Adaptive Reconstruction of Intermediate Views From Stereoscopic Images

Liang Zhang, Demin Wang, and André Vincent

***Abstract*—This paper deals with disparity estimation and the reconstruction of intermediate views from stereoscopic images. Using block-wise *maximum-likelihood* (ML) disparity estimation, it was found that the *Laplacian* model outperformed the *Cauchy* and *Gaussian* models in terms of disparity compensation errors and the number of correspondence matches. The disparity values in occluded regions were then determined using both object-based and reliability-based interpolation. Finally, an adaptive technique was used to interpolate the intermediate views. One distinguishing characteristic of this algorithm is that the left and right-eye images were projected onto the plane of the intermediate view to be reconstructed. This resulted in two projected images. The intermediate view was created using a weighted average of these two projected images with the weights based on the quality of the corresponding areas of the projected images. Subjective examination of the reconstructed images indicate that they have high image quality and good stable depth when viewed stereoscopically. An objective evaluation with the test image sequence "*Flower Garden*" shows that the proposed algorithm can achieve a peak signal-to-noise ratio gain of around 1 dB, when compared to a reference algorithm.**

***Index Terms*—Disparity estimation, intermediate view reconstruction, stereoscopic image, three-dimensional television (3-D TV), virtual viewpoint synthesis.**

I. Introduction

THREE-DIMENSIONAL television (3-D TV) systems may be the next major rung in the evolution of television [1], [2]. Compared to standard television systems, 3-D TV systems have the potential of providing television viewers with an enhanced impression of depth and a greater sense of presence.

With 3-D TV, it is intuitive to record and distribute 3-D signals as two separate video streams. One stream consisting of images captured or created with a camera viewpoint that is intended for the left eye and the other stream with a viewpoint intended for the right eye. However, by restricting the left and right eyes to two fixed camera viewpoints, spatial distortion and sense of presence can be compromised. The stereoscopic view may be inconsistent with the viewer's change in viewpoint [3]. Furthermore, stereoscopic visualization may cause visual discomfort due to the large difference in horizontal disparity between the two streams of images with differing viewpoints [4]. To enable viewer-dependent changes in viewpoint and to reduce disparity between images obtained from two fixed camera viewpoints, an algorithm for intermediate view reconstruction is needed. In this vein, we developed an algorithm with several distinguishing features.

The reconstruction of intermediate views can be achieved by interpolating between the left and right-eye images based on knowledge of the depth information contained in the scene. Although it is not easy, the depth information may be obtained by estimating the disparity between the left and right-eye images [5]. A general approach to disparity estimation and to obtaining disparity maps from the left and right eye images involves locating corresponding points by measuring intensity differences between two images [6]–[10].

A popular method for disparity estimation based on the above-mentioned approach is *maximum likelihood* (ML) disparity estimation. It may be thought of as a special case of *maximum a posteriori* (MAP) disparity estimation with a constant *a priori* model. A statistical model is required to measure how well one pixel in an image with a given disparity value matches another pixel in the other image. Normally, a *Gaussian* model is used in ML disparity estimation [6]–[11]. However, Sebe *et al.* studied three statistical models, *Cauchy*, *Gaussian* and *Laplacian*, and showed that the *Cauchy* model was the best statistical model for pixel-wise ML disparity estimation [12]. We also found that the *Gaussian* model did not perform as well as the *Laplacian* model when we compared the *Laplacian* model and the *Gaussian* model [13]. Furthermore, in that study we also found that block-wise ML disparity estimation was more reliable than pixel-wise ML disparity estimation.

For this particular study, to select a suitable statistical model for block-wise ML disparity estimation, we repeated our previous assessment of statistical models and then extended our analysis to include the *Cauchy* model. Since intermediate view reconstruction requires dense highly accurate disparity maps, we assessed the models in terms of disparity compensation errors and the number of correspondence matches. This assessment along with the final selection of a model for our algorithm for disparity estimation and intermediate view reconstruction is described in Section II of this paper.

Dealing with occluded areas is another challenge in intermediate view reconstruction. Due to the difference in viewpoints of the left- and right-eye images, some areas are occluded. Therefore, some pixels in one image will have no correspondence to any pixel in the other image. The disparity values in occluded areas cannot be found from the left and right-eye images. To determine the depth in occluded areas, a simple method is to assume that all pixels within a rectangular block have the same disparity values [10]. There exist more complex methods that involve extracting objects from the images and performing object-based linear interpolation [14], [15]. However,

Manuscript received January 4, 2004; revised November 3, 2004. This paper was recommended by Associate Editor A. Kot.

The authors are with the Communications Research Centre Canada, Ottawa, ON K2H 8S2, Canada (e-mail: liang.zhang@crc.ca; demin.wang@crc.ca; andre.vincent@crc.ca).

Digital Object Identifier 10.1109/TCSVT.2005.857785

with most of these methods, the disparity values obtained for pixels in occluded areas are based on estimation and, unfortunately, estimated disparities are not always accurate. To alleviate this problem, we surmised that the accuracy of disparity estimates be measured for disparity values within occluded areas. In Section III, we introduce a measure of reliability for disparity estimation and use this measure in conjunction with object segmentation for determining disparity values in occluded areas.

The critical last step in intermediate view reconstruction is the generation of a new intermediate view using a disparity map. An intermediate view can be linearly interpolated from the left and right-eye images [16], [17], or it may be created by nonlinear interpolation [14], [15], [18]. The aforementioned methods use different weighting factors based on the spatial position of the pixel to be interpolated; the method presented in [16] works well for computer graphics while that in [14] is most suitable for images with simple scene content. In the present study, we choose weighting factors that vary with the disparity compensation error because the quality of the reconstructed intermediate view depends on the accuracy of the disparity estimates. Our algorithm for intermediate view reconstruction is described in Section IV.

Experimental results and performance evaluation of our algorithm for disparity estimation and intermediate view reconstruction are presented in Section V. Conclusions are drawn and outlined in the last section.

II. Statistical Models for Block-Wise ML Disparity Estimation

In the case of the parallel stereo camera configuration, given a pixel $\boldsymbol{p}_1$ of coordinates (x_1, y_1) in the first image and its corresponding pixel $\boldsymbol{p}_2$ of coordinates (x_2, y_2) in the second image, where both pixels originate from projections of the same object point in the 3-D world, disparity is defined as a difference of the horizontal coordinates $x_2 - x_1$ [20]. The goal of disparity estimation is to find the value of $\hat{d}(\boldsymbol{p})$, so that the intensity difference

$$w(\boldsymbol{p}) = s_l(\boldsymbol{p}) - s_r\left(\boldsymbol{p} - \hat{d}(\boldsymbol{p})\right) \tag{1}$$

between the left-eye image point (s_l) and the displaced right-eye image point (s_r) is minimized. In the case of a general camera configuration, epipolar geometry is required to find the disparity value [20], [21].

A block-wise ML disparity estimation is defined as

$$\hat{d}(\boldsymbol{p}) = \max_d \left\{ f(\vec{s}_{l,B} | \vec{s}_{r,B}, d) \right\} \tag{2}$$

where $f(\vec{s}_{l,B} | \vec{s}_{r,B}, d)$ is a joint probability density that measures how well a block $\vec{s}_{r,B}$ in the right-eye image with disparity $d(\boldsymbol{p})$ matches the block $\vec{s}_{l,B}$ in the left-eye image and is referred to as the likelihood term. For block-wise ML disparity estimation, a statistical model is required for the likelihood term $f(\vec{s}_{l,B} | \vec{s}_{r,B}, d)$.

Given a statistical model of an intensity difference signal, a block-wise ML disparity estimator can be deduced from (2). The deduced block-wise ML disparity estimator based on a statistical model can be represented as

$$\hat{d}(\boldsymbol{p}) = \min_d \left\{ \delta(\boldsymbol{p}) \cdot c_o + (1 - \delta(\boldsymbol{p})) \cdot c_m \right\} \tag{3}$$

where c_o is the occlusion cost and c_m is the matching cost. (See Appendix A for details.) The costs, c_o and c_m, depend on the statistical model used.

