MMR: MASK BASED MULTI-RESOLUTION IMAGES AND VIDEOS

Siddhartha Chattopadhyay and Suchendra M. Bhandarkar
Department of Computer Science, The University of Georgia, Athens, Georgia 30602-7404, USA
siddh@cs.uga.edu suchi@cs.uga.edu

ABSTRACT

Multi-resolution quad-tree based image representations are useful for image and video encoding at varying bit rates. Existing algorithms use a difference measure of color values to trim branches of the quad-tree, which results in the various spatial regions of the image being represented at different resolutions. However, color difference-based quad-tree branch pruning often leads to sub-optimal multi-resolution images, where interesting features of the image are represented at low resolution, and/or uninteresting image regions are represented at high resolution. In this paper, we present a mask-based method for pruning a quad-tree representation of a multi-resolution image. The masks highlight various features of the image which require the best resolution, allowing tightly controlled pruning of the quad-tree branches. We suggest three main groups of masks, which are suitable for most types of applications requiring multi-resolution images and videos.

Index Terms— Image coding, Video coding, Quadtrees

1. INTRODUCTION

A multi-resolution (MR) image representation [1] is useful for image encoding at multiple bit-rates. MR video, being a sequence of images, uses MR images to encode video at variable bit rates [2]. MR images also have other applications in content-based image retrieval [3], medical imaging [4], storing and manipulating remotely sensed satellite data [5] among others. MR frames of an MR video yield better quality video compared to uniform resolution (UR) frames of the same encoded size, since each MR frame is encoded at multiple resolutions (within the same image), with *average* resolution comparable to that of the UR image, yet with better resulting image quality.

The standard algorithms used to create MR images exploit spatial redundancy within the image to group certain pixels with similar color values to form pixel blocks of uniform color. However, this approach is not sensitive to important spatial image features such as edges, object shapes, and other semantically or visually important objects such as human faces. As a result, MR often leads to sub-optimal image/video representations, where interesting regions of the image/video are represented at low resolution, and/or uninteresting regions of the image/video are represented at high resolution. In addition, since the definition of *interesting regions* of an image/video is subjective and application dependent, the generalized color difference-based schemes used by existing MR algorithms are not adequate to obtain MR representations of images/videos that are best suited for a specific application.

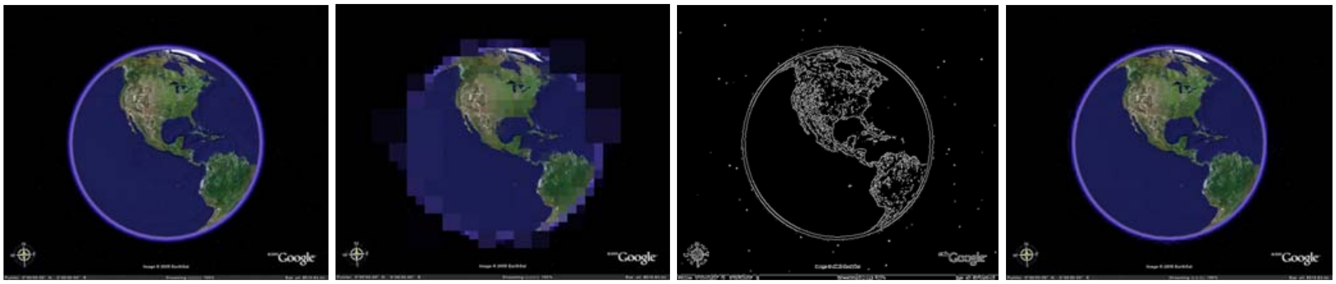
In this paper, a novel mask-based MR (MMR) image/video representation technique is proposed. Rather than strict color based segmentation, a mask is used instead to define regions of interest within the image, where the best resolution is desired. Such a mask, when used appropriately, can be used to encode the desired features of the image at the highest resolution, and encode the rest of the image at the lowest resolution. The introduction of such a mask results in three significant benefits; (a) the quality of the multi-resolution image is completely controlled by the mask; (b) the mask-based multi-resolution image encoding algorithm is parametric in the mask-parameter space, instead of the original MR image space. This is desirable, since the mask is visually more intuitive to control; and finally (c) since the image contents and quality are now completely controlled by the mask, many existing algorithms for low-level feature extraction and semantic feature detection can be used to create the mask.

2. QUAD-TREE FOR MR IMAGES

In this section, we give an overview of the methods used for creating MR images. Note that multi-resolution (MR) videos [2] use the same spatial encoding techniques used for multi-resolution (MR) images. Hence, the techniques for generation of MR images are automatically applicable to MR videos frames.

