Gaze-corrected View Generation Using Stereo Camera System for Immersive Videoconferencing

Sang-Beom Lee, *Student Member*, IEEE, In-Yong Shin, and Yo-Sung Ho, *Senior Member*, IEEE

**Abstract — ***In this paper, we present an immersive videoconferencing system that enables gaze correction between users in the internet protocol TV (IPTV) environment. After we capture the object using stereo cameras, we perform preprocessing techniques, such as camera calibration, color correction, and image rectification. The preprocessed images are down-sampled and disparities are computed by using the down-sampled images. The disparity sequence is then filtered to improve temporal consistency. After central view synthesis, occlusion areas are decided and holes are filled. The entire system is implemented with parallel programming that is executed on the GPU for real-time processing. Finally, the user can observe the gaze-corrected image through display. From experimental results, we have verified that the proposed stereo camera system is sufficient to generate the natural gaze-corrected virtual image and realize immersive videoconferencing*[1].

Index Terms — Immersive videoconferencing, eye contact, gaze correction, stereo camera system, disparity estimation, view synthesis, IPTV.

I. INTRODUCTION

Recent development in computing power, interactive computer graphics, digital transmission, and immersive displays enables us to reproduce reality by computer simulations [1]. When users are exposed to such immersive, interactive, and perceptually realistic media, they report a sense of presence in the mediated environments.

Television realized a human dream of watching a distant world in real-time. Moreover, it has been a great portion of visual system since it was invented. Nowadays, internet protocol TV (IPTV) provides us with more immersive feelings than ever before since the IPTV can service the bidirectional user interaction functionality. Here, IPTV is defined as multimedia services such as video, audio, text, graphics, and data delivered over IP based networks managed to provide the required level of quality of service and experience, security, interactivity, and reliability [2].

One of advantages of IPTV is videoconferencing. It is an interactive telecommunication technology which allows two or more distant users to interact through bidirectional audiovisual transmissions at once. Recently, many researchers are showing their interest in the development of videoconferencing technology for consumer devices such as televisions, laptops, and cell phones, which are equipped with high-quality audio and video [3]–[5].Users can communicate with each other by adding cameras and microphones to these devices. However, these technologies do not realize the natural eye contact between conferees and directional gaze awareness is not considered.

Eye contact is one of the most significant issue and challenging requirement for an immersive videoconferencing system. Although many conventional works have been proposed, these studies have not yet efficiently realized eye contact. One of the recent systems is a remote collaboration system based on a semi-transparent display to create an experience where local and remote users are separated only by a vertical sheet of glass. [6]. However, this approach cannot be widely used due to the expensive costs of hardware. Another videoconferencing system provides nearly resolves the eye contact issue by supporting the system with three cameras mounted on top of the displays [7]. In this approach, metaphor-based design methodology is used. However, due to the short distance from cameras to the displayed head of the remote users and the relatively long distance from local users to the display, the displacement angle regarding the viewing directions can be neglected.

Recently, novel prototype systems using 3D image processing techniques have been proposed. The common issue of these systems is how to synthesize the image of the arbitrary viewpoint that users face. In order to generate the virtual view using 3D image processing techniques, depth information is required. It can be obtained directly from a depth camera using an infrared sensor [8], [9]. It also can be obtained by depth estimation algorithms with multiple cameras [10], [11]. One recent state-of-the-art algorithm proposed in an effort to solve the eye contact problem utilizing 3D image processing [12]–[15]. This prototype system uses three cameras mounted on the top of the display and an additional camera on the bottom of the display. The upper three cameras configure the L-type trifocal system and the acquired images are used as inputs of depth estimation for the small baseline. The lower camera is used for another depth estimation process for the wide baseline. The virtual view synthesis is performed by warping captured images. Still, this system is inappropriate for the practical IPTV environments due to complex setup and emergence of another concern, on how to solve the depth fusion issue.

In this paper, we design an immersive videoconferencing system utilizing stereo cameras with convergent configuration. The main contribution of our work is to

[1] This research was supported in part by MKE under the ITRC support program supervised by NIPA (NIPA-2011-(C1090-1111-0003)), and in part by the R&D program of MKE/KCC (2009-S-001-01(KI002070)).

S. Lee, I. Shin and Y. Ho are with the Department of Information and Communications, Gwangju Institute of Science and Technology (GIST) (e-mail: {sblee, siy0808, hoyo}@gist.ac.kr).

Contributed Paper
Manuscript received 07/15/11
Current version published 09/19/11
Electronic version published 09/19/11.

configuration. The main contribution of our work is to propose a practical eye contact solution for immersive teleconferencing in interactive IPTV environments. The proposed system includes several 3D processing techniques such as image acquisition, preprocessing, disparity estimation, view synthesis, and image display. Furthermore, since one of the most important aspects of IPTV environments is real-time processing, the main processes of this system, which require massive computational time, are implemented by GPU programming.

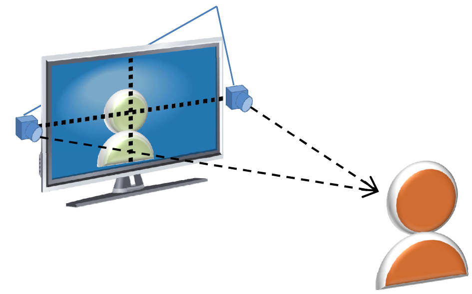


Fig. 1. Specification of the proposed system.

