COLOUR VOLUMETRIC COMPRESSION FOR REALISTIC VIEW SYNTHESIS APPLICATIONS

N. Anantrasirichai, C. Nishan Canagarajah, David W. Redmill and David R. Bull

Department of Electrical & Electronic Engineering, University of Bristol, Bristol, BS8 1UB, UK

ABSTRACT

The colour volumetric data which is constructed from a set of multi-view images is capable of providing realistic immersive experience. However it is not widely applicable due to its manifold increase in bandwidth. This paper presents a novel framework to achieve scalable volumetric compression. Based on wavelet transformation, data rearrangement algorithm is proposed to compact volumetric data leading to high efficiency of transformation. The colour data is also rearranged by using the characteristics of human eve sensitivity. Moreover, the pre-processing for adaptive resolution is proposed in this paper. The low resolution overcomes the limitation of the data transmission at low bit rate, whilst the fine resolution improves the synthesised images' quality. The results of our proposed schemes show significant improvement of the compression performance over the traditional 3D coding.

Index Terms— Multi-view Image, Volumetric representation, View synthesis, Wavelet transforms

1. INTRODUCTION

Recently three-dimensional (3D) visual environments have become one of the emerging problems in communication and computer vision. Multiple cameras are used for obtaining object or scene information from various points of view. The large amount of data has limited many commercial applications, due to its manifold increase in bandwidth over existing monoscopic image system. A *volumetric representation* is a good solution because of its effectiveness in compression ratio and modelling accuracy. However, efficient compression algorithms are vital to reduce the size of volumetric data without sacrificing the image quality.

The compression of the reconstructed volume can be achieved efficiently by exploiting 3D coding and/or geometry coding schemes. The 3D wavelet algorithms for volumetric compression have been developed with high efficiency for both lossless and lossy coding. However, most of these algorithms have been proposed for medical applications of which details are 2D cross sections of the volumes, rather than view synthesis [1][2][3]. Direct coding of 3D voxel models has been proposed and developed by using the idea of video coding in which each third component is operated as a frame of the sequence [4]. For view synthesis applications, this scheme is unsuccessfully applied at low bitrate. Because of losing some model information, the synthesized image contains artefacts and lack of details. Therefore the scheme is designed to encode the model and colours of the volume separately. The model is losslessly coded so that the correct geometry is obtained even at very low bitrate, whilst the colour information is possibly truncated to achieve the target bitrate. This approach also increases the degrees of freedom for designing the algorithm in order to enhance compression efficiency.

In this paper, we present the novel volumetric compression which is suitable for realistic view synthesis. Firstly, the volume with adaptive resolution is constructed. Subsequently, the colour data is prepared to improve the coding performance. Then, it is transformed by wavelet lifting scheme and encoded by the JPEG2000 Part-2 standard which supports multiple compositing layers [5].

Apart from medical applications, the inside of the object is generally occluded and is not essential for view synthesis. Existing methods introduced for compressing volumetric data include this useless information in the coding process causing inefficiency in data transmission and data storage. To achieve the highest compression performance, two data rearrangement schemes are introduced in this paper; one uses the characteristics of human eye sensitivity to diminish the size of chrominance information and the other prepares the dense data for 3D wavelet transformation. Moreover, we propose the adaptive resolution approach for the volume representation which provides scalabilities and increases view synthesis performance. By this approach, the receiver can use only the significant part instead of the entire data stream. Consequently, full-resolution images can be constructed from the half-resolution volume with acceptable quality. In the case that the receiver can support bigger data size or it has a bigger buffer; the double-resolution volume is used for constructing images with better quality.

The rest of paper is organized as follows: Section 2 explains the proposed scheme for the volumetric representation with adaptive resolution. The proposed codec is described in Section 3. The experimental results and discussions are presented in Section 4 in which the performance of the proposed codec is compared to that of the conventional codec. The conclusions and future works are presented in Section 5.

