Available online at www.sciencedirect.com
SCIENCE DIRECT®
Signal Processing: *Image Communication* 18 (2003) 401–416
SIGNAL PROCESSING: IMAGE COMMUNICATION
www.elsevier.com/locate/image

Fast view interpolation of stereo images using image gradient and disparity triangulation

Joon Hong Park, Hyun Wook Park*

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

Received 25 June 2002; received in revised form 20 November 2002; accepted 18 January 2003

Abstract

This 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 of the stereo image disparity. Subsequently, 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 such as occlusion, homogeneous regions, erroneous regions, etc. To construct the intermediate-view images, Delaunay triangulation of the stereo image disparity and image warping with the corresponding disparity are performed. The experimental results show that the proposed algorithm is fast and overcomes the drawbacks of previous methods that use triangular or rectangular patches.

Keywords: View interpolation; Triangulation; Disparity estimation; Stereo image; Image gradient

1. Introduction

Stereo images provide simple means of perceiving the relative depth information in a real world scene. However, 3D television, which probably uses stereoscopic videos, leads to increase visual strain because of imbalance between accommodation and convergence of eyes [6]. 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.

There are generally two main categories of view synthesis algorithms: reconstruction-based methods and interpolation-based methods [5]. Reconstruction-based methods use explicit or implicit 3D structures of the scene to synthesize new views through a fundamental matrix or a trilinear tensor. However, interpolation-based methods do not require 3D structures or camera parameters. The interpolation-based methods are able to generate smooth transitions between reference images by simple interpolation from two stereoscopic images.

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

*Corresponding author. Tel.: +82-42-869-3466; fax: +82-42-869-3410.
E-mail address: hwpark@athena.kaist.ac.kr (H.W. Park).

0923-5965/03/$ - see front matter
doi:10.1016/S0923-5965(03)00013-4

interpolation can be divided into two processes: disparity estimation and intermediate-view generation. The view interpolation of the stereo images may be thought as the extended version of the motion estimation/compensation problem since the disparity of the stereo images can be considered as the motion vector of the monocular moving pictures. Therefore, the motion estimation algorithm of two successive frames in the monocular video could be also applicable to find disparity values of two stereo images [7]. Those approaches of various motion estimation methods can be directly applied to stereo image coding since their main focus is to minimize coded bits and prediction error. However, those approaches of minimizing prediction error has limitations in generating of intermediate view images, and the interpolated images usually have visible artifacts since those are mainly focused on coding. Therefore, finding exact disparities of stereo images is an important work in view interpolation.

Stereo image pairs are obtained using simultaneously recording a scene with two cameras at different positions. The relative position, orientation, and some additional parameters of the cameras generate the disparity vectors of the stereo image pairs [7]. When two pinhole cameras are placed with equal orientation while their positions differ only in direction of the scan line, it is called parallel setup of stereo cameras. This case makes the disparity vector is a one-dimensional value. Therefore, finding disparity vector in stereo image pair is different than finding motion vector in monocular video sequences according to the camera setup.

Fan and Ngan proposed a coding method of disparity map based on adaptive triangular surface modeling [3]. This algorithm consists of two stages: to find a smooth disparity map using block-based hierarchical disparity estimation and to model the acquired disparity map by Delaunay triangulation on a set of nodes. It compresses the set of nodes with disparities by the differential pulse coded modulation and variable length coding. 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 [8]. By partitioning a reference image successively by the generalized quadtree decomposition and finding a disparity value for the partitioned rectangular patch, he achieved very high compression efficiency. However, the drawback of this algorithm stems from having only one disparity value on every rectangular patch. If input stereo images have a continuously varying disparity map, the strategy of just one disparity value per a rectangular patch makes the disparity error large.

Wang and Wang proposed mesh-based analysis and coding of multiview video sequence [9]. In their work, disparity estimation and compensation were performed in such a way that the compensation error of full frame should be minimized. Node points were iteratively moved in the direction of minimizing the prediction error. They proposed a full search method and computationally efficient fast search method also. However, its computational complexity is high because of iterative procedure. In addition, they did not consider the occlusion problem.

In this paper, we present a new view interpolation method for stereoscopic images that brings together computational efficiency and high PSNR of the intermediate-view image compared to the previous view interpolation methods. In the paper, Section 2 shows the details of the proposed method. In Section 3, experimental results are described and some examples are shown. Conclusions are given in Section 4.

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. For example, the reduced-resolution disparity map is required for Fan and Ngan's method, whereas the full-resolution disparity map is required for the generalized quadtree method. 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 homogeneous or noisy regions. Thus, the disparity values of the homogeneous regions are usually interpolated by using reliable disparity values of the textured regions. If the view interpolation and the coding of stereo images are considered together, finding disparity values of homogeneous or noisy regions may be a redundant work because the view-interpolated image quality and the coding efficiency of the homogeneous region are insensitive to error of the disparity values. Therefore, we try to remove these redundancies by estimating disparity values of the high gradient areas only such as object boundaries.

The overall procedure of the proposed algorithm is shown in Fig. 1. The proposed view interpolation algorithm also consists of 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 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.

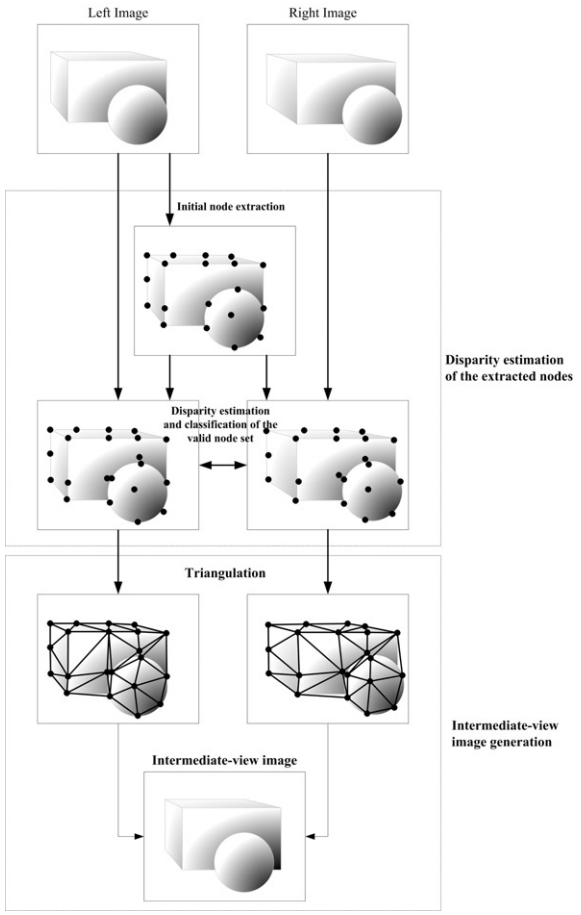

Fig. 1. Overall procedure of the proposed view-interpolation method.

