Attention Model Based Progressive Image Transmission

Yusuo Hu[1*], Xing Xie[2], Zonghai Chen[1], Wei-Ying Ma[2]
[1]*Dept. of Automation, Univ. of Sci. and Tech. of China, Hefei, 230027, P.R. China*
[2]*Microsoft Research Asia, 5F Sigma Center, No. 49, Zhichun Road, Beijing, 100080, P.R. China*
[1]*yshu@mail.ustc.edu.cn,* [1]*chenzh@ustc.edu.cn,* [2]*{xingx,wyma}@microsoft.com*

Abstract

Progressive image transmission provides a convenient user interface when images are transmitted slowly. However, most of the existing PIT techniques only considered the objective quality of the reconstructed image. Here we present an attention model based progressive image transmission approach to improve the subjective quality of the transmission process. We use both bottom-up image features and top-down semantic information to extract the regions of interest and also propose a new ROI coding scheme based on JPEG2000 to control the trade-off between the transmission of ROI and background. Experiments have shown the efficiency of our approach.

1. Introduction

When people view an image through a low speed connection, for example, via a telephone line or via wireless networks, it will take much time to transmit the whole image. Transmitting a losslessly compressed 800x600 24-bit color image over a 56Kbps connection will require about 60s. Even with increased bandwidth, transmitting large images such as pictures captured by digital cameras is still relatively slow. Experiments have shown that if the delay is too long (>5-10s), user will feel nervous and even give up.

Progressive Image Transmission (PIT) techniques have been proposed to alleviate this problem by first sending a coarse version of the original image and then refining it progressively. Using PIT, users can preview the image in advance and therefore decide whether to abort the transferring process or wait for the image to be refined. PIT is especially useful for tele-browsing, tele-medicine and mobile applications.

In our opinion, the main task of PIT is to encode the original image into a code stream which satisfies that: (1) Image can be reconstructed effectively and efficiently by part of the code stream so that it can be transmitted in a progressive way. (2) Bits with more importance should appear earlier in the code stream, so that users will always get the most important information in time and it will provide the best experience in viewing the transmitted image.

While the first goal can be achieved easily by most PIT techniques and even perfectly by some so-called embedded coding techniques, the second goal of optimization is rather more difficult because of the lack of an effective importance criterion for most images.

Existing approaches for PIT have adopted, explicitly or implicitly, the minimal distortion principle to decide the importance. For example, in the SPIHT algorithm [11], the coefficients with larger magnitude are considered more significant for they will cause larger distortion. The algorithm will therefore sort the coefficients by their magnitudes before transmission.

However, in most cases, the minimal distortion principle does not necessarily provide the best viewing experience when the image is transmitted progressively. The first reason is that the image measure based on distortion does not correspond to the human psycho-visual measure. The second reason, more importantly, is that people usually pay most of their attention to only part of the image. Therefore, the distortion within these attended areas should be considered more severe than elsewhere. And the attended areas, or the regions of interest (ROIs), should be considered to be more important.

Some PIT techniques have adopted HVS (human visual system) weighting in spectral domain to improve the perceptual quality of the transmitted image [3]. However, they did not consider the attention change in spatial domain. Popular image standards such as JPEG and JPEG2000 do support ROI coding, but they do not provide any mechanism for automatic ROI definition.

In this paper, we propose a new attention model based approach for progressive image transmission. An attention model [2] is used to extract the ROIs within the image. Unlike previous attention based ROI extraction methods [1],[9],[10], which are mostly based on bottom-up image

* This work was conducted when the first author was a visiting student at Microsoft Research Asia.

0-7803-8603-5/04/$20.00 ©2004 IEEE.

features, we have incorporated both top-down semantic information and bottom-up features to extract the ROI. We use JPEG2000 to encode the image. It is known that the ROI coding of JPEG2000 will lose too much information of the non-ROI (background) part at early stages [1]. To solve this problem, we propose a new ROI coding scheme which is compatible with the JPEG2000 framework.

The rest of this paper is organized as follows: In Section 2, we review the attention model definition as well as the modeling methods, and then propose our ROI extraction algorithm. In Section 3, the progressive coding process is studied. Finally, in Section 4, we show the experimental results on a number of images and give our conclusions in Section 5.

2. Attention Model Based ROI Extraction

In this section, we first give the definition of our attention model and then use it to extract the regions of interest. Previously, this attention model has been successfully used in image adaptation [2], video summarization [5], and mobile picture browsing [7].

