Reference Stream Selection for Multiple Depth Stream Encoding

Sang-Uok Kum
kumsu@cs.unc.edu
Ketan Mayer-Patel
kmp@cs.unc.edu
University of North Carolina at Chapel Hill
CB #3175, Sitterson Hall
Chapel Hill, NC 27599 USA

Abstract

With advances in technology, a dynamic real world scene can be captured, represented, and streamed for realistic interaction in 3D using multiple digital cameras and computers. However, the captured data would be too massive to be streamed uncompressed. Fortunately, these data exhibit spatial and temporal coherence that can be utilized for compression, and research in compression of multiple streams has increased. To facilitate the use of spatial coherence between streams for multiple stream compression, reference streams must be selected. Reference streams are streams that serve as a basis for spatial prediction. Though the selection of reference streams affects encoding efficiency, there has been little research on it. In this paper, we identify the two main approaches for selecting reference streams, and demonstrate that when selecting reference streams, maximizing the volume overlap of reference streams and non-reference streams is more effective than maximizing volume coverage of the reference streams.

1 Introduction

Recently, there has been growing interest in systems for capturing, representing, and transmitting dynamic real world scenes. The acquired real world environments are used in systems for realistic immersive real-time interaction such as 3D immersive systems [5], tele-immersion video conferencing systems [2, 8, 17], 3DTV [4, 14], and medical consultation and surgical training [1].

These systems use multiple cameras for dynamic scene acquisition. One approach for scene acquisition is to use multiple streams captured from varying viewpoints and use image based rendering (IBR) techniques for generating virtual views of the scene [14, 17]. Another approach is to use the imagery from the multiple cameras to derive a 3D model of the environment, which is then transmitted with color [2, 4, 5, 8]. Although using multiple streams requires considerably more data to be transferred than using a 3D model, multiple streams exhibit significant spatial coherence between streams that can be effectively utilized for compression.

A common approach for representing dynamic environments is to use multiple *depth streams* [17, 22]. Multiple depth streams are needed to properly represent a dynamic scene. In general, increasing the number of depth streams generated will increase the quality of the reconstruction from virtual viewpoints. A depth stream is similar to a traditional video stream that has been augmented with per-pixel depth information. Typically this depth information is calculated using stereo-based vision algorithms. The highly parallelizable nature of these algorithms makes real-time depth generation possible [9]. Furthermore, there is active research in developing cameras that can acquire per-pixel depth and color information directly [3, 21].

Uncompressed multiple depth streams, even sparsely sampled, require significant transmission bandwidth. The data set used for this paper has 8 depth streams with 1024x768 resolution and 32 bits per pixel – 1 byte each for RGB and depth – at 15 frames per second (fps). Uncompressed, this would require 2.8 Gb/sec of bandwidth. Therefore different methods have been proposed for compressing and transmitting multiple depth streams.

One such proposed approach is to merge images from multiple depth streams to create a layered depth image (LDI) [15], and use a stream of LDIs [4]. The stream is compressed using traditional video compression techniques, such as MPEG-4 [6], which is sufficient for its target application of 3DTV. However, LDIs impose limits on sampling multiple depth images due to its fixed resolution. It also centralizes processing, making it difficult to scale such a system or increase the number of depth streams used.

Another approach is to exploit the fact that the multiple depth streams will exhibit strong spatial coherence as they are capturing the same environment from different angles. The depth value associated with a pixel in one depth stream (referred to as the *non-reference stream*) can be used to project that pixel into the frame of another depth stream (referred to as the *reference stream*). If the color of the corresponding pixel in the reference stream is similar to the color of the projected pixel of the target stream, encoding of the color information for that pixel in the target stream is no longer required. Simply encoding the depth and indicating that the color can be derived from the reference stream is all that is necessary. Therefore a reference stream must be selected in order to compress multiple depth streams using spatial coherence between streams.

Research has shown that the selection of reference streams affects encoding efficiency [22]. However most of current research on multiple stream compression mainly focuses on encoding of individual streams rather than the problem of selecting reference streams from a set of streams. Often a user is required to pre-select reasonable streams as reference streams. This may be plausible when only a few streams exist, but as the number of streams increases the task becomes more challenging. The group based approach in [10] is the only algorithm for selecting reference streams that we are aware of. Also there have been no studies on evaluating the different approaches to reference stream selection.