II. System Specification and Overall Framework

Before describing the entire process in detail, system specifications and assumptions, which satisfies with IPTV environments and simplifies the process, are introduced.

Figure 1 illustrates the specification of the proposed system. The proposed system utilizes two cameras on the left and right side of a display. The cameras are strictly fixed and the optical axes are faced toward the user. In addition, a single user is positioned within the recommended range of viewing distance from the display and gazes towards the middle of the display where the distant user appears [16]. In this setup, the existence of another user in between the prior user and the display is not allowed. In this paper, we introduce a compact and minimized camera setup for gazed-corrected image generation with stereo cameras satisfying IPTV environments. However, the setup of this system can be easily modified according to the applications and capabilities of other consumer devices such as laptops, cell phones, etc.

Figure 2 represents the overall system framework of the proposed system to generate the gaze-corrected image for IPTV environments. The entire process is divided into five major steps: image acquisition, preprocessing, disparity estimation, view synthesis, and image display.

In the preprocessing step, camera calibration is performed at first. After each camera is calibrated independently, the camera parameter is obtainable. From this, the images can be rectified using the derived camera parameters to adjust geometrical errors, such as horizontal and vertical misalignment of the stereo images. In order to reduce color inconsistency for the stereo images, we conduct the color correction. In the next step, disparity estimation, the preprocessed stereo images are down-sampled to reduce computational time. From the down-sampled images obtained from the current step, disparities for the virtual viewpoint are calculated. Since the disparity sequence is independently computed for each frame, it is temporally inconsistent and causes flickering artifacts. On this account temporal filtering is applied to the disparity sequence. In the fourth step, view synthesis is performed. With stereo images and the disparity map of the virtual viewpoint, the virtual image is synthesized. After occlusion decision and hole filling operation in the view synthesis stage, the gaze-corrected image is generated and finally displayed through the display.

III. Preprocessing of the Stereo Camera System

If the camera setup is built in the system, the following steps can be skipped in the practical environment. The reason is that several parameters computed from the preprocessing stage will be left unchanged, once the camera setup is fixed in the device.

A. Camera and Color Calibration

As mentioned before, the system consists of two cameras with convergent configuration toward the user. In order to make the following processes easier, it is essential to find the relative camera information through camera calibration. For this purpose, we apply the camera calibration algorithm to each camera in our camera system and obtain projection matrices for each camera [17]. The projection matrix is calculated by

$$\mathbf{P}_l = \mathbf{K}_l[\mathbf{R}_l \mid \vec{t}_l] \quad (1)$$

$$\mathbf{P}_r = \mathbf{K}_r[\mathbf{R}_r \mid \vec{t}_r] \quad (2)$$

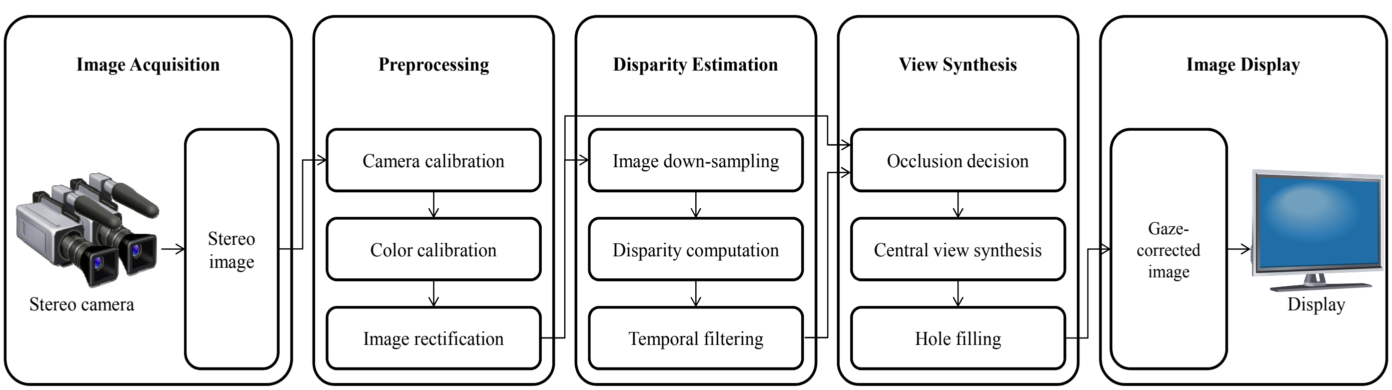


Fig. 2. System framework of gaze-corrected view generation for IPTV.

where $\mathbf{P}_l$ and $\mathbf{P}_r$ are the projection matrices of the left and right cameras, respectively. $\mathbf{K}$, $\mathbf{R}$, and $\vec{t}$ represent the camera intrinsic matrix, rotation matrix, and translation vector, respectively. These projection matrices are used for image rectification.

The color characteristics of captured images are usually inconsistent due to different camera properties, lighting conditions, and independent white balancing, even when the hardware type and the specification of the stereo cameras are the same. Thererfore, we resolve the color mismatch issue of the stereo images by using a color correction method [18].

B. Image Rectification

When we capture images using stereo cameras, there geometric errors exist due to the misalignment of the cameras. This problem causes the pixel mismatch in the vertical and horizontal directions between stereo images. In addition, non-uniform internal characteristics of the cameras, such as focal length and principal point, may also cause the geometric error.

