

The Design and Construction of a Movable Image-Based Rendering System and Its Application to Multiview Conferencing

S. C. Chan · Z. Y. Zhu · K. T. Ng · C. Wang · S. Zhang ·
Z. G. Zhang · Zhongfu Ye · H. Y. Shum

Received: 17 March 2010 / Revised: 16 November 2010 / Accepted: 16 November 2010 / Published online: 3 December 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Image-based rendering (IBR) is an promising technology for rendering photo-realistic views of scenes from a collection of densely sampled images or videos. It provides a framework for developing revolutionary virtual reality and immersive viewing systems. While there has been considerable progress recently in the capturing, storage and transmission of image-based representations, most multiple camera systems are designed to be stationary and hence their ability to cope with moving objects and dynamic environment is somewhat limited. This paper studies the design and construction of a movable image-based rendering system based on a class of dynamic representations called plenoptic videos, its associated video processing algorithms and an application to multiview audio-visual conferencing. It is constructed by mounting a linear array of 8 video cameras on an electrically controllable wheel chair and its motion is controllable manually or remotely through wireless LAN by means of additional hardware circuitry. We also developed a real-time object tracking algorithm and utilize the motion information computed to adjust continuously the azimuth or rotation angle of the movable IBR system

in order to cope with a given moving object in a large environment. Due to imperfection in tracking and mechanical vibration encountered in movable systems, the videos may appear very shaky and a new video stabilization technique is proposed to overcome this problem. The usefulness of the system is illustrated by means of a multiview conferencing application using a multiview TV display. Through this pilot study, we hope to disseminate useful experience for the design and construction of movable IBR systems with improved viewing freedom and ability to cope with moving object in a large environment.

Keywords Image-based rendering · Movable IBR systems · Tracking and stabilization · Multiview conferencing

1 Introduction

Image-based rendering/representation (IBR) [1–16] is a promising technology for rendering new views of scenes from a collection of densely sampled images or videos. It has potential applications in virtual reality, immersive television and visualization systems. Central to IBR is the *plenoptic function* [17], which describes all the radiant energy that can be perceived by the observer at any point (V_x, V_y, V_z) in space and time τ . The plenoptic function is thus a 7-dimensional function of the viewing position (V_x, V_y, V_z) , the azimuth and elevation angle (θ, ϕ) , time τ , and wavelengths λ . Traditional images and videos are just 2D and 3D special cases of the plenoptic function. In principle, one can reconstruct any views in space and time if sufficient number of samples of the plenoptic function is available. In other words, we may generate new views of the scene from a collection of densely sampled images or videos. Depending

S. C. Chan (✉) · Z. Y. Zhu · K. T. Ng · C. Wang · S. Zhang ·
Z. G. Zhang
Department of Electrical and Electronic Engineering,
The University of Hong Kong,
Pokfulam Road,
Hong Kong, China
e-mail: scchan@eee.hku.hk

Z. Ye
University of Science and Technology of China,
Hefei, People's Republic of China

H. Y. Shum
Microsoft Corporation,
Redmond, WA, USA

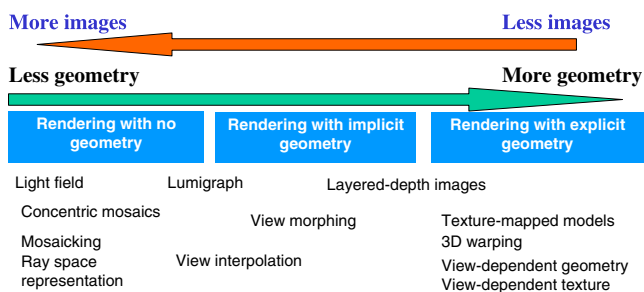


Figure 1 Spectrum of IBR representations.

on the functionality required, there is a spectrum of IBR as shown in Fig. 1. They differ from each other in the amount of geometry information of the scenes/objects being used. A recent survey of IBR can be found in [18, 19].

At one end of the spectrum, like traditional texture mapping, we have very accurate geometric models of the scenes and objects, say generated by animation techniques, but only a few images are required to generate the textures. At the other extreme, light-field [4] or lumigraph [3] rendering relies on dense sampling and very little geometry information in form of depth maps for rendering without recovering the exact 3D models. An important advantage of the latter is its superior image quality, compared with 3D model building for complicated real world scenes.

Since capturing 3D models in real-time is still a very difficult problem, lightfield- or lumigraph-based dynamic IBR representations with little amount of geometry information have received considerable attention in immersive TV (also called 3D or multi-view TVs) applications. Because of the multidimensional nature of the plenoptic function and the scene geometry, much research has been devoted to the efficient capturing, sampling, rendering and compression of IBR. There has been considerably progress in these areas since the pioneer work of *lumigraph* by Gortler et al. [3] and *lightfield* by Levoy and Hanrahan [4]. Other IBR representations include the 2D panorama [6, 7], Chen and Williams' view interpolation [9], McMillan and Bishop's plenoptic modeling [5], layer depth images [8] and the 3D concentric mosaics [10], etc. Motivated by lightfields and lumigraphs, the authors have developed a real-time system for capturing and rendering a simplified dynamic lightfield called the "plenoptic videos" [20–24] with four dimensions. It is a simplified dynamic lightfield, where videos are taken along line segments as shown in Fig. 2, instead of a 2D plane, to simplify the capturing hardware for dynamic scenes.

While there has been considerable progress recently in the capturing, compression and transmission of image-based representations [18, 19, 25], most multiple camera systems are not designed to be movable so that the viewpoints are somewhat limited and usually cannot cope with moving objects in a large environment. Apart from many

system design issues, there are also many important problems and difficulties in realizing these systems such as object tracking, video stabilization and enhancement, etc. This motivates us to study the design and construction of a movable image-based rendering system based on a class of dynamic IBR called plenoptic videos [21–24] and its associated video processing algorithms. In particular, a linear camera array consisting of 8 video cameras is mounted on an electrically controllable wheel chair and its motion can be controlled manually or remotely by means of additional hardware circuitry. The system can potentially provide improved viewing freedom to users and ability to cope with moving objects in a large environment. Moreover, with the advance of technology, multi-view displays are becoming available [26] and their costs have been reducing dramatically. It is predicted that 3D or multi-view TVs will be another trend after high-definition TVs. This breakthrough also motivates us to study in this paper an important application of our movable system in multiview audio-visual conferencing. In particular, we developed an automatic real-time object tracking algorithm and use the motion information computed to adjust continuously the azimuth or rotation angle of the movable IBR system in order to cope with the moving speaker. Due to imperfection in tracking, the videos may appear shaky and new video stabilization technique is proposed to overcome this problem. In particular, feature-based tracking is employed to estimate the global motion at each time instant. The vibration components in the computed velocity are estimated using a novel Kalman filter-based frequency tracker. A time-varying and adaptive notch filter is then proposed to remove these vibration components and obtain a smooth motion path. Finally, an affine model is used to wrap the original images to the stabilized motion path. Through this pilot study, we hope to disseminate useful experience for the design and construction of movable IBR systems with improved viewing freedom and ability to cope with moving object in a large environment.

