Panoramic Depth Imaging: Single Standard Camera Approach

PETER PEER AND FRANC SOLINA

University of Ljubljana, Faculty of Computer and Information Science, Tržaška 25, 1001 Ljubljana, Slovenia

peter.peer@fri.uni-lj.si

franc.solina@fri.uni-lj.si

Abstract. In this paper we present a panoramic depth imaging system. The system is mosaic-based which means that we use a single rotating camera and assemble the captured images in a mosaic. Due to a setoff of the camera's optical center from the rotational center of the system we are able to capture the motion parallax effect which enables stereo reconstruction. The camera is rotating on a circular path with a step defined by the angle, equivalent to one pixel column of the captured image. The equation for depth estimation can be easily extracted from the system geometry. To find the corresponding points on a stereo pair of panoramic images the epipolar geometry needs to be determined. It can be shown that the epipolar geometry is very simple if we are doing the reconstruction based on a symmetric pair of stereo panoramic images. We get a symmetric pair of stereo panoramic images when we take symmetric pixel columns on the left and on the right side from the captured image center column. Epipolar lines of the symmetrical pair of panoramic images are image rows. The search space on the epipolar line can be additionally constrained. The focus of the paper is mainly on the system analysis. Results of the stereo reconstruction procedure and quality evaluation of generated depth images are quite promising. The system performs well for reconstruction of small indoor spaces. Our finall goal is to develop a system for automatic navigation of a mobile robot in a room.

Keywords: stereo vision, reconstruction, panoramic image, mosaicing

1. Introduction

1.1. Motivation

Standard cameras have a limited field of view, which is usually smaller than the human field of view. Because of that people have always tried to generate images with a wider field of view, up to a full 360 degrees panorama.

Under the term stereo reconstruction we understand the generation of depth images from two or more captured images of the scene. A depth image is an image that stores distances to points on the scene. We wish that all points and lines on the scene are visible on all images of the scene. This is the property of panoramic cameras and it presents our prime motivation.

If we tried to build two panoramic images simultaneously by using two standard cameras, we would have problems with non-static scenes. Clearly, one camera would capture the motion of the other camera. So we decided to use only one camera. Our finall goal is to develop a system for autonomous navigation of a mobile robot in a room.

1.2. Basics About the System

In the small photograph within Fig. 1 the hardware part of our system can be seen: a color camera is mounted on a rotational robotic arm so that the optical center of the camera is offset from the vertical axis of rotation. The camera is looking outward from the system's rotational center. Panoramic images are generated by repeatedly shifting the rotational arm for the angle which corresponds to one column of the captured image. By assembling the center columns of these images, we get a mosaiced panoramic image.
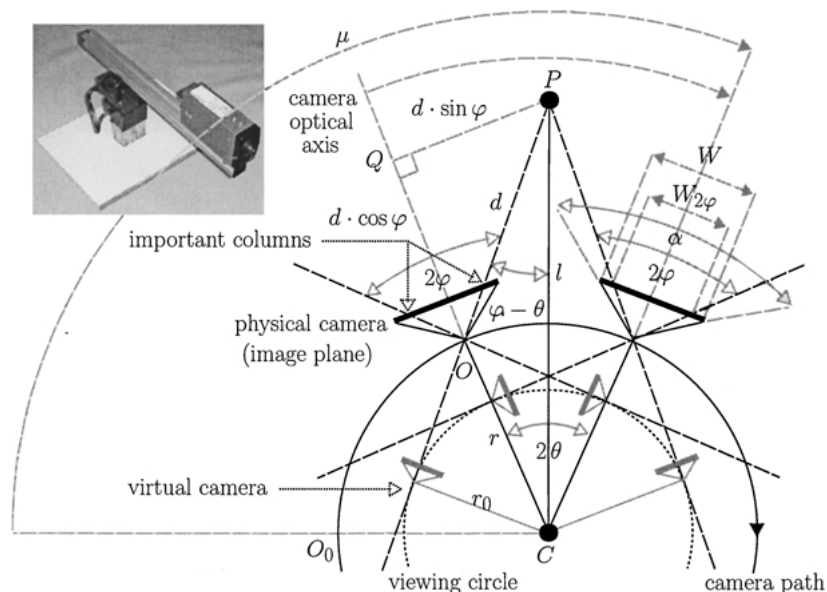
Figure 1.   Geometry of our system for constructing multiperspective panoramic images. Note that a ground-plan is presented the viewing circle extends in 3D to the viewing cylinder. The optical axis of the camera is kept horizontal. In the small photograph is shown the hardware part of the system.

It can be shown that the epipolar geometry is very simple if we are doing the reconstruction based on a symmetric pair of stereo panoramic images. We get a symmetric pair of stereo panoramic images when we take symmetric columns on the left and on the right side from the captured image center column and assemble them in a mosaiced stereo pair.

1.3.   Structure of the Paper

In the next section we give an overview of related work and expose the contribution of our work towards the discussed subject. Section 3 describes the geometry of our system, Section 4 is devoted to the epipolar geometry and Section 5 is about stereo reconstruction. The focus of this paper is on the analysis of system capabilities, given in Section 6. In Section 7 we present some experimental results. At the very end of the paper we summarize the main conclusions and reveal some ideas for future work.

2.   Related Work

One of the best known commercial packages for creating mosaiced panoramic images is QTVR (Quick-Time Virtual Reality). It works on the principle of sewing together a number of standard images captured while rotating the camera (Chen, 1995). Peleg et al. (2000) introduced the method for creation of mosaiced panoramic images from standard images captured with a handheld video camera. A similar method was suggested by Szeliski and Shum (1997), which also does not strictly constraint the camera path to a circle but assumes that there a great motion parallax effect is not present. All methods mentioned so far are used only for visualization purposes since the authors did not try to reconstruct the scene.

Peleg and Ben-Ezra (1999) and Peleg et al. (2000, 2001) introduced a method for creation of stereo panoramic images. Stereo panoramic images are created without actually computing the 3D structure—the depth effect is created in the viewer's brain.

