A Very Low Bit Rate Video Coding for Ubiquitous Services

Ben-Shung Chow
Department of Electrical Engineering, National Sun Yat-sen University
Kaohsiung, Taiwan, ROC.
bschow@mail.ee.nsysu.edu.tw

Abstract

*The video coding with a bit rates below 10k bps has not been considered as a practical application in commercial communication systems. The main reason is that a user in the present multiple access system is assigned a fixed conceptual channel with the bit rate of speech signal, which is much more than 10k bps such as 16k. Video communication below 10k bps using inexpensive resource is indeed of low caliber of visual quality but is still readable in a small display. This visual quality can be tolerable under the consideration of the convenient receiver and the limit bandwidth for communication. Therefore, the proposed coding strategy in this paper to code the 64*64 binary images is suitable for ubiquitous services.*

1. Introduction

The video coding with a bit rates below 1k bps has not been considered as a practical application in commercial communication systems. The main reason is that a user in the present multiple access system is assigned a fixed conceptual channel with the bit rate of speech signal, which is much more than 1k bps such as 20k. Video communication of 1k bps rate under a plain (inexpensive) coding method is indeed of low caliber of visual quality but is still readable in a small display. This visual quality can be tolerable under the consideration of the convenient receiver for carrying and the inexpensive pay rate for communication. As a consequence, a new video coding strategy proposed in this paper can be applied to video coding for ubiquitous services.

Our proposal of data coding is simple but innovative. There are historical reasons for this very low bit rate video coding approach with low visual quality has not been considered before. Conventionally, speech transmission is the major application for the real time communication. Therefore, a user in the present multiple access system is usually assigned a fixed conceptual channel with the bit rate of speech signal, which is several tens of k bps, much more than 1k bps. For this reason, the video coding in the category of very low bit rate is set up to be between 8k to 64k bps such as the ITU –T.H.X standards. The extra low bit rate below 4k bps is usually achieved by advanced coding such as model based methods [1] and is with fairly good visual qualities. These coding methods cost intensive computation or require special instruments for tracking label points in the encoding stage and thus limit their applications in mobile communication.

The fairly good visual quality is generally accepted as a requirement for all standards and thus prevents the plain (inexpensive) coding methods of bit rates below 1k bps. Therefore, the applications of video coding have never considered bi-level image sequences, which are regarded as low quality for dynamic presentation and are only coded for the still images. The progress pace of the above high quality standard video coding is understandable by reviewing the technique developments of these decades. Obviously, the compression of dynamic images is more complicated than that of the still images and is thus developed later than still image compression. The high computational expense for video coding at that time had confined the applications to some products with privileged qualification so that the low quality images such as binary images had not been included. Nowadays, this quality requirement should be reconsidered and is probably not necessary for many unseen applications because of the growing popularity of wireless communication advocated by the growing popular wireless communication.

The dynamic presentation of binary images obviously does not fulfill the above standard. Thus, binary mode is limited to the still presentation of text images and is not extended to dynamic presentation of natural images. However, recently there are some reasons for the emergence of the low quality dynamic presentation: (1) very cheap display of the binary mode liquid crystal material, (2) the trends of down cost of compression computation versus the up needs of low bit-rate communication. For example, the dynamic compression for binary mode is demanded in the applications of deaf communication due to the advantage of the low bit-rate in communication and simple computation structure for both the encoding and decoding procedures. In the cases of monitoring, a one way communication requires only simply-structured decoder. A very inexpensive receiver with a mall binary liquid display can becomes portable, acting as convenient as a wristwatch. As a consequence, a new policy of channel assignment as little as 1k bps may be triggered by this strong demanding of low bit rate in mobile communication for some new applications such as babysitting or factory monitoring.

2. Resolution vs. Quantization

The resolution issue of quality vs. compression has been widely adopted in video coding standards. For example, the ITU-T H-series of video codices have evolved in a variety of applications based upon the resolutions. The H.261 codec, intended for ISDN teleconferencing, supports the CIF (352*288 pixels) and QCIF (176*144 pixels). The H.262 codec is for broadband videoconferencing and thus essentially for the 4CIF (704*576 pixels). The H.263, mainly for the POTS (plain old telephone service) teleconferencing at modem rate, from 14.4 to 56 k bps, supports QCIF. In addition, H.263 supports a wider range of picture formats, including 4CIF and 16CIF (1408*1152 pixels). A more detailed description can be seen in [2].