The paper is organized as follows: Section 2 reviews the concept of object-based approach to plenoptic videos. The design and development of the proposed prototype movable



Figure 2 Plenoptic videos: Multiple linear camera array of 4D simplified dynamic light fields with viewpoints constrained along line segments. The camera arrays developed at [24]. Each consists of 6 JVC video cameras.

plenoptic video system are described in Section 3. The details of other important processing functions such as object tracking and video stabilization will be given in Section 4. Section 5 is devoted to the compression issue and application of the proposed system to the multiview conferencing application. Experimental results will be presented to illustrate the usefulness of the proposed system. Finally conclusions are drawn in Section 6.

2 Object-Based Approach to Plenoptic Videos

In plenoptic videos, multiple linear arrays are employed to capture multiple video sequences in order to render immediate or novel views of the scene at other nearby positions. In [23, 24], two linear arrays, each hosting 6 JVC DR-DVP9ah video cameras, were used as shown in Fig. 2. More arrays can be connected together to form longer segments. As mentioned earlier, an initial segmentation of an IBR object is first obtained using a semi-automatic technique called Lazy snapping [27]. Tracking techniques using the level set method [28–31] are then employed to segment the objects at other video streams and subsequent time instants. From the segmented objects, approximate depth information for each IBR object can be estimated to render new views at different viewpoints. Due to possible segmentation errors around boundaries and finite sampling at depth discontinuities, natural matting is also adopted to improve the rendering quality when mixing IBR objects.

The object-based approach not only reduces the rendering artifacts due to depth discontinuities, it also provides object-based functionalities in coding and other applications. In particular, the IBR objects can be encoded individually by an MPEG-4 like object-based coding scheme [32, 33] which also includes additional information such as depth maps and alpha maps to facilitate rendering. Moreover, the user-defined IBR objects can be flexibly reconstructed at the decoder for rendering and other processing.

In this work, we constructed a movable IBR system by mounting a linear array of cameras on a movable wheel chair. Object tracking techniques are employed to assist the operator to steer the array to track a desirable moving object in a large environment. Due to imperfect tracking and mechanic vibration of the system, the plenoptic video captured may appear very shaky and the video stabilization technique to be described later will be employed to reduce this undesirable effect.

3 Construction of the Proposed Movable IBR System

As mentioned previously, the movable IBR system consists of a linear array of cameras mounted on an electrically

controllable wheel chair so as to cope with moving objects in a large environment and hence improve the viewing freedom of users. Figure 3 shows the movable IBR system that we have constructed. It consists of a linear array of 8 Sony HDR-TGIE high definition (HD) video cameras which is mounted on a FS122LGC wheel chair.

The motion of the wheel chair is originally controlled manually through a VR2 joystick and power controller modules from PG drives technology (PG DRIVE VR2 CONTROLLER URL <http://www.pgdt.com/products/vr2/index.html>). To make it electronically controllable, we examined the output of the joystick and generated the (x -, y -) motion control voltages to the power controller using a Devasys USB-I2C/IO micro-controller unit (MCU) (URL: <http://www.devasys.com/usbi2cio.htm>). By appropriately controlling these voltages, we can control the motion of the wheel chair electronically. Moreover, by using the wireless LAN of a portable notebook mounted on the wheel chair, its motion can be controlled remotely. By improving the mobility of the IBR capturing system, we are able to cope with moving objects in a large environment.

The HD videos are captured in real-time into the storage cards of cam-corders. They can be downloaded to PC for further processing such as calibration, depth estimation, and rendering using the object-based approach [23, 24, 34, 35]. For real-time transmission, the cam-corders are equipped with a composite video output which can be further compressed and transmitted. To illustrate the concept of multiview conferencing, a ThinkSmart IVS-MV02 Intelligent Video surveillance system (www.ivs-tech.com) was used to compress the (320×240) 30 frames/sec videos online, which can be retrieved remotely through the

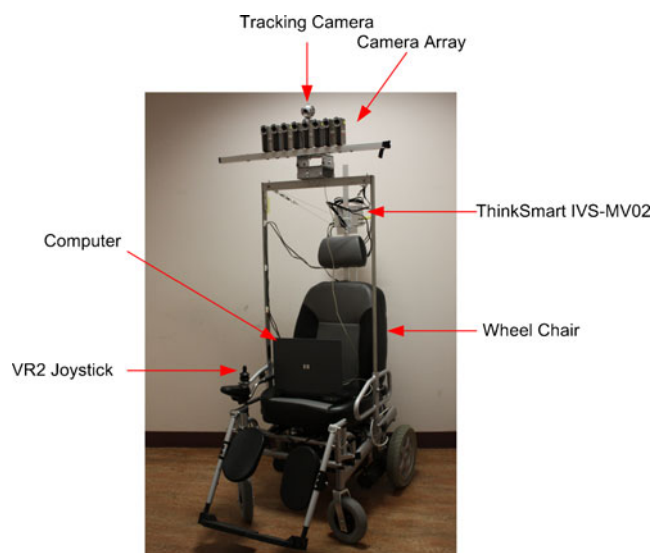


Figure 3 The proposed movable image-based rendering system.

wireless LAN for viewing or further processing. The system is built from Analog Device DSP and real-time compression at a bit rate of 400 kbps.

Before the cameras can be used for depth estimation, they must be calibrated to determine the intrinsic parameters as well as their extrinsic parameters, i.e. their relative positions and poses. This can be accomplished by using a sufficient large checkerboard calibration pattern. We follow the plane-based calibration method [36] to determine the projective matrix of each camera, which connects the world coordinate and the image coordinate. The projection matrix of a camera allows a 3D point in the world coordinate be translated back to the corresponding 2D coordinate in the image captured by that camera. This will facilitate depth estimation.