In selecting a model for our algorithm to determine the best intensity difference signal $w(\boldsymbol{p})$, the performance of three commonly used statistical models, namely, *Cauchy*, *Gaussian*, and *Laplacian* distribution, were compared. Let N_B be the total number of pixels within the block $B(\boldsymbol{p})$ surrounding the pixel point $\boldsymbol{p}$. In the case of the *Cauchy* model

$$c_o(a) = \ln\left(\frac{a}{(f_o \pi)}\right) \tag{4a}$$

$$c_m(a) = \frac{\sum_{\boldsymbol{p}_m \in B(\boldsymbol{p})} \ln\left((s_l(\boldsymbol{p}_m) - s_r(\boldsymbol{p}_m - d(\boldsymbol{p})))^2 + a^2 \right)}{N_B}. \tag{4b}$$

In the case of the Gaussian model

$$c_o(\sigma_w) = \ln\left(\frac{1}{(f_o \cdot \sqrt{2\pi} \cdot \sigma_w)}\right) \tag{5a}$$

$$c_m(\sigma_w) = \frac{\frac{1}{N_B} \cdot \sum_{\boldsymbol{p}_m \in B(\boldsymbol{p})} (s_l(\boldsymbol{p}_m) - s_r(\boldsymbol{p}_m - d(\boldsymbol{p})))^2}{2\sigma_w^2}. \tag{5b}$$

In the case of the *Laplacian* model

$$c_o(\sigma_w) = \ln\left(\frac{1}{(f_o \cdot \sqrt{2} \cdot \sigma_w)}\right) \tag{6a}$$

$$c_m(\sigma_w) = \frac{\frac{1}{N_B} \cdot \sum_{\boldsymbol{p}_m \in B(\boldsymbol{p})} |s_l(\boldsymbol{p}_m) - s_r(\boldsymbol{p}_m - d(\boldsymbol{p}))|}{\frac{\sigma_w}{\sqrt{2}}}. \tag{6b}$$

To implement these block-wise ML disparity estimators, a dynamic programming strategy similar to that in [7] and a hierarchical structure with flexible block sizes presented in [22] were exploited. In this paper, unless specifically noted, an image pyramid with three resolution levels was used for disparity estimation with full-pixel precision. Window sizes of 21 × 21, 11 × 11 and 5 × 5 pixels were used in the hierarchical estimation, from the lowest resolution to the highest resolution. After the disparity map was estimated using the block-wise ML estimator, a bi-directional consistency check was performed in order to detect mismatched estimates of disparity [22]. Pixels associated with mismatched disparities were considered as part of the occluded areas.

To compare the statistical models and because the true disparity maps are unknown, we used the peak signal-to-noise ratio (PSNR) of the disparity-compensation error to evaluate the accuracy of disparity estimation. The disparity-compensation error was calculated based on luminance intensity differences within matching regions between the original right-eye image and the disparity-compensated right-eye image that was obtained from the left-eye image and its estimated disparity map. The *matching region* is used to stand for the regions in which the correspondence between the original left-eye and the

original right-eye images are determined using the method of disparity estimation, and is measured by the matching ratio r

$$r = \frac{\text{number of corresponding pixels}}{\text{total image pixel number}} \times 100\%.$$

Two stereoscopic image sequences, "*Meal*" and "*Tulips*," and one monoscopic image sequence "*Flower Garden*," which can be considered a "stereoscopic" image sequence due to multiple views of the same static scene from camera translation, were used as test stereoscopic image pairs. For the image pair from the "Flower Garden" sequence, we chose one frame for the left-eye image and the next frame for the right-eye image. All images used had a spatial resolution of 720 × 480 pixels.

The disparity estimates of a block-wise ML estimator depend on the value of the estimator parameter, a or σ_w, according to (4), (5), and (6). Given different values of the estimator parameter, each ML estimator can yield a performance comparison of PSNR against the matching ratio. A small value of the estimator parameter, a or σ_w, provides high accuracy of disparity estimates but a lower matching ratio. Fig. 1 shows that the PSNR value decreases as the matching ratio goes up for each estimator. This suggests that large intensity differences due to different viewpoints have to be allowed if a higher matching ratio is desired. A balance between the amount of correspondence matches and accuracy in estimation can be made by appropriate selection of the estimator parameter.

Fig. 1 shows that, at the same matching ratio, the block-wise ML disparity estimator using the *Laplacian* model (*ML-Laplacian*) outperforms both those using the *Gaussian* (*ML-Gaussian*) and *Cauchy* (*ML-Cauchy*) models. We also compared these three models using stereo image pairs from the well-known video sequences "*Aqua*" and "*Tunnel*," and the same results were obtained. Therefore, the findings show that the *Laplacian* model performed the best among the three models tested.

The conclusion from our experimental results is different from that of Sebe [12] who found that the *Cauchy* model was the best among those tested. However, that study was based on pixel-wise ML disparity estimation. The difference between these findings can be explained as follows: for the block-wise *ML-Cauchy* estimator, the matching cost c_m is defined as the sum of the logarithms of intensity differences [see (4b)]. That is similar to the estimation metric, *Lorentzian* function, used in [23] except for a constant that depends on the value of the estimator parameter a. The logarithmic function reduces the relative contribution of large intensity differences to a sum. Large intensity differences usually indicate a mismatch between two blocks. Using (4b), the ability to distinguish two different blocks matched to the reference block is reduced when compared with using (5b) or (6b). Therefore, the block-wise *ML-Cauchy* estimator is less effective than the block-wise *ML-Laplacian* estimator.

Based on our experimental results, we used the block-wise *ML-Laplacian* disparity estimator in the proposed intermediate view reconstruction algorithm.

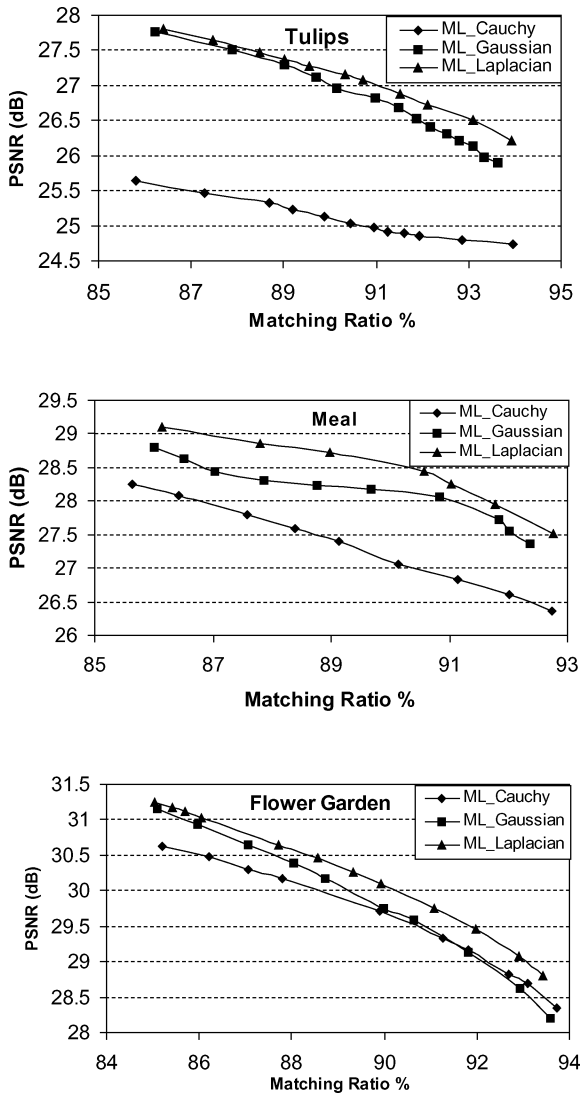


Fig. 1. Comparison of the *Cauchy*, the *Gaussian* and the *Laplacian* models that were used in block-wise ML disparity estimation, for image pairs "*Tulips*", "*Meal*", and "*Flower Garden*".

III. Disparity Interpolation for Occluded Areas

The algorithm to deal with occluded regions consisted of three steps: image object segmentation, calculation of the reliability for the disparity estimates, and object-based and reliability-based disparity interpolation. For the algorithm it was assumed that an image consists of objects and that each object has smooth disparity values. Therefore, the first step was to segment the image into objects and to locate object boundaries. This step was completed using the watershed technique described in detail in [19]. The following subsections describe the other two steps in detail.

A. Reliability Measure of Disparity Estimates

Noise in the original left and right-eye images can lead to disparity estimates that are not accurate, especially in the occluded areas around object boundaries. Therefore, a measure of the reliability of disparity estimates is required for the disparity values in occluded areas. Several methods for measuring the reliability of disparity estimates have been proposed before.

Kanade *et al.* [24] suggested a statistical model of disparity distribution within a window that takes into account disparity variation as well as image intensity variation. Izquierdo [25] used a uniqueness constraint together with an analysis of the curvature of the correlation surface to assess the reliability of disparity estimates. As a reliability measure, Fusiello *et al.* [26] used the variance of disparity estimates obtained with nine asymmetric windows. Murino *et al.* [27] exploited this reliability measure for improving disparity estimates via a *Markov* random map model. Here we put forward another reliability measure.