In this paper, we advance the work of [10] in several ways. We identify two different approaches used for reference stream selection, and present a reference stream selection algorithm based on [10] for maximizing volume overlap between reference streams and non-reference streams. We also compare this method with selecting reference streams for maximizing volume coverage of the reference streams, and show that the reference streams with maximum overlap performs better.

Our contributions presented in this paper include:

- Identifying and evaluating the two different reference stream selection approaches.
- Proposing metrics for evaluating the effectiveness of the reference image generated from the selected reference streams. In consequence, evaluating the selected reference streams.
- Proposing an algorithm based on [10] for selecting reference streams for encoding multiple streams.

2 Related Work

A light field represents a static 3D scene by modeling its light rays with a set of 2D images [11]. These images have very high spatial coherence as a point in the scene is observed in multiple images. Magnor and Girod [12] used this spatial coherence to compress a static light field. A dynamic light field was used for a 3DTV system [14]. The system did not use spatial coherence between the streams and was encoded by processing each image stream separately using standard video encoding for each stream. The MPEG Ad-Hoc Group on 3D Audio and Video (3DAV) [16] are currently investigating standardization for encoding of multiple streams.

Depth streams have also been used for 3DTV [4]. The color and depth are encoded as separate streams in order to be backward compatible with conventional TV transmission. Therefore various different video codecs were investigated for encoding the depth information as stream of grayscale images. In addition, a representation format for depth streams is proposed in MPEG-4 AFX [7], but it does not define the encoding of depth streams.

Multiple depth streams used for tele-immersion [17] were compressed by removing points redundant in the reference depth streams from the non-reference depth streams [10]. The reference streams were selected by dividing the streams into groups and selecting the stream that best represents the group. While this approach is scalable to the number of depth streams, the performance on real world data sets is not excellent due to imperfect depth values [10]. Moreover, it only takes advantage of spatial coherence between different streams and does not utilize the temporal coherence between frames of the same stream. Zitncik et al. [22] compressed multiple depth streams used for 3D video by using temporal coherence to compress reference depth streams and spatial coherence to compress non-reference depth streams. Würmlin et al. [18] used traditional video codecs to encode each streams individually, thus utilizing only the temporal coherence between frames and ignoring spatial coherence between streams.

Spatial coherence between streams has been used to extract 3D models to represent the scene. Examples include the image-based visual hull [13] used in the tele-immersion video conferencing system Coliseum [2] and video fragments [19] used in the 3D immersive system blue-c [5].

3 Reference Stream Encoding

3.1 Reference Stream Encoding

There are two types of streams that can be considered when encoding streams for multiple stream compression. One is the *intra-stream* that is encoded without referring to any other streams and thus can be decoded independently. The other is the *inter-stream* that uses spatial coherence information from other streams (reference streams) to improve its encoding efficiency. Therefore inter-streams cannot be decoded without decoding its reference streams first. This is very similar to intra-frames and inter-frames when encoding frames using motion compensation. Because all reference frames are from the past and have already been decoded, inter-frames can use intra-frames or inter-frames as reference frames without incurring any latency. However, it is different for stream encoding. As the frames from reference streams used for decoding inter-streams, are from the same point in time, there is a latency penalty associated with decoding the inter-streams. The penalty arises because the reference stream's current frame needs to be decoded first. Furthermore if the reference stream had been encoded as an inter-stream, the latency is increased as two reference streams need to be decoded – the reference stream and the reference stream's reference stream. As more reference streams are encoded as inter-streams, the decoding latency increases. Therefore reference streams are only encoded as intra-streams, and non-reference streams as inter-streams.

