Scene Flow Estimation from Sequential Stereo Fisheye Images

Preliminary Study

Arief Suryadi Satyawan[1,4], Junichi Hara[2,3], Hiroshi Watanabe[1,2]
[1]Department of Computer Science and Communications Engineering, Waseda University, Japan
[2]Global Information and Telecommunication Institute, Waseda University, Japan
[3]RICOH Company, LTD, Japan
[4]Research Center for Electronics and Telecommunication, Indonesian Institute of Sciences, Indonesia

***Abstract*—This paper discusses estimating 3D scene flow directly from sequential stereo fisheye images. The 3D scene flow estimation framework consists of optical flow, disparity, and scene flow calculation. The optical flow and disparity calculations are performed by making use of Lucas and Kanade's concept. This concept used to be applied for calculating optical flow. The results show that the 3D scene flow can be generated directly from the sequential stereo fisheye images with certain conditions.**

***Keywords*— scene flow, optical flow, disparity, fisheye, stereo, Lucas and Kanade**

I. Introduction

Research and development on the omnidirectional vision have grown rapidly in the recent year due to the high demand in many applications. These applications are ranging from surveillance, autonomous navigation system, robotics, localization and mapping, and video communication system. In the same time, the wide view used to be composed by aligning multiple narrow views captured by single or arranged general perspective camera now can be created directly from an omnidirectional camera such as a fisheye camera [1].

However, there is a challenge to apply directly the wide view taken from a fisheye camera for the above applications. Due to the characteristic of the fisheye lens, the fisheye image suffers from distortion. Additional effort is needed to overcome this distortion. One of a challenge being investigated in our research is to make use of fisheye images for generating 3D scene flow. The 3D scene flow is very useful for estimating motion in the real world. The state-of-the-art image processing along with sophisticated computer vision technique has seemed to be achieving remarkable performance in estimating the 3D scene flow from narrow view images [3,4,8,9]. However, this achievement might not useful for serving fisheye images directly. We need to modify previous approaches before applying for exploiting the fisheye images.

In this paper, we focus on developing 3D scene flow estimation directly from stereo fisheye images. We propose to make use of optical flow and disparity calculation to solve 3D scene flow approximation. The optical flow and disparity calculations are performed by incorporating Lucas and Kanade's approximation.

II. Related Works

In our observation, the concept of generating 3D scene flow starts from [6]. The approach basically provides a large opportunity for obtaining 3D scene flow with or without taking geometry information continuously from the real 3D surface. That last option will make generating the 3D scene flow more efficient since we can depend on 2D information greatly, although in the same time we will face greater noise that makes the 3D scene flow seems to be unreliable. Furthermore, combining information from 2D optical flow and stereo disparity to obtain scene flow is developed by [5]. This also becomes a basic knowledge for developing 3D scene flow in this research, although we use this for serving sequential stereo fisheye images. In [2], we already describe the theoretical aspect of developing 3D scene flow by make utilization of optical flow and disparity calculation. In this paper, we describe more detile about the experimental results.

III. Design of Scene Flow Estimation

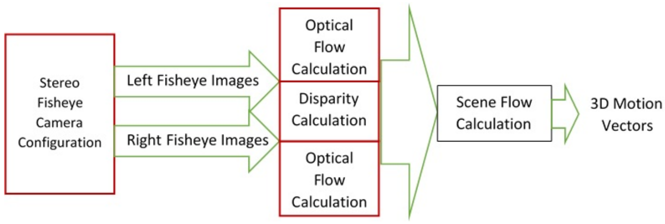

Fig. 1. The main concept of scene flow calculation for stereo fisheye camera

The scheme of 3D scene flow calculation in this research is adopted from our previous experiment [2]. Fig. 1, illustrates the main concept. Firstly, two fisheye cameras are configured into a stereo configuration. These fisheye cameras will produce sequential fisheye images denoted by the left and right sequential fisheye images. From then on, two successive images from each sequential fisheye image are used for obtaining optical flow, while a pair of images from the two sequential images are used for collecting disparity of images. Fig. 2, describes the mechanism of optical flow and disparity calculation. Finally, the optical flow and disparity can be used for finishing scene flow calculation, therefore 3D motion vectors can be generated. These calculations are repeated for the whole sequential fisheye images.