Obviously, the visual quality increases with the resolution. This simple relation holds for the relation with the number of quantization levels too. However, the joint relationship between the resolution and the quantization under a fixed quality has not much been considered. This is probably because the simple formula of bit rates, bits = Resolution2*log (the number of quantization levels), has excluded the interests of the compression by reducing the quantization levels.

An experimental study by Huang [3] has indicated that some images do not improve the image quality by increasing quantization levels. Furthermore, Huang’s isopreference curves demonstrated that the image quality decreases rapidly when the resolution is below 64*64, whereas showing a little relation with the number of quantization levels. Therefore, it can be concluded that the resolution should be considered first for compression till the resolution is reduced down to around 64*64. After that the reduction of the number of quantization levels is more efficient for maintaining a minimum of visual quality. In addition to this quality’s consideration, the edge and the motion information, which are essential to the visual perception, are contained in the binarized images. Therefore, the extension of video coding to the bi-level images is believed to be legitimate.

Our coding procedures are simple: reducing resolution first and then the quantization. That is, the images are first reduced to 64*64 and then threshold to be bi-level images so as to be presented in a very inexpensive and small liquid crystal display. Afterwards, the motion compensation and shape compensation was followed to achieve the further compression. Shape compensation is a new idea proposed in this paper.

3. Review of Mathematical Morphology

Mathematical morphology refers to a branch of nonlinear image processing and analysis developed initially by Serra [4] that concentrates on the geometric structure within an image. The morphological approach is generally based upon the analysis of a two value image with some predetermined geometric shape known as a structuring element. Information about size, spatial distribution, shape, connectivity, convexity, smoothness can be obtained by transforming the image object using different structuring elements.

The simplest morphological transformation is dilation and erosion. In loose terms, these operations cause the swelling or the shrinking of areas when structuring element has a disklike shape. The dilation of A by B, denoted by $A \oplus B$, is defined as

$$A \oplus B = \{y \mid y = a + b,\ a \in A,\ b \in B\} = \bigcup_{b \in B} A_b$$

where $A_b = \{a + b \ : \ a \in B\}$, is the translation of A by the displacement b. Thus the dilation of A by B is equal to the set union of all the translates A_b. Another formulation of dilation involves translates of the rotated structuring element that "hit"(intersect) the input image.

$$A \oplus B = \{x \mid (\check{B})_x \cap A \neq \emptyset\}$$

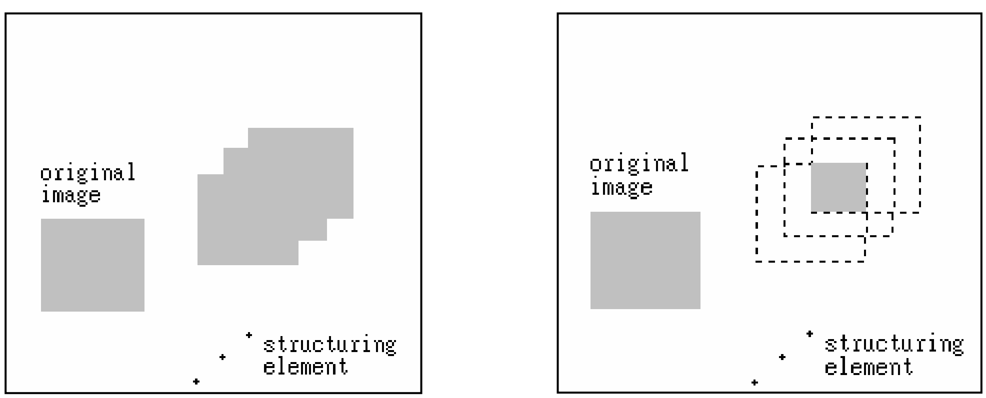
where $\check{B}$ is obtained by rotating B by 180o on the plane so that $\check{B} = \{-b : b \in B\}$. The erosion of A by B, denoted by $A \Theta B$, is defined as

$$A \Theta B = \{x \mid x + b \in A : b \in B\} = \bigcap_{b \in B} A_{-b}$$

Another definition of erosion is the following:

$$A \Theta B = \{x \mid (B)_x \subseteq A\}$$

Which, in words, says that the erosion of A by B is the set of all point x such that B, translate by x, is contained in A. Figure 1 illustrates the dilation and erosion.



(a) (b)

Figure 1. The basic morphological operations (a) dilation (b) erosion

There are certain mathematical properties that the dilation and erosion operations follow. we summarize some of the important properties here and details and proof may be found in Haralick [5].

