2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing

Preliminary Survey of Multiview Synthesis Technology

Ellen Sharma Shijagurumayum and Feng-Cheng Chang
Dept. of Innovative Information and Technology
Tamkang University
Taipei, Taiwan
ellen0sija@hotmail.com, 135170@mail.tku.edu.tw

***Abstract*—With the maturity of digital camera technology, it is feasible to form an array of cameras. The major usage of a camera array is to acquire different views of a scene in one shot. The captured data can be used to analyze the depths of the objects. Once we have the 3D model, we can synthesize virtual views, relight the scene, etc. The potential applications are virtual reality and augmented reality. In order to investigate the multiview technology, we studied the fundamental concepts, including the single lens camera, the eccentric lens camera, the plenoptic camera, and the multiview camera. We also discussed a few application examples for understanding the practical usages. The study showed that the initial depth estimation technology becomes important nowadays in providing photorealistic natural feel to people. The application areas were also extended to entertainment, and also to some crucial tasks such as medical operations and combat missions.**

Keywords-Plenoptic; Multiview;Synthesis; Representation.

I. INTRODUCTION

Stereo imaging has been studied for decades. From the analytic viewpoint, it was focused on estimating the depths of pixels; from the synthetic viewpoint, it was focused on how to present pixels to make humans feel the depths. Display technology has been improved a lot recently. More and more 3D displays and applications are emerging.

To work with 3D equipments, the data representation should be considered. In other words, the images captured from different viewpoints of the same scene should be organized. To design computer vision algorithms that make the best use of the captured data, it is important to thoroughly study the imaging process of a single camera and the geometric relationships involved among a collection of cameras. Two-dimensional images are the primary means by which humans represent the three-dimensional world surrounding them. The introduction of photography resulted in unprecedented levels of realism in these; an often-stated goal in photography has been to increase the amount of visual information that can be acquired. Virtual reality (VR) has become one of the hottest and commercial topics [1]. Two commonly used approaches for building a VR world are model-based approach [2][3] (e.g., AutoCAD, 3D Studio, etc.) and image-based approach [4][5][6] (e.g., Quick-Time VR, Surround Video, Real VR, and IPIX, etc.).

In this paper, we will study the fundamental concepts for representing multi-camera data and the related applications. In Sec. II, fundamentals of plenoptic camera are described. In Sec. III, we discuss the popular multiview concepts and the reason why it is important. In Sec. IV, a few applications that are based on multiview systems are briefly described. At the end, we conclude this study in Sec. V.

II. PLENOPTIC CAMERA

A plenoptic camera (ideally) records information from all possible viewpoints within the lens aperture. It was developed for representing complete views of the received imaging information, possibly including the colors and the intensities. It was invented by Edward Adelson and John Wang of MIT. The original paper was published in 1992. In Feb. 2005, the plenoptic camera design was enhanced by Ren Ng et al of Stanford University. The related theories are also called the light field. The plenoptic camera is based on the photography methods pioneered by Lippman and Ives. A typical plenoptic camera is constructed with an array of micro lens. It focuses those micro lenses at infinity in order to sample the 4D radiance directly at the micro lenses. The consequence is that each micro lens image is completely defocused with respect to the image formed by the main camera lens. As a result, only a single pixel in the final image can be rendered from it, resulting in very low resolution.

A. Single Lens Stereo Design

The design of the plenoptic camera follows the fundamental lens theory, specifically the thin lens formula. We start the study by exploring the optical information available across the aperture of a camera lens. We will see that a single lens camera captures a mixture of images in one shot.

Fig. 1 shows the cases in which the camera is bringing a point object into focus on its sensor plane. Ideally, the in-focus image is a point, and the images corresponding to near and far objects are blurred. Fig. 2 shows the imaging by eccentric lens. When the aperture is displaced to the right, the image of a near object is also displaced to the right and the image of a far object is displaced to the left. In other

978-0-7695-4222-5/10 $26.00 © 2010 IEEE
DOI 10.1109/IIHMSP.2010.48

words, a near object "follows" the aperture displacement, while a far object "opposes". This characteristic can be used in a depth estimator [7].

B. Depth Estimation with Single Lens

The aforementioned characteristics of an eccentric lens can be used to measure the depth of a given point object. By incorporating the thin lens theories, the simple geometric analysis of the displacement is used to derive the depth in a single lens-stereo system. Assume that we have an eccentric lens whose focal length is **F** and the displacement of the aperture is **v**. Given the other parameters as shown in Fig. 3:

- **f** -- distance between lens and sensor plane
- **D** -- distance to a plane conjugate to sensor plane
- ***d*** -- distance of a particular point object
- **g** -- distance to conjugate focus of object
- **h** -- displacement of object's image in sensor plane

The known parameters are **F**, **f**, **D**, and **v**; the measured parameter is **h**; and the synthetic parameter is **g**. We would like to determine the object distance ***d*** by the other known parameters.

