Distance measurement using stereo camera and 3D implementation with 3D display devices

Hyok Song[1], Jinwoo Bae[2], Jongsoo Choi[3], Byeongho Choi[1]
[1]Korea Electronics Technology Institute, 68 Yatap, Bundang, Sungnam, Korea
TEL:82-31-789-7451, e-mail: hsong@keti.re.kr. bhchoi@keti.re.kr
[2]Korea Invention Promotion Association, 647-9 Yeoksam, Kangnam, Seoul, Korea
e-mail: bjw8751@kipa.org
[3]Chung-ang University, 221 Heukseok, Dongjak, Seoul, Korea
e-mail: jschoi@cau.ac.kr

Keywords : *3D, IVR, Depth*

Abstract

Depth data for is very important data for 3D display. Disparity and depth data makes users to feel 3D effect. We used stereo camera to measure depth and made fast algorithm to get in real time. This vision system can be substituted for expensive laser system.

1. Introduction

Technologies from computer graphics, computer vision, multimedia communication is growing new applications such as multi-view video coding which will follow 2D video technology from black and white TV, color TV, and HDTV. Multi-view TV or 3D TV technology allows users to choose viewpoint freely. Depth data or disparity data which is the additional data from multi-view vision processing is used for integration natural video from cameras and CG from computer vision.

Correspondence estimation is one of the fundamental challenges in computer vision lying in the core of many problems from stereo vision system. The goal of stereo matching is to estimate the depth information by finding correspondence pairs from a pair of stereo images. Even though it is a classical problem that has been extensively studied for several decades, it is still one of the most challenging problems in the field of computer vision and attracts a large amount of attention.

They are producing many types of 3D displays and the quality of video is getting better. Therefore the quality of contents is emphasized. 3D contents can be captured using stereo cameras or multi-view cameras. In the process of 3D conversion depth or disparity data is important information. By various display devices, disparity representation method is different. Therefore correct disparity/depth acquisition is emphasized in the thesis. And in real world video data frame is 30 frames in a second. Fast algorithm is proposed.

The rest of this paper is organized as follows. We introduce 3DTV system in section 2. Algorithms are shown in section 3. The experimental result is addressed in section 4. Future works are followed in section 5.

2. 3DTV system and disparity

3D system can have three parts. The first one is the capture system. Stereo camera, parallel multi-view camera, square multi-view camera, and depth cameras are multi-view camera systems. Captured images from camera systems are coded, transferred and decoded. And then the stream is displayed on 3D display devices.

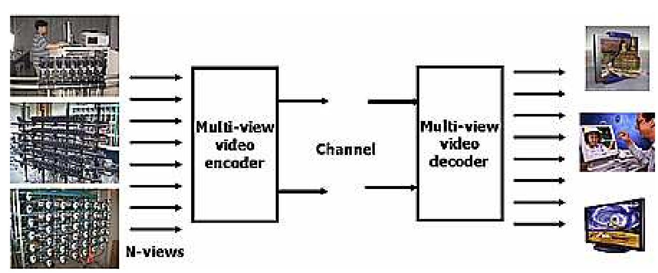
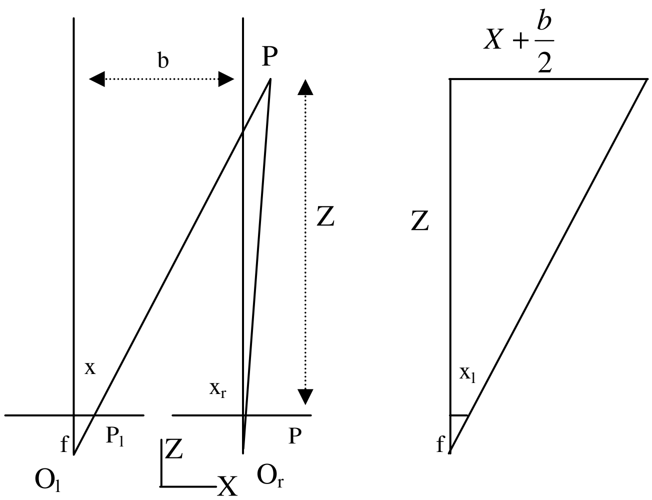