1. Translation invariance

$$(A)_x \oplus B = (A \oplus B)_x$$

$$(A)_x \Theta B = (A \Theta B)_x$$

2. Distributivity

$$(A \cup B) \oplus C = (A \oplus C) \cup (B \oplus C)$$

$$(A \cap B) \Theta C = (A \Theta C) \cap (B \Theta C)$$

3. Iteration

$$A \oplus (B \oplus C) = (A \oplus B) \oplus C$$

$$A \Theta (B \oplus C) = (A \Theta B) \Theta C$$

4. Duality

$$A \Theta B = (A^c \oplus \check{B})^c$$

$$A \oplus B = (A^c \Theta \check{B})^c$$

In this paper, the morphological filter is implemented by dilation only. Applying the duality property of dilation and erosion, we can easily employ both logical and geometric negation to implement erosion.

4. Shape Compensation

Standard dynamic image compression is usually composed of motion compensation and a DCT transformation for the error image after the motion compensation. The DCT coefficients are also classified by the cases coded by Huffman compression. However, the DCT coding, due to its broad dynamic range, will be inefficient for the binary images. It is noted that the binary images are usually described properly by their shapes. In this sense, a novel idea of shape compensation is proposed to replace the DCT processing. More clearly, the binary dynamic images are compressed by the shape compensation after the motion compensation. Our binary images are coded by the motion vectors and the kinds of shape transformations.

In the shape transformation, a morphological filter is selected to modify the shape of the objects in image. Morphological processing [4] is a type of operation, by which the spatial form or structure of objects within an image are modified in a nonlinear way. The computation cost of nonlinear processing for conventional numerical image processing is very expensive. However, the morphology image processing treats the image components as sets and deals with the changes of shapes very efficiently. Thus, the morphology processing has recently been applied successfully to the industry auto-inspection and medical image processing, but has not been seen to be applied to coding except the preprocessing for simplifying images [6]. The efficiency of morphological image processing makes the shape compensation in decoding procedure very simple.

5. Experimental Results

The image frame is divided into blocks of the size of 16*16. Every block, which is referred afterwards as the target block, in the original current frame has a corresponding block with the same location in the motion compensated previous frame. The optimal filter is found to modify the corresponding block in the previous frame (reference frame) to look like the target block in the original current frame. The optimization is based on minimizing the error count, which is the number of different pixels between the output block and the target block. It is noted that the error count is produced in a single scan of the block pairs in the reference and target images.

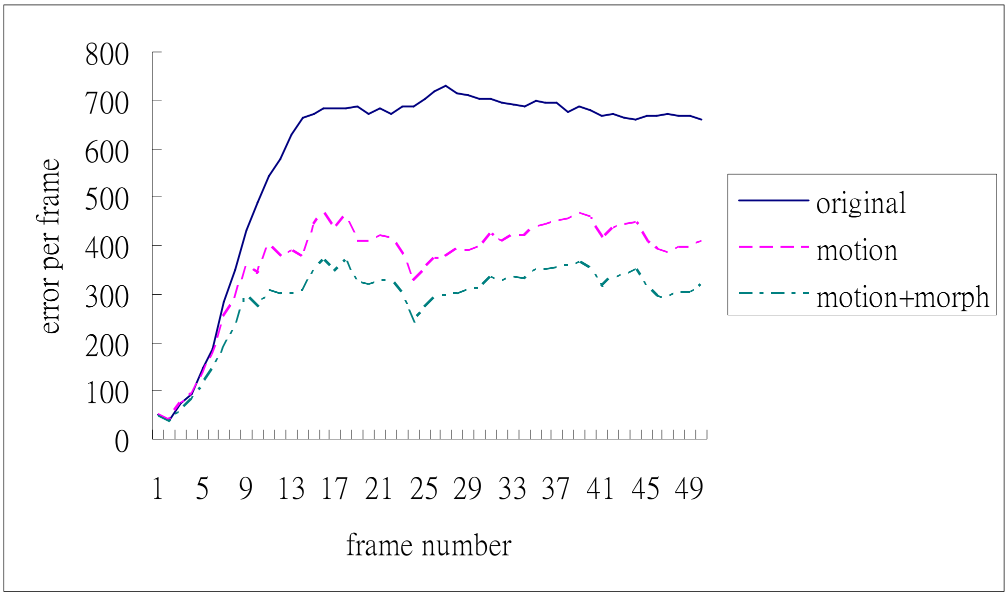
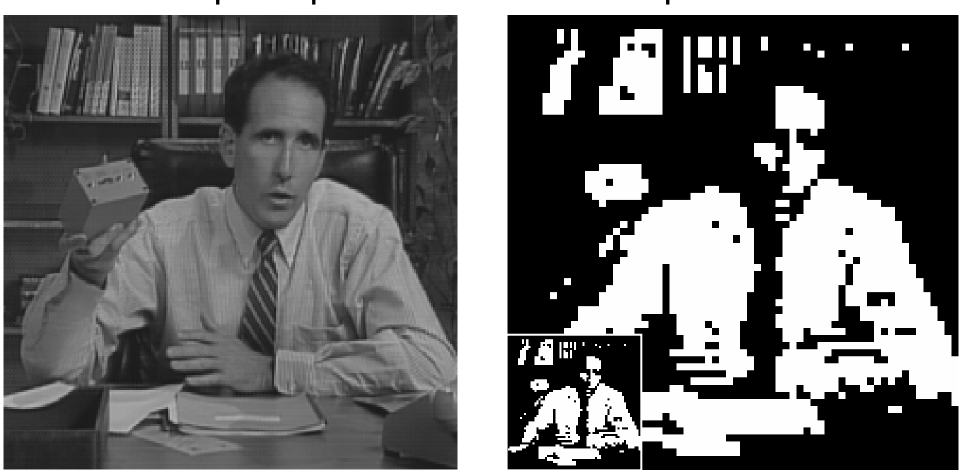