2.2. Disparity estimation of the extracted nodes

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 of 3-D structure. 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. 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.

The left image is divided into multiple blocks with a block size of $M \times M$ and the highest gradient point at each divided block is selected as shown in Fig. 2. To distribute the node points evenly with positioning at high gradient region, we divide the left image into small blocks. Assume that the left and right images are $f_L(x,y)$ and $f_R(x,y)$, respectively, and the size of the input images is $L_x \times L_y$. The gradient of the left image $G_L(x,y)$ is obtained by applying the Sobel operator as follows:

$$\begin{aligned} G_L(x,y) &= G_{L_x}(x,y) + G_{L_y}(x,y), \\ G_{L_x}(x,y) &= |f_L(x,y) \otimes H_x|, \\ G_{L_y}(x,y) &= |f_L(x,y) \otimes H_y|, \end{aligned} \tag{1}$$

where

$$H_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}, \quad H_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix},$$

and the operator $\otimes$ means the linear convolution.

If we define the (k,l)th block of the image as

$$\begin{aligned} \mathbf{B}_{k,l} = \{(x,y) \,|\, &(k-1) \cdot M \leqslant x < k \cdot M, \\ &(l-1) \cdot M \leqslant y < l \cdot M\}, \end{aligned} \tag{2}$$

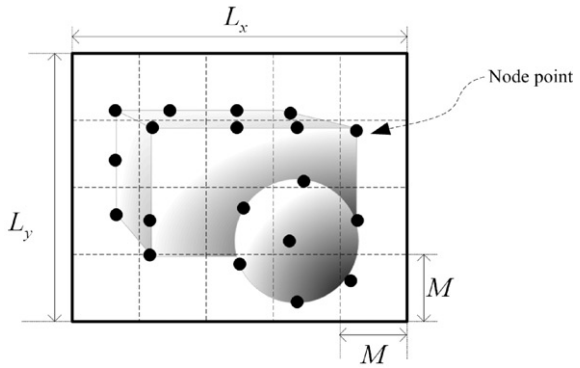

Fig. 2. Extraction of node points from gradient image.

the initial node set $S_0 = \{(x_i, y_i)\}$ is obtained as follows:

$$\begin{aligned} &\text{if} \left(\max_{(x,y) \in \mathbf{B}_{k,l}} G_L(x,y) \geqslant T_G \right), \\ &\quad S_0 \leftarrow \arg\max_{(x,y) \in \mathbf{B}_{k,l}} G_L(x,y) \\ &\qquad \text{for } 0 \leqslant k < L_x/M, \quad 0 \leqslant l < L_y/M. \end{aligned} \tag{3}$$

As shown in Eq. (3), we exclude node points in homogeneous blocks since disparity values of the homogeneous region are redundant as well as vulnerable to noise for view interpolation. If the maximum gradient of a block is larger than a predefined 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 $M \times M$. The relation between the number of node points and T_G or M is described in the experimental results.

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 as follows.

Assume that the block size of the block matching algorithm is $(2l_x+1) \times (2l_y+1)$, a block centered at (x_0, y_0) of the left and the right images can be specified as

$$\begin{aligned} \mathbf{B}_L(x_0,y_0) \equiv \{&f_L(x,y) \,|\, x_0 - l_x \leqslant x \leqslant x_0 + l_x, \\ &y_0 - l_y \leqslant y \leqslant y_0 + l_y\}, \\ \mathbf{B}_R(x_0,y_0) \equiv \{&f_R(x,y) \,|\, x_0 - l_x \leqslant x \leqslant x_0 + l_x, \\ &y_0 - l_y \leqslant y \leqslant y_0 + l_y\}. \end{aligned} \tag{4}$$

In this paper, we use the block size of 5×5 for block matching. The matching error at (x, y) with a horizontal disparity d is defined by a normalized difference between two blocks $\mathbf{B}_L(x, y)$ and $\mathbf{B}_R(x + d, y)$ as follows:

$$\varepsilon^d_{L\to R}(x,y) \equiv \mathrm{diff}(\mathbf{B}_L(x,y), \mathbf{B}_R(x+d,y)) = \frac{\sum_{j=-l_y}^{l_y}\sum_{i=-l_x}^{l_x} |f_L(x+i,y+j) - f_R(x+d+i,v+j)|}{\sum_{j=-l_y}^{l_y}\sum_{i=-l_x}^{l_x} |f_L(x+i,y+j) + f_R(x+d+i,v+j)|}, \quad (5)$$

where the subscript $L\to R$ designates that the disparity is estimated from the left image to the right image and diff$(\cdot)$ denotes the normalized difference function. Then, the estimated disparities from the left image to the right image can be obtained as follows:

$$D_{L\to R}(x,y) = \underset{d\in[D_{\min},D_{\max}]}{\arg\min}\ \varepsilon^d_{L\to R}(x,y), \quad (6)$$

where $D_{\min}$ and $D_{\max}$ are minimum and maximum allowable disparities, respectively. For validity check of the obtained disparity, the disparity from the right image to the left image is also obtained as follows:

$$D_{R\to L}(x,y) = \underset{d\in[-D_{\max},-D_{\min}]}{\arg\min}\ \varepsilon^d_{R\to L}(x,y), \quad (7)$$

where

$$\varepsilon^d_{R\to L}(x,y) \equiv \mathrm{diff}(\mathbf{B}_R(x,y), \mathbf{B}_L(x+d,y)).$$

The two disparity values of $D_{L\to R}(x, y)$ and $D_{R\to L}(x, y)$ are used for classification of the correct disparity in the following subsection.

