3차원 영상을 위한 다초점 방식 영상획득장치

Multi-Focusing Image Capture System for 3D Stereo Image

함 윤 철¹, 권 혁 재², 투멘자르갈 엔크바타르²

Woon-Chul Ham¹, Hyeok-Jae Kwon², Tumenjargal Enkhbaatar²

Abstract In this paper, we suggest a new camera capturing and synthesizing algorithm with the multi-captured left and right images for the better comfortable feeling of 3D depth and also propose 3D image capturing hardware system based on the this new algorithm. We also suggest the simple control algorithm for the calibration of camera capture system with zooming function based on a performance index measure which is used as feedback information for the stabilization of focusing control problem. We also comment on the theoretical mapping theory concerning projection under the assumption that human is sitting 50cm in front of and watching the 3D LCD screen for the captured image based on the modeling of pinhole Camera. We choose 9 segmentations and propose the method to find optimal alignment and focusing based on the measure of alignment and sharpness and propose the synthesizing fusion with the optimized 9 segmentation images for the best 3D depth feeling.

Keywords: Capturing, 3D depth, Calibration, Mapping, Segmentation, Pinhole Camera

1. Introduction

3D reconstruction from 2D images is an active research topic in the computer vision community. Classical algorithms like Structure From Motion (SFM) and Multi-View Stereo (MVS) are known to provide a robust framework for reconstruction, but fail to produce a dense plausible reconstruction. Some recent works have shown that making planar approximations of the scene help obtain more complete and visually pleasing results albeit, still approximate. However, many of these algorithms depend on image level cues like strong edges and lines which may be absent in texture less scenes.

This has led to interactive 3D reconstruction algorithms. The CMU 3D room proposes an algorithm for view-dependent no uniform sampling for image-based rendering. This algorithm may be needed for making a good 2D image for many viewers when each of them has his own different geometric view plane^[1]. A taxonomy and evaluation system proposed by D.S charstein^[2] is widely used in evaluating the disparity accuracy in stereo matching algorithms with a set of benchmark image data sets. There has been many researches for estimation of optical flow especially for estimating of an object by using 2D images under the assumption that the light source is invariant with respect to movement of the object which is not true in real life. Fleet and Weiss provide a tutorial introduction to gradient based optical flow^[4,5] under a image constraint equation which is not always true in 2D image. Stereoscopic video provides many advantages over regular 2D video: Improved image understanding, improved ability to see and judge size and distance.

Received : Sep. 30. 2010; Reviewed : Nov. 18. 2010; Accepted : Jan. 13. 2011

※ This research was supported by post BK21 project of Chonbuk National University and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the ministry of Education, Science and Technology(2011-0005204).

¹ Professor of Division of Electronics and Information Engineering at Chonbuk National University

² Ph.D student of Division of Electronics and Information Engineering at Chonbuk National University

Stereoscopic video also can improve the tele-operation performance and reduce task times and the risk of damage. The perceived depth seen in stereoscopic images varies not only with many geometric parameters but also with the many parameters with human eye such as separation between two eyes and focusing of each eye. It is well known fact that there is a limit to the range of perceived depth. Human cannot detect the depth difference between two objects if two objects are located beyond 350m from the viewer. So it is useless to consider the scene depth of object if its distance is very far from the viewer. This is the reason why the research should be done concerning relationship and optimal mapping between scene depth and perceived depth for the best implementation of stereoscopic image. N. Holliman addresses this problem and propose multiregion algorithm and this algorithm allows different regions of the scene to be mapped to the different ranges of perceived depth on a target 3D display. The main idea of N. Holliman is to finding a smooth mapping in order to give more dynamic range of perceived depth for the region of interest in the scene depth^[6-8]. In his algorithm, he considers the scene outside of the region of interest without neglecting it at all because he thinks that it is important to see the context around the region of interest in some applications. He also considers the visible discontinuity at the boundary of the multiregion algorithm. We also proposed new mapping algorithm between scene depth and perceive depth^[9]. In this paper, we propose theoretical and technical ideas for developing a 3D stereo video camera system based on previous research works. We comments on the important geometric parameters for synthesizing fusion with the multi-captured images. At first we simulate our research with OpenGL software to obtain 9 segmentation of the scene with the assumption that human do not rotate his head and just rolling his eye to watch the scene. We calculate and decide geometric parameters based on this assumption for obtaining virtual 9 segmentations of the scene. We define the measure of the alignment of stereo