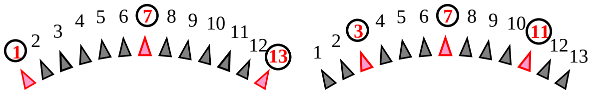
Another restriction for encoding multiple streams using multiple reference streams is the network bandwidth. The data acquisition system for multiple streams uses multiple cameras connected to multiple computers. Therefore the reference streams are usually multicast over the local network bandwidth. Even though local network bandwidth is large it is not infinite, and thus imposes a limit on the number of reference streams that can be used.

3.2 Peripheral and Occluded Pixels

When encoding multiple depth streams, spatial coherence between streams is exploited by projecting pixels from the current stream into the reference stream. One of the following three occurrences will result when projecting pixels. One, the projection produce a similar reference color – the optimal situation. Two, the pixel will project outside the reference stream (referred to as the *peripheral pixels*). Peripheral pixels occur because the current stream and the reference stream do not acquire the environment from the same view. This results in sections of the current stream that is not seen in the reference stream. Therefore a peripheral pixel does not have any spatial reference color to compute its residual.

Finally, the projection will generate a bad reference color. This can occur due to various reasons such as when the depth value used to project a pixel is inaccurate, resulting in an incorrect corresponding pixel and reference color. The inaccurate depth value usually occurs around object borders where multiple objects are visible in a pixel. Thus, the depth value for the border pixel is some average of the depth values of the objects.

A bad reference can also occur due to different sampling rates of the reference stream and the current stream. As the reference stream and the current stream are not acquired from the same view, objects do not occupy the same area in both views. A good example is a poster acquired from the front in the current stream and from an acute angle in the reference stream but at the same distance. In this instance, a pixel of the poster in the reference stream represents a larger area than a pixel of the poster in the current stream. So for a pixel in the reference stream that corresponds to high frequency content areas on the poster, the high frequency content would be blurred. If this high frequency content is detectable in the current stream, when a pixel in this area is projected into the reference stream, the reference color will not match very well.



(a) Maximum coverage (b) Maximum overlap

Figure 1. Example of reference streams.

A third cause of bad reference can be due to occlusion. A surface can be occluded from view in the reference stream but not in the current stream. If a pixel from this surface in the current stream (an *occluded pixel*) were projected into the reference stream, the corresponding pixel's color would be from a different surface. While it is not easy to detect the previous two instances of bad reference colors, detecting occluded pixels are possible using depth values. If the projected pixel's depth is larger than the depth of the corresponding pixel in the reference stream, the pixel can be labeled as an occluded pixel. Like peripheral pixels, occluded pixels do not have any spatial reference color to compute a residual.

When multiple reference streams are used for encoding, a pixel may have multiple reference colors. In such cases, the depth value can be used to select one reference color by using the reference color with the smallest depth difference. Also a pixel may have no spatial reference colors – pixels that are either peripheral pixels or occluded pixels in all reference streams. These pixels can use temporal prediction for its reference color – the color from the previous frame at the same pixel location.

4 Reference Stream Selection

There are two main approaches to selecting reference streams. One approach is to select reference streams that represent the most volume of the scene. The other is to select reference streams such that the volume overlap between the reference streams and the non-reference streams is the maximum. Examples of both approaches are shown in Fig. 1.

Selecting reference streams to maximize volume coverage emphasizes reducing peripheral pixels. For a 1D stream configuration, the streams at both ends are first selected as reference streams. Then for selecting *n* reference streams, the space between the initial reference streams is divided into $n-1$ sections. Streams that are closest to the section borders are then selected as reference streams. Fig. 1a shows an example of selecting three reference streams from a total of thirteen streams, which maximizes volume coverage. For more complex stream configurations selecting reference streams to maximize volume coverage is not trivial.

Selecting reference streams to maximize overlap between reference streams and non-reference streams emphasizes on reducing occluded pixels. However compared to the maximum volume coverage approach, the amount of peripheral pixels increases due to decrease in total volume coverage. Except for some trivial stream configurations, it is difficult to select reference streams to maximize volume overlap between reference streams and non-reference streams. Therefore the issue is simplified to select reference streams that maximize the overlap between non-reference streams and its closest reference stream. This is possible as the effect of the initial spatial reference stream is significantly larger than the subsequent spatial reference streams.

