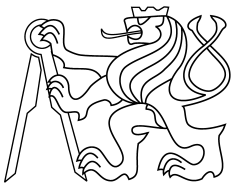
CENTER FOR
MACHINE PERCEPTION



CZECH TECHNICAL
UNIVERSITY IN PRAGUE

MASTER'S THESIS

ISSN 1213-2365

Epipolar Rectification for Stereovision

Jan Sandr

CTU–CMP–2009–04

January, 2009

Available at
ftp://cmp.felk.cvut.cz/pub/cmp/articles/matousek/Sandr-TR-2009-04.pdf

Thesis Advisor: Ing. Martin Matoušek Ph.D.

This work has been conducted under Project 1ET101210406
supported by the Czech Academy of Sciences.

Research Reports of CMP, Czech Technical University in Prague, No. 4, 2009

Published by

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Technická 2, 166 27 Prague 6, Czech Republic
fax +420 2 2435 7385, phone +420 2 2435 7637, www: http://cmp.felk.cvut.cz

Epipolar Rectification for Stereovision

Jan Sandr

January, 2009

Epipolární narovnání obrazů pro stereovidění

Tématem této diplomové práce je epipolární narovnání obrazků pro stereovidění. Na základě teoretického rozboru a znalosti současných metod jsou vybrány dva algoritmy s důrazem na jejich schopnosti epipolárního narovnání obrázků z kamery s širokým zorným úhlem v případě dopředného pohybu. Diplomová práce se dále zabývá možností optimalizace vzorkovaní s přihlédnutím k lokálním spektrálním vlastnostem obrázku. K porovnání přesnosti disparitních map získaných z rektifikovaných obrázků jsou použita data z páru horizontálně uspořádaných kamer.

Epipolar rectification for stereovision

The topic of this diploma thesis is epipolar rectification. The theory and state of the art are discussed emphasizing the rectification methods suitable for wide-filed-of-view cameras with dominant forward motion. Based on the theoretical background two rectification algorithms are chosen for implementation. The issue of optimized resampling is addressed considering local spectral properties of an image. The disparity maps of resulting rectified images are compared to a ground truth measurement obtained from a pair of horizontally aligned cameras.

I would like to thank my advisor, Martin Matoušek, and Dr. Radim Šára for providing me with much needed explanations of consulted topics. I also appreciated very much the support and patience of my family during my work on the thesis. This work has been conducted under Project 1ET101210406 supported by the Czech Academy of Sciences.

Contents

1 Introduction

Imagine a robot who wants to explore the world. He has feet to walk on, but no eyes to see with, just one camera. But there are many obstacles in his way, that he must overcome. Will he be able to do that? Of course! Because, once he starts moving, very exciting things can be done with his camera...

So what can we really do with a moving camera? We can for example retrieve depth maps. This way our robot would be able to detect obstacles in its direction of movement and act accordingly. Darpa Urban Challenge is an example of such application. Autonomous vehicles took part in a competition proving that automobiles can drive in traffic and are capable of maneuvers such as passing, merging and parking without driver's assistance. We should also mention another intersesting application, a view prediction, which allows us to construct images previously unseen.

Another well-explored area is 3D reconstruction, Kamberov [9]. Going back to our robot we could have him walk in a tomb of Egyptian pharaoh and get a complete 3D model of his afterlife pallace.

So what do all these applications have in common? They all make use of stereo matching. Stereo matching is an algorithm which finds corresponding points in given images. In practice this usually leads to some correlation-based search, Ćech and Šára [2]. This proves to be very demanding on both time and computational capacity. In the worst case we have to compute correlations for all the area of the other image shifting a correlation window pixel by pixel in both vertical and horizontal direction.