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

Before introducing the procedure of the node classification, two functions are defined: the minimum matching error (MME) and the cross correspondence (CC). MME measures how much two points, (x, y) in the left image and its corresponding point $(x + d, y)$ in the right image, are similar, which is defined as follows:

$$\mathrm{MME}(x,y) = \varepsilon^d_{L\to R}(x,y)|_{d=D_{L\to R}(x,y)}. \quad (8)$$

CC is the measure used to show whether the acquired disparity is valid or not by comparing two disparities of $D_{L\to R}$ and $D_{R\to L}$ as follows:

$$\mathrm{CC}(x,y) = \begin{cases} 1 & \text{if } |D_{L\to R}(x,y) + D_{R\to L}(x + D_{L\to R}(x,y), y)| \\ & \quad \leqslant 1, \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

If the minimum matching error is smaller than a threshold (i.e., $\mathrm{MME} < T_E$) and the cross correspondence is valid (i.e., $\mathrm{CC} = 1$), it is called 'matched' and the node point in the left image has a corresponding point in the right image. That is, the decision of matching is described as follows:

$$IsMatched(x,y) = \begin{cases} 1 & \text{if } \mathrm{MME}(x,y) \leqslant T_E \text{ and } \mathrm{CC}(x,y) = 1, \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

Finally, the valid node set $S = \{(x_i, y_i, d_i)\}$ is obtained by using MME, CC, and gradient direction, whose pseudocode is described in Fig. 3. If a node is '*matched*', it is added to the

```
Initialize the valid node set and disparities to empty one: S = ∅
for all initial node points (x_i, y_i) ∈ S_0 where 0 ≤ i < n(S_0).
{
  if ( IsMatched(x_i, y_i) = 1 )
      S ← (x_i, y_i, D_{L→R}(x_i, y_i))
  else if ( G_{L_x}(x,y) < G_{L_y}(x,y) ) {   // vertical gradient edge
      if ( IsMatched(x_i − l_x, y_i) = 1 )
          S ← (x_i − 1, y_i, D_{L→R}(x_i − l_x, y_i))
      if ( IsMatched(x_i + l_x, y_i) = 1 )
          S ← (x_i + 1, y_i, D_{L→R}(x_i + l_x, y_i))
  }else if ( G_{L_x}(x,y) ≥ G_{L_y}(x,y) ){   // horizontal gradient edge
      if ( IsMatched(x_i, y_i − l_y) = 1 )
          S ← (x_i, y_i − 1, D_{L→R}(x_i, y_i − l_y))
      if ( IsMatched(x_i, y_i + l_y) = 1)
          S ← (x_i, y_i + 1, D_{L→R}(x_i, y_i + l_y))
  }
}
```

Fig. 3. Pseudocode for construction of the node set S from the initial node set S_0.

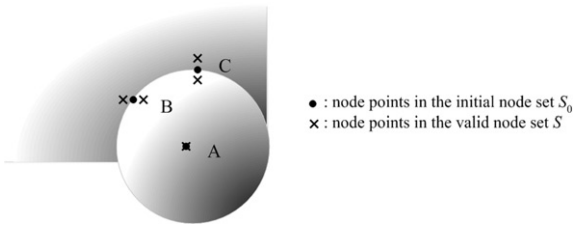

Fig. 4. Placement of the node points around object boundary: textured region inside an object (A), horizontal gradient edge (B), vertical gradient edge (C).

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 (Fig. 4). Then, the matching process in Eq. (10) 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.

Two neighboring nodes with different disparities around the object boundaries approximate the disparity discontinuity properly. 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 as shown in Fig. 5.

In the field of mesh-based motion estimation/compensation which deals only with monocular video sequences, the occluded areas make the motion compensation inefficient like stereo coding. For dealing with this occlusion problem, Altunbasak and Tekalp proposed occlusion adaptive and content-based mesh design method in

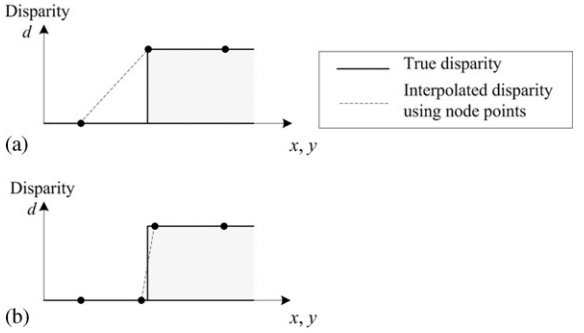

Fig. 5. Effect of adding a new node point on disparity discontinuity. (a) Before adding a node. (b) After adding a node.

coding of a monocular video sequence [1]. In his work, two regions of the background to be covered (BTBC) region and the model failure (MF) region are detected and new mesh nodes are eliminated or inserted into the previous node set. Insertion and deletion of nodes are very similar to our approaches. However, they approximate the occluded region using polygons, whereas we assign two adjacent nodes around the object boundaries to manipulate the occlusions.

2.3. Intermediate-view generation

Delaunay triangulator partitions the left image and the right image into triangular patches with the valid node set S. Two warped images are generated by warping two input images using the obtained disparity values and the triangular patches. Finally, two warped images are averaged to produce a new intermediate-view image.

2.3.1. Delaunay triangulation [4]

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. 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].