The measure that we propose is based on the *a posteriori* probability of disparity estimate d, given images $\vec{s}_{l,B}$ and $\vec{s}_{r,B}$, in a block B denoted by $f(d|\vec{s}_{l,B}\vec{s}_{r,B})$ [28]. From the *Bayesian* theorem, we have

$$f(d|\vec{s}_{l,B},\vec{s}_{r,B}) = \frac{f(\vec{s}_{l,B}|d,\vec{s}_{r,B})f(d|\vec{s}_{r,B})}{f(\vec{s}_{l,B}|\vec{s}_{r,B})}. \tag{7}$$

Replace $f(\vec{s}_{l,B}|\vec{s}_{r,B})$ by constant c_1 since it is not a function of d. $f(\vec{s}_{l,B}|d,\vec{s}_{r,B})$ is the likelihood term that measures how well the left-eye image $\vec{s}_{l,B}$ can be described by the disparity estimate d and the right-eye image $\vec{s}_{r,B}$. The intensity difference for one pixel position $\boldsymbol{p}_m$ within a block $B(\boldsymbol{p})$

$$e_m(d(\boldsymbol{p})) = s_l(\boldsymbol{p}_m) - s_r(\boldsymbol{p}_m - d(\boldsymbol{p}))$$

is then modeled with the *Laplacian* model based on our findings as outlined in Section II and $f(\vec{s}_{l,B}|d,\vec{s}_{r,B})$ can be described as

$$f(\vec{s}_{l,B}|d,\vec{s}_{r,B}) = \frac{1}{\sqrt{2}\sigma_w}\cdot\exp\left(-\frac{\sqrt{2}}{\sigma_w}\sum_{B(\boldsymbol{p})}|e_m(d(\boldsymbol{p}))|\right). \tag{8}$$

$f(d|\vec{s}_{r,B})i$ in (7) is the *a priori* probability of the disparity estimate d. The disparity map was assumed to be a realization of a *Gibbs* random map with an energy function that imposed a local smoothness constraint on the variation of disparity estimates. This constraint assigns a smaller probability to disparity values that are significantly different from its neighboring disparity values. Similar to that used in [29], the *a priori* probability $f(d|\vec{s}_{r,B})$ is expressed as

$$f\left(d|\vec{s}_{r,B}\right) = c_2\cdot\exp\left\{-\frac{1}{\sigma_d^2}\sum_{i\in\Lambda}(d-d_i)^2\right\} \tag{9}$$

where the disparity d is independent of the image signal $\vec{s}_{r,B}$. In (9), σ_d^2 is the variance of the difference $d-d_i$ of disparity estimates, which is calculated from the estimated disparity map as described in Section II. To reduce the computation load, we only compared the disparity value at pixel position $\boldsymbol{p}$ with neighboring values located in the horizontal row and the vertical column through its position. Thus, Λ consists of spatial positions in the horizontal row and the vertical column through the pixel position $\boldsymbol{p}$ within a block $B(\boldsymbol{p})$.

Let c denote the value of the product of c_1 and c_2 and let D be the set of all possible values for d of full-pixel precision. Inserting (8) and (9) into (7), we get the value of the constant c from $\sum_{\zeta\in D} f(\zeta|s_l,s_r)=1$

$$c = \frac{1}{\sum_{\zeta\in D}\exp\left\{-\frac{1}{\frac{\sigma}{\sqrt{2}}}\sum_{B(\boldsymbol{p})}|e_m(\zeta(\boldsymbol{p}))| - \frac{1}{\sigma_d^2}\sum_{i\in\Lambda}(\zeta-\boldsymbol{d}_i)^2\right\}}. \tag{10}$$

From (7)–(9), the reliability of the disparity estimate d, denoted by $r(d)$, is measured by

$$r(d) = c\cdot\exp\left\{-\frac{1}{\frac{\sigma}{\sqrt{2}}}\sum_{B(\boldsymbol{p})}|e_m(d(\boldsymbol{p}))| - \frac{1}{\sigma_d^2}\sum_{i\in\Lambda}(d-d_i)^2\right\} \tag{11}$$

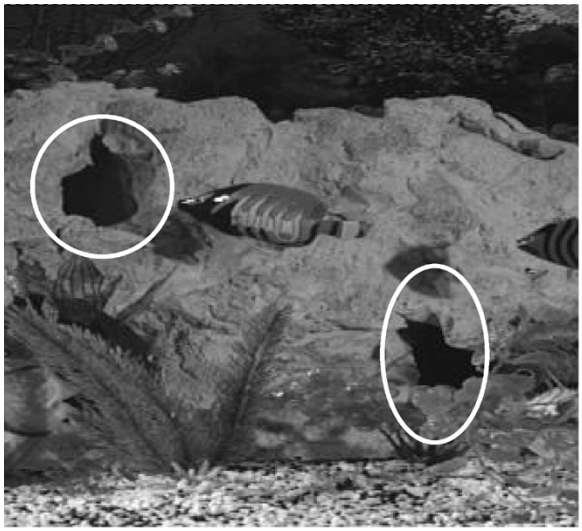
where the constant c is defined as in (10).

Fig. 2 shows the enlarged segments of an estimated disparity map of full-pixel precision. Its reliability map for the stereoscopic image "*Aqua*" uses a block size, $B(\boldsymbol{p})$, of 7×7 pixels and a set Λ of 13 pixels in the horizontal row and the vertical column through the pixel position $\boldsymbol{p}$ and within $B(\boldsymbol{p})$. Fig. 2(a) shows an enlarged segment of the original left-eye image. Fig. 2(b) shows an enlarged segment of the disparity map that was estimated using a block-wise *ML-Laplacian* disparity estimator with a matching ratio of 85%. The estimated disparity values lie in the interval from -31 pixels to 8 pixels. In Fig. 2(b), black areas are the occluded areas where the disparity values have to be determined. Fig. 2(c) shows an enlarged segment of the reliability map that was obtained using (11). The standard deviation σ_d of disparity difference, which was calculated from the estimated disparity map, is 0.27. In the reliability map, the dark areas mean poor estimates with low reliability. In Fig. 2(a), the two circled areas represent holes in the rock of the "*Aqua*" scene. These two holes should have a different depth value than the face of the rock. However, the *ML-Laplacian* disparity estimator assigned the same disparity values to these two holes as the surrounding areas, as shown in Fig. 2(b). From the reliability map in Fig. 2(c), it can be seen that the disparity estimates of points in these holes have low reliability. This demonstrates that the proposed reliability measure is effective.

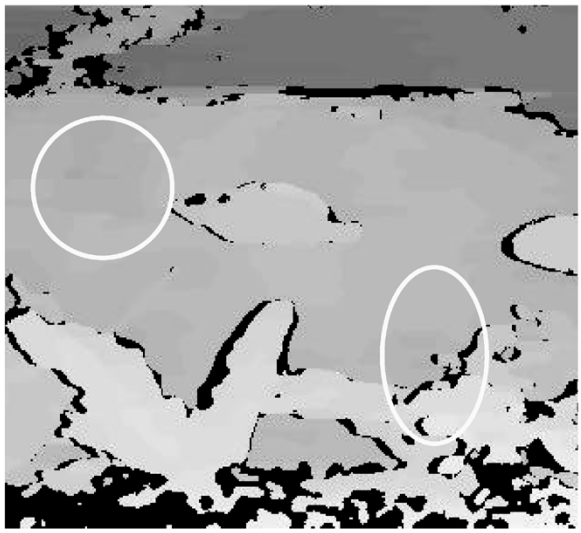
B. Object-Based and Reliability-Based Disparity Interpolation

With the information on object boundaries and a set of reliability measures of disparity estimates, a combined object-based and reliability-based interpolator (Fig. 3) is proposed for determining the disparity values in occluded areas.

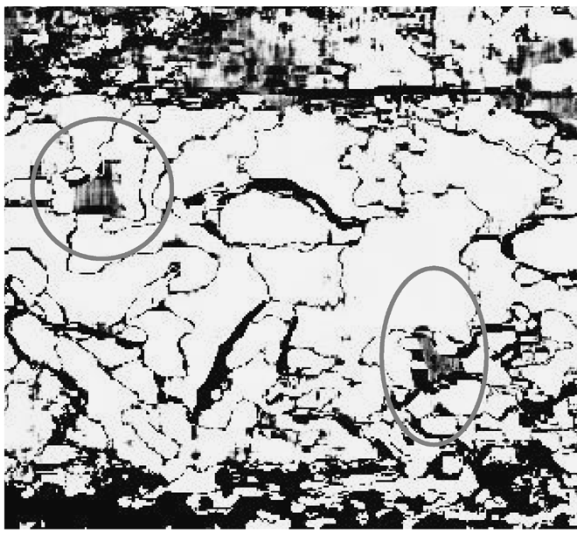
Let $\boldsymbol{p}$ be the position of a pixel in an occluded area and $d(\boldsymbol{p})$ the disparity value to be calculated (see the white regions in Fig. 3). We draw a circle of radius l, centered on pixel position $\boldsymbol{p}$. The radius l of this circle is increased until the circle touches one or more pixels of the nonoccluded area (the gray regions in Fig. 3) that belong to the same object as does the occluded pixel position $\boldsymbol{p}$. These pixels represent the closest pixels in the same object as $\boldsymbol{p}$ with available disparity estimates. Let $\boldsymbol{p}_i$ denote the positions of these closest pixels for $i=1,2,\ldots,N$, $d(\boldsymbol{p}_i)$ their disparity estimates, and $r(d(\boldsymbol{p}_i))$ the reliabilities of these



(a)



(b)



(c)

Fig. 2. Disparity reliability measurements. (a) Enlarged segment of the original left-eye image. (b) Enlarged segment of the estimated disparity map with a matching ratio of 85%. The black regions represent occluded areas. (c) Enlarged segment of a reliability map of the disparity map. Dark areas mean poor estimates with low reliability.

