

3D Image Sensor based on Parallax Motion

Barna Reskó^{1,2}, Dávid Herbay³, Péter Krasznai¹, Péter Korondi³

¹ Budapest Tech

² Computer and Automation Research Institute of the Hungarian Academy of Sciences

³ Budapest University of Technology and Economics

E-mail: rbarna@datatrans.hu

Abstract: For humans and visual animals vision it is the primary and the most sophisticated perceptual modality to get information about the surrounding world. Depth perception is a part of vision allowing to accurately determine the distance to an object which makes it an important visual task. Humans have two eyes with overlapping visual fields that enable stereo vision and thus space perception. Some birds however do not have overlapping visual fields, and compensate this lack by moving their heads, which in turn makes space perception possible using the motion parallax as a visual cue. This paper presents a solution using an opto-mechanical filter that was inspired by the way birds observe their environment. The filtering is done using two different approaches: using motion blur during motion parallax, and using the optical flow algorithm. The two methods have different advantages and drawbacks, which will be discussed in the paper. The proposed system can be used in robotics for 3D space perception.

Keywords: depth perception, parallax motion, rotating persicope, motion blur, optical flow

1 Introduction

Most herbivores, especially hoofed grazers, lack depth perception. Instead, they have their eyes on the side of the head, providing a panoramic, almost 360° view of the horizon - enabling them to notice the approach of predators from any direction. However both avian and mammalian predators have frontal eyes, allowing them to precisely judge distances when they pounce, or swoop down, onto their prey.

In modern terminology, stereopsis is depth perception from binocular vision through exploitation of the two slightly different projections of the world on the two retinas called parallax. Depth perception does indeed rely primarily on binocular vision, but it also uses many other monocular cues to form the final integrated perception. There are monocular cues that would be significant to a

‘one-eyed’ person, and more complex inferred cues, that require both eyes to be perceiving stereo while the monocular cues are noted. This ‘third’ group relies on processing within the brain of the person, as they see a full field of view with both eyes.

Stereopsis has been one of the most widely explored topics of 3D vision. A classical stereo system typically includes two cameras placed side by side to capture stereo image pairs. The depth information of the captured scene is calculated from the disparity map of these two images. Instead of using two or more cameras, one can also sequentially capture image pairs by repositioning a single camera. The advantage of using a single camera over the two-camera systems is that the identical intensity response of the stereo pairs captured with the same camera can improve the accuracy of correspondence matching.

Besides the approach of displacing a single camera, several other single-camera systems have also been investigated. Each of them has its own unique features. Adelson and Wang designed a plenoptic camera system which uses a lenticular array placed in front of the image plane to control the light path, and the depth information is extracted by analyzing a set of sub-images [5]. Lee and Kweon presented a bi-prism stereo camera system which forms the stereo pairs on the left and right halves of a single CCD sensor by a bi-prism [3]. Nene and Nayar investigated stereo imaging using different types of mirrors including planar, ellipsoidal, hyperboloidal and paraboloidal mirror [6]. The subject of this paper is a rotating periscope which is placed in the axis of the camera. By rotating the periscope in a vertical plane around the camera’s optical axis, the image of the scene follows a circular shift on the vertical plane. In the presented experiment, two images corresponding to two fixed poses of the plate are captured. Feature points are extracted by looking for zero-crossings on images filtered by LOG (Laplacian of Gaussian) filter. In the proposed idea the depth information is extracted from the comparison of a sharp and a blurred image.

This paper is organized as follows: in Section 2 the proposed concept is described. This is followed by the computer based simulations of the concept in Section 3. Section 4 introduces the opto-mechanical filter and other hardware implementation. The paper is closed by the test results and conclusions about the results and future perspectives.

2 The Proposed Concept of 3D Perception

2.1 Depth Perception with Single Camera

Although herbivores have sacrificed the depth perception for a much larger view range with eyes on the sides of the head, they are able to judge approximately distances in another way. Actually it is realized by the parallax method as in case of predators, but in this case it is a parallax motion, so the slightly shifted points of view are realized by the motion of the head.

A well know example is the cock, which moves his head forward and backward synchronously with its stepping.

Another example in the nature which seems similar to the method presented in this paper is the method of the ostrich, which has a special method for the depth perception. From time to time it does a circular motion with its head in a vertical plan with the aim of getting better depth information on the great plain. The longer is the baseline (the distance of points of view), the more precise depth information the bird can perceive. This allows to judge the distance in a greater range.

