Compression Issues in Multiview Autostereo Displays

Druti Shah and Neil A. Dodgson

Computer Lab, University of Cambridge, New Museum Site, Pembroke St,
Cambridge CB2 3QG, UK.

ABSTRACT

Image compression for multi-view autostereoscopic displays is one of the major governing factors for the development and acceptance of 3D technology. For example: the multi-view autostereo display developed at Cambridge uses between six and twenty-eight distinct views of the scene, each view being a complete image of the scene taken from a particular view point. Therefore it is of prime importance to use compression methods that would exploit the redundancy present in the view-point direction to allow for 3D image communication since the images require a very high bandwidth for transmission and a large amount of space for storage. In this paper an initial investigation on how the third dimension can be utilised is presented. Entropy measures for multi-view images are derived. It is shown that exploiting the similarities between views can give lower entropy, indicating higher compression rates. The parallel axes geometry of the cameras used for the autostereo display produces only horizontal shifts between stereo images, therefore investigation in using hierarchical row decomposition along with correlation and mean squared error measures for estimating disparity shifts and reducing search spaces respectively are presented.

Keywords: Autostereo image compression, view-point images, entropy measures, hierarchical row decomposition

1. INTRODUCTION

Image compression for multi-view autostereoscopic displays is one of the major governing factors for the development and acceptance of 3D technology. These displays provide a 3D picture which can be viewed without the need for any special glasses or other headgear. This is achieved by displaying a number of different, laterally-spaced views of a scene so that each view is only visible in a limited angular segment in front of the screen. Autostereo display technology thus gives a new dimension in image data – a view-point sequence of images.

The multi-view autostereo display[1–4], developed at Cambridge employs a time-multiplexed system to produce a multi-view autostereoscopic image. Each view is then displayed on a CRT and a dynamic optical system ensures that each view is only visible in a single window in front of the display. It is able to display between six and twenty-eight distinct views of the scene, each view being a complete image of the scene taken from a particular view point. To allow for multi-view autostereo 3D image broadcast, transmission, and storage, a very high bandwidth for transmission and a large storage space is required. High bandwidth can be tolerated for certain applications where the image source and display are close together but, for long distance or broadcast, compression of information is essential. As the number of views and the image size both increase the above problems get worse. Already the amount of data required for the current 28-view display is causing data-handling problems. Therefore compression methods need to be investigated for this technology to be feasible.

An initial investigation on the behaviour of the third dimension with respect to compression issues has been carried out. Section 2 gives a brief review of the Cambridge autostereo display, other works on autostereo displays and compression. Section 3 discusses entropy measures for multi-view images. Section 4 investigates hierarchical row decomposition for estimating disparity shifts with respect to correlation and least mean squared error measures. Section 5 summarises the work.

Further author information: (Send correspondence to Neil Dodgson)
E-mail: Neil.Dodgson@cl.cam.ac.uk

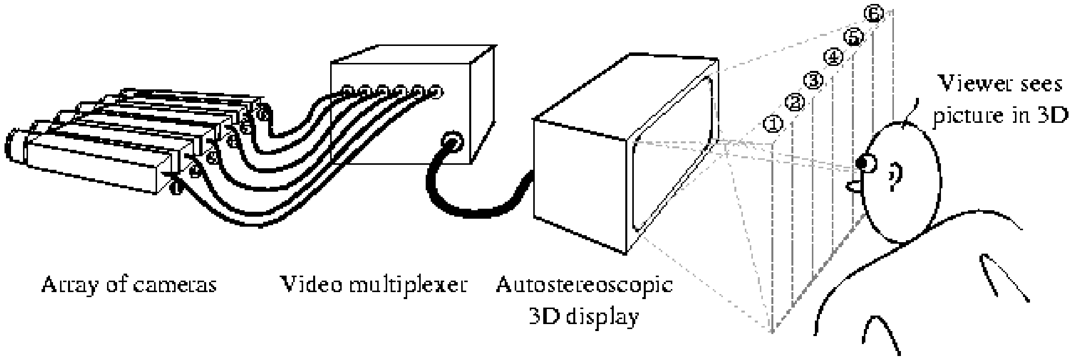

Figure 1. An overview of an autostereo display system.

2. BACKGROUND

2.1. The Display

Three dimensional displays offer added realism over conventional 2D display devices. They are especially desirable for visualisation and remote manipulation, where the extra dimension provides much need depth perception.

An autostereo display (see Figure 1) allows the viewer to see a true 3D picture. Each of the viewer's eyes sees a different image of the displayed scene, just as in real life, and the viewer can move his/her head to "look around" objects in the scene. The result is autostereoscopic vision, and is perfectly natural because there is no need for any special glasses or other headgear.

Over the last few years, a huge amount of work has been developed for displaying autostereo images. Some of the methods proposed include parallax barrier[5,6], lenticular sheets[7,8] and multi-projector systems[9]. However all these displays have certain constraints. For example: in the parallax barrier method, a slight movement of the observer in the horizontal direction within the viewing zone will impair the image in terms of the light density while for lenticular displays, the clear vision of the scene is interrupted due to the reflective light from a lenticular lens sheet. Both the above technologies use spatial multiplexing thus increasing the horizontal resolution of the display. In a multi-projector system, the number of views can be increased with an increase in the number of projectors since each projector displays one view. Such a system can, however, become very expensive.