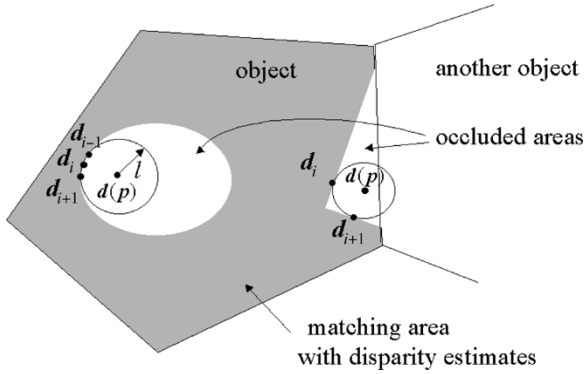
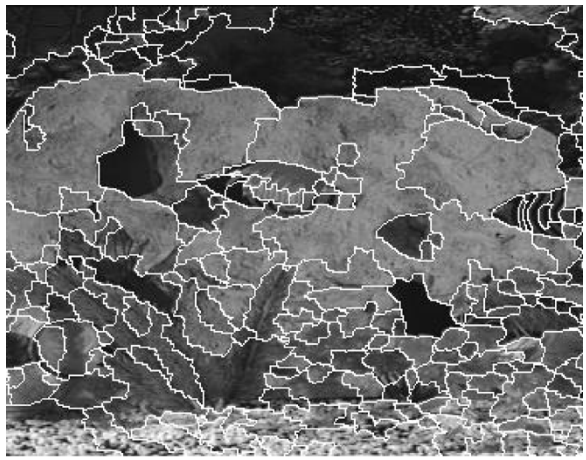


Fig. 3. Object-based and reliability-based disparity interpolation.



(a)

(b)

Fig. 4. Disparity interpolation for occluded areas. (a) Enlarged segment after image segmentation, (b) enlarged segment of disparity map after interpolation for occluded areas.

disparity estimates. The object-based and reliability-based interpolator is then defined as

$$d(\boldsymbol{p}) = \frac{\sum_{i=1}^{N} (r_i(\boldsymbol{p}_i) \cdot d_i(\boldsymbol{p}_i))}{\sum_{i=1}^{N} r_i(\boldsymbol{p}_i)}. \tag{12}$$

Fig. 4 shows the result with the stereoscopic image "*Aqua*" using enlarged segments for better illustration. Fig. 4(a) shows

the segmentation result using the algorithm presented in [19]. The estimated disparity map is shown in Fig. 2(b). The black areas represent occluded areas. Using the segmentation result of Fig. 4(a) and the reliability measures of Fig. 2(c), disparity values in the occluded areas are calculated using the interpolator defined in (12). Fig. 4(b) shows the disparity map after this calculation. Again, the luminance level in this figure represents the magnitude of the disparity values. A lower luminance level represents a smaller disparity value, which means an object is far from the camera. A high luminance level represents a large disparity value, which means an object is closer to the camera. Compared with Fig. 2(b), disparity values in the occluded areas have been successfully determined because there are now no black regions.

IV. Adaptive Intermediate View Reconstruction

The intermediate view reconstruction was carried out by image projection based on disparity estimates and by adaptive combination of the projected images.

A. Image Projection Based on Disparity Map

Let $d_l(\boldsymbol{p})$ denote the disparity map that is estimated by mapping the left-eye image $s_l(\boldsymbol{p})$ to the right-eye image $s_r(\boldsymbol{p})$ using the block-wise *ML-Laplacian* estimator, and $s_\alpha(\boldsymbol{p})$ the intermediate view to be reconstructed. The disparity values for the occluded areas were already obtained using the object-based and reliability-based interpolator described in Section III. Hence, a disparity value is associated with every pixel of the left-eye image. Assume that the distance between the left and right-eye images is 1, and the distance from the left-eye image to the intermediate view $s_\alpha(\boldsymbol{p})$ is α with $0 < a < 1$.

The intermediate view was obtained by projecting the left-eye image $s_l(\boldsymbol{p})$ based on the disparity $d_l(\boldsymbol{p})$ to the plane of the intermediate view. A pixel at position $\boldsymbol{p}$ in the left-eye image is projected to position $\boldsymbol{p}-\alpha\cdot d_l(\boldsymbol{p})$ in the projected image. Let $s_{\alpha,l}(\boldsymbol{p})$ denote this projected image, then $s_{\alpha,l}(\boldsymbol{p}-\alpha\cdot d_l(\boldsymbol{p})) = s_l(\boldsymbol{p})$. Two special situations require additional processing in this projection: multiple projections and holes.

From the disparity map there may be two or more pixels of the left-eye image projected to the same point in $s_{\alpha,l}(\boldsymbol{p})$. In this case, the pixel that belongs to an object that is closer to the camera should be retained because such an object may cover pixels of objects that are farther from the camera. In comparison to other evaluation methods, such as [30], that considers reference images as connected topological meshes to create a back-to-front order, in this paper we adopt a simpler method to cope with this problem. It is known that objects closer to the camera have larger disparities than those farther from the camera in the case of the parallel stereo camera configuration. Therefore, the projection was performed from far to near, i.e., from pixels associated with small disparity values to those with large disparity values.

After all pixels of the left-eye image have been projected, some pixels in $s_{\alpha,l}(\boldsymbol{p})$ may still not have any projection. The areas formed by these pixels are referred to as *holes* in this paper. The pixels in a hole are newly exposed areas and have no correspondence in the left-eye image. The values for these pixels have to be determined from the right-eye image. To this end, another disparity map, denoted by $d_r(\boldsymbol{p})$, is estimated by mapping the right-eye image $s_r(\boldsymbol{p})$ to the left-eye image $s_l(\boldsymbol{p})$. Each disparity value in this map is associated with a pixel in the right-eye image. The holes in $s_{\alpha,l}(\boldsymbol{p})$ are then filled with the projection of the right-eye image based on the disparity map $d_r(\boldsymbol{p})$. That is to say, the pixels in the holes are identical to the pixels projected from the right-eye image. Let R_o denote all holes in $s_{\alpha,l}(\boldsymbol{p})$ which can be described as

$$\boldsymbol{p} = \begin{cases} \boldsymbol{q} - \alpha\cdot d_l(\boldsymbol{q}), & \text{if } \boldsymbol{p} \notin R_o \\ \boldsymbol{q} + (1-\alpha)\cdot d_r(\boldsymbol{q}), & \text{if } \boldsymbol{p} \in R_o \end{cases} \tag{13a}$$

$$s_{\alpha,l}(\boldsymbol{p}) = \begin{cases} s_l(\boldsymbol{q}), & \text{if } \boldsymbol{p} \notin R_o \\ s_r(\boldsymbol{q}), & \text{if } \boldsymbol{p} \in R_o \end{cases}. \tag{13b}$$

Alternatively, we can get another projected image $s_{1-\alpha,r}(\boldsymbol{p})$ by exchanging s_1, d_l, and α with s_r, d_r, and $1-\alpha$, respectively, in (13a) and (13b).

Note: The algorithm proposed above is valid only for stereoscopic image pairs captured or created stereoscopically with a parallel configuration. Otherwise, the stereoscopic image pair has to be rectified, such as with a pre-warping technique as proposed in [31].

B. Adaptive Combination of Projected Images

The final reconstructed intermediate view $s_\alpha(\boldsymbol{p})$ is an adaptive combination of the two projected images $s_{\alpha,l}(\boldsymbol{p})$ and $s_{1-\alpha,r}(\boldsymbol{p})$. These projected images are usually different from one another because $s_{\alpha,l}(\boldsymbol{p})$ mainly comes from the left-eye image, while $s_{1-\alpha,r}(\boldsymbol{p})$ comes from the right-eye image. The difference is also because the multiple projections and holes of $s_{\alpha,l}(\boldsymbol{p})$ are different from those of $s_{1-\alpha,r}(\boldsymbol{p})$. One of the projected images usually has better quality in some areas, while the other is better in other areas. Therefore, an intermediate view of high quality can be obtained by properly combining these two projected images.