2.2 Two Ways to Calculate Depth from Parallax Motion

Two different approaches are proposed for depth perception, both of them based on the motion parallax visual cue.

2.2.1 Using Motion Blur

There are several techniques for stereo vision with 2 or more cameras. The drawback of these techniques is the cost due to the multiple of high quality cameras, the sophisticated programming of image processing and consequently a big hardware background. Several problems turn up in these techniques: opposition of image quality (resolution, colour depth and frame per second) and data transfer and processing ability. The motion blur is proposed, which can be obtained by relatively moving the image sensor and the projected image during exposure. This effect will be used to obtain depth information.

2.2.2 Using Optical Flow

Another way of calculating the depth is to calculate the optical flow between two consecutive frames taken from the scene. The length of the optical flow vectors will be inverse proportional to the distance of the projected 3D point.

2.3 The Geometry of Stereo Vision in Case of a Two-Camera System

Consider the images p and p' of a point P observed by two cameras with optical centres O and O' . These five points all belong to the *epipolar plane* defined by the two intersecting rays OP and $O'P$. In particular, the point p' lies on the line l' , where this plane and the retina P' of the second camera intersect. The line l' is the *epipolar line* associated with the point p , and it passes through the point e' where the baseline joining the optical centres O and O' intersects P' . Likewise, the point p lies on the epipolar line l associated with the point p' , and this line passes through the intersection e of the *baseline* with the plane P (Figure 1).

The points e and e' are called the *epipoles* of the two cameras. The epipole e' is the (virtual) image of the optical centre O of the first camera in the image observed by the second camera, and vice versa. As noted before, if p and p' are images of the same point, then p' must lie on the epipolar line associated with p . This *epipolar constraint* plays a fundamental role in stereo vision and motion analysis. Since the most challenging task in stereo vision is finding the correspondence between points on the two images, using the epipolar constraint for a stereo camera setup shrinks the search area for a point p being the pair of point p' from a plane to a line.

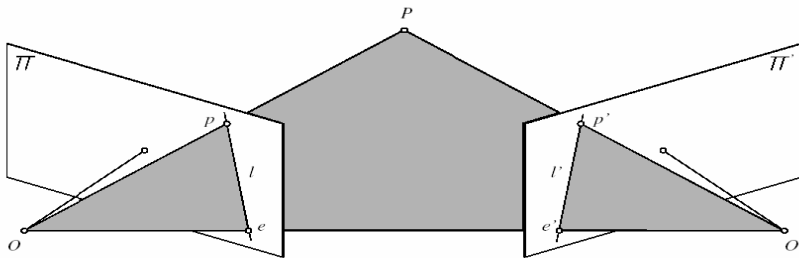


Figure 1

In the previous section it was shown that point P lies on the lines defined by Op and $O'p'$.

If the intrinsic and extrinsic parameters of the cameras are known, the above two lines determine the three-dimensional position of point P . The retrieval of three-dimensional information from two or more pieces of two-dimensional information is referred to as 3D reconstruction.

The reconstruction is realized by matrix transformations. There are several methods in which a common coordinate system is defined, in that coordinate system projection planes and the transformation matrix is to be defined for the 3D reconstruction.

But these are the vast calculations for a high resolution, great fps video.

The idea was inspired to get round this difficult image process by processing in an analog way.

2.4 Motion Blur

It is the apparent streaking of rapidly moving objects in a still image. When a camera creates an image, that image does not represent a single instant of time. Because of technological constraints or artistic requirements, the image represents the scene over a period of time. As objects in a scene move, an image of that scene must represent an integration of all positions of those objects, as well as the camera's viewpoint, over the period of exposure determined by the shutter speed. In such an image, any object moving with respect to the camera, will look blurred or smeared along the direction of relative motion. This smearing may occur on an object that is moving or on a static background if the camera is moving.