Figure 2. Error comparison to illustrate the error improvements by using the proposed shape compensation after motion compensation.



(a) (b)

Figure 3. (a) Frame 1in the form of 256*256 with gray levels of 256 for visual quality's demonstration
(b)Frame 1 in the form of 64*64 with gray levels of 2 also magnified to see the detail

The video streams of the salesman are tested in our experiment. The original streams are of the CIF format with the gray levels of 256. They are first to be reduced to the size of 64*64 and then threshold to be bi-level 64*64 sequences. Binary image processing is then applied to the bi-level sequences in the encoding stage, which includes the motion vectors search and the optimal morphological filters determination. The block would be refreshed if the motion compensated error were above the threshold. Correspondingly, there are two possibilities in the decoding process: block refreshing or shape compensation after motion compensation.

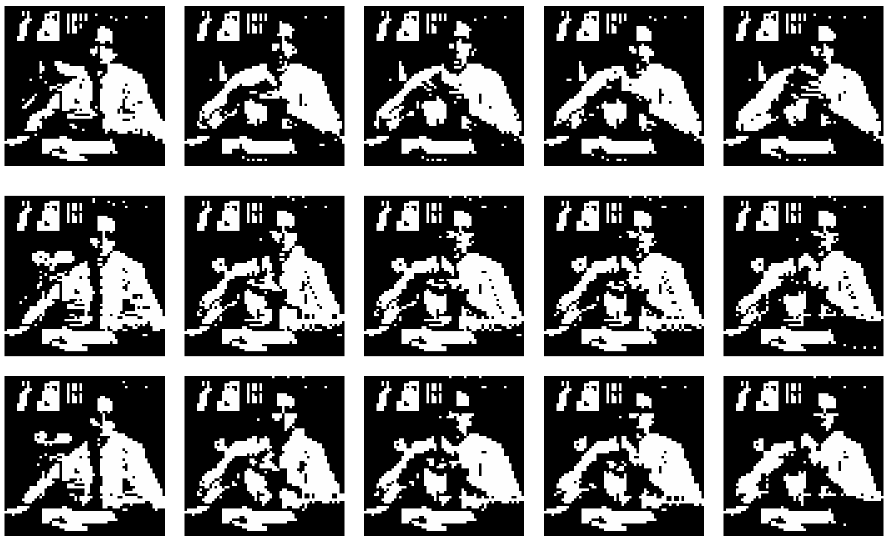


Figure 4. Sample frame 10, 30, 40, and 50 of the original and decoded streams. Upper: the original. Middle: the decoded by motion compensation, lower: the decoded by motion and shape compensation.

For the purpose of comparison, three coding schemes are arranged. The first is a naïve coding by representing all frames using the first one. The second is coding by motion compensation only. The third is coding with motion compensation followed by shape compensation. These three coding methods are compared by the errors and the block refreshing occurrences. The frame errors of the three coding methods for the entire flows of salesman are plotted in Figure 1. For visual quality's demonstration, the first frame of the streams is shown with the forms of grey 256*256 in Figure 2(a) and bi-leveled 64*64 in Figure 2(b). The magnified images for 64*64 are also show in the figure 2(b) for the details. Finally, the image samples of the original streams and the decoded streams from coding II and III are illustrated in the figure 3. The morphological filter selection for frame 40 is listed in Table 1 as one example.

