High-resolution image generation from video image sequences by light field

Kobayashi, Kenkichi, Saito, Hideo

Kenkichi Kobayashi, Hideo Saito, "High-resolution image generation from video image sequences by light field," Proc. SPIE 4309, Videometrics and Optical Methods for 3D Shape Measurement, (22 December 2000); doi: 10.1117/12.410890

Event: Photonics West 2001 - Electronic Imaging, 2001, San Jose, CA, United States

High-resolution image generation from video image sequence by light field

Kenkichi Kobayashi and Hideo Saito

Department of Information and Computer Science, Keio University
3-14-1 Hiyoshi Kouhoku-ku Yokohama 223-8522, Japan

ABSTRACT

We propose a novel method to synthesize high-resolution images by constructing a light field from image sequence taken with a moving video camera. Our method integrates multiple frames from video camera that partly captures the object by constructing a light field, which is quite different from general mosaic methods. In case of light field constructed straightforwardly, blur and discontinuity are introduced into generated images by depth variation of the object. In our method, light field is optimized to remove these blur and discontinuity, so that clear images can be generated. The optimized light field is adapted to the depth variation of object surface, but the exact shape of the object is not necessary. Extremely large resolution images that are impractical in the real system can be virtually generated from the light field. Results of the experiment applied to book surface demonstrate the effectiveness of the proposed method.

Keywords: image-based rendering, light field, high-resolution image, image mosaicing, video image sequence

1. INTRODUCTION

Recent advance of computer and network technologies enables us to be in the world of virtual or augmented reality. Traditional way to generate images for such reality begins with modeling geometry of the object in the environment. Then images of a virtual camera from arbitrary viewpoint can be rendered from the model using computer graphics techniques. Methods with geometry models have great advantages in handling objects in virtual environment. However, despite a great deal of work, small details of complex geometry are still unavailable. Furthermore, realistic images with subtle lighting effects are difficult to render with methods using geometry models. To synthesize images with photographic effect, new approach, which combines and resamples pre-acquired images to generate different views, has emerged. This kind of method[2,3,5] is called image-based rendering and interest in it has been increasing rapidly.

Digital library, which used to be no more than a conception is also starting to work practically. One of key technologies to support digital library is digitization of books and paper works. In such digitization, digital archiving by high-resolution image is desired especially for old and rare books, since shapes of characters, conditions of pages, etc. are considerable information. As a method for digitizing old and rare books, contact types such as digital scanners, which force books to be pressed against its contact-window, are inadequate for the fear of damaging books. Thus, non-contact types such as digital cameras are said to be more suitable for rare historical books. However, the photograph range and the resolution of a digital camera can't be changed separately, and cameras with high resolution are still not general. Therefore, pages have to be partly photographed if the desired resolution for entire page isn't available in one shot. Partial images are then integrated together to reconstruct whole original page. Large numbers of methods for integrating partly photographed images (mosaic) are proposed.[7,8] These mosaic methods basically integrate images together by making use of overlapped region in neighboring images. Regions captured in the images do not form exact rectangles because of the distortion caused by 3D shape of the object. The distortion makes the overlapped region to be observed differently in each image, and it becomes difficult to integrate images in mosaic methods. To remove this image distortion and integrate images without disagreement, adjustments that take account of 3D shape of the object is essential. Methods for these adjustments, sometimes requiring assumptions to the object shape, are proposed[4] , but they are still with much difficulty.

Further author information: (Send correspondence to K.K.)
K.K.: E-mail: ken@ozawa.ics.keio.ac.jp
H.S.: E-mail: saito@ozawa.ics.keio.ac.jp

Videometrics and Optical Methods for 3D Shape Measurement, Sabry F. El-Hakim, Armin Gruen, Editors,
Proceedings of SPIE Vol. 4309 (2001) © 2001 SPIE · 0277-786X/01/$15.00

In this paper we propose a novel method to generate high-resolution images by integrating images from video sequences. Each frame is a high-resolution image of partial of the object. In integrating partial images, our method doesn't explicitly rely on 3D shape of the object. Instead, all the images are integrated into a single light field and high-resolution images of entire object is available from the light field. However, depth variation of the object introduces blur and discontinuity into generated images. To remove these blur and discontinuity, light field is optimized to adapt the depth of object's surface, but does not necessarily match the actual shape of the object.