The adaptive combination is a weighted average

$$s_\alpha(\boldsymbol{p}) = \lambda(\boldsymbol{p},\alpha)\cdot s_{\alpha,l}(\boldsymbol{p}) + [1-\lambda(\boldsymbol{p},\alpha)]\cdot s_{1-\alpha,r}(\boldsymbol{p}) \tag{14}$$

where $\lambda(\boldsymbol{p},\alpha)$ is a weighting factor which is related to the local quality of the projected images and the distance α between the left-eye image and the intermediate view. In one case, if α is close to 0, the intermediate view should be very similar to the left-eye image. Since $s_{\alpha,l}(\boldsymbol{p})$ mainly comes from the left-eye image, $\lambda(\boldsymbol{p},\alpha)$ should be close to 1 and $s_\alpha(\boldsymbol{p}) \approx s_{\alpha,l}(\boldsymbol{p})$. Otherwise, $\lambda(\boldsymbol{p},\alpha)$ should be close to 0 if α is close to 1. In the other case, if the quality of $s_{\alpha,l}(\boldsymbol{p})$ is higher than $s_{1-\alpha,r}(\boldsymbol{p})$ at position $\boldsymbol{p}$, then $\lambda(\boldsymbol{p})$ should be larger than 0.5. Otherwise, $\lambda(\boldsymbol{p},\alpha)$ should be smaller than 0.5.

The local quality of $s_{\alpha,l}(\boldsymbol{p})$ is measured using the disparity compensation errors

$$e_{\alpha,l}(\boldsymbol{p}) = \begin{cases} \left|s_{\alpha,l}(\boldsymbol{p}) - s_r\left[\boldsymbol{q}+d_l(\boldsymbol{q})\right]\right|, & \text{if } \boldsymbol{p} \notin \boldsymbol{R_o} \\ \left|s_{\alpha,l}(\boldsymbol{p}) - s_l\left[\boldsymbol{q}+d_r(\boldsymbol{q})\right]\right|, & \text{if } \boldsymbol{p} \in \boldsymbol{R_o}. \end{cases} \tag{15}$$

The relationship between $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined in (13a). A small $\mathrm{e}_{\alpha,l}(\boldsymbol{p})$ indicates where the left-eye image matches well with the right-eye image. Hence, the quality of $s_{\alpha,l}(\boldsymbol{p})$ should be high at position $\boldsymbol{p}$ if $\mathrm{e}_{\alpha,l}(\boldsymbol{p})$ is small. Let $\mathrm{e}_{1-\alpha,r}(\boldsymbol{p})$ denote the disparity

compensation error associated with $s_{1-\alpha,r}(\boldsymbol{p})$. This error can be obtained by exchanging s_l and d_l with s_r and d_r in (15).

From the above discussion, the weighting factor $\lambda(\boldsymbol{p},\alpha)$ is then defined as

$$\lambda(\boldsymbol{p},\alpha)=\frac{(1-\alpha)\cdot[A+e_{1-\alpha,r}(\boldsymbol{p})]}{A+\alpha\cdot e_{\alpha,l}(\boldsymbol{p})+(1-\alpha)\cdot e_{1-\alpha,r}(\boldsymbol{p})} \tag{16}$$

where A is a predetermined constant with $A>0$. This constant can reduce the impact of small disparity compensation errors on $\lambda(\boldsymbol{p},\alpha)$. Small disparity compensation errors are normally caused by random noise in the original (left and right-eye) images, rather than incorrect disparity estimates. In our experiments, we set $A=2$.

V. Experimental Results

The proposed algorithm was compared with two other algorithms. The first algorithm is the *Hybrid* algorithm. The main difference between the *Hybrid* algorithm and the proposed algorithm is how the disparity map is estimated. In the reference *Hybrid* algorithm, correspondence between the left and right-eye images is established using a block-based and object-based matching technique similar to the algorithm presented in [32] that has been successfully used in frame rate conversion [33]. With this algorithm, disparity-per-pixel (disparity map) can be obtained. After that, the intermediate view is reconstructed by an adaptive combination similar to that presented in this paper. The other algorithm that was used for comparison is the one described in [34] and referred to as *NO-Reli* in the following statement. The main difference between *NO-Reli* and the proposed algorithm is that *NO-Reli* does not use a reliability measure of disparity estimates in the determination of the disparity values in occluded areas. The reason for choosing these two algorithms for comparison was that in both algorithms disparity is estimated in a block-wise manner. In the following experimental results, the parameter, σ_w, for block-wise ML disparity estimation was set to 3.24.

A. Objective Evaluation

For objective comparison, the test image sequence "*Flower Garden*" was used. "*Flower Garden*" can be considered a "stereoscopic" image sequence due to multiple views of a static scene with camera translation. This provides "ground truth" intermediate views for numerical performance evaluation using a PSNR metric. For testing, we chose two image pairs and their "ground truth" center images (i.e., intermediate frames) as shown in Table I. Fig. 5 shows the left-eye images of each of the two image pairs. It should be noted that "*Flower Garden*" is of the interlaced format. Therefore, the intermediate view at position $\alpha=0.5$ was reconstructed field wise and compared to the "ground truth" image. Table II shows the comparison results, in terms of PSNR of the error signals between the reconstructed view and the "ground truth" image at position $\alpha=0.5$.

It can be seen that the proposed algorithm outperforms the *Hybrid* and the *NO-Reli* algorithm in terms of PSNR. For *Image Pair* A, the PSNR gains were, 0.95 and 0.02 dB, respectively. For *Image Pair* B they were 11 and 0.02 dB. For *Image Pair* B, the *Hybrid* algorithm determined that the estimated disparity map was not reliable and therefore copied the left-eye image as the intermediate view to be reconstructed. This resulted in a lower PSNR value.

We also compared the quality of the reconstructed intermediate images. These reconstructed images were created by image projection based on disparity maps and adaptive combination. The images $s_{\alpha,l}$ and $s_{1-\alpha,r}$ are reconstructed by image projection based on disparity maps which mainly use the left-eye image as defined in (13b) and the right-eye image. The image s_α is an adaptive combination of both images as defined in (14). Table III shows the PSNR values of these three reconstructed images compared to the "ground truth" center image. For the *Image Pair* A, the block ML disparity estimator achieved a high matching ratio of around 84%. The high PSNR

TABLE I
Two Image Pairs and Their "*Ground Truth*" Center Images From Image Sequence *Flower Garden*

	Left-eye image (frame #)	Right-eye image (frame #)	"Ground truth" image (frame #)
Image Pair A	3	1	2
Image Pair B	232	230	231

Fig. 5. Left-eye image of *Image Pair* A (top image) and *Image Pair* B (bottom image).

TABLE II
PSNR Comparison Results for the Intermediate View Reconstructed at Position $\alpha=0.5$

	Hybrid (dB)	*NO-Reli* (dB)	*Proposed* (dB)
Image Pair A	28.2486	29.1811	29.2014
Image Pair B	15.8132	27.1442	27.1683

TABLE III
PSNR COMPARISON RESULTS FOR INTERMEDIATE VIEWS RECONSTRUCTED AT POSITION $\alpha = 0.5$ BY IMAGE PROJECTION BASED ON DISPARITY MAP AND BY ADAPTIVE COMBINATION

	image $s_{\alpha,l}$ (dB)	image $s_{1-\alpha,r}$ (dB)	image s_{α} (dB)
Image Pair A	24.4187	29.5505	29.2014
Image Pair B	24.1653	23.9767	27.1683

Fig. 6. Estimated disparity map (top image) corresponding to the left-eye view of the stereoscopic video sequence "*Tulips*" (bottom image).

difference between the images $s_{\alpha,l}$ and $s_{1-\alpha,r}$ mainly stems from the different quality on the right side of the images; where the disparity values are estimated for the right-eye image while they are recovered for the left-eye image by the object-based and reliability-based disparity interpolator as presented in Section III. Fig. 5 shows that the luminance texture on the right side is discontinuous and has a vertical black bar instead of the texture of flowers and houses. This texture discontinuity led to a reduction in image quality due to inaccuracies in the disparity values in generating image $s_{\alpha,l}$. Even so, the adaptively combined image s_{α} still had a higher quality than the image $s_{\alpha,l}$ (a PSNR gain of 4.8 dB), and a similar quality to image $s_{1-\alpha,r}$ (a PSNR difference less than 0.4 dB). For the *Image Pair* B, the block ML disparity estimator provided a matching ratio of only 53%. The remaining disparity values were recovered by the object-based and reliability-based disparity interpolation. For this lower matching ratio, both images, $s_{\alpha,l}$ and $s_{1-\alpha,r}$, had similar quality with a PSNR value of 24 dB. However, the adaptively combined image s_{α} had a higher quality with a PSNR gain of 3 dB. These comparisons show that the use of adaptive combination proposed in this paper can create an intermediate view of better quality than produced by image

