COMPRESSION WITH MOSAIC PREDICTION FOR IMAGE-BASED RENDERING APPLICATIONS

Wing Ho Leung and Tsuhan Chen

Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213
Email: {wingho,tsuhan}@andrew.cmu.edu

ABSTRACT

In this paper, we present a mosaic based compression scheme for image-based rendering applications in which a sequence of images is captured by a camera located at different positions along a circle. The mosaic image is constructed first and then used to predictively encode the original images. Furthermore, motion compensation is applied to provide a closer match at the block level between the prediction image and the original image. We examine this compression scheme and compare its performance with intra and inter coding using a general video codec. Our experimental result shows that mosaic based compression with motion compensation provides a better performance in the rate-distortion sense compared with intra coding while at the same time it has the advantage of random access over inter coding.

1. INTRODUCTION

In image-based rendering applications, light rays from different directions are captured and then resampled in order to generate different views of a scene. In terms of sampling the light rays, Shum and He proposed the idea of concentric mosaics [1] which uses a 3D plenoptic function [2] to represent the light rays. Concentric mosaics are more flexible than 2D panorama [5] because it allows the viewpoint to be moved away from the center. Besides, concentric mosaics can be considered as special cases of Lumigraph [3] or Lightfield [4]. As a result, our current compression work that is being focused on concentric mosaics can easily be extended for Lumigraph or Lightfield.

In image-based rendering, often the number of acquired images is very big, therefore it is necessary to compress the images in order to store them efficiently. In this paper, we propose to apply mosaic based compression on these acquired images. Mosaic based compression has been proposed for video compression [6]. A similar concept also exists in MPEG-4 where global motion compensation is performed based on a static sprite describing a panoramic background [7]. We extend these concepts by utilizing the knowledge that the images are taken from a camera with a known path in image-based rendering applications. Under this scheme, a mosaic is first constructed by merging all images and then used to predict each individual image. The residue between the original image and the prediction is coded using a codec that is similar to a general video codec but with modification to the motion estimation and the motion compensation modules.

This paper is organized as follows. Section 2 describes how the images are acquired. Section 3 describes intra and inter coding using a general video codec. Section 4 describes the mosaic based compression scheme. Our experiments are described and the results are analyzed in Section 5. Finally, we provide a conclusion and outline future work in Section 6.

Work supported in part by IBM University Partnership Program.

2. IMAGE ACQUISITION

To capture images of a real scene, an off-centered camera is rotated along a circle [1] and the images are acquired when the camera is at different positions. Often the camera is facing outward with setup 1 as shown in Figure 1 (a). Once these images are acquired, an image viewed by a virtual camera can be synthesized by taking pixels of the corresponding images that are already acquired as shown in Figure 1 (a). We propose to place the camera in setup 2 as shown in Figure 1 (b). This is because with setup 2 a larger field of view of the scene can be obtained and there will not be any missing pixels when a virtual view is constructed at allowable viewpoint inside the circle. Figure 2 (a) shows some example images taken from a synthetic scene.

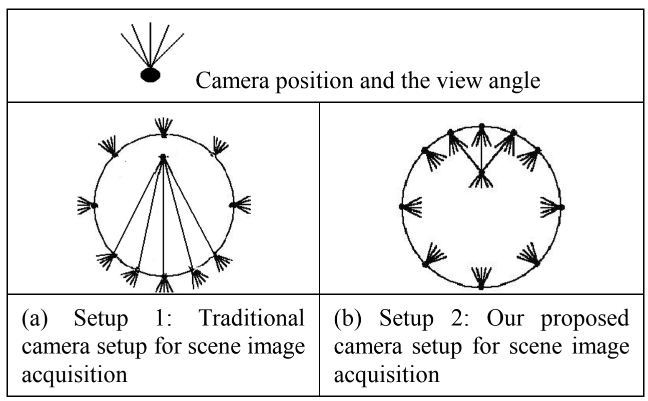

Figure 1 Comparison between traditional camera setup and our proposed camera setup