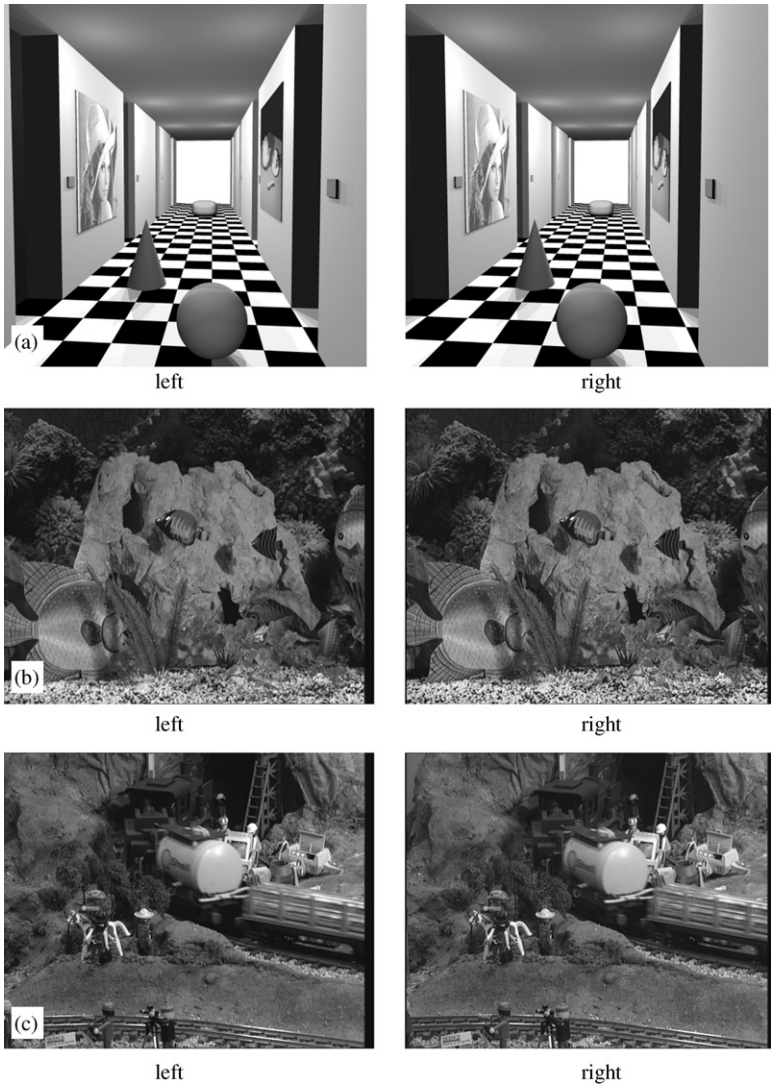

Fig. 6. Test stereo images: (a) corridor images, (b) aqua images, and (c) tunnel images.

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

2.3.2. Warp two images into the intermediate-view image

To make a warped triangle, we assume that the source triangle has three vertices of (x_1, y_1), (x_2, y_2) and (x_3, y_3), and the target triangle has the corresponding three vertices of (u_1, v_1), (u_2, v_2) and (u_3, v_3). Then any arbitrary pixel point (u, v) inside the target triangle is mapped from the point (x, y) inside the source triangle, which can be calculated as follows [10]:

$$x = \sum_{i=1}^{3} (\alpha_i + \beta_i \cdot u + \gamma \cdot_i v) \cdot x_i,$$

$$y = \sum_{i=1}^{3} (\alpha_i + \beta_i \cdot u + \gamma \cdot_i v) \cdot y_i, \tag{11}$$

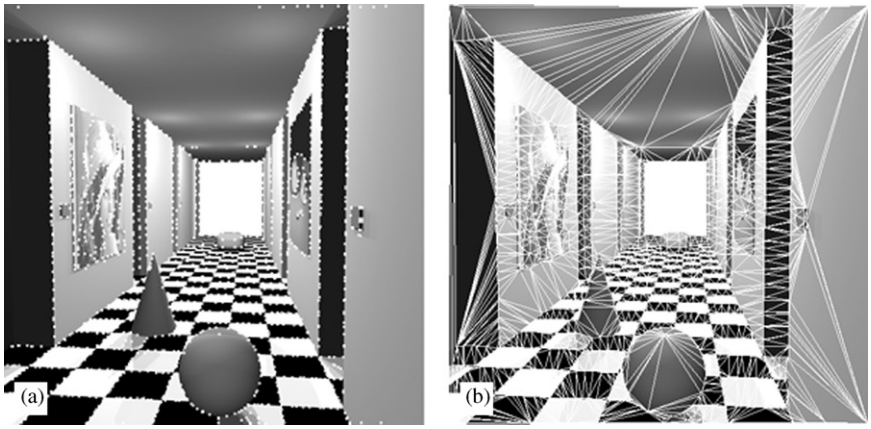

Fig. 7. (a) Extracted node points of the left Corridor image, and (b) Delaunay triangulation of the nodes ($N = 1361$).

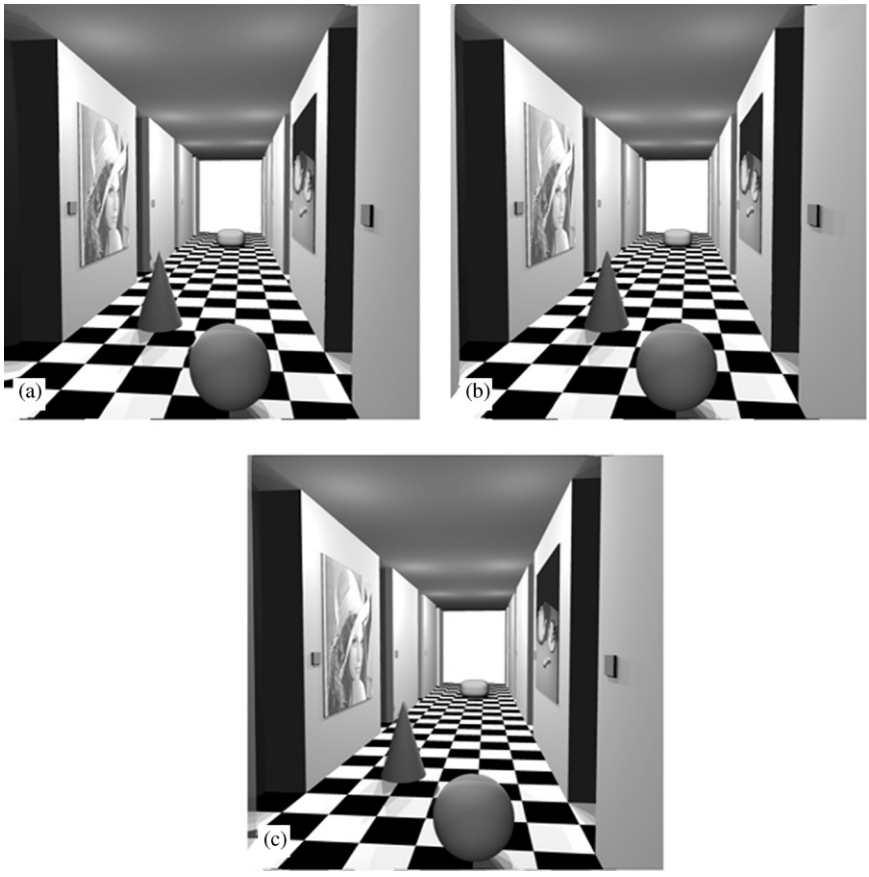

Fig. 8. Warped images and final intermediate-view image of the proposed method ($N = 1361$) for corridor image. (a) Warped image from the left image. (b) Warped image from the right image. (c) Final intermediate-view image by combining two warped images (33.78 dB).

