed method.*

4.2 Bit rate when sending only the person region

Next, we evaluated the total bit rate to be transferred to a receiver by sending only the person region.

Figure 7 shows the transfer bit rates of the existing method and the proposed method. The existing method sends all regions of the image, so the transfer bit rate is always approximately 7-8 [Mbps]. In contrast, the proposed method sends only the region of interest, so the maximum bit rate is approximately 3.6 [Mbps].

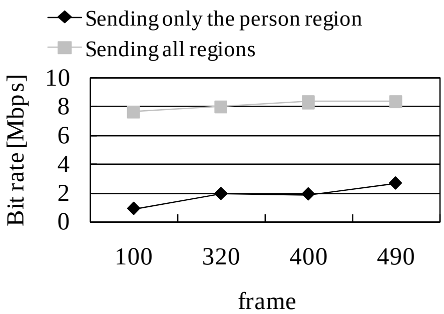


Fig.7: *Effectiveness of sending only the person region.*

Figure 8 shows the respective compression ratios for sending only the person region and for sending all regions. The figure shows that the bit rate is reduced in proportion to the size of the person region. It is confirmed that the proposed method is able to reduce the bit rate to less than half that of the existing method.

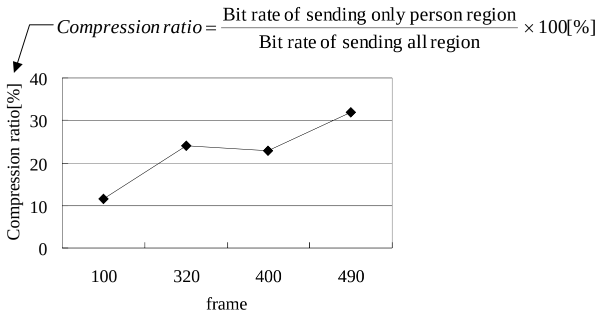


Fig.8: *Compression ratios for sending only the person region and for sending the entire region of the image.*

4.3 Bit rate of sending only part of the bit plane

The total bit rate for sending only part of bit plane is evaluated. In this case, we set both the proposed method and the existing method to send the entire bandwidth and the entire background image under the same conditions.

As shown in Fig. 9, the proposed method (sending only part of the bit plane) can reduce the bit rate to 1/4 that of the existing method (sending the entire bit plane).

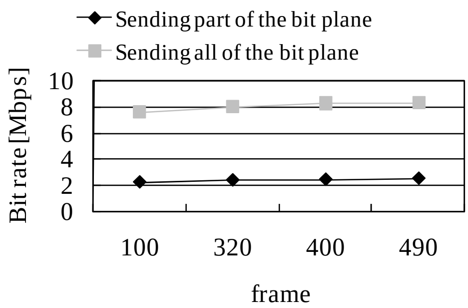


Fig.9: *Effectiveness of sending only part of the bit plane (5-bit plane)*

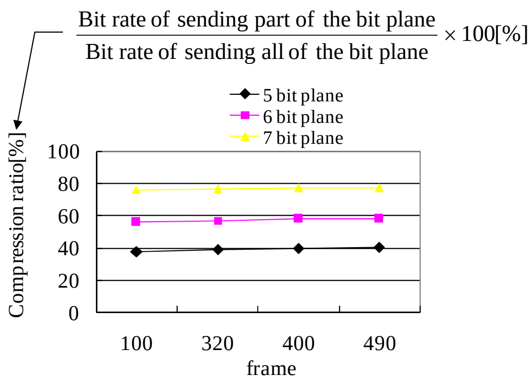


Fig.10: *Compression ratios for sending part of the bit plane and for sending the entire bit plane.*

Figure 10 shows the compression ratio for sending part of the bit plane and that for sending the entire bit plane. The bit rate is reduced in proportion to the number of bits of the plane. It is confirmed that reducing the number of bits of the plane is effective for reducing the bit rate.

4.4 Bit rate for sending only part of the bandwidth

The total bit rate for sending only part of the bandwidth is evaluated. In this case, the entire bit plane and the entire background image is set as the same condition for both the proposed method and the existing method.

Figure 11 shows the bit rate for sending only part of the bandwidth. It is confirmed that sending only part of the bandwidth is able to reduce the bit rate to approximately 1/4 that of the existing method.