Ishiguro et al. (1992) suggested a method which enables the reconstruction of the scene. They used a standard camera which rotated around its optical center. The scene is reconstructed by means of mosaicing together panoramic images from the central column of the captured images and moving the system to another location where the task of mosaicing is repeated. The two created panoramic images are then used as an input in a stereo reconstruction procedure. They also suggested the method described by Shum and Szeliski (1999), but in their case they searched for corresponding points by tracking the features on the scene.

Shum and Szeliski (1999) described two methods used for creation of panoramic depth images from the single center of rotation. Both methods are based on moving the camera on a circular path. Panoramic images are built by taking one pixel column out of a captured image and mosaicing the columns. They call such panoramic images *multiperspective panoramic images*. The crucial property of two or more multiperspective panoramic images is that they capture the information about the motion parallax effect, while the columns contributing to the panoramic images are captured from different perspectives. The authors are using such panoramic images as an input in a stereo reconstruction procedure. They searched for corresponding points using an upgraded *plane sweep stereo* procedure.

However, multiperspective panoramic images are not something new to vision community (Shum and Szeliski, 1999): they are a special case of *multiperspective panoramic images for cel animation* (Wood et al., 1997), they are very similar to images generated with a procedure called *multiple-center-of-projection* (Rademacher and Bishop, 1998), to *manifold projection* procedure (Peleg et al., 2000) and to *circular projection* procedure (Peleg and Ben-Ezra, 1999; Peleg et al., 2000, 2001). The principle of constructing multiperspective panoramic images is also very similar to *the linear pushbroom camera* principle for creating panoramic images (Gupta and Hartley, 1997).

In papers closest to our work (Ishiguro et al., 1992; Shum and Szeliski, 1999) we missed two things: an analysis of system's capabilities and searching for corresponding points using the standard correlation technique and epipolar constraint. Therefore the focus of this paper is on these two issues.

3. System Geometry

Let us begin this section with a description of how the stereo panoramic pair is generated. From the captured images on the camera's circular path we always take only two pixel columns which are equally distant from the middle column. The column on the right side of the captured image is then mosaiced in the left eye panoramic image and the column on the left side of the captured image is mosaiced in the right eye panoramic image. So, we are building panoramic image from only one column of the captured image. Thus, we get a symmetric pair of panoramic images.

The geometry of our system for creating multiperspective panoramic images is shown in Fig. 1.

Panoramic images are then used as an input to create panoramic depth images. Point C denotes the system's rotational center around which the camera is rotated. The offset of the camera's optical center from the rotational center C is denoted as r describing the radius of the circular path of the camera. The camera is looking outward from the rotational center. The optical center of the camera is marked with O. The column of pixels that is sewn in the panoramic image contains the projection of point P on the scene. The distance from point P to point C is the depth l and the distance from point P to point O is denoted with d. θ is the angle between the line defined by point C and point O and the line defined by point C and point P. In the panoramic image the horizontal axis presents the path of the camera. The axis is spanned by μ and defined by point C, a starting point O_0 where we start capturing the panoramic image and the current point O. With φ we denote the angle between the line defined by point O and the middle column of pixels of the image captured by the physical camera looking outward from the rotational center (this column contains the projection of the point Q), and the line defined by point O and the column of pixels that will be mosaiced in panoramic image (this column contains the projection of the point P). Angle φ can be thought of as a reduction of the camera's horizontal view angle α.

The system in Fig. 1 is obviously a non-central since the light rays forming the panoramic image are not intersecting in one point called the viewpoint, but instead are tangent ($\varphi \neq 0$) to a cylinder with radius r_0 called the viewing cylinder. Thus, we are dealing with panoramic images formed by a projection from a number of viewpoints. This means that a captured point on the scene will be seen in the panoramic image only from one viewpoint.

For stereo reconstruction we need two images. If we are looking at only one circle on the viewing cylinder (Fig. 1) then we can conclude that our system is equivalent to a system with two cameras. In our case two virtual cameras are rotating on a circular path, i.e. viewing circle, with radius r_0. The optical axis of a virtual camera is always tangent to the viewing circle. The panoramic image is generated from only one pixel from the middle column of each image captured by a virtual camera. This pixel is determined by the light ray which describes the projection of the scene point on the physical camera image plane. If we observe a point on the scene P, we see that both virtual cameras
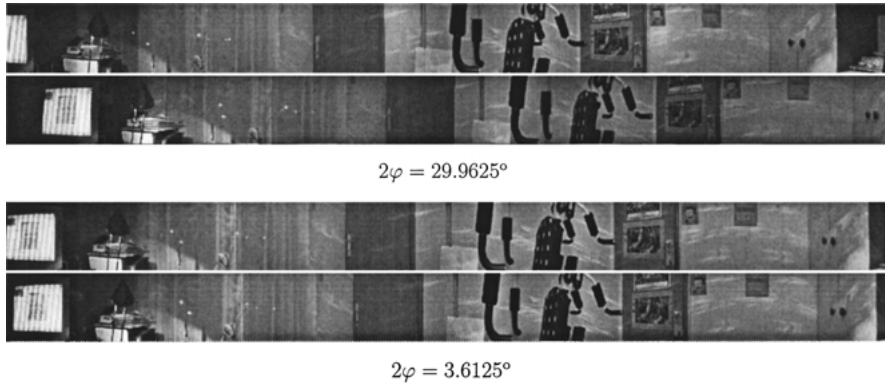
Figure 2. Two symmetric pairs of panoramic images which were generated with the different values of the angle φ. In Section 6.1 we explain where these values for the angle φ come from. Each symmetric pair of panoramic images comprises the motion parallax effect. This fact enables the stereo reconstruction.

which see this point, form a traditional stereo system of converging cameras.

Two images differing in the angle of rotation of the physical camera setup (for example two image planes marked in Fig. 1) are used to simulate a bunch of virtual cameras on the viewing cylinder. Each column of the panoramic image is obtained from different position of the physical camera on a circular path. In Fig. 2 we present two symmetric pairs of panoramic images.