The angle between the view directions of two streams can be used to approximate volume overlap, since volume overlap between two streams is expensive to calculate. Empirically, the view directions of two streams are a good estimate for how much the two stream volumes overlap; the smaller the angle, the bigger the overlap. Fig. 1b shows an example of selecting three reference streams from a total of thirteen streams that maximizes volume overlap of reference streams with non-reference streams.

The group based approach presented in [10] for finding center streams in groups can be used for selecting reference streams to maximize volume overlap. However some modifications is required due to a characteristic difference. In [10], the reference streams change dynamically each frame. Also the reference streams are used to remove redundant points – i.e. reference streams are needed for encoding a stream only but not for decoding. Therefore, reference streams can be encoded as inter-streams without introducing any decoding latencies. Consequently, only one reference stream is encoded as an intra-stream while all other streams, reference and non-reference, are encoded as inter-streams. Thus, the volume overlap between the reference streams affect encoding efficiency. However, for encoding multiple depth streams, all reference streams are encoded as intra-streams to reduce decoding latency. Therefore only the overlap between reference streams and non-reference streams affect encoding efficiency.

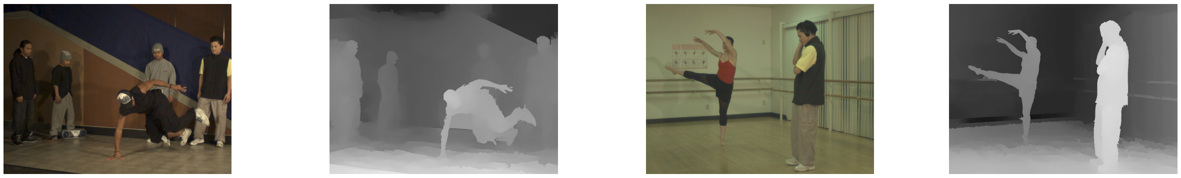
For selecting k reference streams from a total of n streams, the streams are divided into k disjoint groups using the three metrics from [10] – *local squared angle sum* (LSAS), *group squared angle sum* (GSAS), and *total squared angle sum* (TSAS) – are used for reference stream selection. After an acceptable group partition has been achieved, the reference stream of each group is then selected as the resulting reference streams.

Partitioning streams into groups starts with selecting initial reference streams. As the number of reference streams and the total number of streams increase, the total possible number of initial reference streams ($_nC_k$) becomes impractical to completely investigate. Thus, only a small sample of all possible instances can be examined. The chances of finding the same optimal solution as the complete search will increase by intelligently selecting the initial reference streams. A detailed discussion on generating initial reference streams and an algorithm for generating a set of good initial reference stream candidates can be found in [10]. Once a set of good initial reference stream candidates has been generated, the following algorithm can be used to select the final reference streams.

1. Generate a set of initial reference stream candidates. For each set of generated initial reference stream candidates,
2. Partition streams into groups based on the initial reference streams. The criterion used to assign non-reference streams to a group is the absolute angle of the stream and the group's reference stream. The non-reference streams are assigned such that this absolute angle is the smallest. This would result in group partitions where TSAS is minimized.
3. The stream with the lowest LSAS for each group is selected as the new reference stream for the group.
4. The streams are regrouped using the new reference streams using the same criterion used in step 2.
5. Steps 3 and 4 are repeated until the group's reference streams do not change. These reference streams are then selected as one of the possible choices.
6. Repeat steps 2 thru 5 for the other initial reference stream candidates to generate all possible choices of reference streams.
7. From the generated set of possible reference streams, the reference streams with the lowest TSAS are selected as the final reference streams.

5 Result

Two data sets from [22], Breakdancers and Ballet, were used to compare the two different reference stream selection approaches. Both data sets have 8 depth streams, which are 100 frames long at 1024x768 resolution and were captured at 15 fps. Each frame has 24 bit color (RGB) and 8 bit depth information for each pixel. The depth was computed using the technique described in [22]. Example frames from both data sets are shown in Fig. 2.



(a) Breakdancers Color (b) Breakdancers Depth (c) Ballet Color (d) Ballet Depth

Figure 2. Frame from Breakdancers and Ballet Data.