978-1-5386-5689-1/18/$31.00 ©2018 IEEE

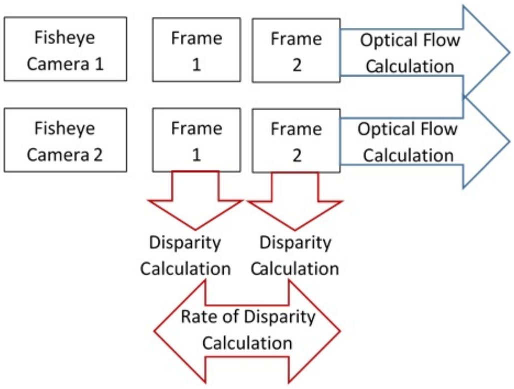

Fig. 2. The mechanism of optical flow and disparity calculations

In this research, either the optical flow or disparity calculation is defined as problems of finding corresponding points in two images. Therefore, we adopt Lucas and Kanade's approach presented in Equation (1) in term of finding the solution for these cases.

$$\begin{pmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{pmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -\sum I_x I_t \\ -\sum I_y I_t \end{bmatrix} \quad (1)$$

Note that I_x and I_y are the horizontal and vertical gradient of the image, while I_t is image difference between two images. In our purpose, for optical flow calculation, the two images mean the two successive images taken from each sequential fisheye images, while for disparity calculation, they are a pair image obtained from the two sequential fisheye images. Additionally, I_x and I_y are defined from the first image (forward calculation).

In similar fashion to optical flow, which is defined as the two-dimensional motion of objects or pixels on an image plane, scene flow is defined as the three-dimensional motion field of objects or points in the real world. This means that simple projection of scene flow onto an image plane is optical flow [6]. The relation can be explained as follow. By considering a 3D point of **A** (X,Y,Z) located in the real world, the projection of this point onto an image coordinate $\mathbf{a}_n$ (x_n,y_n) of camera n is given by (2) and (3), in which $[P_i]_j$ is j^{th} row of the projection matrix P_i.

$$x_n = \frac{[P_i]_1(X,Y,Z,1)^T}{[P_i]_3(X,Y,Z,1)^T} \quad (2)$$

$$y_n = \frac{[P_i]_2(X,Y,Z,1)^T}{[P_i]_3(X,Y,Z,1)^T} \quad (3)$$

If we assume that the cameras are static, the relation between the two-dimensional motion ($\vec{v} = d\mathbf{a}_n/dt$) and the scene flow ($\vec{V} = d\mathbf{A}/dt$) can be expressed as (4).

$$\frac{d\mathbf{a}_n}{dt} = \frac{\partial \mathbf{a}_n}{\partial \mathbf{A}} \frac{d\mathbf{A}}{dt} \quad (4)$$

Based on (5), the scene flow then can be calculated by at least two cameras. Therefore, it should be composed as a linear equation $VB = U$, which is also performed completely by (5).

The 3D coordinate **A** in stereo setup has a relation with disparity (d) and corresponding image coordinates (x_{Right} and y_{Right} for right view, while x_{Left} and y_{Left} for left view), and their relationship is defined by (6). T denotes as baseline (distance between the two cameras), while f denotes as focal length of the cameras (assume that the two cameras have similar focal length, $f = 1$).

$$B = \begin{bmatrix} \frac{\partial x_1}{\partial X} & \frac{\partial x_1}{\partial Y} & \frac{\partial x_1}{\partial Z} \\ \frac{\partial y_1}{\partial X} & \frac{\partial y_1}{\partial Y} & \frac{\partial y_1}{\partial Z} \\ \frac{\partial x_2}{\partial X} & \frac{\partial x_2}{\partial Y} & \frac{\partial x_2}{\partial Z} \\ \frac{\partial y_2}{\partial X} & \frac{\partial y_2}{\partial Y} & \frac{\partial y_2}{\partial Z} \end{bmatrix}, U = \begin{bmatrix} \frac{\partial x_1}{\partial t} \\ \frac{\partial y_1}{\partial t} \\ \frac{\partial x_2}{\partial t} \\ \frac{\partial y_2}{\partial t} \end{bmatrix} \quad (5)$$

$$X = \frac{T(x_{Right} + x_{Left})}{2d}, Y = \frac{T(y_{Right} + y_{Left})}{2d}, Z = \frac{fT}{d} \quad (6)$$