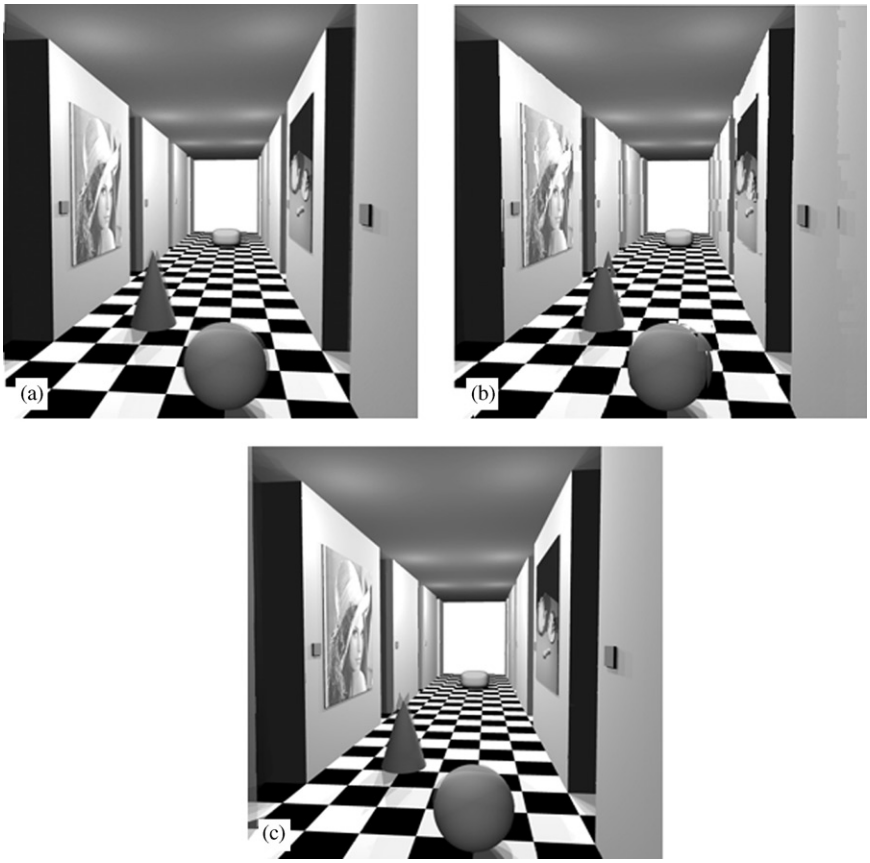

Fig. 9. Intermediate-view images from (a) Fan and Ngan's method (32.9 dB) and (b) the generalized quadtree method (23.5 dB) and (c) Wang and Wang's method (26.5 dB).

where $\begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \gamma_1 & \gamma_2 & \gamma_3 \end{bmatrix} = \begin{bmatrix} 1 & u_1 & v_1 \\ 1 & u_2 & v_2 \\ 1 & u_3 & v_3 \end{bmatrix}^{-1}$. By using Eq. (11), the corresponding pixel position (x, y) to the pixel (u, v) in the target triangle is computed, where (x, y) is not usually an integer value for the corresponding integer pixel (u, v). Thus, the interpolation process is required to obtain the pixel value at (x, y) in the source image that is mapped to the pixel value at (u, v) in the target image.

Since we assumed only horizontal disparity in the paper, three vertices of the target triangle (u_i, v_i) are written as $(x_i + \lambda \cdot d_i, y_i)$ for $i = 1, 2, 3$, where the λ value corresponds to the camera position of the intermediate view. The intermediate-view image is the right image when $\lambda = 1$, the left image when $\lambda = 0$, and the center of two images when $\lambda = 0.5$.

Finally, averaging two warped images from the left and right images, the intermediate-view image is finally obtained.

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. Applications of the proposed method can be the view interpolation of stereo images as well as the compression of stereo images. For efficient compression of the

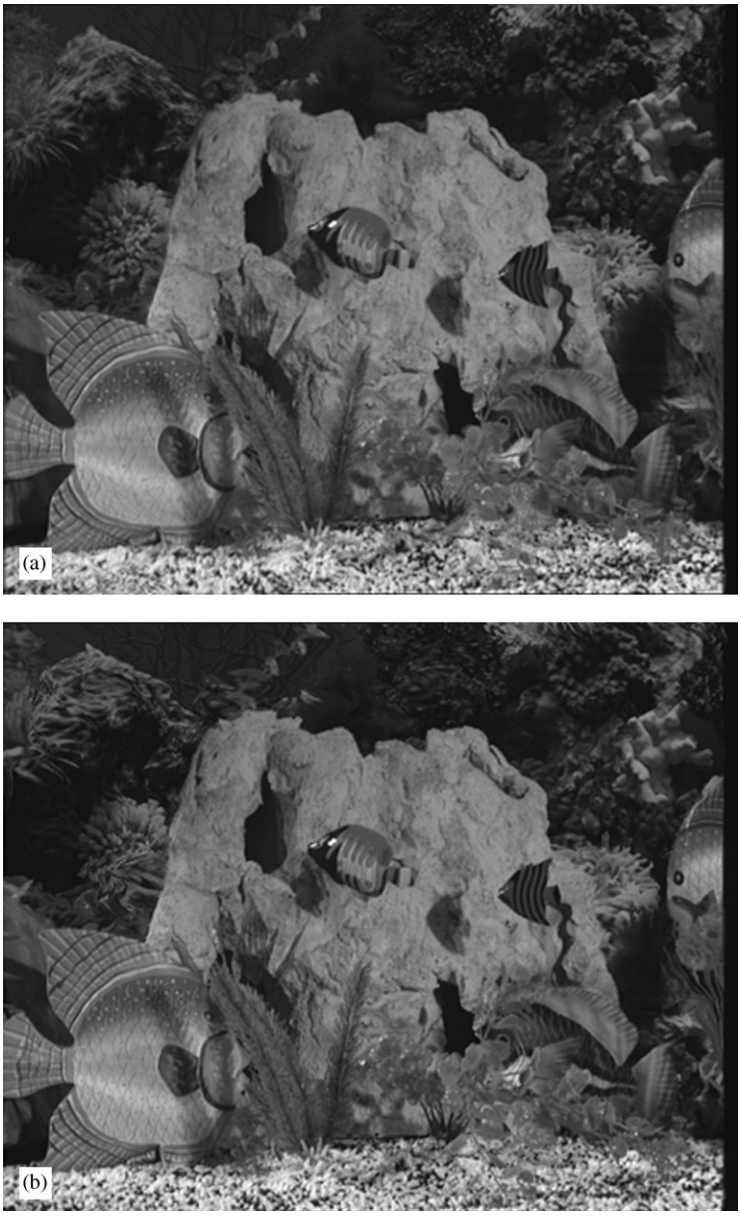

Fig. 10. Intermediate-view image from (a) the proposed method (Aqua, $N = 1580$) and (b) Wang and Wang's method (Aqua, $N = 1708$).