Therefore, it is essential to correct geometric errors in the stereo images. Image rectification is one of image processing techniques to reduce the geometric errors [19]. Image rectification is the process for transforming the original image pairs into rectified image pairs which have parallel epipolar lines. From the image rectification process, we can rearrange the optical axes of the stereo cameras so that the optical axes are parallel and only horizontal disparities in the stereo images.

In order to minimize geometric errors, we first find the baseline of two cameras, the first basis of new axis $\vec{v}_1$.Then, we obtain the second basis $\vec{v}_2$ by calculating cross product of the optical axis vector of the left camera $\vec{z}_l$ and $\vec{v}_1$. The third basis $\vec{v}_3$ is easily obtained by calculating cross product of $\vec{v}_1$ and $\vec{v}_2$. Consequently, we can obtain the new projection matrices of two cameras, $\mathbf{P}_l'$ and $\mathbf{P}_r'$.

$$\vec{v}_1 = c_r - c_l \tag{3}$$

$$\vec{v}_2 = \vec{z}_l \times \vec{v}_1 \tag{4}$$

$$\vec{v}_3 = \vec{v}_1 \times \vec{v}_2 \tag{5}$$

(a) Original image (b) Rectified image

Fig. 3. Image rectification results.

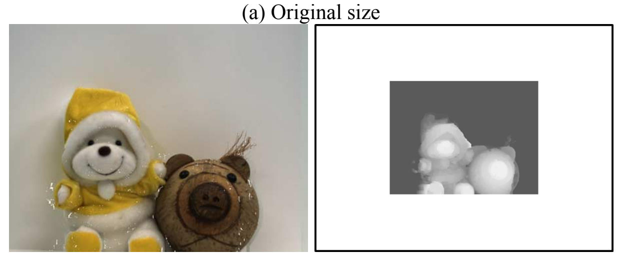
Figure 3 shows image rectification results with overlaid images of stereo cameras. As shown in the enlarged figures, the image rectification algorithm corrects the geometric errors.

IV. Disparity Estimation

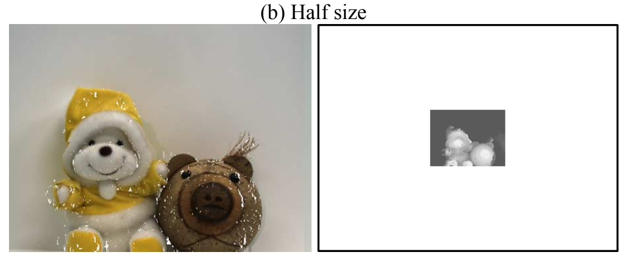
A. Image Down-sampling

In order to reduce computational time and realize real-time processing, we first perform image down-sampling prior to disparity computation. Since the characteristics of the disparity map are simpler than those of the color image, it is sufficient that we apply a view synthesis algorithm using reliable down-sampled disparity map. Figure 4 shows the synthesized images with various down-sample rates. The white dots as shown in left figures are holes caused by incorrect disparities and truncation errors. From the synthesized images, proper down-sampling does not induce the serious degradation of image quality.

(a) Original size



(b) Half size



(c) Quarter size

Fig. 4. synthesized images with various down-sample rates.

B. Disparity Computation

With the down-sampled stereo images, disparities are computed. In the proposed system, we apply a global stereo matching algorithm using energy minimization to find corresponding points [11]. The energy function contains two assumptions. One is that a corresponding pixel should have the similar intensity. The other is that neighboring pixels are likely to have the similar disparities. These two assumptions are represented as the data term and the piecewise smoothness term of the energy function. The energy function $E_{\mathbf{X}}(d)$ is defined by

$$E_{\mathbf{X}}(d) = E_{data,\mathbf{X}}(d) + E_{smooth,\mathbf{X}}(d) \tag{6}$$

where $E_{data,\mathbf{X}}$ and $E_{smooth,\mathbf{X}}$ represent the data term and the piecewise smoothness term, respectively. d is the disparity for the virtual-viewpoint.

Unlike the conventional stereo matching algorithms, we directly estimate the disparity map of the virtual viewpoint to further reduce computational time. Figure 5 describes the disparity estimation process of virtual-viewpoint.

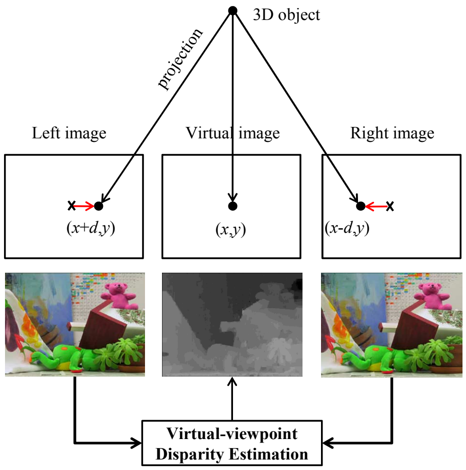


Fig. 5.Disparity estimation of virtual viewpoint.