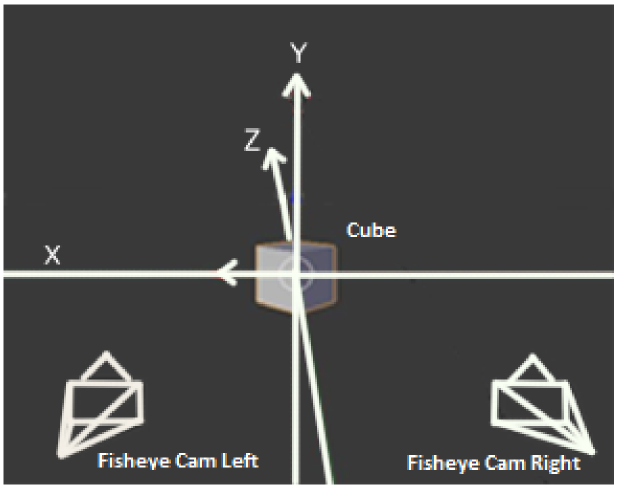

Fig. 3. The camera configuration with an object developed by using blender

TABLE I. The design of scene flow algorithm

Input: two successive fish-eye images from either left or right fisheye camera
Calculating optical flow
*1: **for** each two successive images, **do***
2: $im1_{st}$ & $im2_{nd}$ ← converting the two images to grayscale
3: $im1_{st}$ & $im2_{nd}$ ← smoothing the two grayscale images by using Gaussian smooth filter
4: I_x, I_y & I_t ← obtaining spatial and temporal gradient of the images
5: u & v ← calculating vector flows by using matrix equation 1.
*6: **end for***
*7: **Return***
Input: two stereo fish-eye images from either t=1 or t=2
Calculating disparity
*1: **for** each two successive images, **do***
2: im_{Left} & im_{Right} ← converting the two images to grayscale
3: im_{Left} & im_{Right} ← smoothing the two grayscale images by using Gaussian smooth filter
4: I_x, I_y & I_t (im_{Left} & im_{Right}) ← obtaining spatial and stereo gradient of the images
5: d ← calculating disparity by using equation 2 and calculating rate of disparity
6: end for
7: Return
Input: optical flow and disparity
1: Calculating equation 5
2: Displaying 3D scene flow
3: Calculating AAE and SAE
4: Return

In this experiment, the sequential stereo fisheye images is made synthetically by using blander. This synthetic images can be set up ideally, therefore both qualitative and quantitative evaluation can be obtained easily. The configuration of the camera with an object used for the experiment is illustrated in Fig. 3. Two fisheye cameras are composed in a stereo configuration with the distance between the two cameras is 4 or 6 cm. An object, a cube with an edge of 10 cm, is located in front of the stereo camera. The distance between the stereo camera and the cube is 15 cm. This cube is moved 1, 2, 3, or 4 cm forward started from a left side position to a right side position. The stereo fisheye camera then captures the movement of the cube (e.g., the first image shows the cube at the original position, the second image shows the cube after one cm movement, the third image displays the cube after 2 cm movement, etc.). From this scenario, there are 190 fisheye images from each sequential fisheye images. Finally, these fisheye images are applied to the scene flow calculation framework written in TABLE 1.

IV. Results and Discussion

In this section, the qualitative and quantitative performance of scene flow estimation are discussed comprehensively. In term of qualitative evaluation, due to page limitation, there are only three types of fisheye images displayed in this paper. The first is fisheye images showing the cube at the left side position, which means the cube is located in front of camera, but at the left side. The second is fisheye figures presenting the cube in the middle position, which means the cube exactly in front of the stereo camera. The last is fisheye images displaying the cube at the right side position, which means the cube is located in front of the camera, but at the right side. On the other hand, the quantitative evaluation for the sequential fisheye images describing the cube position from the left to the right side of the stereo camera can be reported in this paper.