Experimental Results Figure 4(a) shows the snapshots of the cameras taken at several time instants. Using an initial segmentation obtained by Lazy snapping [27], the object at other time instants and views are tracked using level-set method [28–31, 34, 35]. Some tracking results are shown in Fig. 4(b) where the speaker is tracked with the boundary marked in green color. The depth maps of each object are then estimated and are shown in Fig. 4(c). Some renderings at other locations are also shown in Fig. 4(d). Since the stick has the same color as the background, it is rather difficult to extract it from the background. Fortunately, since it also has similar color as the background, the quality of renderings is not affected significantly.

Next, we shall discuss the object tracking technique for steering the array in order to track a desirable moving object in a large environment. Details about the video stabilization technique to compensate for the undesirable shaking effects during tracking motion of the system will also be described.

4 Object Tracking and Video Stabilization

4.1 Real-Time Object Tracking

In principle, the proposed system has two degrees of freedom. For simplicity, we only explore the angular domain so that complicated path planning of the movable IBR system can be avoided. Our tracking algorithm is based on the combination of the mean shift algorithm [37] and the Kalman filter [38]. At each frame, the Kalman filter is used to predict the object position, and the mean shift algorithm is used to obtain a more accurate position. The tracking starts by defining the object to be tracked by means of a user specified rectangular window in the screen. A separate webcam is connected to a Lenovo ThinkPad

T400 notebook computer for object tracking, since its interfacing is considerably simplified. Using the x -position of the object in the screen, a feedback signal is generated to steer the wheel chair and linear array angularly so as to position the object as close to the center of the screen as possible. Although it is interesting and useful to be able to estimate the focus of each video camera when they are set to auto-focus mode, the images so captured may not be focused on the given object. For simplicity, we shall focus on the fix focus case and the more difficult problem of self-calibration will be addressed in the future. In our current implementation, the tracking can be done completely in real-time in the ThinkPad T400 notebook computer.

Experimental Results Figure 5 shows the example tracking results of a moving object in a video conferencing application. It can be seen that the speaker can be satisfactorily tracked.

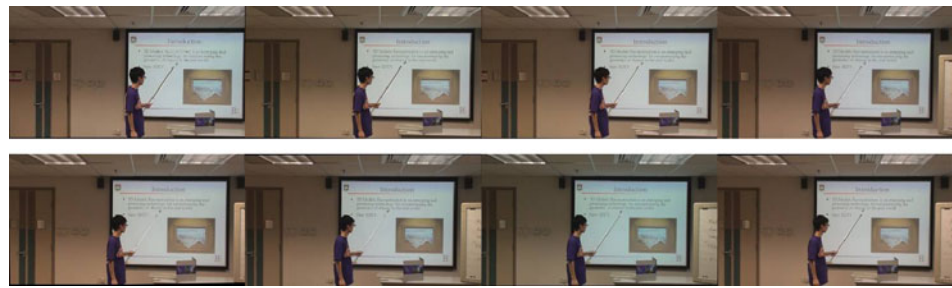
4.2 Video Stabilization

When the camera array is rotated either manually by the operator or automatically by the tracking algorithm, the video captured may be very shaky. To reduce this annoying effect, video stabilization should be employed [37, 39, 40]. The basic idea of video stabilization is to estimate the global motion of the camera, say by means of optical flow on the video sequence, so that this annoying motion can be compensated and hence the videos of the scenes can be stabilized. In conventional video stabilization algorithm for handheld devices, long term smoothing of global motion is performed to stabilize the videos. In moving mechanical systems, oscillations may arise and it is not easy to remove them completely by simple smoothing. In the proposed method, we adopt a Kalman filter-based method to estimate the vibration frequencies so that time-varying notch filtering can be applied to suppress them to a reasonably low level.

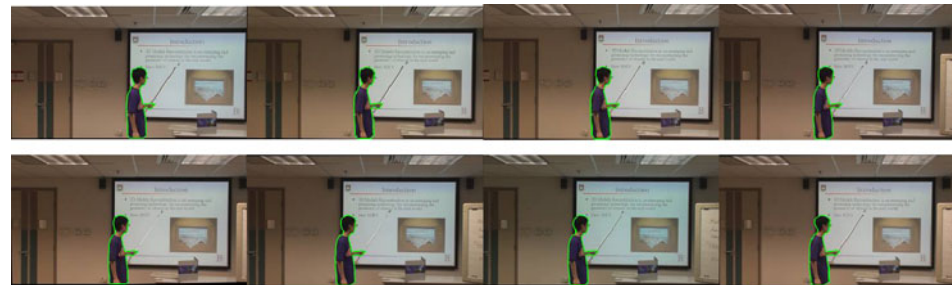
In the proposed algorithm, the global motion is first estimated by tracking feature points of the scenes. The Kanade-Lucas-Tomasi feature tracker [41] is employed. The histograms of the x - and y - velocities of these feature points at frames 21 and 450 are shown in Fig. 6. It can be seen that there is a major peak in the histograms which correspond to the global motion. Small isolated peaks usually correspond to the features extracted from the speaker. Since majority of the feature points are coming from the background, all the feature points are used to compute an affine model for global motion. The linear translation computed from the affine model gives the final x - and y - velocities of the camera.

Figure 7(a) shows the extracted global motion over time (in blue color) and the smoothed global motion (in red

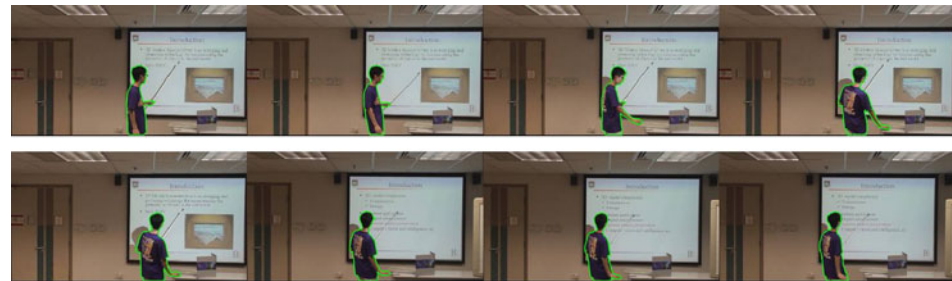
Figure 4 **a** Snapshot captured by the movable IBR system at a given time instant. **b** Segmented objects from the scene (i) different views at the same time instant, and (ii) same view at different time instants. **c** Depth maps computed at a time instant. **d** Upper: original cameras views 1 to 4. Lower: rendered views between (Left) cameras 1 and 2, and (Right) cameras 3 and 4. Note the rendered views are moved forward and away from the camera array.



