Region Classification and Image Compression Based on Region-Based Prediction (RBP) Model

Cassio M. Yorozuya Yu Liu Masayuki Nakajima

Department of Computer Science
Graduate School of Information Science and Engineering
Tokyo Institute of Technology
{yorozuya,liuyu,nakajima} @cs.titech.ac.jp

Abstract: This paper presents a new prediction method RBP *region-based prediction* model where the context used for prediction contains regions instead of individual pixels. There is a meaningful property that RBP can partition a cartoon image into two distinctive types of regions, one containing full-color backgrounds and the other containing boundaries, edges and homo-chromatic areas. With the development of computer techniques, synthetic images created with CG (computer graphics) becomes attractive. Like the demand on data compression, it is imperative to efficiently compress synthetic images such as cartoon animation generated with CG for storage of finite capacity and transmission of narrow bandwidth. This paper utilizes the property of RBP to partition cartoon images into two regions and then apply a lossy compression method to full-color regions and a lossless compression method to homo-chromatic and boundaries regions. Two criteria for partitioning are described, constant criterion and variable criterion. The latter criterion, in form of a linear function, gives the different threshold for classification in terms of contents of the image of interest. We carry out experiments by applying our method to a sequence of cartoon animation. Compared with the available image compression standard MPEG-1, our method gives the superior results in both compression ratio and complexity.

1 Introduction

With the development of computer technologies, synthetic digital images generated with CG (computer graphics) become more and more attractive. Their applications include commercial advertisement, education and entertainment. An active branch in the entertainment applications is cartoon production. Consequently, compressing synthetic images such as CG cartoon animation is imperative for finite capacity storage and narrow bandwidth transmission. This paper focuses on the compression of cartoon sequences.

In this paper we first presents a new prediction method RBP *region-based prediction model*. In conventional prediction coding algorithms, the context used for prediction only contains several individual pixels which are contiguous to a pixel to be predicted. The prediction value is the linear or nonlinear combination of the values of these neighboring pixels [1]. In our proposal, the context used for prediction contains regions instead of individual pixels. This prediction model can exactly depict the uniform regions, geometric boundaries, and regular edges for cartoon images. When the proposed prediction model is applied to a cartoon image or a cartoon sequence, an interesting and important property is exposed. That is, the prediction result manifests two distinctive regions, one being composed of full-color background or large moving objects and the other being uniformly-distributed characters or objects, edges and boundaries. We utilize the property to partition cartoon images into two regions and exploit two different methods to compress the two distinctive regions. For instance, each image is divided into blocks of size 8×8. All blocks are classified into two groups, lossless group and lossy group, according to the result of RBP and a classification criterion. The blocks of lossless group are compressed by lossless coding while the blocks of the lossy group are compressed by the lossy coding. We give two classification criteria, constant criterion and variable criterion. The former criterion is simple and intuitive while the latter criterion, in form of a linear function, gives the different threshold in terms of contents of the image of interest. The parameters for the latter are generated by linear regression with a sequence of training images. We carry out experiments by applying our method to a cartoon animation. Compared with the available compression standards MPEG-1, our method gives the superior results in both compression ratio and complexity.

2 RBP Model

This section describes a new concept **region-based prediction**, briefly RBP, which utilizes the contiguous regions, instead of individual pixels, to predict the pixel of interest. The concept is originated from the neighborhood-search method of SRC (structure run-length coding) where the pixel being predicted is depicted by one of neighboring color regions instead of individual pixels[2]. It is based on such an assumption that

Assumption 1 *An image to be processed is a multi-color image. Under the* $180°$ *type context, the number of regions contiguous to the pixel being predicted is less or equal to* N.

An image restricted to the assumption 1 implies the image without noise and color gradient. Algorithm 1 depicts the RBP.

