Margins of Allowable Stereoscopic Geometry Errors

Jongyoung Kim†, Namgyu Kim††, and Seiwoong Oh†

†Stereoscopic Imaging Research Center, Busan IT Industry Promotion Agency, Busan, Korea
††Department of Game Engineering, Dong-Eui University, Busan, Korea

Summary
Preparation for stereoscopic 3D image shooting is tedious and time-consuming. Unsuccessful preparational setup prior to actual stereoscopic shooting can result in some "bad" disparities such as vertical, scale, rotational differences in left and right images. The ideal case would be where we have zero disparities except the horizontal one, which is expensive and time-consuming to realize. In this paper, we investigated the margins of allowable stereoscopic geometry errors. In our experiments, subjects were asked to adjust the three kinds of disparities (vertical, scale, rotation), looking at images with "bad" disparities, until they think it is comfortable to watch. We made Gaussian distribution over the experimental data and performed parameter estimation, including Maximum Likelihood Estimation and EM algorithm. From experimental result, rotational disparity showed the narrowest "allowable" margin, while scale and vertical disparities showed some variabilities.

Key words:
Geometry error, Maximum likelihood estimator, Gaussian mixture model, EM algorithm, Stereoscopic errors

1. Introduction

Stereoscopic 3D is a result of attempting to creating or controlling the "feeling of depth" in an image by means of binocular disparity. Binocular disparity is natural and inevitable since human eyes are approximately 6.5cm apart, which is called interocular distance, and the images for left and right eyes are slightly different. The fusion process within human brain plus with binocular disparity create the feeling of depth. When shooting a 3D image, two cameras, which are separated at roughly human eye interocular distance, are used to take separate images from slightly different angles to get binocular disparity. This separation is horizontal and produces a binocular disparity.

When shooting stereoscopic 3D images, some errors can happen, which are main causes of visual discomfort when watching stereoscopic 3D images. These errors include vertical misalignment, rotational error, scale mismatch, color mismatch, reversed left and right images, retinal rivalry, ghosting, focus mismatch, keystone and edge violation. Specifically, vertical alignment, rotational error and scale mismatch between left and right images are of main concern in this paper. Vertical misalignment occurs when the vertical alignment of left and right images are different, thus human eyes have to move vertically to fuse the images, which is very uncomfortable for our eyes. Rotational error can happen usually in combinations of pitch, yaw and roll elements. One of primary reasons for scale mismatch between left and right images is that two cameras are at different focal lengths.

In [1], they showed how distortions caused by camera convergence or toed-in affects the ability to fuse and perceive stereoscopic images. In [2], they suggested comfortable depth budget as a form of guideline. In [3], they tried to determine the discomfort ranges for the kinds of natural image that people are likely to take with 3D cameras rather than the artificial line and dot stimuli typically used for laboratory studies. They assessed visual discomfort on a five-point scale for artificial misalignment disparities applied to a set of full-resolution images. They modeled the data with a second-order hyperbolic compression function incorporating a term for the basic discomfort of the 3D display in the absence of any misalignments through a Minkowski norm.

2. Backgrounds

2.1 Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is a method by which a set of parameters can be computed for a given probabilistic model and data set. MLE, originally developed by R. A. Fisher in his seminal paper [4], states that the desirable probability model is the one which renders the observed data "most likely" in terms of data generation viewpoint. In other words, the output of MLE is the point estimation of parameter(s) for which the likelihood of observed data set is at its maximum. Suppose we have a data set X of n independent and identically distributed (iid) observations $X_1, \dots X_n$ which is assumed to be generated from an unknown probability density function $f(\cdot\,|\,\theta)$ where θ is a set of parameter(s) for the corresponding probability density. Generally, one makes an assumption about the functional form of the probability density function. For an independent

Manuscript received January 5, 2013
Manuscript revised January 20, 2013

and identically distributed data points $x_1, \ldots x_n$ and $\boldsymbol{\theta}$, the joint density is written as follows:

$$\begin{aligned} & f(x_1, \ldots, x_n \mid \theta) \\ & = f(x_1 \mid \theta) \times f(x_2 \mid \theta) \times \cdots \times f(x_n \mid \theta). \end{aligned}$$

