Stereo Imaging Using a Camera with Stereoscopic Adapter

Woontack Woo, Namgyu Kim* and Yuichi Iwadate**

ATR MIC Labs,
Kyoto 619-0288, Japan
Email:wwoo{ngkim,yiwadate}@mic.atr.co.jp

ABSTRACT

In this paper, we analyze the characteristics of the stereoscopic adapter, which is a cost-effective way to generate stereo video sequences with a camera. We also propose an efficient way to compensate for the inherent distortions. In general, stereo sequences can be captured using a pair of cameras, but the resulting sequences tend to yield various well-known problems due to different characteristics of the pair of stereo cameras. Meanwhile, a camera with the stereoscopic adapter provides a natural way to capture and display stereoscopic video. It allows users to access all the functions built into the camera, e.g. zoom, auto-focus, auto-exposure, special effects, etc. The cost however is the reduced quality of the videos since the adapter allows capturing stereo video sequences in the field sequential format, i.e. left and right images in different scan lines, respectively. In addition, it generates size and color distortions due to the physical configuration of the mirror in the adapter. We analyze and compensate for such distortions to reduce possible errors in vision applications exploiting the stereo images. According to our preliminary study, the adapter with the proposed compensation scheme will pave the way for various low-cost image-based virtual reality applications at hand.

Index terms – stereo camera, calibration, rectification, disparity estimation, stereoscopic adapter.

I. INTRODUCTION

Exploiting stereo images provides various advantages over using an image. For example, stereo vision provides 3D information, such as orientations and distances, of the objects in the scene. In addition, well-designed stereoscopic displays convey a very compelling sense of 3D depth. Although 3D perception can be achieved through various other cues (such as geometric perspective, relative size, shade, texture gradient, occlusion, motion, disparity, etc.), binocular depth perception is considered to be much more powerful. Even in 2D display environments, exploiting the 3D depth information can lighten the burden of image/video processing, analysis and communication in various levels, *e.g.* help segment objects from the background [10,11], which has been one of the hardest computational vision problems.

The difficulties in stereo imaging mainly stem from capturing well-controlled stereo images, which is a key step toward accurate depth estimation. In general, stereo images can be captured using a pair of stereo cameras, where each camera captures a scene from a slightly different perspective. However, several well-known problems arise from capturing stereo images/video sequences, since two cameras will generally have slightly different physical characteristics. Without accurate camera calibration, we may fail to estimate accurate 3D information and to provide realistic 3D effects on the screen.

Meanwhile, a camera with a stereoscopic adapter, *e.g.* NuView system, can be considered as a new way to capture stereo images/video sequences [1]. The optical adapter, placed in front of the lens of a camera, allows for the camera to capture stereo video sequences. As a result, it can alleviate the problems, which arise from the different characteristics of a pair of stereo cameras. It also allows users to access all the functions built into the camcorder, e.g. zoom, auto-focus, auto-exposure, special effects, etc. Note however that the cost of this single lens-based approach is the reduction in quality of the resulting stereo video sequences. The adapter captures each image of the stereo pair in the different line, *i.e.* field sequential format. In addition, inherent distortions due to the mirror in the stereoscopic adapter may deter the usage of the resulting stereoscopic images in various applications exploiting 3D depth information. Therefore, a simple yet successful way to compensate for these distortions is essential for the adapter to be used in real vision applications.

In this paper, we focus on compensation for these 3D distortions in vision applications, rather than the 3D display itself. However, it is worthy noting that compensation is also essential for comfortable stereoscopic 3D display. First we analyze the characteristics of the adapter, then we explain how to compensate for the resulting distortions based on analyzed results. To analyze the characteristics of the adapter, we perform a three-step procedure as follows. First, we separate the test sequence in the field sequential format into the above/below format and then transform to side-by-side format by bilinear interpolation. Second, we rectify the image, *i.e.* compensate for the size distortion, by using the parameters obtained from the Tsai algorithm [8]. Finally, we compensate for the color degradation, which is inevitable due to the projection through the mirror in the adaptor, based on the analyzed statistics. According to our preliminary study, the stereoscopic adapter with the proposed

* N. Kim is a Ph.D. candidate at Pohang University of Science and Technology (POSTECH), Pohang, Korea.
** Y. Iwadate is now with NHK Science and Technology Research Labs, Tokyo, Japan. Email: yiwadate@strl.nhk.or.jp

0-7803-6583-6/00/$10.00 © 2000 IEEE

compensation scheme will pave the way for various low-cost image-based virtual reality applications at hand.

