CONTENT-BASED RATE CONTROL SCHEME FOR VERY LOW BIT-RATE VIDEO CODING

Chun-Hung Lin[1] and Ja-Ling Wu[1,2]
[1]Communication and Multimedia Lab.
Department of Computer Science and Information Engineering
National Taiwan University, Taipei, 106, Taiwan, R.O.C.
[2]Department of Information Engineering
National Chi Nan University
Puli, 545, Taiwan, R.O.C.

ABSTRACT —

In the applications of very low bit-rate video coding, the quality of the encoded pictures always suffers from serious degradation due to limited bandwidth. The quantization levels, in typical video coding systems, are adjusted depending only on the fullness of the output buffer while the content of the video is not taken into account. This explains why the picture quality varies saliently in very low bit-rate applications.

In this paper, a content-based bit-rate control scheme is proposed. In this approach, the content of the coded picture is first analyzed to find the main attributes of each macroblock. The attributes of the macroblocks are then consulted while adjusting the quantization levels. In other words, the quantization levels are adjusted not only depending on the fullness of the output buffer but also on the content of the coded material. Simulation results show that the perceptual quality of the coded pictures, by using the proposed method, is better while comparing with the traditional non-content-based approach.

1 INTRODUCTION

Many international video coding standards have been applied to different kinds of video coding applications, such as ITU-I H.261 for audiovisual services at $p \times 64$ kbits [1], ISO MPEG-1 for digital storage media at 1.5 Mbit/s [2], and ISO MPEG-2 for a wide range of applications at 2–15 Mbit/s [3]. The most recent video coding standardization focuses especially on the very low bit-rate coding techniques, such as H.263 [4] and MPEG-4 [5].

The basic structure of an H.263 coder still comes from the traditional MC/DCT hybrid coding approaches. Although it has been shown that an H.263 coder can meet the very low bit-rate constraint, but it is widely aware of that the picture quality and resolution of the coded video degraded. However, since the other subtle methods have not been set and the complexity of the H.263 codec is not high, it is still the best choice nowadays for very low bit-rate video coding applications.

When the hybrid coding method is applied, both the temporal frequency and the quantization levels can be adjusted depending on the still available bit-rates. The movement of objects in the video sequence will not be smooth if the temporal frequency is not static. Similarly, the picture quality will not be stable if the quantization level varies frequently. In a typical video coding system, the quantization levels are adjusted according to the fullness of the output buffer. Because the adjustment does not take the content of the coded pictures into account, the picture quality will not be stable especially when the given bit-rate is insufficient. In the proposed method, the temporal frequency is fixed because the audiences are more sensible to the unsteady movement of objects, and the unstable picture quality will be improved by a content-based bit-rate control scheme.

The coding material in very low bit-rate applications is often composed of the head-and-shoulder images. It can be found that the picture quality of the critical areas, such as eyes, mouth, and the face area, is more important and noticeable. One interesting and unique fact for very low bit-rate video applications is that: the lightly moving regions are not so sensible as their

Contributed Paper
Manuscript received January 31, 1997 0098 3063/97 $10.00 © 1997 IEEE

heavy moving counterparts. This is because the whole temporal frame rate is relative low as compared with the medium- or high- rate video applications. In other words, the heavy moving regions are more critical than the lightly moving regions in the very low bit-rate video applications.

In the proposed content-based bit-rate control scheme, the content of the coded material is taken into consideration while adjusting the quantization levels. The picture quality of the lightly moving regions are sacrificed to enhance that of the heavy moving regions. Because most people will concentrate more on the critical areas and the corresponding perceptual picture quality should be better.

This paper is organized as follows. In Section 2, an overview about the system structure of the proposed bit-rate control scheme is depicted. In Section 3, the method of separating the active regions from the background region and the method of extracting the face region are elaborated. A genetic search method is introduced in Section 4 to find the feature points among the face region. The experimental results are shown in Section 5 and the conclusion is given in Section 6.

2 OVERVIEW OF THE SYSTEM STRUCTURE

Fig. 1(a) shows a traditional video coding system [6]. In order to meet the constant bit-rate constraint, the quantization levels in the coding system are adjusted according to the fullness of the output buffer. That is, the quantization level of the current encoded unit (a macroblock in most of the cases) depends on how many bit-rates are consumed by the previous encoded units. In medium or high bit-rate video coding systems, this kind of bit-rate control scheme works successfully. The variation of video quality caused by the changing of quantization levels is not so prominent because the quality of the whole video frames is much higher than the perceptible level. Whereas, some of the video frames will suffer from serious quality degradation due to the unsteady changing of quantization levels in very low bit-rate video applications. In order to improve the whole quality of the sequence, different kinds of bit-rate control schemes have to be investigated. Since different regions of the video frames will receive different levels of concentration, the bit-rates can therefore be assigned depending not only on the fullness of the buffer but also on the importance of the coded regions. This kind of control scheme is called the content-based bit-rate control scheme (CBCS). The block diagram of the CBCS is depicted in Fig. 1(b). A video frame is first analyzed by a content analyzer to extract active regions (e.g., features, head and shoulder regions). Each region of the frame is then encoded based on the importance of that region and on the fullness of the output buffer.