2. REPRESENTATION

Light field is a representation of flow of light through a volume of space. Images from different viewpoints can be generated from light field, which is constructed from pre-acquired images. Representation of the flow of light requires 5 parameters, i.e. position (x, y, z) and direction (θ, ϕ). However, if we consider radiance to be constant along a line, it is well known that 5D representation of light field can be reduced to 4D. For uniformity of sampling and efficiency of calculation, we adopt the representation of 4D light field that parameterize lines by their intersections with two parallel planes[2,5] , as shown in Fig. 1. By convention, each plane has an orthogonal coordinate, (s, t) for the first plane and (u, v) for the second plane. In the direction vertical to the planes, we define z-axis with origin on st plane. Connecting a point on st plane to a point on uv plane defines a line identified by (s, t, u, v). Light is assumed to propagate straight, and radiance of a ray entering one plane and exiting another plane is represented by $L(s, t, u, v)$.

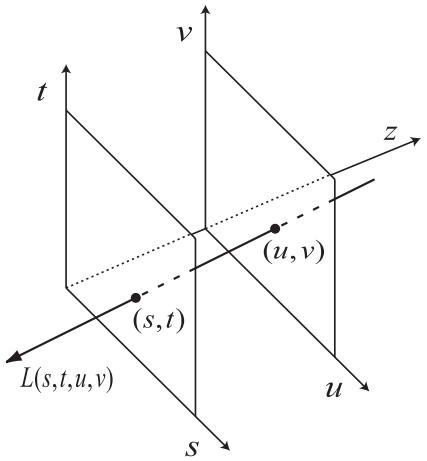


Figure 1. Representation of a ray.

Light field can be considered as a database of $L(s, t, u, v)$. Intensity of each pixel is regarded as radiance of a ray which pass the pixel and reach the center of camera lens, so given images are samples of ray to construct light field. Conversely, new image from arbitrary viewpoint can be generated by querying appropriate $L(s, t, u, v)$ from the database for each pixel. Concept of this representation can be well understood with Fig 2.

3. CONSTRUCTING A LIGHT FIELD

We construct light field from video image sequences, which are captured as followings. As shown in Fig. 3, we place st plane at the position of camera. Parallel to st plane, uv plane is placed near the object. Now, we can parameterize rays that enter uv plane and exit from st plane. Camera motion for capturing image sequences is like raster scanning an image. A video camera is moved along the s-axis in constant speed to sweep the object. We keep on making uniform sweep along scan lines with even interval. An image sequence is captured for each scan line and we extract frames from each image sequence by constant frame rate to get a 2D array of images. These are the images captured from st plane in even grid interval and we refer to these (s, t) points as grid points.

Each image is captured from grid point (s, t) respectively, and each pixel in images corresponds to a (u, v) coordinate respectively. So each pixel in each image is mapped to an individual (s, t, u, v), and the intensity of a pixel is treated as a radiance $L(s, t, u, v)$ of the ray. This correspondence of a pixel with (s, t, u, v) constructs

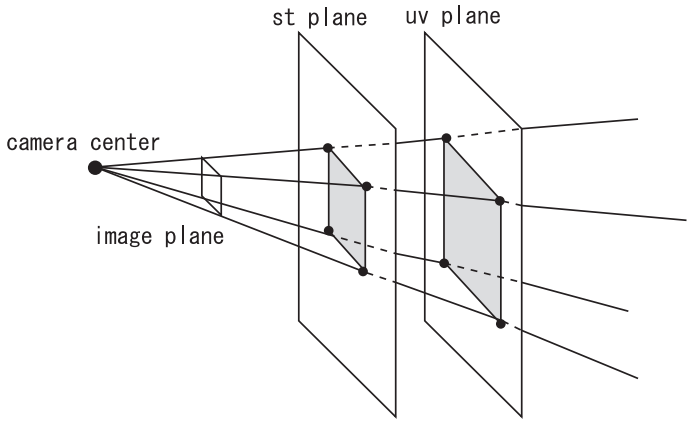


Figure 2. Relationship between an image and a light field.