Fortunately a concept of epipolar geometry can be employed to reduce the search dimension by looking for matches only along corresponding epipolar lines. Some stereo matching algorithms work directly on the original images, but these methods typically do not produce a large number of found correspondences. To obtain a very large number of corresponding points we need to employ a dense stereo matching algorithm. In order for it to perform efficiently, the images first have to be preprocessed by *epipolar rectification*. Epipolar rectification is a geometric transformation of a pair of images mapping their epipolar lines onto the scanlines, such that corresponding epipolar lines are on the same row.

The focus of this thesis is epipolar rectification of an image pair with dominant forward displacement. This is a typical camera setup in monocular stereovison, which easily finds use in applications mentioned above. The thesis is structured as follows: in Chapter 2 we see a quick overview of camera models and epipolar geometry. Chapter 3 discusses the rectification state of the art and based on the theoretical background suitable rectification methods are chosen and implementation details are further developed. Chapter 4 presents an approach to optimized resampling of the image considering its spectral properties. Experimental results are shown in Chapter 5, where disparity maps are compared to ground truth measurements.

2 Cameras and epipolar geometry

2.1 Perspective camera

A perspective camera realizes a perspective projection of a homogeneous scene point $\mathbf{X}$ with entries $\left(x,\, y,\, z,\, 1\right)^\top$ onto the homogeneous camera sensor point $\mathbf{u}$ with entries $\left(u,\, v,\, 1\right)^\top$ via a ray connecting $\mathbf{X}$ and camera center $\mathbf{C}$ as follows:

$$\alpha\,\mathbf{u} = \mathbf{P}\,\mathbf{X}\,, \tag{2.1}$$

where $\mathbf{P} = \mathbf{K}\big[\,\mathbf{R} \mid -\mathbf{C}\,\big]$ denotes camera projection matrix, $\mathbf{K}$, $\mathbf{R}$ are calibration and rotation matrices respectively and $\mathbf{C}$ is the inhomogeneous camera center. The calibration matrix $\mathbf{K}$ bears an information about the intrinsic parameters of the camera, such as pixel shape and camera center. The matrix $\mathbf{R}$ defines the orientation of camera-related coordinate frame with respect to world coordinate frame. Note that in this camera model any scene point lying on the ray through the camera center, is projected onto a single point in the image, regardless of whether it is in front of or behind the camera.

Given two projection matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ for reference and target camera respectively, a scene point $\mathbf{X}$ is projected as a pair of points $\mathbf{u}_1$ and $\mathbf{u}_2$. These corresponding points are bound by epipolar geometry as follows:

$$\mathbf{u}_2^\top\,\mathbf{F}\,\mathbf{u}_1 = 0\,. \tag{2.2}$$

Epipolar geometry is characterised by a fundamental matrix $\mathbf{F}$, which can be computed as:

$$\mathbf{F} = [\mathbf{e}_2]_\times\,\mathbf{P}_2\,\mathbf{P}_1^+\,, \tag{2.3}$$

where epipole $\mathbf{e}_2$ is a projection of $\mathbf{C}_1$ into the target camera, $\mathbf{e}_2 = \mathbf{P}_2\,\mathbf{C}_1$. Equation (2.2) states that a projection $\mathbf{u}_2$ of a scene point $\mathbf{X}$ in the target view lies on the epipolar line $\mathbf{l}_2 = \mathbf{F}\,\mathbf{u}_1$. An epipolar line $\mathbf{l}_2$ can be seen as an image of ray connecting $\mathbf{X}$ and $\mathbf{C}_1$. Similarly $\mathbf{l}_1$ in the reference view is a projection of ray connecting $\mathbf{X}$ and $\mathbf{C}_2$. We see that points $\mathbf{X}$, $\mathbf{C}_1$ and $\mathbf{C}_2$ define a particular plane in space, which is called epipolar plane. Such plane intersects the image planes in corresponding epipolar lines $\mathbf{l}_1$ and $\mathbf{l}_2$. A line through camera centers $\mathbf{C}_1$ and $\mathbf{C}_2$ is called a baseline and it intersects image planes of the cameras in epipoles $\mathbf{e}_1$ and $\mathbf{e}_2$.