As we have discussed in the previous section, the first step of the scene flow estimation is to determine optical flow directly from two successive fisheye images taken from each camera in the stereo configuration. These fisheye images can be seen from the two upper rows of Fig. 4. Specifically, both (a) and (b) consist of two consecutive fisheye images taken from left and right camera (LC and RC) respectively. They display the position of the cube on the left side. Both (c) and (d) contain two consecutive fisheye images taken from LC and RC consecutively. They show the cube in the middle position. (e) and (f) are composed of two consecutive fisheye images taken from LC and RC successively. They present the position of the cube on the right side. Moreover, the optical flow resulted from the calculation can be seen from the two lower rows of Fig. 3. (g), (h), (i), (j), (k), and (l) are optical flow for the consecutive images of (a), (b), (c), (d), (e), and (f) respectively.

As a result, the optical flow occurs in the border area of the cube due to the strong difference in intensity between the cube and background. However, since the intensity around the edge of the cube (in the middle area of the cube) is almost similar, optical flow cannot be obtained at this area. This phenomenon is visible clearly when the cube is located in front of the stereo camera. On the other hand, when the cube is located at the left or right side position, the optical flow occurs almost similar, even though it appears strongly only around the left and right edge of the cube. This means that there is a specific threshold of the difference in intensity between the two areas, in which the optical flow can still be obtained by the system calculation. The Fig. 4, shows the optical flow when the distance between two cameras is 6 cm.

The second step is the calculation of disparity between two fisheye images provided by the LC and RC. These stereo fisheye images are shown by the first row of Fig. 5. (a) and (b) are a stereo fisheye images showing the cube at the left side position. (c) and (d) are a stereo fisheye images showing the cube in the middle position. (e) and (f) are a stereo fisheye image showing the cube at the right side position. Furthermore, the disparity of each stereo fisheye images can be seen from the last two rows of Fig. 5. In this calculation, the distance between two cameras is 6 cm.

The disparity of fisheye images occurs prevalently around the border areas between the cube and background. A constant result can be obtained when the cube is located in the middle area, even though it is not visible at the edge area of the cube with similar intensity, especially when the cube is located in the middle position. On the contrary, the disparity obtained when the fisheye images show the cube at the left side position is different with the disparity occurs when the fisheye images show the cube at the right side position. It seems like there are so many fake vectors occurs when the cube at the right side position. However, they will be neglected by the opposite one during the calculation of scene flow.

Scene flow calculation finally performs in the last step. Fig. 6, describe the visual performance of scene flow estimation. (a), (b), and (c) show the scene flow obtained when the cube in the stereo fisheye image is located at the left, center and right side position respectively. The distance between two cameras is 6 cm. From that figure, it can be seen that the scene flow can be generated although the position of the cube in the stereo fisheye images is located at the left, middle, or right side. However, when the fisheye images show the cube at the left or right side position, they occur sparsely. The scene flow estimation is affected greatly by the results of optical flow and disparity calculation.

Furthermore, the qualitative evaluation indicated by the average angular error (AAE) and the standard deviation average angular error (SAE) is presented in Fig. 7. While (a) and (c) show AAE and SAE when the distance between two cameras is 4 cm, (b) and (d) illustrate the same quantitative evaluation when the distance between two cameras is 6 cm. The frame number indicated by the horizontal axis describes the cube position in each fisheye images. It starts from 0 (left side position), and it finishes to 190 (right side position). The evaluation is applied to different types of cube movements (1, 2, 3, and 4 cm). It can be seen that AAE and SAE are lower than around 0.1^{o} and 1.5^{o} respectively, especially when the movement of the cube is 1 cm. However, these parameters increase slightly to around 0.2^{o} and 2.5^{o} for AAE and SAE consecutively, when the movement rises from 1 cm until 4 cm.