This paper is organized as follows. In Section 2, we introduce a way to capture stereo video sequences using a camera with the stereo adapter. In Section 3, we analyze the characteristics of the stereoscopic adapter and explain how to compensate for the resulting embedded distortions. Some experimental results and discussion are given in Section 4 and 5.

II. STEREOSCOPIC VIDEO WITH A CAMERA

A. Field Sequential 3D Video

As shown in Figure 1, the stereoscopic adapter (*e.g.* Nu-View system [1]) consists of a sturdy black plastic housing, a reflecting mirror and liquid crystal shutters (LCS). The prismatic beam splitter and the orthogonally positioned polarizing surfaces (1.45"x1.25") in the LCS open and close the light valves, to record either the direct image or the mirror reflected image on alternate fields of the video. As a result, the left image is recorded during the "odd" field and the right image during the "even" field, or vice versa. As shown in Figure 1, the synchronization of the light valves with the alternating fields of the camcorder is achieved through the cable connecting the video-out of the camcorder and the connector in the adapter.

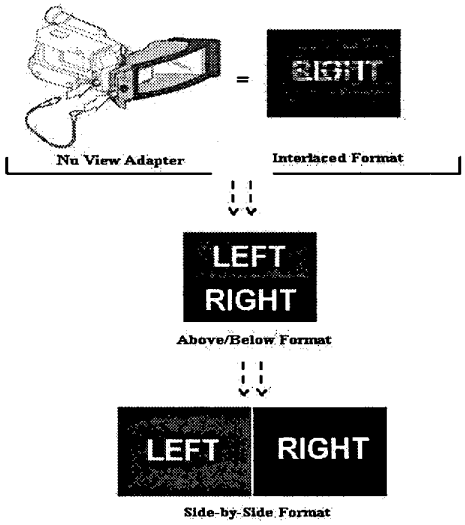


Figure 1. Capturing stereo video sequences using a camera with stereoscopic adapter.

B. Field Separation and Interpolation

It is necessary to convert the video sequence to above/below format. As explained earlier, the adapter produces a field sequential stereoscopic 3-D video by simultaneously recording the second eye view to the camcorder. The resulting field sequential video can be display on a 2D (TV) monitor or 2D screen with special stereo glasses. The field sequential format, however, is an inconvenient format to be used in various other vision applications. For example, applying processing, such as filtering or transformation, to the field sequential video can cause a loss in quality of the stereo images because such processing propagates the effects into the interlaced lines and thus produces 3D artifacts. For the same reason, the available video compression scheme cannot be exploited to save the hard disk space or limited channel bandwidth. Therefore, we first separate the field sequential format to above/below format, where the left image is placed to the top part of the image and the right image to the bottom part, or vice versa.

After the field separation, we transform the image to side-by-side format. We now need to perform temporal or spatial interpolation to each image to provide a high quality of 2D/3D images/video sequences. We experience the flickering effects when we display the stereoscopic 3D videos, which are captured using the adapter in 60Hz. The stereoscopic 3D video in 60Hz is not as smooth as compared to the 2D video in 60Hz, because the 2D monitor allocates 30Hz to the right image and the other 30Hz to the left image. In addition, displays (such as head-mounted display, polarized screen or auto-stereoscopic display) require projecting an image in the original size to provide comfortable 3D display. The spatial interpolation is also required in 2D applications exploiting only 3D depth information (e.g. z-keying). The spatial interpolation is achieved by line copy, doubling of the size, or linear interpolation between lines as follows.

$$\begin{cases} F_L^{2i} &= G_L^i \\ F_L^{2i+1} &= (G_L^i + G_L^{i+1})/2 \end{cases} \quad (1)$$

where F_L and G_L denote the left images in side-by-side and above/below formats, respectively. The superscript i represents the index of the row in the image. The right image can be interpolated in a similar way.

III. DISTORTION AND COMPENSATION

A. Geometry of a Pinhole Camera

Figure 2 shows the basic geometry of the pinhole camera model and its projection from the 3D world coordinate of a feature point (in mm) $[x_w\ y_w\ z_w]$ to 2D image coordinate (in pixels) $[u\ v]$. The 3D camera coordinate is centered at the optical center, O_c, with the c_z-axis being the same as the optical axis. The image coordinate is centered at the intersection point with the optical axis in the image plane. The distance between the optical center and the image plane is the focal length f.