stereo images, the right image (left image) should be predicted as accurately as possible from the left image (right image) and the disparity values. Thus, we analyzed the prediction accuracy also. We used three images to test our algorithm and their results were compared with those of the Fan and Ngan's method [3], the generalized quadtree method [8] and Wang and Wang's method [9]. In the experiments, a 1 GHz Pentium-III system equipped with 512 MByte RAM was used.

The test stereo images are shown in Fig. 6. The 'corridor' image is a synthetic one and any

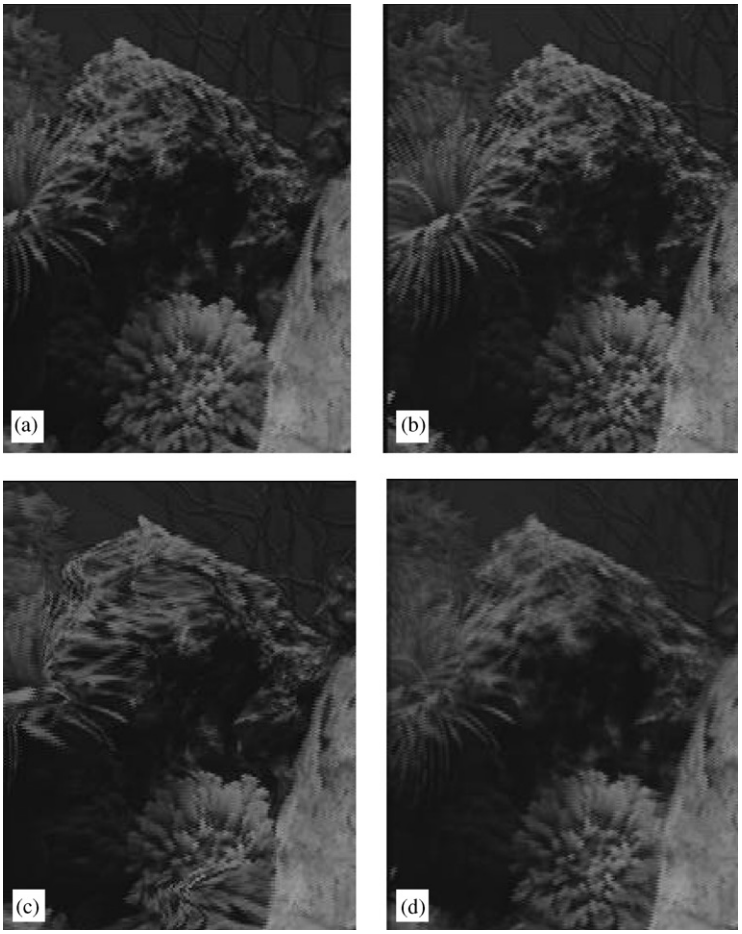

Fig. 11. Top-left regions of original and intermediate-view images for subjective visual comparison: (a) the original left image, (b) the original right image, (c) the intermediate-view image from Wang and Wang's method, and (d) the intermediate-view image from the proposed method.

intermediate-view images can be generated exactly. Therefore, this image set is used for comparing the objective quality (PSNR) of the intermediate-view images. The extracted node points and the triangulation of the left image are shown in Fig. 7 when M is eight, T_G is 170 and T_E is 0.1. The two warped images from the left and right images and the final intermediate-view image from the proposed method are shown in Fig. 8. In addition, the intermediate-view images of three other methods are shown in Fig. 9. The intermediate-view images of 'aqua' image set from the proposed method and the Wang and Wang's method are also shown in Fig. 10 and their partially magnified images are shown in Fig. 11 for visual comparison. All the intermediate-view images in Figs. 8–11 are obtained at the center view with λ of 0.5. The top-left regions of original left and right images are shown in Figs. 11(a) and (b), and the corresponding portions of the intermediate-view images from the proposed method and the Wang and Wang's methods are shown in Fig. 11(c) and (d), respectively. By comparing two intermediate-view images in considering the original left and right images, the intermediate-view image from the proposed method is more natural and looks better than the Wang and Wang's method.

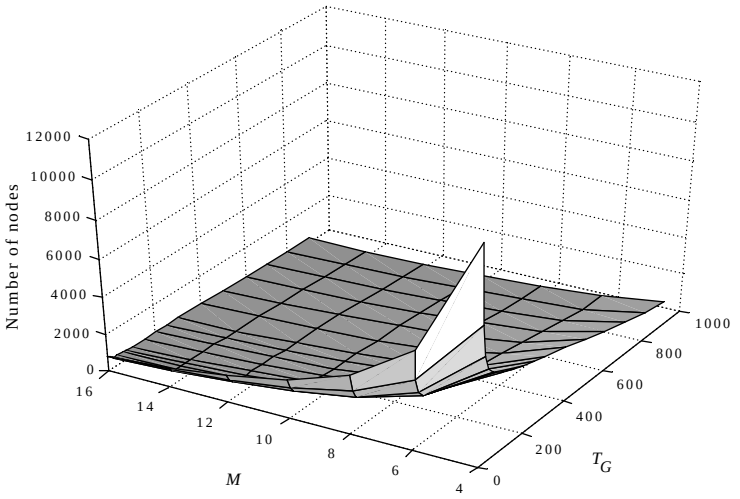

Fig. 12. Number of nodes with respect to the block size M and T_G when $T_E = 0.1$.

