Scene Analysis Based on Horse Vision System

Hanchao Jia
Tottori University
4-101, Koyama, Tottori, 680-8552, Japan
E-mail: m09t3009@faraday.ele.tottori-u.ac.jp

Shigang Li
Tottori University
4-101, Koyama, Tottori, 680-8552, Japan
E-mail: li@ele.tottori-u.ac.jp

Abstract

In this paper we propose a method of scene analysis based on a horse vision system. A horse vision system (HVS) consists of a pair of fisheye cameras which have a hemispherical field of view, respectively, and are laid to overlap each other partially. The characteristics of the HVS result in a representation which enables a wide omnidirectional monocular vision and a limited-field-of-view binocular vision simultaneously. We present the method of the rectification of the proposed HVS and the preliminary experimental results of scene analysis based on the HVS.

1. Introduction

Vision-based scene analysis is a basic research topic in machine vision, and many approaches have been developed until now for this task. What kind of approaches should be used is dependent on the concrete task of scene analysis. For example, for motion detection by a stationary single camera, the method of background subtraction is popular and effective; for the construction of 3D structure of environment, a stereo method is usually used. While the processing of the background subtraction by a single camera is simple, the construction of 3D structure of environment by a stereo method not only needs multiple cameras, but also costs a large amount of computation relatively. The scene analysis using a single camera is referred to as monocular vision while that using two cameras is referred to as binocular vision.

What should a vision system be like when we construct a vision system for a general scene analysis task? Biological vision systems often give us a hint. Human beings have a binocular vision system; however, the scene in the rear cannot be observed. To cope with a dynamic environment in real time, a wide field of view (FOV) brings definite merits to a vision system.

A horse has two big eyes, which located on either side of its head. On one hand, like people, it can see the same scene with both eyes at once, resulting in better depth perception and a more concentrated field of vision. On the other hand, a horse can also use each eye to see separate scene, which enables it to get a view of surroundings on both sides. It is greatly important for horses to detect stalking carnivorous animals sneaking up from behind in order to avoid danger. So different from people, a horse holds a wide, circular view with a range of vision of more than 350°. Approximately 65° among that is binocular vision, while the remaining 285° is monocular vision [1], [2], as shown in Fig. 1. It implies that by imitating a horse's eyes, we can construct a vision system, called a horse vision system (HVS), with two modes of vision, monocular vision and binocular vision. Using the HVS, a mobile robot can detect obstacles by the narrow FOV binocular vision, and meanwhile, monitor the surrounding environment by the wide omnidirectional monocular vision.

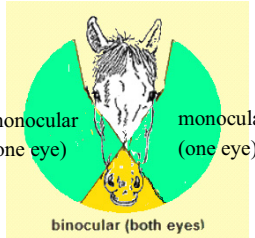
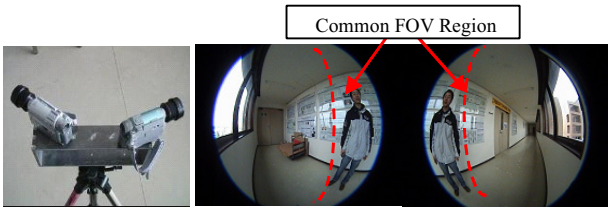


Fig. 1. A sketch of the explanation of the function of horse vision.

In this paper, we imitate a horse's eyes using a pair of fisheye cameras, as shown in Fig. 2(a). Each of them has a hemispherical FOV. The pair of fisheye cameras is mounted on a rig with the overlapping region of observation fields so as to acquire the two modes of vision, monocular vision and binocular vision, simultaneously. The method of the rectification of the proposed HVS and the preliminary experimental results of scene analysis based on the HVS are presented.

The remainder of this paper is organized as follows: Related research is introduced in the next section. In Section 3, the realization of the HVS is described. An application of scene analysis by the HVS is presented in Section 4. Finally, conclusions are given in Section 5.



(a) (b)

Fig. 2. (a) The horse vision system proposed in this paper. (b) A pair of sample fisheye images captured by the HVS.