We first transform the augmented 3D world coordinate of the object $[x_w\ y_w\ z_w\ 1]^T$ to the 3D camera coordinate $[x_c\ y_c\ z_c]^T$. The transformation can be performed by a 3x3 rotation matrix R and a 3x1 translation vector $T=[t_x\ t_y\ t_z]^T$, as shown in (2).

$$[x_c \quad y_c \quad z_c]^T = [R \quad T]\cdot[x_w \quad y_w \quad z_w \quad 1]^T \quad (2)$$

where the rotation matrix R is represented as a product of three 3x3 rotation matrixes, *i.e.* $R_\phi R_\theta R_\xi$, corresponding to the rotation along roll (ϕ), pitch (θ) and yaw (ξ), respectively.

$$R=\begin{bmatrix}1 & 0 & 0\\ 0 & c\phi & -s\phi\\ 0 & s\phi & c\phi\end{bmatrix}\begin{bmatrix}c\theta & 0 & s\phi\\ 0 & 1 & 0\\ -s\theta & 0 & c\theta\end{bmatrix}\begin{bmatrix}c\xi & -s\xi & 0\\ s\xi & c\xi & 0\\ 0 & 0 & 1\end{bmatrix}$$

where s and c denote sine and cosine, respectively. The 3D camera coordinates are projected onto the augmented ideal image coordinate $[x_i\ y_i\ 1]^T$ by the perspective projection under pinhole camera model.

$$\begin{bmatrix}x_i\\ y_i\\ 1\end{bmatrix}=\frac{1}{z_c}\begin{bmatrix}f & 0 & 0\\ 0 & f & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_c\\ y_c\\ z_c\end{bmatrix} \qquad (3)$$

where f denotes the effective focal length of the camera. As usual, the origin of the image coordinate is in the upper left corner of the image array and the unit of the image coordinate is not "meter" but the number of "pixels". Therefore, the augmented actual (or computer) image coordinates in pixels $[u_i\ v_i\ 1]^T$ are obtained from $[x_i\ y_i\ 1]^T$ by applying the following transformation.

$$\begin{bmatrix}u\\ v\\ 1\end{bmatrix}=\begin{bmatrix}s_x & 0 & u_0\\ 0 & s_y & v_0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_i\\ y_i\\ 1\end{bmatrix} \qquad (4)$$

where the (u_0, v_0) denotes the offset (pixels) in the image and s denotes the scaling (pixels/m). The scaling is defined as, $s=a$ x k x d x N_s/N_p, where d is the center-to-center distance between adjacent sensor element, and N_s and N_p, respectively, denote the number of sensor elements in horizontal (vertical) direction in the CCD and the number of pixels in an image scan (vertical) line. Note that, in general, both the uncertainty parameter a and the lens distortion parameter k are set to one, but in a real situation both should be introduced [2,8]. By combining (2)-(4), the 3D world coordinate is related to the 2D image coordinate and vice versa.

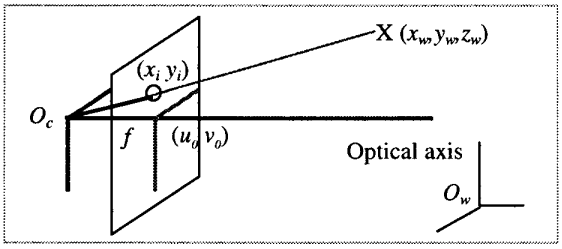


Figure 2. Camera geometry and image projection.

B. Stereo Calibration

Now we are ready to investigate the virtual camera configuration of the camera with stereoscopic adapter. We first perform camera calibration to capture the relationship between the 3D world coordinate and its 2D perspective projection onto the virtual stereo cameras. The calibration of a camera is performed by observing a calibration object, *e.g.* one or two plane model, whose geometry is known with a very good precision with respect to a 3D coordinate system attached to this apparatus. The resulting camera parameters allow us to learn the distortions that occurred during projection through the stereoscopic adapter.

In general, standard stereo camera calibration techniques follow 3-step procedures. First, they establish a list of 3D world coordinates and corresponding 2D image coordinates. Given the list, camera parameters are estimated for each camera using a set of equations. Finally, the epipolar geometry is constructed from the projection matrices.