(a): Snapshot captured by the movable IBR system at a given time instant.

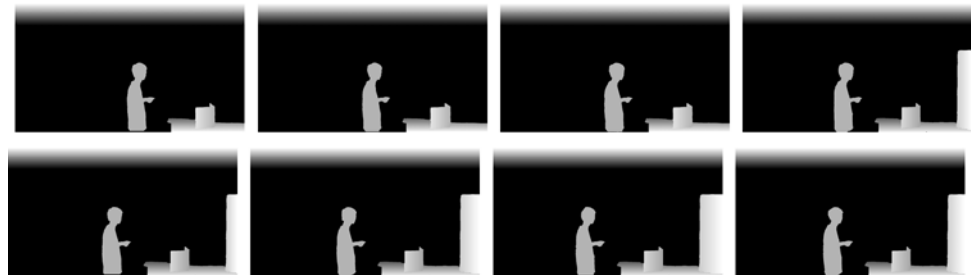


(i)



(ii)

(b): Segmented objects from the scene (i) different views at the same time instant, and (ii) same view at different time instants.

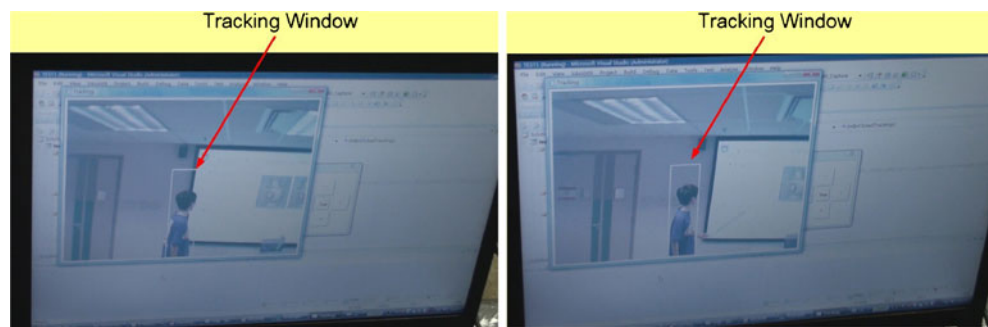


(c): Depth maps computed at a time instant.



(d): Upper: original cameras views 1 to 4. Lower: rendered views between (Left) cameras 1 and 2, and (Right) cameras 3 and 4. Note the rendered views are moved forward and away from the camera array.

Figure 5 Example tracking results of a moving object at two time instants.



color). As mentioned, oscillations are observed, especially when the system is moving and about to settle down. To effectively remove these oscillations and obtain a smooth motion path, we estimate the frequency of the oscillation using the Kalman filter-based (KF) frequency tracking algorithm proposed in [42]. Using the AIC criterion, it was found that the order M is equal to 3 and 2 frequency components should be used in tracking the x - and y -velocities. The tracking results are shown in Fig. 7(b). It can be seen that the two high-frequency components (4–5 Hz and 8–10 Hz) of the x - and y -directions are similar. They seem to come from the natural fundamental vibration frequency of the system and its 2nd harmonic. These two undesirable components can be effectively removed by

applying a time-varying adaptive notch filter to the original x - and y -velocities signals. More precisely, the two frequencies detected in y are used to construct a notch filter to filter out oscillations in the x - and y -velocity signals. A 2nd-order IIR notch filter is employed twice to remove the fundamental and its harmonic with a Q factor of 1, i.e. the bandwidth of filter is around $f_{\text{notch}}/4$, where f_{notch} is the notch frequency of the filter. After that, the x - and y -velocity signals are further smoothed using a first order IIR filter with a pole of 0.9. The smoothed velocity signals are shown in Fig. 7(a) in red line. They are then used to modify the translation term of the affine model computed previously. We found that the rotational parameters are quite stable and hence their values are not compensated. Using

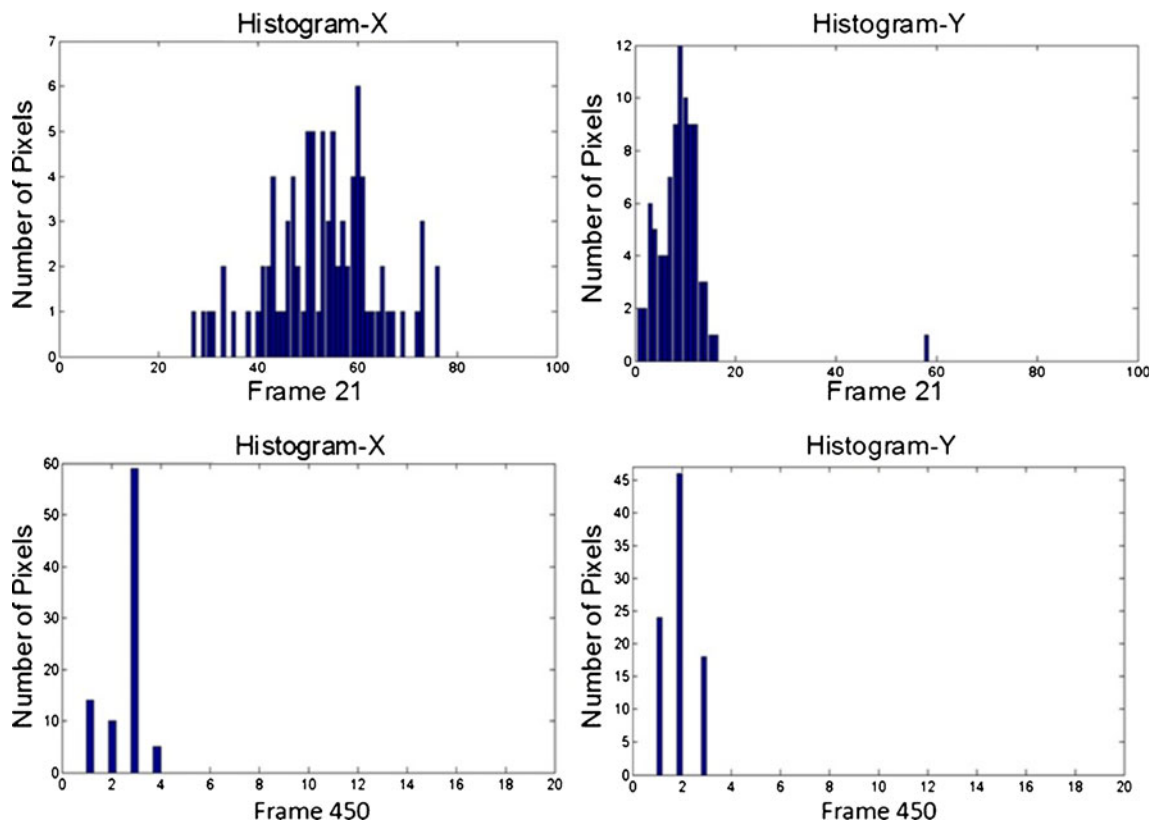


Figure 6 Histograms of the x - and y - velocities at frames 21 and 450.