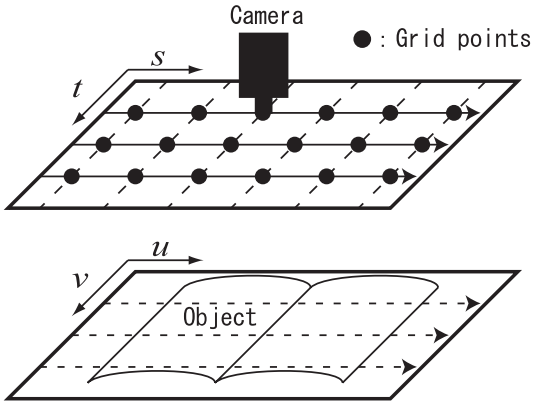


Figure 3. System arrangement.

a database of $L(s,t,u,v)$, a light field. Intrinsic parameter of camera is necessary for estimating correspondence of each pixel to a (u,v) coordinate. We fix the lens while capturing image sequence so that the intrinsic camera parameter remains constant. Therefore, camera calibration[6] takes place only once before the image acquisition stage. The correspondence of a pixel to (s,t,u,v) is estimated with an assumption that the input images are captured by pinhole camera. Lens distortion in captured images induces error in mapping, so all the input images are undistorted using intrinsic parameter estimated by calibration.

Constructing a light field integrates images that are captured partly. Fig. 4(a) represents rays for a single image which captures only a part of the object. Rays covering entire object are obtained from all the input images, as shown in (b). Finally in (c), an image of entire object can be rendered by querying desired rays from light field. Radiance of each ray derived from different input images. In other words, light field integrates input images.

4. OPTIMIZATION OF LIGHT FIELD

As mentioned previously, querying radiance of appropriate ray for each pixel from light field renders images of desired camera. The radiance of the exact ray can be queried from continuous light field. Unfortunately, light field in practice is a database that holds data only at the grid points, and we need to interpolate the radiance of desired ray from data of neighboring grid points. In our case, we identify 4 grid points in (s,t), which are nearest from the intersection of desired ray. A ray that is closest in (u,v) is queried from each 4 grid points, and the radiance of desired ray is interpolated from these 4 rays. In doing so, blur and discontinuity caused by the depth variation of the object are introduced to generated images. Fig. 5 shows this aspect of light field. In Fig. 5 each line from grid points represents

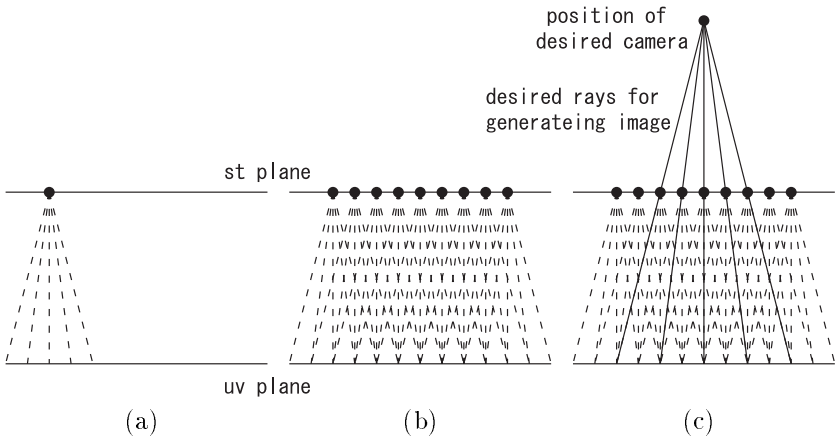


Figure 4. Images are integrated by constructing a light field.

a queried ray from database that best approximates the desired ray. In Fig. 5 (a), surface of the object lies on uv plane and the radiances of all the queried rays will agree, since they intersect the object surface at the same position. In cases where the object surface is away from uv plane like (b) and (c), queried rays intersect the object surface at different points and radiances don't agree. This disagreement makes generated images blurred and discontinued.