By taking a perspective that the joint density is a function of $\boldsymbol{\theta}$, not of "fixed" observed data $x_1, \ldots x_n$, $f(x_1, \ldots, x_n \mid \theta)$ is called likelihood and written as follows:

$$\begin{aligned} & L(\theta \mid x_1, \ldots x_n) \\ & = f(x_1, \ldots x_n \mid \theta) \\ & = \prod_{i=1}^{n} f(x_i \mid \theta) \end{aligned}$$

Usually, it is often convenient to take logarithm of the likelihood function and it is called the log-likelihood:

$$\ln L(\theta \mid x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln f(x_i \mid \theta).$$

Maximum likelihood estimator is defined as:

$$\underset{\theta}{\arg\max} \sum_{i=1}^{n} \ln f(x_i \mid \theta).$$

For the normal distribution $N(\mu, \sigma^2)$, the probability density function is:

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right).$$

For a set of n independent and identically distributed normal random variables, the likelihood is given as;

$$\begin{aligned} & f(x_1, \ldots x_n \mid \mu, \sigma^2) \\ & = \prod_{i=1}^{n} f(x_i \mid \mu, \sigma^2) \\ & = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp\left(-\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2\sigma^2} \right) \end{aligned}$$

Since Gaussian distribution has two parameters, we need to maximize the likelihood with regard to two parameters:

$$\begin{aligned} & \frac{\partial}{\partial \mu} \log f(x_1, \ldots, x_n \mid \mu, \sigma^2) \\ & = \frac{\partial}{\partial \mu} \left(\log\left(\frac{1}{2\pi\sigma^2} \right)^{n/2} - \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2} \right) \\ & = 0 - \frac{-2n(\bar{x} - \mu)}{2\sigma^2} \end{aligned}$$

which is solved by

$$\mu = \sum_{i=1}^{n} x_i / n$$

and

$$\begin{aligned} & \frac{\partial}{\partial \sigma} \log f(x_1, \ldots, x_n \mid \mu, \sigma^2) \\ & = \frac{\partial}{\partial \mu} \left(\frac{n}{2} \log\left(\frac{1}{2\pi\sigma^2} \right) - \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2} \right) \\ & = -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{\sigma^3} \end{aligned}$$

which is solved by

$$\begin{aligned} \sigma^2 & = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} \\ & = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \end{aligned}$$

The logarithm is a continuous and strictly increasing, values maximizing the likelihood will maximize the logarithm.

2.2 Expectation Maximization (EM) Algorithm

EM algorithm is an iterative method for finding maximum likelihood estimates of parameters of concern[5]. In E-step, it updates a function of expectation of the log likelihood and in M-step, it updates parameters maximizing the function of expectation of the log likelihood found on the E-step. EM algorithm is useful when maximum likelihood estimates can't be solved analytically. Generally, when applying EM algorithm, one assumes the existence of latent variable in addition to unknown parameters and observed data.

A representative application of EM algorithm is density estimation problem. Given a set of N data points $X = (x_1, \ldots x_N)$ in D dimensions, one finds probability density f that is most likely to have generated the given points. For example, the probability density is form of Gaussian mixtures and it has following form:

$$f(x;\theta) = \sum_{k=1}^{K} \pi_k g(x; \mu_k, \sigma_k)$$

$$g(x;\mu_k,\sigma_k) = \frac{1}{\left(\sqrt{2\pi}\sigma_k\right)D} \exp\left(-\frac{1}{2}\left(\frac{||\,x-\mu_k\,||}{\sigma_k}\right)^2\right)$$

Where π_k is the mixing coefficient. EM iterates the following computations until convergence to a local maximum of the likelihood function:

- E-Step

$$p^{(i)}(k \mid n) = \frac{p_k^{(i)} g(x_n; \mu_k^{(i)}, \sigma_k^{(i)})}{\sum_{m=1}^{K} p_m^{(i)} g(x_n; \mu_m^{(i)}, \sigma_m^{(i)})}$$