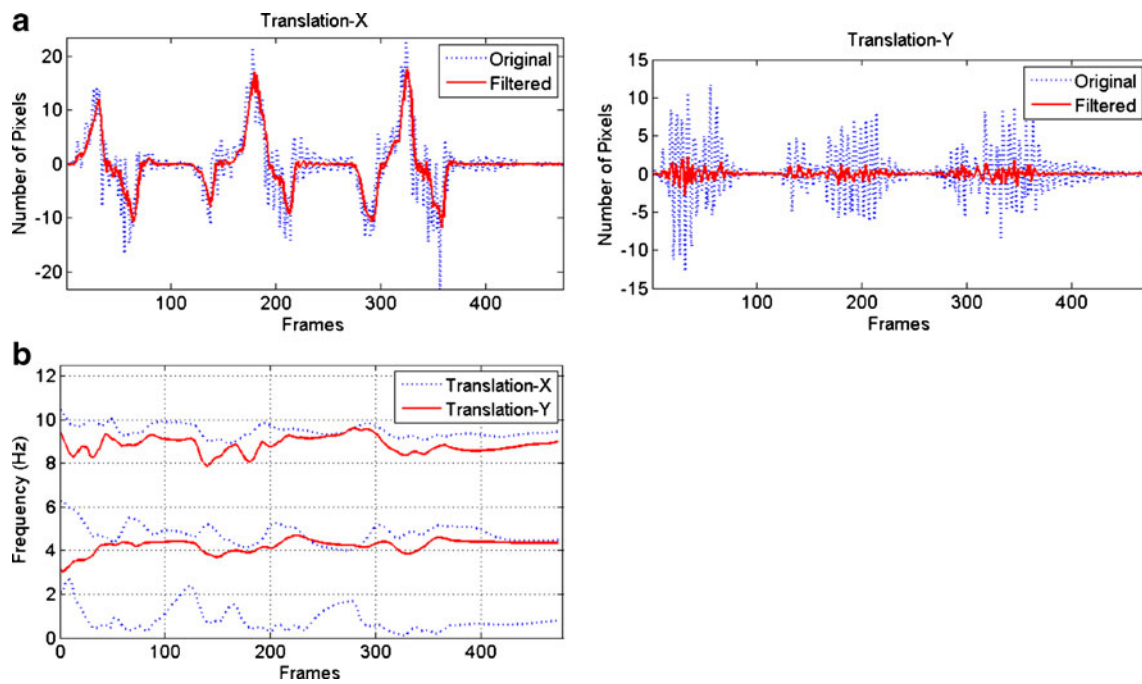


Figure 7 Motion estimation and stabilization results: **a** extracted global motion over time (*in blue*) and smoothed motions after adaptive notch filtering and 1st order recursive smoothing (*in red*), **(b)**.

this affine model between consecutive images, full-frame warps are performed on the original images to the filtered motion model so as to stabilize the videos.

After motion compensation, some parts of the compensated images at the boundary may be missing. These missing areas can be filled by images or motion inpainting techniques. For simplicity and avoiding different inpainting algorithms from affecting the compression results, we simply reduce slightly the resolution of the video to avoid this problem (Fig. 8).

5 Compression and Multiview Audio-Visual Conferencing

As mentioned earlier, multi-view displays are becoming more assessable recently and their costs have been reducing dramatically. This motivates us to study in this paper an important application of our movable system for multiview conferencing. More precisely, using the automatic real-time object tracking algorithm described in Section 4, the position of the object on the screen was estimated and the rotation of the movable IBR system was adjusted continuously in order to track the moving speaker. The multiple videos can either be recorded or compressed online using the ThinkSmart IVS-MV02 Intelligent Video surveillance systems. The compressed videos are decoded in a PC and are filtered and multiplexed together in order to be displayed on a Newsight 42" multiview AD3 TV. To speed

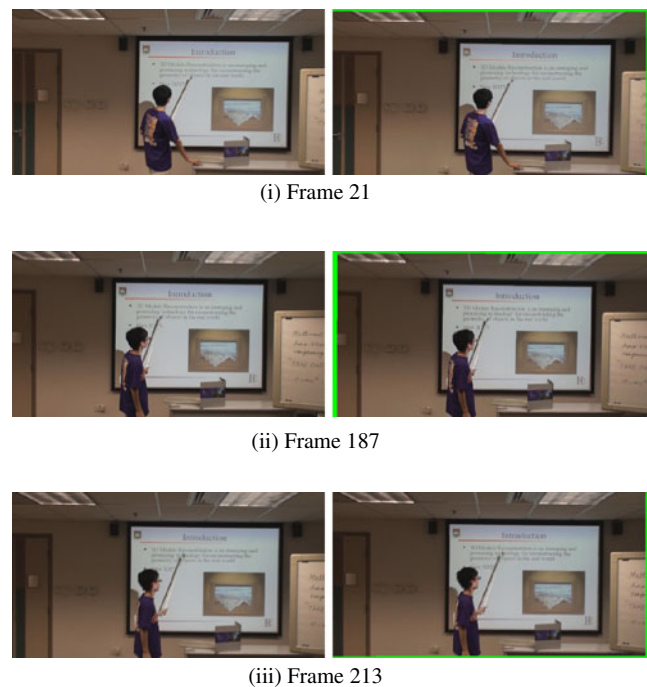


Figure 8 Video stabilization results: *left*—original image in frames (i) 21, (ii) 187, and (iii) 213; *right*—stabilized image in frames (i) 21, (ii) 187, and (iii) 213. The extended boundary pixels are marked in *green color*.

up the filtering operation, it is carried out using the graphic processing unit (GPU) in a NVIDIA GTX260+ graphic card. Since completely automatic segmentation is still difficult to achieve in real-time, view synthesis is not performed online. Instead, the videos, after compensating for slight differences in relative locations and rotation, are displayed directly in the multiview display. Moreover, since the linear array is moved in such a way that it is always facing the speaker, the centered microphone located on the linear array is used to pick up the speaker’s signal. Future work will focus on using directive microphone or arrays to suppress possible undesirable interference from other directions.