The mathematical form of the integration of the light captured on the sensor during the motion blur (Figure 2):

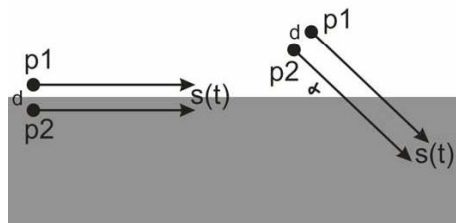


Figure 2

$$\left. \begin{array}{l} s_x(t) \\ s_y(t) \end{array} \right\} s(t) \text{ is the path of the point observed on the sensor}$$

$$m(p) = \int_0^t I(p + s(t)) dt, \text{ equation of the image of horizontal motion}$$

blur of any point is the time integration the light intensity on time with the motion of the point

$$m_\alpha(p) = \int_0^t I(p + s_\alpha(t)) dt, \text{ the same with an inclination } \alpha$$

The equation of the whole image is the integration of the equation 3 on the entire surface:

$$M_{\alpha}(x_c, y_c) = \int_{p \in A_{x_c, y_c}} m_{\alpha}(p) dp$$

This way by using motion blur effect an analog integration of points can be gained without using any computer.



Figure 3

We can see also that the edges are reinforced which are oriented to the direction of the motion blur and those are blurred which are oriented in other direction. So we can reinforce the contour of object in a certain orientation with a straight motion (Figure 3). But if we want to reinforce the contour in all directions then we have to apply a circular motion.

This leads to the final idea, that is by moving one camera on a circular path creates a blurred image with reinforced contours from which spatial information can be extracted based on that consideration that the closer the object is the more it is blurred, and so the further the object is the less it is blurred. It means the more intense the contour is the closer the object is.

It is already a technical consideration that instead of moving a camera, it is advisable to use an optical device creating a turning image. The most suitable such a device is the periscope. It shifts the image parallel as much as it is designed. Fortunately it does not rotate the image as it turns around just shift the view point and describe a circular path. In an idealistic case it does not cause even any image distortion.

An image may contain edges of different strengths. Applying the circular blur on the images may result in similar edge strength values, if the blurred edges were not similar and they were in a different distance. This explains why the resulting blurred image has to be compared to a reference image containing the non-blurred edge values (Figure 4).

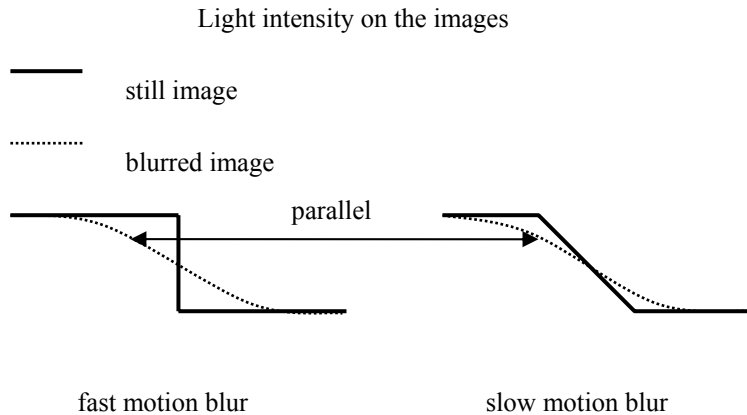


Figure 4

2.5 Optical Flow

Optical flow is the velocity field which warps one image into another. This can be described by a vector field, which contains vectors representing the relative position of corresponding pixels (features).

Estimating the optical flow is useful in pattern recognition, computer vision, and other image processing applications. It is closely related to motion estimation and motion compensation. Often the term optical flow is used to describe a dense motion field with vectors at each pixel, as opposed to motion estimation or compensation which uses vectors for blocks of pixels, as in video compression methods such as MPEG.

2.5.1 Optical Flow Estimation by Lucas-Kanade Method

The Lucas-Kanade method for optical flow is described in [8]. The problem statement of optical flow is as follows: let A and B be two 2D greyscale images. The two quantities $A(x) = A(x; y)$ and $B(x) = B(x; y)$ are then the greyscale values of the two images at the location $x = [x, y]^T$, where x and y are the two pixel coordinates of a generic image point x . The images A and B are discrete functions (or arrays), and the upper left corner pixel coordinate vector is $[0, 0]^T$. Let w and h be the width and height of the two images. Then the lower right pixel coordinate vector is $[w-1, h-1]^T$.

Consider an image point $u = [u_x, u_y]^T$ on the image A . The goal of feature tracking is to find the location $v = u + d = [u_x + dx, u_y + dy]^T$ on the image B . The vector $d = [dx, dy]^T$ is the image velocity at x , also known as the optical flow at x .