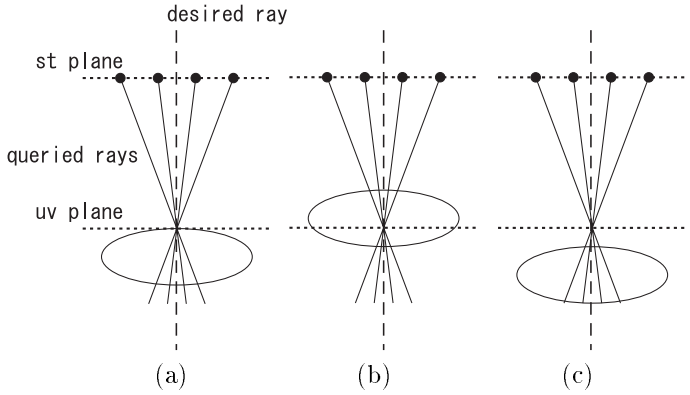


Figure 5. Disagreement on radiance of rays.

In images captured in the real world, regions out of focus are also blurred. Disagreement on the radiance may properly induce out-of-focus effect in images generated from light field. However, generated images may be too much blurred than expected, and blur is not desired at all in our method. Blur, which comes from this disagreement, should be avoided. Disagreement is due to discrete data structure of light field, and we can reduce it by sampling denser data from more images. At the same time, excessively increasing the number of input images is irrelevant, since it requires total data size to be enormous. To reduce disagreement as possible with reasonable sampling rate, optimization of light field is introduced.

4.1. Optimization Process

Basic policy of our optimization process is reparameterizing light field[1] , that is, to shift uv plane in the direction of z-axis to meet the object surface. Though data in light field is discrete, valid radiance for desired ray is available if uv plane coincides surface of the object, as shown in Fig. 5. By shifting uv plane, correspondence of each pixel of input images with (u, v), and parameters of desired ray vary. Consider the two cases in Fig. 6, where z coordinate of uv plane differs. Fig. 6(a) is the case in which the coordinate of uv plane is z_1 and desired ray is identified by

(s, t, u_1, v_1). Similarly in (b), the same ray is identified by (s, t, u_2, v_2) with uv plane at z_2. In each case, the ray queried from each grid point is different. Queried ray corresponds to a pixel of input image in our method, and shifting uv plane changes the pixel which desired ray refers. It also can be seen in Fig. 6.

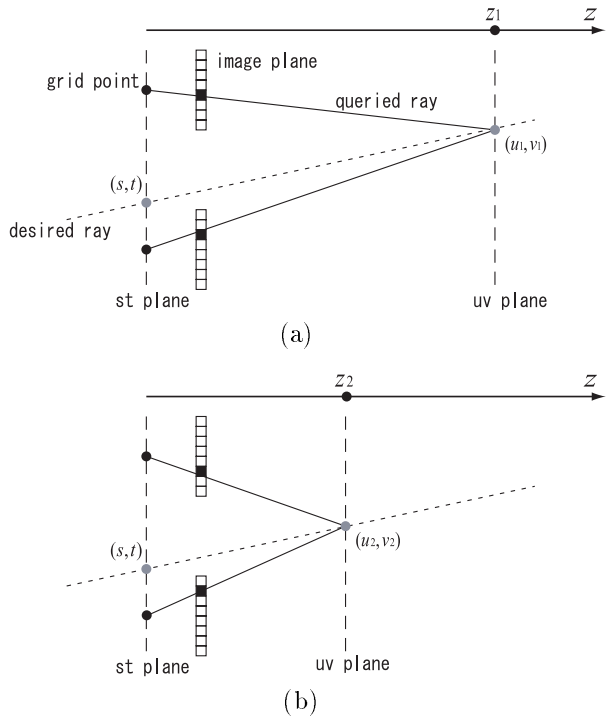


Figure 6. Correspondence of desired ray with a pixel of input image.