For the captured plenoptic videos, the multiple videos can be compressed offline using the object-based coder we have proposed in [32, 33]. It is based on the MPEG-4 coder and the picture frame structure is shown in Fig. 9. It employs prediction in both temporal and spatial directions. For simplicity, only three video object (VO) streams are shown. In each VO stream, we have a view of the IBR object, which we refer to as the video object plane (VOP). There are two types of VO streams associated with each dynamic IBR object: main video object stream and secondary video object stream. The main VO stream is encoded similar to the MPEG-4 algorithm, which can be decoded without reference to other VO streams. For better performance, bi-directional prediction is also employed for the B-VOPs. To provide random access to individual VOP, we adopt the Group of VOPs (GOVOP) structure of MPEG-4 in the main VO stream. A GOVOP contains an I-VOP and possibly P-VOPs and/or B-VOPs between this I-VOP and the subsequent I-VOPs. I-VOPs are coded using intra-frame coding to provide random access point without reference to any other VOPs, while P-VOPs are coded by

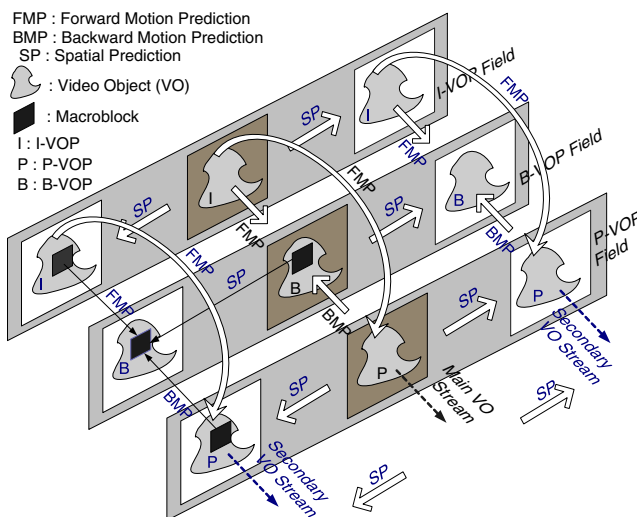


Figure 9 Picture frame structure and basic coding method for the texture coding of an IBR object in a PV.

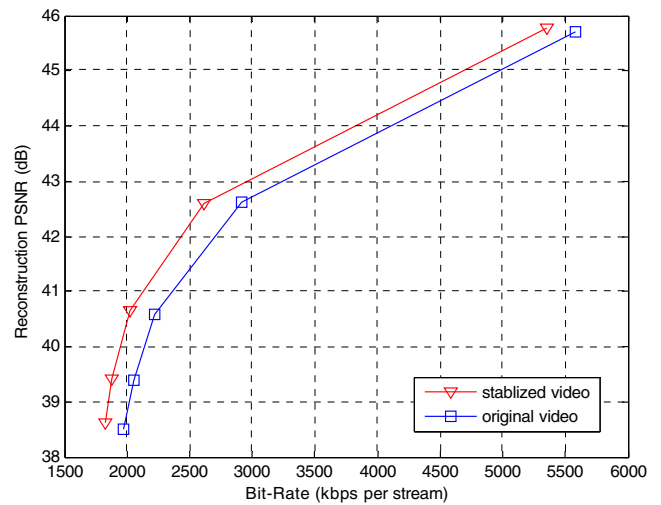


Figure 10 Coding performance of the original and stabilized multiple videos.

motion-predictive coding using previous I- or P-VOPs as references. B-VOPs are coded by a similar method except that forward and backward motion compensations are performed by using nearby I- or P-VOPs as references, which are indicated by the block arrow in Fig. 9. The VOPs captured at the same time instant as the I-VOP in a main stream constitute an I-VOP field. Similarly, we define the P- and B-VOP fields, which contain respectively the P- and B-VOPs of the main VO stream. A VOP from the secondary stream in an I-VOP field is encoded using disparity compensation prediction (DCP) from the reference I-VOP in the I-VOP field. Similarly, apart from using temporal prediction in the same stream, the P/B-VOPs in a secondary stream also employ spatial prediction from their adjacent P/B-VOPs in the main stream for better performance. The concept of GOVOP in the main stream can be extended to the VOP fields covering all the streams, which will be called a group of VOP fields (GOVOPF), to provide random access points in a PV. By extending the concept of GOVOPF, group of frame fields can be collected to form group of frame fields (GOFF). Interested readers are referred to [32, 33] for more details.

Experimental Results The performance of the proposed system is now evaluated. The multiview video captured in Section 3 above is down-sampled to a resolution of 4-CIF to evaluate the compression performance of the proposed movable system. The frame-based coding mode is employed because it is more suitable for video conferencing applications. To explore the spatial redundancy in images from adjacent views, three videos are encoded in a group as showed in Fig. 9 and only P-pictures are employed. The reconstructed peak signal-to-noise ratios (PSNR) of the original and stabilized videos versus the averaged bit rate per stream are plotted in Fig. 10. It can be

seen that due to reduced motion of the stabilized videos, the coding performance is slightly better than the original ones.

For the real-time transmission experiment, we use a lower resolution of (320×240) and two views to demonstrate real-time streaming of the stereo videos over a wireless LAN. The videos are compressed using two ThinkSmart IVS-MV02 Intelligent Video surveillance systems at 400 kbps each. The compressed videos are decoded using a PC and are displayed in the multiview TV for real-time multiview streaming and audio-visual conferencing. For simplicity, the audios are not compressed.

6 Conclusion

The design and construction of a movable image-based rendering system based on a class of dynamic representations called plenoptic videos, and its associated video processing algorithms have been presented. An example application to multiview conferencing is also presented. The system consists of a linear array of 8 video cameras mounted on an electrically controllable wheel chair with its motion being controllable manually or remotely through wireless LAN by means of additional hardware circuitry. A real-time object tracking algorithm is implemented and is utilized to adjust continuously the azimuth or rotation angle of the movable IBR system in order to track moving object in a large environment. A new video stabilization technique based on the estimation of the vibration velocities and adaptive notch filtering is developed to overcome the problem of imperfect tracking and mechanical vibration of the system during object tracking. The usefulness of the system is demonstrated by means of a multiview audio-visual conferencing application using a multiview TV display. The system developed provides useful experience for the design and construction of movable IBR systems with improved viewing freedom and ability to cope with moving object in a large environment.