To automatically register captured images directly from knowing the camera's viewing direction, the camera lens' horizontal view angle α and vertical view angle β are required. If we know this information, we can calculate the resolution of one angular degree, i.e. we can calculate how many columns and rows are within an angle of one degree. The horizontal view angle is especially important in our case, since we are moving the rotational arm only around it's vertical axis. To calculate these two parameters, we use an algorithm described in Prihavec and Solina (1998). It is designed to work with cameras where zoom settings and other internal camera parameters are unknown. The algorithm is based on the mechanical accuracy of the rotational arm; the basic step of our rotational arm corresponds to an angle of $0.051428\bar{5}°$.

We were using the camera with horizontal view angle $\alpha = 34°$ and vertical view angle $\beta = 25°$. In the process of the construction of panoramic images we did not vary this two parameters.

4. Epipolar Geometry

Searching for the corresponding points in two images is a difficult problem. In general, the corresponding point can be anywhere on the second image. That is why we would like to constraint the search space as much as possible. With the epipolar constraint we reduce the search space from 2D to 1D, i.e. to an epipolar line (Faugeras, 1993). In Section 6.2 we prove that in our system we can effectively reduce the search space even on the epipolar line.

In this section we will only illustrate the procedure of the proof that epipolar lines of the symmetrical pair of panoramic images are image rows (Huang and Pajdla, 2000; Huang et al., 2001; Shum and Szeliski, 1999). This statement is true for our system geometry. For proof see Huang and Pajdla (2000), Huang et al. (2001).

The proof is based on radius r_0 of the viewing cylinder (Fig. 1). We can express r_0 in terms of known quantities r and φ as: $r_0 = r \cdot \sin \varphi$.

We carry out the proof in three steps: *first*, we have to execute the projection equation for the line camera, *then* we have to write the projection equation for multiperspective panoramic image and in the *final* step we prove the property of epipolar lines for the case of a symmetrical pair of panoramic images. In the first step we are interested in how the point on the scene is projected to the camera's image plane (Faugeras, 1993), which has in our case, since we are dealing with a line camera, a dimension of $n \times 1$ pixels. In the second step, we have to write the relation between different notations of a point on the scene and the projection of this point on the panoramic image: notation of the scene point in Euclidean coordinates of the world coordinate system and in cylindrical coordinates of the world coordinate system, notation of the projected point in angular coordinates of the (2D) panoramic image coordinate system and in pixel coordinates of the (2D)

panoramic image coordinate system. When we know the relations between the mentioned coordinate systems, we can write the equation for projection of scene points on the cylindrical image plane of the panorama. Based on angular coordinates of the panoramic image coordinate system property, we can in the third step show that the epipolar lines of the symmetrical pair of panoramic images are actually rows of panoramic images. The basic idea for the last step of the proof is as follows:

If we are given an image point on one panoramic image, we can express the optical ray defined by a given point and the optical center of the camera in 3D world coordinate system. If we project this optical ray described in world coordinate system on the second panoramic image, we get an epipolar line corresponding to the given image point in the first panoramic image.

5.   Stereo Reconstruction

Let us go back to Fig. 1. Using trigonometric relations evident from the sketch we can write the equation for depth estimation l of point P on the scene. Using the basic law of sines for triangles, we have:

$$\frac{r}{\sin(\varphi - \theta)} = \frac{d}{\sin \theta} = \frac{l}{\sin(180° - \varphi)},$$

and from this equation we can express the equation for depth estimation l as:

$$l = \frac{r \cdot \sin(180° - \varphi)}{\sin(\varphi - \theta)} = \frac{r \cdot \sin \varphi}{\sin(\varphi - \theta)}. \qquad (1)$$

From Eq. (1) follows that we can estimate depth l only if we know three parameters: r, φ and θ. r is given. Angle φ can be calculated with regard to camera's horizontal view angle α as:

$$2\varphi = \frac{\alpha}{W} \cdot W_{2\varphi}, \qquad (2)$$

where W is the width of the captured image in pixels and $W_{2\varphi}$ is the width of the captured image between columns contributing to the symmetrical pair of panoramic images, given also in pixels. To calculate the angle θ we have to find corresponding points on panoramic images. Our system works by moving the camera for the angle corresponding to one column of captured image. If we denote this angle with θ_0, we can

write angle θ as:

$$\theta = dx \cdot \frac{\theta_0}{2}, \qquad (3)$$

where dx is the absolute value of difference between corresponding points image coordinates on horizontal axis x of the panoramic images.

We are using a procedure called *normalized correlation* to search for corresponding points (Faugeras, 1993), because it is one of the most commonly used technique for searching the corresponding point. Paar and Pölzleitner (1992) described other interesting methods than just those based on correlation.

Procedure of the normalized correlation works on the principle of similarity of scene parts within two scene images. The basic idea of the procedure is to find the part of the scene on the second image which is most similar to the given part of the scene in the first image. The procedure is using a window within which the similarity is measured with help of the correlation technique.

To increase the confidence in estimated depth we are using the procedure called *back-correlation* (Faugeras, 1993). The main idea of this procedure is to first find a point $\mathbf{m}_2$ in the second image which corresponds to a point $\mathbf{m}_1$ given in the first image. Then we have to find the corresponding point for the point $\mathbf{m}_2$ in the first image. Let us mark this corresponding point with $\mathbf{m}_1'$. If the point $\mathbf{m}_1$ is equal to the point $\mathbf{m}_1'$ then we keep the estimated depth value. Otherwise, we do not keep the estimated depth value. This means that the point $\mathbf{m}_1$ for which the back-correlation was not successful has no depth estimation associated with it in the depth image. With back-correlation we are also solving the problem of occlusions.

6.   Analysis of System Capabilities

6.1.   Time Complexity of Creating a Panoramic Image

The biggest disadvantage of our system is that it can not produce panoramic images in real time since we are creating them by rotating the camera for a very small angle. Because of mechanical vibrations of the system, we also have to be sure to capture an image when the system is completely still. The time that the system needs to create a panoramic image is much too long to make it work in real time.

In one circle around the system's vertical axis our system constructs 11 panoramic images: 5 symmetric pairs and a panoramic image from the middle columns of the captured images. It captures 1501 images with resolution of 160×120 pixels, where radius is $r = 30$ cm and the shift angle is $\theta_0 = 0.2°$. We can not capture $360/0.2 = 1800$ images because of the limitation of the rotational arm. The rotational arm can not turn for 360 degrees around it's vertical axis.

