PERCEPTION-BASED IMAGE TRANSCODING FOR UNIVERSAL MULTIMEDIA ACCESS

Keansub Lee, *Hyun Sung Chang, Seong Soo Chun, Hyungseok Choi, Sanghoon Sull

School of Electrical Engineering Korea University, Seoul, Korea
**Electronics and Telecommunications Research Institute (ETRI), Korea*
*{ leeks, sschun, chs, sull }@mpeg.korea.ac.kr *hyunsung@computer.org*

ABSTRACT

In this paper, we propose a novel scheme for generating transcoded (e.g. scaled and cropped) image to fit the size of the respective client display when an image is transmitted to a variety of client devices with different display sizes. The scheme has two key components; i) perceptual hint for each image block, and ii) an image transcoding algorithm. For a given semantically important block in an image, the perceptual hint provides the information on the minimum allowable spatial resolution reduction. The image transcoding algorithm that is basically a content adaptation process selects the best image representation to meet the client capabilities while delivering the largest content value. The content value is defined as a quantitative measure of the information on importance and spatial resolution for the transcoded version of an image.

Experimental results demonstrate a feasibility of the proposed scheme with a variety of client devices having different display sizes.

1. INTRODUCTION

With the advance of information technology, such as the popularity of the Internet, multimedia presentation proliferates into ever increasing kinds of media including wireless media. Multimedia data are accessed by ever increasing kinds of devices such as hand-held computers (HHC's), personal digital assistants (PDA's), and smart cellular phones. There is a need for accessing multimedia content in a universal fashion from a wide variety of devices [1].

Several approaches have been made to effectively enable such universal multimedia access (UMA). A data representation, the *InfoPyramid*, is a framework for aggregating the individual components of multimedia content with content descriptions, and methods and rules for handling the content and content descriptions [2]. The InfoPyramid describes content in different modalities, at different resolutions and at multiple abstractions. Then a transcoding tools dynamically selects the resolutions or modalities that best meet the client capabilities from the InfoPyramids. And J. R. Smith et al. proposed a notion of *importance* value for each of the regions of an image as a hint to reduce the overall data size in bits of the transcoded image [3,4,5]. The importance value describes the relative importance of the region/block in the image presentation compared with the other regions. This value ranges from 0 to 1, where 1 stands for the highest important region and 0 for lowest. For example, the regions of high importance are compressed with a lower compression factor than the remaining part of the image. Then, the other parts of the image are first blurred and compressed with a higher compression factor in order to reduce the overall data size of the compressed image.

When an image is transmitted to a variety of client devices with different display sizes, a *scaling mechanism*, such as format/resolution change, bit-wise data size reduction, and object dropping, is needed. More specifically, when an image is transmitted to a variety of client devices with different display sizes, a system should generate transcoded (e.g. scaled and cropped) image to fit the size of the respective client display. The extent of transcoding depends on the type of objects embodied in the image, such as cars, bridges, face, and so forth. Consider, for example, an image containing an embedded text or a human face. If the display size of a client device is smaller than the size of the image, the spatial resolution of the image must be reduced by sub-sampling and/or cropping to fit the client display. Users very often in such a case have difficulty in recognizing the text or human face due to the excessive resolution reduction. Although the importance value may be used to provide information on which part of the image can be cropped, it does not provide a quantified measure of perceptibility indicating the degree of allowable transcoding. For example, it does not provide the quantitative information on the allowable

This work was partially supported by grant No. (98-0102-04-01-3) from the Basic Research Program of the Korea Science & Engineering Foundation.

compression factor with which the important regions can be compressed while preserving the minimum fidelity that an author or a publisher intended. And the InfoPyramid does not either provide the quantitative information on how much the spatial resolution of the image can be reduced while ensuring that the user will perceive the transcoded image as the author or publisher initially wanted to represent it.

