Binocular Disparity Estimation Algorithm Using Multiple Spatial Frequency Information and a Neural Network

Ryoka Sato
Graduate School of Information Science and Technology, Osaka Institute of Technology,
1-79-1 Kitayama, Hirakata, Osaka 573-0196, Japan

Hirotsugu Okuno
Faculty of Information Science and Technology, Osaka Institute of Technology,
1-79-1 Kitayama, Hirakata, Osaka 573-0196, Japan
E-mail: hirotsugu.okuno@oit.ac.jp
www.oit.ac.jp

Abstract

We developed a disparity estimation algorithm that uses a disparity energy, which is computed based on a model of neurons responding selectively to a particular disparity. Disparity energy values depend on the spatial frequency of input images because of the Gabor filters employed. To reduce the frequency dependency, our algorithm uses disparity energy values computed from multiple spatial frequencies for neural-network-based regression. The algorithm was successful in estimating a disparity from images with a certain range of spatial frequencies.

Keywords: neural network, image processing, binocular disparity, disparity energy, visual cortex

1. Introduction

Depth information is one of the most important information for animals, humans, and robots to move in real environments. The visual nervous system can acquire depth information from disparity, which is the difference in the position of images reflected on the left and right retinas. Depth estimation from binocular disparity consists of the following two procedures: searching for corresponding points in the left and right images, and calculation of the gap between the corresponding points. Similar feature points located close to each other can cause incorrect matching between the left and right images, which leads to a failed depth estimation. Therefore, it is important to reduce candidates for the corresponding points with a similar feature.

The visual nervous system reduces candidates with similar features by applying a spatial bandpass filter to images on the retina. The disparity energy model, which is a model of the visual nervous system that estimate a corresponding points utilizing a bandpass filter, has been proposed based on physiological studies [1], and output neurons of the model respond selectively to a particular disparity.

However, the disparity energy model, which uses a single bandpass filter, has a problem with disparity estimation in regions where the amplitude of the spatial band of the bandpass filter is small. To overcome this drawback, previous studies proposed a use of disparity energies computed using multiple bandpass filters [2],[3].

The purpose of this study is to develop a disparity estimation algorithm that uses neural-network-based regression whose input signals are disparity energy values with multiple spatial frequency information.

2. Disparity Energy Model

The disparity energy model is a computational neuronal network model that explains the activity of cortical neurons responding selectively to a particular binocular disparity.

Fig. 1 shows the processing flow of the disparity energy model used in this study. The disparity energy is computed from two sets of images, which are acquired by the left and right eyes and are filtered by four Gabor filters whose phases are different by $\pi/2$ radian. The disparity energy I is given by:

$$\begin{aligned} I(x,y,\phi) = \{&G(x,y,0) * L(x,y) \\ &+G(x,y,\phi) * R(x,y)\}^2 \\ &+\{G\left(x,y,\frac{\pi}{2}\right) * L(x,y) \\ &+G\left(x,y,\phi+\frac{\pi}{2}\right) * R(x,y)\}^2. \end{aligned} \quad (1)$$

$G(x,y,\phi)$ represents a Gabor filter with phase offset ϕ. $L(x,y)$ and $R(x,y)$ represents the left and right input images at coordinates (x,y). The disparity energy is maximized when the phase offset ϕ of the Gabor filter matches the disparity of the input image. Therefore, phase ϕ that maximizes I is the estimated disparity in radians. The estimated disparity d in pixels are given by

$$d = \frac{\phi l}{2\pi}, \quad (2)$$

where l is the wavelength of the sinusoidal function in the Gabor filter. The estimated disparity depends on the frequency, and therefore, the detectable range also depends on the wavelength; the range is $-l/2 < d < l/2$.

Fig. 2 shows an example of the output of the disparity energy model when the input images are grating images and when the right image is shifted by +3 pixels with respect to the left image. The wavelength l of the Gabor filter used is 12 pixels, and the shift of the right image corresponds to $\pi/2$. Therefore, disparity energy $I(\pi/2)$ has the maximum value in a large region among all disparity energy images. This result tells that the estimated disparity in radians is $\pi/2$.

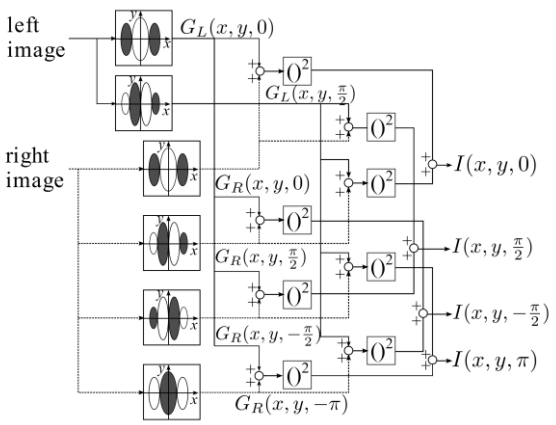

Fig. 1. Processing flow of the disparity energy model used in this study. The black and white ellipses represent the weights of the Gabor filter. The phase φ of the Gabor filter used is (0, $\pi/2$) for the left image and (0, $\pi/2$, $-\pi/2$, $-\pi$) for the right image.