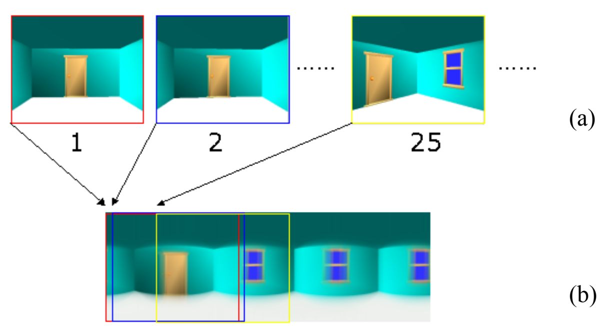

Figure 2 Examples of IBR images and the generation of the mosaic

3. GENERAL VIDEO CODEC

Figure 3 shows the block diagram of a general video codec. Each image can be coded either in the intra or inter mode. In the intra mode, each switch is connected to the top node and the feedback loop is not used. Discrete Cosine Transform (DCT) and quantization (Q) are performed on the current image directly and the result is passed to an entropy encoder. The decoder reverses the process with inverse quantization (IQ) followed by inverse DCT (IDCT) to reconstruct the images. In the inter mode, each switch is connected to the bottom node. The prediction is made by motion estimation (ME) which is the process of finding a motion vector (MV) that gives a good match between each block on the current image and a region in the previously decoded image. Motion compensation (MC) is then performed to use the best-match region as the prediction. The prediction is subtracted from the original block in the current image to produce the residue. DCT and quantization are performed on this residue and then the result is passed to the entropy encoder. On the decoder side, the process is reversed except that motion estimation is not required.

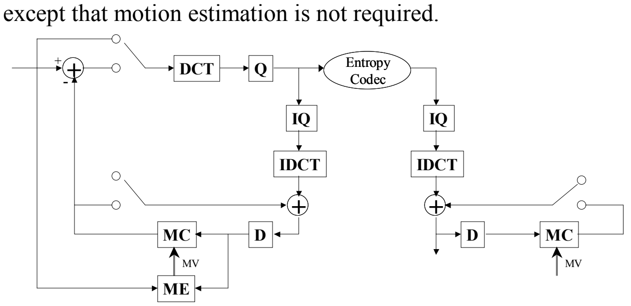

Figure 3 Block diagram of a general video codec

In our application, the acquired images have high correlation. As a result, intra coding does not yield optimal compression. On the other hand, inter coding takes advantage of exploiting the correlation between successive images by using the previously decoded image to predict the current image, thus the result yields a higher compression ratio. However, the problem with inter coding is that the current image depends on the previously decoded image. Often in image-based rendering applications, the images to be decoded are not in a fixed sequence. For example, consider the two cases where the new views are generated by rotating a virtual camera clockwise in one case and counter-clockwise in another case. These two cases represent two different sequences of the same set of images. Decoding will be very inefficient if we use inter coding because of the lack of random access to each individual image.

In order to maintain the random access nature, yet at the same time to result in a good compression ratio by exploiting the correlation between images, we will make use of mosaic based compression scheme that is described in the next section.

4. MOSAIC BASED COMPRESSION

4.1 Generation of mosaic

The first step in mosaic based compression is the generation of a mosaic. The images are assumed to align vertically and successive images are shifted horizontally until the mean absolute difference (MAD) between every two successive images, including between the last image and the first image, in the overlapping area is minimum given the constraint that the amount of overlap between successive images cannot be too small to avoid the trivial solutions. We call these horizontal shifts as *offsets* and the process of finding the offsets is shown in Figure 4 (a). The mosaic image is then composed by averaging the overlapped images using the offsets as shown in Figure 4 (b). It can be observed that the generated mosaic is a circularly wrapped around image whose size is data dependent. Figure 2 (b) shows an example of a mosaic generated in this manner.

The constructed mosaic image is similar to a 2D panorama except that the images are captured by positioning the camera along a circle instead of fixing the camera at the center. Therefore, the mosaic image is blurred due to the averaging effect. On the other hand, we make use of the fact that the camera position is along a circle.

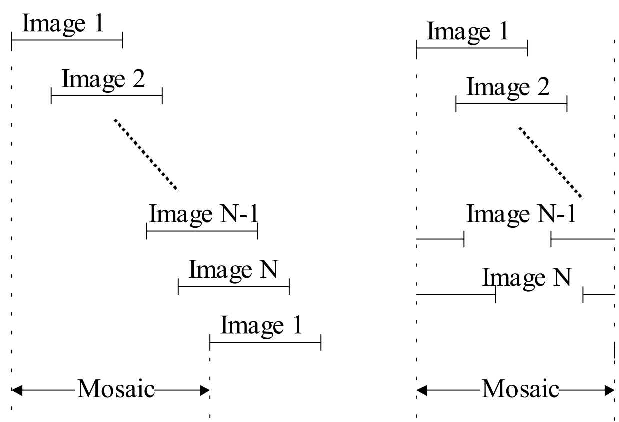

Figure 4 (Left) Finding the offsets. (Right) Constructing the mosaic image by averaging

4.2 Mosaic based compression without motion compensation

The codec for mosaic based compression can be obtained by modifying the general video codec described in Section 3. The general video codec is modified such that a mosaic image is generated first and then the corresponding part of the mosaic image is taken as the prediction image. The decoder receives the mosaic image at the beginning therefore the mosaic image is present on both the encoder and decoder, and there is no dependency on the previously decoded image. The corresponding part of the mosaic image is selected as the prediction image according to the offset. Figure 5 shows the block diagram of the modified scheme.