camera capture system and propose the method how to decide the optimal alignment based on this measure. We also define the measure of sharpness of an captured image and try to solve the optimal focusing problem. In this paper, we present 5 degree of freedom (DOF) stereo camera system which is the first version of our laboratory, as shown in Figure 1. We use 4 DC servo motor and one stepper motor for mechanical control. We also develop a control algorithm to align the two left and right camera capture system by measuring the offset distances. In next version, we will add 2 more DOF for each left and right camera system for initial calibration which is very important to obtain a good stereo image. We begin with a brief review and mathematical preliminary for the technology and theory for 3D image processing and propose a decision algorithm whether an object is within the trapezoidal region or not and also comment on the projection theory. In section 3, we discuss the technical issue for calibration and suggest a simple control law for adjusting the focus and alignment for capturing several segmentations of image. We also propose a new performance index measure for focus and alignment problem in capturing the left and right images. In section 4, we suggest simple fusion algorithm to make a whole 3D stereo image from the several segments of image which are captured by stereo camera system operating with the proposed control algorithm. In section 5, we simulate our proposed algorithm with the virtual objects synthesized on circular polarized LCD screen by using OpenGL programming. We conclude with a summary of this paper and a discussion of planned future work.

2. Mathematical Preliminary

In this section, we introduce the mapping theory that is necessary to understand and design the parameters for the 2D projection view of 3D object. In stereo vision signal processing, we should synthesize the 3D object into the 2D circular polarized LCD screen. Therefore we should be good at the special planar mapping as shown in figure 1.

Fig. 1. 2 Dimensional Projective Image

The key idea is that planar mapping function is dependent on the geometric parameters between LCD and the eyes of human who is sitting in front of LCD. In this paper, we comments on the geometric parameters which should be considered in synthesizing the 3D stereo image from the image data of two left and right camera system. As we can observe from figure 1, the relation between the projective view v_1 of camera c_1 and the projected view v_2 of camera c_2 cannot be obtained from a linear transformation because the projective operation is nonlinear transform operation. If the distance between the two cameras is very close to each other compared with the distance of the object from the camera, then we can approximate and find the linear transformation between two projective images and this linear transformation become identity matrix. But in real situation, this assumption is not always satisfied. This is the reason why we comments on the geometric parameters between object and camera. We insist that these geometric parameters play very essential role for synthesizing a good quality of 3D stereo image.

Now we introduce the mathematics for projection process from the 3D image space to 2D image plane for synthesizing 3D stereo image into 2D polarized LCD screen. We also propose the algorithm whether or not a point is inside of trapezoidal region to decide whether a point should be considered or not in synthesizing a 3D stereo image.

2.1 Geometric Parameters

The geometric coordinate relation between the eye of the viewer and virtual LCD screen can be expressed as follows.

$$T_{se} \begin{pmatrix} x_s \\ y_s \\ z_s \end{pmatrix} = \begin{pmatrix} x_e \\ y_e \\ z_e \end{pmatrix} \quad (1)$$

Where

$$T_{se} = R_{(x, -\theta_x)} R_{(y, -\theta_y)} T_{(y, -d_y)} T_{(x, -d_x)} T_{(z, -d_z)} \quad (2)$$

In this paper, the virtual LCD denotes the 2D screen window through which a viewer watch the objects and we should synthesize this virtual scene into real LCD screen located at with the same orientation and position as those of virtual LCD. The scene synthesized with this suggested method is the optimized scene under the assumption that a person is going to watch this scene with the same geometric relation with real LCD screen. In there, (x_s, y_s, z_s) denotes the coordinate of the virtual LCD screen coordinate system and (x_e, y_e, z_e) denotes the coordinate of the eye coordinate system of the viewer as shown in figure 2.