When uv plane coincides the surface of the object, the light field is optimized and appropriate radiances to generate fine images will be interpolated from queried rays. As a matter of fact, we can't make a plane to fit every part of the object unless the object is a single plane. A solution we found for the problem is to move uv plane during the rendering process to coincide every point of the object surface. So the z coordinate of uv plane has to be decided at every point of the object. Our optimization process is to find z coordinate of uv plane that generates the sharpest image at each point of the object. First, we decide z coordinate of uv plane for several points on the object surface respectively. Then z coordinate of other points are interpolated. This process resembles surface reconstruction process, which estimates 3D coordinates of feature points to form a mesh. However, the optimization process doesn't explicitly estimate the depth of the object, since the optimum position of uv plane is decided only by the sharpness of generated image. Points to estimate z coordinate don't have to be the feature points on the object. So we prefer to make estimations for evenly spaced points. Method for deciding z coordinate of single point is described from next section. This surface reconstruction like process has to be brought out only once when the light field is constructed.

4.2. Finding Optimum Position of uv Plane

4.2.1. Shifting uv plane

A virtual camera with extremely small image plane is placed far away from the planes (see Fig. 7). Then uv plane is shifted in the direction of z-axis little by little, with virtual camera capturing an image sequence. A small region of the object is captured in this sequence. The sharpness of the image changes gradually, since each frame is generated using uv plane at different z coordinate. The sharpest frame in the sequence is supposed to be generated when uv plane coincides the small region of the object surface. We assume the small region to be a single point, and optimum z coordinate for the point is decided by the position of uv plane which is used to generate the sharpest frame in the sequence. The sharpest frame in the sequence is decided by evaluating high-frequency energy.

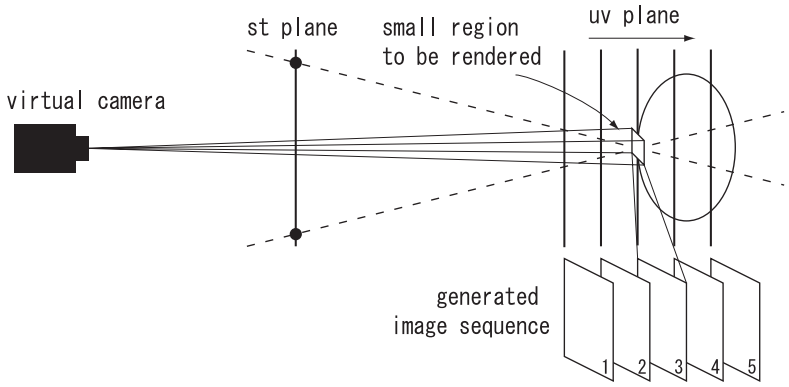


Figure 7. Shifting uv plane.

4.2.2. Evaluation of high-frequency energy

We apply a high pass filter represented by Equation (1) over each frame in the sequence.

$$g(i,j) = 1 - exp\left(-\frac{i^2+j^2}{\lambda}\right) \quad (1)$$

$$\left(\lambda \ : \ \text{constant} \right)$$

High-frequency energy in each frame is extracted in filtered images. We define sum of all pixel values in filtered image as high-frequency score for the frame. In the greater number of cases, a distinct peak can be found in score variation, and the frame that corresponds to the peak is the sharpest. There are cases that score variation is not in desirable way. Evaluating regions with little texture is in such a case, then z coordinate of uv plane can't be matched with object surface by evaluating high-frequency energy. However, it is not critical in our method since disagreement of rays doesn't make significant difference in generated images for such regions. We simply define z coordinate by finding the frame with highest score on the sharpness of the image, and we do not care whether it matches the actual depth of the object or not.