In this paper, we propose a novel scheme for transcoding an image to fit the size of the respective client display when an image is transmitted to a variety of client devices with different display sizes. We first introduce the notion of perceptual hint for each image block, and then present an image transcoding algorithm. The perceptual hint provides the information on the minimum allowable spatial resolution reduction for a given semantically important block in an image. The image transcoding algorithm selects the best image representation to meet the client capabilities while delivering the largest content value. The content value is defined as a quantitative measure of the information on importance and spatial resolution for the transcoded version of an image.

This paper is organized as follows: Section 2 introduces a novel notion of perceptual hint for each block in an image that provides information on the minimum perceptually allowable resolution. Section 3 describes an algorithm for image transcoding based on perceptual hint. Section 4 presents the experimental results and section 5 summarizes the paper.

2. PERCEPTUAL HINT FOR IMAGE TRANSCODING

2.1. Spatial Resolution Reduction value

A *Spatial Resolution Reduction (SRR)* value is determined by either the author or publisher and can also be updated after each user interaction. It specifies a scale factor for the maximum spatial resolution reduction of each semantically important block within an image. A block is defined as a spatial segment/region within image that often corresponds to the area of an image that depicts a semantic object such as car, bridge, face, and so forth. The SRR value represents the information on the minimum allowable spatial resolution, namely, width and height in pixels, of each block at which users can perceptually recognize according to the author's expectation. The SRR value for each block can be used as a threshold that determines whether the block is to be sub-sampled or dropped when the block is transcoded.

Consider the n number of blocks of users' interests within an image I_A. If we denote the ith block as B_i, $I_A = \{B_i\}$, $i=1, 2,\ldots,n$. Then, the SRR value $\boldsymbol{r_i}$ of B_i is modeled as follows:

$$r_i \equiv \frac{r_i^{\min}}{r_i^o},$$

where $r_i^{\min}$ is the minimum spatial resolution that human can percept and r_i^o is the original spatial resolution of B_i, respectively. For simplicity, the spatial resolution is defined as the length in pixels of either the width or height in a block.

The SRR value ranges from 0 to 1 where 0.5 indicates that the resolution can be reduced by half and 1 indicates the resolution cannot be reduced. For a 100x100 block whose SRR value is 0.7, for example, the author indicates that the resolution of the block could be reduced up to the size of 70x70 (thus, minimum allowable resolution) without degrading the perceptibility of users. The SRR value also provides a quantitative measure of how much the important blocks in an image can be compressed to reduce the overall data size of the compressed image while preserving the image fidelity that the author intended.

2.2. Transcoding Hint for Each Image Block

The SRR value can be best used with the importance value in [3,4,5]. Both SRR value (r_i) and importance value (s_i) are associated with each B_i. Thus, we have

$$I_A = \{B_i\} = \{(r_i, s_i)\}, \quad i = 1, 2,\ldots,n.$$

3. IMAGE TRANSCODING ALGORITHM BASED ON PERCEPTUAL HINT

3.1. Content Value Function V

Image transcoding can be viewed in a sense as adapting the content to meet resource constraints. Rakesh Mohan et al. modeled the content adaptation process as a resource allocation in a generalized rate-distortion framework [5,6,7]. This framework has been built on the Shannon's rate-distortion (R-D) theory [8] that determines the minimum bit-rate R needed to represent a source with desired distortion D, or alternately, given a bit-rate R, the distortion D in the compressed version of the source. They generalized the rate-distortion theory to a value-resource framework by considering different versions of a content item in an InfoPyramid as analogous to compressions, and different client resources as analogous to the bit-rates, respectively. But, this value-resource framework does not provide the quantitative information on the allowable factor with which blocks can be compressed while preserving the minimum fidelity that an author or a publisher intended. In other words, it does not provide the quantified measure of perceptibility indicating the degree

of allowable transcoding. For example, it is difficult to measure the loss of perceptibility when an image is transcoded to a set of a cropped and/or scaled ones.

