An Efficient Compound Image Compression by Exploiting Spatial Correlation

Meenakshi. P

M.E-Embedded System Technologies

SIET Coimbatore, Tamil Nadu,India

Jabanesh. P

M.E-Embedded System Technologies

SIET Coimbatore, Tamil Nadu,India

Abstract

Combination of text, graphics and natural image is called the compound image. Compound image is difficult to compression with normal schemes because for text, graphics and natural images needs separate schemes to compress. So we need a scheme which will compress the compound image. A coding scheme from the H.264 intraframe coding is proposed in this paper. In this scheme two intramodes are developed to better exploit the spatial correlation in compound images. Those two modes are residual scalar quantization (RSQ) and base colors and index map (BCIM). Both modes are used to better exploit spatial correlation in compound images. Proposed scheme improves the compression efficiency more than 10db at most bit rates for compound images. Rate-Distortion Optimization is used to select the mode

1. Introduction

The number of connected computers and other digital devices keeps growing; there has been a critical need for real-time computer screen image transmission technologies. Remote control software, such as AT&T virtual network computing (VNC), allows a person at a remote computer (the client, maybe a Linux machine) to view and interact with another computer (the server, maybe a Windows PC) across a network, as if sitting in front of the other computer. A smart display device, such as Microsoft Mira, acts as a portable screen with 802.11b wireless connection to a nearby desktop PC, enabling people to surf the web or browse pictures that are stored on the desktop PC. Another application is wireless projector which provides the flexibility to site anywhere in the room without cable connecting to the presentation computer. JPEG [1] is an international standard for the lossy and lossless compression of images. In lossy JPEG, image pixels are divided first into non overlapping 8 x 8 blocks. Each block is transformed into the frequency domain using the discrete cosine transform (DCT) and the output of the DCT [5] is quantized and coded using a combination of run-length and Huffman coding. Baseline JPEG [1], [2] allows for two quantization tables per image; one for the luminance and one for the chrominance components. Data compression algorithms are essential for these real-time applications, since a huge amount of image data is to be transmitted in real time. One 800 600 true color screen image has a size of 1.44 MB, and 85 frames/second produce more than 100-MB data. Without data compression algorithms, it is definitely impossible to transmit such a large volume of data over the state-of-the-art bandwidth-limited networks in real time. Although the network bandwidth keeps growing, compression algorithms can achieve more efficient data transmission, especially for smart devices and wireless projectors [3]. Taking the extreme anisotropic features of text and graphics into account, some schemes are proposed for compound image coding.

1.1. Image-coding-based approaches

They adopt conventional image coding schemes but improve the bit allocation between text/graphics and natural image areas because the text/graphics areas are often blurred after compression [7]. Thus, the quantization steps in text/graphics areas are decreased and more bits are allocated to them. For a fixed bit budget, it would correspondingly decrease bits for the coding of natural image areas. Consequently, the overall quality after compression is still not good.

1.2. Layer-based approaches

They adopt the mixed raster content (MRC) image model for compression [8], where one compound image is decomposed into a foreground layer, a background layer, and a binary mask plane at block or image level. The mask plane indicates which layer each pixel belongs to and can be compressed by mature binary coding schemes, such as JBIG and JBIG2 [9]. The foreground and background layers are smoothed by data filling algorithms and then compressed by conventional image coding schemes. It demonstrates significant gains over conventional image coding schemes.

1.3. Block-based approaches

They first classify blocks in compound images into different types according to their spatial properties. Image features, such as histogram [13], gradient and the number of colors, are often used for classification. Then different type blocks are compressed by different coding schemes to better adapt to their statistical properties. Considering the sparse histogram distribution of colors in text/graphics blocks, a novel method is proposed to represent a text/graphics block by several base colors and an index map [4]. Thus, the coding performance of that scheme on compressing compound images is significantly improved.