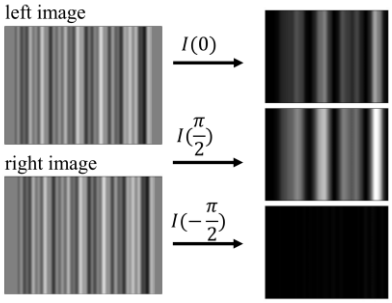

Fig. 2. Examples of disparity energy model output when the input is grating images. The right image is shifted by +3 pixels with respect to the left image. The right column shows disparity energy $I(0)$, $I(\pi/2)$, and $I(-\pi/2)$. Their corresponding disparities in pixels are $d = 0$, +3, and -3, respectively.

3. Disparity Estimation Algorithm

3.1. *Processing flow*

Fig. 3 shows the processing flow of the disparity estimation algorithm proposed in this study. First, the disparity energy is computed for two spatial frequencies using Gabor filters whose wavelength is 16 pixels and 32 pixels. Hereinafter, these energy values are referred to as high-frequency disparity energy I_h and low-frequency disparity energy I_l. These are transferred to the neural network after normalized respectively.

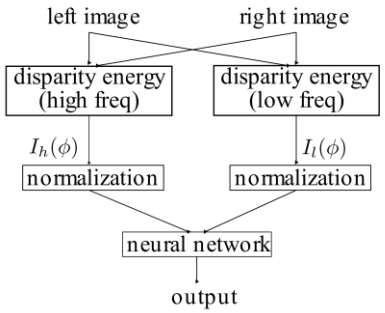

Fig. 3. Processing flow diagram of the proposed disparity estimation model with neural- network-based regression. I_h and I_l represents energy computing using high and low frequencies, respectively.

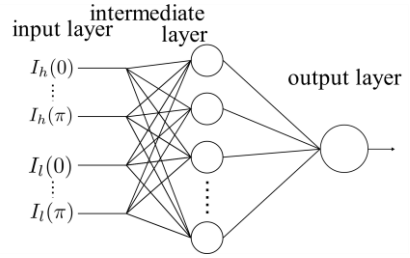

Fig. 4. Neural network used for regression. The number of neurons in the intermediate layer is 20.

3.2. *Neural-network-based regression*

Disparity estimation was performed using neural-network-based regression whose input signals are high-frequency and low-frequency disparity energies. Fig. 4 shows the neural network topology used for regression. The output of the neural network is the estimated disparity.

We used sinusoidal wave grating images as shown in Fig. 2 as training data of the neural network. The function of the wave is given by

$$S(x) = \sum_{i=0}^{N} A_i \sin\left(\frac{2\pi x}{\lambda_i} + \theta_i\right), \qquad (3)$$

where N in an integer from 7 to 15. We generated a pair of grating images; one is the wave grating generated by Eq. (3) directory, and the other is the image obtained by shifting the first image along x axis. The second one simulates an image with a certain disparity. We generated 100 images for each disparity ($-6<d<+6$; disparity d is an integer) and extracted 100 points from each image to prepare training data, resulting in 10,000 data points for each disparity.

4. Experiments and Results

4.1. *Experimental environment*

We implemented the model described in the previous section and trained the neural network using Python and Pytorch. The trained models were evaluated in the following two ways. First, we evaluated the model by using wave grating images generated by Eq. (3). The generated data were separated into training data and test data. Second, we evaluated the model by using images acquired by a binocular camera system.

4.2. *Evaluation using grating images*

We used two types of regression models to examine effects of using multiple frequencies: a model using only low-frequency disparity energy and a model using both frequencies. We compared the estimation results at 10,000 points in the grating images.

Fig. 5 shows the estimation results of the regression model that uses only low-frequency disparity energy. Fig. 5(a) shows that most of the estimated disparity is close to the correct disparity. The absolute error is smaller for a small disparity. Fig. 5(b) shows that all the absolute errors were within 0.5 pixels when the correct disparity is 0, but the error increased for a larger correct disparity.

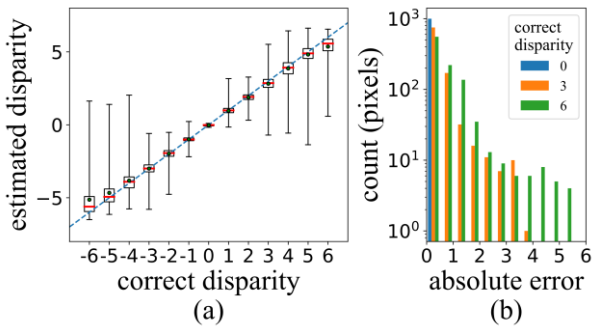

Fig. 5. Estimation results of the regression model that use only low-frequency disparity energy. (a) Box-and-whisker plot of the estimation result. The horizontal and vertical axis represent the correct and the estimated disparity, respectively. The red line and green dots show the median and mean of the estimated disparities, and the dashed line plots the correct disparity. (b) Histogram of the absolute errors. The histogram has 12 bins in the range from 0, to 6 pixels.