2. Related Research

In this section, after introducing the related research on omnidirecitonal vision, we explain the characteristics of the proposed HVS.

2.1. Omnidirectional monocular vision

Omnidirectional image sensors are widely used for visual surveillance and vision-based robot navigation. An omnidirectional image can be obtained by a catadioptric image sensor [3], a fisheye camera [4], or a camera cluster [5]. Thanks to the wide FOV of an omnidirectional camera, it is very effective to detect the motion from the surrounding environment speedily. However, it is difficult to measure the 3D information of the surrounding environment from a single omnidirectional camera.

2.2. Omnidirectional binocular vision

To measure 3D information of the surrounding environment, multiple omnidirectional cameras are usually used by modifying the conventional stereo method [6], [7]. In order to obtain the 3D information of the surrounding environment as much as possible, the FOV of the multiple omnidirectional cameras are laid to overlay each other as largely as possible. In comparison with the motion detection by the background subtraction, the stereo method costs a lot of computation.

2.3. Characteristics of the HVS

In this paper our goal is to construct a biologically-inspired horse-like vision system. Concretely, a pair of fisheye cameras is used to realize binocular vision and omnidirectional monocular vision simultaneously like a horse, as shown in Fig. 2(a).

There are following characteristics in the HVS.

1) The pair of fisheye cameras point to different directions, respectively. Meanwhile, between them there are overlapping regions, i.e., the common FOV region (CFOVR), in the fisheye images for the realization of binocular vision. Obviously, this setup is different from that of the conventional canonical stereo.
2) The two visual modes, omnidirectional monocular vision and binocular vision, of the HVS result in reasonable distribution of limited computation power of a computer. By the HVS, a robot can monitor the surrounding environment by the omnidirectional monocular vision mode, and measure the 3D information by the binocular vision mode only for a part of environment which necessitates detailed investigation.

The main contribution of this paper is as follows.

1) Construct a biologically-inspired vision system, the HVS, like horses' eyes. According to our best knowledge, it may be the first system to try to imitate the visual function of a horse's eye.
2) Present an algorithm of identifying the CFOVR and rectifying the CFOVRs for the binocular vision of the HVS.
3) Present the preliminary experimental results of the scene analysis based on the proposed HVS.

3. Realization of the HVS

In this section, we describe the method of realizing the HVS, that is, the two visual modes, omnidirectional monocular vision and binocular vision, using a pair of fisheye cameras.

3.1. Omnidirectional monocular vision by the HVS

The intrinsic parameters of the pair of fisheye cameras used in the VHS are calibrated using the method of [4] beforehand. Given the correspondence of features between the pair of fisheye images, the relative pose between the pair of fisheye cameras can be computed using the method of [8]. Using the intrinsic and extrinsic parameters of the pair of fisheye cameras, the pair of fisheye images can be integrated into a wider omnidirectional image or a spherical image.

Figure 3(a) shows the spherical image obtained by mapping the pair of fisheye images onto a sphere. The duplication of the scenes appears in the CFOVR. Fig. 3(b) shows the image obtained by extending the spherical image along the longitudinal and latitudinal directions. For the omnidirectional monocular vision of the HVS of this paper, the motion detection is first carried out for each fisheye camera, and then, the relations of the detected motion regions are integrated in the spherical image.

(a) (b)

Fig. 3. (a) The integrated spherical image from a pair of fisheye images in Fig. 2(b). (b) The flatten image of (a).

3.2. Binocular vision by the HVS

Here, we focus on the explanation of the difference of the rectification of the HVS in comparison with the conventional canonical stereo.

3.2.1 Determination of the CFOVRs

First, we explain the idea of determining the boundaries of the CFOVRs. Fig. 4(a) shows two hemispherical images which correspond to the pair of fisheye images of the HVS. The boundary of the hemispherical images is the great circle on the XY plane of the camera coordinate system, respectively. The CFOVRs to be determined are indicated in blue color. Obviously, one part of the boundaries of the CFOVRs is the boundary of the hemispherical images. Thus, only the other part needs to be determined.