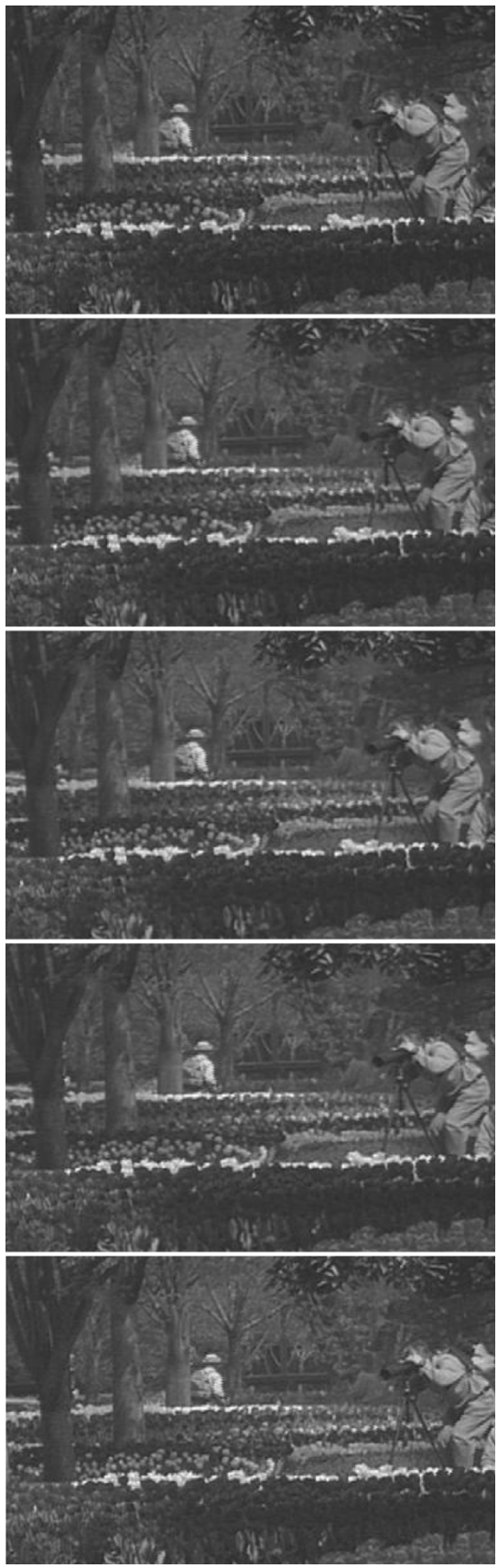


Fig. 7. Enlarged segments of intermediate views reconstructed from an image pair of stereoscopic video sequence "*Tulips*". From top to bottom, the images are left-eye image, intermediate views at position $\alpha = 0.25$, 0.5, 0.75, and right-eye image.

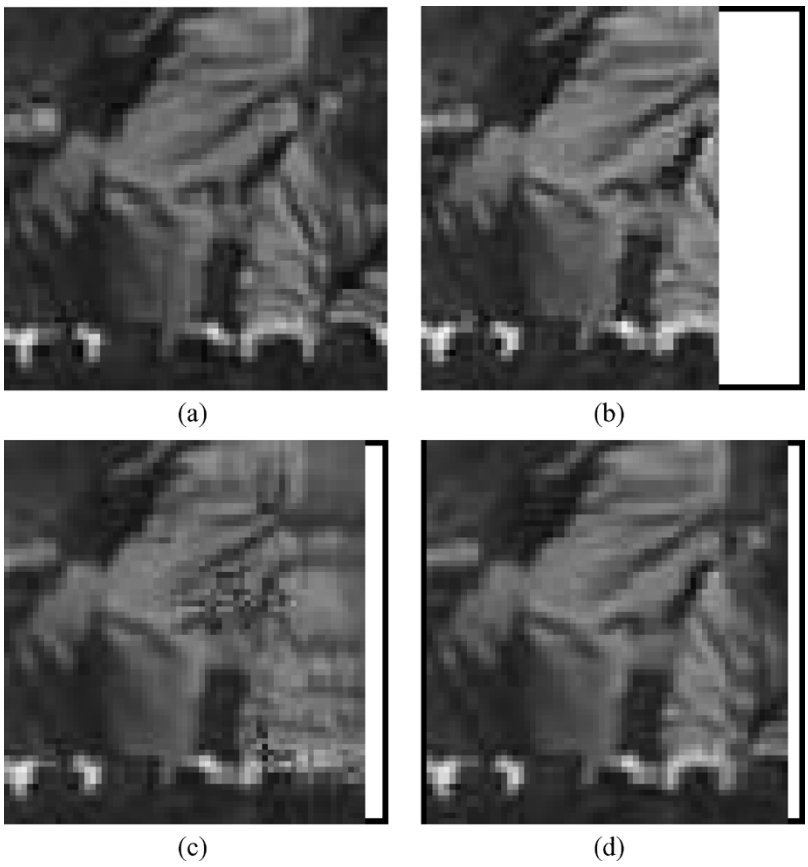


Fig. 8. Enlarged segments of *Tulips* and its intermediate views reconstructed at position $\alpha = 0.5$. (a) Original left-eye image, (b) original right-eye image, (c) result of the *Hybrid* algorithm, (d) result of the proposed algorithm.

projection based on disparity maps using either the left or right-eye image.

As stated in Section II, we could increase the value of the estimator parameter σ_w to improve the matching ratio. However, tests with different natural stereoscopic sequences showed that, given the estimator parameter, matching ratios depend on scene content. Additionally, when the scene depth changes, it is difficult to predict which image, $s_{\alpha,l}$ or $s_{1-\alpha,r}$, has better quality. Thus, the proposed technique of adaptive combination is simple and practical for generating intermediate views of high quality.

B. Subjective Evaluation

For subjective comparison, the results with the test image sequence "*Tulips*" were used. Fig. 6 shows the disparity map corresponding to the left-eye view of "*Tulips*." The block ML disparity estimator provided a matching ratio of around 88%. The remaining disparity values were recovered by object-based and reliability-based disparity interpolation. The range of the estimated disparity values lies in the interval from -27 pixels to $+10$ pixels. It should be noted that there are some estimation errors at the left of the person. Fig. 7 shows the results of intermediate view reconstruction with five images using enlarged segments for clear illustration. From top to bottom, these images are the original left-eye image, the intermediate views reconstructed at position $\alpha = 0.25$, 0.5, and 0.75, and the original right-eye image. From these images, it can be seen that the proposed algorithm reconstructs image details well, producing "sharp" images. Overall, the quality of the reconstructed images is excellent.

Note the position of the cameramen on the right side of Fig. 7. It changes gradually in the horizontal direction from the top image to the bottom image. This indicates that the depth of this person in the reconstructed stereoscopic view changes smoothly and in a consistent manner. A stereoscopic sequence created with the reconstructed intermediate view and the left or right-eye original image sequence exhibits very clean and stable depth.

The difference in subjective quality between the proposed algorithm and the *Hybrid* algorithm can be observed by comparing the images in Fig. 8. It shows an enlarged segment of "*Tulips*." We can see the legs of the cameraman and the arm of another person. Fig. 8(a) and Fig. 8(b) are the segments from the original left and right-eye images, respectively. Fig. 8(c) is the segment of the intermediate view that was reconstructed using the *Hybrid* algorithm. Fig. 8(d) was reconstructed using the proposed algorithm. It can be seen that the *Hybrid* algorithm introduced some texture artifacts between the cameraman's legs and the arm of the other person, while the proposed algorithm strongly reduced these texture artifacts and produced an intermediate view of higher quality.

Fig. 9 shows differences in the reconstructed views based on the proposed algorithm and the *NO-Reli* algorithm. Fig. 9(a) shows an enlarged segment of the result produced by the *NO-Reli* algorithm, and Fig. 9(b) is an enlarged segment of

(a) (b)

Fig. 9. Intermediate views reconstructed at position $\alpha = 0.5$ for an enlarged region of "*Tulips*". (a) Result based on *NO-Reli*, (b) result based on the proposed algorithm.

the result by the proposed algorithm. Texture artifacts can be observed along the boundary of the woman's face in Fig. 9(a), and are noticeably less visible in (9b). The artifacts are due to the fact that the *NO-Reli* algorithm does not use a measure of reliability for disparity estimates, so the impact of inaccurate estimates is significant.