By the use of similar triangles $\frac{h}{v}=\frac{g-f}{g}$,

we have $\frac{1}{g}=\frac{1}{f}\left(1-\frac{h}{v}\right)$.

Combining with the thin lens equation $\frac{1}{F}=\frac{1}{g}+\frac{1}{d}$,

it leads to $\frac{1}{d}=\frac{1}{F}-\frac{1}{f}\left(1-\frac{h}{v}\right)$

or $\frac{1}{d}=\frac{h}{v}\left(\frac{1}{F}-\frac{1}{D}\right)+\frac{1}{D}$.

The left-hand side is the reciprocal of the object distance **d**; the right-hand side consists of the system parameters **F**, **D**, and **v**, along with the measured displacement **h** [8]. The single-lens-stereo approach works theoretically. It has a practical disadvantage that numerous snapshots are required for complete depth estimation (i.e., for all point objects). We prefer to acquire all of the image information with a single snapshot. The plenoptic camera is thus invented to solve this problem.

C. Plenoptic Camera Design

In an ordinary camera, the lens captures a lot of viewpoints at once. Each view point forms an image on the sensor plane. Therefore, the sensed image is actually a mixture (average) of the captured images. In other words, each sensor element receives more than one input ray from the scene.

A plenoptic camera is designed in a different way. There is an array of micro lens between the main lens and the sensor plane. If the micro lens array is properly placed, it

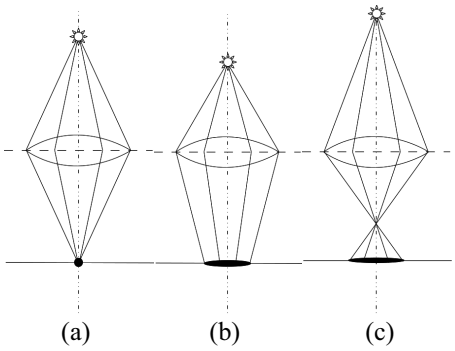

Figure 1. Single lens image: (a) in-focus (b) near (c) far

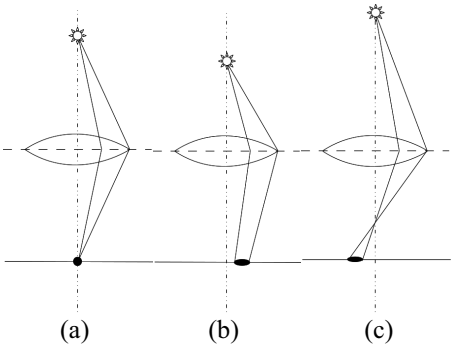

Figure 2. Eccentric lens image: (a) in-focus (b) near (c) far

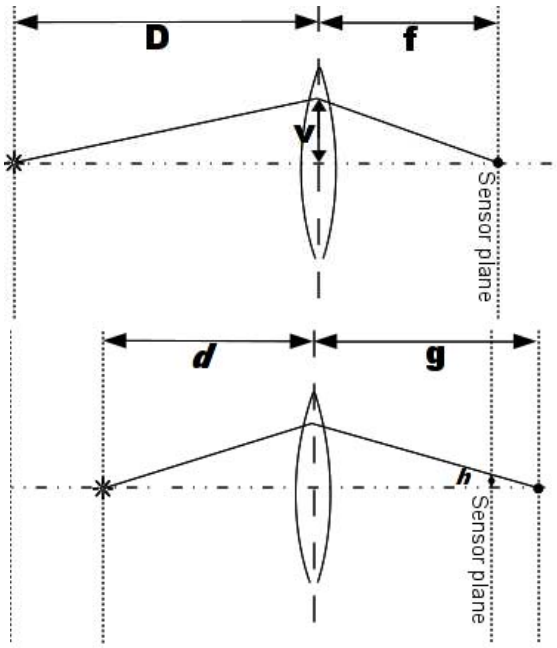

Figure 3. Depth estimation with eccentric lens

effectively divides the sensor plane into macro-pixels. Each macro-pixel corresponds to a small input aperture at the main lens. In other words, a macro-pixel collects a limited range of rays directed from the corresponding aperture to the micro lens. Therefore, a plenoptic camera can capture an image with the directional lighting distribution at each sensor location. With the direction information, we have additional

dimension to manipulate the captured pixels. In a plenoptic image, a pixel is not the intensity of the mixed arriving rays. It represents the set of arriving rays (the direction and the intensity). Therefore, rearranging the pixels is equivalent to rearranging the rays.