2.2 Wide field of view camera

Wide field of view cameras are used in applications, where large visual field has to be covered. Fish-eye lens and catadioptric system are typical examples. Using either of them results in a warped image compared to linear camera model and we must take this into account before we decide on rectification method.

Perspective camera with radial distortion

Wide field of view cameras covering an angle lower than 180∘ can be described as a perspective cameras with radial distortion. Such camera projection is modeled in two steps. First the scene point $\mathbf{X}$ is perspectively projected according to (2.1) onto a point $\mathbf{u}$. Second the point $\mathbf{u}$ is transformed by some non-linear function r, Fitzgibbon [3], to obtain a radially distorted point $\mathbf{u}^{rd}$.

If we retrieve parameters of the radial distortion function r, we are able to remove the image distortion introduced by it, thus leaving us with perspective camera where epipolar geometry is described by (2.2).

Omnidirectional camera

Omnidirectional cameras are able to cover an angular field of view of up to 360 degrees. They are often realized as a catadioptric system, i. e. an optical system consisting of a mirror and a lens, such that their optical axes are aligned. Here a line through the camera center has two intersections with camera image plane, as the camera sees all around itself. According to [4] we may describe a projection of such camera by two projection steps, where first we project onto the sphere via central projection and second we project, depending on the mirror used in the system, from a specific point lying between the center and the north pole of the sphere onto the plane whose normal is aligned with optical axis.

Considering an epipolar plane in the first step we see that it is projected onto a great circle. Projecting such great circle according to the second step onto the image plane would yeild a general curve. We may imagine this projection as an intersection of an image plane and a cone containing the great circle and whose vertex is a point of projection. It is shown that in a case of parabolic catadioptric camera the point of projection is at the north pole and that the great circles project onto circular arcs.

Given a point correspondences $\mathbf{u}_1$ and $\mathbf{u}_2$ in respective images the epipolar geometry for parabolic catadioptric system may be defined similarly to (2.2) as follows:

$$\tilde{\mathbf{u}}_2^\top \mathbf{F}\, \tilde{\mathbf{u}}_1 = 0\,, \tag{2.4}$$

$$\tilde{\mathbf{u}} = \big(u,\, v,\, u^2 + v^2 - \frac{1}{4},\, 1\big)^\top, \tag{2.5}$$

where $\tilde{\mathbf{u}}_1$ and $\tilde{\mathbf{u}}_2$ are the liftings of image points $\mathbf{u}_1$, $\mathbf{u}_2$ to a paraboloid via (2.5) and $\mathbf{F}$ is a catadioptric fundamental matrix as introduced by Geyer and Daniilidis in [5].

2.3 Epipolar configurations arising from camera motion

Let us now briefly consider a moving perspective camera with fixed intrinsic parameters. The camera at any moment is then defined by its rotation matrix $\mathbf{R}$ and center $\mathbf{C}$. Given two cameras $\mathbf{P}_1$ and $\mathbf{P}_2$ and connecting their centers $\mathbf{C}_1$ and $\mathbf{C}_2$ we obtain a baseline. Now depending on the cameras' relative rotations different epipolar configurations arise. We may study such coonfigurations by examining the positions of epipoles in the image planes.

In what follows we will assume that optical axes of the cameras are aligned with the baseline. We will now consider a rotation about an axis perpendicular to the baseline. We see that there are certain intervals of angle $\theta \in (-\alpha + k\,\pi,\ +\alpha + k\,\pi)$ for which the baseline intersects the bounded image area and intervals of angle $\theta \in (+\alpha + k\,\pi,\ \alpha + (k + + 1)\,\pi)$ for which it does not. If $\theta = \pm\alpha + k\,\pi$ then the epipole lies on the boundary of the image. We should differentiate between a case, in which the epipole is located inside the image, and a case, in which it is ouside. As we will see in the following chapter, this has a fundamental impact on the rectification approach taken. Having one epipole per image there exist four possible groups of epipolar configurations.