The main problem to solve in this paper is to obtain the equation that makes the objects which is behind the

Fig. 2. Geometric Parameters

virtual LCD to be projected onto the virtual LCD screen in the direction of the eye view point of viewer. We define the trapezoidal region R which will be projected onto the virtual LCD by using the 4 corner points P_1, P_2, P_3, P_4 and the range of scale k as follows in the eye coordinate system.

$$R = \text{trapezoidal}(p_1, p_2, p_3, p_4) \quad (3)$$

The trapezoidal region R is shown in figure 3.

In the next subsection, we propose an algorithm whether a point is within the trapezoidal or not.

Fig. 3. Trapezoidal Region

2.2 Belonging Decision Algorithm

In this subsection, we propose an algorithm whether or not a point is inside of trapezoidal region to decide whether a point should be considered or not in synthesizing a 3D stereo image. This algorithm is needed to make a stereo 3D image from the synthesized scene and objects made by OpenGL software. Let a point in side of trapezoidal be P_0 and its coordinate in eye coordinate system be (x_0, y_0, z_0) . We try to find 6 planar equations for the trapezoidal with its value at P_0 are positive. The equation of each plane S_n is determined by any arbitrary three corner points of the corresponding surface of the trapezoidal. Let (a_n, b_n, c_n, d_n) be the parameter vector of each surface. Then the equation of

each surface can be expressed as follow.

$$S_n(x, y, z) = a_n x + b_n y + c_n z + d_n = 0, n = 1 \dots 6 \quad (4)$$

Then we propose an algorithm whether a point is inside of trapezoidal or not as a theorem 1.

Theorem 1: If all the values of $S_n(x, y, z)$ at a point P are positive, then the point P is located inside of trapezoidal.

2.3 Projective Mapping Function

We propose a projective mapping as a theorem as follow under the assumption that a point is inside of trapezoidal region we are interested in. In here, we only consider a object which lie in side of the trapezoidal characterized by 4 vertex which denote the virtual LCD screen in the eye coordinate system and scale factor k expressed in equation (3).

Theorem 2: For a given 4 vertex P_1, P_2, P_3, P_4 with the scale factor k , the projective point P'_e of the point P_e which is located in the corresponding trapezoidal can be expressed as follow.

$$p'_e = P_{roj} p_e \quad (5)$$

Let P'_s is the corresponding coordinate of P'_e with respect to the coordinate system of virtual LCD screen. Then we obtain

$$p'_s = P_{roj} p_s \quad (6)$$

Where

$$P_{roj} = T_{se}^{-1} P_{roj} T_{se} \quad (7)$$

And P_{roj} is the projective operation in the eye coordinate system with respect to trapezoidal region.

Proof: From the definition of T_{se} expressed in equation (2), we have

$$\begin{aligned} p'_s &= T_{se}^{-1} p'_e \\ &= T_{se}^{-1} P_{roj} p'_e \\ &= T_{se}^{-1} P_{roj} T_{se} p_e \end{aligned} \quad (8)$$

Therefore, we prove equation (7).

3. Experimental Environment

We build the stereo camera system using advanced embedded system based on the ARM1176 MPU core and we use MS WinCE as a operating system for developing camera device driver and application program based on USB OTG for sending image data file to the host computer system that is operating under the Windows XP and we use the OpenGL as a software tool and implement the 3D stereo image on the circular polarized 17 inch LCD Screen. The overall block diagram of camera capture system implemented in our laboratory is shown in figure 4.

We use the MT9M112 CMOS camera module as key element for stereo image capture system. The main feature of the CMOS camera module is that the depth of field (DOF) is enough to have all the objects in the

scene to be in focus when the objects are located 20cm far from the CMOS camera. We try to obtain the geometric parameters and the relations between camera capture system and 3D circular polarized LCD display system to make comfortable 3D real time video play. We assume that only one person is watching at center and 50cm in front of LCD which is 33.5cm width and 27.0cm height. We also assume the distance of the left eye and right eye of a person is 6.0cm. In our camera capture system, we set the distance between the right and left camera as 6cm. The following table I show the dynamic range of the view angle of MT9M112 which we find by analyzing the standard captured image set in our laboratory. Its dynamic range is enough to capture the left and right camera images for synthesizing its stereo 3D image on the 3D circular polarized LCD by using image processing and OpenGL programming.