The horizontal view angle of our camera was $34°$. The middle column of the captured image was in our case the 80th column. We assume that the middle column that we are referring to in this paper, is the middle column of the captured image. The distances between the columns contributing to symmetric pairs of panoramic images were 141, 125, 89, 53 and 17 columns. These numbers include two columns that contribute to each pair. This means that the value of angle 2φ (Eq. (2)) corresponds to $29.9625°$ (141 columns), $26.5625°$ (125 columns), $18.9125°$ (89 columns), $11.2625°$ (53 columns) and $3.6125°$ (17 columns).

The acquisition process takes a little over 15 minutes on PC Intel PII/350 MHz. The steps of the acquisition process are as follows: 1) Move the rotational arm to it's initial position. 2) Capture the image. 3) Contribute image parts to the panoramic images. 4) Move the arm to the new position. 5) Check in the loop if the arm is already in the new position. The communication between the program and the arm is written in a file for debugging purposes. After the program exits the loop, it waits for 300 ms. This is done in order to stabilize the arm in the new position. 6) Repeat steps 2 to 5 until the last image is captured. 7) When the last image is captured, contribute image parts to the panoramic images and save them.

We could achieve faster execution since our code is not optimized. For example, we did not optimize the time of waiting (300 ms) after the arm moves to a new position. No computing is done in parallel.

6.2. Constraining the Search Space on the Epipolar Line

Knowing that the width of the panoramic image is much bigger than the width of the captured image, we would have to search for a corresponding point along a very long epipolar line (Fig. 3(a)). Therefore we would like to constraint the search space on the epipolar line as much as possible. A side effect is also an increased confidence in estimated depth.

If we consider Eq. (1) we can ascertain two things, which nicely constraint the search space:

We have to wait until the point imaged in the column contributing to the left eye panorama does not move in time to the column contributing to the right eye panorama. If θ_0 presents the angle for which the camera is shifted, then $2\theta_{\min} = \theta_0$. This means that we
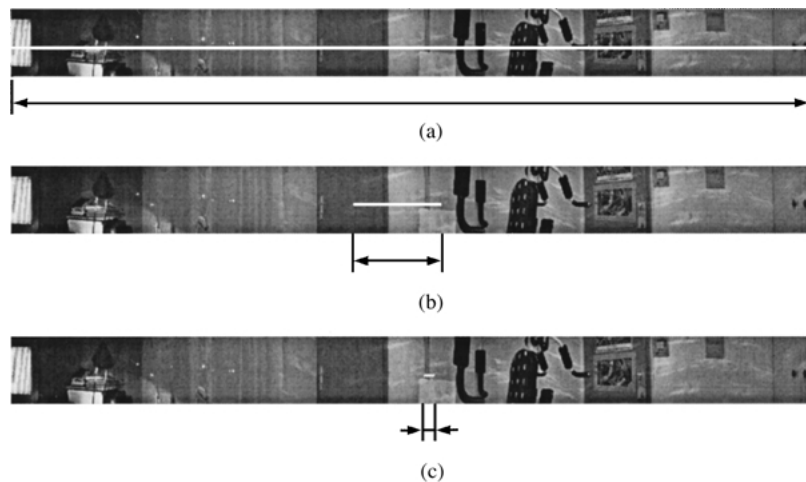

Figure 3. We can effectively constraint the search space on the epipolar line. (a) Unconstrained length of the epipolar line: 1501 pixels; (b) constrained length of the epipolar line: 149 pixels, $2\varphi = 29.9625°$; (c) constrained length of the epipolar line: 18 pixels, $2\varphi = 3.6125°$.

have to make at least one basic shift of the camera to get a scene point projected in a right column of the captured image contributing to the left eye panorama, to be seen in the left column of the captured image contributing to the right eye panorama. Based on this fact, we can search for the corresponding point in the right eye panorama starting from the horizontal image coordinate $x + \frac{2\theta_{min}}{\theta_0} = x + 1$ forward, where x is the horizontal image coordinate of the point on the left eye panorama for which we are searching the corresponding point. Thus, we get value $+1$ since the shift for angle θ_0 describes the shift of the camera for one column of the captured image.

Theoretically, the estimation of depth is not constrained upwards, but from Eq. (1) it is evident that the denominator must be non-zero. We can write this fact as: $\theta_{max} = n \cdot \frac{\theta_0}{2}$, where $n = \varphi$ div $\frac{\theta_0}{2}$ and φ mod $\frac{\theta_0}{2} \neq 0$.

If we write the constraint for the last point, which can be a corresponding point on the epipolar line, in analogy with the case of determining the starting point, which can be a corresponding point on the epipolar line, we have to search for corresponding point on the right eye panorama to including horizontal image coordinate $x + \frac{2\theta_{max}}{\theta_0} = x + n$. x is the horizontal image coordinate of the point on the left eye panorama for which we are searching the corresponding point.

In the following sections we will show that we can not trust the depth estimates near the last point of the epipolar line search space, but we have proven that we can effectively constraint the search space.

To illustrate the use of specified constraints on real data, let us write the following example which describes the working process of our system: while the width of the panorama is 1501 pixels, we have to check only $n = 149$ pixels in case of $2\varphi = 29.9625°$ (Figs. 3(b) and 4) and only $n = 18$ pixels in case of $2\varphi = 3.6125°$ (Fig. 3(c)), when searching for corresponding point.

From the last paragraph we could conclude that the stereo reconstruction procedure is much faster for a smaller angle φ. But we will show in the next section that a smaller angle φ, unfortunately, has also a negative property.