2. PRE-PROCESSING FOR ADAPTIVE RESOLUTION

The construction of volume which can be used in three resolutions, namely half, full and double resolutions, is proposed in this section. The final result is the double-resolution volume which is capable of providing good quality when the volume is used as the full or half resolution. In order to give priority to original image size, the full-resolution volume is first constructed by using the algorithm proposed in [6]. The existing voxels which are parts of the object contain intensity Y and colours U and V, whilst the voxels which are outside the object or invisible are transparent. Then, the proper voxels are inserted so as to complete the volume when it is used at smaller or larger scale.

2.1 Pre-processing for Half Resolution

For lower resolution, the subsampling causes significant information loss; therefore, a pre-processing step is introduced. By subsampling process, the voxels at $(2\hat{i}-1, 2\hat{j}-1, 2\hat{k}-1)$, where $1 \le 2\hat{i}-1 \le h$, $1 \le 2\hat{j}-1 \le w$, $1 \le 2\hat{k}-1 \le z$, and $\hat{i}, \hat{j}, \hat{k} \in \{1, 2, 3, ...\}$ are included in the half-resolution volume. Subsequently, the positions containing the transparent voxels are checked. If one of eight adjacent voxels of the full-resolution volume is not transparent, the transparent voxel at $(2\hat{i}-1, 2\hat{j}-1, 2\hat{k}-1)$ is replaced by the median of the existing voxels of these eight adjacent positions. Next, the new voxels which are located outside the volume are removed to ensure that the new inserted voxels do not occlude or affect any point of the original volume at the full resolution. The pre-processed data contains more voxels, hence the holes in synthesized views are disappeared and the image quality is improved.

When the decoder receives the first part of bitstream, the half-resolution model is reconstructed, and then the colour volume is rebuilt. The reconstructed images or the virtual views with the half size of the original images can be generated by directly employing this half-resolution model and colour volume. However, to create a full-resolution image, the 3D interpolation process is required. The full-resolution reconstructed image from the half-resolution volume is shown in Fig. 1.

Comparing to the wavelet transformation, the subsampling from the prepared volume produces better shape contour at small scale. This is because filtering the binary data of the model possibly generates the floating-point data which is hard to define whether it is the existing or transparent voxel. Moreover, high different values between the transparent voxels and the existing voxels cause high frequencies. Therefore, details are significantly lost if only low-pass signal of transformed volume is used. As a result the synthesised views at the low bitrate possibly have poor perceptual quality.



Fig. 1 (a) Original image. (b)-(c) The reconstructed image at full resolution by interpolating from the half-resolution volume; (b) without preparation process (PSNR = 28.86 dB) and (c) with preparation process (PSNR = 32.04 dB).

2.2 Pre-processing for Double Resolution

To increase the resolution, the voxel at (i, j) of the fullresolution volume is mapped to the location at (2i-1, 2j-1) of the double-resolution volume. The other voxels are inspected whether they are parts of volume or transparent by operating the same construction approach of the full-resolution volume. That is, the depth maps of the references are first registered into the double-resolution volume. Then the constructed volume is iteratively refined by utilizing the brightness information of the reference views. Finally, the colour of each voxel is properly selected in order to render a smooth and realistic image.

The finer-resolution volume improves the view synthesis performance, because it provides more details and more accuracy of voxel positions. Basically, the imaging system is operated as discrete signals, so the computed positions are adjusted to the nearest integers. Some non-corresponding pixels from the different reference views which have different spatial resolutions are possibly mapped to the same voxel. Consequently, the synthesized views might obtain the absolutely wrong values of brightness and colours leading to poor subjective quality if the resolution of the representing volume is not fine enough.

3. PROPOSED CODING SCHEME

After obtaining the 3D volume from a set of multi-view sequences, the model and texture are encoded to reduce the data size with the highest perceptual quality. The proposed coding scheme is shown in Fig. 2.

The depth or the position of each voxel in a 3D volume is vital for image warping and view synthesis. Hence, the model is losslessly coded. The luminance and chrominance data are reorganised so that the compression method can attain the highest efficiency. The lossy coding is then applied to achieve target bits.