Note that for a scene point which has a big distance from the pair of fisheye cameras of the HVS in comparison with the displacement between the pair of fisheye cameras, the rays from the distant scene points to the pair of fisheye cameras of the HVS are approximately the same. This implies that we can determine the other part of the boundary of the CFOVIs by mapping the boundary of one of the two hemispherical images to the other one in terms of the

relative orientation between the pair of the fisheye cameras of the HVS by ignoring the displacement between the pair of fisheye cameras.

Next, we give the detail of the algorithm of determining the boundaries of the CFOVRs.

In the camera coordinate systems, respectively, for corresponding points $p_l(u_l, v_l, s_l)$ and $p_r(u_r, v_r, s_r)$ [(u, v, s) are the Cartesian coordinates on the spherical image] on the circumference of the great circle of the hemisphere, we have:

$$\begin{cases} p_l^T A p_l = 0 \\ s_l = 0 \end{cases}, \qquad \begin{cases} p_r^T A p_r = 0 \\ s_r = 0 \end{cases} \tag{1}$$

where A is identity matrix.

Using the relative pose (R_{rl}, t) between the pair of fisheye cameras computed in Section 3.1, we have the following equation:

$$p_l = R_{rl} p_r + t \tag{2}$$

By ignoring the displacement t between the pair of fisheye cameras of the HVS, the equation above is represented as follows:

$$p_l = R_{rl} p_r \tag{3}$$

Then, the boundaries of the left CFOVR can be presented in two parts. One is the same as the original image (indicated in red color),

$$p_l^T A p_l = 0, s_l = 0, u_l > 0 \tag{4}$$

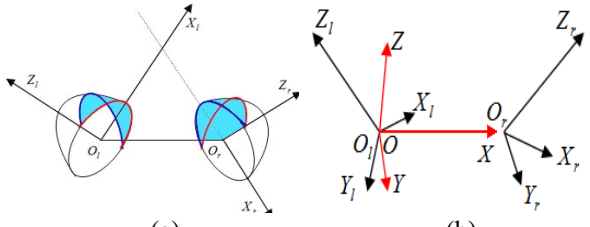
and the other is the projection of the half circumference of the great circle of the hemisphere in the right image (indicated in dark blue color). Using the equation (3), it can be presented as follows:

$$\begin{cases} (R_{rl}^{-1} p_l)^T A (R_{rl}^{-1} p_l) = 0 \\ s_l = 0, \quad u_l < 0 \end{cases} \tag{5}$$

Similarly, the right CFOVR can be determined. Finally, we have the equations of the boundaries of two CFOVRs, presented respectively as follows:

$$\begin{cases} p_l^T A p_l = 0, s_l = 0, u_l > 0 \\ (R_{rl}^{-1} p_l)^T A (R_{rl}^{-1} p_l) = 0, s_l = 0, u_l < 0 \end{cases}$$
$$\begin{cases} p_r^T A p_r = 0, s_r = 0, u_r < 0 \\ (R_{rl} p_r)^T A (R_{rl} p_r) = 0, s_r = 0, u_r > 0 \end{cases} \tag{6}$$

Concretely, we first determine the boundaries of the CFOVRs in hemispherical images based on a spherical camera model as described above, and then map the determined boundaries of the CFOVRs onto the fisheye image of the HVS by the known intrinsic parameters of the fisheye cameras.



(a) (b)

Fig. 4. (a) The CFOVRs of two cameras in our system. (b) The sketch of the rectification of the CFOVRs. $O-XYZ$ indicates the rectified camera coordinate system.

3.2.2 Rectification of the CFOVRs

Suppose that a scene point $P(X, Y, Z)$ is projected onto the pair of the original hemispherical image of the HVS as p_{slo} and p_{sro}, respectively. Using the relative pose (R_{rl}, t) between the pair of fisheye cameras of the HVS computed in Section 3.1, we have the following equations:

$$p_{slo} \cong [A \quad 0]\tilde{P}, \quad p_{sro} \cong [R_{rl}^{-1} \quad -R_{rl}^{-1} t]\tilde{P} \tag{7}$$

where $\tilde{P}$ is the homogeneous coordinate of P. The symbol $\cong$ means that p_{slo} and p_{sro} are equal to $\tilde{P}$ multiplied by a scale factor.