In this disparity computation, the data term can be defined by

$$E_{data,\mathbf{X}}(d) = E_{data}(x, y, d) = \left| I_R(x-d, y) - I_L\{x+d, y\} \right| \quad (7)$$

In order to minimize the energy function, hierarchical belief propagation (HBP) is used [20], [21]. It passes messages called belief around the graph constructed by four-connected image grid. The message can be updated by several iterations. For a single iteration step, each neighbor pixel computes its message using the messages of the previous iteration. The smoothness term of the energy function is defined by

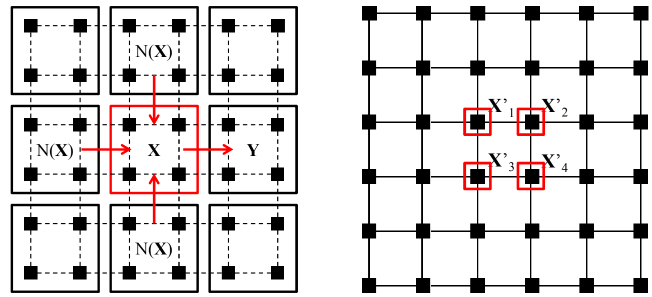
$$E_{smooth,\mathbf{X}}(d) = \sum_{\mathbf{Y}\in N(\mathbf{X})} M_{\mathbf{Y},\mathbf{X}}(d) \quad (8)$$

$$M^{t}_{\mathbf{X},\mathbf{Y}}(d) = \arg\min_{d_{\mathbf{X}}}\{E_{data,\mathbf{X}}(d_{\mathbf{X}}) + \sum_{\mathbf{P}\in N(\mathbf{X}),\mathbf{X}\neq\mathbf{Y}} M^{t-1}_{\mathbf{P},\mathbf{X}}(d_{\mathbf{x}}) + h(d_{\mathbf{X}}, d)\} \quad (9)$$

$$h(d_{\mathbf{X}}, d) = \rho \cdot \min\{|d_{\mathbf{X}} - d|, \eta\} \quad (10)$$

where $M^{t}_{\mathbf{X},\mathbf{Y}}(d)$ represents the message passed from the pixel **X** to its neighbor pixel **Y** and $h(d_{\mathbf{X}}, d)$ represents the jump cost function that adopted the truncated linear model. ρ and η represents the weighting factor and the constant value controlling the increase of the jump cost function. In this paper, we set ρ=10 and η=L/8, where L represents the number of disparity levels. Figure 6(a) shows the process of message propagation.

The HBP algorithm is performed from the coarse-to-fine manner to reduce the complexity of long range propagation and iteration of messages. This is called by multi-grid belief propagation. First, the initial message is calculated at the coarsest level. Then, the computed messages are used at the next finer level. Figure 6 shows two-level multi-grid BP. The message of pixel **X** at level *i* is used to compute four messages of **X***'* at the next level *i*-1.



(a) level *i* (b) level *i*-1

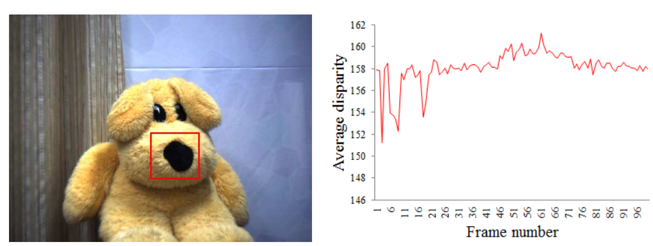
Fig. 6. Process of message propagation with multi-grid.

C. *Temporal Filtering of Disparity Sequences*

The stereo matching algorithm was originally developed to compute the disparity map for the still images. However, since we deal with the stereo sequences and there is no consideration of temporal consistency in the HBP algorithm, the appliance of the same algorithm to the sequences is not desired. This algorithm fails to generate the temporally consistent disparity sequence.

Ideally, in the case of static objects or background, disparities should be the same for each frame if the camera configuration is fixed. However, since the disparity computation is independently performed for each frame, disparities of the static regions fluctuate. The temporal inconsistency problem of the disparity sequences affects the rendering quality of the synthesized images and leads to flickering artifact discomforting viewers.

In order to quantitatively investigate the disparity fluctuation, we calculated the average disparity of the static region, as shown in Fig. 7. Figure 7(a) shows the static region for 100 frames and Fig. 7(b) shows the average disparity. As shown in Fig. 7(b), the average disparities of 100 frames are severely fluctuated even though disparities of the static region are unchanged.



(a) Static region for 100 frames (b) Disparity fluctuation

Fig. 7. Disparity fluctuation of static regions.

As mentioned before, the disparity sequences are temporally inconsistent since the original stereo matching algorithm is independently operated for each frame. Therefore, we modify the energy function by adding a temporal consistency term [22]. The temporal consistency term refers to the previous and next disparities when estimating the current disparities. The modified version of energy function and the temporal consistency term $E_{temp,\mathbf{x}}(d)$ are defined by

$$E_{\mathbf{x}}(d) = E_{data,\mathbf{x}}(d) + E_{smooth,\mathbf{x}}(d) + E_{temp,\mathbf{x}}(d) \tag{11}$$

$$E_{temp,\mathbf{x}}(d) = \lambda \frac{|d - d_{prev,\mathbf{x}}| + |d - d_{next,\mathbf{x}}|}{2} \tag{12}$$

where λ is the weighting factor. $d_{prev,\mathbf{x}}$ and $d_{next,\mathbf{x}}$ represent the disparities of the previous and the next frame, respectively.

V. Gaze-corrected View Synthesis

With the disparity map of the virtual viewpoint, we can synthesize the gaze-corrected image. The detailed process of view synthesis is depicted in Fig. 8. Left and right images are shifted to virtual viewpoint using the disparity map. Then, the occlusion areas are decided. After two synthesized images are integrated and the remaining holes are filled, we can finally obtain the gaze-corrected image.