The Cambridge autostereoscopic display[1–4] uses time-multiplexed system to produce a multi-view autostereo image. These multiple pictures are flashed up on the CRT very quickly, one after another. At the same time as one of the pictures is being displayed, one of a set of liquid crystal shutters is opened, making the picture visible to part of the area in front of the display. The shutters determine where the observer can see each of the pictures. This whole process is repeated very rapidly, sixty times a second. Each of the observer's eyes thus sees a series of very short, very bright image of one of the pictures. The eye integrates these short bursts of image to give the effect of a continuously displayed picture.

Because each eye sees a different picture, the observer gets one of the important 3D depth cues: stereo parallax. Because she sees different pictures when she moves her head, she gets another important 3D depth cue: movement parallax. These two combine to give an effective illusion of real depth in the 3D image.

The multi-view images used in the display have been generated either by computer graphics or sourced by an array of cameras. The cameras were arranged in a horizontal row, a suitable distance apart from each other. A parallel configuration for mounting the cameras was adopted (see Figure 2). The reason for selecting the above configurations was that the common viewplane was displayed on a single screen thus avoiding distortions. This resulted in a good 3D effect with the depth of view limited to prevent view problems[10,11].

2.2. Compression Overview

Several 3D displays have been developed to enhance the reality of visual communication through the provision of natural depth sensation. For this technology to be realised issues such as image capture, image representation and

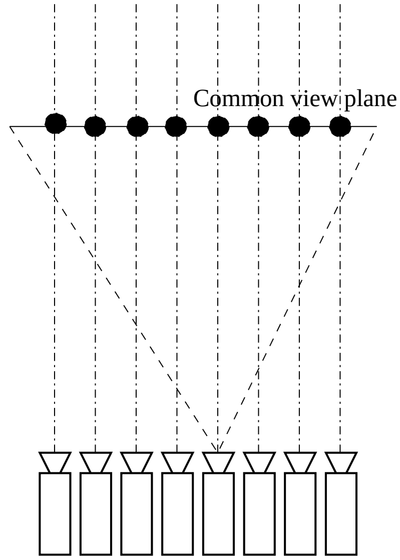

Figure 2. Camera geometry.

image compression need to be discussed. One of the major factors governing this technology is the acceptance of multi-view compression.

It is known that current image compression techniques are applicable to 2D static images and time sequences of images. These techniques could be applied to the individual views in a view-point sequence, however this would not take advantage of the high degree of similarity between views. Most of the work to date on compression has been for stereo images. The issues concerning stereo images such as disparity estimation, spatial correlation, MPEG compatible coding issues, the human visual system and standard 2D image and video coding techniques can be utilized for multi-view images.

One of the first published work on stereo compression methods was by Perkins[12]. He worked on a conditional stereo pair coder structure which described a scheme that first coded one image of a stereo pair and then the other based on the first. He also presented two compression techniques for stereo pairs. The first technique attempted to minimise the mean squared error between the original stereo pair and the compressed stereo pair. The second technique was based on the known facts of human stereovision to code stereo pairs in a subjectively acceptable manner.

A few years on, a European project DISTIMA[13], explored stereoscopic systems. The project developed a stereoscopic display based on twin projectors. The group worked on various factors dealing with coding and compression of stereo images.

Research groups at the University of Tokyo have worked with different algorithms. In brief, Fujii et al[14] showed how a multi-view sequence can be reduced to a mesh structure (generated by placing vertices at points of minimum variance) and a texture map (dependent on disparity estimation). Later on Naemura et al[15] worked on improving Fujii's method. They used a 3D segmentation algorithm, before transforming the data into a mesh structure and texture map.

Aydinoglu et al[16] worked on coding the images using local orthogonal bases. They adopted two different frameworks, a bidirectional predictive coding scheme and a unidirectional predictor. The actual coding was performed using subspace projection technique and a locally adaptive incomplete transform approach. It was shown both the frameworks were able to overcome problems related to occlusion.

A recent overview of 3D compression and representation is given by Naemura et al[17].

3. THIRD DIMENSION INVESTIGATIONS WITH ENTROPY MEASURE

For autostereo images, it is wise to investigate how the third dimension can be utilised in compression. Many 2D compression schemes treat a 2D image as a 1D stream of data. The order in which the 2D image data is input to the 1D compression algorithm can impinge on the compression ratio. With autostereo 3D imagery the extra dimension allows for many more options. At its simplest, where a 2D image could be scanned as either (x, y) or (y, x), a 3D image could be scanned up to six ways including (x, y, θ), (y, θ, x) and (θ, x, y). In addition, the behaviour of data in the x and y dimensions can be expected to be similar, while in the θ dimension the data may behave in quite a different manner. Thus to investigate how the behavior differs, zero order and first order entropy measures were employed to see the amount of redundant information in each frame. Zero order entropy indicates the amount of compression possible if each pixel is treated independently. First order entropy indicates the compression which is possible if each pixels' value is predicted from the previous pixel, where "previous" could mean previous of any one of x, y or θ.