- M-Step

$$\mu_k^{(i+1)} = \frac{\sum_{n=1}^{N} p^{(i)}(k \mid n) x_n}{\sum_{n=1}^{N} p^{(i)}(k \mid n)}$$

$$\sigma_k^{(i+1)} = \sqrt{\frac{1}{D} \frac{\sum_{n=1}^{N} p^{(i)}(k \mid n) \,|\, x_n - \mu_k^{(i+1)} \,|^2}{\sum_{n=1}^{N} p^{(i)}(k \mid n)}}$$

$$p_k^{(i+1)} = \frac{1}{N} \sum_{n=1}^{N} p^{(i)}(k \mid n)$$

3. Experiment

3.1 Procedure

The purpose of experiment was to find out the “allowable” margins of each of geometric errors (vertical misalignment, rotation error, scale mismatch) within which subjects didn’t feel visual fatigue. We used two dslr cameras (Canon 5D Mark III) to take left and right images. These images were saved as JPEG file, which were imported by Assimilate Scratch in order to convey left and right images to Sony MPE 200 in the form of SDI (Serial Digital Interface) signal. Leaving geometric values of right image unchanged, considerable amount of geometric errors were applied to left image, resulting in very uncomfortable viewing condition.

Fig. 1 SpaceMouse™ from 3DCONNEXION

SpaceMouse™ in figure 1 is a prodcut from 3DCONNEXION. It has jog-like dial on it. It was connected to Sony MPE 200 and when subjects turned the dial-like button on the device, Sony MPE 200 updated and displayed the left image according to changed geometric values (vertical alignment, rotational degree, zoom scale) in real time. Subjects were asked to stop when they think it was comfortable to watch the stereoscopic image.

Fig. 2 Test image

Figure 2 show the test image used in the experiment. Experimental data set was collected using Cel-Scope 3D stereoscopic analyzer. Cel-Scope 3D is a software providing stereoscopic 3D monitoring and depth budget analysis. When subjects stopped turning the jog-like dial, we recorded the geometric difference between left and right images. Specifically we recorded vertical misalignment in pixel, scale mismatch in percentage and rotational error in degrees.

Fig. 3 Analysis of disparities

3.2 Experimental Result

We collected data from experiment. The minimum values were all 0s for vertical, scale and rotational cases. The maximum values were 39 pixel, 6.1% and 2 degrees in cases of vertical, scale and rotational experiment.

The maximum likelihood estimates of mean and variance, under the assumption that the data points are i.i.d. sample from Gaussian distribution, is given in table 1. From the result, it can be said approximately that subjects showed more variability in adjusting vertical alignment, compared to rotational difference between left and right images.

Table 1. Maximum Likelihood Estimates

	mean	variance
vertical	12.7986	9.4583
scale	1.6536	1.4356
rotation	0.7071	0.43598

Under the assumption that the data points has been"generated" by 2-component Gaussian mixture, EM algorithm was used to estimate those parameters. The result is given in table 2. In the case of rotational difference data, interesting was that the maximum likelihood estimates of mean was identical with the result of analysis under 2-Gaussian mixture model assumption. The mixing coefficients were 0.5s and mean was same as 0.7071 for both components. In case of rotational experiment, experimental result was unimodal.

Table 2. Parameters for 2-Gaussian mixture model

	component 1 mean(weight)	component 2 mean(weight)
vertical in pixel	22.6210 (0.354912)	7.3945 (0.645088)
scale in percentage	1.0512 (0.764275)	3.6067 (0.235725)
rotation in degree	0.7071 (0.5)	0.7071 (0.5)

In the case of vertical aignment data, the preference seemed to be at somewhere near 7.3945 rather than near 22.6210, since the mixing weight was 0.645088 for the component 2. This can be said “compatible” with the maximum likelihood estimates which is 12.7986. However, it was not clear enough to conclude that subjects showed unimodal tendency in adjusting vertical alignment of left and right images. In the case of scale data, the preference seemed to be at somewhere near 1.0512 rather than at 3.6067, since the mixing weight was 0.764275 for the component 1. Like vertical alignment data, this can be said “compatible” with the maximum likelihood estimates which is 1.6536. Compared to vertical alignment experimnet, subjects showed relatively strong unimodality. The figure 4 shows the histogram of scale alignment data.