Table 1. Dynamic range of camera module (unit: degree)

θ_{LM}	θ_{RM}	θ_{MU}	θ_{MD}	θ_{LU}	θ_{RD}
-27.3	27.3	22.1	-22.1	33.3	33.3

3.1 Calibration of Pinhole Camera

In this paper, we briefly discuss about the calibration of pinhole under the assumption that the eye mechanism of human and the camera module we use in experiment are very similar to the pinhole camera^[10]. The geometric relation among world coordinate system, camera coordinate system and pixel image coordinate system are shown in figure 5.

Usually the focal length f_x and f_y of real camera module are not exactly the same and has an asymmetric optical property which is expressed by parameter γ between x and y axis in camera coordinate system. Also there exist a little planar offset(x_0, y_0) in pixel image coordinate system because of the misalignment between camera coordinate system and pixel image coordinate system. Let the 2D position in pixel image coordinate system and the 3D position in world coordinate system be $(u \ v \ 1)^T$ and $(x_w \ y_w \ z_w \ 1)^T$ by using notation of Homogeneous coordinate. Then we obtain following

Fig. 4. Block Diagram of the Environment

Fig. 5. Geometric Parameters of Pinhole Camera

equation.

$$\begin{pmatrix} uz_c \\ vz_c \\ z_c \end{pmatrix} = A_{cp} A_{wc} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \quad (9)$$

Where

$$A_{cp} = \begin{pmatrix} f_x & \gamma & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}, \quad A_{wc} = (R:T) \quad (10)$$

And R and T are 4×4 rotational matrix and 3×1 translational vector between world coordinate system and camera coordinate system. The calibration problem in pinhole camera is to find the parameters for matrix A_{cp} and A_{wc} . The number of parameters A_{cp} and A_{wc} are value is 5 and 6 respectively. Therefore the calibration problem in pinhole camera is to find or estimate these 11 parameters. In general case, CMOS camera module is designed and manufactured so that the values of f_x equals to f_y and, x_0 and y_0 should be near to zero. In this paper, we assume that the values of f_x equals to f_y and x_0 and y_0 are zero. In this case, the calibration problem is to find and calculate f_x and the 6 parameters of A_{wc} . Therefore, we try to control a camera system such that

$$\begin{aligned} A_{wc}^L &= \begin{pmatrix} 1 & 0 & 0 & +3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\ A_{wc}^R &= \begin{pmatrix} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{aligned} \quad (11)$$

Where A_{wc}^L and A_{wc}^R denote the A_{wc} for left and right camera module and 3 means 3cm in above matrices.

3.2 Calibration Problem

In this section we discuss how to calibrate the stereo camera system for measuring the depth information of the scene. We make mechanical system as shown in figure 6 which has two identical patterns.

We also develop a control algorithm to align the two left and right camera capture system by measuring the offset distances. We define the offset distances as follows where $d_1(t)$, $d_2(t)$ are x component and y component of the offsets and $d_3(t)$ is the length difference of upper and lower line L_T and L_B and $d_4(t)$ is the length difference of left and right line L_L and L_R of the calibration pattern as shown in figure 7. We propose the following simple control law for position command vector $P_c(t)$. In real experimental system, $P_c(t)$ is the number of the counter clockwise clock in step motor system and increment of duty cycle in the DC servo system.

$$p_c(t) = K_p d(t) \quad (12)$$

Fig. 6. Mechanical System for Calibration

Fig. 7. Calibration Pattern

Where K_p is the 4×4 diagonal matrix and

$$d(t) = [d_1(t) d_2(t) d_3(t) d_4(t)]^T \quad (13)$$

The following table II is the tuning parameters of K_p for our camera capture system. We use two DC servo system for up-down and left-right rotating motions and two step motor system for up-down and left-right translational motions for each right and left camera capture system.