Experiments were carried out on different computer generated and real autostereo images. Due to space limitations, only the results for the computer generated sequence known as Granny are discussed here. A single frame of the Granny sequence (which has 14 views in the view point direction) was considered.

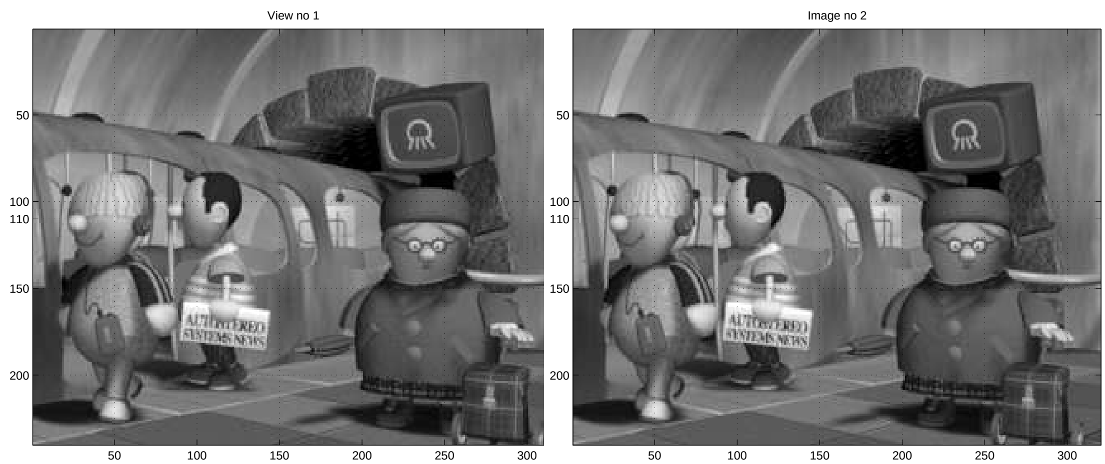

Figure 3. Images from a single frame Granny sequence.

Figure 3 shows the first and second views of the Granny single frame sequence.

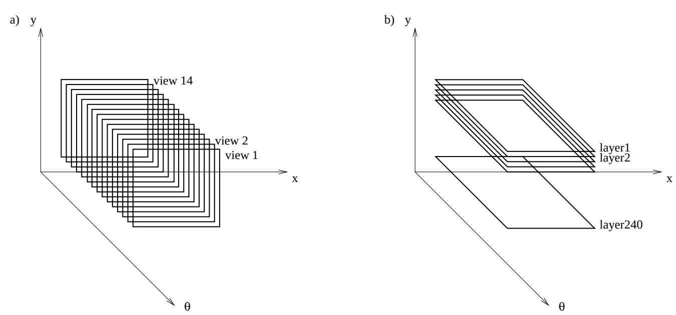

Figure 4. Two different possibilities of scanning the x, y and θ dimensions: a) view by view b) layer by layer.

The first experiment compared view with view (see Figure 4a), to ascertain how good a predictor one view is for the next. Figure 5 shows the entropy measure for a single view sequence. It can be seen that the entropy varies in small steps between consecutive views implying that a large amount of information can be discarded giving good compression. It should be noted that the difference between two views increases as the distance between views increases. If a lossless encoding method were to be utilised it would be best to code each view with respect to the previous view, rather than with respect to the first view.

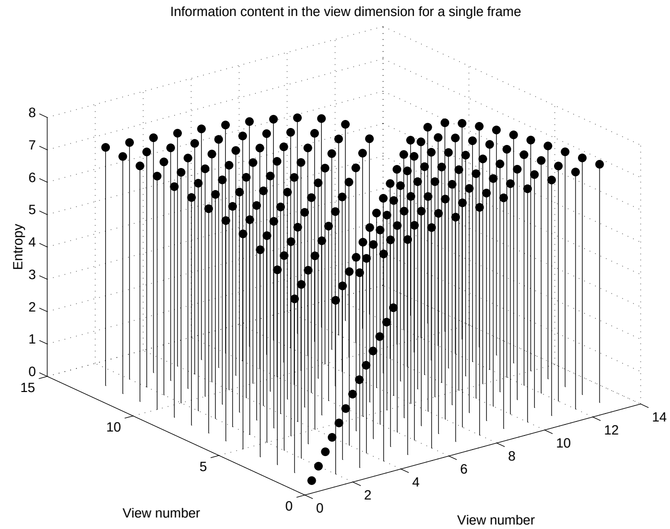

Figure 5. Entropy measure of a single frame(14 views) of the Granny sequence.