2.1 Attention Model

Definition 1: The visual attention model for an image is defined as a set of attention objects.

$$\{AO_i\} = \{(RECT_i, AV_i, MPS_i)\}, \quad 1 \le i \le N \qquad (1)$$

where

AO_i,	the ith *attention object* within the image
$RECT_i$,	position and size of AO_i
AV_i,	attention value of AO_i
MPS_i,	minimal perceptible size of AO_i
N,	total number of attention objects

The *MPS* property is not considered in this work. We are only interested in the position, size and attention value of the objects.

Neurological research has shown that human visual attention is not only affected by low level image features but also guided by high level semantic information. Therefore, we use both bottom-up and top-down methods to model the attention.

We apply the fuzzy growing method [8] to extract the salient regions based on image contrast. Face and text in an image are considered to be important top-down information. We use face and text detection algorithms to obtain these two kinds of attention objects. The attention values of all the attention objects are calculated by a number of heuristic rules. The detail of the attention modeling algorithm can be found in [2].

After attention modeling, we will get a set of attention objects with different attention values. The regions of interest will then be extracted from these attention objects.

2.2 ROI Extraction

We first sort all the attention objects by their attention values in a descending order, i.e. $AV_1 > AV_2 > \ldots > AV_n$. Experiments have shown that the total area of the ROIs should not be too large and the number of the ROIs should be small to keep the encoding/decoding efficient [1]. Therefore, we try to find the maximal M that satisfies:

$$\text{(a)} \quad Area\left(\bigcup_{i=1}^{M} RECT_i\right) < \frac{1}{4} ImageSize \qquad (2)$$

$$\text{(b)} \quad M < N_{\max} \qquad (3)$$

where $N_{\max}$ is a predefined threshold. In our experiments we set $N_{\max}$ =6.

The region of interest R is then defined as the union of the selected AOs, i.e.

$$R = \bigcup_{i=1}^{M} RECT_i \qquad (4)$$

An example of ROI extraction on one of the tested images is shown in Fig. 1. We can see that the most interested parts of the image have been successfully extracted.

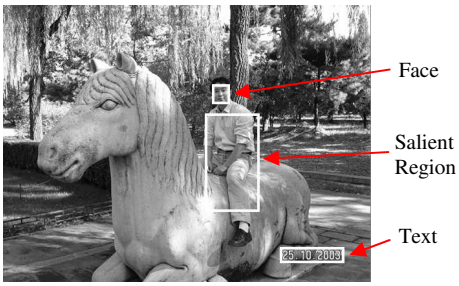


Fig.1 An example of ROI extraction

The region of interest R may contain multiple disconnected sub-regions and some of them may be irregularly shaped because of the overlapping of attention objects. As we will see, the JPEG2000 encoder can efficiently deal with multiple ROIs with arbitrary shapes, so we directly build the ROI mask bitmap from R and use the map to encode the final bit-stream.

3. Progressive Image Coding

JPEG2000 is used to encode the image with the extracted ROIs. In this section, we first give a brief description of the JPEG2000 coding standard, and then discuss the ROI coding part.

3.1 JPEG2000 Architecture

JPEG2000 [13] is an efficient coding standard for lossy or lossless multi-component image coding. It has a highly scalable structure. The encoding process consists of following stages: First, for each component, the pixel data is transformed using reversible or irreversible wavelet transformation and an orientation tree sub-band structure is generated. The wavelet transform coefficients are then quantized into integer indices. Afterwards, the indices of each sub-band are divided into small code blocks (e.g.

32x32 pixels) and bit-plane coding is performed in each code block independently. The coded data constructs several quality layers. Finally, the code blocks are also grouped into precincts with a nominal size for each sub-band. The code coming from each precinct, layer, resolution level and component will be wrapped into a packet and all the packets are organized to form the final bitstream in a certain progressive order.

Five progressive orders have been defined in JPEG2000. Among them, the layer progressive ordering is the most effective because it can provide successive improving image quality. However, the ROI coding in JPEG2000 is not well suited for the layer progressive transmission. We will discuss it further in the next section.