2. Relative works

In our scheme, two new intra-modes (RLSQ and BCIM) are proposed to exploit the spatial correlation in compound images. In the RLSQ mode, intra-prediction residues are directly scalar quantized and coded by CABAC without transform [10]. The idea of directly compressing intra-prediction residues has been reported in for lossless [6] coding of natural images. In the BCIM mode, the conversion from an image block to several base colors and an index map is extended to different size blocks to adapt to the non-stationary properties of compound images. Moreover, the rate cost is considered in the proposed BCIM mode and the number of base colors is determined in the sense of rate-distortion optimization [4]. Following the coding of base colors, each index in the map is coded by context-based entropy coding. With the two new modes and the RDO mode selection method, the proposed scheme significantly outperforms all existing schemes on compressing compound images; meanwhile, it keeps the comparable performance to H.264 on coding natural images.

2.1 Spatial orientation

Normally, most of the image's energy is concentrated in the low frequency components. Consequently, the variance decreases as one move from the highest to the lowest of the sub band pyramid [19]. There is a spatial self similarity between sub bands, and the coefficients are expected to be better magnitude-ordered as one move downward in the pyramid following the same spatial orientation. A tree structure, called spatial orientation tree, naturally defines the spatial relationship on the hierarchical pyramid.
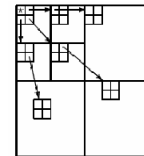
"**Figure 1** Parent –offspring dependencies in spatial orientation tree"

Fig. 1 shows how the spatial orientation tree is defined in a pyramid constructed with recursive four-band splitting. Each node of the tree corresponds to a pixel, and is identified by the pixel coordinate. Its direct descendants (offspring) correspond to the pixels of the same spatial orientation in the next finer level of the pyramid. The tree is defined in such a way that each node has either no offspring or four off-springs [20], which always form a group of 2X2 adjacent pixels. The pixels in the highest level of the pyramid are the tree roots and are also grouped in 2X2 adjacent pixels. However, their offspring branching is different, and in each group one of them (indicated by the star in Fig 1) has no descendants. Parts of the spatial orientation trees are used as the partitioning subsets in the sorting.

Spatial prediction in H.264/AVC [3] uses several spatial prediction block sizes including 16×16, 8×8, and 4×4 for luma and 16×16 (for 4:4:4 video), 16×8 (for 4:2:2 video), or 8×8 (for 4:2:0 video) for chroma [19]. For each case, the encoder selects a directional spatial prediction mode which governs the creation of a prediction of the complete block of samples using the values of samples in neighboring blocks that have previously been decoded specifically, the column of samples immediately to the left of the block to be predicted and the row of samples immediately above the block to be predicted.

2.2 Residual scalar quantization

The text blocks and graphics blocks containing edges of many directions, intra-prediction along a single direction cannot completely remove the

directional correlation among samples. After intra-prediction, residues still preserve strong anisotropic correlation [14]. In this case, it is not efficient to perform a transform on them. One method is to skip the transform and directly code prediction residues, which is similar to traditional pulse-code modulation (PCM). However, the question is whether the performance of PCM is better than that of a transform for text and graphics residual blocks. To answer it, we introduce the method proposed to analyze the coding gain of PCM over a transform. Given the same rate, the coding gain is defined as the ratio of distortions on transform coefficients and residual samples, respectively, as

$$G_{PCM/TC} = \frac{D_{TC}}{D_{PCM}} \qquad (1)$$

Is the distortion of PCM and is the distortion on transform coefficients. We assume the distortions result from the optimal quantization.

If $G_{PCM/TC}$ is larger than 1, it indicates that non-transform coding is more efficient than transform coding. To statistically investigate the coding gain of PCM over a transform on text and graphics residual blocks, we collect 8297 residual blocks of size 8×8 from text and graphics parts.

2.3 Base color index map

The text/graphics blocks can be expressed concisely by several base colors together with an index map. It is somewhat like color quantization that is a process of choosing a representative set of colors to approximate all the colors of an image [12]. In the BCIM mode, we first get the base colors of a block by using a clustering algorithm. All the base colors constitute a base color table. Then, each sample in the block will be quantized to its nearest base color. The index map indicates which base color is used by each sample. Different from color quantization, each text/graphics block, but not an entire image, has its own base colors and an index map for representation in our scheme. Thus, it is content adaptive for each block [5]. In addition, since the base color number of a block is small, fewer bits are required to represent each mapped index. Let us take the luminance plane of a 16×16 text/graphics block as an example, with each sample expressed by 8 bits. If four base colors are selected to approximate colors of that block, only two bits are required to represent each sample's index without compression

2.4 Mode selection and mode structure

In this mode selection, each block select its mode by RDO algorithm. Each mode has its advantages at

dealing with blocks of different features. One question that arises here is how to fully take advantage of each mode in the proposed scheme. It can be solved by the RDO algorithm [22] that has been adopted by H.264. The best mode with the best block partition having the minimum rate-distortion cost will be selected to compress the current block.