Acknowledgment This project is supported by a GRF grant from the Hong Kong Research Grant Council and a tier-3 grant from the Hong Kong Innovative Technology Fund. The technical support of Mr. James L. C. Koo is gratefully acknowledged.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Chen, S. E. (1995). QuickTime VR—an image-based approach to virtual environment navigation. In *Proc. of the annual conference on Computer Graphics (SIGGRAPH'95)* (pp. 29–38), Aug.
- Debevec, P. E., Taylor, C. J., & Malik, J. (1996). Modeling and rendering architecture from photographs : a hybrid geometry—and image-based approach. In *Proc. of the annual conference on Computer Graphics (SIGGRAPH'96)* (pp. 11–20), Aug.
- Gortler, S. J., Grzeszczuk, R., Szeliski, R., & Cohen, M. F. (1996). The lumigraph. In *Proc. of the annual conference on Computer Graphics (SIGGRAPH'96)* (pp. 43–54), Aug.
- Levoy, M., & Hanrahan, P. (1996). Light field rendering. In *Proc. of the annual conference on Computer Graphics (SIGGRAPH'96)* (pp. 31–42), Aug.
- McMillan, L., & Bishop, G. (1995). Plenoptic modeling: An image-based rendering system. In *Proc. of the annual conference on Computer Graphics (SIGGRAPH'95)* (pp. 39–46), Aug.
- Peleg, S., & Herman, J. (1997). Panoramic mosaics by manifold projection. In *Proc. of IEEE computer society conference on Computer Vision and Pattern Recognition (CVPR'97)* (pp. 338–343), June.
- Szeliski, R., & Shum, H. Y. (1997). Creating full view panoramic image mosaics and environment maps. In *Proc. of the annual conference on Computer Graphics (SIGGRAPH'97)* (pp. 251–258), Aug.
- Shade, J., Gortler, S., He, L.-W., & Szeliski, R. (1998). Layered depth images. In *Computer Graphics (SIGGRAPH'98) Proceedings* (pp. 231–242), Orlando: ACM SIGGRAPH, July.
- Chen, S. E., & Williams, L. (1993). View interpolation for image synthesis. In *Proc. of the annual conference on Computer Graphics (SIGGRAPH'93)* (pp. 279–288), Aug.
- Shum, H. Y., & He, L. W. (1999). Rendering with concentric mosaics. In *Proc. of the annual conference on Computer Graphics (SIGGRAPH'99)* (pp. 299–306), Aug.
- Zhou, K., Hu, Y., Lin, S., Guo, B., & Shum, H. Y. Precomputed shadow fields for dynamic scenes. In *SIGGRAPH'05*.
- Fujii, T., Kimoto, T., & Tanimoto, M. (1996). Ray space coding for 3D visual communication. In *Proc. Picture Coding Symp. '96* (pp. 447–451), Mar.
- Chai, J. X., Tong, X., Chan, S. C., & Shum, H. Y. (2000). Plenoptic sampling. In *Proc. of the annual conference on Computer Graphics (SIGGRAPH'00)* (pp. 307–318), July.
- Shum, H. Y., Sun, J., Yamazaki, S., Lin, Y., & Tang, C. K. (2004). Pop-up light field: an interactive image-based modeling and rendering system. *ACM Transactions on Graphics*, 23(2), 143–162.
- Shum, H. Y., Ng, K. T., & Chan, S. C. (2005). A virtual reality system using the concentric mosaic: construction, rendering and data compression. *IEEE Transactions on Multimedia*, 7(1), 85–95.
- Ng, K. T., Chan, S. C., & Shum, H. Y. (2005). The data compression and transmission aspects of panoramic videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(1), 82–95.
- Adelson, E. H., & Bergen, J. (1991). The plenoptic function and the elements of early vision. In *Computational models of visual processing* (pp. 3–20). Cambridge: MIT.
- Shum, H. Y., Chan, S. C., & Kang, S. B. (2007). Image-based rendering. Springer
- (2007). *IEEE Signal Processing Magazine: Special Issue on MVI and 3DTV*, 24(6), Nov.
- Chan, S. C., Ng, K. T., Gan, Z. F., Chan, K. L., & Shum, H. Y. (2003). The data compression of simplified dynamic light fields. In *Proc. of IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing (ICASSP'03)* (Vol. 3, pp. 653–656), April.
- Chan, S. C., Ng, K. T., Gan, Z. F., Chan, K. L., & Shum, H. Y. (2004). The plenoptic videos: capturing, rendering and compression. In *Proc. of IEEE Int'l Symposium on Circuits and Systems (ISCAS'04)* (Vol. 3, pp. 905–908), Vancouver, May 23–26.