Because of the aperture problem, it is essential to define the notion of similarity in a 2D neighbourhood sense. Let w_{inx} and w_{iny} be two integers.

We define the image velocity d as being the vector that minimizes the residual function defined as follows:

EQUATION 1

$$\varepsilon(dx, dy) = \sum_{x=u_x-w_{inx}}^{u_x+w_{inx}} \sum_{y=u_y-w_{iny}}^{u_y+w_{iny}} (A(x, y) - B(x + dx, y + dy))^2$$

The similarity function is measured on a image neighbourhood of size $(2w_{inx} + 1) \times (2w_{iny} + 1)$. This neighbourhood will be also called integration window.

Minimising ε with least squares method

At the optimum, the first derivative of $\varepsilon(dx, dy)$ with respect to $d=(dx, dy)$ is zero:

EQUATION 2

$$\frac{\partial \varepsilon(d)}{\partial d} = [0, 0]$$

After expansion of derivate:

EQUATION 3

$$\frac{\partial \varepsilon(d)}{\partial d} = 2 \sum_{x=u_x-w_{inx}}^{u_x+w_{inx}} \sum_{y=u_y-w_{iny}}^{u_y+w_{iny}} (A(x, y) - B(x + dx, y + dy)) \begin{bmatrix} \frac{\partial B}{\partial x} & \frac{\partial B}{\partial y} \end{bmatrix}$$

Let now substitute $B(x+dx, y+dy)$ by its first order Taylor expansion about the point $d = [0, 0]^T$.

EQUATION 4

$$\frac{\partial \varepsilon(d)}{\partial d} \approx 2 \sum_{x=u_x-w_{inx}}^{u_x+w_{inx}} \sum_{y=u_y-w_{iny}}^{u_y+w_{iny}} (A(x, y) - B(x, y) - \begin{bmatrix} \frac{\partial B}{\partial x} & \frac{\partial B}{\partial y} \end{bmatrix} d) \begin{bmatrix} \frac{\partial B}{\partial x} & \frac{\partial B}{\partial y} \end{bmatrix}$$

The meaning of $A(x, y) - B(x, y)$ can be interpreted as the temporal image derivative at the point $[x \ y]^T$:

EQUATION 5

$$\delta I = A(x, y) - B(x, y)$$

The matrix $\begin{bmatrix} \frac{\partial B}{\partial x} & \frac{\partial B}{\partial y} \end{bmatrix}$ is the image gradient vector. Now make a new notation:

EQUATION 6

$$\nabla I = \begin{bmatrix} I_x \\ I_y \end{bmatrix} = \begin{bmatrix} \frac{\partial B}{\partial x} & \frac{\partial B}{\partial y} \end{bmatrix}$$

If we use central difference operator to make derivation, we can express I_x , I_y :

EQUATION 7

$$I_x(x, y) = \frac{B(x+1, y) - B(x-1, y)}{2}$$

EQUATION 8

$$I_y(x, y) = \frac{B(x, y+1) - B(x, y-1)}{2}$$

Following this new notation, the equation can be written:

EQUATION 9

$$\frac{1}{2} \frac{\partial s(d)}{\partial d} = \sum_{x=ux-winx}^{ux+winx} \sum_{y=uy-winy}^{uy+winy} (\delta I - \nabla I d) \nabla I$$

EQUATION 10

$$\frac{1}{2} \frac{\partial s(d)}{\partial d} = \sum_{x=ux-winx}^{ux+winx} \sum_{y=uy-winy}^{uy+winy} \left(\begin{bmatrix} \delta I I_x \\ \delta I I_y \end{bmatrix} - \begin{bmatrix} I_x^2 & I_x I_y \\ I_y I_x & I_y^2 \end{bmatrix} d \right)$$

Make new notations:

EQUATION 11

$$G = \sum_{x=ux-winx}^{ux+winx} \sum_{y=uy-winy}^{uy+winy} \begin{bmatrix} I_x^2 & I_x I_y \\ I_y I_x & I_y^2 \end{bmatrix} \quad \text{and} \quad b = \sum_{x=ux-winx}^{ux+winx} \sum_{y=uy-winy}^{uy+winy} \begin{bmatrix} \delta I I_x \\ \delta I I_y \end{bmatrix}$$

Then the equation can be written:

EQUATION 12

$$\frac{1}{2} \frac{\partial s(d)}{\partial d} = b - Gd \quad \text{and at optimum} \quad d = -G^{-1}b$$

This equation can be solved if G is invertible. This means that the gradient is large enough under point u , which is the corresponding position on the other image we are looking for.