1. Both epipoles are outside the image area and both angles of rotation $\theta_{1,2}$ lie within $(+\alpha + k\,\pi,\ \alpha + (k + 1)\,\pi)$. We should note a rotation of $\pm\frac{pi}{2}$ as, if both cameras are subject to such rotation, the epipolar lines become parallel to the baseline in both views. Furthermore corresponding epipolar lines will be horizontally alligned in both views, thus producing a rectified pair of images.
2. Both epipoles are inside the image, thus the angles of rotation are bound by $(-\alpha + + k\,\pi,\ +\alpha + k\,\pi)$. In this case the epipolar planes project onto a set of half-lines emanating from the epipole.
3. The epipole in the reference view is outside and $\theta_1 \in (+\alpha + k\,\pi,\ \alpha + (k + 1)\,\pi)$, while in the target view it is inside the image and $\theta_2 \in (-\alpha + k\,\pi,\ +\alpha + k\,\pi)$.
4. The epipole in the target view is outside and $\theta_2 \in (+\alpha + k\,\pi,\ \alpha + (k + 1)\,\pi)$, while in the reference view it is inside the image and $\theta_1 \in (-\alpha + k\,\pi,\ +\alpha + k\,\pi)$.

Setups 3. and 4. are considered identical as we only swap reference and target camera. If we now, instead of the baseline, fix the orientation of the reference camera, we will also come to four configurations mentioned above. Views with both epipoles far outside the image arise from camera motion with dominant lateral component. For views where both epipoles are inside the image is characteristic dominant forward motion. In the case where target epipole is inside and reference epipole outside, the camera moved in lateral direction and rotated so that the baseline intersects the image plane of the target view.

2.4 Our camera

In our application the images are obtained by reversing camera with a horizontal angle of view 130°. Figure 2.1 shows an example. This camera most closely resembles a model of perspective camera with radial distortion. Assuming that the radial distortion of an image point is induced by a paraboloid of revolution and based on precise measurements on the calibration grid the radial distortion function r was determined as follows:

$$r(u, \|\mathbf{u}\|) = u\,\frac{\frac{2\,b}{f}}{1 + \sqrt{1 + \frac{\|\mathbf{u}\|}{f}^2}}\,, \tag{2.6}$$

where u is the first coordinate of perspective image point $\mathbf{u}$, b is a distance between the center and vertex of the paraboloid and f is the focal distance of the perspective camera and so the first coordinate of radially distorted image point $u^{rd} = r(\mathbf{u}, \|\mathbf{u}\|)$. Seeing that the paraboloid is radially symmetric, the second entry of the distorted point is computed in the same manner $v^{rd} = r(v, \|\mathbf{u}\|)$, leaving us with a radially distorted image point $\mathbf{u}^{rd} = (u^{rd},\, v^{rd},\, 1)^\top$.

Figure 2.1 A radially distorted image obtained by reversing camera.

We shall not go more into detail of radial distortion model used, as it is beyond the scope of this thesis. More may be found in the work of Fitzgibbon in [3].

Three important things should be noted about the reversing camera in our application:

1. The radial distortion function is known and thus we may obtain a perspective camera image and apply perspective epipolar geometry.
2. The radial distortion function is derived using a projection via paraboloid making the distorted images suitable for epipolar geometry introduced by Geyer and Daniilidis in [4].
3. In our application the camera's most dominant motion component is in the direction of optical axes.

This chapter gave us a quick overview of camera models and also discussed suitable epipolar geometry for each of them. The topics intoduced here will be of importance in the following chapter.

3 Rectification methods

The goal of all rectification algorithms is to simplify the stereo matching algorithm task of correspondence finding. This is done by some transformation mapping corresponding epipolar lines onto the same row in both rectified images.