3.2 ROI Coding

A simple algorithm called MAXSHIFT has been adopted in the JPEG2000 standard [4]. The MAXSHIFT coding method can deal with multiple ROIs with arbitrary shapes efficiently. However, because it has separated the ROIs from the background by using different bit-planes, when the image is transmitted in a layer progressive manner, no background information will be transmitted until all the ROI parts have been completely reconstructed (Fig. 2(a)). This will cause unnecessary delay during the transmission.

Although some alternative ROI coding techniques [6][12] have been proposed to solve this problem, they are not implemented by the standard JPEG2000 baseline coder. In addition, it is difficult to choose their parameters to achieve the best result. We will show a simple yet efficient solution which is compatible with the JPEG2000 framework.

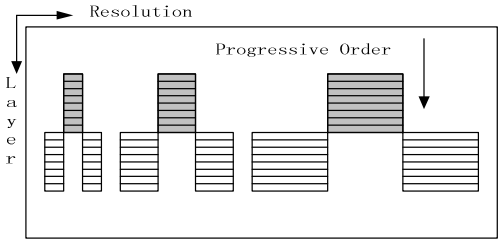
We first introduce the concept of Most Appropriate Resolution (*MAR*) of an image. The *MAR* of an image is related with its size and presentation context. A large image can have a *MAR* lower than its actual resolution and may even lower if the image is embedded in a webpage. The value of *MAR* is defined as follows:

$$MAR = \max(H - \max(\lfloor \log_2(1/K) \rfloor 0), 0) \qquad (5)$$

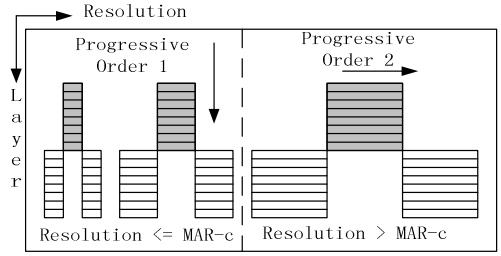
where H is the maximal level of the wavelet transform, and K stands for the desired zooming ratio, which can be decided by the display size or directly specified by the image author or the web publisher.

We first transmit the ROI data whose resolution level is no higher than *MAR-c*, then switch to transmit the background data. When all the packets whose resolution level is no higher than *MAR-c* are transmitted, we start to send the remaining data progressively. Here the constant c is used to control the trade-off. We found that for most images, $c=1$ produces the best result.

We use the POC marker segment defined in the JPEG2000 standard [13] to change the progressive order. The encoding process is illustrated as Fig. 2(b).



(a) MAXSHIFT with layer progressive ordering



(b) *MAR* ROI coding method

Fig. 2 Comparison of ROI coding schemes

The *MAR* approach ensures that users will get a fine enough view of the ROI part at early stages and also permits the background information to be transmitted in time. The concept of *MAR* is understandable and its value is easy to choose.

It is also very convenient for web publishers to use *MAR*. Images are often embedded into a web page with a layout size smaller than its actual size. By associating *MAR* with image layout size and encoding the source image in the proposed manner, the progressively transmitted image will always provide the best user experience without sacrificing the quality or maintaining different image versions.

4. Experiments

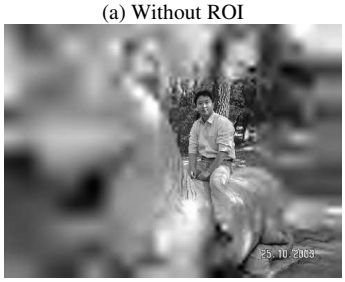
The performance of our method is tested on a dataset of 20 color images with their sizes varying from 780*600 to 2048*1536. We extract the region of interest of all the images and then encode them into JPEG2000 code streams.

We compare three coding methods in the experiment: the standard JPEG2000 coding without ROI, the standard MAXSHIFT ROI coding and the proposed *MAR* ROI coding. The *MAR* is set to be the finest resolution, i.e. the images are shown in their original size. We use a relatively slow bit-rate (1200bps) to demonstrate the effect of the progressive transmission and use JJ2000 [14] as the JPEG2000 encoder.

Through the experiments, it was found that, at early stages, both MAXSHIFT and *MAR* code streams can quickly render the ROI parts and users will get the most important information first. However, while the *MAR* code stream soon switches to refine the background, the MAXSHIFT code stream is still transmitting hardly noticeable details of the ROI parts and slows down the

improvement of the overall image quality. Some intermediate results are shown in Fig 3. We can see that the image reconstructed from *MAR* code stream can provide enough details of the ROI while also provide competitively good quality for the background.