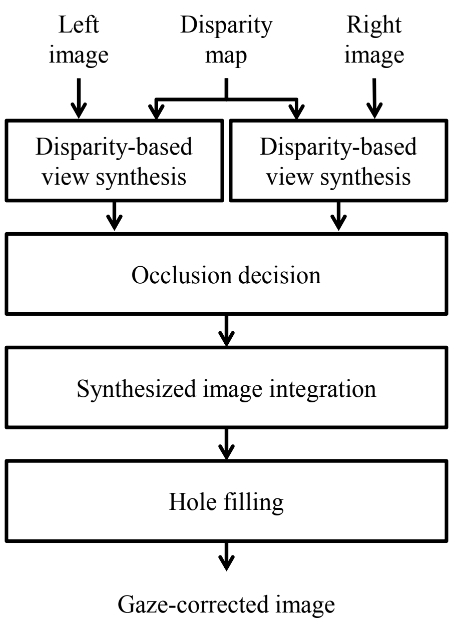


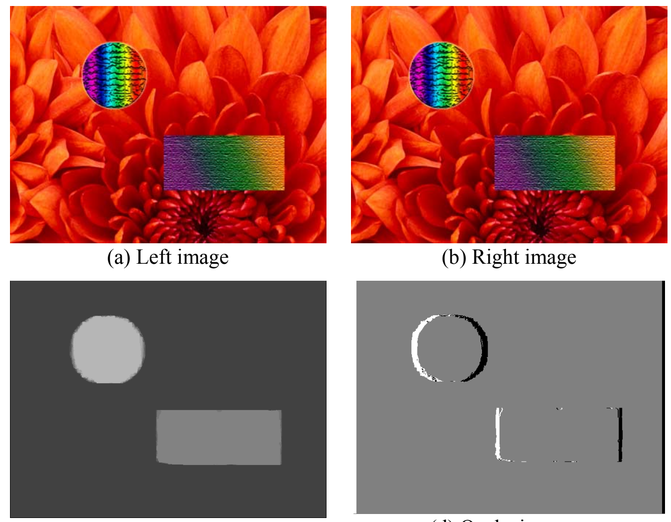
Fig. 8. Flow chart of view synthesis.

A. Occlusion Decision

Since we compute the disparity map of virtual viewpoint, we can obtain the virtual image using disparities. In order to avoid the view shift of occlusion areas, it is required to decide occlusion areas of virtual viewpoint before shifting left and right images. Therefore, we make the occlusion map with four labels as defined by

$$O_v(x,y) = \begin{cases} A & \text{Occluded from } I_L \\ B & \text{No occlusion} \\ C & \text{Occluded from } I_R \\ D & \text{Occluded from } I_L \text{ and } I_R \end{cases} \tag{13}$$

Occluded pixels are labeled by shifting view from the virtual viewpoint to the left and right viewpoints. Figure 9 shows the example of occlusion map decision which has the following four labels: A=255, B=128, C=0, and D=1.



(a) Left image (b) Right image

(c) Disparity map (d) Occlusion map

Fig. 9. Example of occlusion map decision.

B. Synthesized Image Integration

The second step of view synthesis is to integrate the left and right images to the virtual viewpoint. Although two shifted images contain disocclusion areas and holes, many of them can be reduced via image integration. There are three methods for image integration: left reference, right reference and view blending. The left/right reference method initially fills with the left-/right- shifted image, then fills the disocclusion areas and holes with the image of the other viewpoint. The position of pixels can be easily calculated by

$$I_v(x,y) = I_L\{x + D(x,y), y\} = I_R\{x - D(x,y), y\} \tag{14}$$

where $D(x,y)$ represents the disparity at pixel (x, y).

A view blending method uses the weighting sum of two images. It is represented by

$$\begin{aligned} I_v(x,y) = (1-\alpha) I_L\{x + D(x,y), y\} + \\ \alpha I_R\{x - D(x,y), y\} \end{aligned} \tag{15}$$

where α represents the weighting factor. In this process, we set α=0.5. Since the view blending method causes boundary mismatch artifact as many conventional works mentioned, visual quality is degraded. Therefore, we use the left/right reference using the occlusion map in this process.

C. Hole Filling

Although occlusion decision and synthesized image integration process fills the outmost of pixels of the virtual viewpoint, unknown holes which cannot find the same fetch from the reference images due to wrong disparities or truncation error still remain. Thus, we have to find the most plausible disparities by using the information of neighbor pixels. The representative hole-filling methods are image interpolation and inpainting [23], [24]. In order to obtain the