Algorithm 1 (RBP)

1. *Search a specific neighborhood of the pixel of interest along the order shown in Fig.1 in the case of still pictures or along the order shown in Fig.2 in the case of moving pictures. Find a color set* $\mathbf{C}$ *including* N *different kinds of colors* $\{\mathbf{C_1}, \mathbf{C_2}, ..., \mathbf{C_N}\}$. *The* N *colors appear along the search order and do not repeat any preceding color. That is*

$$\mathbf{C}_i \neq \mathbf{C}_{i-1}, \mathbf{C}_{i-2}, ..., \mathbf{C}_1 \quad i = 1, ..., N \qquad (1)$$

The color set $\mathbf{C}$ *is called* **prediction set**. *Each element* $\mathbf{C}_i$ *is represented as a vector of color components* (R_i, G_i, B_i) *or* $(Y_i, (R-Y)_i, (B-Y)_i)$.

2. *Label the pixel of the color* $\mathbf{C}_p$ *as P*

$$P(\mathbf{C}_p) = \begin{cases} k & \text{if} \quad \mathbf{C}_p \equiv \mathbf{C}_k \\ & \text{and} \quad \mathbf{C}_k \in \mathbf{C} \\ 0 & \text{if} \quad \mathbf{C}_p \neq \mathbf{C}_1, ..., \mathbf{C}_N \end{cases} \qquad (2)$$

where $P(\mathbf{C}_p)$ *is the label value with range from 0 to* N.

We apply the RBP algorithm to a cartoon image shown in the left of Fig.3. The result is shown in the right of Fig.3 which represents the original cartoon image as label values. The size of color set is 3.

In Fig.3, the character, a calling boy, is of multi-colors whereas the right rear background is a full-color scenery. Label value 0 occupies most of the background. Intuitively, RBP brings us a meaningful property as the following:

...	...	...	...	...	...	...	...	...
...	22	21	20	13	19	18	17	...
...	23	11	10	5	9	8	16	...
...	24	12	4	1	3	7	15	...
...	14	6	2	@				

Figure 1: Search order in the case of still pictures. @: the pixel being% predicted.

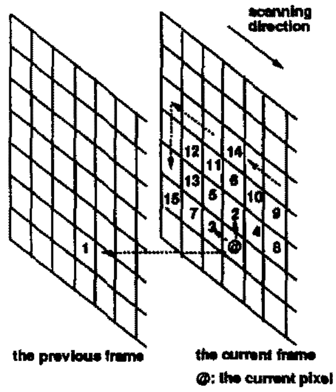

Figure 2: Search order in the case of moving pictures. @: the pixel being predicted.

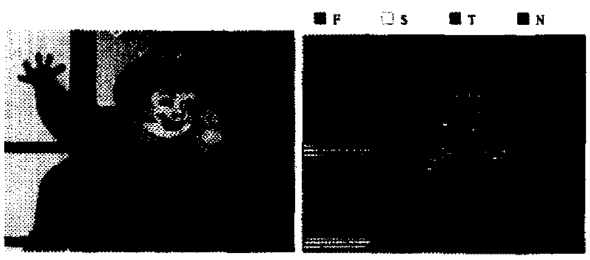

Figure 3: Original cartoon image(left) & label value image (right) where grey stands for label value 1, white for 2, dark for 3 and black for 0.

Property 1 *For an image containing homo-chromatic regions, the label values derived from RBP as described in Algorithm 1 expose two distinctive areas. One area corresponds to homo-chromatic regions, edges and boundaries while the other area corresponds to the remainder of regions. A homo-chromatic region can be represented as label values* $\in [1, N]$ *while the remainder of the image can be represented as number* 0.

3 RBP Coding Scheme for Cartoon Sequences

As well known, regions containing edges and boundaries are much sensitive to any available lossy compression algorithms. That is, any existing lossy compression easily gives rise to visible impairment to the edges and boundaries of an image. There are many schemes involving the issue of how to protect the edge regions. These schemes, mostly, start with extracting edge information or recognizing edges and then only apply heavy distortion to the regions which do not contain edges. ADCT(adaptive discrete cosine transform) coding is a well established technique. This scheme divides an image into subblocks and classifies them into several classes such as a class containing smooth regions with uniformly-distributed intensity and a class containing edge information [3][4]. Each class is matched to the best DCT encoder. Ran and Rarvardin presented a three-component image model and used it to decompose an image into three components, texture information, smooth regions and primary components of the image[5]. Lossy DCT is applied only to the textures and the smooth regions. Carlsson described a sketch-based compression scheme where contours are extracted to form sketch information [6]. An intelligible image can be reconstructed from the sketch information alone while the detailed texture is compensated by coding residual. Lynch et al. proposed an edge compensated transform coding where edges were removed from an image ahead and hence DCT with coarse quantization did not give rise to visible impairment [7]. When reconstructing the encoded image, the edges are compensated for it by the geometric edge information which is appended and encoded together. Another related field for reduction of boundary artifacts is worth referring, the image restoration where window function, or filtering technique, is utilized [8][9][10].