Table 2. Parameter gain for electric servo

K_{p1}	K_{p2}	K_{p3}	K_{p4}
10.0	10.0	20.0	20.0

3.3 Zooming and Alignment

We try to obtain and use 9 images from each left and right camera capture system for synthesizing fusion 3D image for the more comfortable 3D depth feeling. The geometric view of each segmentation 9 images can be shown in figure 8. Each image is a captured projected image of left or right camera which is adjusted optimally with alignment and focusing of corresponding segmentation. In later, these 9 segmented images are synthesized for making one whole image. By considering the dynamic range of the rotation of human eye shown in table III, we choose the following pairs of dynamic range of each image shown in table IV and we developed a control algorithm for finding the exact optimal capturing angles based on the sharpness of stereo images for each segmentation images and record the angle and distance offset parameters for rendering

Fig. 8. Segmentation of Image

Table 3. Angle parameters of human eye (unit: degree)

Segmentation	θ_{L1}	θ_{L2}	θ_{R1}	θ_{R2}
LT	11.7	11.1	18.1	10.8
CT	-3.4	11.3	3.4	11.2
RT	-18.1	10.8	-11	11.1
LM	11.7	0	18.1	0
C	-3.4	0	3.4	0
RM	-18.1	0	-11	0
LB	11.7	-11	18.1	-10.8
CB	-3.4	-11	3.4	-11.3
RB	-18.1	-10.8	-11.7	-11.1

Table 4. Dynamic range of segmentation (unit: degree)

Segmentation	θ_{L1}	θ_{L2}	θ_{R1}	θ_{R2}
LT	(10,13)	(10,13)	(17,20)	(9,12)
CT	(-5,-2)	(+10,+13)	(2,5)	(10,13)
RT	(-20,-17)	(9,12)	(-13,-10)	(10,13)
LM	(10,13)	(-2,+2)	(17,20)	(-2,+2)
C	(-5,-2)	(-2,+2)	(+2,+5)	(-2,+2)
RM	(-20,-17)	(-2,+2)	(-13,-10)	(-2,+2)
LB	(10,13)	(-13,-10)	(17,20)	(-12,-9)
CB	(-5,-2)	(-13,-10)	(10,13)	(-13,-10)
RB	(-20,-17)	(-12,-9)	(-13,-10)	(-13,-10)

later. The angle parameters of human eye can be expressed as follows under the assumption that human right and left eyes are located at $(-3,0,0)$, $(3,0,0)$ respectively.

$$\begin{aligned} \theta_{R1} &= a \tan 2(x+3, z) \\ \theta_{R2} &= a \tan 2(y, \sqrt{(x+3)^2 + z^2}) \\ \theta_{L1} &= a \tan 2(x-3, z) \\ \theta_{L2} &= a \tan 2(y, \sqrt{(x-3)^2 + z^2}) \end{aligned} \quad (14)$$

Segmentation symbols of each images are left-top(LT), center-top(CT), right-top(RT), left-middle(LM), center (C), right-middle(RM), left-bottom(LB), center-bottom(CB), right-bottom(RB) respectively.

In the above table, θ_{L1} , θ_{L2} , θ_{R1} , θ_{R2} denote the left-right, up-down rotating angle of the left and right camera capture system respectively. In this paper, we assume that a person who watches the synthesized LCD screen does not move his neck and just make rolling his eyes only. We try to propose a new image processing in the next paper which should be considered when a person move his neck left/right and up/down. This case is usually happened when the size of LCD is large compared with the distance between a person and LCD screen. The measure of the alignment is calculated by the weighting sum of the average m and deviation σ of the left and right image around the region of center and we consider 21×21 pixels for calculating these parameters. Before doing above process for calibration of the camera module that has the zooming functioning, we also consider to find the method to correct for the different camera characteristics by performing a pre-processing step for bias-gain or histogram equalization^[3].

$$\begin{aligned} m &= \sum_{i=-10}^{10} \sum_{j=-10}^{10} (p_L(i, j) - p_R(i, j)) / 121 \\ \sigma^2 &= \sum_{i=-10}^{10} \sum_{j=-10}^{10} ((p_L(i, j) - p_R(i, j)) - m)^2 / 121 \end{aligned} \quad (15)$$

We define the measure of alignment l as

$$l = \lambda m + (1 - \lambda) \sigma \quad (16)$$

Where $0 \leq \lambda \leq 1$.