5.1 Reference Stream Selection

Since both data sets have eight streams, two streams were selected as reference streams. Fig. 3 shows the two reference streams selected using the reference stream selection algorithm for maximizing volume overlap. Since there are only eight streams all possible combinations, ${}_8C_2$, were tried as initial reference streams. Fig. 4 shows the generated reference streams, the associated total squared angle sum (TSAS), and the number of initial reference streams that resulted in the given reference streams. For the Ballet data set, two possible reference stream selections resulted and the one with the smaller TSAS was selected – streams 2 and 5. This also had more initial starting points. For the Breakdancers data set all 28 possible starting reference streams resulted in one possible reference stream selection – streams 2 and 6. Fig. 4 also shows the TSAS for reference streams that maximize volume coverage – streams 0 and 7. Compared to the reference streams that maximize volume overlap, the TSAS is significantly higher.

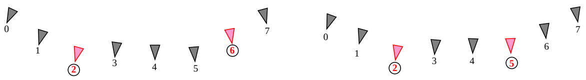
5.2 Reference Images

The two different reference stream selection methods are compared using two different methods. One is the reference image generated from the reference streams. The other is the compression ratio of the residuals. In this section, results for the generated reference images are presented.

The reference image for computing the residuals is generated by projecting each pixel to find the corresponding pixel in the reference streams. When corresponding pixels exist in multiple reference streams, the closest one (smallest depth difference) is chosen. However, there are also pixels that do not project to any correspondences (peripheral and occluded pixels), which are shown as green in Fig. 5d-f. Therefore, the reference colors for these peripheral and occluded pixels are temporally predicted: color of the pixel at the same location in the previous frame. However, this temporal prediction does not always results in good estimated values (Fig. 6). Fig. 6a is a portion of Fig. 5e where no spatial prediction was available. Therefore, these pixels were predicted by the temporal reference frame. However, in the previous frame, the wall was occluded by the arm, resulting in bad predicted values (Fig. 6b).

When reference images are used to evaluate the effectiveness of the reference streams, much consideration has to be given when using temporal prediction. Since the goal is to evaluate spatial prediction, spatially predicted value should be chosen when available even if a better temporal prediction exists. Additionally, using temporal reference for pixels that do not have spatial reference may distort results. Assume significant number of pixels was peripheral pixels, but the temporal predictions for these pixels were good predicted values. If every pixel were considered in evaluating this reference image, this would be identified as a good reference image since most of the pixels have good estimates. However, significant number of the pixels was peripheral pixels, which would indicate that the reference streams were not optimal. If only the pixels with spatial references were considered, the reference images with many peripheral and occluded pixels would not be penalized. Therefore, results for reference images with peripheral and occluded pixels (labeled as ‘All’) and without peripheral and occluded pixels (labeled as ‘Spatial’) are both examined.

All frames in this experiment were encoded as inter-frames with the previous frame encoded as an intra-frame. Also all frames of the reference stream and depth were encoded as intra-frames. This was done to eliminate error



(a) Breakdancers (b) Ballet

Figure 3. Data reference streams.

	Ref. Str.	TSAS	Initial
Ballet	1, 5	187.655	11
	2, 5	181.25	17
	0, 7	416.502	0
Breakdancers	2, 6	145.84	28
	0, 7	332.88	0

Figure 4. Reference streams selection.

(a) Reference Stream 2 (b) Temporal Reference (c) Reference Stream 5

(d) Projected Reference Stream 2 (e) Spatial Only Reference Frame (f) Projected Reference Stream 5

(g) Original Frame to Encode (h) Reference Frame (i) Decoded Frame

Figure 5. Frame images at color and depth quantizer of 5.

propagation when an inter-frame is encoded using an inter-frame as reference. Also the quantizer for all blocks in a frame was kept constant for the same components. This insures that the encoding is affected similarly across blocks within the same frame. Finally the color quantizer for all streams were kept same, as well as the depth quantizer.