There is however no guarantee that the θ dimension contains the greatest redundancy. Therefore the second experiment compares "layer" with layer, where a layer consists of a single row taken from all views (see Figure 4b). Figure 6 shows results from this experiment. It is observed that the entropy measure in the y dimension is very small, as seen in the top left graph, suggesting the possibility of a greater degree of compression compared to the previous results. The remaining three graphs show small portions of the layers and it is seen, for example, in the bottom left graph, that the first order entropy of the first few layers is relatively small since there is not much motion information in the first few rows of the views (see Figure 3).

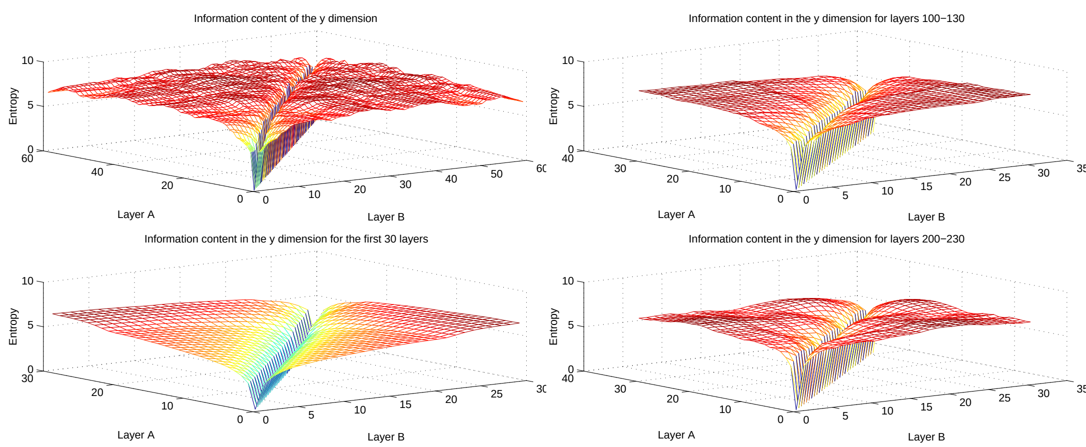

Figure 6. Entropy measure in the y dimension for Granny sequence.

These experiments show that there is better coherence between layers than views. Therefore, naively using one view to predict the next is not going to be as good as simply predicting within a view. However we know that there is a disparity shift between views, thus from our knowledge we now consider estimating disparity shifts within a view and between views, using different search space measures.

4. ESTIMATION OF DISPARITY SHIFTS

Differences in images of the real and computer generated scenes may be caused by 1) the relative motion of the camera and the scene, 2) the relative displacement of the cameras or 3) the motion of objects in a scene. These differences are important because they encode information that often allows a partial reconstruction of the 3D scene structure from 2D projections. Such differences between two images can be represented as a vector field mapping one image onto another. An example can be found in Figure 7.

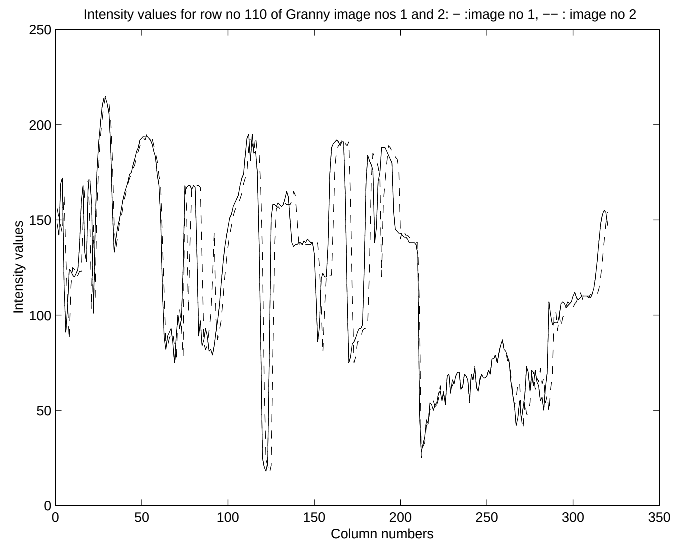

Figure 7. Delta shift between images 1 and 2 for row no. 110 where image 1 is represented by a line (—) and image 2 by line segments (- -).

Matching between the sets of points from two images gives rise to two important issues. Firstly, how to select points for matching and secondly how to determine which matches are correct. In this section the above issues are discussed with respect to space search measures and hierarchical row decomposition.

4.1. Search space measures

Search space measures are measures of similarity, they work by considering a small region in image 1 which surrounds an interesting point, a search is then made in image 2 for the region of maximum similarity. The two measures compared are the correlation measure and least mean square error.

Correlation measure (CM):

In correlation based measurements, the elements to match are image windows of fixed size, the similarity criterion is a measure of the correlation between the windows in the two images. The best match element is given by the window that maximises the correlation within the search region.

Experiments were carried out for the Granny images. A maximum of a 50 pixel search region in image 2 with respect to the reference image 1 was considered. The outcome showed that most of the difference was within [-4,4] pixel range. At some positions other than the true match positions, spikes were observed which indicated false matches. In the next set of experiments, correlation between rows from images 1 and 2 were obtained and this proved to give better results as well as better entropy values. Figure 8 shows the correlation vs delta shift between the images for a specific row. It is observed that the maximum correlation value lies near the origin (at a pixel shift of 2) proving that the difference between the two views lies within the specified pixel range.