To support very low bitrate applications, firstly the half resolution volume is sent, followed by the additional data in the full resolution excluded from the preceded transmission. Finally, the remaining data used for composing the doubleresolution volume is included in data stream. Noticeably, this scheme provides scalability. The camera parameters are separately coded by Huffman coding, and stored or transmitted as the overhead of the data stream. The background of the same geometry of the volume is separately coded from the volumetric data and selective for transmission. This gains the benefit of bit allocation for region of interest coding. The background can be coded with a very small size and a different background can be used depending on the users.



Fig. 2 Coding diagram of the volumetric compression (m_x =coded model, T_x =coded texture, $x \in \{h, f, d\}$, h=half resolution, f=added to full resolution, d=added to double resolution)

3.1. Data Rearrangement

Two data rearrangement approaches are proposed, namely chroma subsampling and component reduction. The first approach reduces the resolution of chrominance and the other approach prepares the dense data which is suitable for 3D wavelet transformation.

3.1.1 Chroma Subsampling

Due to the fact that the human's eyes are less sensitive to colour than brightness, the chrominance of an image can be defined with lower resolution than the luminance. However, for the volume, the intensity and colours are defined only on the existing voxels. Direct subsampling for the volume then causes significant information loss. To preserve good colour information at lower resolution, one colour value for eight adjacent corners is defined. It is taken as a transparency if no existing voxel presents in any of the eight corners, or the median value of the existing voxels if any of them are presented.

In the decoding process, the missing values are interpolated from neighbouring voxels or repeated from the preceding sample. Since viewers' sense of colour is barely affected, the repeating approach is employed in our simulation.

3.1.2 Component Reduction

The general volume is formed with the surface of the object. Compared to a solid volume where the inside voxels are filled in order to reduce the high frequencies after transformation, the number of voxels along the surface is much smaller than those of the whole object. Therefore, instead of coding a solid volume with original shape, the volume is deformed to reduce the number of components by shifting the voxels forward to fill the empty component in each width. The deformation approach is applicable since the model is coded independently. Using this approach, not only is the number of components reduced, but the holes are also disappeared leading to a decrease in high frequencies after transformation. The proposed algorithm is simply explained by pictures as illustrated in Fig. 3.



Fig. 3 The example of a cross section of a volume showing data rearrangement idea. (a) the original volume with 8 components (b) rearranged volume with 4 components (c) volume filled all transparent voxels with the medium value.

In the decoding process, the model is used for identifying the position of each voxel. By reversing the chroma subsampling and component reduction processes, the original position of the shrunk luminance and chrominance data are retrieved.

3.2 Wavelet Coding

A 3D volume is an image where one sample is composed of multiple components. This characteristic meets the definition of the JPEG2000 Part-2 standard. Thus the multi-component transformation can be employed. The wavelet lifting scheme is applied to each pixel in the image, as well as that in the third direction.

In the proposed scheme, the new compact volumes are transformed to wavelet coefficients. Along the component axis, the Haar and 5/3 wavelet lifting schemes are applied to remove correlations among consecutive components of the luminance Y and chrominance U and V, respectively. Basically, the longer filters achieve the higher compression efficiency but, in the case that the correlations among the successive components are not high enough and/or the total number of components is too small, the long filter will generate high frequencies and cause the ghosting artefacts. A short filter, such as Haar, is therefore selected for the luminance Y rather than the 5/3 wavelet which is used for transforming the chrominance U and V. After applying the multi-component transformation with particular number of levels, each frame is compressed using the JPEG2000 Part-1 standard.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

The proposed scheme is tested with the volume of Leo^{t} sequence. This multiview sequence composes of five views with different camera parameters that cause different spatial

¹ The multiview *Leo* sequence was captured at University of Bristol

resolutions.

4.1 Compression Performance

In this section, the qualities of five reconstructed views were compared to the original views by calculating the average PSNR. The background of each view was regenerated by using the estimation depth map from dynamic programming proposed in [7]. For this section, the background was not encoded, so all bitrates shown in graphs comprise only the information and details used for reconstructing the volume, i.e. binary model, brightness and colour of texture surface and camera/geometry parameters. Note that the coding performance depends on the accuracy of the available geometry model [8]. The lossless quality implies the best quality of the coding results.