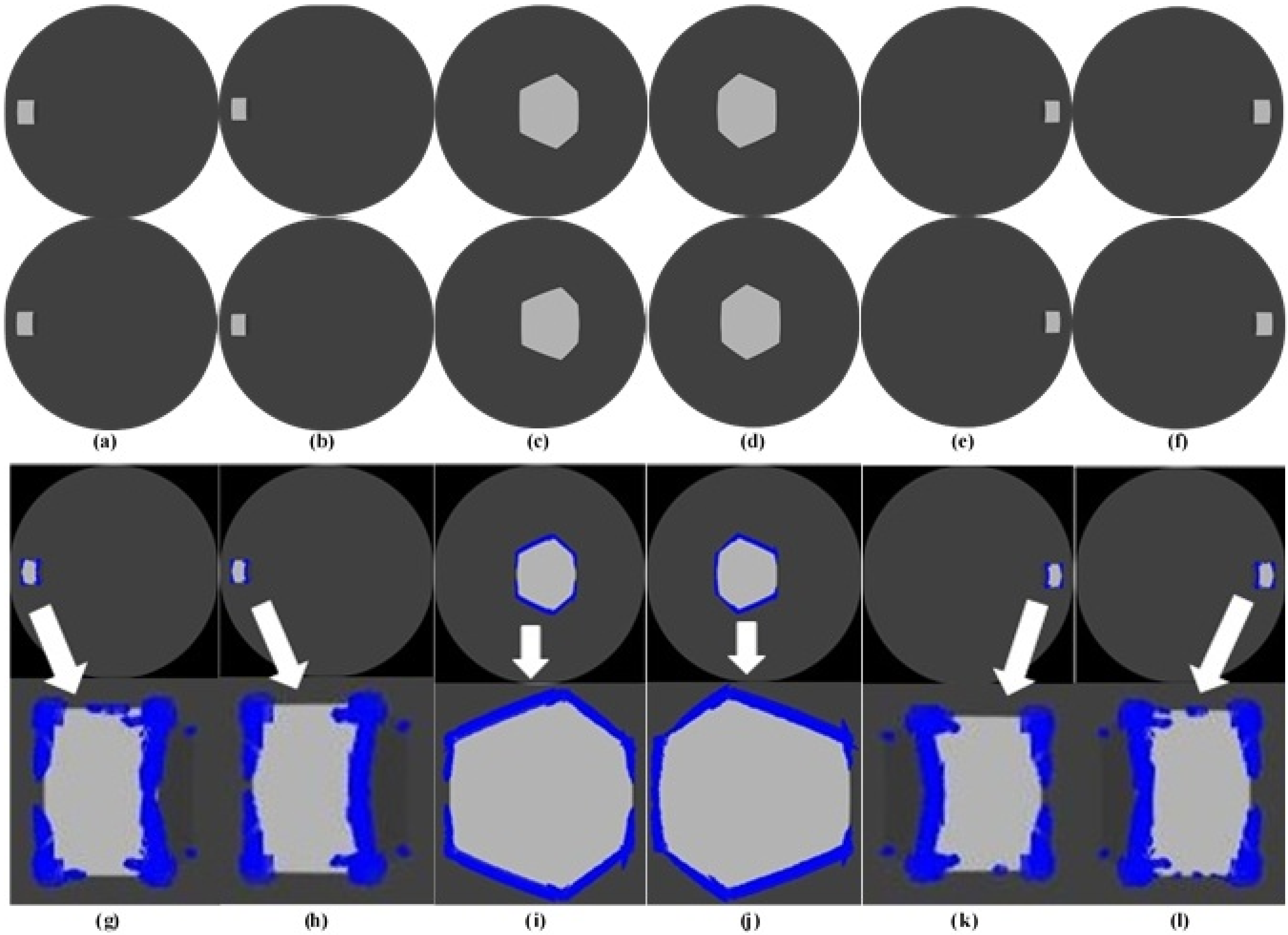

Fig. 4. The example of consecutive images produced by each camera showing a cube at a different location, and the optical flow obtained from the calculation involving each of the consecutive images. The distance between two cameras is 6 cm. (a), (c), and (e) consists of consecutive images taken by the LC, while (b), (d), and (f) composed of two consecutive images released by the RC. In the last two rows, (g), (h), (i), (j), (k), and (l) are optical flow for the consecutive images of (a), (b), (c), (d), (e), and (f) consecutively.

This condition is achieved when the distance between two cameras is set up to 4 or 6 cm, and the cube in the stereo fisheye images is located at the right or left side position.

Nevertheless, these quantitative parameters are higher, especially when the cube in the stereo fisheye images is located in the middle position. When the cube movement is 1 cm, AAE and SAE reach 0.22^{o} and 2.7^{o} respectively. This situation occurs when the distance between two cameras is set up to 4 or 6 cm. Moreover, the results become worst when the cube movement is increased to 4 cm. The AAE and SAE rise to around 0.7^{o} and 5^{o} respectively, when the distance between two cameras is set up to 4 cm. For the same cube position (in the middle), and the distance between two cameras is set up to 6 cm, the AAE and SAE increase to around 0.5^{o} and 4.5^{o} consecutively.