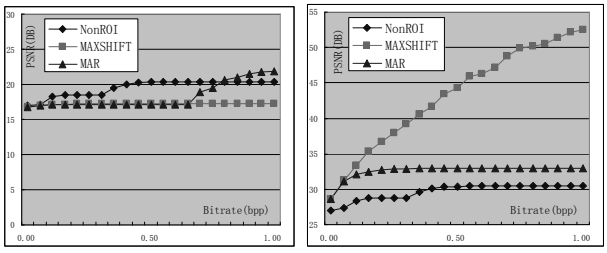
(a) Without ROI



(b)MAXSHIFT ROI coding

(c)*MAR* ROI coding

Fig. 3 Reconstructed images at 1bpp bit rate by different coding methods



(a) PSNR of the whole image (b) PSNR of the ROIs

Fig. 4 Comparison of the PSNR curves of different coding methods

The PSNR curves of the three code streams are shown in Fig. 4. We can see that the quality of the whole image has to be sacrificed in order to transmit the ROIs with higher priority. However, for the *MAR* code stream, once the PSNR of the ROIs reaches a certain level, it will stop transmitting the ROI data and start to refine the background. Therefore, the PSNR of the whole image will increase much earlier than the standard MAXSHIFT. That will result in a more convenient viewing experience during the progressive image transmission.

5. Conclusions

We have proposed a progressive image transmission scheme based on human attention. The image attention model is used to extract the region of interest and an efficient ROI coding approach has been implemented using JPEG2000. The ROI coding approach based on Most Appropriate Resolution (*MAR*) is efficient and easy to use. Experiments show that our method provides satisfactory results for progressive image transmission.

6. References

[1] Bradley A.P., Stentiford F.W.M., "Visual Attention For Region of Interest Coding in JPEG2000", *Journal of Vision Communications & Image Representation*, 14(2003) 232-250.

[2] Chen L.Q., Xie X., et al, "A visual attention model for adapting images on small displays", *ACM Multimedia Systems Journal*, vol. 9, no.4, pp. 353-364, 2003.

[3] Chitprasert B., Rao K.R, "Human Visual Weighted Progressive Image Transmission", *IEEE Trans. on Communications*, vol. 38, no. 7, pp. 1040-1044, July 1990.

[4] Christopoulos C., Askelf J. et al., "Efficient Methods for encoding Regions of Interest in the upcoming JPEG2000 still image coding standard", *IEEE Signal Processing Letters,* vol. 7, pp. 247-249, Sept. 2000.

[5] Liu H., Xie X., et al, "Automatic Browsing of Large Pictures on Mobile Devices", *ACM Multimedia 2003*.

[6] Liu L., Fan G., "A new JPEG2000 region-of-interest image coding method: partial significant bitplanes shift", *IEEE Signal Processing Letters*, vol. 10, no.2, pp. 35-38, Feb. 2003.

[7] Ma Y.F., Lu L., et al, "An Attention Model for Video Summarization", *ACM Multimedia 2002*.

[8] Ma Y.F., Zhang H.J, "Contrast-based Image Attention Analysis by Using Fuzzy Growing", *ACM Multimedia 2003*.

[9] Osberger W., Maeder A.J., "Automatic Identification of Perceptually Important Regions in an Image", *Proceedings ICPR 1998*, vol.1,16-20, pp.701 -704, Aug. 1998.

[10] Privitera C.M., Stark L.W., "Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations", *IEEE Tran. PAMI*, vol.22, no.9, Sept. 2000.

[11] Said A., Pearlman W.A., "A New, Fast, and Efficient Image Codec Based on Set Parttitioning in Hierarchical Trees", *IEEE Trans. On Circuits and Systems for Video Technology*, vol. 6. no. 3, pp. 243-250, June 1996.

[12] Wang Z., Bovik A. C., "Bitplane-by-Bitplane Shift (BbBShift) - A Suggestion for JPEG 2000 Region of Interest Coding", *IEEE Signal Processing Letters*, vol. 9, no. 5, pp.160-162, May 2002.

[13] *ISO/IEC JTC 1/SC 29/WG 1 (ITU-T SG8) JPEG2000 Part I Final Cmittee Draft Version 1.0*, Mar. 2000.

[14] JJ2000 Site: *http://jj2000.epfl.ch/.*