Fig. 1. Structure of 3DTV system

Figure 1 shows the entire system of 3DTV. KETI designed and made 8-camera system which is parallel 1 dimensional structure. CMU used 48 cameras of 2 dimensional structure[1]. Nagoya University used 100 cameras of 1D and 2D. Multi-camera system requires calibration and rectification from captured images. It makes easier to code video streams and gives more realistic 3D effect to TV users. Multi-view video codec is under standardization based on H.264 codec in MPEG. In decoder stage, multi-view video is decoded and modified to display on various 3D displays.

3. Disparity searching algorithm

A camera is a mapping between the 3D world and a 2D image. The most specialized and simplest camera model is the basic pinhole camera model and stereo camera model is shown in Figure 2. The model is principally designed for CCD like sensors. p is the image point of world point P in the image plane. x_r and x_l are the x-coordinate of p in each image plane.



(a) (b)

Fig 2. Stereo system and depth by triangulation

In Fig 2. (b), there is the relation between big triangle and small triangle. Equations below shows the relation

$$x_l \times Z = f \times (X + \frac{b}{2}), \frac{x_l}{f} = \frac{X + \frac{b}{2}}{Z} \quad (1)$$

$$x_r \times Z = f \times (X + \frac{b}{2}), \frac{x_r}{f} = \frac{X + \frac{b}{2}}{Z} \quad (2)$$

$$\frac{x_l - x_r}{f} = \frac{b}{Z}, Z = \frac{bf}{x_l - x_r} \quad (3)$$

We can solve equation and get x_l-x_r using triangulation method. x_l-x_r is disparity.

We propose a new algorithm that improves speed up using predictors which are disparity vectors already found on PMVFAST for finding motion vectors[2]. Predictors in adjacent blocks which are up, up-right and left blocks are candidates of disparity vector in current block. We used median vector of candidates as a predictor. Many disparity vectors of blocks in an object in the image have similar direction and distance. This homogeneity character reduces calculation. However, there is noise from camera and luminance illumination between stereo cameras. It makes cost higher and can cause error in finding disparity vector. It is "Trade-off relation". If we want find correct disparity, the cost will be higher. This algorithm finds appropriate algorithm.

We define a threshold T1 which is used in condition 1. Median disparity vector of adjacent vectors is used for disparity vector of current vector, and if the cost is smaller than T1, the vector value is the disparity vector.

In second, The vectors of adjacent blocks are the same, and if the cost of the disparity is smaller than the second threshold T2, it is the disparity vector.

```
IF( SAD = (MedianDV)<T1)                              //Condition 1.
   {
   DV=MedianDV;
   Terminate;
   }
ELSE IF( (DV_L==DV_U==DV_UR) < T2 )                   //Condition 2.
   {
   DV=MedianDV;
   Terminate;
   }
ELSE IF( (SAD<min{SAD(DV_L), SAD(DV_U), SAD(DV_UR)}) && (SAD<T3) )   //Condition 3.
   {
   DV=MedianDV;
   Terminate;
   }
ELSE                                                  //Condition 4.
   {
   Full search ();
   }
```

Fig 3. Pseudo code of predictive disparity estimation

The next process is that we use the vector of the minimum cost in adjacent blocks. If the cost is smaller than T3, it is disparity. And then the next process is full search of quad block. In disparity searching using stereo images, errors happen in occluded area. Blocks in the area usually don't match in the conditions. Moreover the area doesn't exist in the other image or partially exist. In occluded area interpolation method or adjacent disparity method is used.

3D display devices have various 3D effect methods like parallax barrier, lenticular, polarized LCD, hologram, etc. And there are many kind of display devices in view number. For example, stereo, 8 views, 2D 16views, etc. In that case, intermediate views are reconstructed.