We apply Tsai algorithm to determine the distortion model and model parameters [Tsai98]. The algorithm estimates 11 model parameters: five intrinsic (also called internal or interior) and six extrinsic (also called external or exterior) parameters. The intrinsic camera parameters include the effective focal length f, the first order radial lens distortion coefficient κ_1, the principal point (the center of radial lens) $[c_x, c_y]$, the scale factor to account for any uncertainty due to frame grabber horizontal scanline resampling s_x. The extrinsic parameters include the rotation matrix R (rotation angles for the transform between the world and camera coordinate frames) and transformation matrix T (translational components for the transformation between the world and camera coordinate frames).

C. Stereo Rectification

Figure 3 shows the epipolar geometry between a pair of images. In 3D video shooting, the tradeoffs between 3D effects and 3D distortions are unavoidable. The objects at the convergence point, corresponding to zero-disparity, appear on the 2D screen and others will appear with relative depth. As the convergence point moves toward the camera, the 3D distortions such as "keystone effects" increase [9]. Even if we move the convergence point backward, it would be very difficult to eliminate such distortion at all object points. If the convergence point is not in the infinite point, the stereoscopic adaptor causes the rotation, as well as translation, of the virtual camera. As a result, epipolar lines are not aligned with coordinate axis and are not parallel, which makes disparity estimation difficult.

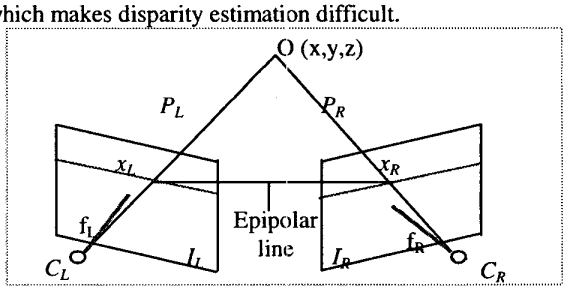


Figure 3. Epipolar geometry between a pair of images.

Once the camera model parameters have been identified, we compensate for the size distortion by an operation known as stereo rectification [4-5]. The rectification brings the two image planes to be coplanar to a common plane in space [7]. Let C_L and C_R be a pair of pinhole cameras in 3D space. Each

camera is represented by a 3x4 homogeneous transform matrix, $H=[P| -PC]$, where the vector C denotes the position of the camera's optical center. The 3x3 projection matrixes, P_L and P_R, between the augmented 3D world coordinates O and the image coordinates, x_L and x_R, can be calculated using Tsai algorithm. The rectification can be accomplished by the transformation matrix obtained from the relationship between the two matrices, P_L and P_R [7].

Given two projection matrix, $H_L=[P_L| -P_LC_L]$ and $H_R=[P_R| -P_RC_R]$, the corresponding points in the images, x_L and x_R, are related by $x_R=P_RP_L^{-1}x_L$. The projective matrix, $P_RP_L^{-1}$, projects the image plane of I_L to that of I_R. More generally, the matrix, $P_L'P_L^{-1}$, represents a planar projective transformation onto a new image I_L'. It is convenient to choose the world coordinate system so that both C_L and C_R lie on the world X-axis, i.e. $C_L=[X_L\ 0\ 0]^T$ and $C_R=[X_R\ 0\ 0]^T$. The remaining axes are chosen in a way that reduces the distortion incurred by image reprojection. After a proper rectification process, the rectified images have the following properties; (a) all epipolar lines are parallel to the horizontal scan line and thus (b) corresponding points have identical vertical coordinates. As a result, the matching process of two images can be simplified and efficient.

D. Color Histogram Modification

Before we exploit stereo images, we need color modification. The color distortion occurs due to another inherent weakness of capturing stereo video with the stereoscopic adapter. The orthogonally positioned polarizing surfaces in the adapter yield stereo video sequences with different levels of color. Note that color equalization not only allows comfortable stereoscopic 3D display, but also helps to estimate accurate 3D depth information, especially when the depth is estimated based on the intensity level.

To equalize the color levels of both images in the stereo pairs, we use 3 test pairs, where each pair contains only one color, *i.e.* red, green or blue. We first select the region of interest from the given pairs of stereo images and then estimate statistics regarding the color distortion. Given the statistics, we normalize and modify the color histograms. The new intensity (color) of the left image, the projected image through the mirror, F_{L_N}, is normalized using the formula defined as follows.

$$F_{L\_N}=\frac{\sigma_R}{\sigma_L}(F_L-m_L)+m_R \qquad (5)$$

where F, σ and m represent the image, standard deviation and average, respectively. The subscript L and R denote left and right images, respectively.