6.3. Meaning of the One-Pixel Error in Estimation of the Angle θ

Let us first define what we mean under the term one-pixel error. The images are discrete. Therefore, we would like to know what is the value of the error in the depth estimation if we miss the right corre-
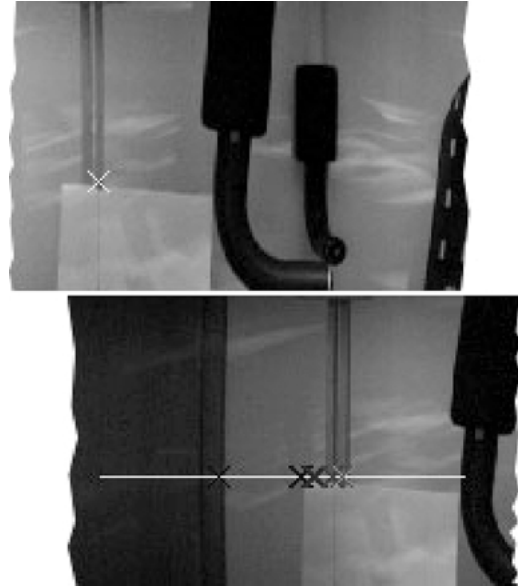

Figure 4. Constraining the search space on the epipolar line in the case of $2\varphi = 29.9625°$. On the left eye panorama (top image) we denoted the point for which we are searching the corresponding point with the cross. On the right eye panorama (bottom image) we marked with the same color the part of the epipolar line on which the corresponding point must lie. The best corresponding point is marked with the lightest cross.

sponding point for only a pixel. And we would like to have this information for various values of the angle φ.

Before we illustrate the meaning of the one-pixel error in estimation of the angle θ, let us take a look at functions in Fig. 5. The functions are showing the dependence of depth l from angle θ for two different values of angle φ. It is evident that depth l is rising slower in case of a bigger angle φ. This property decreases the error in depth estimation l when using bigger angle φ, but this decrease in the error becomes even more evident if we know that the horizontal axis is discrete and the intervals on the axis are $\frac{\theta_0}{2}$ degrees wide. If we compare the width of the interval on both graphs with respect to the width of interval that θ is defined on ($\theta \in [0, \varphi]$), we can see that the interval, whose width is $\frac{\theta_0}{2}$ degrees, is much smaller when using bigger angle φ. This subsequently means that the one-pixel error in estimation of the angle θ is much smaller when using bigger angle φ, since a shift for the angle θ_0 describes the shift of the camera for one column of pixels.

Because of a discrete horizontal axis θ (Fig. 5), with intervals which are $\frac{\theta_0}{2}$ degrees wide (in our case $\theta_0 = 0.2°$), the number of possible depth estimation
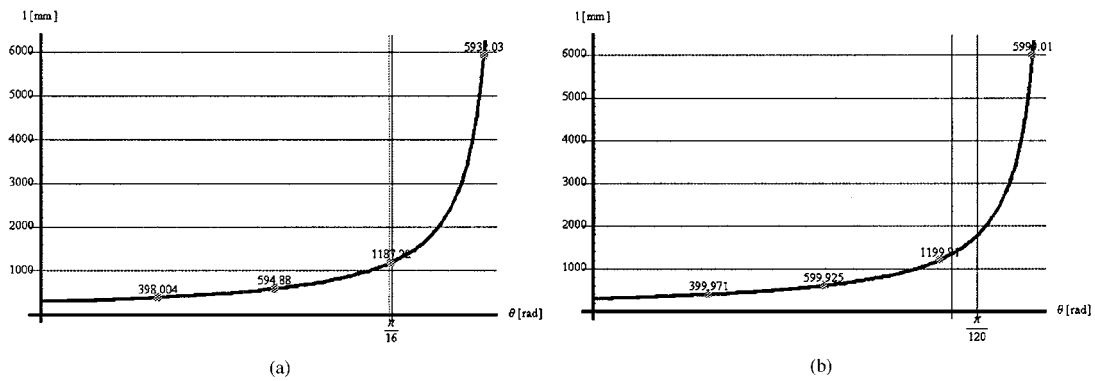
Figure 5. Graphs showing dependence of depth function l from the angle θ while radius $r = 30$ cm and using different values of the angle φ. To ease the comparison of the one-pixel error in estimation of the angle θ we showed the interval of width $\frac{\theta_0}{2} = 0.1°$ between the vertical lines around the third point. (a) $2\varphi = 29.9625°$; (b) $2\varphi = 3.6125°$.
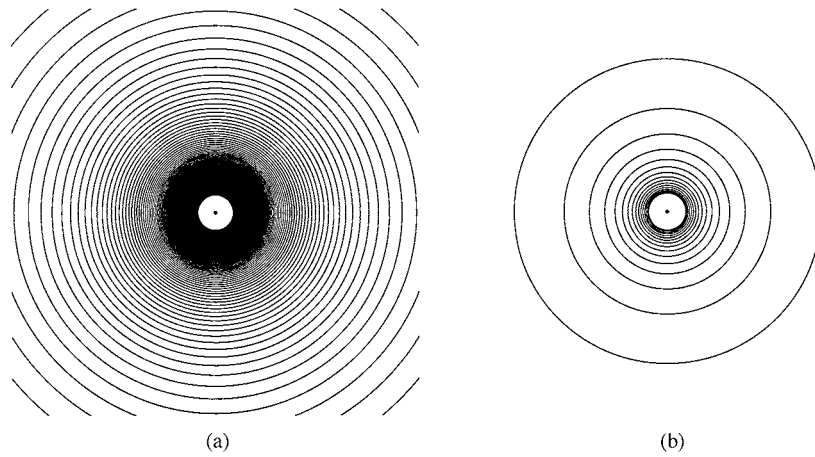

Figure 6. The number of possible depth estimation values is proportional to the angle φ. Each circle denotes possible depth estimation value. (a) $2\varphi = 29.9625°$; (b) $2\varphi = 3.6125°$.

values is proportional to the angle φ: we can calculate φ div $\frac{\theta_0}{2} = 149$ different depth values if we are using angle $2\varphi = 29.9625°$ (Fig. 6(a)) and only 18 different depth values if we are using the angle $2\varphi = 3.6125°$ (Fig. 6(b)). This is the disadvantage of using small angles φ.

Let us illustrate the meaning of the one-pixel error in estimation of angle θ: We would like to know what is the error of the angle θ if θ is at the beginning of the interval on which it is defined ($\theta \in [0, \varphi]$) and what is the error of the angle θ which is near the end of this interval?