V. CONCLUSION

Estimating scene flow directly from stereo fisheye images can be done by incorporating optical flow and disparity calculation. Both can be calculated by using Lucas and Kanade's approximation with different composition of input images. However, visual investigation shows that the scene flow can be estimated very well only in the areas where intensity differences occur strong enough, otherwise it is difficult to generate 3D vectors. The performance of the scene flow also depends on the motion of the object, location of object and distance between two cameras. It is important to do further investigation deeply, therefore the system calculation can adapt with these three aspects, and finally, it can be used for serving real fisheye images.

ACKNOWLEDGMENT

A special credit goes to the Ministry of Research, Technology and Higher Education of Republic of Indonesia, which grants me the Riset-Pro scholarship for pursuing a doctoral degree at Waseda University.

REFERENCES

[1] https://theta360.com, accessed on 1 April 2017
[2] A.S. Satyawan, J. Hara and H. Watanabe: "Scene Flow from Stereo Fisheye Images," International Workshop on Advanced Image Technology (IWAIT2018), No.77, pp.1-4, Jan. 2018.

[3] Z. Yan and X. Xiang, "Scene Flow Estimation: A Survey", arXiv preprint arXiv:1612.02590 (2016).

[4] A. Wedel and D. Cremers, "Stereo Scene Flow for 3D Motion Analysis", DOI 10.1007/978-0-85729-965-9_2, Springer-Verlag 2011.

[5] R. Li and S. Sclaroff, "Multi-scale 3D scene flow from binocular stereo sequences", Computer Vision and Image Understanding, Vol.110, No.1, pp.75-90, Apr. 2008.

[6] S. Vedula, P. Rander, R. Collins, and T. Kanade, "Three-dimensional scene flow", IEEE Transactions on Pattern Analysis and Machine Intelligence,Vol. 27, No.3, Mar. 2005.

[7] S. Baker and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework", International Journal of Computer Vision, Vol.56, No.3, pp.221–255, 2004.

[8] M. Manze, and A. Geiger, "Object Scene Flow for Autonomous Vehicles,' In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2015.

[9] L. Zhang, B. Curless, and S.M. Seitz, "Spacetime Stereo: Shape Recovery for Dynamic Scenes," In Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2003.