Facial features, including eyes and mouth, have the highest priority to be assigned more bits, so the quantization levels in these regions should be finest among all. The face region has the second finer quantization level, and the other active regions (such as shoulder and head silhouette) have the coarsest quantization level. The static background regions can be directly skipped for saving both bit-rate and encoding time.

The content analyzer is composed of four components, as shown in Fig. 2. The function of each component is summarized as follows.

1) **Extraction of Active Regions:** The active regions are separated from the background regions in the first component. The distinctions of different regions are decided based on the error image generated from the current encoding picture and its preceding one.

2) **Thresholding of Active Regions:** The intensity values of the active regions are converted to binary values. The active regions are then divided into two kinds of subregions. The subregion with higher intensity value will be expected to contain the face region because lighting always focuses on that region.

3) **Extraction of Face Region:** A starting point is first determined and the region growing scheme [7] is performed. A candidate face region is then grown

from the starting point. A rectangle region is selected from the candidate region as the face region.

4) **Genetic Search of Facial Features:** The critical facial feature points are searched among the face region based on an evaluation function. In order to reduce the computational complexity, a fast search scheme developed from the genetic algorithms (GAs) [8] is applied.

3 EXTRACTION OF ACTIVE AND FACE REGIONS

The active regions that contain the moving objects, i.e., the protagonist in the video scene, are extracted from the coding frame. The other regions are then considered as the background regions. In typical video coding systems, each macroblock has to be motion compensated, encoded, and decoded to determine its coding type. In the proposed system, the attribute of the coded macroblock has been known at the initial stage, so the redundant computations can be saved. Because most of the scenes, in a head-and-shoulder image, are covered by background regions, the computations saved in this part will compensate most of the computations spent on the content analyzer.

In order to determine the active regions, an error image E is generated as,

$$E_{x,y} = \begin{cases} 1, & |I_{x,y} - I'_{x,y}| > T, \\ 0, & else, \end{cases} \tag{1}$$

where (x, y) represents the coordinate of the processed pixel, I and I' respectively represent the current and the previous pictures, and T is a predetermined intensity threshold ($T = 3$ in our implementation). Each macroblock B has a difference-value $D(B)$ which is defined as,

$$D(B) = \sum_{(x,y)\in B} E_{x,y}. \tag{2}$$

The macroblocks with $D(B)$ larger than a predetermined threshold H (which is set to be 5 in our implementation) are selected as the active macroblocks. A macroblock is also assigned to be an active macroblock if both its horizontal neighboring macroblocks or vertical neighboring macroblocks are active macroblocks. Therefore, the active macroblocks can be defined as

$$A(B_{i,j}) = \begin{cases} 1, & D(B_{i,j}) > H \text{ or} \\ & A(B_{i-1,j}) \cdot A(B_{i+1,j}) + A(B_{i,j-1}) \cdot A(B_{i,j+1}) > 0, \\ 0, & else. \end{cases} \tag{3}$$

The active regions are further divided into two classes as,

$$C_{x,y} = \begin{cases} 1, & I_{x,y} > K \text{ and } A(B) = 1 \text{ for } (x,y) \in B, \\ 0, & else. \end{cases} \tag{4}$$

The threshold value K is calculated from the intensity of the overall active macroblocks to minimize the *within group variance* σ_w^2 [9], that is

$$K = \min_t\{\sigma_w^2(t)\} = \min_t\{q_1(t)\sigma_1^2(t) + q_2(t)\sigma_2^2(t)\}, \tag{5}$$

and

$$q_1(t) = \sum_{i=0}^{t} H(t), \tag{6}$$

$$q_2(t) = \sum_{i=t+1}^{255} H(t), \tag{7}$$

$$\mu_1(t) = \sum_{i=0}^{t} iH(i)/q_1(t), \tag{8}$$

$$\mu_2(t) = \sum_{i=t+1}^{255} iH(i)/q_2(t), \tag{9}$$

$$\sigma_1^2(t) = \sum_{i=0}^{t}[i - \mu_1(t)]^2 H(i)/q_1(t), \tag{10}$$

$$\sigma_2^2(t) = \sum_{i=t+1}^{255}[i - \mu_2(t)]^2 H(i)/q_2(t), \tag{11}$$

where H is the histogram statistics of the active regions, that is

$$H(i) = \#\{(x,y)|A(B_{\lfloor x/16\rfloor,\lfloor y/16\rfloor}) = 1 \text{ and } I_{x,y} = i\}. \tag{12}$$

The thresholded image C is first processed by the closing operation and then the opening operation [9]. A starting pixel has to be found for growing the face region. The macroblock with the maximum error energy is selected from the active macroblocks based on the energy function defined as

$$E(B) = [\sum_{(x,y)\in B} C_{x,y}] \cdot \frac{D(B)}{\sum_{A(B')=1} D(B')}. \quad (13)$$

Because the intensity and the density of edges in the face area are usually higher than the other areas among the video scene, the selected macroblock will locate on the face region. The starting pixel is set to be the central point of the selected macroblock. By applying the region growing techniques [7], a region can be generated from the starting pixel. A rectangle is then constructed to surround the grown area. In order to exclude the neck region from the grown area, the lower boundary of the rectangle is adjusted so that the height of the rectangle is equal to its width. The so-obtained square area is defined as the face region.

