FAST VIEW INTERPOLATION OF STEREO IMAGES
USING IMAGE GRADIENT AND DISPARITY TRIANGULATION

JoonHong Park and HyunWook Park

Department of Electrical Engineering
Korea Advanced Institute of Science and Technology (KAIST)
373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea

ABSTRACT

The paper proposes a fast view interpolation method based on image gradient and disparity triangulation. The image gradient is used to select the node points for triangulation and each node point is evaluated by its matching errors and cross correspondence of the disparity values. To model the abrupt changes of disparity on the object boundaries, new node points are added along the image gradient direction. In addition, some node points are removed by consideration of unreliable matching conditions. To construct the intermediate-view images, Delaunay triangulation and image warping are performed. The experimental results show that the proposed algorithm is fast and overcomes the drawbacks of previous methods.

1. INTRODUCTION

Stereo images provide simple means of perceiving the relative depth information in a real world scene. However, three-dimensional television, which probably uses stereoscopic videos, leads to increase visual strain because of imbalance between accommodation and convergence of eyes [1]. The view synthesis technique can be used to overcome this problem [2]. The look-around capability of the view synthesis makes viewers comfortable and produces photo-realistic images.

In consideration of compression for multiview image sequences, we must consider coding efficiency, computational complexity and the ease of generating the intermediate-view images.

Fan and Ngan proposed a disparity map coding method based on adaptive triangular surface modeling [3]. Because the disparity map is modeled by a finite number of nodes, the acquired disparity map must be smooth so that the disparity error can be neglected. If the disparity map is not smooth enough, i.e., there exist some disparity discontinuities, large number of triangulation nodes must be placed very heavily around the discontinuity or noisy area to reduce the disparity error.

Sethuraman proposed a compression method of multiview image sequences using a generalized quadtree [4]. The drawback of this algorithm stems from having only one disparity value on every rectangular patch. If input image has a continuously varying disparity map, the strategy of just one disparity value per a rectangular patch makes the disparity error large.

R.-S. Wang and Y. Wang proposed mesh-based analysis and coding of multiview video sequence [5]. In their work, disparity estimation and compensation are performed in such a way that the compensation error of full frame should be minimized. Node points are iteratively moved in the direction of minimizing prediction error. Their method may be optimal for minimizing disparity prediction error. However, its computational complexity is high because of iterative procedure.

In the paper, we present a new view interpolation method for stereoscopic images, which brings together computational efficiency and high PSNR of the intermediate-view image compared to the previous view interpolation methods.

2. PROPOSED METHOD

2.1. Overview

The view interpolation algorithm can be composed of two major processes: the disparity estimation and the intermediate-view image generation. Most of the view interpolation algorithms require disparity maps with pixel-size full resolution or with reduced resolution. No matter which disparity estimation algorithms we choose, we could expect reliable results solely on textured areas of the image. However, it is difficult to find correct disparity values of homogenous or noisy regions. Thus, the disparity values of the homogeneous regions are usually interpolated by using reliable disparity values of the textured regions. If only the view interpolation and the coding of stereo images are considered, finding disparity values of homogeneous or noisy regions may be a redundant work because the view-interpolated image
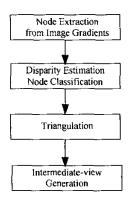
Figure 1. Overall procedure

Figure 2. Node extraction using image gradient

quality and the coding efficiency of the homogenous region are insensitive to error of the disparity values. Therefore, we try to remove these redundancies by estimating disparity values of the high gradient points only such as object boundaries.

The overall procedure of the proposed algorithm is shown in Fig. 1. The proposed view interpolation algorithm consists of two major parts: disparity estimation of the extracted nodes and intermediate-view image generation. The disparity estimation part is to extract node points in the left image, to find disparities of the extracted node points, and to classify a valid set of node points and their disparities from the extracted node points. The valid node points and their disparity values are used for synthesizing the intermediate-view images. The intermediate-view image generation part is to synthesize new intermediate-view images from two input stereo images by using the disparity information that is derived from the disparity estimation.

We assume that the stereo images used in this paper are obtained by a stereo camera and preprocessed by using rectification. Therefore, there exist horizontal disparities only. This assumption makes disparity estimation simple.