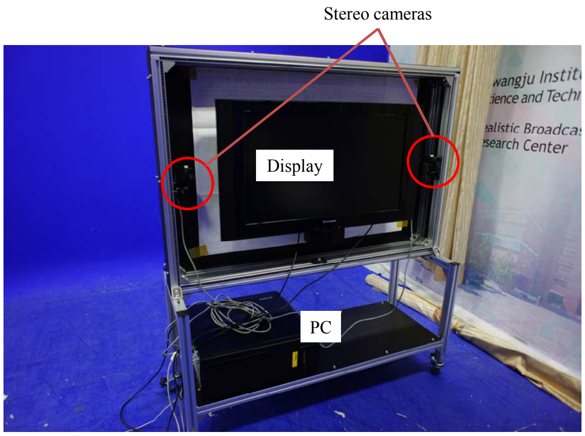
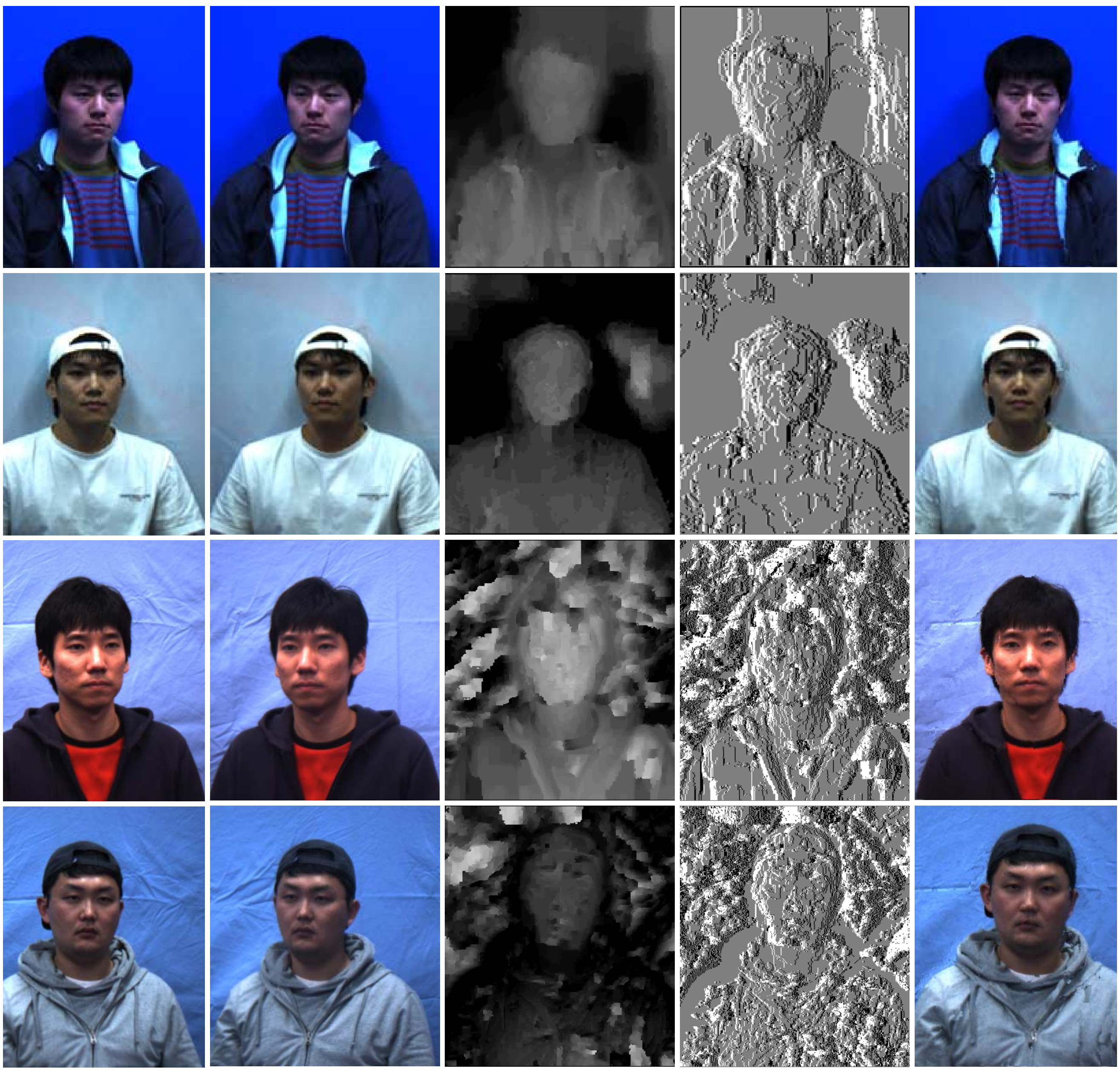


Fig. 10. System construction.

best quality of the hole-filled image, we use the neighboring background and their geometric information. This is because the information of background pixels is more reliable compared to that of foreground pixels by definition of the disocclusion. Hence, we fill out holes with neighboring pixels containing background disparity.

VI. Experimental Results and Analysis

In order to evaluate our algorithm, we constructed the stereo camera system, as shown in Fig. 10. The display size is 42 inches and the viewing distance is 2.7m, as the recommendation denoted [16]. The convergence point of the stereo cameras is at the maximum viewing distance. The camera baseline is dependent to the display size, where it is around 1.0m in these experiments. The resolution of the stereo cameras is 1280×960 at 15 fps. For real-time disparity estimation, the stereo images are down-sampled up to 640×480.



(a) Left image (b) Right image (c) Disparity map (d) Occlusion map (e) Gaze-corrected image

Fig. 11. Results of gaze-corrected view synthesis.

A. Disparity Estimation and View Synthesis

Figure 11 shows the results of gaze-corrected view synthesis. These results are obtained by different lighting conditions and the background. Although we can observe some errors near the object boundary, we have realized the natural gaze-corrected images, as shown in Fig. 11(e). Figure 12 shows the enlarged figures of gaze-corrected images. From Fig. 12(c), we noticed that the eye contact problem is efficiently resolved.