3. Proposed scheme

H.264 intraframe coding provides a flexible and efficient framework to accommodate them. The block diagram of the proposed scheme for coding compound images as well as natural images is depicted in Fig. 2. Similar to H.264, an input image is partitioned into regular and non-overlapped blocks. There are three possible paths for coding each block. The first path is the existing method in H.264 intraframe coding [18]. Intra-directional prediction and transform coding contribute much to the high compression efficiency on natural image parts.
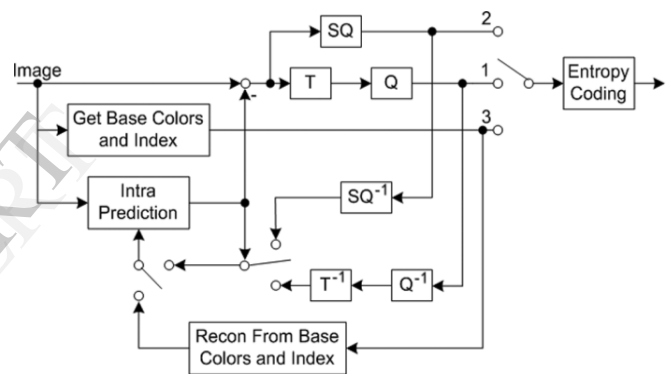

"**Figure 2.** Structure of proposed scheme"

The second path is the proposed Residual Lossless Scalar Quantization (RLSQ) method [14]. After the directional prediction, the residual is directly quantized and entropy coded. The third path is the proposed BCIM method. It does not use intra-prediction. The input block is converted to a compact representation of several base colors and an index map. The above three methods can applied to different block sizes of 16×16, 8×8, and 4×4. The mode selection is completed by the lossless video coding techniques similar to that in H.264 [12]. In addition, no matter which mode is selected, the reconstructed samples will update the reference buffer as the reference for subsequent samples.

All modes in the proposed scheme can be categorized into two types: spatial domain (SD) and DCT frequency domain (FD) [17]. There is a flag in the bit stream to distinguish them. The organized structure of all the modes is shown in Fig. 2. FD indicates the original intramodes in H.264 [21],

where the compression is performed in the DCT domain. SD indicates our proposed RLSQ and BCIM modes [15]. To adapt to the local non-stationary property of compound images, the spectral active measure (SAM) modes are applied to 16×16, 8×8, and 4×4 block sizes as those DCT spatial frequency measure (SFM) modes [17] shown in figure 3. The best mode in the spatial domain is compared with the best mode in the DCT frequency domain for the same size block in the rate-distortion sense. The better one is selected. For 8×8, 4× 4 size blocks, 8 prediction directions are designed for the RLSQ mode. The DC mode in the spatial domain is replaced by the BCIM mode in the stream syntax. For those small size blocks, the BCIM mode is only
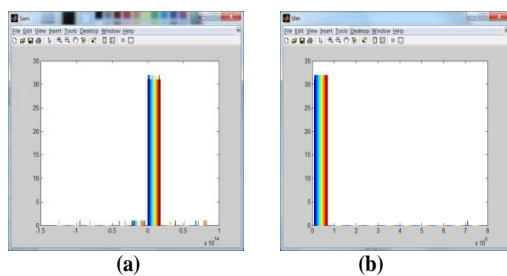
"**Fig 3.a)** Histogram of SAM **b)** Histogram of SFM"

performed on the luminance component in our scheme for simplicity. In 16×16 blocks, the BCIM mode takes the place of the DC intra-mode.

The better one between Dim3 and Dim1 is selected based on the rate-distortion criteria [11]. For the Dim3 case, when the input image format is YUV 4:2:0, interpolation will be performed on UV color planes to get the same size color planes to facilitate the 3-D clustering.