To sum up, the Initial motivation of plenoptic cameras was depth estimation of a scene. One of its applications is called synthetic refocusing [9]. It is based on rearranging the rays and synthesizing the virtual views. More specifically, synthetic refocusing means to capture an image once with the plenoptic camera and generate numerous images focused at different depths. The major usage of plenoptic cameras is to capture a natural 3D content for autostereoscopic displays. With the advances in hardware technology, a plenoptic camera will be a practical apparatus in that it could capture an image and calculate the corresponding depth map with on board chip.

III. MULTIVIEW CAMERA

Plenoptic cameras are simple and compact for capturing 3D information. However, the resolution is limited by both the density of sensor elements and the number of micro lens. If the number of micro lens is not large enough, the efficiency of capturing directional information is degraded; if the resolution of each macro-pixel is not sufficient, the accuracy of depth estimation is also degraded.

Because single-lens digital camera system is mature, the physical size of a digital camera becomes smaller and smaller. In addition, the resolution of the captured image is higher and higher, which means high-density sensors are available. This leads to an alternative approach to obtain the directional information: a multi-camera system (or a multiview camera system). In fact, this approach is an extension to the classic stereo imaging problem. The human brain is quite adept at deriving the two images captured by each eye's retina to the depth information. It is believed that humans use many visual cues to define the sense of depth and make spatial connections between objects in an environment.

The multiview approach has attracted a great deal of interest for quite a while, partly because of the stronger demands of virtual reality and/or augmented reality. It is not difficult to arrange a large number of cameras with high-density sensors. While the problem can be viewed as a generalization of stereo imaging, it is considerably harder. The major reason of the difficulty is the reasoning about visibility. In stereo matching, most scene elements are visible from both cameras because the cameras are typically close enough. However, the occlusion problem becomes severe due to the physical span of the camera array. That is, the visibility constraints are considerably important in a multiview camera system. One of the multiview camera systems is based on the rayspace representation [10]. The application is called Freeview TV and is part of the MPEG 3DAV specifications.

IV. MULTIVIEW APPLICATIONS

A camera array is useful for capturing a variety of 3D information, depending on how the cameras are arranged and configured. In the previous section, we mentioned that it is important to reasoning the visibility of an object. However, if the number of cameras is large enough, some image-based rendering researches [11][12] showed that one can accurately synthesize virtual views without much knowledge of the 3D geometry (specifically the depths) of the captured scene. In the following sections, we will describe a few multiview applications based on the image-based approach:

- Aerial Vehicles
- Facial Recognition
- Surgical Applications

A. Aerial Vehicles

Unmanned aerial vehicles (UAV) have been used in a reconnaissance and intelligence-gathering role since the 1950s. UAVs are remotely piloted or self-piloted aircraft that can carry cameras, sensors, communications equipment or other payloads. They are very useful for acquiring surveillance data in dangerous environments or circumstances, including the combat missions. For remote imaging applications, a traditional UAV takes pictures at a distant position. To reconstruct a specific 3D scene, it has to take a number of pictures at different time instants. If it equips with an array of cameras, it is possible to acquire all the necessary pictures at one time instant.

To providing stereo capabilities on a UAV, it requires that the size and weight of the cameras are minimized to fit within restrictions. The other important factor is how to arrange the positions of the cameras. A typical arrangement is to disperse the cameras across the wingspan. An operator would use one of the views for navigation or use all the views to synthesize the stereo scene.

B. Facial Recognition

The second example is for a mid-level scene. It means that the distance from the camera array to the objects is only a few meters and the separation between two cameras is less than one 13 meter. A typical application for a mid-level scene is the 3D facial recognition for security and possibly entertainment.

In general, face recognition systems are highly sensitive to the environmental illumination of the captured images. In particular, changes in lighting conditions can increase both false rejection rates (FRR) and false acceptance rates (FAR). In addition to the illumination, the orientation of the taken picture also affects the recognition precision. Most facial recognition algorithms are designed to match the features extracted from the frontal view. It is obvious that a slight change in the shot orientation would produce an "abnormal" frontal face picture. This implies that a pre-processing of the picture is required to compensate the orientation distortion.

A camera array could be used to solve the problem. On the one hand, multiview images can be used to analyze the illumination of the environment, especially for the lighting conditions around the face. This eliminates the unstable analysis of only one facial picture. On the other hand, the images taken from the camera array can be used to reconstruct the 3D model of the face surface. Once we have the 3D information of the face, we can synthesize the frontal

view of the face. The conventional 2D face recognition can be applied and achieve higher precision.

The estimated stereo model of a face also provides opportunities to develop 3D face recognition algorithms. In theory, we can extract more features from a 3D model, and these features leads to a better matching accuracy. The 3D facial surface models can also be used to predict facial changes, facial orientation, and possibly add aging effects. One of the traditional 3D face model construction methods is by means of laser scanning. It comes with the lighting issue which the multiview approach does not have.

C. Surgical Applications

The third application focuses on hyperstereo imaging for micro close-ups in recording surgical procedures. This application is important for both teaching/training and for telemedicine.