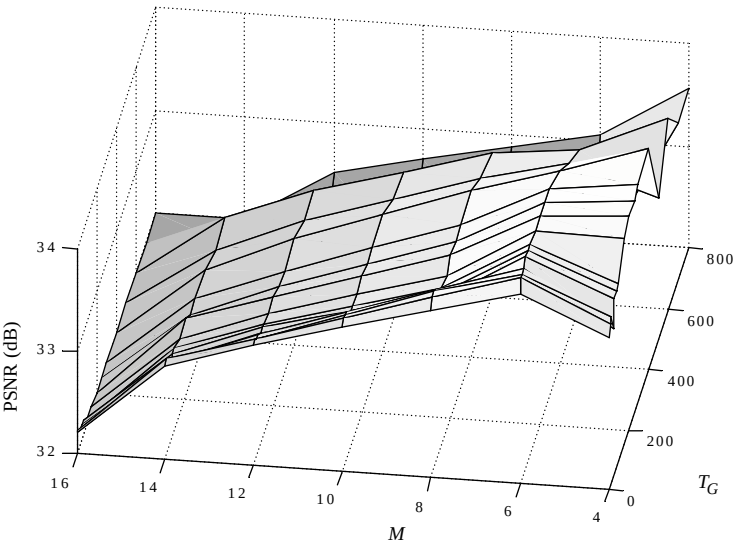

Fig. 13. PSNR of the intermediate-view image with respect to the block size M and T_G when $T_E = 0.1$.

The number of node points according to the block size M and the threshold value T_G are shown in Fig. 12 for the 'corridor' image. These two parameters could control the number of node points. The PSNR of the intermediate-view image according to the block size M and the threshold value T_G are shown in Fig. 13. This figure shows that the PSNR of the intermediate-view image is insensitive to the threshold T_G in the range of $50 \leqslant T_G \leqslant 600$. In addition, the quality of the intermediate-view image according to the threshold value T_E is shown in Fig. 14. The threshold value T_E is around 0.1 where the high PSNR occurs.

In the experiments for PSNR comparison, the block size for node extraction was 8×8, T_G was 170 and the threshold T_E was 0.1. Then the number of node points was 1361 for the corridor image.

The computation times of four algorithms are shown in Table 1. Table 1 shows that the Fan and Ngan's method takes longer computation time

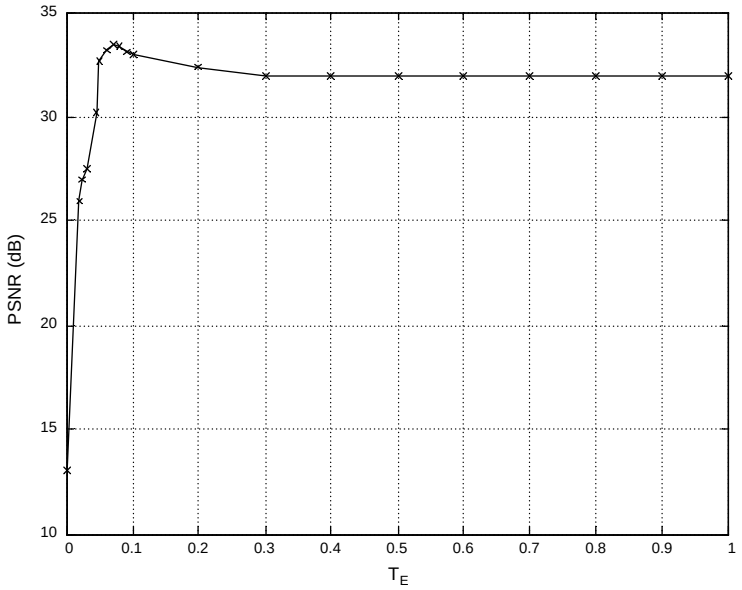

Fig. 14. PSNR of the intermediate-view image with respect to T_E when $T_G = 170$ and $M = 8$.

Table 1
Computation times of various view-interpolation methods for the 'corridor' image

	Generalized quadtree (1400)	Fan and Ngan (1500)	Wang and Wang (1500)	Proposed method (1400)
Average computation time	0.33 s	10 s	5 s	0.35 s

The numbers in parenthesis are the number of nodes (Fan and Ngan, Wang and Wang, and the proposed methods) and the number of rectangular patches (generalized quadtree method), respectively.

than the other three methods. It takes long time for iterative interpolation of the disparity map and finding the position of minimum error, and inserting a new vertex. In addition, the PSNR curve of the intermediate-view images is shown in Fig. 15. These show that the proposed method shows the best PSNR among four methods.

The prediction performances of three image sets are shown in Figs. 16–18. The PSNR of the quadtree method is worst for 'corridor' image since the 'corridor' image has a gradually varying disparity map. In the case of the 'aqua' and 'tunnel' images, the PSNRs of the Fan and Ngan's method are worse than other methods since these images have many disparity discontinuities of object boundaries.

The Fan and Ngan's method is good for stereo images of gradually varying disparity because triangulation of disparity 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.

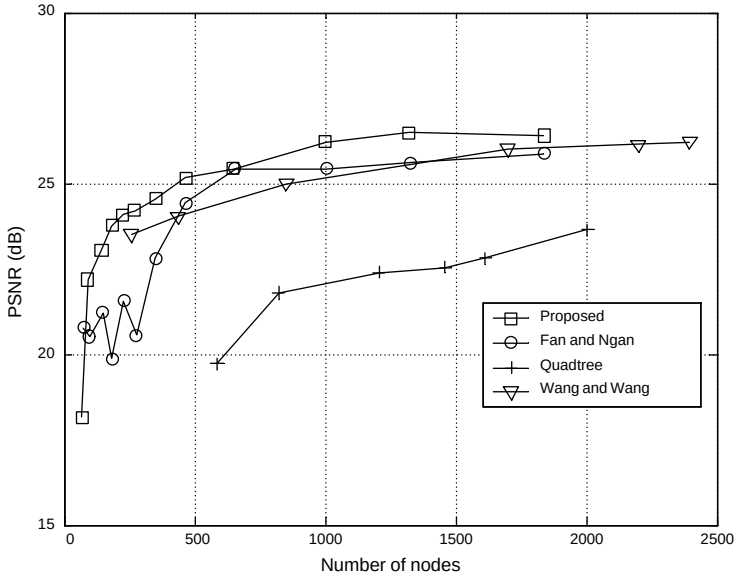

Fig. 15. PSNR curve of the intermediate-view image for 'corridor' image.