Also we define the measure of sharpness as

$$s = \sum_{i=-10}^{10} \sum_{j=-10}^{10} (|\Delta_x p(i, j)| + |\Delta_y p(i, j)|) \quad (17)$$

Where

$$\begin{aligned} \Delta_x p(i, j) &= p(i+1, j) - p(i, j) \\ \Delta_y p(i, j) &= p(i, j+1) - p(i, j) \end{aligned} \quad (18)$$

For the left or right image around the region of center and we consider 21×21 pixels. The first problem is the focusing problem and this problem is to find optimal zooming f_z for any chosen angle θ_{L1} , θ_{L2} within the dynamic range given in table IV that maximize the measure of sharpness of left or right image and we choose left image first in this paper. The next problem is to find the optimal zooming f_z that maximize the measure of sharpness of the right image and at the same time to find the optimal angles θ_{R1} , θ_{R2} within the dynamic range shown in table IV that minimize the measure of alignment. Because of the hardware difficulty, we can't implement the hardware of the zooming mechanism for the time being. Therefore we just comment on and don't consider the measure of the sharpness in the signal processing.

4. Fusion Based On Interpolation

In this subsection, we propose simple fusion methods that make a whole stereo 3D image with the 9 segmented/captured images from the left and right camera module. The simple method is to crop each segmented image with the same size, such as 1/9 of whole LCD image area, based on the angle data which are acquired when the corresponding segmented image is captured. Under the assumption that the angle data are the same as those of eyes of person who watch the real LCD screen, we just union the same pixel size of each segmentation to make a whole image for left and right stereo images as follows.

$$I_w = \bigcup_{i=1, \dots, 9} \text{Crop}(I_{\text{seg}, i}) \quad (19)$$

Where $I_{\text{seg}, i}$ denotes each captured segmented image and I_w denotes whole synthesized image. For example, let assume that the LCD Screen size is 1280×1024 .

Then we crop each segmented image into $(1280/3) \times (1024/3)$ size image around its center and then combine these 9 cropped images into a whole 1280×1024 image. We do this operation twice for obtaining left and right whole image.

If the above assumption concerning the angle data is not satisfied, we should do adopt a filtering and smoothing signal processing with the captured images. In this case, we first calculate the geometric position vector $p_v = (x_v, y_v, z_v)$ of center pixel of each captured segmented image in virtual LCD screen based on the capturing angle data and also calculate the corresponding position $p_L = (x_L, y_L)$ in the real LCD screen. Then we crop the image as shown in figure 9 by considering this position P_L for each of corresponding segmentation. In next future work, we try to find more advanced algorithm for interpolation of segmentation images for fusion.

Fig. 9. Cropping and Union for Fusion

5. Simulation

As the experimental environment is now being constructed and not finished yet, we simulate our proposed algorithm based on the virtual objects made by programming with OpenGL software tool. We synthesize a 3D stereo image from the 9 segments of image

captured by virtual left and right camera which is separated 6cm with each other. This separation length is the same length of distance between two left and right eyes of average person. We choose several cubes as objects located around 1m behind of virtual LCD screen. In this simulation, we assume that a person who watches the synthesized image does not move and rotate his neck and moves only his eye balls. We use the angle parameters of human eye shown in table III for obtaining 9 segmentations of image. We also suggest a camera model for general CMOS camera. Camera model of general CMOS camera can be approximated as an image sum of several/many ideal pinhole cameras.