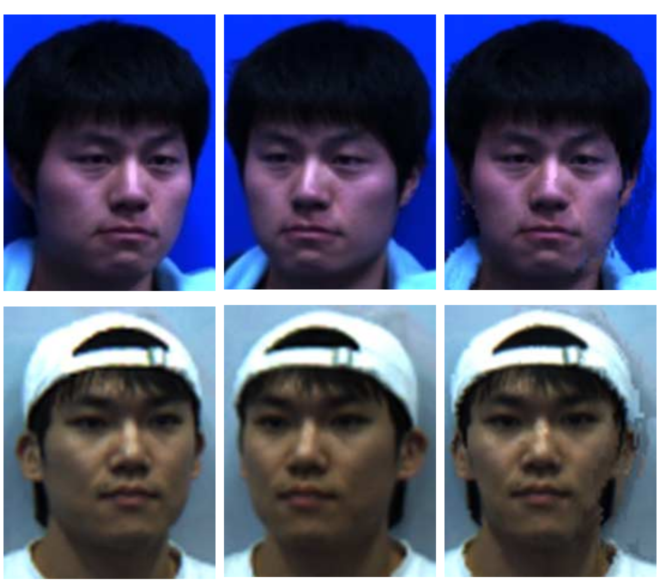
Since the large disparity fluctuation can lead to flickering artifact of the virtual image, quantitative verification is operated by calculating the average disparity of static regions and checking the amount of disparity fluctuation.

Figure 13 shows the average disparity of the static regions, as denoted in Fig. 7. The dash and solid lines represent the average disparity of the independent disparity estimation and the result with temporal filtering, respectively. As shown in Fig. 13, in the case of independent operation, average disparities are severely fluctuated, whereas the proposed disparity estimation with temporal filtering remarkably reduced the disparity fluctuation.

B. GPU Implementation

We implemented parallel programming that is executed on the GPU for real-time implementation. The architecture of the CPU and GPU are different. While the GPU possesses many cores capable of calculating floating point operation, it has a small number of instruction control units. Thus, we call the GPU structure a Single Instruction Multiple Threads (SIMT). Accordingly, GPU programming is greatly suitable for general image processing algorithm due to the fact that all pixels are likely to have the same operation. The data independency is an important condition of SIMT parallel processing. That is, all pixels should be executed simultaneously.

In these experiments, we implemented the entire process with the parallel GPU programming. Table I shows the execution time of each process. From the results, the proposed system enabled the real-time processing up to 15.6 fps.



(a) left image (b) right image (c) gaze-corrected image

Fig. 12. Enlarged images of view synthesis results.

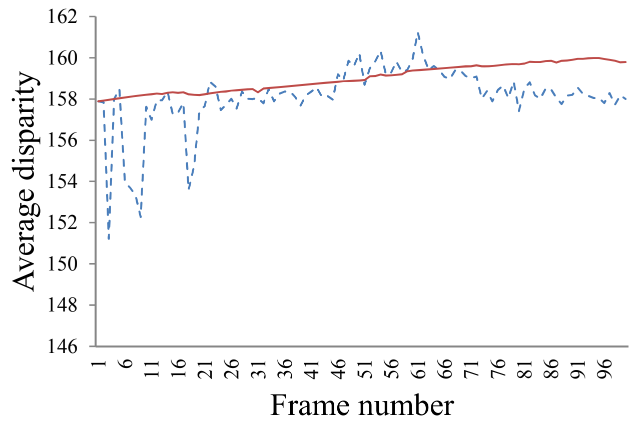


Fig. 13. Average disparity variation.TABLE I

TABLE I
EXECUTION TIME

Process	Computational time (ms)
Image down-sampling	4.19
Image rectification	0.4
Disparity computation	1.33
Belief propagation	53.6
View synthesis	4.6
Total	**64.12 (15.60 fps)**

VII. CONCLUSIONS

In this paper, we have presented a new approach to generate a gaze-corrected image from the stereo camera system in the IPTV environments. We have used 3D image processing techniques, disparity estimation and view synthesis, to solve the eye contact problem. In order to reduce computational time, we have modified the energy function in the disparity computation step. We then performed the temporal filtering of disparity sequences to reduce the disparity fluctuation and flickering artifacts of the virtual images. In the last step, we finally synthesized a virtual image using the disparity map of the virtual viewpoint and the occlusion map. Experimental results have verified that our proposed system realized the natural gaze-corrected image efficiently in real-time. Therefore, our proposed system could be helpful for various applications since the system can be easily modified according to the applications and capabilities of other consumer devices.

REFERENCES

[1] J. Freeman and S. E. Avons, "Focus group exploration of presence through advanced broadcast services," *In Proceedings of SPIE, Human Vision and Electronic Imaging*, vol. 3959, pp. 530–539, 2000.

[2] J. Y. Choi, W. D. Neve, and Y. M. Ro, "Towards an automatic face indexing system for actor-based video services in an IPTV environment," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 1, pp. 147–155, Feb. 2010.

[3] S. Huang and J. Wang, "A low-cost desktop videoconferencing codec: an adaptive Motion-JPEG design," *IEEE Transactions on Consumer Electronics*, vol. 40, no. 4, pp. 944–950, Nov. 1994.

[4] S. M. Kuo, Y. C. Huang, and Z. Pan, "Acoustic noise and echo cancellation microphone system for videoconferencing," *IEEE*

Transactions on Consumer Electronics, vol. 41, no. 4, pp. 1150–1158, Nov. 1995.