In this chapter we will mention the basics of rectification transformations and discuss few of them in more detail and choose such, that are suitable for a reversing camera in our application. Their implementation will be described.

3.1 State of the art

We may differentiate between the rectification algorithms based on the transformation they perform to obtain rectified images.

Linear methods

The linear rectification methods are based on finding a linear projective transformation. The advantage of such approach is that only one transformation matrix needs to be found for each view, thus resulting in high speed and simplicity of algorithms. Another advantage is that rectified images still comply with perspective camera model. According to [7] we look for a transformation that both sends the epipole to infinity, thus aligning epipolar lines horizontally, and matches corresponding epipolar lines on the same row. However a significant drawback of linear algrithms is, that they fail to perform in such configurations, when one or both epipoles are inside the image, typically yielding a rectified image stretching to infinity. In that case we may only process a part of an image at a time, in order to keep the size of the rectified image within feasible boundaries.

Non-linear methods

As opposed to linear methods we do not look for a linear transformation that would be aplicable to the whole region of the image. There are several approaches to non-linear rectification algorithms:

1. geometric transformations in 3D,
2. direct image domain transformations,
3. use of special camera geometry.

First approach involves some kind of geometric construction, where original images are remapped onto a specific geometric figure to obtain rectified images. Roy, Meunier and Cox [12] chose a cylinder which is oriented such that its axis is coincident with a baseline. This way all straight lines on the surface of the cylinder become parallel with focus of expansion, making them suitable as rectification targets. Now the object is to find a map that takes corresponding epipolar lines and places them onto the surface of the cylinder on the same angular position. This is done by first rotating the epipolar

line in an epipolar plane so that it becomes parallel to field of expansion. Then a transformation of coordinate system is performed to obtain the space point coordinates with respect to a basis common for both images. The last step of the process is a scaling, which projects the line onto the surface of the unit cylinder. Since the initial rotation of corresponding epipolar lines is done in the same epipolar plane, these lines will not only become parallel but also located on the same angular position. As we see the three steps may be described by linear transformations in projective space. This results in necessity of finding the three transformation matrices per each epipolar line. The algorithm preserves the original epipolar line lengths, thus minimizing the lost information in the process of resampling and is capable of rectifying images that arise from a general motion of a camera always producing a bounded image.

Second approach is characteristic by strictly working in an image domain, choosing corresponding epipolar lines and aligning them in an image domain of rectified images. This approach proceeds sequentially building the rectified image line by line, enabling us to control a the distance of two consecutive epipolar lines. Non-linear polar rectification was proposed by Pollefeys, Koch and van Gool [11]. We may describe this algorithm as partially linear, as we only apply a linear projective transformation to a very small wedge part of the image with vertex at the epipole. A linear transformation has to be found for every such wedge, which may be computationaly little more expensive, but on the other hand leads to the ability of rectifying whole image for every possible epipolar configuration, while keeping the size of resulting image reasonably large for further stereo matching. The algorithm is designed in such way that no pixel loss is guaranteed. Also the length of original epipolar lines is preserved, therefore the resulting image is upper-bounded by $2\,(W + H) \times \sqrt{W^2 + H^2}$, where W is original image width and H its height. An important property of this method is, that it uses a concept of oriented epipolar geometry to reduce matching ambiguity to a half of the epipolar line in a case of epipole inside the image. To do this one point correspondence is needed in both views. The method uses a polar parameterization around the epipole to obtain coordinates for rectified image points. Given two rectified images a stereo matching algorithm performs $\frac{1}{2}R\,c_1\,c_2$ correlations instead of $R\,c_1\,c_2$ needed with images obtained by Roy et al. [12], where R stands for number of rows and c_1, c_2 are the numbers of columns in respective rectified images.