5. RESULTS

Computer-controlled camera gantry we used for our method is shown in Fig. 8. Image sequences are captured while controlling the camera motion with the gantry. Frames in sequences we captured are all 256×256 pixels of 8bit intensity resolution. Though each frame is a partly photographed image of the object, every part of the object can be seen in the sequence.

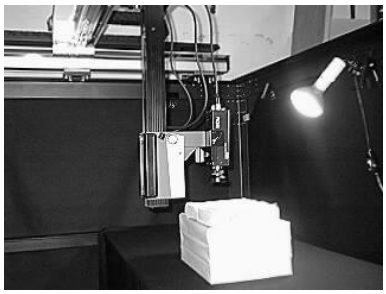


Figure 8. Camera motion is controlled by the gantry.

Figure 9. Examples of input frames. (256 × 256 pixels)

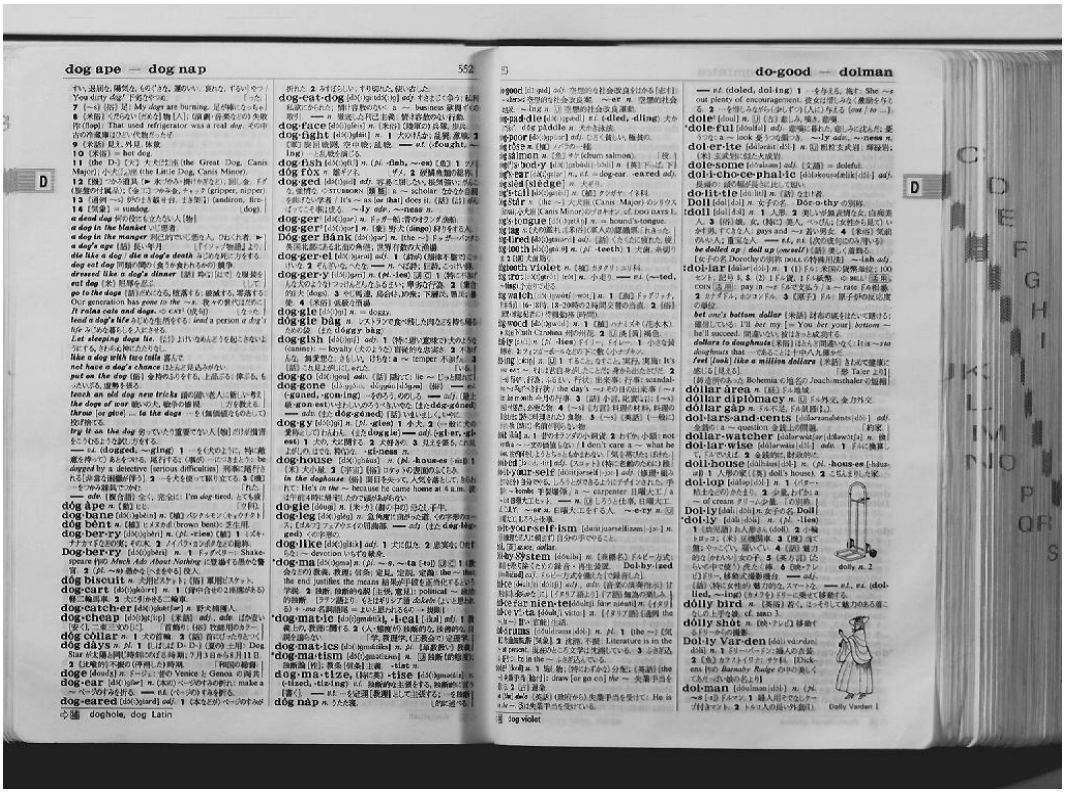


Figure 10. Image of entire object generated by proposed method. (2000 × 1500 pixels)

Image sequence of 1,600 frames were used to construct light field. These frames are images captured from 50×32 evenly spaced grid points at interval of $6mm$. Examples of input frames are shown in Fig. 9. An image of entire object generated by the proposed method is shown in Fig. 10.