The objective evaluation in the previous subsection shows that the proposed algorithm achieves similar quality with the *NO-Reli* algorithm in terms of PSNR. However, the subjective evaluation presented in this subsection indicates that the quality of the intermediate views can be quite different and the proposed algorithm outperforms the *NO-Reli* algorithm.

VI. Conclusion

In this paper, an algorithm for disparity estimation and intermediate view reconstruction for stereoscopic imagery was presented. Based on objective criteria, the *Laplacian* model was selected over the *Cauchy* and *Gaussian* models for block-wise ML disparity estimation. We note that this choice of model is different from that of Sebe [12] for pixel-wise ML disparity estimation.

The introduction of a reliability measure of disparity estimation combined with object segmentation reduced the impact of inaccurate disparity estimates on the quality of the reconstructed view in occluded areas.

Furthermore, the introduction of an adaptive technique for reconstructing intermediate views whereby the intermediate view is a weighted average of the two projected images from the left and right-eye views, with weights based on the local quality of the projected image, contributed to a favorable performance of the proposed algorithm when compared to a reference *Hybrid* algorithm.

An objective evaluation with the test sequence "*Flower Garden*" showed that the proposed algorithm could achieve a PSNR gain of around 1 dB compared to the reference *Hybrid* algorithm.

We conclude that the introduction of a reliability measure for disparity estimation as well as the use of projection errors to determine the weights for view interpolation are practical and useful. This conclusion is supported by subjective evaluation with a stereoscopic test sequence, "*Tulips*," showing that the proposed algorithm reconstructs intermediate views of high quality with very clean and stable depth when viewed stereoscopically.

Appendix A
Block-Wise ML Disparity Estimation

For block-wise ML disparity estimation, neighboring pixel points surrounding the pixel point $\boldsymbol{p}$ are taken into account. Let $B(\boldsymbol{p})$ be a block centered at the pixel position $\boldsymbol{p}$ in the left-eye image, N_B the total number of pixels within the block $B(\boldsymbol{p})$, and $\vec{s}_{l,B}$ and $\vec{s}_{r,B}$ denote a set of pixel intensity levels in the left and right-eye images, respectively. $d(\boldsymbol{p})$ is the disparity value for the block $B(\boldsymbol{p})$. It is assumed that additive noise signal intensity $w(\boldsymbol{p}_{\mathrm{m}})$

$$w(\boldsymbol{p}_m) = s_l(\boldsymbol{p}_m) - s_r\left(\boldsymbol{p}_m - \hat{d}(\boldsymbol{p})\right) \tag{A-1}$$

at any pixel point $\boldsymbol{p}_{\mathrm{m}}$ within the block $B(\boldsymbol{p})$ is independent of its neighboring noise signal intensity. The likelihood function $f(\vec{s}_{l,B}|\vec{s}_{r,B}, d)$ for block-wise ML estimation is then defined as

$$f(\vec{s}_{l,B}|\vec{s}_{r,B}, d) = \prod_{\boldsymbol{p}_m \in B(\boldsymbol{p})} \left(f(s_l(\boldsymbol{p}_m)|s_r(\boldsymbol{p}_m), d(\boldsymbol{p})\right). \tag{A-2}$$

where $f(s_l(\boldsymbol{p}_m)|s_r(\boldsymbol{p}_m), d(\boldsymbol{p}))$ is a likelihood function that describes how well the right-eye image pixel $s_r(\boldsymbol{p}_{\mathrm{m}})$ with the disparity $d(\boldsymbol{p})$ matches the left-eye image pixel $s_l(\boldsymbol{p}_{\mathrm{m}})$. For a correspondence pixel, $f(s_l(\boldsymbol{p}_m)|s_r(\boldsymbol{p}_m), d(\boldsymbol{p}))$ can be expressed as

$$f\left(s_l(\boldsymbol{p}_m)|s_r(\boldsymbol{p}_m), d(\boldsymbol{p})\right) = f\left(w(\boldsymbol{p}_m)\right) \tag{A-3}$$

according to (A-1). In the case of occlusion, the left-eye image intensity $s_l(\boldsymbol{p}_{\mathrm{m}})$ is independent of the right-eye image intensity $s_r(\boldsymbol{p}_{\mathrm{m}})$ and also of the disparity $d(\boldsymbol{p})$. The likelihood function

$f(s_l(\boldsymbol{p}_m)|s_r(\boldsymbol{p}_m), d(\boldsymbol{p}))$ for an occluded point can then be simplified as

$$f\left(s_l(\boldsymbol{p}_m)|s_r(\boldsymbol{p}_m), d(\boldsymbol{p})\right) = f\left(s_l(\boldsymbol{p}_m)\right). \tag{A-4}$$

Furthermore, if it is assumed that the probability distribution function $f(s_l(\boldsymbol{p}_{\mathrm{m}}))$ is a constant and under the consideration that intensity values of image pixel points lie in the interval [0, 255], $f(s_l(\boldsymbol{p}_m))$ is then equal to

$$f\left(s_l(\boldsymbol{p}_m)\right) = f_o = \frac{1}{256} \tag{A-5}$$

which is the same as that proposed in [8]. Considering correspondence and occlusion cases, the likelihood function can be written as

$$f\left(s_l(\boldsymbol{p}_m)|s_r(\boldsymbol{p}_m), d(\boldsymbol{p})\right) = (f_o)^{\delta(\boldsymbol{p})} f\left(w(\boldsymbol{p}_m)\right)^{1-\delta(\boldsymbol{p})} \tag{A-6}$$

with an indicator variable $\delta(\boldsymbol{p})$ that is defined as

$$\delta(\boldsymbol{p}) = \begin{cases} 1, & \text{if } \boldsymbol{p} \in \text{occlusion} \\ 0, & \text{if } \boldsymbol{p} \notin \text{occlusion.} \end{cases} \tag{A-7}$$

Maximizing the likelihood function $f(\vec{s}_{l,B}|\vec{s}_{r,B}, d)$ results in the block-wise ML estimator

$$\hat{d}(\boldsymbol{p}) = \max_d \left\{ f(\vec{s}_{l,B}|\vec{s}_{r,B}, d) \right\}. \tag{A-8}$$

To simplify the optimization, (A-8) is modified as

$$\max_d \left\{ f(\vec{s}_{l,B}|\vec{s}_{r,B}, d) \right\} \rightarrow \min_d \left\{ -\ln\left(f(\vec{s}_{l,B}|\vec{s}_{r,B}, d) \right) \right\}. \tag{A-9}$$

To model noise signal $w(\boldsymbol{p}_m)$, three statistical models $f(w(\boldsymbol{p}_m))$, namely, the *Cauchy* model with a parameter a

$$f\left(w(\boldsymbol{p}_m)\right) = f_C(w, a) = \frac{a}{\pi} \frac{1}{a^2 + w^2} \tag{A-10}$$

the *Gaussian* model with a variance σ_w^2

$$f\left(w(\boldsymbol{p}_m)\right) = f_G(w, \sigma_w) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left\{ -\frac{w^2}{2\sigma_w^2} \right\} \tag{A-11}$$

and the *Laplacian* model with a variance σ_w^2

$$f\left(w(\boldsymbol{p}_m)\right) = f_L(w, \sigma_w) = \frac{1}{\sqrt{2}\sigma_w} \exp\left\{ -\frac{|w|}{\frac{\sigma_w}{\sqrt{2}}} \right\} \tag{A-12}$$

were studied in this paper because they are the commonly used models in the literature. They were also examined in [12]. A block-wise ML disparity estimator can then be deduced from (A-8) $\sim$ (A-12) by inserting each $f(w(\boldsymbol{p}_m))$ into (A-6), (A-2) and then into (A-9). The deduced block-wise ML disparity estimator, based on different models, can be represented as

$$\hat{d}(\boldsymbol{p}) = \min_d \left\{ \delta(\boldsymbol{p}) \cdot c_o + (1 - \delta(\boldsymbol{p})) \cdot c_m \right\} \tag{A-13}$$

where c_o is the occlusion cost and c_m is the matching cost. The costs, c_o and c_m, depend on the statistical model used [see (4), (5) and (6)].

ACKNOWLEDGMENT

The authors would like to thank NHK for supplying the natural stereoscopic test sequence *Tulips*. Thanks are also due to Dr. W. J. Tam, Mr. P. Blanchfield and Dr. F. Speranza for their valuable comments on an earlier version of this manuscript.

REFERENCES

[1] R. Bushmann, "Stereoscopic and 3-D visual communications for the future," *SPIE Stereoscopic Display Virtual Reality Syst. VI*, vol. 3639, pp. 232–241, May 1999.