2.1. Creating quad-trees for multi-resolution images

Quad-trees are, by far, the most popular data structures to represent MR images [1] [6]. Quad-tree decomposition is a simple technique for representing images at multiple resolutions. In this technique, the image is recursively divided into four equal-size square regions depending on the contents of the blocks. For example, a 2^nx2^n image is represented as a tree of depth n. The root of the tree represents the original image at resolution level zero, and



(a) *(b)* *(c)* *(d)*

Figure 1. MR vs MMR for a google earth image. (a) The original image (b) Standard MR image; δ_{MR} = 76.1% (c) MMR *edge mask* (d) MMR image; δ_{MMR} = 78.8% i.e. δ_{MMR} and δ_{MR} are comparable; which means that the MMR image has approximately the same average resolution as that of MR image, but significantly better image quality.

the four equally-sized squares represent its children at resolution level one. Each node at every resolution level encodes its own color values (RGB); the parent node color is the mean of the colors of its child nodes. The details of the quad-tree decomposition algorithm can be found in [1] and [6].

2.2. Color difference-based quad-tree pruning

In most standard MR algorithms, at each node of the quad-tree corresponding to an image, a decision is made as to whether to decompose the corresponding block into four equal-size squares or to halt the decomposition process. In order to arrive at a decision, the most widely used measure is a difference measure of color values [6]. At each node, the value of the difference measure is compared with a *threshold value*; if the absolute difference is smaller than the *threshold value*, the recursive decomposition procedure at that node is halted. Otherwise, the node is further decomposed into four equal-size squares. The data structure for each node of the quad-tree can be represented by:

```
struct QuadNode
        ENUM type: NODE | LEAF
        QuadNode child[0, 1][0, 1]
        COLOR: RGBA
        REDUNDANCY: 0 or 1
end struct
```

If the node type is LEAF, then each child is NULL, and the COLOR is the color of the corresponding pixel. Otherwise, the node type is NODE, and each child is a pointer to the four equal sections of the block. The COLOR attribute in this case is the average color of the four children. The attribute REDUNDANCY, which is a binary number, is used to describe the degree of redundancy. REDUNDANCY = 1 in a node means that this node (and its children) can be eliminated, whereas REDUNDANCY = 0 means that this node (and consequently its children) cannot be eliminated at all.

3. USING MASKS FOR QUAD-TREE PRUNING

The quad-tree pruning mask is defined as follows:

Mask-Bin(i, j) = 1 if that pixel should not change at all
= 0 otherwise

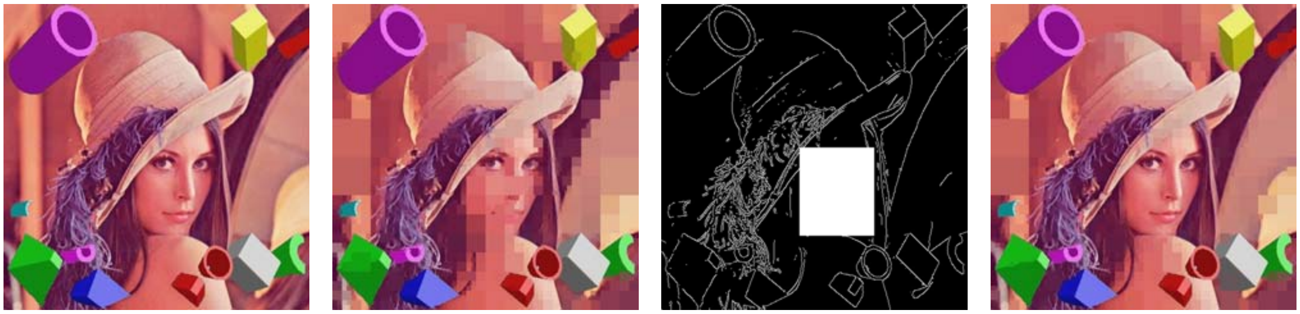
Assuming that the quad-tree and the desired mask, *Mask-Bin*, exist, the quad-tree branches can be pruned using the given mask by the following recursive pseudo-code:

```
function Mask_Prune(QuadNode p) returns 0 or 1
        if p. type = LEAF then
                i, j ← pixel corresponding to this leaf
                return (1 - Mask-Bin(i, j))
        end if
        redundancy = 1 // assume this node is redundant
        for each i, j = 0 and 1
                flag = Mask_Prune(p.child[i][j])
                if flag = 0 then redundancy = 0
        end for
        p.REDUNDANCY = redundancy
        return redundancy
end function
```

The root of the quad-tree is passed to the Mask_Prune function. The function Mask_Prune results in the best resolution (by preserving the original pixels) in the smallest-size image blocks corresponding to non-zero mask values. It also results in the maximum number of low resolution blocks in the other regions of the image. In the next section, we discuss some plausible masks for MMR images and MMR videos.