To overcome this problem, we introduce an objective measure of fidelity that models the human perceptual system which is called a content value function V for any transcoding configuration $\boldsymbol{C}$:

$$C = \{ I, r \}, \tag{1}$$

where $I \subset \{1, 2, \ldots, n\}$ is a set of indices of blocks to be contained in the transcoded image and r is the spatial resolution reduction factor of the transcoded image. The content value function V can be defined as;

$$\begin{aligned} V &= V(I, r) \\ &= \sum_{i \in I} V_i(r) \\ &= \sum_{i \in I} \left(s_i \cdot u(r - r_i) \right), \end{aligned} \tag{2}$$

where

$$u(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{elsewhere} \end{cases}.$$

The above definition of V now provides a measure of fidelity that is applicable to the transcoding of an image at different resolution and different sub-image modalities. In other words, V defines the quantitative measure of how much the transcoded version of an image can have both importance and perceptual information. The V takes a value from 0 to 1, where 1 indicates that all of important blocks can be perceptible in the transcoded version of image and 0 indicates that none can be perceptible. The value function is assumed to have the following property.

Property 1: The value V is monotonically increasing in proportion to r and I. Thus we have

1. For a fixed I, $V(I, r_1) \leq V(I, r_2)$ if $r_1 < r_2$,
2. For a fixed r, $V(I_1, r) \leq V(I_2, r)$ if $I_1 \subset I_2$.

3.2. Content Adaptation Algorithm

Denoting the width and height of the client display size by W and H, respectively, the content adaptation is modeled as the following resource allocation problem:

$$\text{maximize}\left(V(I, r)\right) \quad \text{such that} \quad \begin{cases} r|x_u - x_l| \leq W \\ \quad \text{and} \\ r|y_u - y_l| \leq H \end{cases},$$

where the transcoded image is represented by a rectangle bounding box whose lower and upper bound points are (x_l, y_l) and (x_u, y_u), respectively.

Lemma 1: For any I, the maximum resolution factor is given by

$$r_{\max}^{I} = \min_{i,j \in I} r_{ij}, \tag{3}$$

where

$$r_{ij} \equiv \min\left(\frac{W}{|x_i - x_j|}, \frac{H}{|y_i - y_j|} \right). \tag{4}$$

The *Lemma 1* says that only those configurations $C=\{I, r\}$ with $r \leq r_{\max}^{I}$ are feasible. Combined with *property 1.1*, this implies that for a given I, the maximum value is attainable when $C=\{I, r_{\max}^{I}\}$. Therefore other feasible configurations $C=\{I, r\}$, $r < r_{\max}^{I}$ do not need to be searched.

At this moment, we have a naive algorithm for finding an optimal solution: For all possible $I \subset \{1, 2, \ldots, n\}$, Calculate $r_{\max}^{I}$ by (3) and again $V(I, r_{\max}^{I})$ by (2) to find an optimal configuration C_{opt}.

The algorithm can be realized by considering a graph

$$R = [r_{ij}], \quad 1 \leq i, j \leq n.$$

And noting that an I corresponds to a complete subgraph (clique) of R, and then $r_{\max}^{I}$ is the minimum edge or node value in I. Let us assume I be a clique of degree K ($K \geq 2$). It is easily shown that among the cliques, denoted by S, of I, there are at least 2^{K-2} cliques whose $r_{\max}^{S}$ is equal to $r_{\max}^{I}$, which, according to *Property 1.2*, need not be examined to find the maximum value of V. Therefore, only maximal clique will be searched.

Initially, r is set to $r_{\max}^{R}$ so that all of the blocks could be contained in the transcoded image. Then, r is increased discretely and for the given r, the maximal cliques are only examined. We maintain a minimum heap H to store and track maximal cliques with $r_{\max}$ as a sorting criterion.

Enqueue R into H.
WHILE H is not empty
  I is dequeued from H
  Calculate V(I, $r_{\max}^{I}$).
  Enqueue maximal cliques inducible from I after removing the critical (minimum) edge or node.
END_WHILE
Print optimal configuration that maximizes V.