Intermediate view reconstruction is a big issue for 3D researchers. In view reconstruction for 3D display devices, we used view conversion using cubic convolution filter. We have stereo images and disparity information. Disparity can be converted to depth. Calculated depth data is varied by the position of TV users.

Interpolation is very useful in image processing. Scale conversion, rotation, translation are some usage of in geometric process[3]. In 3D intermediate image reconstruction, hole and overlapping happen frequently. Hole means no data in a pixel and overlap is duplicating two data in a pixel. Cubic convolution has static coefficient in an image. The equation is shown blow

$$\hat{f}(x) = \sum_k c_k \beta(x - x_k)$$

$$\beta(x) = \begin{Bmatrix} (a+2)|x|^3 - (a+3)|x|^2 + 1, x \le |x| \le 1 \\ a|x|^3 - 5a|x|^2 + 8a|x| - 4a, 1 \le |x| \le 2 \end{Bmatrix}$$

On the point of x=1, $\beta(x)$ to be similar with sinc(x), a=-1[4][5].

$$\begin{aligned} \hat{f}(x) &= f(x_{k-1})(\alpha s^3 - 2\alpha s^2 + \alpha s) \\ &+ f(xk)((\alpha+2)s^3 - (3+\alpha)s^2 + 1) \\ &+ f(x_{k+1})(-(\alpha+2)s^3 + (2\alpha+3)s^2 - \alpha s \\ &+ f(x_{k+2})(-\alpha s^3 + \alpha s^2) \end{aligned}$$

This equation can be written as

$$\begin{aligned} f(x) &= \alpha_k (f(x_{k-1}) - f(x_{k+1}))A(s) \\ &+ \alpha_k (f(x_k) - f(x_{k+2}))B(s) \\ &+ (f(x_k) - f(x_{k+1}))C(s) \\ &+ f(x_k) \end{aligned}$$

Where

$$A(s) = (s^3 - 2s^2 + s)$$
$$B(s) = (s^3 - s^2)$$
$$C(s) = (2s^3 - 3s^2)$$

Cubic convolution filter makes analog function of the image. The function has the same value in sampled positions with the original image data. In resampling process, disparity can be made.

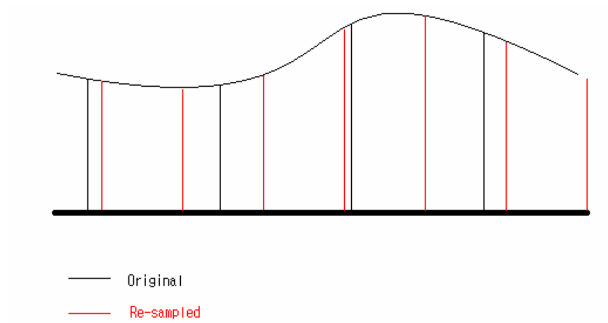


Fig 4. Resampling of a line in an image

Figure 4. shows resampling of a line in an image. If re-sampled position is the left side of the original sampled position, re-sampled image is shifted right. A shifted image is a reconstructed image which has disparity. Resampling can be done in the whole image, in a line and in the region of interest.

We normally recommend the use of 1.0 (single) line spacing. However, when typing complicated mathematical text it is important to increase the space between text lines in order to prevent sub- and super-script fonts overlapping one another and making your printed matter illegible.
Tables and illustrations should be originals or sharp images. They should be arranged throughout the text

4. Results

Depth data is acquired using a stereo camera and the specification is shown below.