Let I_{CMOS} and be the image of CMOS camera and let $I_{pin}(x, y, z)$ be the image of pinhole camera whose aperture is located at (x, y, z) from the center. We can approximate I_{CMOS} as follows by using several images obtained from the ideal pinhole cameras as shown in figure 10.

$$I_{CMOS} = \frac{1}{N} \sum_{i=1, N} I_{pin}(x_i, y_i, z_i) \quad (20)$$

For the simplicity of simulation, we assume that all the ideal pinhole cameras have the same focal length $f_{x0} = f_{y0}$ and share the same image plane and have the same aperture plane, that is, z components of aperture are the

Fig. 10. CMOS Camera Model

same. We obtain camera parameters of each pinhole camera as follows.

$$A_{cp,i} = \begin{pmatrix} f_x & \gamma & x_{i0} \\ 0 & f_{y0} & y_{i0} \\ 0 & 0 & 1 \end{pmatrix}, \quad A_{wc,i} = \begin{pmatrix} 1 & 0 & 0 & -x_{i0} \\ 0 & 1 & 0 & -y_{i0} \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (21)$$

From the above equation, we can see the above transformation is not linear with respect to aperture offset x_{i0} and y_{i0} . This kind of nonlinear effect makes it difficult to find a good preprocessing method for the captured CMOS camera image. In this paper, we set N as 5 for the simplicity of simulation and (x_i, y_i) , $i = 1, \dots, 5$ be center point and 4 corner points of rhombus around the center of aperture. We set the distance of a vertex of rhombus d from the center as 0.1[mm] in this paper. We set the focal length f_{x0} as 5[mm]. Therefore the percentage of the aperture offset compared with focal length is about 2 % in this simulation.

We use the OpenGL real time graphics as a software tool. The test scene consists of same size of 19 cubes and 2 axes. The width, length and height of cube are 8cm each and surrounding 8 cubes are located 12cm far from the centered cube on each axis. We assume that distance between 3D LCD screen and human viewer is 50cm and the distance between the centered cube and human viewer is about 100cm. So we set the near and far clipping planes into the z-axis as follows.

$$z_n = -50cm \quad z_f = -200cm$$

Next we move the cubes 100cm in the direction of negative z-axis by using `glTranslatef()` API function of OpenGL. Figure 11 shows the image captured by using one ideal pinhole camera and the image synthesized by

using 5 ideal pin hole camera model proposed above when the distance of a vertex of rhombus d from the center are 10.0, 5.0, 1.0[mm] respectively. We also calculate the difference of the image in region of interest R_{ROI} by using the following equation.

$$I_d = \sum_{(i,j) \in R_{ROI}} |I_{d,RGB}(i,j)| \quad (22)$$

Where, $I_{d,RGB}(i,j)$ means the absolute sum of difference of RGB color components in pixel $p(i,j)$ between two images. We select R_{ROI} one of segmentation (about 420×330 pixel size window patch) around the center of synthesized images. The simulation results in this region of interest are shown in table IV. From the data in table IV, we can see that the proposed CMOS camera module can be used for synthesizing general CMOS camera whose characteristics are very similar to the ideal pinhole camera. From the simulation results, we can simulate the general CMOS camera by choosing appropriate number of pinhole camera model and selecting appropriate offset value. We found that 1.0[mm] is appropriate for offset value in simulating CMOS camera model. Also another important characteristic of general CMOS camera is that the resolution of image with respect to the solid angle is nonlinear. Its image resolution at solid angle zero is better and bigger than at any other angle. This is the same reason why we should control and focus our eye on an object to see more clearly. When our eye is focused at an object, we can get larger image by rotating eye with the same amount of linear angle. It means that the sensitivity of our eye with respect to rotating angle is relatively high when it is focused at an object. Therefore the disparity of the left and right captured images depends on the alignment of the left and right camera module with respect to an object which we are interested in. The disparity d_L of the left and right image for some region of interest can be decided such that the two left and right images match as closely as possible as follow.