Figure 1. Pseudo-code to find the optimal configuration.

Workstation	Color PC	TV	HHC	PDA
				 **(a)**
Content Value: 1.0	1.0	0	0	0
				 **(b)**
Content Value: 1.0	1.0	0.53	0.53	0.53

(a)Without using the SRR value (b)Using the SRR value

Figure 2. First example of image transcoding.

Workstation	Color PC	TV	HHC	PDA
				 **(a)**
Content Value: 1.0	1.0	0	0	0
				 **(b)**
Content Value: 1.0	1.0	1.0	0.78	0.44

(a)Without using the SRR value (b)Using the SRR value

Figure 3. Second example of image transcoding

4. EXPERIMENTAL RESULTS

We used test images with the size of 352 by 288 that were derived from one (news.mpg) of the contents contributed to MPEG-7. The images were transcoded to match several types of client devices of having the following display sizes; a workstation (256 x 208), color PC (192 x 156), TV browser (128 x 104), hand-held computer (HHC: 96 x 76), and personal digital assistant (64 x52) in the order of decreasing display sizes.

We demonstrate the feasibility of the proposed scheme by comparing our results (See Figures 2(b) and 3(b)) for a variety of client devices having a different display size with those obtained from resizing the entire image (See Figures 2(a) and 3(a)). In Figure 2, there are two important regions indicated by red boxes corresponding to a face (r_1=0.47, s_1=1.0) and text (r_2=0.44, s_2=0.9). In Figure 3, there are three important regions corresponding to face$_1$ (r_1=0.54, s_1=1.0), face$_2$ (r_2=0.6, s_2=0.75) and face$_3$ (r_3=0.45, s_3=0.5).

Figures 2(a) and 3(a) show that users can't recognize the transcoded images from TV to PDA, namely the content value is zero. But Figures 2(b) and 3(b) show that the perceptibility of a transcoded image is preserved while delivering the largest content value.

5. SUMMARY

We have proposed a novel scheme for transcoding an image to fit the size of the respective client display when an image is transmitted to a variety of client devices with different display sizes.

We first introduced the notion of perceptual hint for each image block, and then present an optimal image transcoding algorithm. Experimental results were shown to demonstrate the effectiveness of the proposed scheme.

6. REFERENCES

[1] J. R. Smith, R. Mohan, and C.-S. Li, "Transcoding Internet Content for Heterogeneous Client Devices," in *Proc. ISCAS,* Monterey, CA, 1998.

[2] C.-S. Li, R. Mohan, and J. R. Smith, "Multimedia content description in the InfoPyramid," in *Proc. IEEE Intern. Conf. on Acoustics, Speech and Signal Processing*, May 1998.

[3] J. R. Smith, R. Mohan, and C.-S. Li, "Content-based Transcoding of Images in the Internet," in *Proc. IEEE Intern. Conf. on Image Processing*, Oct. 1998.

[4] S. Paek and J. R. Smith, "Detecting Image Purpose in World-Wide Web Documents," in *Proc. SPIE/IS&T Photonics West, Document Recognition*, Jan. 1998.

[5] R. Mohan, J. R. Smith and C.-S. Li, "Adapting Multimedia Internet Content for Universal Access," *IEEE Trans. on Multimedia*, Vol. 1, No. 1, pp. 104-114, Mar. 1999.

[6] R. Mohan, J. R. Smith and C.-S. Li, "Multimedia Content Customization for Universal Access," in *Multimedia Storage and Archiving Systems.* Boston, MA:SPIE, Vol. 3527, Nov. 1998.

[7] R. Mohan, J. R. Smith and C.-S. Li, "Adapting Content to Client Resources in the Internet," in *Proc. IEEE Intern. Conf. on Multimedia Comp. and Systems ICMCS99*, Florence, Jun. 1999.

[8] C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J., vol. 27, pp. 379-423, 1948.