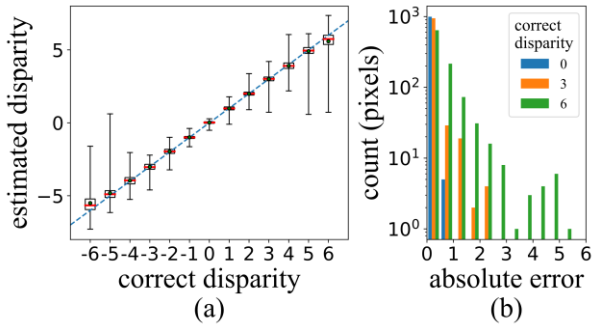

Fig. 6. Estimation results of the regression model using both low and high frequency disparity energies. The method of creating the figures is the same as that in Fig. 5.

Fig. 6 shows the estimation results of the regression model that uses both low and high frequency disparity energies information. Error conditions are similar to Fig. 5, but the error is smaller for a larger correct disparity compared to Fig. 5.

4.3. *Evaluation using a binocular camera system*

Fig. 7 shows the experimental environment used to investigate the disparity estimated from images obtained by a binocular camera system. A sinusoidal wave grating image was placed in front of the binocular system.

Fig. 8 shows the disparity map image estimated by the model that was trained to obtain Fig. 6. Although apparent spatial frequency of the image changes depending on the distance to the object, the results in Fig. 8(a)(b)(c) show that the proposed model estimated the correct disparity irrespective of the input spatial frequency.

5. Conclusion

In this study, we developed an algorithm that uses the disparity energy and neural-network-based regression. The simulation results showed that the use of two spatial frequencies is effective reducing errors. The results obtained by the binocular camera experiments suggest that our method is applicable to robotic vision.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 19K12916.

References

1. I. Ohzawa, G.C. DeAngelis, and R.D. Freeman, “Stereoscopic depth discrimination in the visual cortex: neural ideally suited as disparity detectors”, Science, vol. 249, No.4972, pp.1037-1041, Aug, 1990.
2. D. J. Fleet, H. Wagner, and D. J. Heeger, “Neural encoding of binocular disparity: energy models, position shifts and phase shifts”, Vision Research, Vol. 36, pp. 1839-1857, June, 1996.
3. J.J. Tsai, and J.D. Victor, “Reading a population code: a multi-scale neural model for representing binocular disparity”, Vision Research, Vol. 43, pp. 445-466, Feb, 2003.

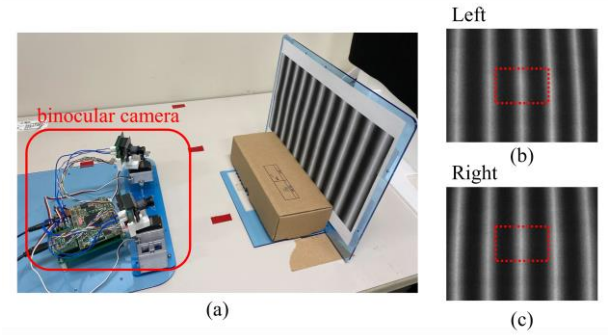

Fig. 7. (a) Experimental environment. A binocular camera system acquires the grating image. (b)(c) Left and right images acquired by the system. The estimated disparity within the red dashed rectangle in (b)(c) was used to obtain the disparity map image in Fig. 8.

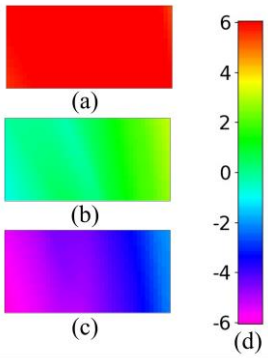

Fig. 8. Disparity map estimated by the proposed model in the dashed red rectangle of Fig. 7(b)(c). (a)(b)(c) Disparity map when the distance to the grating image is far, intermediate, and near, respectively. The color represents estimated disparity, and the correspondence between colors and estimated values of the disparity is shown in (d).

Authors Introduction

Mr. Ryoka Sato

He received his B.S. degree from the Department of Information Science and Technology, Osaka Institute of Technology, Japan in 2022. He is currently a Master's course student in Osaka Institute of Technology, Japan.

Dr. Hirotsugu Okuno

He received the Ph.D degree in electrical, electronic and information engineering from Osaka University, in 2008. He is currently an Associate Professor at the Faculty of Information Science and Technology, Osaka Institute of Technology. His research interests include visual information processing in the nervous system and their applications to robotics.