As an alternative to the above schemes, an approach is proposed in this section that utilizes the property 1 to partition images into lossy group which contains full-color backgrounds and lossless group which contains edge information and homo-chromatic regions. JPEG-like algorithm is applied to the lossy group.

Algorithm 2 (RBP Coding)

The algorithm takes a two-pass procedure. Each pass begins with scanning the current frame in the raster scan order from the upper-left to the lower-right corner. For a pixel being scanned, its neighborhood is searched for associating the pixel with its neighbors. The further processing is dependent on the association or the label value.

Pass 1:

1. *Search the neighborhood and label each pixel as described in Algorithm 1.*
2. *Divide the current frame into blocks of size* 8×8 *and classify blocks into two groups,* lossy-group *and* lossless-group, *according to the following criterion:*

   $lossy\_group = \{Block\ B \mid$ $the\ number\ of\ label\ 0\ in\ B \geq threshold\}$

   $lossless\_group = \{Block\ B \mid$ $the\ number\ of\ label\ 0\ in\ B < threshold\}$

3. *Make a header file which records a block of* lossless-group *as digit 0 or a block of* lossy-group *as digit 1.*

Pass 2:

1. *Search the neighborhood of the pixels which are in the blocks of the* lossless-group *and label these pixels again as done in Pass 1 except that the neighborhood excludes the pixels located in the blocks of the* lossy-group.
2. *Entropy-encode the label values from the upper-left to the lower-right in each block of the* lossless-group *where a corresponding color value is appended to label value 0.*
3. *Encode each block of the* lossy-group *with DCT and coarse quantization like lossy JPEG.*

Threshold Determination

Two factors affect the quality of image and the compression ratio. One is the *threshold* for classifying blocks and the other is the quantization of DCT coefficients for the blocks of lossy group. The higher the *threshold*, the higher the quality of image but the lower the compression ratio. On the other hand, the coarser the quantization of DCT coefficients, the higher the compression ratio but the lower the quality of image. SRC[2] and JPEG can be treated as two extreme cases when the *threshold* is set as $8 \times 8 + 1$ and 0 respectively. In the case of a multi-color image, the number of label value 0 is small while for a natural image like any photograph picture the number of label value 0 approximates to the size of the image. Therefore, the optimal *threshold* is dependent on the image. We give two criteria for the *threshold*, constant and variable

In the constant criterion, we choose the *threshold* as a constant 0.5, which implies that

> *When a block contains more than 50% of pixels which belong to the full-color background, it is classified into the lossy group, that is, the block is compressed in lossy form. When a block contains at most 50% of pixels which belong to the multi-color area, it is classified into the lossless group, that is, the block is compressed in lossless form.*

In the variable criterion, we assume that the *threshold* is dependent on the global frequency of the label value 0 like the following expression:

$$threshold = \Psi(f_o) \quad (3)$$

where f_0 is the global frequency of label value 0 and is estimated in the first pass of Algorithm 2. A simple expression for $\Psi(\cdot)$ is

$$\Psi(f_o) = a \times f_o + b \quad (4)$$

where the coefficients a and b can be determined by linear regression with a training set of images.

4 Experiment & Preliminary Result

We used a cartoon sequence named Goalkeeper-2 in our experiments. In the cartoon sequence, an animal character plays soccer in a green soccer field. Some frames contain the still background while other frames contain the background with zoom or pan motions. The character always tries to catch the speedy ball by quickly moving his body. The background contains audience in full colors and the field with green grass and the goal gate with regular frames. Compared with cartoons generated with computer painting tools, Goalkeeper-2 contains both homo-chromatic regions and gradient regions.