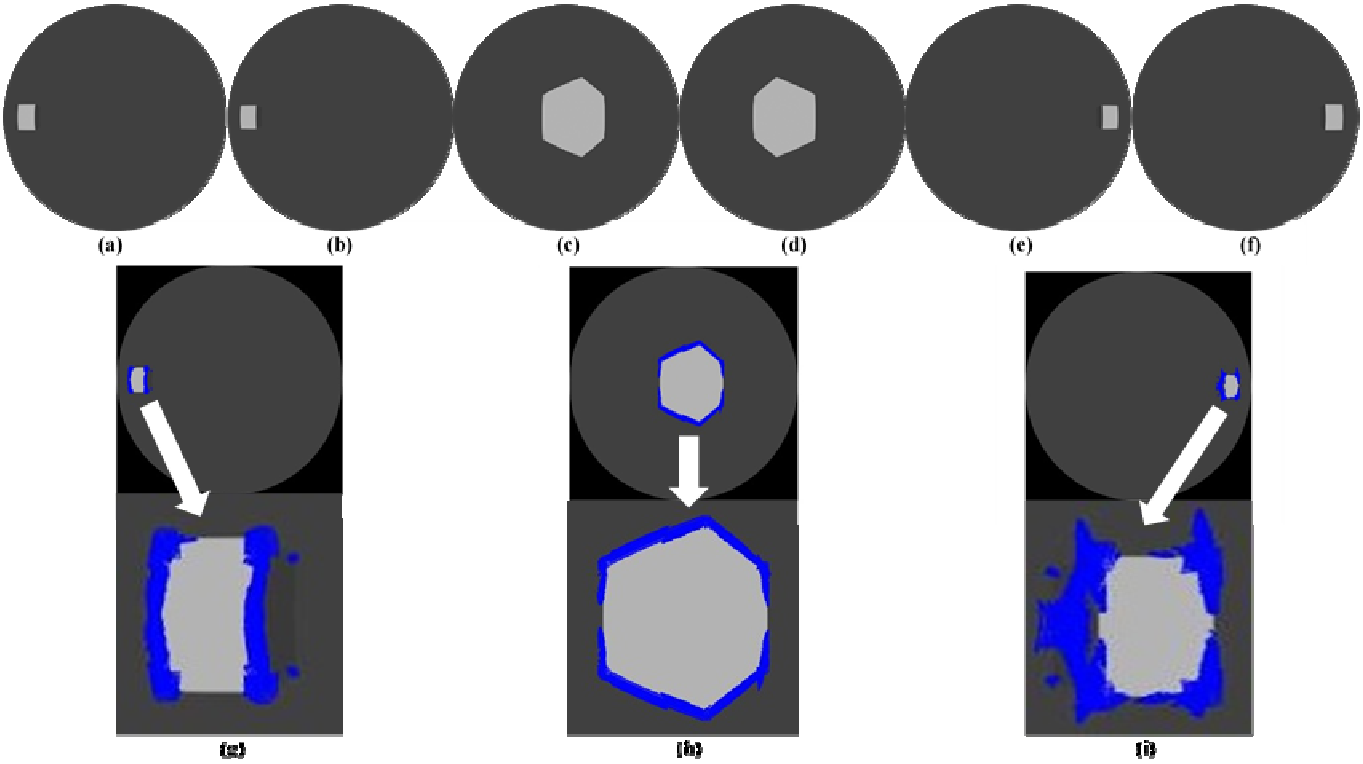

Fig. 5. The example of the pair of images produced by the stereo camera showing a cube at a different location. The distance between two cameras is 6 cm. In the first row, (a) and (b), (c) and (d), and (e) and (f) are stereo images showing the position of the cube at the left, middle, and right side respectively. In the last two rows, (g), (h), and (i) show disparity of fisheye images yielded from the calculation.

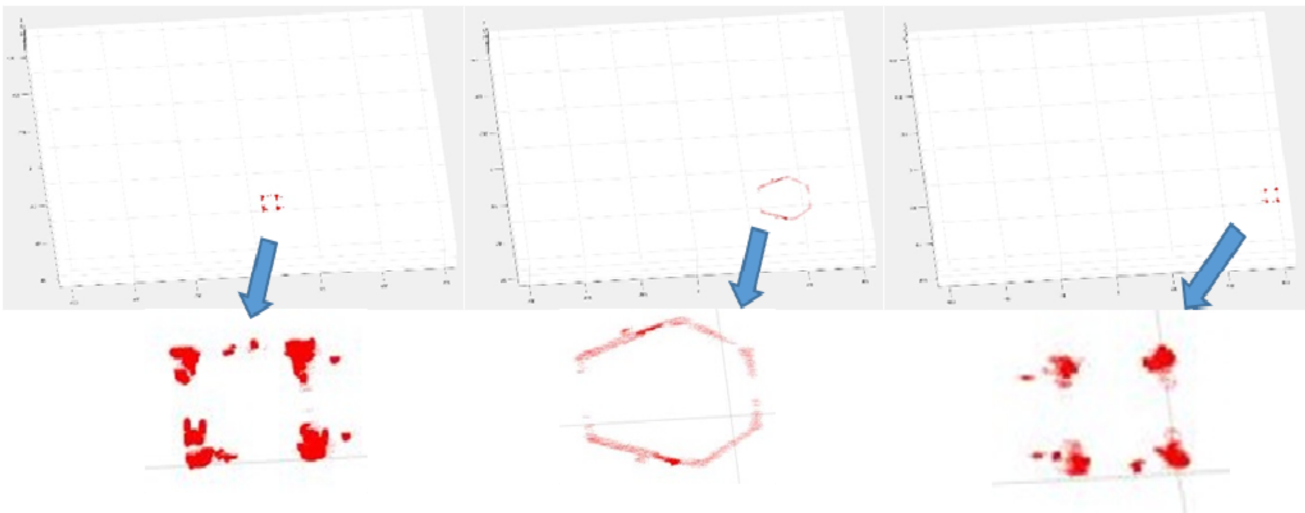

Fig. 6. Visual performance of 3D scene flow according to different object location. (a), (b), and (c) show the scene flow obtained when the cube is located at the left, center and right of the stereo camera respectively. The distance between two cameras is 6 cm.