Third approach uses a special type of camera geometry. A non-linear rectification method proposed by Geyer and Daniilidis [6] is specifically designed to rectify a pair of images obtained by parabolic catadioptric cameras. It uses a concept of parabolic catadioptric epipolar geometry to find a closed-form rectification formula. A point in the image is represented as a complex number and a bipolar system of coordinates is employed to obtain the coordinates of rectified point. A bipolar system consists of two mutually orthogonal sets of coaxal circles.

The equivalent to epipolar line in perspective image is an epipolar circle in parabolic catadioptric image. As we have seen in the previous chapter the projections of antipodal points are separate, and therefore a projection of camera center of one view has two images in the other view, two epipoles $\mathbf{e}_1$ and $\mathbf{e}_2$. Any circle is defined by three points in plane. All epipolar circles must contain both epipoles. Thus given a point $\mathbf{z}$ we have a unique epipolar circle. An angle at vertex $\mathbf{z}$ of the triangle $\mathbf{e}_1\,\mathbf{z}\,\mathbf{e}_2$ parameterizes the set of points lying on considered epipolar circle. Therefore we have one system of coaxal

circles, where each circle is determined by an angle. For the other set of coaxal circles, to comply with bipolar coordinate system, must hold that for any point $\mathbf{w}$ on a given circle a ratio of distances $\frac{\|\mathbf{w}\,\mathbf{e}_1\|}{\|\mathbf{w}\,\mathbf{e}_2\|}$ is constant. In the rectified image we want the epipolar circle to be mapped onto a straight line. Indeed the authors had found such mapping and furthermore it is conformal and therefore locally preserves angles. The range of this transformation is $-\infty$ to ∞ in the real component and $-\pi$ to π in imaginary component. The epipole is a singularity of this mapping and so we must exclude the circular area around. The size and resolution of the resulting image is determined by the position of the epipoles.

3.2 Methods suitable for our application.

We now have enough information to make a qualified choice of the suitable rectification method. We must exclude linear methods as they do not give reasonable results when epipole is inside the image. From non-linear methods we have chosen the algorithms introduced by Pollefyes et al. [11] and by Geyer and Daniilidis [6].

The selection of the first is justified by the fact that it reduces the matching ambiguity to a half of the epipolar line in a case of epipole inside the image. Also it gives us a solid framework for further optimization of the image resampling as will be shown in following chapter. The latter choice has an advantage of working directly on the image obtained by parabolic-distorted images, therefore we only need to resample the image once to get the rectified images. Furthermore it is conformal and thus locally angle-preserving. This may have a significant impact on the performance of the stereo matching algorithm.

3.3 Geyer and Daniilidis algorithm implementation

An existing implementation of this method may be found in a rectification toolbox *OmniRect* created by Jan Heller [8]. We decided to use this toolbox as a basis for our experiments.

3.4 Pollefeys' algorithm implementation

Here the implementation details of the rectification method are discussed. At the end of this section rectification capabilities are shown on synthetic images.

Common region detection

The task of this stage is to determine a pair of corresponding half-epipolar lines defining a common region of the images. We must deal with three possible cases depending on whether the epipoles are located inside or outside the image, Figure 3.1. Using a concept of oriented epipolar geometry we can say that a pencil of epipolar planes intersects the image in a set of half-epipolar lines emanating from the epipole. If the epipole is located inside the image we need to resolve half-epipolar lines pointing in oposite direction, which up to a non-zero scale factor are equivalent. Before we proceed further let us state that entities belonging to or derived from the reference image are denoted by 1 in entity subscript and similarly for entities which correspond to the target image 2 in entity subscript is used.