An example of face region locating is illustrated in Fig. 6. The original picture selected from the "Miss America" sequence is shown in Fig. 6(a). The error image generated from the original picture and its preceding one is shown in Fig. 6(b). Fig. 6(c) shows the active macroblocks in the error image. It can be seen that the whole body of the lady is totally covered by the active macroblocks. The thresholded image is shown in Fig. 6(d). Because the image is processed by closing and opening, the face area is regular. The face region can therefore be found from the thresholded image. In Fig. 6(e), it is shown that the face region is surrounded by a rectangle.

4 GENETIC SEARCH OF FACIAL FEATURES

Basically, the face region can be divided into three subregions. As shown in Fig. 3, the subregion R_a is expected to contain the right eye feature point, while the left eye feature point is expected to appear in R_b ,and the mouth feature point is in R_c.

To extract the critical feature points from the face region, a generic feature cost function is defined as,

$$F(p_0) = C_m(p_0) + U_{I_{p_1}-I_{p_0}-\kappa} \cdot U_{I_{p_2}-I_{p_0}-\kappa} \cdot C_c(p_0), \quad (14)$$

where p_0 is the evaluated candidate feature point, C_m is a mean crossing function (which is used to accumulate the intensity variation within the masked area), C_c is a central weighting function (which is used to compute the intensity difference between the central part and the boundary parts of the masked area), κ is a constant, I_p means the intensity of the point p, and U is a unit step function.

The mean crossing function is applied to detect the intensity variation within the square area around the evaluating point. Let d_t represents the length of the square area, $(d_t + 1)$ 1-D signals can be taken from the square area row-by-row, and another (d_t+1) 1-D signals column-by-column. The corresponding mean crossing values for these 1-D signals are calculated. The total mean crossing value C_m is defined to be the sum of these mean crossing values, that is

$$C_m(x,y) = \sum_{i=-d_t/2}^{d_t/2} [V_r(x, y+i) + V_c(x+i, y)], \quad (15)$$

where (x, y) is the coordinate of the evaluated point, and V_r and V_c are the mean crossing values of the row signal and the column signal which are respectively defined as:

$$V_r(x,y) = \sum_{i=-d_t/2+1}^{d_t/2} U_{-S_r(x+i,y)\cdot S_r(x+i-1,y)-1}, \quad (16)$$

$$V_c(x,y) = \sum_{i=-d_t/2+1}^{d_t/2} U_{-S_c(x,y+i)\cdot S_c(x,y+i-1)-1}, \quad (17)$$

where

$$S_r(x+i,y) = \begin{cases} 0, & i = -\frac{d_t}{2}, \\ 1, & I(x+i,y) - \mu_m > K_m, \text{ and } i > -\frac{d_t}{2}, \\ -1, & I(x+i,y) - \mu_m < -K_m, \text{ and } i > -\frac{d_t}{2}, \\ S_r(x+i-1,y), & |I(x+i,y) - \mu_m| \le K_m, \text{ and } i > -\frac{d_t}{2}, \end{cases} \quad (18)$$

and

$$S_c(x,y+i) = \begin{cases} 0, & i = -\frac{d_t}{2}, \\ 1, & I(x,y+i) - \mu_m > K_m, \text{ and } i > -\frac{d_t}{2}, \\ -1, & I(x,y+i) - \mu_m < -K_m, \text{ and } i > -\frac{d_t}{2}, \\ S_c(x,y+i-1), & |I(x,y+i) - \mu_m| \le K_m, \text{ and } i > -\frac{d_t}{2}. \end{cases} \quad (19)$$

The mean value μ_m is computed for each evaluated 1-D signal, and a bar area $(\mu_m - K_m, \mu_m + K_m)$ is formed, where K_m is a constant. The number of μ_m-crossings of the 1-D signal within the bar area is calculated. If the intensity variation of the signal is large, the mean crossing value will also be large. Because the crossing value is computed on the bar area, slight intensity variation caused by noise will be eliminated. While comparing with the other non-feature areas, the intensity variations in the feature areas are relatively large, and so are the corresponding mean crossing values. For the sake of saving computations, we only extract ten lines (five for each direction) near the center of the square area to calculate the crossing values.