Currently there are several 3D visualization methods in the medical field. One category of the 3D imaging devices is in radiology ranging from ultrasound, X-ray, nuclear imaging, computed tomography (CT) scanning, and magnetic resonance imaging (MRI). The other category of 3D visualization is to analyze either 3D or 2D data and create the corresponding 3D models. The disadvantages of the current methods are (1) they are not real-time and thus not suitable for surgical training; (2) the real-time issue also prevents them to be the implementation for recording purpose. If multiview solution can be realized as medical devices, it is easier to acquire/render real-time 3D data. The consequence would be a more convenient telemedicine system.

D. Discussions

Using multiview camera arrays to obtain the images and then constructing the 3D environments is expected to be one of the practical implementation of virtual reality and augmented reality. Instead of making the environments from models, one can use multiview images to synthesize the desired view. The additional advantage is that the estimated model or the synthesized view can be naturally photorealistic rendered. The ability to deliver the feel of real places thus expands the applications of entertainment, virtual tourism, telemedicine, telecollaboration, and teleoperation. Multiview images can also be used to re-illuminate scenes, synthesize images from virtual viewpoints, and derive geometric 3D models.

V. CONCLUSIONS

In this paper, we studied the concepts of plenoptic cameras. A single lens camera treats rays from different directions equally and the sensor effectively captures a mixture of the images from different directions. An eccentric lens camera can distinguish the depth of an object by the displacement. To obtain multiple pictures in one shot, two approaches can be used. The first is inserting a micro-lens array between the main lens and the sensor. This is called a plenoptic camera. On the sensor plane, the sensor elements are grouped as macro-pixels to capture directional rays. The resolution of this design is limited by the number of micro-lens and the density of the sensor elements. The second approach is called multiview camera system. It arranges a number of single-lens cameras to form the array. Some image-based rendering researches showed that it is not difficult to estimate the 3D scene as long as the number of cameras is sufficient.

Multiview camera system is a very promising approach to optics technology. It can be used to provide high-resolution 3D capturing and rendering. The applications include 3D remote sensing, precise facial recognition, real-time surgical recording, etc.

ACKNOWLEDGMENT

This work was partially supported by the NSC, Taiwan, under Grants NSC 98-2218-E-032-008.

REFERENCES

[1] J.M. Moshell, “Three view of virtual reality: Virtual environments in the U.S. military,” IEEE Computer, vol. 26, no. 2, pp. 81-82, Feb. 1993.

[2] C. H. Sequin and R. W. Bukowski, “Interactive virtual building environment,” in *Proc. of Pacific Graphics '95*, 1995, pp. 159-179.

[3] Y.-W. Lei, The SpaceWalker Walkthrough system for unrestricted three-dimensional polygon environments, PhD thesis, National Taiwan University, Taipei, Taiwan, ROC, 1996.

[4] S. E. Chen, “QuickTime VR an image-based approach to virtual environment navigation,” in *Proc. SIGGRAPH Computer Graphics, Annual Confernece Series*, 1995, pp. 29-38.

[5] W.-K. Tsao et al., “Photo VR: A system of rendering high quality images for virtual environments using sphere-like polyhedral environment maps,” in *Proc. of Second Workshop on Real-Time and Media Systems, RAMS'96*, Taipei, Taiwan, ROC, July 1996, pp. 397-403.

[6] R. Szeliski and H.-Y. Shum, “Creating full view panoramic image mosaics and environment maps,” in Proc. SIGGRAPH '97, 1997, pp. 251-258.

[7] T. Adelson and John Y. A. Wang, “Single lens stereo with a plenoptic camera,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pp. 99-106, Feb. 1992.

[8] R. NG et al., “Light field photography with a hand-held plenoptic Camera,” Tech. Rep. CSTR 2005-02, Stanford Computer Science, 2005.

[9] Scalable Multi-view Stereo Camera Array for Real World Real-Time Image Capture and Three-Dimensional Displays, the Massachusetts Institute of Technology June 2004.

[10] P. Nabangchang, T. Fuji, and M. Tanimoto, “Experimental System of Free Viewpoint Television,” in Proc. of SPIE, vol. 5006, May 2003, pp. 554-563.

[11] C. Zhang and T. Chen. “A survey on image-based rendering - representation, sampling and compression,” EURASIP Signal Processing: Image Communication, vol. 19, no. 1, pp. 1–28, 2004.

[12] M.W. Halle, “Multiple viewpoint rendering for 3-Dimensional displays, ” Ph.D. Thesis, Program in Media Arts and Sciences, Massachusetts Institute of Technology, 1997.

[13] K.I Chang, K.W. Bowyer, P.J. Flynn, “Face recognition using 2D and 3D facial data,” Workshop in Multimodal User Authentication, 2003, pp. 25-32.