4. MASKS FOR MMR IMAGES AND VIDEOS

In this section, we present three broad categories of masks, which can be used for most applications requiring MR images and videos. Note that the advantage of mask based multi-resolution (MMR) image is that the average resolution of the image is much lower than that of the original image.



(a) *(b)* *(c)* *(d)*

Figure 2. MR vs MMR for the Lena image with embedded 3D objects. (a) The original image (b) MR image; δ_{MR} = 72.4% (c) MMR combination of *edge mask* and *face object mask* (d) MMR image; δ_{MMR} = 78.4%.

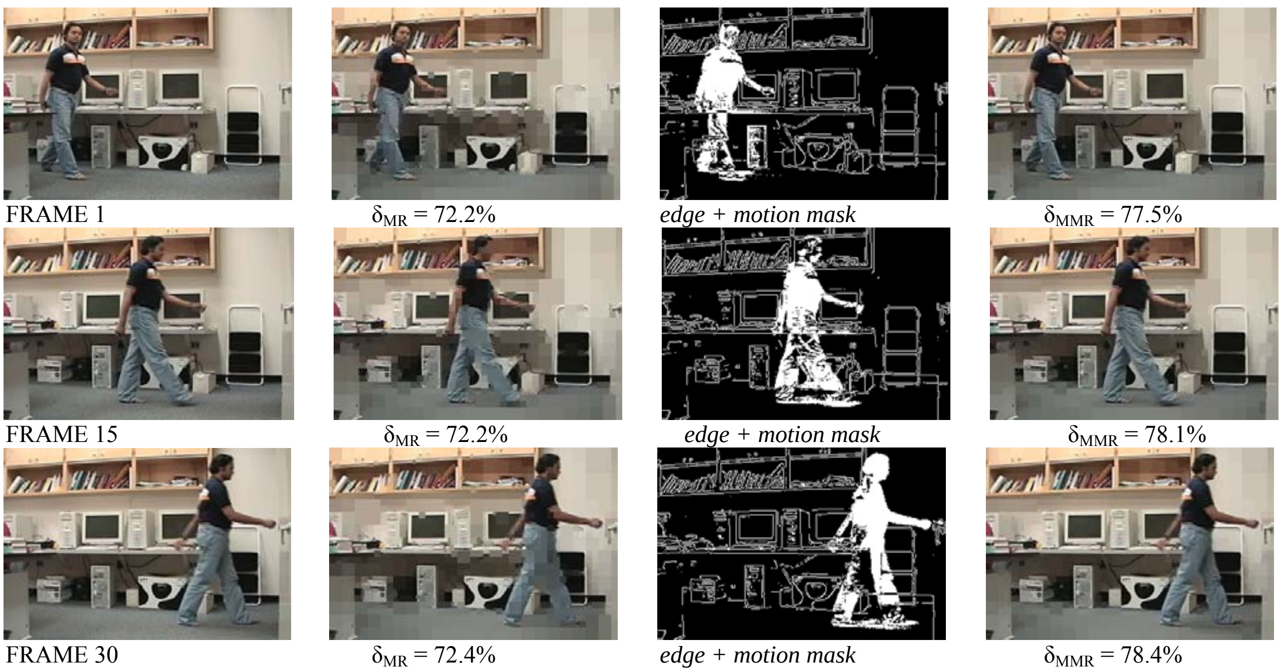


Figure 3. Three frames from a video sequence. (*Column 1*) The original frames (*Column 2*): Standard MR frames (*Column 3*): Mask for MMR: *edge + motion mask*; *motion mask* obtained by background separation (*Column 4*): MMR frames.

Hence, the MMR image when encoded using a standard encoding technique for quad-tree based image representation, results in much less data compared to the original image, thus proving to be an enhanced image/video compression method.

4.1. Feature masks

Features masks are useful in preserving low level spatial image features such as edges, boundaries, texture etc. For example, edges can be preserved in MMR using *edge masks*. An interesting application of *edge masks* is in multi-resolution streaming video applications, such as *Google-Earth*, where the basic edge features are maintained while zooming in on a country or region. We used an *edge mask* on a *Google-Earth* image (**Figure 1**), obtained by using the standard Canny edge detector [7]. The value of the Canny edge detector parameter, σ, is obtained semi-empirically; a detailed discussion on this issue is beyond the scope of this paper. Note that the MMR image quality is significantly superior compared to the MR image quality. Detailed quantitative analysis on the comparison of MMR and MR image qualities is given in **Section 5**.