The central weighting function is included in the cost function because of the fact that the intensity difference between the central part and the boundary parts of the feature area is large. As shown in Fig. 4, three mean values μ_c, μ_u, and μ_l are calculated for the three different subregions (B_0, B_1, and B_2) within the square area. The central weighting function is defined as,

$$C_c(p_0) = [K_c \cdot (\frac{\mu_u + \mu_l}{2}) + \mu_l - \mu_u]/\mu_c, \quad (20)$$

where

$$\mu_c = \frac{1}{|B_0|} \sum_{p \in B_0} I(p), \quad (21)$$

$$\mu_u = \frac{1}{|B_1|} \sum_{p \in B_1} I(p), \quad (22)$$

$$\mu_l = \frac{1}{|B_2|} \sum_{p \in B_2} I(p), \quad (23)$$

and K_c is a constant.

It is obvious that the intensity difference between main feature points and those points below and near the main feature areas are large. Therefore, unit step functions are included in (14) to prevent from selecting the other features, such as eyebrows and nose.

Each possible point in the face region is evaluated based on the cost function given in (14). The points with maximum matching values in different subregions are selected as the feature points. Fig. 5 shows an example of the extracted feature points. The computational complexity will be very high if each point of the face region is evaluated to find the best matches. Therefore, a genetic search algorithm was proposed to relieve the computational burdens [10].

Let S be a solution space and all the elements in S have their associated fitness values. A straightforward way to find the element with the maximum fitness value is to search among all the elements and to compare their fitness values. However, the computational complexity will be very high if the space size is large. In order to reduce the computational complexity, an efficient search algorithm should be applied.

If GAs are applied to search for the global maximum in S, a population P is maintained which consists of N elements, where N is the population size. Each element in P is called a chromosome which is composed of a list of genes. The population P will evolve into another population P' by performing some genetic operations. The chromosomes with higher fitness values will have more probability to be kept in the population of the next generation and to propagate their offspring. On the other hand, the weak chromosomes whose fitness values are small will be replaced by another stronger chromosomes. Therefore, the quality of the chromosomes in the population will be better and better. After a suitable number of generations, the mature population will be expected to contain the element with the global maximum value.

5 RATE CONTROL SCHEME

To meet the designate bit-rate constraint in the coding system, a suitable bit-rate control scheme is a must. In our system, the adopted bit-rate control scheme is modified from the one given in [13] which, basically, is a frame-based approach and is called frame-based bit-rate control scheme (FBCS) in this paper. The quantizer parameter Q_p of each macroblock is identical in the same picture. The value of Q_p will be modified while encoding a new picture as

$$Q_p = \begin{cases} \min\{31, Q'_p + \bar{Q}\}, & b_L > b_R \cdot 1.15, \\ \max\{1, Q'_p - \bar{Q}\}, & b_L < b_R/1.15, \\ Q'_p, & else, \end{cases} \quad (24)$$

and

$$\bar{Q} = \max\{1, 0.1 \cdot Q'_p\}, \quad (25)$$

where Q'_p is the quantizer parameter of the previous picture, b_L is the number of bits used for the previous picture, and b_R is the current average number of bits for each picture, that is

$$b_R = b - (f-1) \cdot b_L, \quad (26)$$

where b represents the target bit-rate and f is the frame rate.

Different quantizer parameters are set for different macroblock types. The macroblocks which belong to the feature areas require the finest quantization steps. Their quantizer parameters are set to be $\max\{1, Q_p - d_1\}$. The quantizer parameters of the macroblocks in the face region are set to be $\max\{1, Q_p - d_2\}$, and that of the other active macroblocks are Q_p. Because the maximum allowable differential value of the quantizer parameters between two neighboring coded macroblocks is 2 in H.263 coders, d_1 and d_2 are respectively selected as 4 and 2 in our implementation.

6 EXPERIMENTAL RESULTS

The proposed CBCS was tested on two QCIF (176×144 pixels) video sequences, "Miss America" and "Claire", with fixed temporal frequency 12.5 frames per second (fps). Three different target bit-rates are tested in the simulations: 28.8, 14.4, and 8 kbps. The frame-based bit-rate control scheme (FBCS) presented in [13] was also implemented for comparison.

Table 1 shows the average PSNR values of the coded video sequences by using different bit-rate control schemes for various target bit-rates. Because the bit-rates are diminished in the lightly moving regions, it follows from Table 1 that the average PSNR values using the CBCS are a little less than that of the FBCS. Nevertheless, the PSNR values can not evaluate the picture quality exactly in this case because each pixel was assigned with different number of bits but has the same weight in the evaluation.