To rectify the original CFOVRs, the original hemispherical images must be rotated so that the axes of the two new camera coordinate systems of the rectified CFOVRs are parallel to each other. Suppose that, after the rectification, the rotation matrix R_{rl} is changed to $R' = [r_1 \; r_2 \; r_3]^T$. The projections of scene point P onto the two rectified spherical images p_{sln} and p_{srn} can be represented as follows in term of (7):

$$p_{sln} \cong [R' \quad 0]\tilde{P}, \quad p_{srn} \cong [R' \quad -R't]\tilde{P} \tag{8}$$

According to the characteristics of the HVS, R' is determined as follows (see Fig. 4(b)).

1) Let X-axis be parallel to the line joining the two centers of images. Thus, we have:

$$r_1 = t \tag{9}$$

2) Let the vector sum of the unit vectors z_1 and z_2 of Z-axis of two original images be k, and r_2 corresponding to Y-axis is presented as follows:

$$r_2 = k \times r_1 \tag{10}$$

where

$$k = \frac{(z_1 + z_2)}{\|z_1 + z_2\|} \tag{11}$$

3) r_3 corresponding to Z-axis is determined based on r_1 and r_2, which is shown as follows:

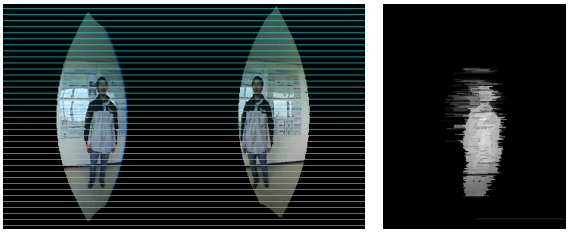
$$r_3 = r_1 \times r_2 \tag{12}$$

According to the method of [7], for (7), (8) we have:

$$p_{sln} = R' p_{slo}, \quad p_{srn} = R' R_{rl} p_{sro} \tag{13}$$

3.2.3 Acquisition of dense disparity map

The rectified CFOVRs are first transformed to longitude-latitude representation so that the epipolar lines of the CFOVRs are parallel to each other, as shown in Fig. 5(a). Then, the conventional area-correlation-based method is used to compute the disparity of the CFOVRs, as shown in Fig. 5(b). The brighter pixel has a bigger distance from the HVS. Using the disparity values, the 3D information of environments can be computed further according the disparity of spherical stereo defined in [7].



(a) (b)

Fig. 5. (a) The longitude-latitude image of the CFOVRs. (b) The dense disparity map is computed using the longitude-latitude image.

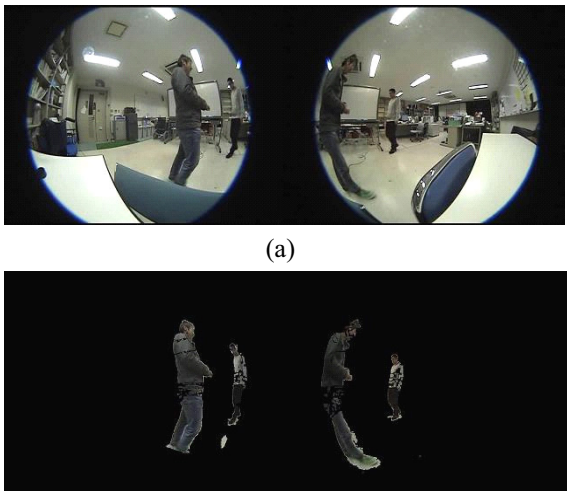
4. Scene Analysis by the HVS

The configuration of the proposed HVS has been introduced in Section 1 as shown in Fig. 2(a). The pair of hemispherical FOV fisheye images is acquired by the video cameras, Sony DCR-HC30, mounted with a fisheye conversion lens, Olympus FCON-02. The size of the captured fish-eye images is 640×480 pixels. A pair of sample images captured by the HVS is shown in Fig. 2(b).