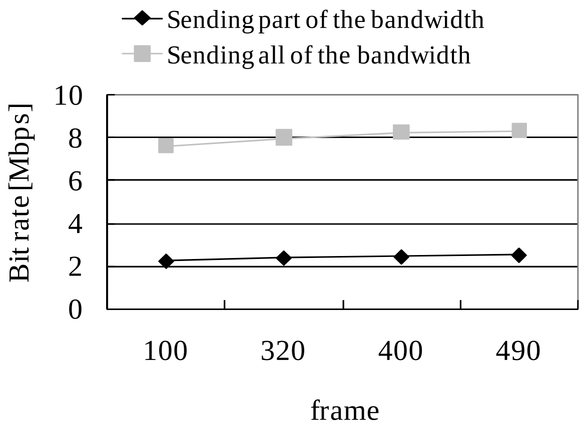


Fig.11: *Effectiveness of sending only part of the bandwidth (3LL)*

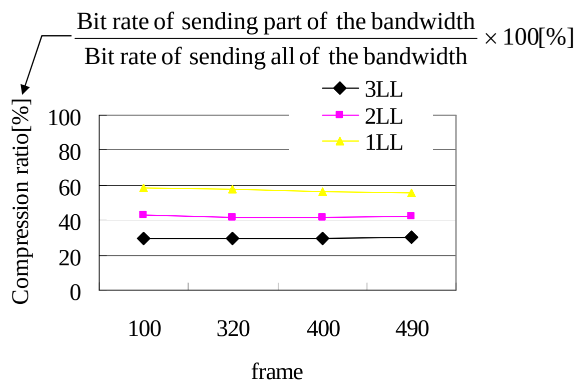


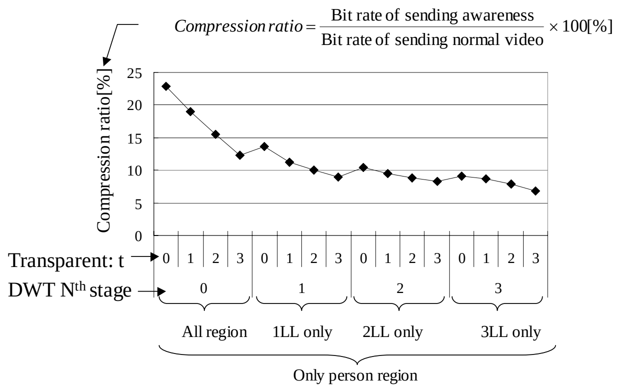
Fig.12: *Compression ratios for sending part of the bandwidth and for sending all of the bandwidth.*

Figure 12 shows that the bit rate is reduced in proportion to the stage of the wavelet transform. It is confirmed that the system is able to reduce the bit rate to below 60% of that of the existing method.

4.5 Combination of compression ratio

Figure 13 shows the combination of compression ratio for sending awareness and normal video. It is confirmed that the bit rate for sending awareness information can be reduced to 1/10 that of sending normal video. This compression ratio is much better than sending awareness in the existing system called distortion filtration [3].

The distortion filtration system is proposed as a possible means for striking a balance between the desire to provide teleawareness and the need to preserve privacy. It is the process of algorithmically manipulating the contents of video streams to selectively obscure various levels of detail in the picture. Many filtration effects are possible: pixelization; smoothing; convolutions (e.g., Sobel, Laplace operations); overlays; and, translatory distortions (e.g., wave, ripple, and fish-eye) are available. The bit rate for sending awareness information of distortion filtration system can be reduced to just 1/3 that of sending normal video.



Transparent: t=*0 means non-transparent*
Transparent: t=*1 means 1/2 transparent*
Transparent: t=*2 means 1/4 transparent*
Transparent: t=*3 means 1/8 transparent*

Fig.13: *Combination of compression ratio for sending awareness and normal video.*

4.6 Complexity

We used the original kakadu software [8] as a basic JPEG2000. On the other hand, we modified the same software and used as a Modified JPEG2000 (proposed method). We ran video sequences which are in avi format. Table 3 compares the computational complexity of basic JPEG2000 and Modified JPEG2000.

The results have been obtained by profiling the software, and therefore depend on the optimization level of the respective codes.

Table 3: *The execution time of basic JPEG2000 and Modified JPEG2000.*

	JPEG2000	Modified JPEG2000 (proposed)
Encoder	78	54
Decoder	34	31

CPU time ([msec] per frame).

We can see that basic JPEG2000 encoding is approximately 1.4 times more complex than Modified JPEG2000, while decoding is almost of equal complexity. Modified JPEG2000 encoding is less complex because of replacing zero value to the non-person region.

In addition, examples of displaying awareness images are shown in Figs. 14 and 15.

Fig.14: *Examples of displaying awareness (blurred and semi-transparent (3LL, 2LL, 1LL, and 1/2 transparent)).*

Fig.15: *Examples of displaying awareness (contour, blurred, and semi-transparent (1HH, 1HL, 1LH, and 2LL)).*

5. CONCLUSIONS

The present paper proposes a new video communication system that transmits awareness at a very low bit rate. The proposed system decomposes the person region into (a) a basic layer, which contains awareness, and (b) an enhanced layer, which contains other detailed information. This processing is implemented by basic constituent technology of the JPEG2000, such as wavelet transform and bit plane decomposition. The basic layer contains (1) the person region, (2) the bit plane component, and (3) the bandwidth image component. It is confirmed by quantitative measurement that the bit rate can be reduced to 1/10 that of the exiting method.

In the future, we intend to adapt quantization and bit truncation, namely lossy compression, to the proposed system for further bit rate reduction.

6. ACKNOWLEDGEMENT

This research was supported in part by the Ministry of Education, Science, Sports, and Culture through a 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.13-18 April, 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 14*, 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.