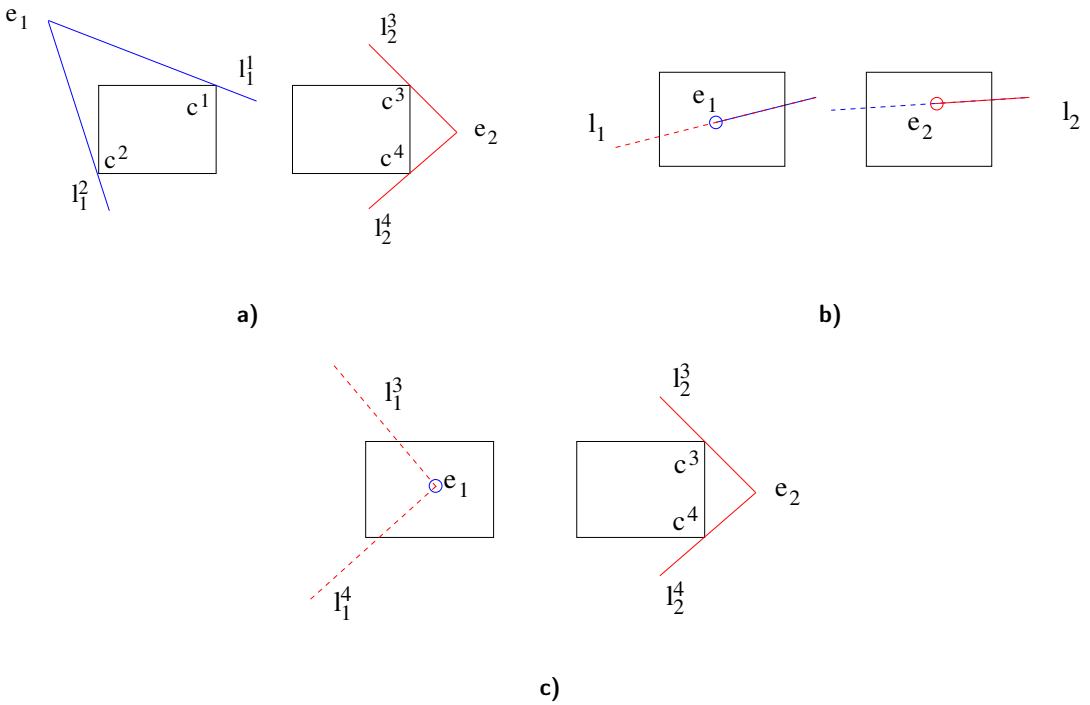


Figure 3.1 Different position of epipole with respect to image leads to different common region detection scheme. Note that in a and c the location of the epipole also determines the outer corners of the images which are to be considered.

Both epipoles are outside.

Let us consider reference image with epipole $\mathbf{e}_1$. Regardless of the location of the epipole extremal epipolar lines always touch the outer corners of the image $\mathbf{c}^1$, $\mathbf{c}^2$. First we obtain extremal epipolar lines $\mathbf{l}_1^1$ and $\mathbf{l}_1^2$, such that $\mathbf{l}_1^{1,2} = [\mathbf{e_1}]_\times \mathbf{c}^{1,2}$. Then we compute $\mathbf{l}_1^3$ and $\mathbf{l}_1^4$, such that $\mathbf{l}_1^{3,4} = \mathbf{F}^\top \mathbf{c}^{3,4}$, where $\mathbf{c}^{3,4}$ are the two outer corners in the target image. Checking whether the lines $\mathbf{l}_1^{3,4}$ intersect the picture area gives us three possibilities, Figure 3.4:

1. Neither of the lines intersect the image, Figure 3.2a. The common region is defined by lines $\mathbf{l}_1^{1,2}$ in the reference image and their corresponding lines $\mathbf{l}_2^{1,2} = \mathbf{F}\,\mathbf{c}^{1,2}$ in the target image.
2. Both lines intersect the image, Figure 3.2b. The common region is defined by lines $\mathbf{l}_1^{3,4}$ in the reference image and their corresponding lines $\mathbf{l}_2^{1,2} = [\mathbf{e}_2]_\times \mathbf{c}^{3,4}$ in the target image.
3. Only one line intersects the image, Figure 3.2c. We now have three lines intersecting the image $\mathbf{l}_1^{1,2}$ and $\mathbf{l}_1^j$ defining regions in the image, where $j \in \{3, 4\}$. To determine the lines bounding the common region we compute the lines $\mathbf{l}_2^{1,2} = \mathbf{F}\,\mathbf{c}^{1,2}$ in the target image. Only one line intersects the image area. Let $\mathbf{l}_2^i$ be this line. Then common region is bound by lines $\mathbf{l}_1^i$ and $\mathbf{l}_1^j$ in the reference image and by $\mathbf{l}_2^i$ and $\mathbf{l}_2^j$ in the target image.