In this paper, we present the preliminary experiment of scene analysis by the HVS as an application and also to test the effectiveness of the proposed method. The main characteristic of the experiment by the HVS system is that it can both perform motion detection by the omnidirectional monocular vision for near full FOV and motion estimation by binocular vision for a relative narrow FOV.

The processing is as follows:

1) Calibrate the intrinsic and extrinsic parameters of the pair of fisheye cameras in the HVS by the method in Section 3.1.
2) The motion detect is taken for the video captured by the whole vision of our system using background subtraction algorithm [9].
3) For every frame of the detected motion video above, the corresponding dense disparity map for the binocular vision is generated, using the algorithm in Section 3.2.
4) Then, the depth map is computed from each disparity map until the end of the video, which indicates the estimated location of the movement. Finally, we can get the whole depth video of detected moving object's motion.



(a)

(b)

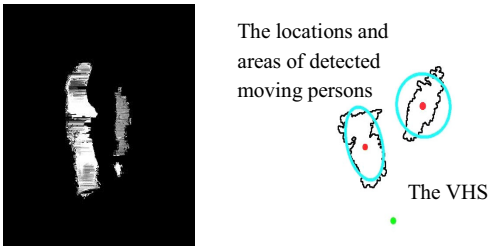
Fig. 6. (a) one pair of original images of the video captured by our system. (b) The result of motion detection.

Two person walking from opposite direction in front of the VHS are captured as moving objects in our experiment. We use the methods described above to deal with the captured video. Fig. 6 illustrates one pair of original frames and the detected persons. Afterwards the disparity map is computed as shown in Fig. 7(a). As the result of motion estimation, Fig. 7(b) shows the area of detected persons in a depth map. The centers of gravity and the location of the VHS are also indicated.

5. Conclusions

In this paper, a biologically-inspired vision system, the VHS, like horses' eyes is constructed. The algorithm of identifying the CFOVR and rectifying the CFOVRs for the binocular vision of the HVS is presented also. Finally, the preliminary experimental results of scene analysis based on the two vision modes of the HVS are presented to show effectiveness of the proposed method.

As well as more quantitative evaluation, applying the proposed HVS to a mobile robot which can cope with dynamic environments is our future work.



(a) (b)

Fig. 7. (a) The computed dense disparity map of the CFOVRs in the detected images in Fig. 6.(b). (b) The result of the depth map for motion estimation. Both the location of the VHS and the centroid of the detected persons are shown in the map.The areas of the persons detected are surrounded by black curve.

References

[1] B. Timney, T. Macuda, "Vision and hearing in horses," *Journal of the American Veterinary Medical Association*, vol. 218, no. 10, pp. 1567-1574, May. 2001.

[2] G. Waring, *Horse Behavior*. Norwich, NY: William Andrew Press, 2007, pp. 18-25.

[3] J. C. Bazin, et al, "Automatic scene structure and camera motion using a catadioptic system,"*Computer Vision and Image Understanding*, vol.109, pp.186-203, Feb. 2008.

[4] S. Li, "Monitoring around a vehicle by a full-view spherical image sensor", IEEE Trans. on ITS, Vol.7, No.4, pp.541-550, 2006.

[5] C. Chen, et al, "Cooperative mapping of multiple PTZ cameras in automated surveillance systems," in *Proc. CVPR,* 2009, pp. 1078-1084.

[6] J. Heller, T. Pajdla, "Stereoraphic rectification of omnidirectional stereo pairs," in *Proc. CVPR,* 2009, pp. 1414-1421.

[7] S. Li, "Binocular spherical stereo," *IEEE Trans. Intell. Transp. Syst.*, vol. 9 no. 4, pp. 589-600, Dec. 2008.

[8] S. Li, K. Fukumori, "Spherical stereo for the construction of immersive VR environments," in *Proc. IEEEVR Conf.*, 2005, pp. 217-222.

[9] A. Elgammal, et al, "Background and Foreground Modeling Using Nonparametric Kernel Density for Visual Surveillance," *Proceedings of the IEEE*, vol. 90, no.7, pp. 1151-1163, 2002.