Problems with this approximation

This method has two key assumptions:

- **colour constancy:** a point in frame A looks the same in B (for greyscale images, this is brightness constancy)
- **small motion:** points do not move very far

By using first order Taylor series we can track only small motion between frames. By using image gradients we are assuming, that corresponding pixels have the same intensity.

3 Computer Simulations of Motion Blur-based Depth Perception

In the first step with the purpose of ensuring these theories, a computer simulation was elaborated.

With a Matlab based program, two different distances were simulated by applying two different motion radiuses on a given image.

After applying the Sobel operator and comparing the changes in intensity, the following results were obtained.

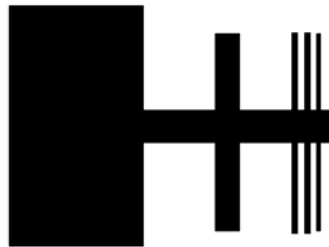


Figure 5

The Figure 5 shows the still, reference image of the original test object.

The first motion blur was made on a circle of a radius 10 units (Figure 7), the second on a radius of 30 units (Figure 6) which are equivalent with two different distances from the camera.

On the right of the Figures 6 and 7 the index numbers show the difference of the light intensity changes.

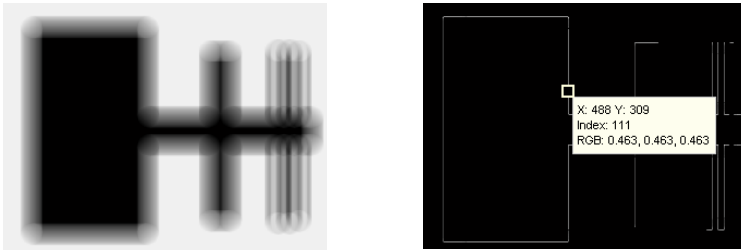


Figure 6

After checking several point pairs on the two results, we got that the index number is always larger in the second case. This means that the proposed method is able to compare distances. We can say that the theory is proved by the simulation.

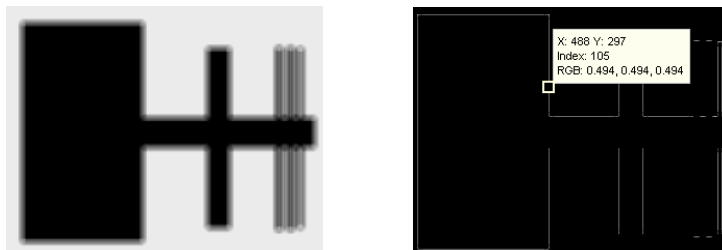


Figure 7

4 The Rotating Periscope

A mechanical frame was designed to implement the rotating periscope concept. During the design a compromise had to be made between the length of the periscope and the angular width of the field of view. The result was a 20 degrees symmetrical field of view and a periscope length 68 millimeters. This length has been chosen because it coincides the average distance between the eyes of a human.

The camera is positioned in front of the periscope, which is mounted using a large diameter bearing around the optical axis of the camera. This allows the periscope to be rotated using an external motor, and also reduces vibration and any other mechanical noise.

4 mm thick mirrors have been used, which have a very low distortion and bending radius. The mirrors have to be parallel with a very low error causing a disalignment of the image less than one pixel.

The drawings and rendered view of the periscope are shown in Figure 8, while the implemented

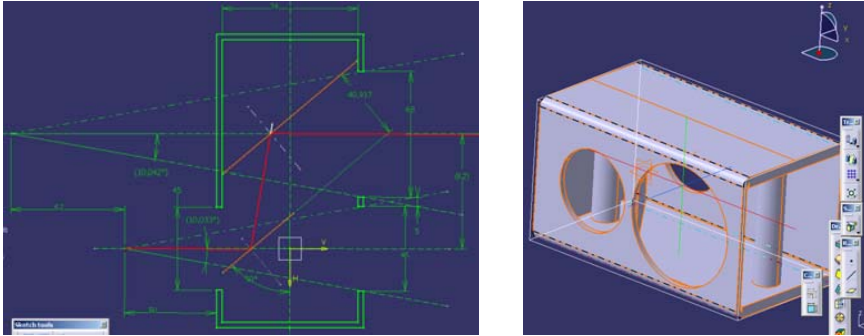


Figure 8



Figure 9

5 Testing and Evaluation

5.1 The Motion Blur-based Method

To set up the test, it is only needed to give the supply voltage for the motor and the camera and connect the camera to a PC by a fire-wire cable (Figure 10).