4.3 Mosaic based compression with motion compensation

In order to have a better match in the block level from the prediction, motion compensation can be performed on the prediction image from the mosaic. The motion vectors are required in this case. Figure 6 shows the block diagram of this scheme.

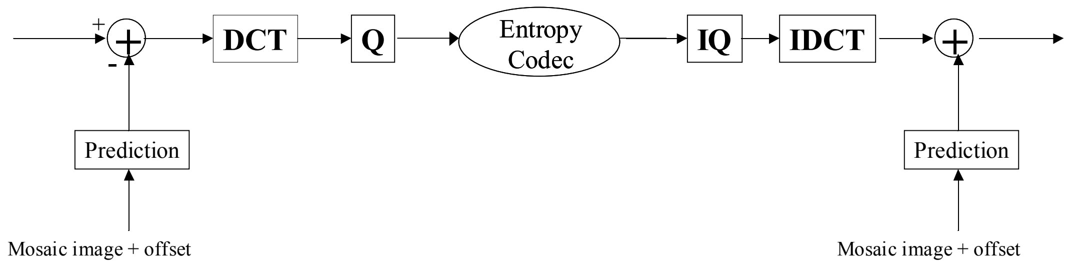

Figure 5 Block diagram of the modified compression scheme using prediction from mosaic without motion compensation

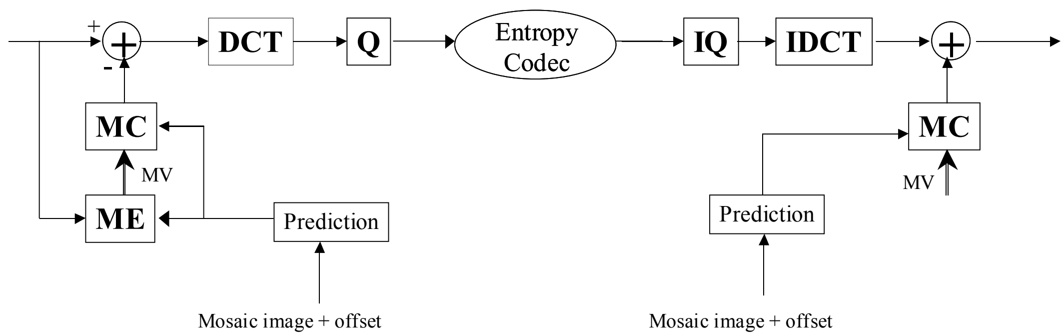

Figure 6 Block diagram of the modified compression scheme using prediction from mosaic with motion compensation

5. EXPERIMENT

A total of 200 images of the synthetic scene of a virtual room used in Networked Intelligent Collaborative Environment (NetICE) [8] is obtained by positioning the camera along a circle. The images are stored in YUV 4:2:0 CIF format (352×288 pixels per frame). We conduct experiments to compare methods proposed in previous sections.

The first method is intra coding and the second method is inter coding. For these two methods, we make use of the H.263 codec we developed [9]. The third and fourth methods use the mosaic based compression schemes. The resulting mosaic image has 857×288 pixels. The mosaic image is chopped into a sequence of images in the CIF format. The remaining columns in the last frame are repeated with the last column of the mosaic image for padding. The resulting mosaic image sequence is compressed using the H.263 codec. The offsets are losslessly compressed using LZW. Then the prediction image is obtained by taking the corresponding part of the decoded mosaic image according to the offset. The residue between the original image and the prediction image is coded. The difference between the third and the fourth methods is that the third method does not include motion compensation but the fourth method does. We examine the results in terms of average mean reduced mean absolute energy (mrMAE) under different coding schemes. Small values of average mrMAE are equivalent to better results because they often result in smaller bit rates. This measure of mrMAE is defined in the following:

$$\text{mrMAE} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|z(i,j) - \bar{z}\right|$$

with $z(i,j) = x(i,j) - y(i,j)$ and $\bar{z} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[x(i,j) - y(i,j)\right]$

where $N = 16$

$x(i, j)$ is the intensity at the (i, j) position of the original macroblock

$y(i, j)$ is the intensity at the (i, j) position of the prediction macroblock

Note : when Intra coding is used, $y(i, j)$ is set to 0 for all i and j.

The values of average mrMAE under different schemes are shown in Table 1:

	Average mrMAE
Intra coding	8.52
Inter coding	1.17
Mosaic wthout MC	8.67
Mosaic with MC	5.97

Table 1 Average mrMAE under different schemes

It is found that in terms of average mrMAE, mosaic based compression with motion compensation is better than all other schemes except inter coding.

In addition, rate-distortion curves are generated by varying the DCT quantization levels. Bit-streams are generated under each scheme and the Peak Signal-to-Noise Ratio (PSNR) of the luminance is computed. The rate is measured by the average number of bits used to represent one pixel (bpp). Under the mosaic based compression schemes with and without motion compensation, the bit count includes both the bit count of the reconstructed sequence, the bit count of the compressed mosaic image, and the bit count of offsets.