[5] H. Kwon, H. Han, S. Lee, W. Choi, and B. Kang, "New video enhancement preprocessor using the region-of-interest for the videoconferencing," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2644–2651, Nov. 2010.

[6] K. Tan, I. N. Robinson, B. Culbertson, and J. Apostolopoulos, "Enabling genuine eye contact and accurate gaze in remote collaboration," *In Proceedings of IEEE International Conference on Multimedia and Expo,* pp. 430–435, July 2010.

[7] D. Nguyen and J. Canny, "Multiview: spatially faithful group video conferencing," *In Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, pp. 799–808, April 2005.

[8] S. Kim, S. Lee, and Y. Ho, "Three-dimensional natural video system based on layered representation of depth maps," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 3, pp. 1035–1042, Aug. 2006.

[9] E. Lee and Y. Ho, "Generation of multi-view video using a fusion camera system for 3D displays," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2797–2805, Nov. 2010.

[10] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *In Proceedings of SIGGRAPH'04*, pp. 600–608, Aug. 2004.

[11] D. Sharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *In Proceedings of IEEE Workshop on Stereo and Multi-Baseline Vision*, pp. 131–140, Dec. 2001.

[12] O. Schreer, N. Atzapadin, and I. Feldmann, "Multi-baseline disparity fusion for immersive videoconferencing," *In Proceedings of International Conference on Immersive Telecommunications*, pp. 27–29, May 2009.

[13] P. Kauff and O. Schreer, "An immersive 3D video-conferencing system using shared virtual team user environments," *In Proceedings of International Conference on Collaborative Virtual Environments*, pp. 105–112, Oct. 2002.

[14] O. Schreer, N. Brandenburg, S. Askar, and M. Trucco, "A virtual 3D video-conference system providing semi-immersive telepresence: a real-time solution in hardware and software," *In Proceedings of eBusiness and eWork*, pp. 184–190, Oct. 2001.

[15] N. Atzpadin, P. Kauff, and O. Schreer, "Stereo analysis by hybrid recursive matching for real-time immersive video conferencing," *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Immersive Telecommunications*, vol. 14, no. 3, pp. 321–334, March 2004.

[16] M. Sugawara, K. Mitani, M. Kanazawa, F. Okano, and Y. Nishida, "Future prospects of HDTV – technical trends toward 1080p," *SMPTE Journal*, vol. 115, pp. 10–15, 2006.

[17] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[18] J. Jung and Y. Ho, "Color correction method using gray gradient bar for multi-view camera system," *In Proceedings of International Workshop on Advanced Image Technology*, pp. MP.C4(1–6), Jan. 2009.

[19] Y. Kang and Y. Ho, "Geometrical compensation for multi-view video in multiple camera array," *In Proceedings of International Symposium on Electronics and Marine*, vol. 1, pp. 83–86, Sep. 2008.

[20] J. Sun, N. Zheng, and H. Shum, "Stereo matching using belief propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* , vol. 25, no. 7, pp. 787–800, July 2003.

[21] Q. Yang, L. Wang, and N. Ahuja, "A constant-space belief propagation algorithm for stereo matching," *In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1458–1465, June 2010.

[22] S. Lee and Y. Ho, "Temporally consistent depth map estimation using motion estimation for 3DTV," *In Proceedings of International Workshop on Advanced Image Technology*, pp.149(1–6), Jan. 2010.

[23] K. Oh, S. Yea, and Y. Ho, "Hole filling method using depth based in-painting for view synthesis in free viewpoint television and 3-D video," *In Proceedings of Picture Coding Symposium*, pp. 39(1–4), May 2009.

[24] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of Graphics Tools*, vol. 9, no. 1, pp. 25–36, 2004.

BIOGRAPHIES

Sang-Beom Lee (S'06) received his B.S. degree in Electric & Electronic Engineering from Kyungpook National University (KNU), Korea, in 2004 and M.S. degree in Information and Communication Engineering at the Gwangju Institute of Science and Technology (GIST), Korea, in 2006. He is currently working towards his Ph.D. degree in the Department of Information and Communications at GIST, Korea. His research interests include digital image processing, video coding, 3D television, depth estimation, 3D video coding, and realistic broadcasting.

In-Yong Shin received his B.S. degree in Electric & Electronic Engineering from Inha University, Korea, in 2009. He is currently working towards his M.S. degree in the Department of Information and Communications at GIST, Korea. His research interests include digital image processing, GPU programming, 3D television, depth estimation, 3D video coding, and realistic broadcasting.

Yo-Sung Ho (M'81-SM'06) received both B.S. and M.S. degrees in electronic engineering from Seoul National University (SNU), Korea, in 1981 and 1983, respectively, and Ph.D. degree in Electrical and Computer Engineering from the University of California, Santa Barbara, in 1990. He joined the Electronics and Telecommunications Research Institute (ETRI), Korea, in 1983. From 1990 to 1993, he was with Philips Laboratories, Briarcliff Manor, New York, where he was involved in development of the advanced digital high-definition television (AD-HDTV) system. In 1993, he rejoined the technical staff of ETRI and was involved in development of the Korea direct broadcast satellite (DBS) digital television and high-definition television systems. Since 1995, he has been with the Gwangju Institute of Science and Technology (GIST), where he is currently a professor in the Information and Communications Department. His research interests include digital image and video coding, image analysis and image restoration, advanced coding techniques, digital video and audio broadcasting, 3D television, and realistic broadcasting.