The shape compensation can improve about 20% of errors compared to the coding with the motion compensation. The number of refreshing blocks for salesman is 10 from the total 800 (16*50) blocks for the stream. According to Fig. 1, the average error rate of coding III is 6%. The refreshing rates can go down to 1% if error rate release to 7%.

Table 1. Filter selection results for table 1.

Block #	Structuring elements		
Block 1	no operation	Block 9	00100 00010 erosion
Block 2	no operation	Block 10	10001 01010 dilation
Block 3	00010 00001 dilation	Block 11	10110 01011 erosion
Block 4	no operation	Block 12	10000 00100 erosion
Block 5	01000 00010 dilation	Block 13	10010 01100 dilation
Block 6	01001 00110 erosion	Block 14	10001 00100 dilation
Block 7	10100 0100 1 dilation	Block 15	10110 01110 erosion
Block 8	no operation	Block 16	10110 01101 erosion

6. Discussion and Conclusions

The testing images are of the size 64*64. The block sizes of the motion compensation and shape compensation are both 16*16. The range of motion vectors is of 16*16, at the expense of 0.128 k bits per frame for coding. The number of kinds of shape transformation is about 256 for two erosions or dilations with the mask of 4-neighbours, at the expense of around 0.128 k bits per frame for coding the shape transform. In our experiments, the average block (or frame) refreshing ratio is 0.01 for the streams of salesman. Therefore, the total bit rate per second for this binary video coding is 0.256k*frame-rate*Huffman compression ratio + 4k*0.05*frame-rate*still compression ratio where 0.256k is the bit rate for dynamic coding by motion compensation and shape compensation, 4k is the number of bits per frame, and 0.01 is the refreshing ratio. A reasonable frame rate for low bit rate video coding is 4 to 8 frames per second. It is noteworthy that there is a significant correlation between neighboring blocks for both shape and motion compensation operators. Also, the still image coding for the refreshing image can be coded by quadtree with an assumption of data compression rate of 4. Thus, a conservative estimate of the Huffman compression ratio and the still compression ratio is 0.25. As a summary, a total of 1k bits per second can be achieved in our coding system for the image size 64*64. It should be noted that the visual quality appears better in video form than that demonstrated in the still form.

Our coding procedures are simple and plain: reducing resolution first and then the quantization. The coding strategy is natural: motion compensation first and then the shape compensation. This visual quality can be tolerable due to the demands of convenient receivers for carrying and the inexpensive pay rate for communication. Thus, this coding strategy is suitable for video coding for ubiquitous services

7. References

[1] J. Lee and A. Eleftheriadis "Spatio-temporal model-assisted very low-bit-rate coding with compatibility," IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 12, pp. 1517–1532, May 2005.
[2] B. G. Haskell, P. G. Howard, Y. A. LeCun, A. Pury, J. Ostermann, M. R. Civanlar, L. Rabiner, L. Bottu, and P. Haffner, "Image and video coding - emerging standards and beyond," IEEE Trans. Circuits. Syst. Video Technol., vol. 8, no. 7, pp. 814-837. Nov. 1998.
[3] T. S. Huang, "PCM picture transmission," IEEE Spectrum, vol. 2, no 12, pp. 57-63. 1965.
[4] J. Serra, Image Analysis and Mathematical Morphology, New York: Academic Press,1983
[5] R. M. Haralick, S.R. Sternberg, and X. Zhung, "Image analysis using mathematical morphology," IEEE Trans. Pattern Anal. Machine Intell., vol. 9, pp. 532–550, Jul. 1987.
[6] D. Wang, C. Claude Labit, and J. Ronsin "Segmentation-based motion-compensated video coding using morphological filters," IEEE Trans. Circuits Syst. Video Technol., vol. 7, No. 3, pp. 549-555. 1997.

Authors

Ben-shung Chow received the B.S. degree in Electrical Engineering from National Taiwan University, the M.S. degree in electrical engineering from University of Florida, Gainesville, FL, U.S.A. and the Ph.D. degree from University of Wisconsin, Madison, WI, U.S.A. in 1978, 1982, and 1988, respectively. He is presently Associate Professor in the Department of Electrical Engineering National Sun Yat-Sen University, Taiwan. His research interests include the fields of image processing, video communication and communication theory. Dr. Chow is a member of IEEE, SPIE, and OSA.