Fig. 10 is virtually captured image using a camera with resolution that is impracticable, or hardly available. Details of real camera which captures the input images and virtual camera which captures image in Fig. 10 are shown in Table 1. Images of other scenes that are also virtually captured in this way are shown in Fig. 11.

Table 1. Details of the real and virtual camera

	resolution (pixels)	approximate distance to the object (cm)	captured region (cm^2)
real camera	256×256	25	4.5×3.5
virtual camera	2000×1500	200	28×20

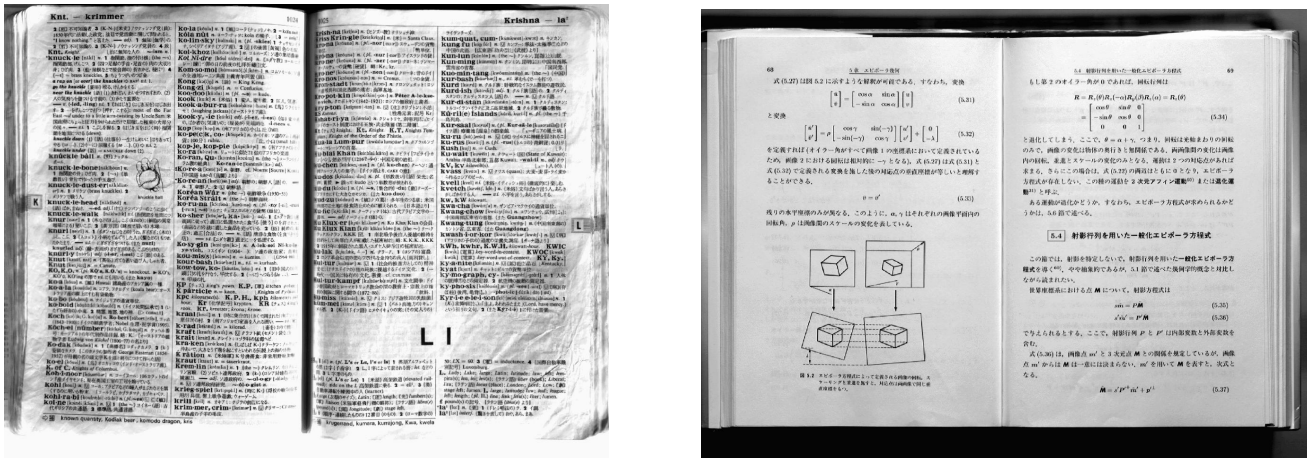


Figure 11. Results of other scenes.

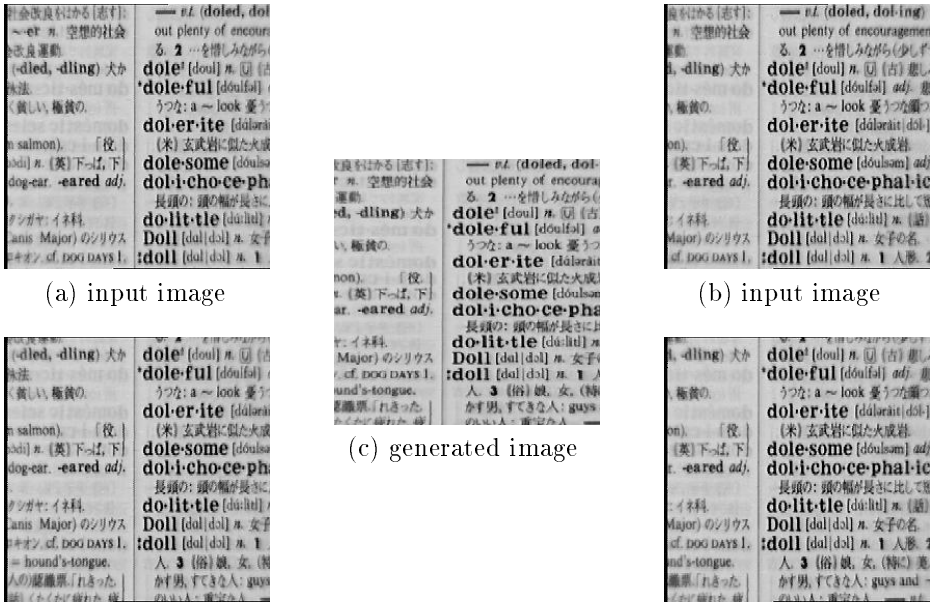
In Fig. 12, (a), (b), (d) and (e) are input frames captured from neighboring grid points, and (c) is a generated image in which viewpoint of virtual camera is located at the center of those of 4 input frames. Image in (c) is virtually captured image using light field constructed from input frames, and is not same as image generated by integrating 4 input images by 2D shifting.