2.2. Node extraction and classification

A set $S = \{(x_i, y_i, d_i)\}$ that is composed of positions of node points and their associated disparity values is obtained at first. The set S represents geometric information of the input scene such as the node positions (x_i, y_i) that are feature points in the left image and the disparity values d_i that are features from the stereo images. The set S is used to synthesize new intermediate-view images. The detailed procedure of obtaining the set S from input stereo images is explained as follows.
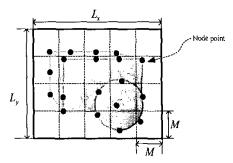
2.2.1. Extraction of node points

An initial node set S_0 is extracted from the original left image as shown in Fig. 2. Because the edges that can be representatives of objects in an image are usually located at high gradient regions, we use the image gradient to find the initial node points. To distribute the node points evenly with positioning at high gradient region, we divide the left image into small blocks of $M \times M$. At every partitioned block, a node is placed at the highest gradient point. However, we exclude node points in homogeneous blocks since disparity values of the homogenous region are redundant as well as vulnerable to noise for view interpolation. If the maximum gradient of a block is larger than a pre-defined threshold value T_G, the block has a node point. On the other hand, if the maximum gradient of the block is smaller than the threshold value, the block is considered as a homogeneous region that does not have a node point. The total number of node points is dependent on the threshold value T_G and the block size of $M \times M$.

2.2.2. Disparity estimation

The disparity estimation is to find correspondences between two stereo images. It has no general solution and it is very difficult to estimate correct disparities because of occlusion, camera noise, and homogeneous regions. Some of node points may be occluded in the right image, may be on the boundary of two different objects with different disparity values, or may not have reliable disparity values due to errors. To identify and estimate the correct disparity, the obtained disparity values from simple block matching algorithm are evaluated. In this paper, we use the block size of 5×5 for block matching. For validity check of the obtained disparity, the disparity from the right image to the left image is also obtained. The two disparity values of both left-to-right and right-to-left estimates are compared to identify the correct disparity.

2.2.3. Construct a valid node set S from the initial node set S_0

As explained previously, disparity values may not be reliable because of various erroneous conditions. Therefore, classification of nodes is required. To do this, matching error of disparity estimation and cross correspondence of disparity values are evaluated. If the minimum matching error of disparity estimation is smaller than a threshold and the cross correspondence is valid, it is called 'matched' and the node point in the left image has a corresponding point in the right image.

Finally, the valid node set $S = \{(x_i, y_i, d_i)\}$ is obtained as follows. If a node is '*matched*', it is added to the set S. However, if the node is not '*matched*', the node is tested for whether it is on the object boundary. The disparity values can abruptly vary around the object boundary. This abrupt change of the disparity values degrades the performance of the intermediate-view interpolation. In order to reduce this problem, two adjacent node points around the object boundary are placed along the image gradient direction. The two adjacent node points are placed with distance of two pixels in horizontal direction for horizontal gradient edge or in vertical direction for vertical gradient edge. Then, the matching process is applied to two adjacent node points. In the case that both disparities are invalid, there is no addition of the node point to S. This case is when the disparity is ambiguous because of various reasons such as occlusion, repeated texture, erroneous region, etc. Meanwhile, if one or two disparities of the two adjacent node points are valid, then the valid nodes are added to S.

2.3. Triangulation and intermediate-view generation

Delaunay triangulation is used to partition each stereo image into triangular patches using the set of nodes. An important property of Delaunay triangulation is that the triangulation with an arbitrary set of points in a plane is unique under simple constraints [6]. In other words, the Delaunay triangulation can be exactly reconstructed by the set of points without the connection information. This is an advantage of triangular mesh coding [3].

After partitioning the left and right images by Delaunay triangulation, the intermediate images are obtained by warping every triangular patch into the intermediate-view plane.

Because Delaunay triangulation is used for modeling the true disparity map by a set of nodes, two neighboring nodes with different disparities could approximate the discontinuity of disparities well.

3. EXPERIMENTAL RESULT

To verify the performance of the proposed view-interpolation method, the computation time, the objective

quality of the intermediate-view image and the prediction accuracy of the right image from the left image were analyzed. We used three images to test our algorithm and their results were compared with those of the Fan and Ngan method, the generalized quadtree method and Wang's method. In this experiment, a 1 GHz Pentium-III system equipped with 512 MByte RAM was used.