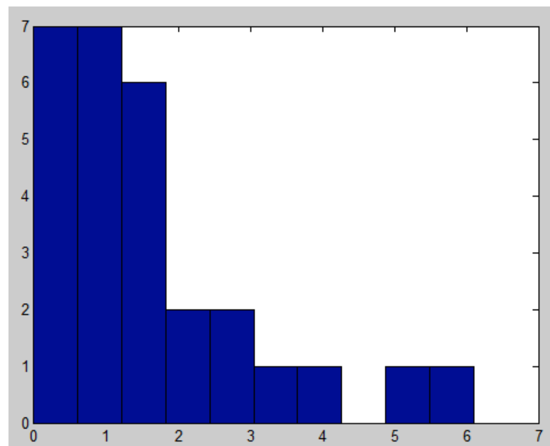

Fig. 4 Histogram of scale alignment data

4. Conclusion

Among some categories of stereoscopic errors, we chose so-called “geometric error” which includes vertical misalignment between left and right images, scale (or zoom) mismatch between left and right images and rotational error between left and right images for answering the question: what is the allowable margin or range of these errors in the context of visual comfort?

We made Gaussian assumption for the experimental data and performed statistical analysis. From our experiment, subjects showed narrower margin for the scale differences when compared to margin for the vertical aignments. In the case of rotational error, there’s no difference between one-component and two-component mixture analysis which can be interpreted that subjects showed unified preference over margin of rotational errors.

These experimental result can be used as a guideline for stereoscopic content production. Stereoscopic content producers can save time and effort for configuring the shooting environment. They manage to stay within margins of allowable stereoscopic errors.

Acknowledgment

This research is supported by Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Joint Research Project 2012.

References

[1] R. S. Allison, “Analysis of the influence of vertical disparities arising in toed-in stereoscopic cameras,” J. of Imaging Sci. Technol, 51(4), 317-327, 2007.
[2] 3DC Safety guidelines for dissemination of human-friendly 3D, 3D Consortium, Japan, 2011.
[3] C. W. Tyler, L. T. Likova, K. Atanassov, V. Ramachandra and S. Goma, ‘3D discomfort from vertical and torsional disparities in natural images,” Proc. SPIE 8291, 2012.
[4] R. A. Fisher, On the mathematical foundations of theoretical statistics, Philos. Trans. Roy. Soc. London Ser. A, 222, 309-368, 1922.
[5] A. P. Dempster, N. M. Laird, D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” Journal of the Royal Statistical Society, Series B, 39(1), 1-38, 1977.

Jongyoung Kim received the B.S. and M.S. degrees in Computer Science Department from Hanyang University, Seoul, Korea. Now he is pursuing his Ph.D. degree in Computer Science Department from Hanyang University, Seoul, Korea. He joined the Stereoscopic Imaging Research Center (SIRC) in 2011. He is a research engineer.

Namgyu Kim received the B.S. degree in Computer Science from KAIST, Daejeon, Korea, and earned the M.S. and Ph.D. degrees in Computer Engineering from POSTECH, Pohang, Korea. He joined Advanced Telecommunications Research Institute International, Kyoto, Japan and Korea Telecommunication research center, Daejeon, Korea. He is now a full-time lecturer in the department of game engineering, Dong-Eui University, Busan, Korea.

Seiwoong Oh received the B.S. and M.S. degrees in Electronic Engineering from Hanyang University, Seoul, Korea and earned the Ph.D. degree in Information Engineering from Osaka University, Osaka, Japan. He joined Electronics and Telecommunications Research Institute, Daejeon, Korea. He is now a professor in the department of game engineering, Dong-Eui University, Busan, Korea.