By appropriately specifying parameters of virtual camera, image of arbitrary region of the object is generated without degrading resolution. Fig. 13 shows another example of image generated by our method. Though it is a partial image of the object, the range is larger than the input images.

Fig. 14 shows effectiveness of our optimization process. Fig. 14(a) is an image generated without optimization, that means generated with fixed uv plane, at $z = 50mm$ in this case. It is seriously blurred, while fairly improved image can be generated with optimization, as shown in (b). Image shown in (a) might be better if we put uv plane at different position, but note that rendering with fixed uv plane can't generate an image in which everywhere on the object is clearly seen. In our method, uv plane is shifted during the rendering process to adapt the depth variation.

6. CONCLUSIONS

We have proposed a method to generate high-resolution images from partly photographed images by constructing a light field to integrate these images. Optimizing light field successfully removes blur and discontinuity caused by depth variation of the object. Our method can generate images of arbitrary region of object, which are virtually captured by camera with extremely high resolution, that isn't available in real world. However, our optimization is insufficient for regions with rapid depth variation such as binding part of books. So one of the future works for us is to refine optimization method. Besides, images can't be rendered in real time yet and rendering process should be reconsidered.



(a) input image (b) input image (c) generated image (d) input image (e) input image

Figure 12. Generated image from arbitrary viewpoint.

Figure 13. Image of arbitrary region can be generated by light field. Resolution of generated image shown here is 512×512 pixels.

(a) without optimization (b) with optimization

Figure 14. Effectiveness of optimization.

REFERENCES

1. Aaron Isaksen, Leonard McMillan, Steven J. Gortler, "Dynamically Reparameterized Light Fields," in *Technical Report MIT-LCS-TR* 778, 1999.
2. Marc Levoy, Pat Hanrahan, "Light Field Rendering," in *SIGGRAPH96*, Computer Graphics Proceeding, pp. 31–42, 1996.
3. Leonard McMillan, Gary Bishop, "Plenoptic Modeling : An Image-Based Rendering System," in *ACM SIGGRAPH*, Computer Graphics, Annual Conference Series, pp. 39–46, 1995.
4. T. Nakajima, M. Kashimura, S. Ozawa, "Compensation of Partly Photographed Page-Images Using 3-D Shape Information," in *IEEE ICMCS'99*, pp. 179–183, 1999.
5. Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, Michael F. Cohen, "The Lumigraph," in *SIGGRAPH96*, Computer Graphics Proceeding, pp. 43–54, 1996.
6. R. Y. Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," in *IEEE Journal of Robotics and Automation*, RA-3 No.4, pp. 323–344, 1987.
7. R. Szeliski "Video mosaicing for virtual environments," in *IEEE Computer Graphics and Applications*, 1996.
8. Anthony Zappara, Andrew Gee, Michael Taylor "Document Mosaicing," in *BMVC97*, Vol.2, pp. 600–610, 1997.