22. Chan, S. C., Ng, K. T., Gan, Z. F., Chan, K. L., & Shum, H. Y. (2005). The plenoptic videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(12), 1650–2659.
23. Gan, Z. F., Chan, S. C., Ng, K. T., & Shum, H. Y. (2005). An object-based approach to plenoptic videos. In *Proc. of IEEE Int'l Symposium on Circuits and Systems (ISCAS'05)* (pp. 3435–3438), May.
24. Chan, S. C., Gan, Z. F., Ng, K. T., & Shum, H. Y. (2009). An object-based approach to image/video-based synthesis and processing for 3-D and multiview televisions. *IEEE Transactions on Circuits and Systems for Video Technology*, 19(6), 821–831.
25. Chan, S. C., Shum, H. Y., & Ng, K. T. (2007). Image-based rendering and synthesis: technological advances and challenges. *IEEE Signal Processing Magazine: Special Issue on MVI and 3DTV*, 24(6), 22–33.
26. Konrad, J., & Halle, M. (2007). 3-D Displays and signal processing. *IEEE Signal Processing Magazine*, 24(6), 97–111.
27. Li, Y., Sun, J., Tang, C. K., & Shum, H. Y. (2004). Lazy snapping. In *Proc. of the annual conference on Computer Graphics (SIGGRAPH'04)* (pp.303–308)
28. Sapiro, G. (2001). *Geometric partial differential equations and image analysis*. Cambridge: Cambridge University Press.
29. Sethian, J. A. (1996). *Level set methods: Evolving interfaces in geometry, fluid mechanics, computer vision and materials sciences*. Cambridge: Cambridge University Press.
30. Osher, S., & Paragios, N. (2003). *Geometric level set methods in imaging, vision, and graphics*. Springer Verlag.
31. Chan, T. F., & Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing*, 10(2), 266–277.
32. Wu, Q., Ng, K. T., Chan, S. C., & Shum, H. Y. (2005). On object-based compression for a class of dynamic image-based representations. In *Proc. ICIP'2005* (pp. 405–408), Sept.
33. Ng, K. T., Wu, Q., Chan, S. C., & Shum, H. Y. Object-based coding for plenoptic videos, to appear. In *IEEE Trans. Circuits and Systems for Video Technology*.
34. Gan, Z. F., Chan, S. C., Ng, K. T., & Shum, H. Y. (2005). Object tracking for a class of dynamic image-based representations. In *Proc. Visual Communications and Image Processing*. (pp. 1267–1274), July.
35. Gan, Z. F., Chan, S. C., & Shum, H. Y. (2005). Object tracking and matting for a class of dynamic image-based representations. In *Proc. IEEE Advanced Video and Signal-Based Surveillance* (pp. 81–86), Sep.
36. Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330–1334.
37. Comaniciu, D., Ramesh, V., & Meer, P. (2003). Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, 564–575.
38. Bar-Shalom, Y., Li, X. R., & Kirubarajan, T. (2001). *Estimation with applications to tracking and navigation*. Wiley.
39. Matsushita, Y., Ofek, E., Tang, X., & Shum, H. Y. (2005). Full-frame video stabilization. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 1, 50–57.
40. Liu, F., Gleicher, M., Jin, H., & Agarwala, A. (2009). Content-preserving warps for 3D video stabilization. *ACM SIGGRAPH*.
41. Bouguet, J. (2005). *Pyramidal implementation of the Lucas Kanade feature tracker description of the algorithm*. Microprocessor Research Labs, Intel Corporation
42. Zhang, Z. G., Chan, S. C., & Tsui, K. M. (2008). A recursive frequency estimator using linear prediction and A Kalman filter-based iterative algorithm. *IEEE Transactions on Circuits Systems II*, 55(6), 576–580.



S. C. Chan received the B.S. (Eng.) and Ph.D. degrees from the University of Hong Kong, Pokfulam, Hong Kong, in 1986 and 1992, respectively. He joined the University of Hong Kong at 1994 and is now a Professor in the department of electrical and electronic engineering. He was a Visiting Researcher at Microsoft Corporation, Redmond, WA and Microsoft, Beijing, China, in 1998 and 1999, respectively. His research interests include fast transform algorithms, filter design and realization, multirate and array signal processing, biomedical signal processing and image-based rendering. Dr. Chan is an associate editor of *Journal of Signal Processing Systems*, and a Member of the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society. He was an associate editor of the *IEEE Transactions on Circuits and Systems—I: Regular Papers* from 2008-2009 and was the special session and tutorial chairs of IEEE ICASSP2003 AND ICIP2010, respectively.



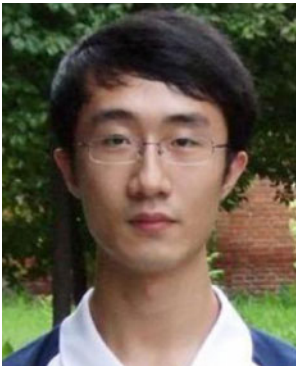
Z. Y. Zhu received the B.Eng. degree from Huazhong University of Science and Technology in 2008. He is currently pursuing the degree of Ph.D. at the department of Electrical and Electronic Engineering, The University of Hong Kong. His main research interests are in digital image processing, multiple view geometry and 3D reconstruction.



King-To Ng received the B.Eng. degree in computer engineering from the City University of Hong Kong, Hong Kong, China, in 1994, and the M. Phil. and Ph.D. degrees in electrical and electronic engineering from the University of Hong Kong, in 1998 and 2003, respectively. In 2004, he was with Microsoft Research Asia, Beijing, China, as a Visiting Associate Researcher. From 2004 to 2009, he was a Postdoctoral Fellow in the Department of Electrical and Electronic Engineering, University of Hong Kong. His research interests include visual communication, image-based rendering, and video broadcast and transmission.



S. Zhang received his B.Sc(CE) degree from the Yanshan University in 2008. He received his M.Eng degree from The University of Hong Kong at 2010 and is now a research assistance in the department of electrical and electronic engineering, The University of Hong Kong. His research interests focus on video processing and computer vision.



C. Wang received the B.Eng. degree from Zhejiang University of Technology in 2007, and the M.Eng. degree from University of Science and Technology of China in 2010. He is currently pursuing the degree of Ph.D at the department of Electrical and Electronic Engineering, The University of Hong Kong. His main research interests are in digital signal processing, video de-noising and super-resolution.



Z. G. Zhang received the B.Sc. and M.Eng. degrees in Electrical and Electronic Engineering from Tianjin University and the University of Science and Technology of China in 2000 and 2003, respectively. He obtained the Ph.D. degree in 2008 from the Department of Electrical and Electronic Engineering at the University of Hong Kong, where he is currently a Postdoctoral Fellow. His research interests are in general area of statistical signal processing and digital signal processing, and in particular adaptive filtering, nonparametric regression, time frequency analysis, and biomedical signal processing.



Zhongfu Ye was born in 1959. He has received the B.E., M.Sc from HeFei University of Technology in 1982, 1986, respectively, and the Ph.D degree in electronics and information engineering from University of Science and Technology of China in 1995. His current research interests include statistical and array signal processing, image processing.



Heung-Yeung Shum received the Ph.D. degree in robotics from the School of Computer Science, Carnegie Mellon University, Pittsburgh, PA. Currently, he is the Corporate Vice President responsible for search product development with Microsoft Corporation, Redmond, WA. Previously, he oversaw research activities at Microsoft Research Asia, Beijing, China, as well as the lab's collaborations with universities in the Asia Pacific region. He was responsible for the Internet Services Research Center, an applied research organization dedicated to long-term and short term technology investments in search and advertising at Microsoft, and he was a Researcher with Microsoft Research in 1996, based in Redmond. He moved to Beijing, China as one of the founding members of Microsoft Research, China (later renamed Microsoft Research Asia). There he began a nine-year tenure as a Research Manager, subsequently moving on to become Assistant Managing Director, Managing Director of Microsoft Research Asia, Distinguished Engineer, and Corporate Vice President. He has published more than 100 papers about computer vision, computer graphics, pattern recognition, statistical learning, and robotics. He holds more than 50 U.S. patents. Dr. Shum is a Fellow of the Association for Computing Machinery.