(a) (b)

(c) (d)

(e) (f)
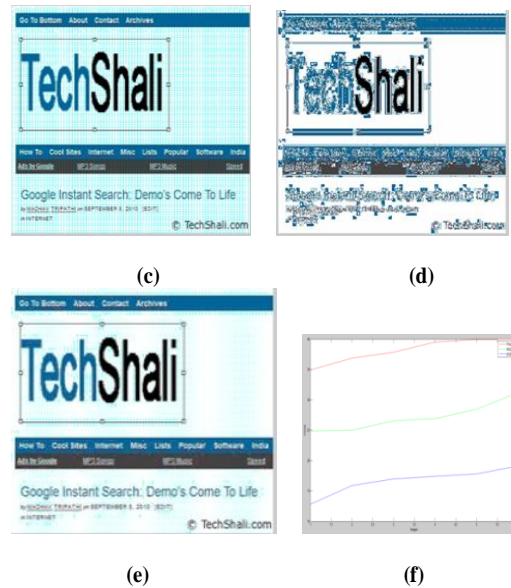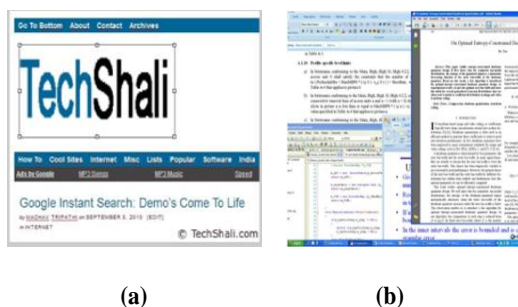
"**Figure 4**.
a) Sample Compound image **b)** Sample file image
c) Sample image of RLSQ **d)** sample image of BCIM
e) Proposed scheme **f)** Graph for proposed"

4. Implementation Result

In this paper, a lossless [6] intra coding method based on sample wise mode structure has been presented in the context of the block based H.264/AVC design. As shown in Fig. 4, three captured screen images and a compound document are used as test images: (a) sample compound image (b) is a sample file image (c) is a sample image of RLSQ. Their size is 1280×1024. Text and graphics are different in different compound images or in different regions of the same image. Some text and graphics blocks have no transition, whereas others have rich shadows. Some symbols on them are small and only occupy a 16×16 or smaller size block but others may take up several blocks. The percentages of the RLSQ and BCIM modes in all testing images are given in Table I. They are obtained at three different rates: 0.4, 0.8, and 1.2 bpp. One can observe that the percentages of the RLSQ mode increase with rate increasing, whereas the percentages of the BCIM mode decrease. But the phenomenon is not clear in Natural Image. With the proposed two modes, the perceptual quality is greatly improved, close to the original image. By combining both RLSQ and BCIM we can get output more efficiently, that is proposed scheme in this paper, that output is shown in fig-4(e). We can observe the proposed and two modes in the graph which is fig-4(f)

5. Conclusion

We propose a compound image compression scheme by fully exploring spatial domain properties of compound images. Two spatial domain modes, called residual lossless scalar quantization (RLSQ) and base colors and the index map (BCIM), are integrated into H.264 intraframe coding; they achieve significant gains at all bit rates. The RLSQ mode can cope with complicated text and graphics blocks in a simple way, which is just to quantize the intra-prediction residues without a transform. The BCIM mode provides the ability to have a high performance improvement for the efficient representation form of the text/graphics block. They are both able to preserve the spatial structures of the text and graphics parts, important to visual quality. A rate distortion optimal method, similar to that in H.264, simplifies the mode selection and avoids the performance loss imported by the inaccurateness of segmentation. In short, this paper points out a good way to extend H.264 to compress compound images with simple technical extensions and to moderate complexity increasing because of addition mode selections.

6. References

[1] W. P. Penne baker and J. L. Mitchell, JPEG: Still Image Compression Standard. New York: Van Nostrand Reinhold, 1993.

[2] D. Taubman and M. Marcelline, JPEG2000: Image Compression Fundamentals, Standards, and Practice. Norwell, MA: Kluwer, 2001.

[3] Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264/ISO/IEC 14496-10 AVC), 2003 JVT of ISO/IEC MPEG and ITU-T, JVTG-050.

[4] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 560–576, Jul. 2003.

[5] W. Ding, F. Wu, X. Wu, S. Li, and H. Li, "Adaptive directional lifting-based wavelet transform for image coding," IEEE Trans. Image Process., vol. 16, no. 2, pp. 416–427, Feb. 2007.