E. Disparity Estimation

Disparity estimation is recognized as the most difficult step in stereo imaging. The task of disparity estimation in a 3D reconstruction is to find correspondence in a pair of stereo images and estimate 3D position by triangulation. Many estimation algorithms have been proposed but the resulting disparity is not accurate enough to be used in real applications. To overcome the weakness of available estimation schemes, we adopt a hierarchical block matching scheme with hybrid cues, which exploit edge, as well as intensity similarity, based on MRF framework [10-12]. We start with a block to maintain robustness of the estimation and then segment the block. Since the estimation error within the block is non-uniform, we segment the block according to the estimation error and variance level. The edge information is exploited to estimate accurate disparity along the object boundaries.

IV. Experimental Results

After mounting the Nu-view adapter on the SONY TRV900, we capture test patterns and sequences in field sequential format. We then digitize the sequences using the video card, DV Rapter, with the capturing software, Adobe Premier. To maintain a stereo sync signal, the captured images are set to full size (720x480 at 30.0 fps) [1]. The even lines contains the picture for left-eye and the odd lines for right eye. As explained in Section II and III, we first transform the images in the field sequential format to the side-by-side format and then interpolate using (1) to recover the original size.

In our experiments, a non-coplanar pattern is used to estimate camera calibration parameters. As shown in Figure 4 (a), the 12 rectangles in the calibration wedge are projected on to the 2D image plane. The corner points of the rectangles on the wedge are projected onto image planes and the image coordinates corresponding to the calibrated points on the wedge are determined using simple image processing techniques. The calibration is performed using Tsai algorithm [8]. The resulting camera parameters are shown in Table 1.

Figure 5 shows the resulting virtual stereo camera configuration of the camera with the stereoscopic adapter. As shown in Figure 5 (a), the virtual camera is not positioned on the same plane as the reference (right) camera. The set back of the virtual (left) camera results in the size distortions in a pair of stereo images and the rotation of the camera yields additional "keystone" error. According to our experiments, the virtual camera is set back about 50 mm from the reference camera due to the mirror in the adapter. The closer convergence point causes more size distortion. The set back and rotation of the virtual camera is compensated for using the camera parameters obtained by Tsai algorithm, which are shown in Table 1.

After the camera calibration and image rectification, we perform color modification using (5). To analyze the characteristics of color distortion, we capture three test images, each containing only one color component, i.e. red, green and blue, respectively. Figure 6 (a), (c) and (e) show the histograms of uncompensated images for each color component. As shown in Figure 6 (b), (d) and (f), the left video sequences can be compensated by using (5) for red, green and blue components, respectively. Note that the statistics (mean and variance of each sequence) can be estimated during camera calibration process and be applied on the fly.

Figure 7 (a) and (b) show a pair of stereo images in field sequential format and above/below format, respectively. We first rectify the images using the calibration parameters to compensate for size distortion. We then modify each of the color components since, as expected, the pair of stereo images has slightly different color levels. Figure 8 compares cumulative histograms of the original and the compensated pairs of stereo images, in terms of each color component, *i.e.* red, green and blue, respectively. According to our experimental results on the outdoor scenes, the polarization of LCD in the adapter causes color shift in the red component on the right image. The comparison between the original and the calibrated right image is shown in Figure 9. The calibration and rectification will improve disparity estimation dramatically.

V. DISCUSSION

We first analyzed the characteristics of the stereoscopic adapter, and then proposed an efficient scheme to recover the original quality of video. As shown in our experimental results, in addition to the classical lens distortions, the stereoscopic adapter generates different kinds of distortions in terms of size and color. Those distortions can be compensated for by the stereo rectification and color modification. According to our preliminary study on the characteristics of the stereoscopic adapter, we believe the adapter with the proposed compensation scheme can provide a practical solution to capture 3D video at hand with reasonable quality. As a result, with the portability and effectiveness, the adapter with the proposed compensation scheme will play a key role in generating image-based photo-realistic virtual environment with depth information. The remaining work for the adapter to be used in real applications is to analyze the effects according to the changing focused point and zooming environments [8].

Acknowledgement

The authors would like to thank Mr. David Chersky and 3-D Video, Inc., CA, USA, for providing the Nu-View stereoscopic adaptor used in this paper.