For this purpose we will choose angles $\theta_1 = \frac{\varphi}{4}$ and $\theta_2 = \frac{7\varphi}{8}$. We are also interested in the nature of the

error for different values of angle φ. In this example we will use our already standard values for angle φ: $2\varphi = 29.9625°$ and $2\varphi = 3.6125°$. Results in Table 1 give values of the one-pixel error in estimation of the angle θ for different values of parameters θ and φ.

From results in Table 1 we can conclude that the error is much bigger in case of smaller angle φ than in case of bigger angle φ. The second conclusion is that the value of the error is getting bigger when the value of the angle θ is getting closer to the value of the angle φ. This is true regardless of the value of the angle φ. This two conclusions are also evident from Fig. 6: possible depth estimations lie on concentric circles centered in the center of the system and the distance between circles

Table 1. The one-pixel error in estimation of the angle θ, where $r = 30$ cm and $\theta_0 = 0.2°$ (Eqs. (1) and (3)).

	$\theta - \frac{\theta_0}{2}$		θ		$\theta + \frac{\theta_0}{2}$
(a) $\theta = \theta_1 = \frac{\varphi}{4}$, $2\varphi = 29.9625°$					
l (mm)	394.5		398		401.5
Δl (mm)		3.5		3.5	
(b) $\theta = \theta_1 = \frac{\varphi}{4}$, $2\varphi = 3.6125°$					
l (mm)	372.5		400		431.8
Δl (mm)		27.5		31.8	
(c) $\theta = \theta_2 = \frac{7\varphi}{8}$, $2\varphi = 29.9625°$					
l [mm]	2252.9		2373.2		2507
Δl (mm)		120.3		133.8	
(d) $\theta = \theta_2 = \frac{7\varphi}{8}$, $2\varphi = 3.6125°$					
l [mm]	1663		2399.6		4307.4
Δl (mm)		736.6		1907.8	

Table 2. The one-pixel error Δl in estimation of the angle θ for the minimal possible depth estimation l_{min} and the maximal possible depth estimation l_{max} regarding the angle φ.

	$2\varphi = 29.9625°$	$2\varphi = 3.6125°$
Δl_{min}	2 mm	19 mm
Δl_{max}	30172 mm	81587 mm

is increasing the further away they lie from the center. The figure nicely illustrates the fact that in case of a small angle φ, we can estimate only a few different depths and the fact that the one-pixel error in estimation of angle θ increases if we move away from the center of the system.

We would like to get reliable depth estimates but at the same time we would like that the reconstruction procedure executes fast. Here we are faced with two contradicting requirements, since we have to make a compromise between the accuracy of the system and the speed of the reconstruction procedure. Namely, if we like to achieve the maximal possible accuracy then we would use the maximal possible angle φ. But this means that we would have to search for corresponding points on a larger segment of the epipolar line. Consequently, the speed of the reconstruction process would be lower. We would come to the same conclusion if we want to achieve a higher speed of the reconstruction procedure. The speed of the reconstruction process is inversely proportional to its accuracy.

By varying the parameters θ_0 and r, we are changing the size of the error.

6.4. Definition of the Maximal Reliable Depth Value

In Section 6.2 we defined the minimal possible depth estimation l_{min} and the maximal possible depth estimation l_{max}, but we did not write anything about the meaning of the one-pixel error in estimation of the angle θ for these two estimated depths. Let us examine the size of error Δl for these two estimated depths. We calculate Δl_{min} as an absolute value of difference between the depth l_{min} and the depth l for which the

angle θ is bigger from the angle θ_{min} for the angle $\frac{\theta_0}{2}$:

$$\Delta l_{min} = \left| l_{min}(\theta_{min}) - l\left(\theta_{min} + \frac{\theta_0}{2}\right) \right|$$
$$= \left| l_{min}\left(\frac{\theta_0}{2}\right) - l(\theta_0) \right|.$$

Similarly, we calculate the error Δl_{max} as an absolute value of difference between the depth l_{max} and the depth l for which the angle θ is smaller from the angle θ_{max} for the angle $\frac{\theta_0}{2}$:

$$\Delta l_{max} = \left| l_{max}(\theta_{max}) - l\left(\theta_{max} - \frac{\theta_0}{2}\right) \right|$$
$$= \left| l_{max}\left(n\frac{\theta_0}{2}\right) - l\left((n-1)\frac{\theta_0}{2}\right) \right|,$$

where variable n denotes a positive number in equation: $n = \varphi \operatorname{div} \frac{\theta_0}{2}$.

In Table 2 we gathered the error sizes for different values of angle φ. The results confirm statements in Section 6.3. We can add two additional conclusions: 1) The value of error Δl_{max} is unacceptable high and this is true regardless of the value of the angle φ. This is why we have to sensibly decrease the maximal possible depth estimation l_{max}. In practice this leads us to define the upper boundary of the allowed one-pixel error Δl in estimation of the angle θ and with it we subsequently define the maximal reliable depth value. 2) Angle φ always depends upon the horizontal view angle α of the camera (Eq. (2)). While the angle α is limited to around $40°$ for standard cameras, our system is limited with the angle α when estimating the depth, since in the best case we have: $\varphi_{max} = \frac{\alpha}{2}$. Thus our system can be used only for 3D reconstruction of small spaces.