On the camera's lens the aperture and the focal distance can be adjusted. The user interface supplied for the camera the exposure time, the resolution, the colour depth, etc. are adjustable.



Figure 10

At the first experiments the mechanism worked well, but the image was distorted. The point of the test was that if the camera looks at an infinite distance (600 meters across the river) the image should not change at all. But the result was a fluctuating distortion of the image. It was obviously the consequence of the too low quality mirrors. This way the evaluation of the image is not possible because the blurred image is not a consequence of the relative motion.

Evidently the mirrors were to be changed.

After changing to 4 mm thick much more rigid mirrors the result was better, not distorted, but still not perfect. The test was the same as previously, a still image was expected to appear. The resulting image was however blurred again (Figure 11). It was realized that there was also a misalignment of the mirrors. If they are not perfectly parallel then the axis of the periscope does not coincide with the axis of the camera and the rotation axis, so the image is blurred proportionally to the misalignment, and not to the displacement of the view point.

Short estimation of the misalignment: the distance from the buildings on the other side of the river is about 700 m, the relative displacement of them is about 2 m, $\arctan(2/700)=0,1$ degree. So the periscope is to be redesigned a bit to be able to regulate the inclination of the mirrors. Until the submission of this paper it was not worked out.

After designing and fabricating the periscope the theoretical conception was tested. It was realized that the concept is relatively very sensitive to the precision of the optical device but this precision requirement is still far from the precision of general optical device like for e.g. a handy cam. It can be stated that the image quality depends from the applied camera, and the image processing is faster than in case of the binocular stereo matching techniques.



Figure 11

If good test results will be obtained in the near future, then this concept has a great chance to be worked-out in a more compact and applicable form, and has a good chance for interest in the industrial applications, or for e.g. intelligent space applications and robotics.

5.2 The Optical Flow-based Method

The method using optical flow has been implemented in a C++ software. Figures 12-14 show the obtained results. The first three images were taken with a camera, below left shows the optical flow vectors, and on the right their length is encoded in color, blue meaning the far points, red meaning the close points.