The front six frames of Goalkeeper-2 contain pan and zoom motions of the background. Fig.4 shows the 6th frame of Goalkeeper-2. Fig.5 shows the labeling result. Based on the labeling result the image

Table 1: MPEG configuration.

IPB pattern	IBBPBBPBBPBBPB
FRAME RATE	30 frames/sec
Q-Scale values for I picture	3
Q-Scale values for P picture	4
Q-Scale values for Q picture	7
PSEARCH algorithm	LOGARITHMIC
BSEARCH algorithm	CROSS2
PIXEL	FULL
Bit Rate Limiter	——

is classified into lossy group and lossless group as shown in Fig.6. From frame 7 to 27, the background remains still while the animal character quickly moves for catching the ball. Fig.7 shows the 10th frame of Goalkeeper-2. Fig.8 shows the labeling result whereby the image is partitioned into the lossy group and the lossless group as shown in Fig.9.

The comparison between MPEG-1 and our RBP coding is shown in Fig.10 where RBP coding takes on the constant criterion and PSNR's of Y signals for two methods are in terms of the bit rate. MPEG-1 codec software is offered by the University of California at Berkeley whose compression group has greatly contributed to the progress of MPEG standard [11]. MPEG parameters are set as Table 4. The bit rate of MPEG compression is adjusted to fit the RBP coding by setting such parameter as Q-Scale (quantization scale) of the MPEG encoder until obtaining the same bit rate. As shown in Fig.10, The RBP coding achieves better PSNR than MPEG by 5dB on average. Peak improvements can be observed at the frames 0 whereas worse PSNR are observed at the frames 16, 32 and 34.

Fig.11 shows the result of comparison between RBP coding with constant criterion and RBP coding with variable criterion. Coefficient a takes on 3.3×10^{-5} and b takes on 21.29 in the formula 4. They are estimated using linear regression with a sequence of training images. a approaches 0, which means the threshold is not so closely dependent on f_0 in the linear sense. $b = 21.29$ is equivalent to 32% in the percent form of expressing the threshold.

5 Conclusions & Future Work

In this paper we first presented a new concept RBP model and showed an interesting property of RBP. Then, we applied RBP to cartoon sequences and obtained a meaningful representation of the image: homo-chromatic regions, edges, and boundaries can be represented as the label values with range $[1, N]$ while full-color background can be represented as the label value 0. We employed the block coding technique to divide a whole image into sub-blocks and classified sub-blocks into two different groups based on the representation result. A well-known lossy method DCT with coarse quantization is employed to compress the lossy group. The experimental result showed that to some degree our proposal's performance surpass MPEG-1. We also gave two criteria for classifying images, constant criterion and variable criterion. Although the latter criterion did not gave the better PSNR than the former criterion according to the preliminary result shown in Fig.11, the latter criterion described an approach to theoretic determination of the threshold. Consequently, our future work includes

1. collecting larger samples of cartoon sequences as training images to determine the coefficients of formula 4.
2. substituting for formula 4 the other possible function which is more reasonable to practically fit cartoon images.

References

[1] Nasir D. Memon and Khalid Sayood: "Lossless Image Compression: A Comparative Study", SPIE Vol.2418, pp.8-20

[2] Y.ARAI, T.AGUI and M.NAKAJIMA: "Structure Run-length Coding for Digital Image", IECE Japan, Vol.E69, No.11, pp.1225-1230, 1986

[3] W.H.Chen and C.H.Smith: "Adaptive Coding of Monochrome and Color Images", IEEE COM-25, pp.1285-1292, Nov.1977

[4] W.A.Pearlman: "Adaptive Cosine Transform Image Coding with Constant Block Distortion", IEEE COM-38, pp.698-703, May 1990

[5] Xiaonong Ran and Nariman Rarvardin: "Adaptive DCT Image Coding Based on A Three-component Image Model", IEEE ICASSP, Vol.3, pp.III-201, 1992

[6] Stefan Carlsson: "Sketch Based Coding of Grey Level Images", Vol.15, No.1, pp.57-83, 1988

[7] William E. Lynch, Amy R.Reibman, and Bede Liu: "Edge Compensated Transform Coding", Proc. IEEE ICIP (inter. conf. image processing), pp.105-109, 1994,