In order to evaluate the picture quality more correctly, another evaluation function considering the importance of different image regions has to be used. A weighted SNR (WSNR) evaluation function, which is similar to the one used for evaluating the quality of color images [14], is defined in this work as

$$WSNR = 10 \log \frac{255^2}{\text{WMSE}}, \quad (27)$$

and

$$\begin{aligned} WMSE = & \frac{1}{N} \cdot [(\frac{\tau_1}{\tau_1+\tau_2+\tau_3}) \cdot \sum_{p \in S_1} (I_p - I'_p)^2 + \\ & (\frac{\tau_2}{\tau_1+\tau_2+\tau_3}) \cdot \sum_{p \in S_2} (I_p - I'_p)^2 + \\ & (\frac{\tau_3}{\tau_1+\tau_2+\tau_3}) \cdot \sum_{p \in S_3} (I_p - I'_p)^2], \end{aligned} \quad (28)$$

where N is the number of pixels belonging to the active regions, S_1 represents the feature regions, S_2 represents the face regions, S_3 represents the other active regions, I_p and I'_p respectively represent the original intensity and the reconstructed intensity of the point p, and τ_1, τ_2, and τ_3 are the average bit counts per pixel of the regions S_1, S_2, and S_3. Since the WMSE, defined in (28), has taken the allocated bit counts of each active region into account for picture quality evaluation, a better objective evaluation result (which is expected to be closer to the subjective result) can be obtained. The average WSNR values of different schemes are shown in Table 2. It follows from Table 2 that the average WSNR values of the CBCS are better than that of the FBCS when the target bit-rate is greater than 8 kbps.

The performance of the CBCS can also be testified by the subjective observation of the encoded pictures. Fig. 7 shows the subjective quality of the decoded pictures by using different bit-rate control schemes. The figure is a magnified version of one frame selected from the simulation results It is clear that, CBCS gives better perceptual/subjective picture quality than FBCS does (especially on the feature regions) although its WSNR value is worse than that of the FBMA at 8 kbps. Fig. 7 shows that the coded picture of the CBCS is free from serious blocking effects and the feature areas (e.g., eyes and mouth) are more clear. Although there is a quality degradation on the edge of hair, it is always negligible.

The mean square error (MSE) ratios [1] of different types of regions, in the "Miss America" sequence, are shown in Fig. 8. The MSE ratio of the critical area (especially the feature area) decreases while the CBCS is applied. And this implies that the picture quality around the face region will be well improved. When the target bit-rate is high, the above improvement will be more prominent as shown in Fig. 8(c). However, the improvement will not be so obvious when the target bit-rate is less than 8 kpbs. This is because the flexibility of quantization parameters adjusting will be reduced due to tight limitation of the total bit-rate.

Table 3 shows the average bit count ratios [2] of different types of regions. It is shown in the table that the bit count ratio of the critical area increases while the CBCS is applied. This result shows once again that the picture quality of the critical regions will be enhanced when the proposed bit-rate control scheme is adopted.

7 CONCLUSION

A content-based bit-rate control scheme is proposed, in this paper, to improve the perceptual quality of the picture in very low bit-rate video applications. In the proposed scheme, each coded picture is analyzed first and the critical regions (including feature regions, face regions, and the other active regions) are extracted. The macroblocks of each frame are then classified into different types according to which regions they belong to. The quantization level of each macroblock is adjusted depending both on the fullness of the output buffer and the types of that macroblock. Because finer quantizers are used in the critical regions, the overall perceptual quality of the proposed control scheme will be better than that of the traditional non-content-based bit-rate control schemes.

[1] The mean square error ratio is defined as the individual mean square error normalized by the mean square error of the whole picture.

[2] The average bit count ratio is defined as the individual average bit count normalized by the average bit count of the whole picture.

The proposed approach has been implemented and embedded in an H.263 coder. The simulation results show that the proposed method promotes the perceptual picture quality of the coded pictures in the very low bit-rate applications. The computational complexity of the proposed scheme is not high. The execution time of the H.263 coder using the proposed scheme is only about 6% higher than that of the original one. Therefore, it is our belief that the proposed scheme is feasible and very suitable for very low bit-rate video coding applications.

References

[1] ITU-T SGXV Recommendation H.261, *Video codec for audio visual services at p×64 Kbits/s.* July 1990.

[2] ISO/IEC MPEG, *Coding of moving pictures and associated audio.* Committee Draft of Standard ISO11172, Dec. 1991.

[3] ITU-T Recommendation H.262, ISO/IEC 13818-2, *Generic coding of moving pictures and associated audio information.* Draft International Standard, Mar. 1995.

[4] Draft ITU-T Recommendation H.263, *Video coding for low bitrate communication.* July 1995.

[5] ISO/IEC JTC/SC29/WG11 N937, *MPEG-4—Proposal Package Description (PPD)—Revision 2.* Mar. 1995.

[6] "Description of reference model 8 (RM8)," CCITT SGXV WG4, Specialists Group on Coding for Visual Telephony, Doc. 525, June 1989.

[7] R. C. Gonzalez and P. Wintz, *Digital Image Processing.* 2nd edition, Addison-Wesley Publishing Company, Inc., 1987.

[8] D. E. Goldberg, *Genetic Algorithms in Search, Optimization & Machine Learning.* Reading: Addison-Wesley, 1989.