7. Experimental Results

Figure 7 shows some results of our system. In the case denoted with (b), we constructed the dense panoramic
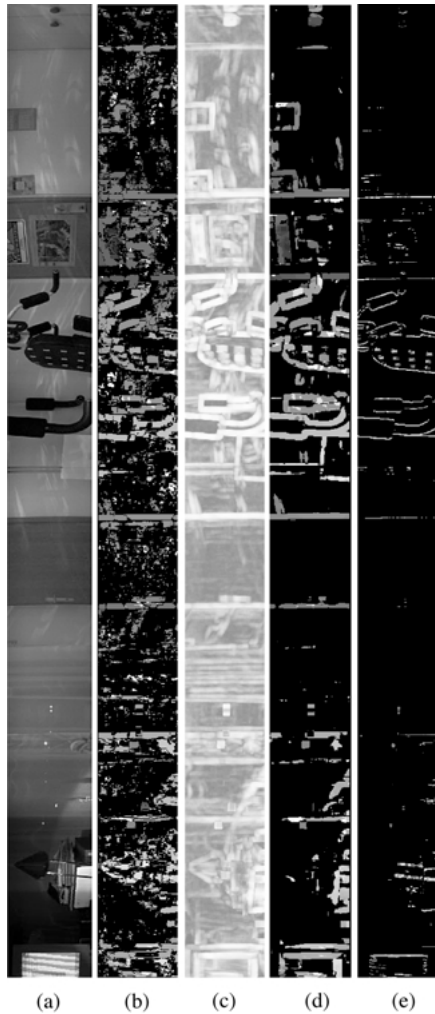
(a)   (b)   (c)   (d)   (e)

Figure 7. Some results of stereo reconstruction when creating the depth image for the left eye while angle $2\varphi = 29.9625°$: (a) left eye panorama, (b) dense depth image/using back-correlation reconstruction time: 6 hours, 42 min., 20 sec., (c) confidence of estimated depth, (d) dense depth image after weighting/without back-correlation/reconstruction time: 3 hours, 21 min., 56 sec., (e) sparse depth image/without back-correlation/reconstruction time: 38 seconds. The time needed for the acquisition of panoramic images is not included in the reconstruction time.

image, which means that we tried to find a corresponding point on the right eye panorama for every point on the left eye panorama. Black color marks the points on the scene with no depth estimation associated. Otherwise, the nearer the point on the scene is to the rotational center of the system, the lighter the point appears in the depth image.

In the case denoted with (d), we used the information about the confidence in estimated depth (case (c)),

which we get from the normalized correlation estimations. In this way, we eliminated from the dense depth image all the associated depth estimates which do not have a high enough associated confidence estimation. Lighter the point appears in case (c), more we trust in the estimation of normalized correlation for this point.

In the case marked with (e), we created a sparse depth image by searching only for the correspondences of feature points on input panoramic images. The feature points we used were vertical edges on the scene, which were derived by filtering the panoramic images with the Sobel filter for searching vertical edges (Faugeras, 1993; Ishiguro et al., 1992).

If we use a smaller value for angle φ, i.e. $\varphi = 3.6125°$, the reconstruction time would be up to eight times smaller from presented ones. All results were generated by using a correlation window of size $(2n+1) \times (2n+1), n = 4$. We searched for corresponding points only on the panoramic image row, which was determined by the epipolar geometry.

Since it is hard to evaluate the quality of generated depth images given in Fig. 7, we will present the reconstruction of the room from the generated depth image. Then we will be able to evaluate the quality of generated depth image and consequently the quality of the system. The result of the (3D) reconstruction process is a ground-plan of the scene. The following properties can be observed in Fig. 8: 1) Big dot near the center of the reconstruction shows the center of our system. 2) Other big dots denote features on the scene for which we measured the actual depth by hand. 3) Small black dots are reconstructed points on the scene. 4) Lines between black dots denote links between two successively reconstructed points.

The result of the reconstruction process based on the 68th row of the depth image when we used back-correlation and weighting is given in Fig. 8. Black dots are reconstructed on the basis of estimated depth values, which are stored in the same row of the depth image. The features on the scene marked with big dots are not necessarily visible in the same row.

Let us end this section with the demonstration of the reconstruction error. The error function on the manually measured points on the scene is evaluated in Table 3.

8. Summary and Future Work

We presented an exhaustive analysis of our mosaic-based system for construction of depth panoramic

Table 3. The comparison between the manually measured actual distances and estimated depths (Eq. (1)) for some features on panoramic images. It is given for different values of the angle φ and $r = 30$ cm. Please remember that we can calculate only 149 different depth values if we are using the angle $\varphi = 29.9625°$ and only 18 different depth values if we are using the angle $\varphi = 3.6125°$ (Section 6.3).

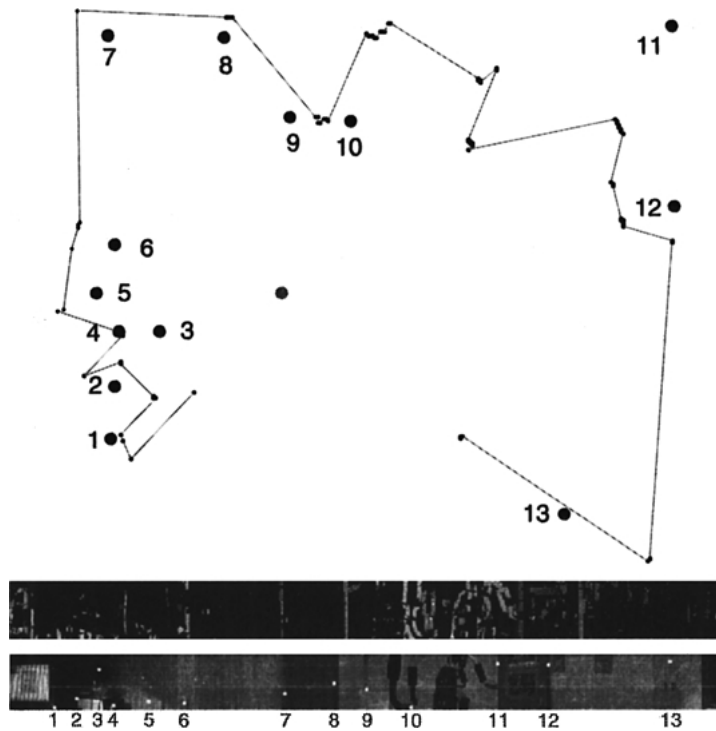| Feature marked in Fig. 8 with	Actual distance d (cm)	Estimated depth l (cm) for 2φ		Difference $l - d$ (cm (% of d)) for 2φ	
		3.6125°	29.9625°	3.6125°	29.9625°
1	111.5	89.4	109	−22.1 (−19.8%)	−2.5 (−2.2%)
2	95.5	76.7	89.3	−18.8 (−19.6%)	−6.2 (−6.5%)
3	64	53.8	59.6	−10.2 (−15.9%)	−4.4 (−6.9%)
4	83.5	76.7	78.3	−6.8 (−8.1%)	−5.2 (−6.2%)
5	92	89.4	89.3	−2.6 (−2.8%)	−2.7 (−2.9%)
6	86.5	76.7	82.7	−9.8 (−11.3%)	−3.8 (−4.4%)
7	153	133.4	159.8	−19.6 (−12.8%)	6.8 (4.4%)
8	130.5	133.4	135.5	2.9 (2.2%)	5 (3.8%)
9	88	76.7	87.6	−11.3 (−12.8%)	−0.4 (−0.5%)
10	92	89.4	89.3	−2.6 (−2.8%)	−2.7 (−2.9%)
11	234.5	176.9	213.5	−57.6 (−24.6%)	−21 (−9%)
12	198	176.9	179.1	−21.1 (−10.7%)	−18.9 (−9.5%)
13	177	176.9	186.7	−0.1 (−0.1%)	9.7 (5.5%)