Table 5. Simulation results of cmos camera model

offset[mm]	I_d	offset[mm]	I_d
10	2,502,362	2	645,503
5	1,365,189	1	300,149

Fig. 11. Comparisons of CMOS Camera Model

$$I_L(i, j) \approx I_R(i - d_L, j) \text{ for } (i, j) \in R_{ROI} \quad (23)$$

Where $I_L(i, j)$ and $I_R(i, j)$ is the left and right captured images respectively. Figure 11-a is an image captured when camera module is well aligned with the 3D object and figure 11-b is an image captured when camera module is a little misaligned. The values of d_L which depends on the alignment between camera module and object is very important for the 3D depth synthesizing. This idea is the main key point of this paper. Therefore multi-focusing image capturing system can improve the depth quality in 3D image processing.

6. Conclusion

In this paper, we propose a new fusion method to combine a whole stereo image from the 9 captured images. We also introduce and discuss a projective mapping which plays an important role for handling signal processing. We suggest a control algorithm based on the measure of alignment and sharpness which is needed for calibrating camera module before capturing the image. It is needed to have good knowledge concerning the geometric parameters not only in capturing a left and right image but also in synthesizing stereo image for designing of a good stereo vision system. We introduce mathematical preliminary which is

needed to analyze and synthesize stereo image. We do simulate by using OpenGL software and get some results that verify the validity of our proposed algorithm. We are now building real experimental environment based on ARM11 core embedded system with WinCE operating system but it is not completely finished yet. We will try and hope to make another proposal for stereo capturing system with real experimental data in near future.

References

- [1] C. Zhang and T. Chen, "View-dependent Non-uniform Sampling for Image-based Rendering", ICIIP 2004
- [2] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", International Journal of Computer Vision, 47(1/2/3).pp.7-42, April-June 2002.
- [3] I.J.Cox, S.Roy, and S.L.Hingorani, "Dynamic histogram warping of image pairs for constant image brightness," IEEE International Conference on Image Processing, Vol.2, pp. 366-369, 1995.
- [4] Fleet, D.J. and Weiss, Y., "Optical flow estimation," in "Mathematical models for Computer Vision: The Handbook," N. Paragios, Y. Chen, and O. Faugeras (editors.), chapter 15, Springer, 2005
- [5] J.L. Barron, D.J.Fleet, and S.S. Beauchemin, "Performance of optical flow techniques," International Journal of Computer Vision, 12(1), PP. 43-77, 1994
- [6] N. Holliman, "Smoothing Region Boundaries in Variable Depth Mapping for Real Time Stereoscopic Images," *Stereoscopic Displays and Virtual reality System XII*, (2005), vol.5664A of *Proceedings of SPIE-IST Symposium on Electronic Imaging*.
- [7] N. Holliman, "3D Display Systems", Handbook of Opto-electronics, IOP Press, Spring, 2005.
- [8] N. Holliman, "Mapping Perceived Depth to Regions of Interest in Stereoscopic Images," *Stereoscopic Displays and Virtual reality System*

XI, Proceedings of SPIE 5291, 2004.

- [9] S.Kim and W.ham, "A New Algorithm for Depth Perception in 3D Screen and Its Implementation," *2007 International Symposium on Information Technology Convergence*, Nov, 2007.
- [10] T.Dang, C. Hoffmann, and C.Stiller, "Self-calibration for Active Automotive Stereo Vision," in *IEEE Intelligent Vehicles Symposium*, pp.364-369, 2006.

권혁재

2011 전북대학교 전자공학과
(공학사)

2011~현재 전북대학교 전자공
학과 석박사통합과정

관심분야: 입체영상, 로봇제어,
임베디드 시스템

함운철

1979 서울대학교 전자공학과
(공학사)

1982 서울대학교 전자공학과
(공학석사)

1988 서울대학교 전자공학과
(공학박사)

1988~현재 전북대학교 전자공학과 교수

관심분야: 비선형제어, 적응제어, 로봇제어, 임베디드 시스템

투멘자르갈 엔크바타르

2001 몽골과학기술대학교 전
자공학과(공학사)

2003 몽골과학기술대학교 전
자공학과(공학석사)

2009~현재 전북대학교 전자
공학과 박사과정

관심분야: 디지털 회로설계, 임베디드 시스템