[6] X. Li and S. Lei, "On the study of lossless compression of computer generated compound images," in Proc. Int. Conf. Image Processing, vol.3, 2001, pp. 446–449.

[7] F. Ono, I. Ueno, T. Takahashi, and T. Semasa, "Efficient coding of computer generated images with acceptable picture quality," in Proc. Int.Conf. Image Processing, vol. 2, 2002, pp. 653–656.

[8] "Mixed raster content (MRC)," ITU-T Study Group 8, Draft Recommendation T.44, 1997.

[9] R. de Queiroz, R. Buckley, and M. Xu, "Mixed raster content (MRC) model for compound image compression," Proc. SPIE, vol. 3653, pp.1106–1117, 1999.

[10] L. Bottou, P. Haffner, P. G. Howard, P. Simard, Y. Bengio, and Y. LeCun "High quality document image compression with DjVu," J. ElectronImage., vol. 7, no. 3, pp. 410–425, Jul. 1998.

[11] P. Haffner, L. Bottou, P. G. Howard, and Y. LeCun, "DjVu: Analyzingand compressing scanned documents for Internet distribution," presented at the Int. Conf. Document Analysis and Recognition, Sep. 1999.

[12] D. huttenlocher, P. Felzenszwalb, and W. Rucklidge, "DigiPaper: A versatile color document image representation," in Proc. Int. Conf. Image Processing, vol. I, Oct. 1999, pp. 219–223.

[13] J. Huang, Y. Wang, and E. K. Wong, "Check image compression using a layered coding method," J. Electron. Image. vol. 7, no. 3, pp. 426–442, Jul.1998.

[14] ITU-T Recommendation H.264 and ISO/IEC 14496-10, Advanced Video Coding for Generic Audiovisual Services, May 2003.

[15] A. Luthra, G. J. Sullivan, and T. Wiegand, "Introduction to the special issue on the H.264/AVC video coding standard," IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 557–559,Jul. 2003.

[16] G. J. Sullivan and T. Wiegand, "Video compression—from concepts to the H.264/AVC standard, " Proc. IEEE, no. 1, pp. 18–31, Jan. 2005.

[17] ITU-T Video Coding Experts Group (VCEG), Video Codec Test Model Near-Term, Version 10 (TMN10) Draft 1, Apr. 1998.

[18] ITU-T Recommendation H.262 and ISO/IEC 13818-2, Information Technology—Generic Coding of Moving Pictures and Associated Audio Information: Video, Jul. 1995.

[19] G. J. Sullivan, T. McMahon, T. Wiegand, and A. Luthra, Eds., Draft Text of H.264/AVC Fidelity Range Extensions Amendment to ITU-T Rec. H.264 j ISO/IEC 14496-10 AVC, ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16 Joint Video Team document JVT-L047, Jul.2004.

[20] H. Yu, Ed. Draft Text of H.264/AVC Advanced 4:4:4 Profile Amendment to ITU-T Rec. H.264 j ISO/IEC 14496-10 AVC, ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16 Joint Video Team document JVT-Q209, Oct. 2005.

[21] Cuiling Lan, Guangming Shi, Member, IEEE, and Feng Wu, Senior Member, IEEE "Compress Compound Images in H.264/MPGE-4 AVC by Exploiting Spatial Correlation" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 19, NO. 4, APRIL 2010.

[22] http://research.microsoft.com/en-us/people/fengwu/screen_tip_10.pdf

P.Meenakshi did her bachelor of engineering in Electroincs and Communication Engineering at Jayaram College of Engineering and Technology,Trichy and doing Master of Engineering in Embedded System Technologies in Sri Shakthi Institute of Engineering and Technoloy, Coimbatotre, India. Her research interest includes Image Processing and Embedded System Design. She has presented a paper in Dindugul Anna university, Dindugul.

P.Jabanesh did his bachelor of engineering in Electrical and Electronics Engineering at Francis Xavier Engineering College, Tirunelveli and doing Master of Engineering in Embedded System Technologiens in Sri Shakthi Institute Of Engineering and Technology, Coimbatore, India. His research interest includes Image Processing and Embedded System Design.

.