[8] Farzin Aghdasi and Rabab K. Ward: "Reduction of Boundary Artifacts in Image Restoration", IEEE Trans. Image Processing, Vol.5, No.4, pp.611-618, April 1996

[9] Hock Lim, Kah-Chye Tan, and B.T.G.Tan: "Edge Errors in Inverse and Wiener Filter Restorations of Motion-Blurred Images and their Windowing Treatment", CVGIP: Graphical Models and Image Processing, Vol.53, No.2, pp.186-195, Mar. 1991

[10] Kah-Chye Tan, Hock Lim, and B.T.G.Tan: "Windowing Techniques for Image Restoration", CVGIP:Graphical Models and Image Processing, Vol.53, No.5, pp.491-500, Sep.1991

[11] Available MPEG codec software, ftp://mm-ftp.cs.berkeley.edu/pub/multimedia/mpeg/encode

Figure 4: Frame 6 of Goalkeeper-2.

Figure 5: Labeling representation for frame 6. black= 0, white= 1, gray= 2, dark= 3.

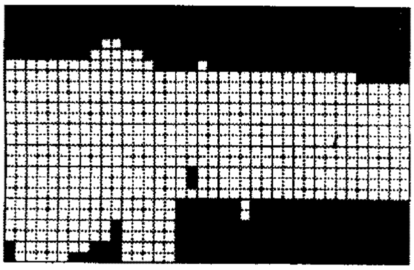

Figure 6: Partitioning result: Blocks of lossy group (black) and blocks of lossless group (white).

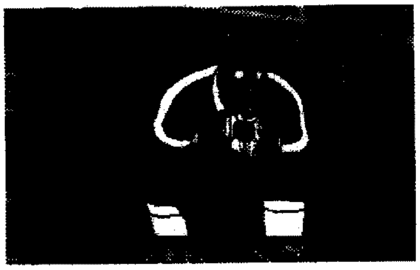

Figure 7: Frame 10 of Goalkeeper-2.

Figure 8: Labeling representation of frame 10. black= 0, white= 1, gray= 2, dark= 3.

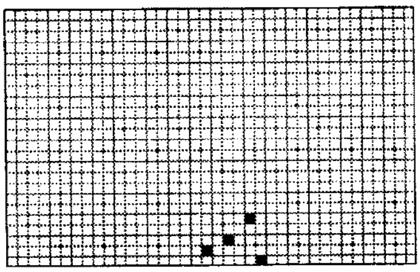

Figure 9: Partitioning result: Blocks of lossy group (black) and blocks of lossless group (white).

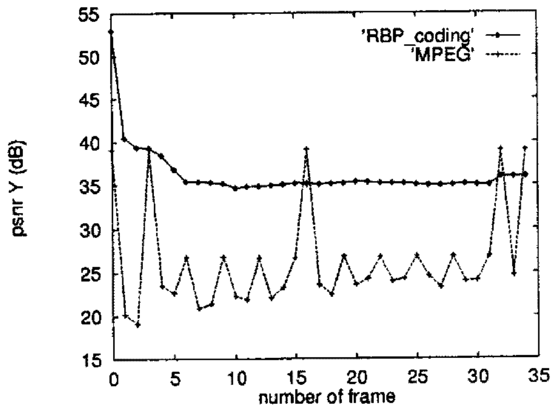

Figure 10: PSNR: MPEG vs. RBP Coding with constant criterion.

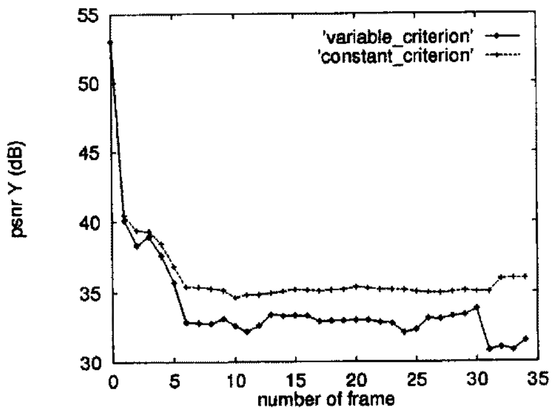

Figure 11: PSNR(RBP Coding): Constant criterion vs. variable criterion.