References

[1] Nu-View Stereoscopic Adapter, 3-D Video, Inc., http://www.3-dvideo.com

[2] C. Cheung and W. Brown, "3D Shape Measurement using Three Camera Stereopsis in Optics," Proc. SPIE Illumination and Image Sensing for Machine Vision II, vol. 850, pp. 128-139, 1987.

[3] U. Dhond and J. Aggarwal, "Binocular Versus Trinocular Stereo," in IEEE Proc. Int. Conf. on Robotics and Automation, pp. 2045-2050, 1990

[4] O. Faugeras, " Three-dimensional Computer Vision: A Geometric Viewpoint," MIT Press, Cambridge, MA, 1993.

[5] C. Loop and Z. Zhang, "Computing Rectifying Homographies for Stereo Vision," in IEEE Proc. Int. Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 125-131, Jun. 1999.

[6] J. Park and C. Lee, "Robust Estimation of Camera Parameters from Image Sequence for Video Composition," Signal Processing: Image Communication, vol. 9, pp.43-53, 1996.

[7] S. Seitz and C. Dyer, "View Morphing," SIGGRAPH 96, pp. 21-30, 1996.

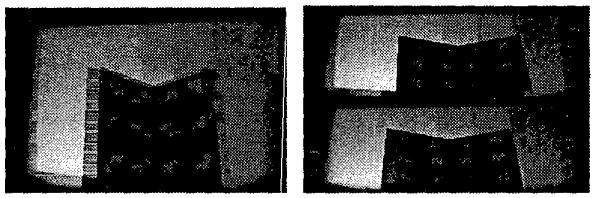
[8] R. Y. Tsai, A versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", IEEE Journal of Robotics and Automation, Vol. RA-3, No. 4, August 1987, pages 323-344.

[9] W. Woo, "Rate Distortion Based on Dependent Coding for Stereo Images and Video: Disparity Estimation and Dependent Bit Allocation," Ph.D. Dissertation, USA LA, CA, USA, 1998.

[10] W. Woo and Y. Iwadate, "Object-oriented Hybrid Segmentation Using Stereo Images," Proc. IVCP'00, Jan. 2000.

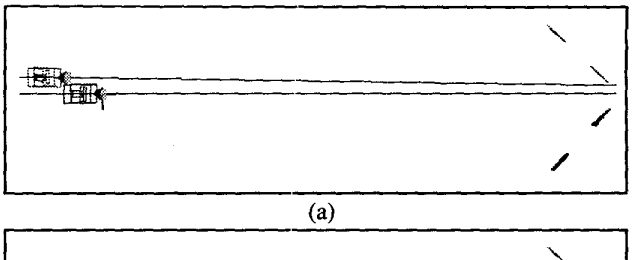
[11] W. Woo and A. Ortega, " Overlapped Block Disparity Compensation with Adaptive Windows for Stereo Image Coding," IEEE Tr. on CSVT, vol. 9, no.6, pp. 194-200, Mar. 2000.

[12] W. Woo, N. Kim and Y. Iwadate, "Object Segmentation for Z-keying Using Stereo Images," in Proc. WCC-ICSP'00, Aug. 2000.



(a) (b)

Figure 4. Test pattern for stereo calibration and rectification. (right image, 720x240) (a) field sequential format (b) above/below format.



(a)

(b)

Figure 5. Virtual camera configuration. (a) before calibration (b) after calibration.

	Left	Right		Left	Right
f [mm]	68.70	65.88	Rx [deg]	151.06	150.67
k1 [1/mm^2]	2.5E-05	4.5E-05	Ry [deg]	-42.59	-42.77
Cx [pixs]	359.06	364.89	Rz [deg]	18.94	19.46
Cy [pixls]	237.54	232.73	Tx [mm]	-232.64	-293.02
s_x	0.85	0.85	Ty [mm]	-42.54	-34.04
			Tz [mm]	1618.91	1491.88

Table 1. The resulting camera parameters. (a) right camera (b) virtual left camera.