Figure 8. On top is a ground-plan showing the result of the reconstruction process based on the 68th row of the depth image. We used back-correlation and weighting for angle $2\varphi = 29.9625°$. The corresponding depth image is shown in the middle picture. For orientation, the reconstructed row and the features on the scene for which we measured the actual depth by hand are shown in the bottom picture. Note that the features on the scene marked with big dots and associated numbers are not necessarily visible in this row.

images using only one standard camera. We demonstrated the following: the procedure for creating panoramic images is very long and can not be executed in real time under any circumstances (using only one camera); epipolar lines of symmetrical pair of panoramic images are image rows; based on the equation for estimation of depth l (Eq. (1)), we can effectively constraint the search space on the epipolar line; confidence in estimated depth is changing: bigger is the slope of the function l curve, smaller is the confidence in estimated depth; if we observe the reconstruction time, we can conclude that the creation of dense panoramic images is very expensive.

The essential conclusions are: 1) Such systems can be used for 3D reconstruction of small spaces. 2) With respect to the presented reconstruction times (Fig. 7) we could conclude that the reconstruction procedure could work in nearly real time, if we work with 8-bit grayscale images, with lower resolution, if we create the sparse depth image of only part of the scene and/or simply if we use a faster computer. This could be used for robot navigation (Ishiguro, 1992).

A further time reduction in panorama building can be achieved: instead of building the panorama from only one column of the captured image, we could build the panorama from the wider stripe of the captured image (Prihavec and Solina, 1998). Thus, we would increase the speed of the building process. If we use this idea in our system, we know that within the stripe the angle φ is changing. However, the question is how this influences the reconstruction procedure.

In the future we intend to enlarge the vertical field of view of panoramic images, address the precision of vertical reconstruction and use the sub-pixel accuracy procedure.

Our future work is directed primarily in the development of an application for the real time automatic navigation of a mobile robot in a room.

Acknowledgment

References

Chen, S. 1995. Quicktime VR—an image-based approach to virtual environment navigation. In *ACM SIGGRAPH*, Los Angeles, USA, Aug. 1995, pp. 29–38.

Faugeras, O. 1993. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press: Cambridge, Massachusetts.

Gupta, R. and Hartley, R.I. 1997. Linear pushbroom cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(9):963–975.

Hartley, R. and Zisserman, A. 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press: Cambridge, UK.

Huang, F. and Pajdla, T. 2000. Epipolar geometry in concentric panoramas. Technical Report CTU-CMP-2000-07, Center for Machine Perception, Czech Technical Univer-sity, Prague, Czech Republic, March 2000. Available at: ftp://cmp.felk.cvut.cz/pub/cmp/articles/pajdla/Huang-TR-2000-07.ps.gz

Huang, F., Wei, S.K., and Klette, R. 2001. Geometrical fundamentals of polycentric panoramas. In *IEEE International Conference on Computer Vision*, Vancouver, Canada, July 2001, vol. I, pp. 560–565.

Ishiguro, H., Yamamoto, M., and Tsuji, S. 1992. Omni-directional stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):257–262.

Paar, G. and Pölzleitner, W. 1992. Robust disparity estimation in terrain modeling for spacecraft navigation. In *IEEE International Conference on Pattern Recognition*, The Hague, The Netherlands, Aug./Sept. 1992, vol. I, pp. 738–741.

Peleg, S. and Ben-Ezra, M. 1999. Stereo panorama with a single camera. In *IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, USA, June 1999, vol. I, pp. 395–401.

Peleg, S., Ben-Ezra, M., and Pritch, Y. 2001. Omnistereo: Panoramic stereo imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3):279–290.

Peleg, S., Pritch, Y., and Ben-Ezra, M. 2000. Cameras for stereo panoramic imaging. In *IEEE Conference on Computer Vision and Pattern Recognition*, Hilton Head Island, USA, June 2000, vol. I, pp. 208–214.

Peleg, S., Rousso, B., Rav-Acha, A., and Zomet, A. 2000. Mosaicing on adaptive manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1144–1154.

Prihavec, B. and Solina, F. 1998. User interface for video observation over the internet. *Journal of Network and Computer Applications*, 21:219–237.

Rademacher, P. and Bishop, G. 1998. Multiple-center-of-projection images. In *Computer Graphics (ACM SIGGRAPH)*, Orlando, USA, July 1998, pp. 199–206.

Shum, H.Y. and Szeliski, R. 1999. Stereo reconstruction from multiperspective panoramas. In *IEEE International Conference on Computer Vision*, Kerkyra, Greece, Sept. 1999, vol. I, pp. 14–21.

Szeliski, R. and Shum, H.Y. 1997. Creating full view panoramic image mosaics and texture-mapped models. In *Computer Graphics (ACM SIGGRAPH)*, Los Angeles, CA, USA, Aug. 1997, pp. 251–258.

Wood, D., Finkelstein, A., Hughes, J., Thayer, C., and Salesin, D. 1997. Multiperspective panoramas for cel animation. In *Computer Graphics (ACM SIGGRAPH)*, Los Angeles, USA, Aug. 1997, pp. 243–250.