In the following simulation, use of windows of size (1x16) and (1x8) were employed and the correlation between rows was found. It was observed that using a (1x16) window did not give good matches since the information to be compared with was too varied. Good matches were obtained from a (1x8) window. For both the experiments, a search space of (1x24) pixels was used. Figure 9 shows the outcome for a specific row.

Least mean square error measure (LMSE):

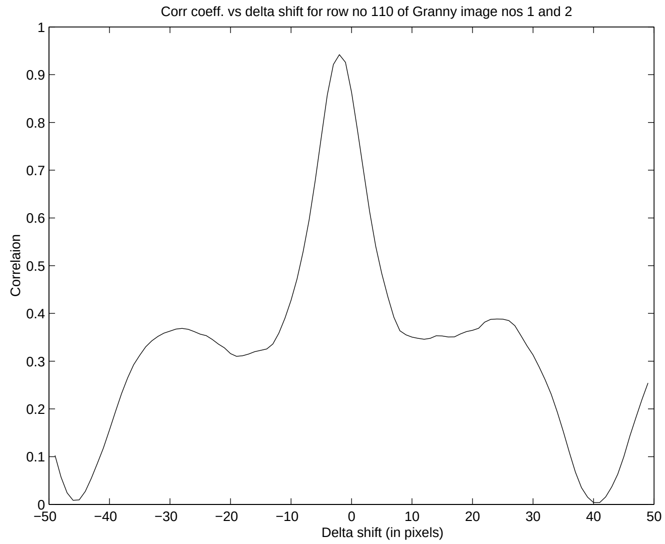

Figure 8. Correlation measure wrt. delta shifts for row 110 between images 1 and 2. Note that the maximum correlation value is at a pixel shift of 2 thus a true match for row 110 of image 2 is obtained by shifting the corresponding row of image 1 by 2 pixels.

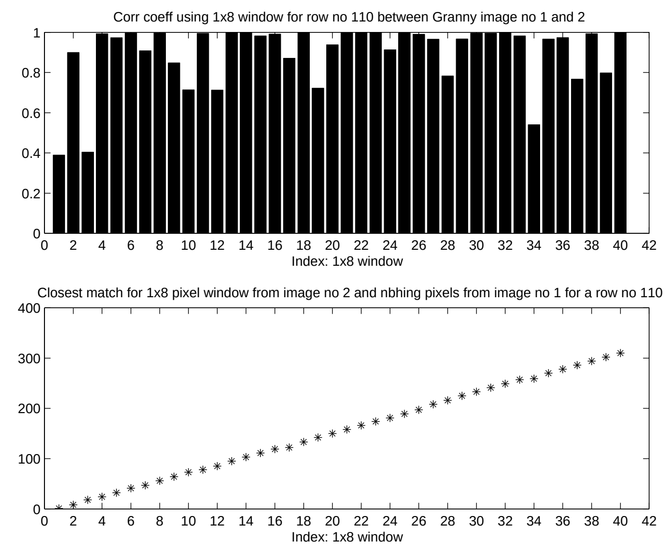

Figure 9. Correlation measure and the best match between images 1 and 2 for row 110.

The process of searching matching regions between images frequently leads to spurious matches. The measure of a successful match is an error measure between image 1 and the selected subpart of image 2. The mean-square error is thus defined by:

$$E(\hat{I}2, I1) = 1/N \times M \sum_{N-1} \sum_{M-1}((\hat{I}2 - I1)^2)$$

where $I1$ is the reference image and $\hat{I}2$ is the estimated image.

Similar sets of experiments were performed, the outcome is shown in Figure 10. It was concluded that a window of size (1x8) gave the best results.

Finally, entropy measures were calculated for delta shift and the closest matches as shown in Table 1. As observed the LMSE measure works a fraction better than the CM measure.

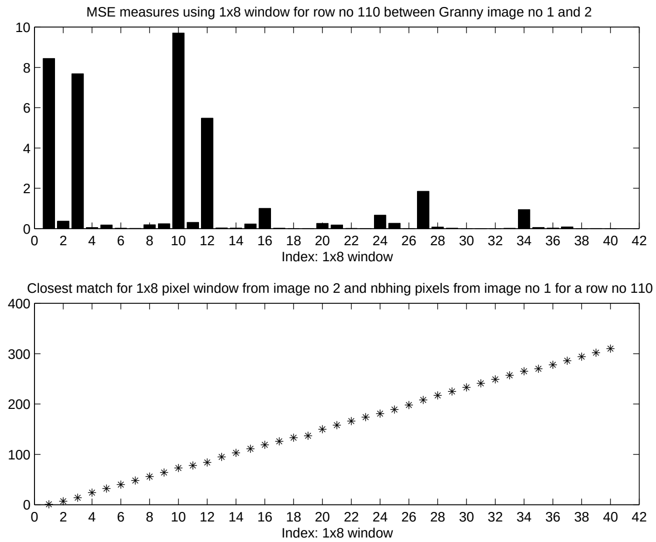

Figure 10. Mean square measure and the best match between images 1 and 2 for row 110.