5.2.1 Peripheral and Occluded Pixels

Fig. 7 shows the portion of pixels in a reference frame that are either peripheral pixels or occluded pixels. It is apparent that the quality of depth affects the number of peripheral and occluded pixels, which are to be expected since the quality of depth affect pixel projection. Also the figures show, for both data sets, reference streams with maximum coverage have more peripheral and occluded pixels than reference streams with maximum overlap.

(a) Spatial prediction (b) After temporal prediction

Figure 6. Bad temporal prediction.

5.2.2 Residual Accumulative Histogram

The *accumulative residual histogram*, which shows the number of pixels equal or less than the given residual, is used to compare the effectiveness of the reference images. Fig. 8 shows the average accumulative residual histogram of Breakdancers and Ballet at color quantizer (cq) and depth quantizer (dq) of 5. For both Breakdancers and Ballet data, the reference streams with maximum overlap have more pixels with smaller residuals than reference streams with maximum coverage. This holds true for reference images with peripheral and occluded pixels and reference images without peripheral and occluded pixels.

5.2.3 Residual Average and Standard Deviation

Fig. 9 shows the average and standard deviation of the residuals. The Ballet data, for both reference images with and without peripheral and occluded pixels, has a smaller residual average and standard deviation for reference streams with maximum overlap than reference streams with maximum coverage. This suggests that the reference images generated from reference streams with maximum overlap

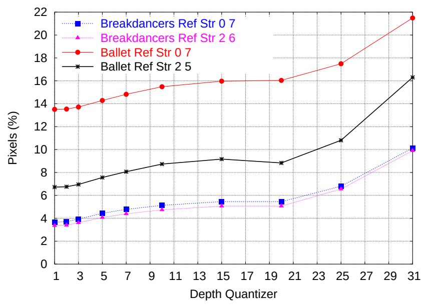
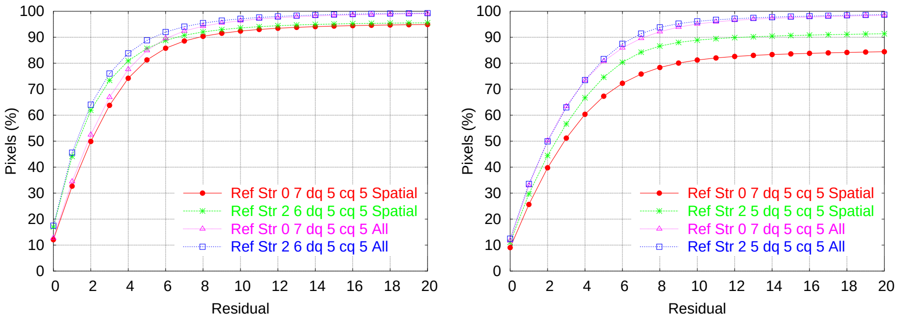
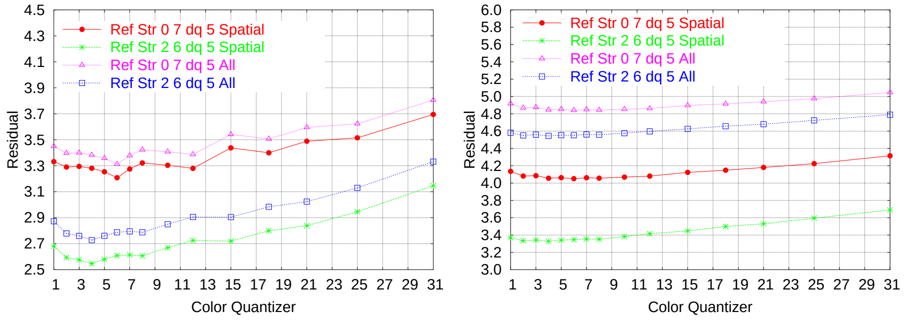


Figure 7. Peripheral and occluded pixels.