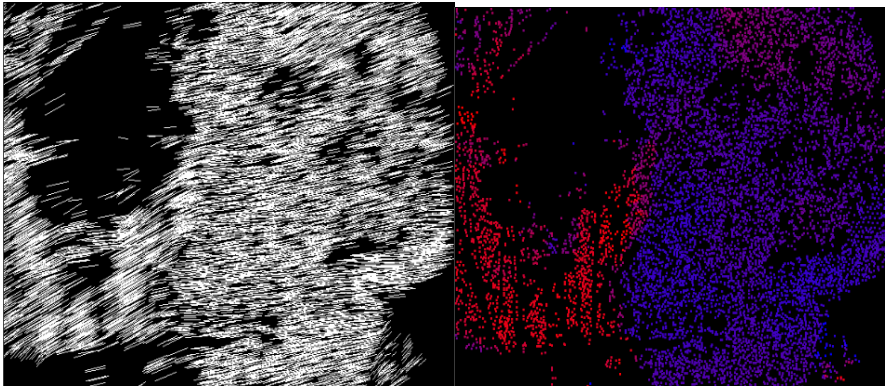


Figure 12

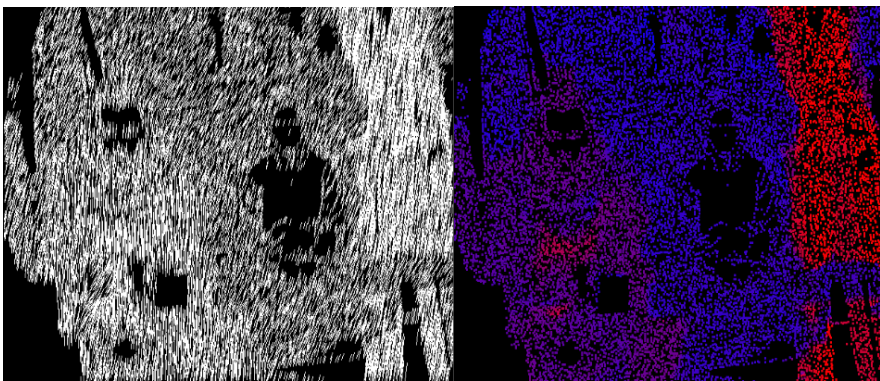


Figure 13

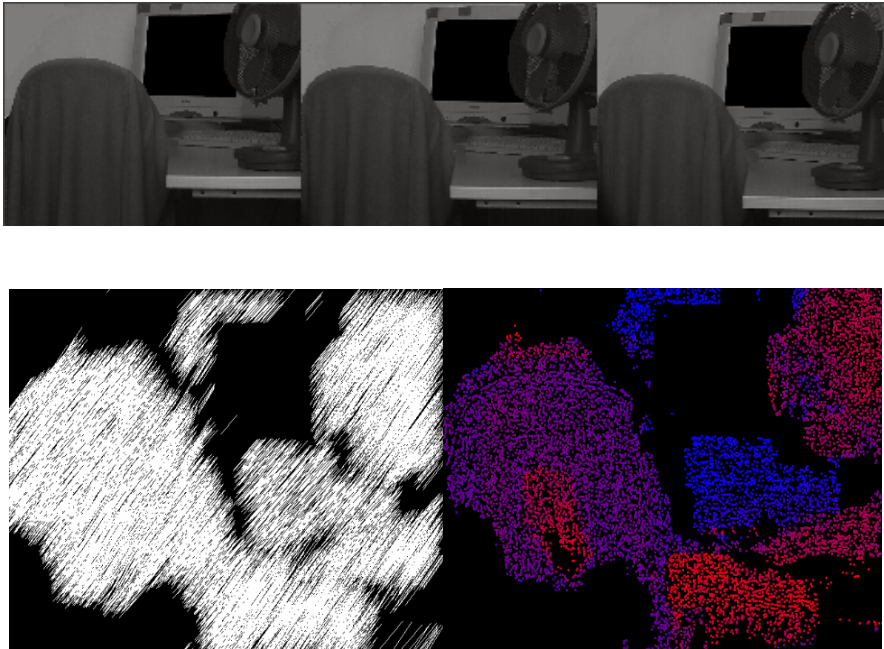


Figure 14

Conclusion

The obtained results show that the proposed methods are able to extract spatial information from motion parallax cues. The motion blur based method is computationally very unexpensive, but it is very sensible to the accuracy of the optical devices installed in the hardware equipment. The optical flow based method is on the other hand very robust to inaccuracies of the mirrors, but computationally more complex. This complexity is due to the optical flow algorithm. The applicability of the proposed method could be enhanced by using special hardware based optical flow calculation devices. Such devices have recently emerged on the market, and the authors intend to test their results using such hardware tools.

Acknowledgement

The authors wish to thank the National Science Research Fund (OTKA K62836) for their financial support, and the Rancz Ltd. for the support in manufacturing the periscope presented in this paper.

References

- [1] SKF.com, <http://www.skf.com/portal/skf/home>
- [2] Introduction to stereo vision system, <http://www.cs.mcgill.ca/~gstamm/P644/stereo.html>

- [3] Doo Hyun Lee, In So Kweon, Roberto Cipolla, "A Biprism-Stereo Camera System," *cvpr*, p. 1082, 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'99) 1, 1999
- [4] Stereovision, <http://axon.physik.uni-bremen.de/research/stereo/>
- [5] Adelson, E. H., Wang, J. Y. A., "Single Lens Stereo with a Plenoptic Camera", *PAMI*,14(2), pp. 99-106, 1992
- [6] S. A. Nene, S. K. Nayar. Stereo with Mirrors. In Proc. ICCV, pp. 1087-1094, 1998
- [7] Barna Reskó. Optical Motion Tracking for Industrial Robot Programming, Master's thesis, 2004
- [8] Lucas B D, Kanade: An Iterative Image Registration Technique with an Application to Stereo Vision