[2] "Applications and Requirements for 3DAV," ISO/IEC JTC1/SC29/WG11 N4982, Klagenfurt, Austria, Jul. 2002.

[3] S. Pastoor, "Human factors of 3-D displays in advanced image communications," *Displays*, vol. 14, no. 3, pp. 150–157, 1993.

[4] J. Konrad, "Enhancement of viewer comfort in stereoscopic viewing: parallax adjustment," *SPIE Stereoscopic Displays Virtual Reality Syst.*, vol. 3639, pp. 179–190, Jan. 1999.

[5] A. Redert, E. Hendriks, and J. Biemond, "Correspondence estimation in image pairs," *IEEE Signal Process. Mag.*, vol. 16, no. 3, pp. 29–46, May 1999.

[6] P. N. Belhumeur, "A bayesian approach to binocular stereopsis," *Int. J. Comput. Vis.*, vol. 19, no. 3, pp. 237–262, 1996.

[7] I. Cox, S. Hingoraini, and S. Rao, "A maximum likelihood stereo algorithm," *Comput. Vis. Image Understanding*, vol. 63, no. 3, pp. 542–567, May 1996.

[8] L. Falkenhagen, "Blockbasierte disparitaets-schaetzung unter berueck-sichtigung statistischer abhaengigkeiten der disparitaeten," Ph.D. dissertation, Univ. Hannover, Hannover, Germany, 2001.

[9] J. Karathanasis, D. Kalivas, and J. Vlontzos, "Disparity estimation using block matching and dynamic programming," in *IEEE Conf. Electronics, Circuits and Systems*, Rhodes, Greece, Oct. 13–16, 1996, pp. 728–731.

[10] A. Mansouri and J. Konrad, "Bayesian winner-take-all reconstruction of intermediate views from stereoscopic images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1710–1722, Oct. 2000.

[11] L. Matthies, "Stereovision for planetary rovers: stochastic modeling to near real-time implementation," *Int. J. Comp. Vis.*, vol. 8, no. 1, pp. 71–91, 1992.

[12] N. Sebe, M. S. Lew, and D. P. Huijsmans, "Toward improved ranking metrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1132–1143, Oct. 2000.

[13] L. Zhang, "Statistical model for intensity differences of corresponding points between stereo image pairs," in *Proc. IEEE Int. Conf. Multimedia and Expo (ICME)*, vol. I, Baltimore, MD, Jul. 6–9, 2003, pp. 365–368.

[14] E. Izquierdo, "Stereo matching for enhanced tele-presence in three-dimensional video communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 4, pp. 629–643, Aug. 1997.

[15] E. Izquierdo and J.-R. Ohm, "Image-based rendering and 3-D modeling: a complete framework," *Signal Process. Image Commun.*, vol. 15, pp. 817–858, 2000.

[16] S. Chen and L. Williams, "View interpolation for image synthesis," in *Proc. SIGGRAPH 93*, Anaheim, CA, Aug. 1993, pp. 279–288.

[17] T. Werner, R. D. Hersch, and V. Hlavac, "Rendering real-world objects using view interpolation," in *Proc. IEEE Int. Conf. Comp. Vis.*, Boston, MA, 1995, pp. 957–962.

[18] R. Krishnamurthy, J. W. Woods, and P. Moulin, "Frame interpolation and Bi-directional prediction of video using compactly encoded optical-flow fields and label fields," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 5, pp. 713–725, Aug. 1999.

[19] D. Wang, "Unsupervised video segmentation based on watersheds and temporal tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 539–546, Sep. 1998.

[20] O. Faugeras, *Three-Dimensional Computer Visualization: A Geometric Viewpoint*. Cambridge, MA: MIT Press, 1993.

[21] Z. Zhang, R. Deriche, O. D. Faugeras, and Q. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artif. Intell.*, vol. 78, pp. 87–119, 1995.

[22] L. Zhang, "Hierarchical block-based disparity estimation using mean absolute difference and dynamic programming," in *Proc, Int. Workshop Very Low Bit-Rate Video Coding (VLBV01)*, Athens, Greece, Oct. 11–12, 2001, pp. 114–117.

[23] M. Black and A. D. Jepson, "Estimating optical flow in segmented images using variable-order parametric models with local deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 972–986, Oct. 1996.
[24] T. Kanada and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 9, pp. 920–932, Sep. 1994.
[25] E. Izquierdo, "Disparity/segmentation analysis: Matching with an adaptive window and depth-driven segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 4, pp. 589–607, Jun. 1999.
[26] A. Fusiello, V. Roberto, and E. Trucco, "Symmetric stereo with multiple windowing," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 14, no. 8, pp. 1053–1066, Dec. 2000.
[27] V. Murino, U. Castellani, and A. Fusiello, "Disparity map restoration by integration of confidence in Markov random fields models," in *IEEE Int. Conf. Image Process.*, vol. II, Thessaloniki, Greece, Oct. 7–10, 2001, pp. 29–32.
[28] L. Zhang, D. Wang, and A. Vincent, "Reliability measure of disparity estimates for intermediate view reconstruction," in *Proc. IEEE Int. Conf. Image Process.*, vol. III, Rochester, NY, Sep. 22–25, 2003, pp. 837–840.
[29] J. Konard and E. Dubois, "Byesian estimation of motion vector fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 9, pp. 910–927, Sep. 1992.
[30] L. McMillan and G. Bishop, "Head-tracked stereo display using image warping," *Proc. Stereoscopic Displays and Virtual Reality Systems II*, vol. 2409, pp. 21–30, Feb. 1995.
[31] S. M. Seitz and C. R. Dyer, "View morphing: synthesizing 3-D metamorphoses using image transforms," in *Proc. SIGGRAPH*, Aug. 1996, pp. 21–30.
[32] D. Wang and D. Lauzon, "Hybrid algorithm for estimating true motion fields," *Opt. Eng.*, vol. 39, no. 11, pp. 2876–2881, Nov. 2000.
[33] D. Wang, A. Vincent, and P. Blanchfield, "CRC-FRC: Advanced Frame Rate Converter," Tech. Rep, Communications Research Centre Canada, Feb. 2003.
[34] L. Zhang, D. Wang, and A. Vincent, "An adaptive object-based reconstruction of intermediate views from stereoscopic images," in *Proc. IEEE Int. Conf. Image Process.*, vol. III, Thessaloniki, Greece, Oct. 2001, pp. 923–926.

Liang Zhang received the B.Sc. degree from Chengdu Institute of Radio Engineering, Chengdu, China, in 1982, the M.Sc. degree from Shanghai Jiaotong University, Shanghai, in 1986, and the Ph.D. degree in electrical engineering from the University of Hannover, Hannover, Germany, in 2000.

He was working as an Assistant from 1987 to 1988 and as a Lecturer from 1989 to 1992 in the Department of Electrical Engineering, Shanghai Jiaotong University. From 1992 to 2000, he was a Research Assistant at the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover. Since 2000, he has been with Communications Research Centre, Ottawa, ON, Canada. His research interests include image and video coding, image and video processing, stereo image processing and 3-D TV.

Demin Wang received the B.S. and M.S. degrees in electrical engineering from Shandong University of Technology, Shandong, China, in 1982 and 1985, respectively, and the Ph.D. degree from the Institut National des Sciences Appliquees (INSA) de Rennes, Rennes, France, in 1992.

In 1985, he joined Shandong University of Technology where he served as a Professor of electrical and computer engineering from 1992 to 1993. He was a Visiting Researcher at the University of Sherbrooke, Sherbroke, QC, Canada, from 1993 to 1994, and a Visiting Professor at the IRISA Rennes, France, from 1994 to 1995. Since 1996, he has been with the Communications Research Centre, Ottawa, ON, Canada, where he is currently senior scientist and the project leader of video coding and processing. His research interests include image coding and processing, video processing and coding, digital TV and broadcasting. He has published over 70 journal and conference papers.

Dr. Wang is an Associate Editor for IEEE TRANSACTIONS ON BROADCASTING.

André Vincent received the B.Sc. degree in electrical engineering from l'École Polytechnique de Montréal, Montréal, QC, Canada, in 1975.

From 1975 to 1977, he worked at the Department of National Defence, in the design of maritime communications systems. From 1977 to 1979, he worked at Canadian Marconi in the design and development of mobile radio communications systems. He joined the Communications Research Centre in Ottawa, ON, Canada, in 1979, where he has conducted research in the areas of Teletext, data transmission, television channel characterization, and digital mobile radio. Since 1986, he has been involved in research in HDTV, video processing, video compression and 3-D video. He is currently Manager of the Advanced Video Systems Group, Communications Research Centre.