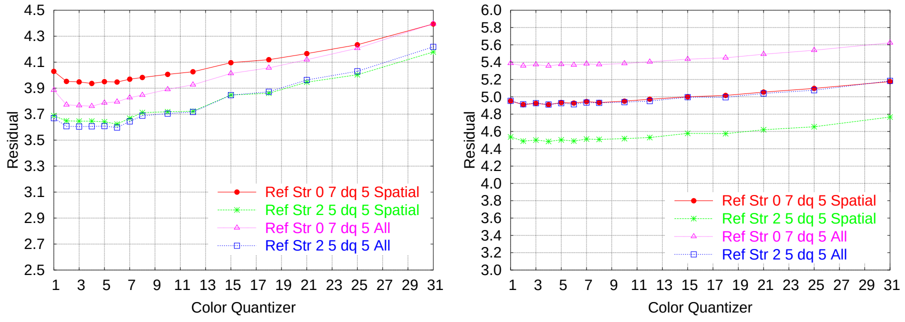


(a) Breakdancers (b) Ballet

Figure 8. Residual accumulative histogram.

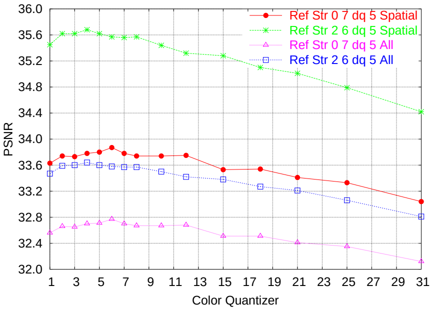


(a) Breakdancers average (b) Breakdancers standard deviation

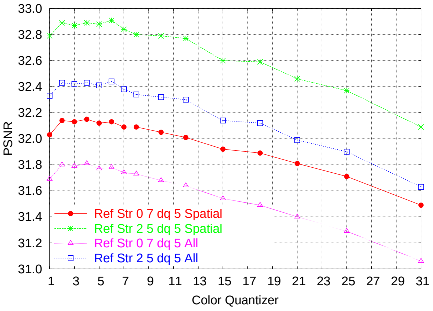


(c) Ballet average (d) Ballet standard deviation

Figure 9. Residual average and standard deviation.



(a) Breakdancers



(b) Ballet

Figure 10. PSNR.

are better. This is also true for the Breakdancers data.

5.2.4 Reference Image Quality

The image quality of the reference image is assessed by comparing it with the image to be encoded. Fig. 10 shows the PSNR of the reference images at depth quantizer (dq) 5. PSNR for reference images with peripheral and occluded pixels, and reference image without them are presented. For both reference images and data sets, reference streams with maximum overlap have slightly better PSNR than reference streams with maximum coverage.

Results of comparing different metrics for reference images – peripheral and occluded pixels, residual accumulative histogram, residual average and standard deviation, and PSNR – show that reference streams with maximum overlap performs better than reference streams with maximum coverage. This is true for both Breakdancers and Ballet data sets, as well as for reference images with and without peripheral and occluded pixels.

5.3 Compression Ratio

In this section, the compression ratio of the residuals is examined. The reference image used for computing the residuals are generated from the two reference streams and

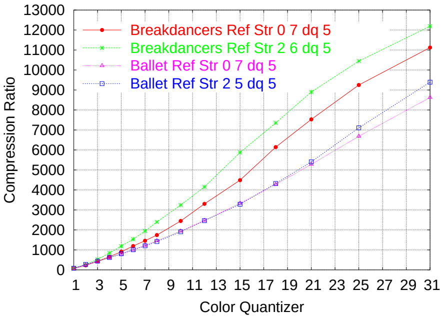


Figure 11. Compression ratio.

the temporal reference frame (Fig. 5h). The residuals are encoded with XviD [20]. XviD is an open source ISO MPEG-4 [6] compliant video codec. XviD was selected mostly because of source code availability, which can be modified without any restrictions. The residuals are encoded by XviD as if it had been computed using motion compensation. The resulting bitstream encoding only has information of the residual coefficients – i.e. no motion vectors encoding.

Fig. 11 shows the average compression ratio for encoding residuals of the six non-reference streams. The figures compare selecting reference streams with maximum coverage and selecting reference streams with maximum overlap for depth quantizer (dq) 5. For both data sets, selecting reference streams with maximum overlap performs better. Furthermore, the accuracy of the depth has an effect on the compression ratio – more accurate the depth is the better the compression ratio. This is to be expected because as the depth values gets more accurate, so does the projection of the pixel to the spatial reference stream.