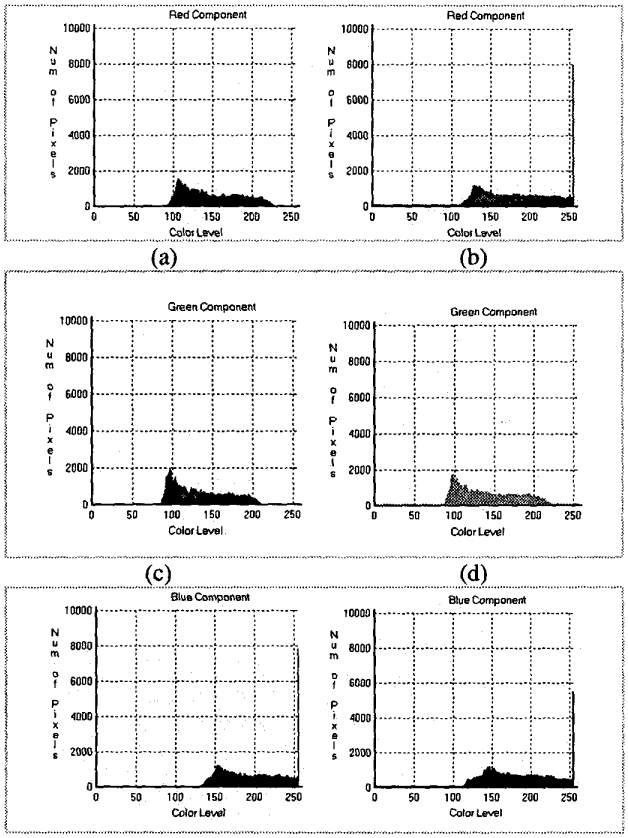


Figure 6. Histogram of test color images. (a) red-left (b) red-right (c) green-left (d) green-right (e) blue-left (f) blue-right.

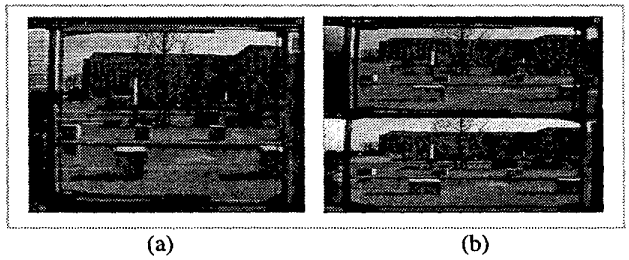


Figure 7. Test images (720x480). (a) field sequential format (b) above/below format.

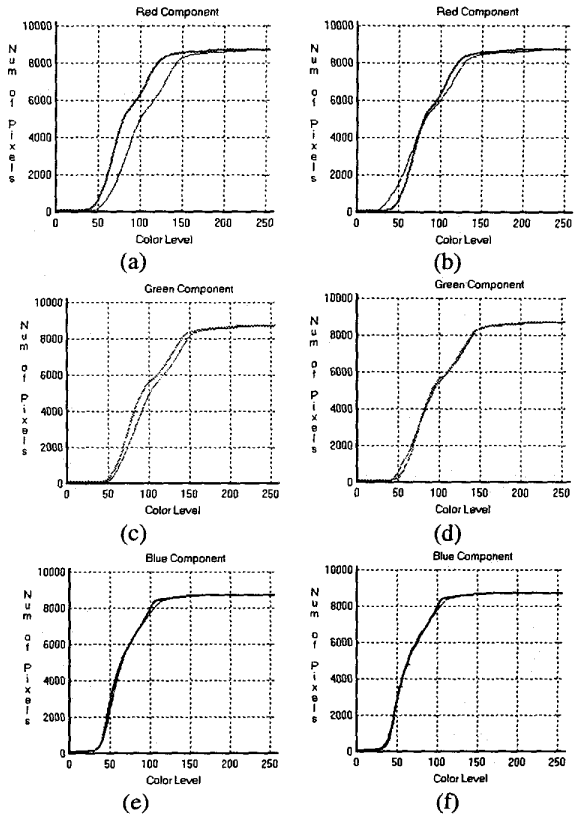


Figure 8. Comparison of Histogram. (a) red-original (b) red-modified (c) green- original (d) green-modified (e) blue- original (f) blue- modified.

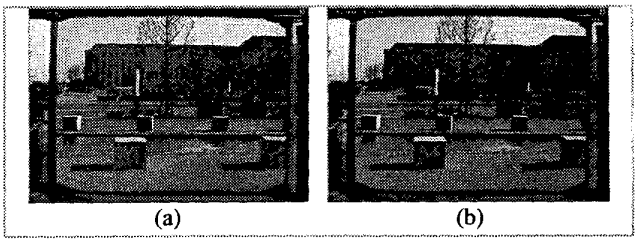


Figure 9. Color compensation. (left image, 720x480) (a) the original (b) the compensated image.