We now consider ways of combining disparity estimation with our knowledge that there is often good correlation between layers.

4.2. Hierarchical row decomposition (HRD)

Hierarchical row decomposition attempts to combine the advantages of large similar blocks with those of single rows. The reliability of matching is influenced by block size. Large blocks are more likely to track overall movement than small ones and hence are less likely to converge on local minima or maxima depending on the search space measure used. Although such matches are reliable, the quality of matches for large blocks may not be as good compared to small blocks. Thus hierarchical row decomposition exploits these factors to obtain an overall better match between two images.

The HRD algorithm initially starts by finding the correlation coefficient between two lines of image 2 (i.e. the image that is to be matched with reference image 1). If the coefficient value is greater than a specified threshold, then the two rows are combined to form one block. The resulting block is then correlated with the next row and depending on the value obtained, may be adjoined with the block or not. For the experiments a threshold value of 0.95 was used. The final outcome for Granny image 2 is shown in Figure 11. The top graph in the figure shows the blocks, in black, obtained using a threshold of 0.95. In the second graph, the block correlation coefficients are displayed. It is observed that all the true matches lie between [-4 4] pixel range.

Combining this technique with search space measures to detect local matches, significant results in terms of entropy and matching were obtained. This is due to the similarity between and within rows of consecutive images.

Figures 12 and 13 show the frequency distribution of greyscale and delta shift with respect to search space measures. It can be seen that the range for the greyscale and the delta shift both lie near the origin indicating a possibility of a high degree of compression. These results can be verified from Table 1.

Finally, Figure 14 gives a comparison of the HRD match with respect to CM and LMSE methods. As can be seen both the methods have estimated closely to the original reference image.

To prove that these issues are relevant to other images, entropy measures were calculated (see Table 2) for the Porsche car sequence shown in Figure 15.

5. CONCLUSION AND FUTURE WORK

This paper has investigated how the third dimension can be utilised and its effect on information loss and compression. It has been shown that exploiting the similarities between views gives lower entropy, indicating higher compression

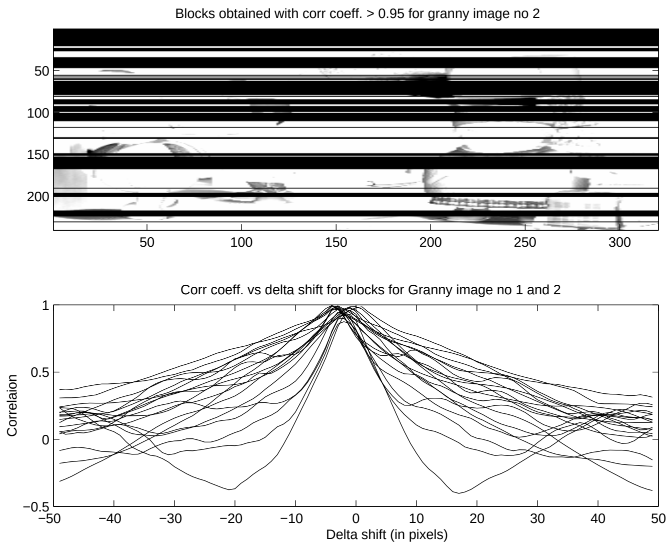

Figure 11. Outcome of row decomposition algorithm, showing blocks in black obtained using 0.95 correlation threshold and the corresponding block correlation coefficient. The maximum correlation lies within the [-4 4] pixel range.

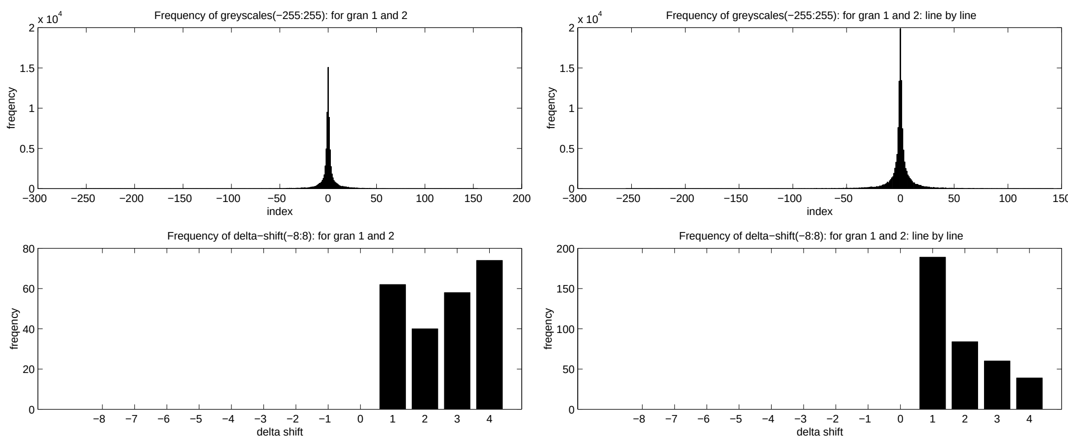

Figure 12. Frequency and delta shift distribution for images 1 and 2 for blocks and lines wrt. correlation measure.

rates. The parallel axes geometry of the cameras produces only horizontal shifts in images. Therefore hierarchical row decomposition and search space measures can be employed as a means of selecting points for matching, to find areas and finally to assess potential for data compression. It was found that using the search space methods and windows of size (1x8) gave the lowest entropy values indicating the likely possibility of high compression.