[9] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision, Volume 1.* Addison-Wesley Publishing Company, Inc., 1992.

[10] Chun-Hung Lin and Ja-Ling Wu, "Genetic block matching algorithm for video coding," *in Proc. IEEE Int. Conf. on Multimedia Computing and Systems*, Hiroshima Japan, pp. 544–548, June 1996.

[11] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics 2nd edition.* Addison-Wesley Publishing Company, Inc., 1994.

[12] G. Syswerda, "Uniform crossover in genetic algorithms," *in Proc. 3rd Int. Conf. Genetic Algorithms and Applications*, CA, 1989, pp. 2–9.

[13] K. O. Lillevold, "H.263 anchors for MPEG-4 tests," *in Int. Organization for Standardization*, Dallas, Nov. 1995, ISO/IEC JTC1/SC29/WG11 MPEG95/0322.

[14] J. W. Woods, *Subband Image Coding.* Boston: Kluwer Academic Publishers, 1991.

PSNR	CBCS	FBCS
Miss (8 kbps)	34.22	34.23
Miss (14.4 kbps)	34.65	36.29
Miss (28.8 kbps)	37.61	37.69
Claire (8 kbps)	32.82	32.99
Claire (14.4 kbps)	34.41	34.95
Claire (28.8 kbps)	35.91	37.08

Table 1: *Comparisons of the average PSNR values using different bit-rate control schemes.*

WSNR	CBCS	FBCS
Miss (8 kbps)	35.38	36.99
Miss (14.4 kbps)	37.95	37.74
Miss (28.8 kbps)	40.73	39.43
Claire (8 kbps)	33.39	34.11
Claire (14.4 kbps)	34.80	34.62
Claire (28.8 kbps)	37.50	37.14

Table 2: *Comparisons of the average WSNR values using different bit-rate control schemes.*

	cbcs_ft	cbcs_fa	cbcs_at	cbcs_bg
Miss (8 kbps)	0.58	0.22	0.20	0
Miss (14.4 kbps)	0.53	0.23	0.24	0
Miss (28.8 kbps)	0.62	0.21	0.17	0
Claire (8 kbps)	0.49	0.23	0.28	0
Claire (14.4 kbps)	0.48	0.27	0.25	0
Claire (28.8 kbps)	0.51	0.26	0.23	0

(a)

	fbcs_ft	fbcs_fa	fbcs_at	fbcs_bg
Miss (8 kbps)	0.49	0.24	0.26	0.01
Miss (14.4 kbps)	0.48	0.23	0.26	0.03
Miss (28.8 kbps)	0.54	0.20	0.21	0.05
Claire (8 kbps)	0.45	0.26	0.28	0.01
Claire (14.4 kbps)	0.41	0.29	0.26	0.04
Claire (28.8 kbps)	0.42	0.30	0.22	0.06

(b)

Table 3: *Bit count ratio lists (ft: feature region, fa: face region, at: the other active regions, bg: background region.)*

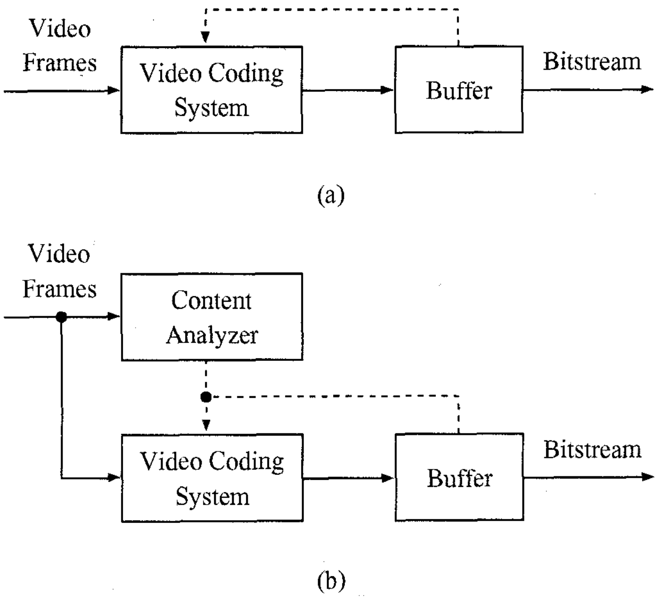


Figure 1: *Comparison of different rate control schemes.*

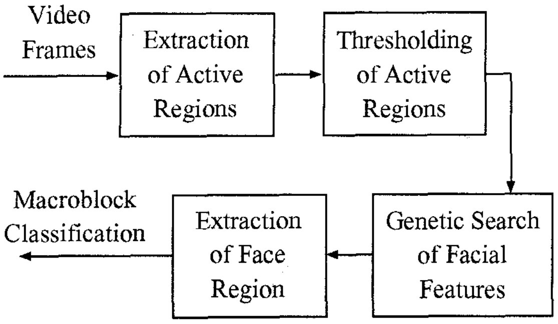


Figure 2: *The basic structure of the content analyzer.*

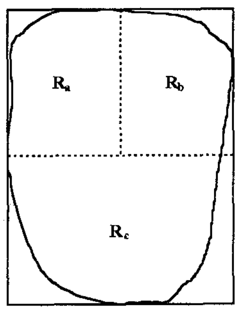


Figure 3: *Three subregions are selected from the face region to extract the main features in the feature extracting process.*