As illustrated in Fig. 4, the maximum quality of the halfresolution volume is approximately 1.1 dB less than that of the full-resolution volume, whilst the double-resolution volume improves the compression performance up to 0.3 dB. This is because the more details and higher accuracy are obtained from the finer resolution. The details of the volume are not able to cover every pixel in the projected images if the resolution of the volume is coarser than that of the reconstructed images. In this case, the holes appearing in these images are filled by interpolating from the neighbouring voxels instead of being directly rendered from the finer-resolution volume.

4.2 View Synthesis Performance

The view synthesis performance was tested by taking off one of the five reference views to be used for view synthesis assessment. The volume was constructed by the remaining four views and coded by the proposed schemes. The virtual view was then synthesized with the same geometry of the removed reference view and the image quality was compared to the original view. Fig. 5 illustrates the synthesized images. Noticeably, the proposed scheme significantly outperforms the conventional scheme.

5. CONCLUSIONS AND FUTURE WORK

This paper proposes a novel approach for colour volumetric coding which is suitable for realistic view synthesis application. The volumetric data is rearranged by two proposed algorithms: one uses the characteristics of human eye sensitivity to diminish the size of chrominance information and the other reduces the number of components to compact volumetric data. Moreover, the adaptive resolution is introduced to support the scalability and improve quality for large image as well as view synthesis.

The proposed approaches show great improvement in a wavelet coding environment and possibly enhance the performance of other coding methods, such as H.264. Moreover, the proposed schemes are possibly applied to 4D data which represents a time-varying volumetric imaging system.



Fig. 4 Comparison of the performance of thee resolution volumes and the conventional approach.



Fig. 5 The synthesized *Leo* images with the proposed rearrangement approaches coded at (a) lossless (535kbits, PSNR=32.2 dB) and (b) 300kbits (31.15 dB). The synthesized view without data rearrangement coded at (c) 580kbits (29.43 dB) and (d) 320kbits (24.87 dB).

6. REFERENCES

[1] F.F. Rodler, "Wavelet based 3D compression with fast random access for very large volume data," *IEEE Proc. 7th Pacific Conf. on Computer Graphics and Applications*, pp. 108-117, 1999.

[2] Z. Xiong, X. Wu, S. Cheng and J. Hua, "Lossy-to-Lossless Compression of Medical Volumetric Data Using Three-Dimentional Integer Wavelet Transforms," *IEEE Trans.on Medical Imaging*, vol.22, pp. 459-470, 2003.

[3] P. Schelkens, A. Munteanu, J. Barbarien, M. Galca, X. Giro-Nieto, J. Cornelis, "Wavelet coding of volumetric medical datasets," *IEEE Trans. on Medical Imaging*, vol.22, pp.441-458, 2003.

[4] Y. Gao and H. Radha, "Multi-View Image Coding Using 3-D Voxel Models," *IEEE Proc. ICIP* '05, vol.2, pp. 257-260, 2005.

[5] "Information Technology - JPEG 2000 Image Coding System: Extensions," ISO/IEC JTC1/SC29/WG1, 15444-2, May 2004.

[6] N. Anantrasirichai, C. Nishan Canagarajah, David W. Redmill and David R. Bull, "Volumetric Representation for Sparse Multiviews," to be published in *IEEE Proc. ICIP'06*, 2006.

[7] N. Anantrasirichai, C. Nishan Canagarajah, David W. Redmill and David R. Bull, "Dynamic Programming for Multi-view Disparity/Depth Estimation," *IEEE Proc. ICASSP'06*, vol.2, pp.269-272, 2006.

[8] M. Magnor, P. Ramanathan, B. Girod, "Multi-View Coding for Image-Based Rendering Using 3-D Scene Geometry," *IEEE Trans. on Circuits and Sys. for Video Tech.*, vol.13, pp.1092-1106, 2003.