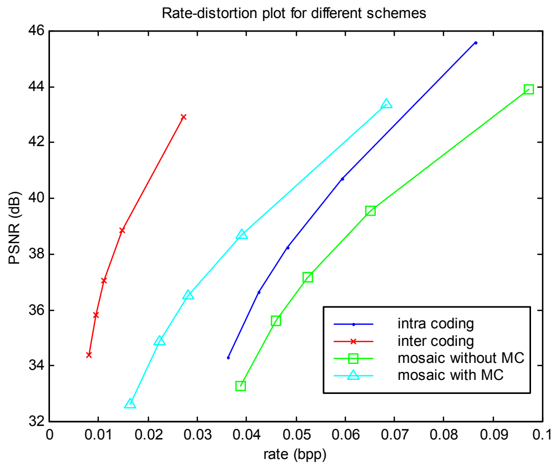

Figure 7 Rate-distortion plot for different schemes

The resulting rate-distortion curves are show in Figure 7. It can be seen that for the same PSNR value, inter coding results in the lowest bpp. Mosaic based compression with motion compensation requires smaller bpp than mosaic based compression without motion compensation. This shows that by using mosaic based compression with motion compensation, it is possible to find a region in the mosaic prediction image closely matched to a block in the original image. On the other hand, mosaic based compression without motion compensation performs worse than intra coding. This may be due to the fact that there is too much averaging effect on the mosaic image which makes the corresponding block of the mosaic prediction image quite different from the original image. The rate-distortion curves in Figure 7 are consistent with the values of average mrMAE shown in Table 1.

Although the results for inter coding are better than all the other three methods, both in terms of average mrMAE and in rate-distortion sense, we still propose to use mosaic based compression with motion compensation because it allows random access of individual images as described in Section 4.2.

6. CONCLUSION AND FUTURE WORK

In this paper, we have presented how mosaic based compression can be used in image-based rendering applications to compress a sequence of images captured by a camera located at different positions along a circle. It is concluded that the mosaic based compression with motion compensation provides a better compression ratio than intra coding while it has the advantage of random access of each image over inter coding.

We are working on taking advantage of the low bit rate of the inter coding by modifying it to allow random access of each individual image. This can be done by picking some images spreading evenly to be intra coded and then inter coding the remaining images. Another possible research issue is to examine other ways of constructing of the mosaic image, for instance, we can try to construct the mosaic image by using several vertical lines of each acquired image and pasted them together to form a mosaic, which is equivalent to forming the mosaic image by using simple 2D panorama.

ACKNOWLEGDEMENTS

Special thanks to Ta-Chien Lin, Carnegie Mellon University and Institute for Information Industry, for his help in modifying the motion estimation and motion compensation modules of the H.263 codec to implement the mosaic based compression schemes.

REFERENCES

[1] Shum, H-Y. and He L-W. "Rendering with Concentric Mosaic", *Computer Graphics Proceedings, Annual Conference Series, Proc. SIGGRAPH'99,* pages 299-306, Los Angeles, CA, August 1999.

[2] Adelson E. H. and Bergen J. "The plenoptic function and the early vision". *Computational Models of Visual Processing,* pages 3-20, MIT Press, Cambridge, MA, 1991.

[3] Gortler S. J., Grzeszczuk R., Szeliski R., and Cohen M. F.. "The lumigraph". *Computer Graphics Proceedings, Annual Conference Sereies*, *Proc. SIGGRAPH'96*, pages 43-54, New Orleans, August 1996.

[4] Levoy M. and Hanrahan P. "Light field rendering". *Computer Graphics Proceedings, Annual conference Series*, *Proc. SIGGRAPH'96*, pages 31-42, New Orleans, August 1996.

[5] Chen S.E. "QuickTime VR – an image-based approach to virtual environment navigation". *Computer Graphics (SIGGRAPH'95)*, pages 29-38, August 1995.

[6] Irani M., Anandan, P., Bergen, J., Kumar, R., and Hsu, S. "Efficient Representations of Video Sequences and Their Applications", *Signal Processing: Image Communication*, Vol. 8, No. 4, pages 327-351, May 1996.

[7] MPEG-4, Overview of the MPEG-4 Standard, ISO/IEC JTC 1/SC 29/WG 11, Melbourne, Australia, October 1999.

[8] Leung W. H. and Chen T. "Networked Collaborative Environment with animated 3D avatar", *IEEE Workshop on Multimedia Signal Processing*, Los Angeles, CA, December 1998.

[9] Advanced Multimedia Processing Lab, Carnegie Mellon University, CMU H.263+ Video Codec, http://amp.ece.cmu.edu/