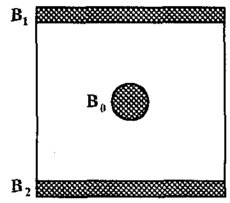


Figure 4: *The evaluation of the central weighting.*

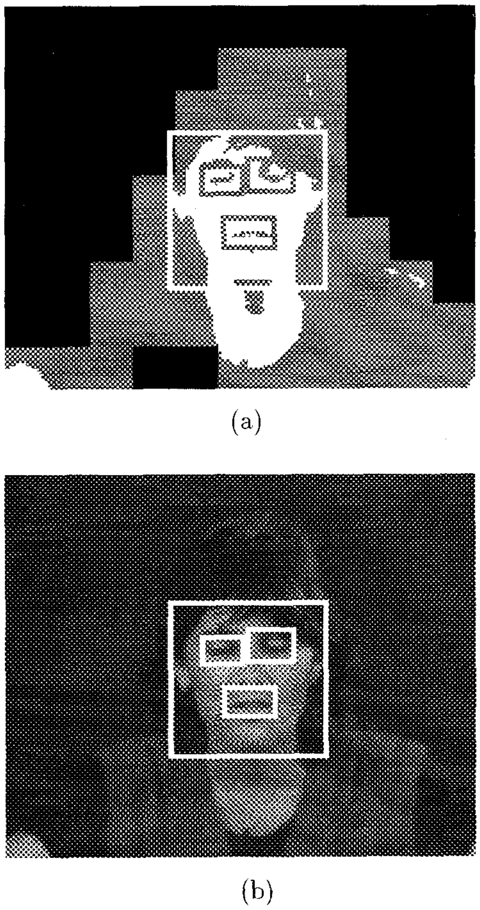
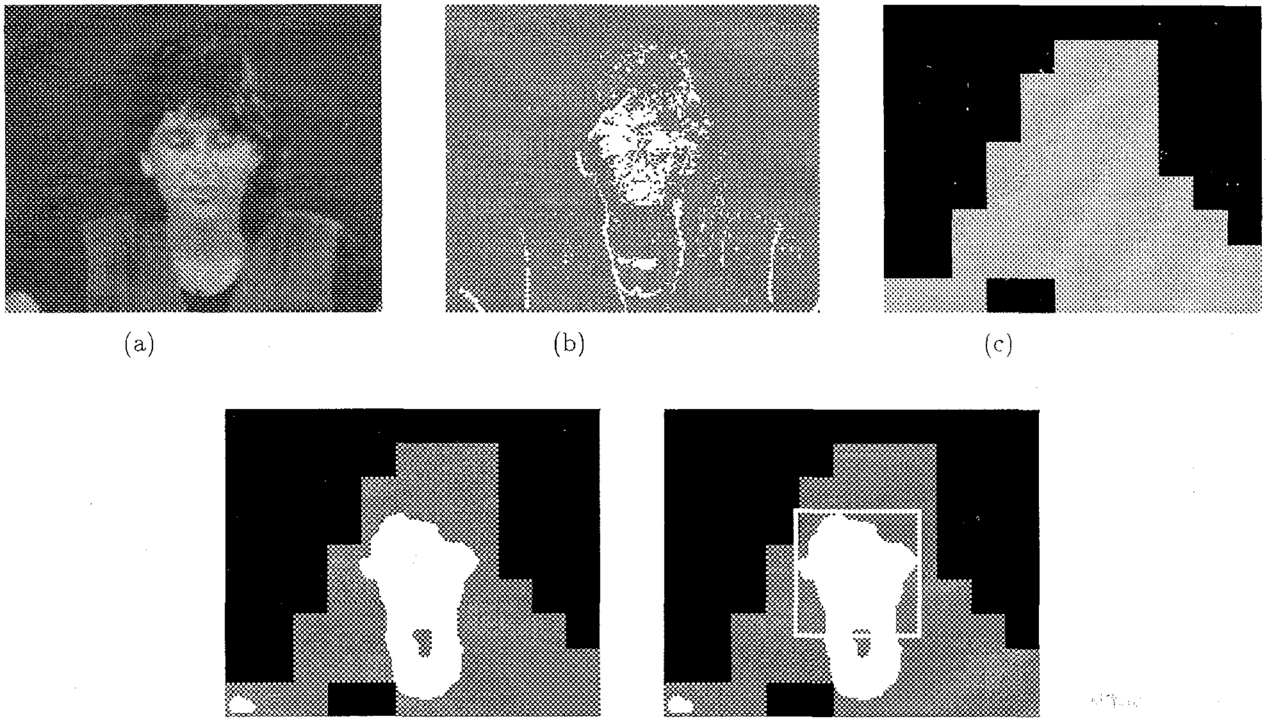