4.2. Object masks

Object masks exploit the basic human psychology of focusing on semantically important objects while ignoring other artifacts in an image/video. For example, in a video requiring attention to be drawn to human faces, a *face mask* can be used for each frame in the video stream. A *face mask* can be obtained using a face-detection algorithm [8]. Other *object masks*, such as *shape masks*, emphasize certain shapes in an image, which can be useful for robotic applications. Such a mask may be obtained using pattern recognition algorithms [9]. **Figure 2** illustrates a combination of an *edge mask* and a *face mask*. Note that *visually interesting regions*, such as the face, hair, and the shapes of the 3D objects, are preserved at their original resolution, whereas other areas are displayed at the lowest local resolution possible.

4.3. Motion Masks

Motion masks can be used to render temporally changing objects in a video at higher resolution compared to the rest of the video. An example, using a combination of an *edge mask* and a *motion mask*, obtained by simple background subtraction, is illustrated in **Figure 3**. The MMR frames show a significant improvement in quality of the *interesting video regions* such as the moving object and background edges; these specific enhancements are absent in the general MR frames.

5. MMR VS MR

A comparison between MMR and MR involves two fundamental attributes; the visual quality obtained and the size in bytes of the encoded image/video. The image/video quality resulting from both MMR-based and MR-based rendering can only be assessed subjectively by the user, based on the requirements of a particular application. To have a numerical estimation of the image/video size obtained using MR or MMR, we have used a percentage metric δ such that:

$$\delta = 100*n / N \qquad (1)$$

where n is the number of pixels in the multi-resolution image, obtained by treating each block as a single pixel, and N is the total number of pixels in the original image. To ensure a fair comparison between MR and MMR, we compute δ_{MR} and δ_{MMR} for MR and MMR respectively, and report the resulting image quality where δ_{MR} is comparable to δ_{MMR}. **Figures 1**, **Figure 2** and **Figure 3** illustrate the difference in the image/video quality obtained by using MR and MMR, where δ_{MR} and δ_{MMR} are comparable. The MMR images/videos retain all the desired visually and semantically meaningful features of the images/video-frames. The MR images/videos, on the other hand, are not subjected to the same quality control, resulting in visual distortion in some critical regions.

6. CONCLUSIONS AND FUTURE WORK

A mask-based multi-resolution (MMR) image/video representation technique has been discussed. The MMR representation uses a mask to control the quad-tree pruning for generation of multi-resolution (MR) quad-tree image/videos. The masks are defined such that a value of 1 indicates that the pixel is totally redundant (i.e. pixel color can be changed at will) whereas a value of 0 indicates that the pixel has to remain unaltered in the image. Three groups of masks i.e. *feature masks*, *object masks* and *motion masks* have been suggested. Examples and results have shown the considerable advantage of using MMR compared to MR in terms of image/video quality control.

Using masks provides enormous opportunity for future work. Various masks suitable for different types of images/videos can be proposed. Application-specific masks can be developed in future to aid in compression, representation and querying of images and videos.

7. REFERENCES

[1] Wilson, R., Todd, M. and Calway, A. D., "Generalised quadtrees: a unified approach to multiresolution image analysis and coding", *SPIE Visual Communication and Image Processing*, pages 1-2, May 1990.

[2] Finkelstein, A., Jacobs, C. E., and Salesin, D. H., "Multiresolution video", *In ACM Computer Graphics Proceedings, SIGGRAPH'96*, Annual Conference Series, pages 281-290, 1996.

[3] Jacobs, C. E., Finkelstein, A. and Salesin, D. H., "Fast multiresolution image querying", *Proceedings of SIGGRAPH'95,* pages 277-286. ACM, 1995.

[4] Liu, S., Babbs, C. F., and Delp, E. J., "Multiresolution detection of spiculated lesions in digital mammograms," *IEEE Transactions on Image Processing*, Vol. 10, No. 6, pp. 874-884. June 2001.

[5] Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., and Heynen, M., "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information", *ISPRS Journal of Photogrammetry and Remote Sensing,* Volume 58, Issues 3-4, Pages 239-258. 2004.

[6] Samet H., *Application of Spatial Data Structures: Computer Graphics, Image Processing, and GIS*, Addison Wesley, Reading, MA, 1990.

[7] Canny, J., "A computational approach to edge detection" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679--698, 1986.

[8] Rowley, H., Baluja, S., and Kanade, T., "Neural Network-Based Face Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, number 1, pp. 23-38, Jan 1998.

[9] Jain, A.K., Duin R.P.W., and Mao, J., "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4-37, 2000.