- CCD : Stereo camera Sony progressive scan
- Baseline : 12Cm
- Focal length : 6mm

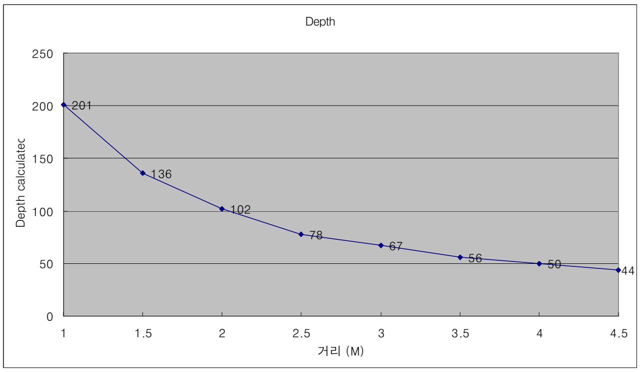


Fig 5. Disparity value

Figure 5. shows disparity value by the distance.

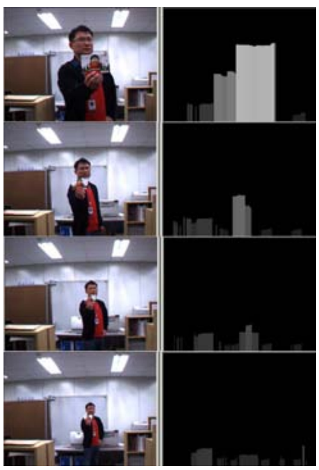


Fig 7. Original images and depth images

We tested in an office environment. Tested point is the center of the image. Fig 7. shows the original right images and depth images at several distances. We tested in distances of 1m, 1.5m, 2m, 2.5m, 3m, 3.5m, 4m and 4.5m.

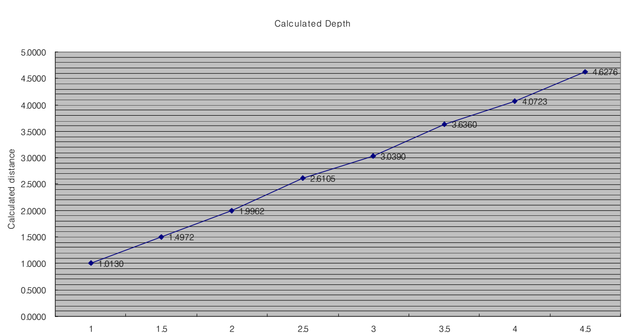


Fig 8. Real distance and measured distance.

We compared depth data and real distance. The error was 1% to 4.42%. We adopted this data to real 3D display device to reconstruct views.

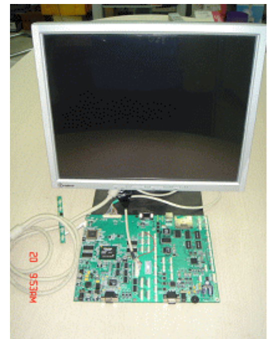


Fig 9. 3D display device

In the hardware board, images are reconstructed in real time. We used Actel FPGA and the language is VHDL.

5. Future works

The error ratio was measured in the center point. Error ratios in every pixel point give reliability.

For more correct depth data, bi-directional searching, occluded area processing and sub-pixel interpolation is under progress.

6. References

1. Ted Square, Tsuhan Chen, „CMU ultimate eye multi-camera data“, *ISO/IEC JTC 1/SC 29/WG 11, Document M12003*, Busan, Korea, April (2005)
2. A.M. Tourapis, O.C.Au, and M,L,Liou,“Predictive motion Vector Field Adaptive Search technique Enhancing Block Based Motion Estimation“, *Proc. SPIE Visual Commun. Image Process*, San Jose, CA, Jan. (2001)
3. Randy Crane, *A simplified approach to image processing* Prentice hall, (1997)
4. S.S Rifman, *Digital rectification of ERTS Multispectral imager*, Proc. Symposium significant results obtained from ERTS-1 (NASA SP-327), I, Sec. B, pp. 1131-1142 (1973)
5. R. Bernstein, *Digital Image Processing of Earth Observation Sensor Data*, IBM, J.Res. Devel, Vol 20, pp.40-57 (1976)

– This work was supported by the IT R&D program of MIC/IITA [2007-S-050-01, Audio-visual Signal Processing SoC for Intelligent Robot]