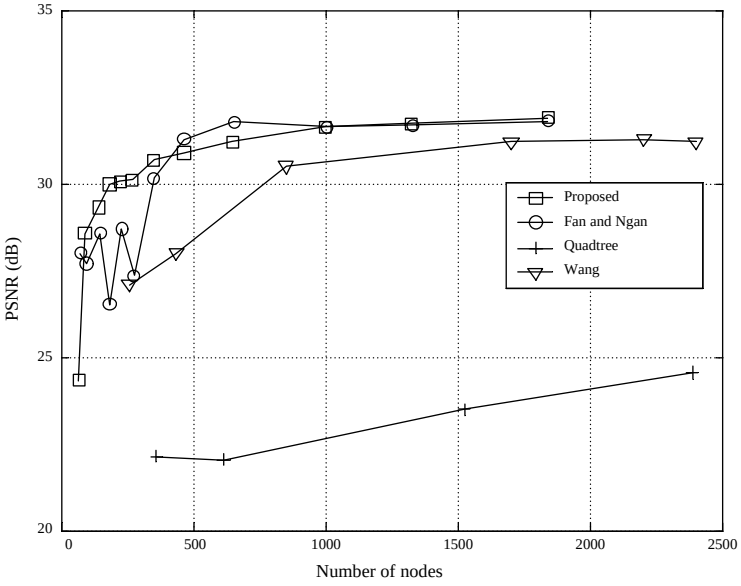

Fig. 16. PSNR curve of the predicted right image from left image for 'corridor' image.

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 proposed algorithm is composed of the node extraction on image gradient, node classification with reliability measures (the minimum matching error, cross correspondence, and

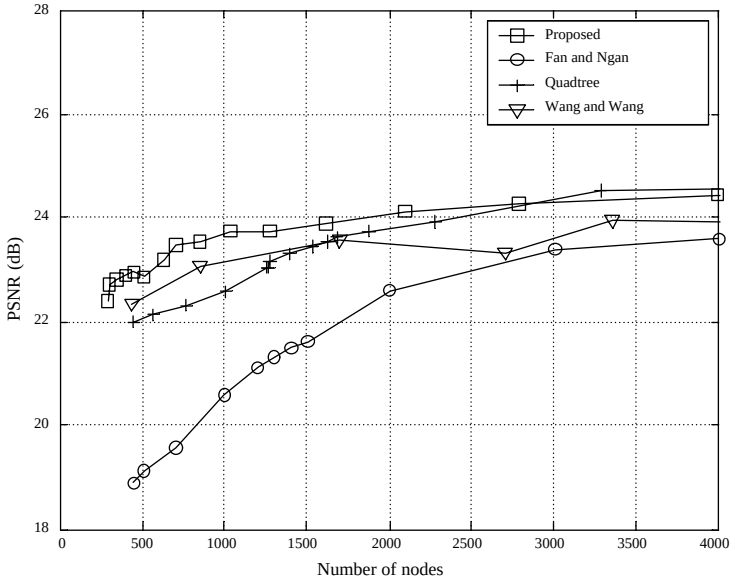

Fig. 17. PSNR curve of the predicted right image from left image for 'aqua' image.

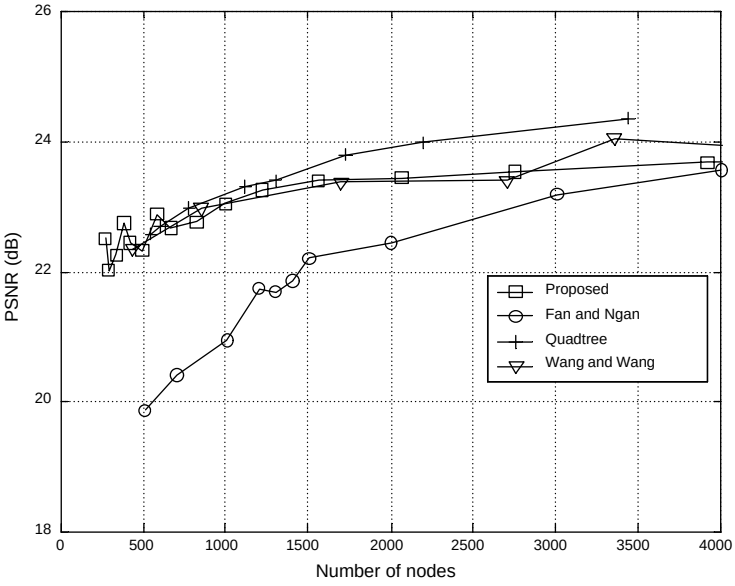

Fig. 18. PSNR curve of the predicted right image from left image for 'tunnel' image.

gradient direction), view interpolation by combining two warped images that are projected from the left and right images with consideration of occlusions. 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.

References

[1] Y. Altunbasak, A.M. Tekalp, Occlusion-adaptive, content-based mesh design and forward tracking, IEEE Trans. Image Process. 6 (9) (September 1997) 1270–1280.

[2] S.E. Chen, L. Williams, View interpolation for image synthesis, Proceedings of the SIGGRAPH '93, Anaheim, CA, 1–6 August, 1993, pp. 279–288.

[3] H. Fan, K.N. Ngan, Disparity map coding based on adaptive triangular surface modeling, Signal Process.: Image Commun. 14 (2) (November 1998) 119–130.

[4] L. Guibas, J. Stolfi, Primitives for the manipulation of general subdivisions and the computation of Voronoi diagrams, ACM Trans. Graphics 4 (2) (April 1985) 74–123.

[5] M. Lhuillier, L. Quan, Image interpolation by joint view interpolation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, CO, 23–25 June, 1999, pp. 139–145.

[6] S. Pastoor, 3D-television: a survey of recent research results on subjective requirements, Signal Process.: Image Commun. 4 (1) (November 1991) 21–32.

[7] A. Rebert, E. Hendriks, J. Biemond, Correspondence estimation in image pairs, IEEE Signal Process. Mag. 16 (3) (May 1999) 29–46.

[8] S. Sethuraman, Stereoscopic image sequence compression using multiresolution and quadtree decomposition based disparity—and motion-adaptive segmentation, Ph.D Dissertation, Carnegie Mellon University, 1996.

[9] R.-S. Wang, Y. Wang, Multiview video sequence analysis, compression, and virtual viewpoint synthesis, IEEE Trans. Circuits Systems Video Technol. 10 (3) (April 2000) 397–410.

[10] G. Wolberg, Digital Image Warping, IEEE Computer Society Press, Los Alamitos, CA, 1990, pp. 41–94 (Chapter 3).