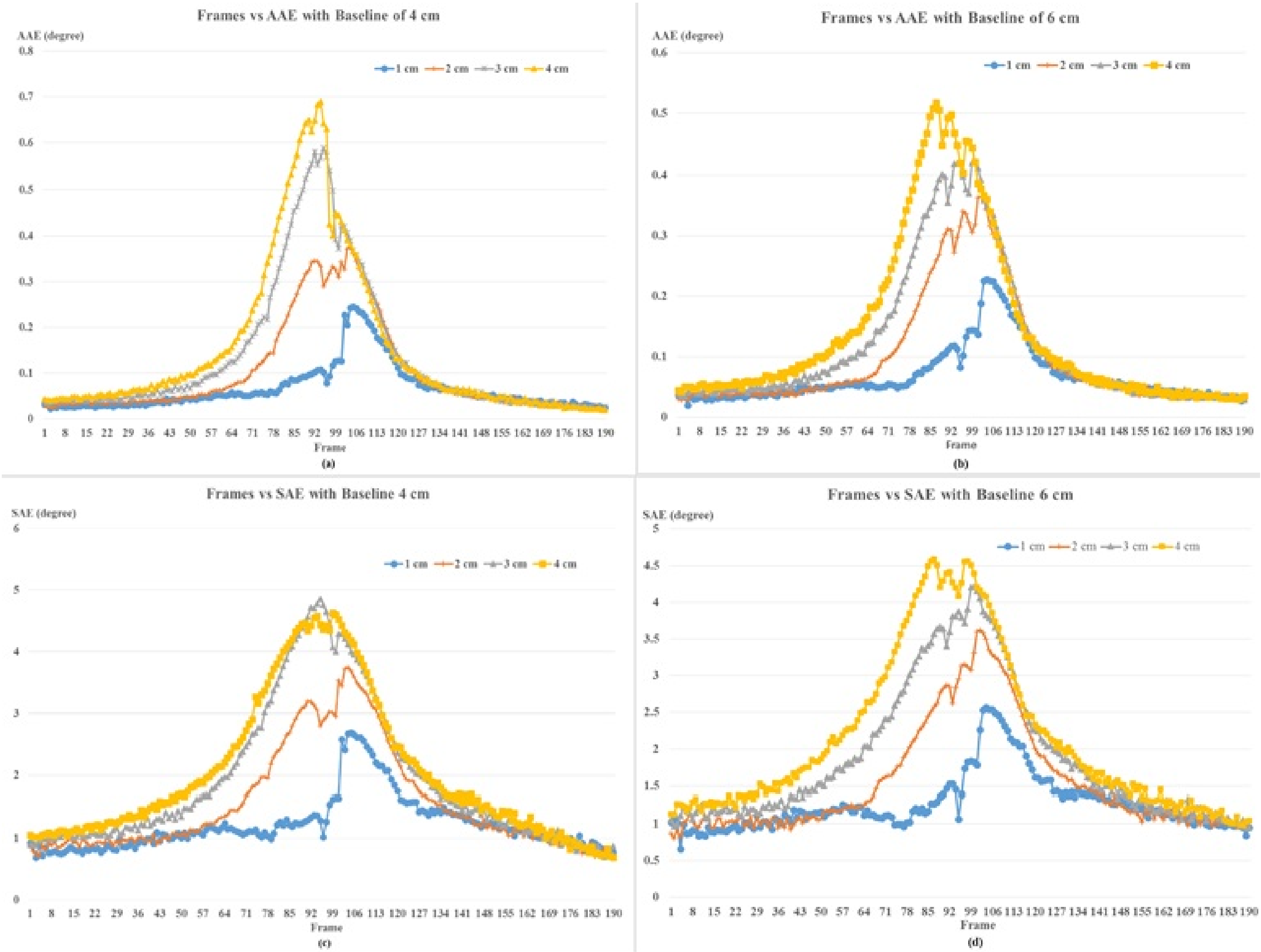

Fig. 7. Quantitative performance of scene flow estimation from sequential fisheye images. (a) and (b) are AAE scores, while (c) and (d) are SAE scores.