Figure 5: *Feature extraction results which are demonstrated on, (a) the thresholded image before performing closing and opening and (b) the original image.*



(a) (b) (c)

(d) (e)

Figure 6: *Intermediate results in different stages of the active region finding processes: (a) the original picture, (b) the difference image generated from the current picture and the previous picture, (c) the active regions, (d) the thresholded image after performing opening and closing, (e) the face region indicated by a rectangle.*

(a) (b)

Figure 7: *Subjective results of Frame 66 of the "Miss America" sequence at 8 kbps, (a) using the FBCS, (b) using the CBCS.*

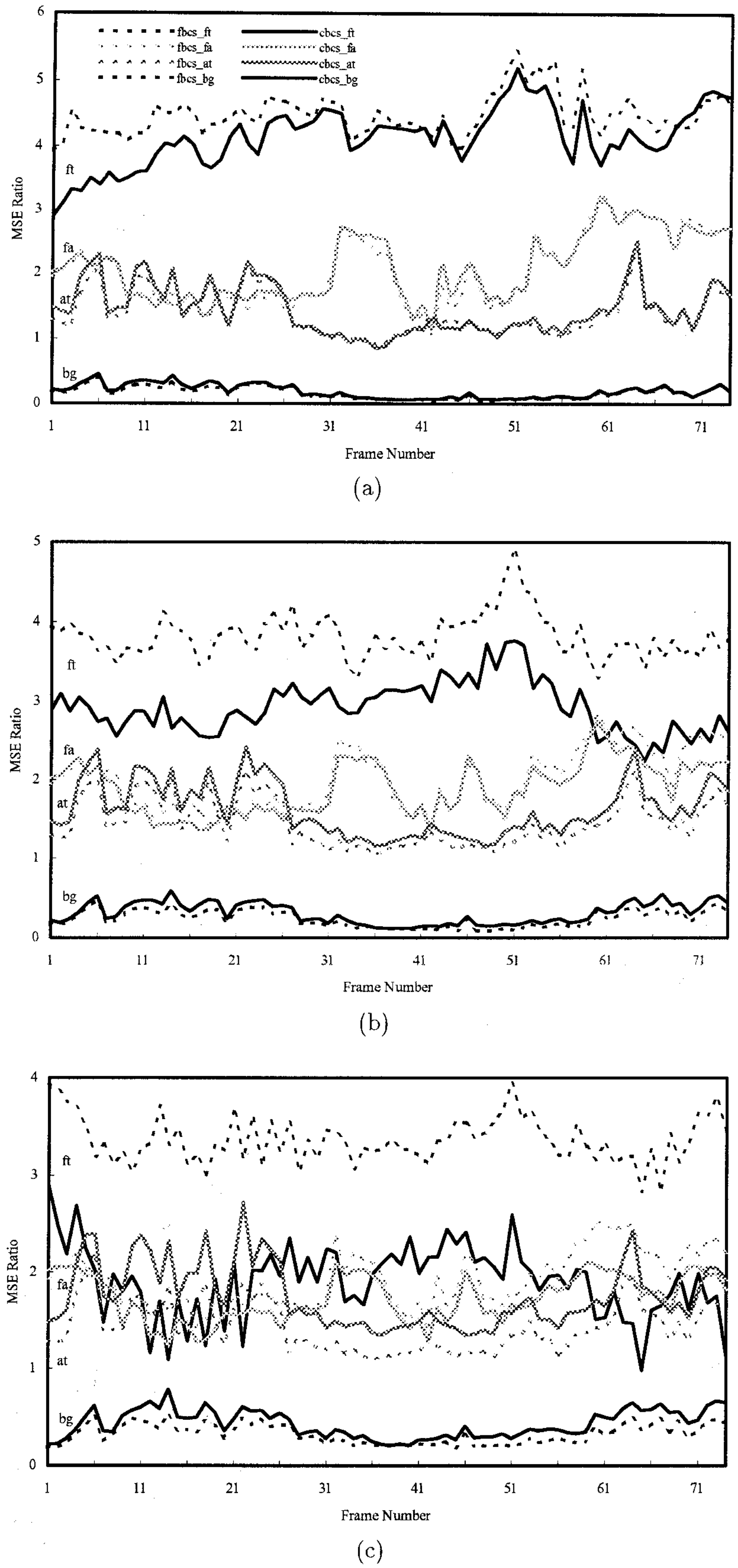


Figure 8: *MSE ratio plots of the "Miss America" sequence for, (a) 8 kbps, (b) 14.4 kbps, (c) 28.8 kbps (ft: feature region, fa: face region, at: the other active regions, bg: background region.)*