These preliminary results indicate that multi-view point sequences of images can be coded efficiently. Future work will involve automatic generation of intermediate images in the sequence and the investigation of lossy image compression methods.

ACKNOWLEDGMENTS

The authors would like to acknowledge the ASD Systems for their use of the Granny image sequence. This project is funded by EPSRC research grant GR/L82359.

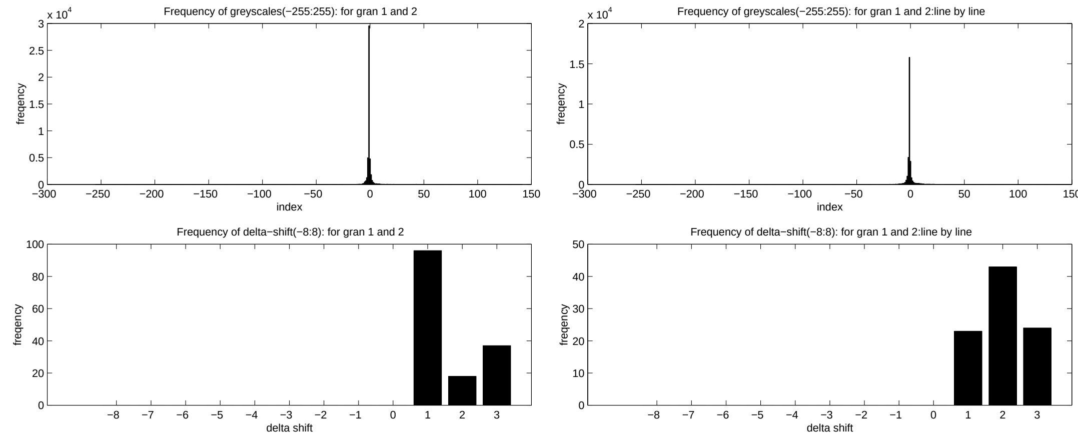

Figure 13. Frequency and delta shift distribution for images 1 and 2 for blocks and lines wrt. least mean square measure.

	Correlation Measure		LMSE Measure	
Matching Criteria	Matching Entropy	Shift Entropy/8	Matching Entropy	Shift Entropy/8
Full Image:				
wrt. row	4.7803	1.9298	4.7893	1.9298
wrt. window (1x8)	3.7374	3.0561	3.3717	2.7387
Decomposed Image:				
wrt. blocks	4.5849	1.9674	4.5849	1.9674
wrt. row	4.9394	1.7468	4.9597	1.7468
wrt. window (1x8)	3.9600	3.0463	3.5956	2.7393

Table 1. Entropy Measure for Granny images.

REFERENCES

1. J. R. Moore, N. A. Dodgson, A. R. L. Travis, and S. R. Lang. Time-multiplexed color autostereoscopic display. In *Proc. SPIE*, volume 2653, pages 10–19, 1996.
2. N. A. Dodgson, J. R. Moore, S. R. Lang, G. Martin, and P. Canepa. Time-sequential multi-projector autostereoscopic 3d display. *submitted to J. Soc. for Information Display*, 1999.
3. N. A. Dodgson, J. R. Moore, S. R. Lang, G. Martin, and P. Canepa. A 50” time-multiplexed autostereoscopic display. In *Proc. SPIE*, volume 3957, 2000.
4. G. Martin, A. Smeyne, J. R. Moore, S. R. Lang, and N. A. Dodgson. 3-d visualization without glasses - a large-screen autostereoscopic display. In *Aerosense*, 2000.

	Correlation Measure		LMSE Measure	
Matching Criteria	Matching Entropy	Shift Entropy/8	Matching Entropy	Shift Entropy/8
Full Image:				
wrt. row	2.4909	1.5256	1.5841	1.5613
wrt. window (1x8)	2.260	3.4860	2.3167	3.4340
Decomposed Image:				
wrt. blocks	2.3445	1.5556	2.4221	1.2784
wrt. row	2.8543	1.5301	2.8872	1.5206
wrt. window (1x8)	2.5720	3.3980	2.5752	3.4301

Table 2. Entropy Measure for Porsche images.