6 Conclusion

The spatial coherence between streams greatly aid in effective encoding of multiple streams. In order to leverage this spatial coherence between streams, reference streams must be selected. Previously, reasonable streams were preselected by the user as reference streams. However, as the number of streams increase, the task will become more difficult. In this paper, we have identified the two different approaches used for reference stream selection. Furthermore, we have proposed metrics for evaluating the different methods of reference stream selection, and shown that the reference streams with maximum overlap is better than reference streams with maximum coverage.

For future work we would like to examine the following:

- Using temporal prediction with spatial prediction leads to good results. However, as seen in Fig. 6, it does not always predict well. Investigating methods to incorporate more advanced temporal prediction techniques, such as using motion vectors, should lead to better results.
- It was shown that the quality of depth values affects the efficiency of encoding. However encoding depth at high quality for better encoding of color increases the bitrate for encoding depth. Identifying the relationship between depth and color would help in designing an overall guideline for encoding multiple streams.

References

[1] 3D Telepresence for Medical Consultation. `http://www.cs.unc.edu/Research/nlm`.
[2] H. H. Baker et al. The Coliseum immersive teleconferencing system. In *ITP 2002*, Juan Les Pins, France, Dec. 2002.
[3] CanestaVision. `http://www.canesta.com`.
[4] C. Fehn et al. An evolutionary and optimised approach on 3D-TV. In *IBC*, pp. 357–365, Amsterdam, The Netherlands, 2002.
[5] M. Gross et al. blue-c: A spatially immersive display and 3D video portal for telepresence. In *SIGGRAPH 2003*, pp. 819–827, San Diego, CA.
[6] ISO/IEC JTC1/SC29/WG11. *Coding of Audio-Visual Objects – Part2: Visual*, 2004.
[7] ISO/IEC JTC1/SC29/WG11. *Coding of Audio-Visual Objects – Part16: Animation Framework eXtension*, 2005.
[8] P. Kauff and O. Schreer. An immersive 3D videoconferencing system using shared virtual team user environments. In *4th ACM CVE*, pp. 105–112, Bonn, Sept. 2002.
[9] N. Kelshikar et al. Real-time terascale implementation of tele-immersion. In *the TPA*, Melbourne, Australia, 2003.
[10] S.-U. Kum and K. Mayer-Patel. Real-time multidepth stream compression. *ACM TOMCCAP*, 1(2):128–150, 2005.
[11] M. Levoy and P. Hanrahan. Light field rendering. In *SIGGRAPH 1996*, pp. 31–42, New Orleans, LA.
[12] M. Magnor and B. Girod. Data compression in image-based rendering. *CSVT*, 10(3):338–343.
[13] W. Matusik et al. Image-based visual hulls. In *SIGGRAPH 2000*, pp. 369–374, New Orleans, LA.
[14] W. Matusik and H. Pfister. 3D TV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. In *SIGGRAPH 2004*, pp. 814–824, Los Angeles, CA.
[15] J. Shade et al. Layered depth images. In *SIGGRAPH 1998*, pp. 231–242, Orlando, FL.
[16] A. Smolic and P. Kauff. Interactive 3-D video representation and coding technologies. *Proc. of the IEEE*, 93(1):98–110.
[17] H. Towles et al. 3D tele-immersion over internet2. In *ITP 2002*, Juan Les Pins, France, Dec. 2002.
[18] S. Würmlin et al. Image-space free-viewpoint video. In *VMV*, pp. 453–460, Erlangen, Germany, Nov. 2005.
[19] S. Würmlin, E. Lamboray, and M. Gross. 3D video fragments: Dynamic point samples for real-time free-viewpoint video. *Computers & Graphics*, 28(1):3–14.
[20] XviD. `http://www.xvid.org`.
[21] ZCam Camera. `http://www.3dvsystems.com`.
[22] C. L. Zitnick et al. High-quality video view interpolation using a layered representation. In *SIGGRAPH 2004*, pp. 600–608, Los Angeles, CA.