(a) Corridor image

(b) Aqua image   (c) Tunnel image

Figure. 3 Left images of stereo images used for experiments

The test stereo images are shown in Fig. 3. The 'corridor' image is a synthetic one and any intermediate-view images can be generated exactly. Therefore, this image set is used for comparing the objective quality (PSNR) of the intermediate-view images.

	Fan and Ngan [3] (1500)	Generalized quadtree [4] (1400)	Wang [5] (1500)	Proposed method (1400)
Average computation time	10 sec.	0.33 sec.	5 sec.	0.35 sec.

Figure 4. Computation time

The execution times of three algorithms are shown in Fig. 4. Fig. 4 shows that the Fan and Ngan method and Wang's method takes longer than the other two methods. It takes long time for procedure. In addition, the PSNR curve of the intermediate-view images of 'corridor' image is shown in Fig. 5 (a). This show that the proposed method is comparable to the Fan and Ngan method, and it surpasses the generalized quadtree method and Wang's method.

The prediction performances of three stereo image sets are shown in Figs. 5(b), (c) and (d). In the case of

'corridor' image, the performance of the proposed method is similar to the Fan and Ngan method and better than other two methods. The PSNR of the quadtree method is worst since the 'corridor' image has a gradually varying disparity map. In the cases of the 'aqua' and 'tunnel' images, the PSNR of the Fan and Ngan method is worse than other methods since these images have many disparity discontinuities of object boundaries.

The Fan and Ngan method and Wang's method are good for stereo images of gradually varying disparity because mesh-based approach without considering occlusion is efficient when modeling such a gradually varying disparity map. However, it is inefficient for images with disparity discontinuities. In case of the generalized quadtree algorithm, it uses rectangular patches with single disparity values so that it has advantages for the modeling of disparity discontinuities and disadvantages for modeling gradually varying disparity. However, the proposed method produces good results in both cases. By node point extraction with consideration of image gradients and disparity discontinuities, the proposed algorithm overcomes the drawbacks of the previous methods.
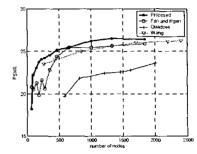
4. CONCLUSIONS

In this paper, we present a new view interpolation method for stereo images that brings together computational efficiency and accurate interpolation compared to previous view interpolation methods. The experimental results show that the proposed method is computationally efficient and produces better quality of the intermediate-view images and the predicted images than the previous view-interpolation methods.
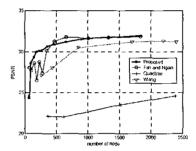
5. REFERENCES

[1] S. Pastoor, "3D-television: A survey of recent research results on subjective requirements," *Signal Processing: Image Communication*, Vol. 4, No.1, pp.21-32, Nov. 1991.

[2] S.E. Chen and L. Williams, "View interpolation for image synthesis," *SIGGRAPH'* 93, pp 279-288, 1993.

[3] H. Fan and K. N. Ngan, "Disparity map coding based on adaptive triangular surface modeling," *Signal Processing: Image Communication*, Vol. 14, No. 2, pp. 119-130, 1998.

[4] S. Sethuraman, "Stereoscopic image sequence compression using multiresolution and Quadtree decomposition based disparity- and motion-adaptive segmentation," Ph.D Dissertation, Carnegie Mellon Univ., 1996.

[5] R.-S. Wang and Y. Wang, "Multiview Video Sequence Analysis, Compression, and Virtual Viewpoint Synthesis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, No.3, pp. 397-410, 2000.
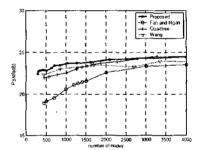
[6] L. Guibas and J. Stolfi, "Primitives for the Manipulation of General Subdivisions and the Computation of Voronoi Diagrams," *ACM Transactions on Graphics*, Vol.4, No.2, pp. 74-123, April 1985.
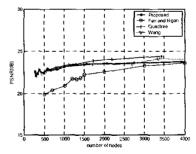
(a) PSNR graph of intermediate-view images (corridor)

(b) PSNR graph of predicted images (corridor)

(c) PSNR graph of predicted images (aqua)

(d) PSNR graph of predicted images (tunnel)

Figure 5. PSNR of the predicted mages from various view interpolation methods.