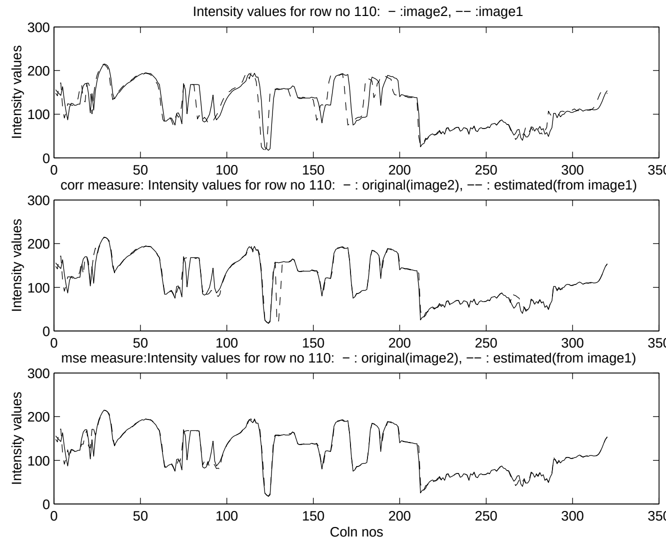

Figure 14. Comparison of the hierarchical matching of image 2 wrt. to image 1 using correlation measure and mean square measure.

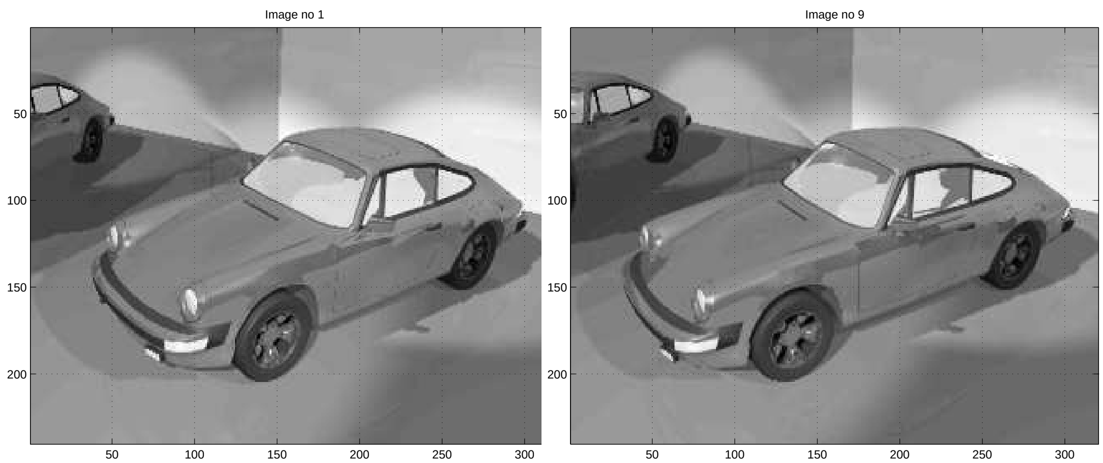

Figure 15. Images from a single view Porsche sequence.

5. H. Isonon, M. Yasuda, D. Takemori, H. Kanayama, C. Yamada, and K. Chiba. 50-inch autostereo full-color 3-d display system. In *Proc. SPIE*, volume 1669, pages 176–185, 1992.
6. H. Morishima, H. Nose, N. Taniguchi, K. Inoguchi, and S. Matsumura. Rear cross lenticular 3d display without eyeglasses. In *Proc. SPIE*, volume 3295, pages 193–202, 1998.
7. T. Okosi. *Three-Dimensional Imaging Techniques.* Academic Press, 1996.
8. A. Arimoto, T. Ooshima, T. Tani, and Y. Kaneko. Wide viewing area glassless stereoscopic display using multiple projectors. In *Proc. SPIE*, volume 3295, pages 186–192, 1998.
9. G. R. Little, C. Gustafson, and V. E. Nikolaou. Multiperspective autostereoscopic display. In *Proc. SPIE*, volume 2219, pages 388–394, 1994.
10. N. A. Dodgson and J. R. Moore. Design and implementation of an autostereoscopic camera system. Technical Report 398, University of Cambridge, June 1996.
11. N. A . Dodgson, J. R. Moore, and S. R. Lang. Time-multiplexed autostereoscopic camera systemy. In *Proc. SPIE*, volume 3012, pages 72–83, 1996.
12. M. G. Perkins. Data compression of stereopairs. *IEEE trans. on Communications*, 40(4):684–696, April 1992.

13. European Project. Race 2045 - distima. http://www.tnt.uni-hannover.de/project/eu/distima/overview.html, 1995.
14. T. Fujii and H. Harashima. Data compression of an autostereoscopic 3-d image. In *Proc. SPIE*, volume 2308, pages 930–941, 1994.
15. T. Naemura, M. Kaneko, and H. Harashima. 3-d object based coding of multi-view images. In *Picture Coding Symposium*, pages 459–464, 1996.
16. H. Aydinoglu and M.H. Hayes. Multi-view image coding using local orthogonal bases. In *Proc. SPIE*, volume 3024, pages 340–351, 1997.
17. T. Naemura, M. Kaneko, and H. Harashima. Compression and representation of 3-d images. *IEICE trans. INf. & Syst.*, E82-D(3):558–567, March 1999.