Both epipoles are inside.

Here the issue of common region is simply resolved, because every epipolar line $\mathbf{l}_1$ in the reference image has its corresponding epipolar line $\mathbf{l}_2$ in the target image, making all

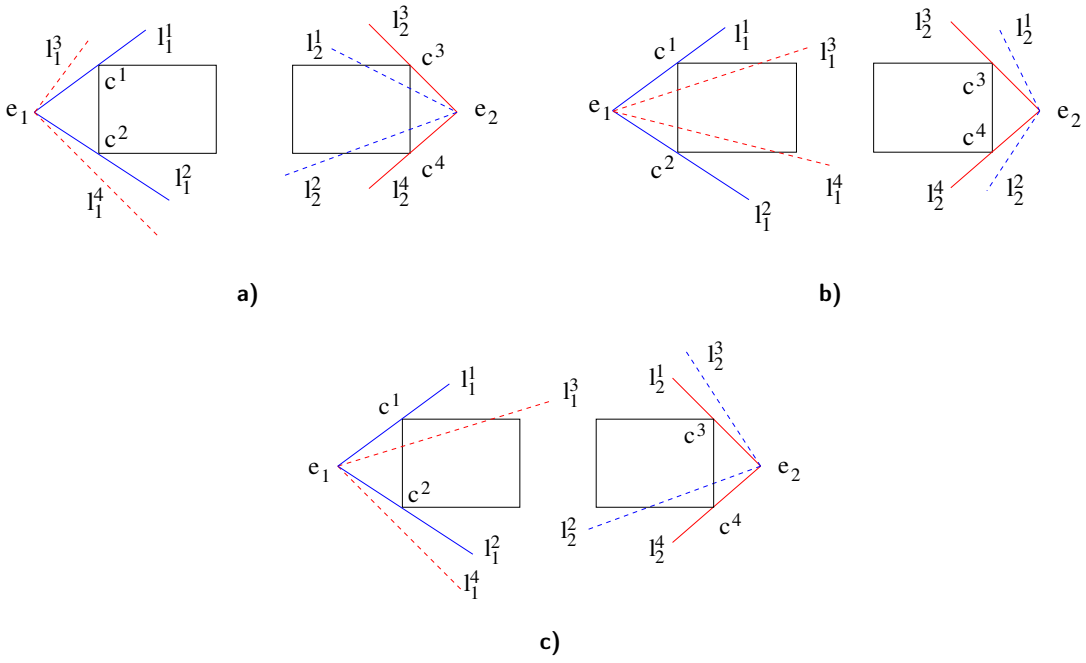


Figure 3.2 Possible epipolar line configurations when both epipoles are outside the image.

the image a common region. Now the problem of half-epipolar correspondence must be reasoned. In general to arbitrary point $\mathbf{u}_1$ in reference image there is a corresponding epipolar line $\mathbf{l}_2$ in target image. The epipole divides $\mathbf{l}_2$ in two half-epipolar lines. A point correspondence to $\mathbf{u}_1$ must lie on one of them. To resolve on which of the two half-epipolar lines point $\mathbf{u}_2$ lies we will employ a concept of oriented epipolar geometry. Given a line $\mathbf{l}_1$ in the reference image its corresponding line $\mathbf{l}_2$ may be found by homography $\mathbf{H}$ as:

$$\mathbf{l}_2 = \mathbf{H}^{-\top}\,\mathbf{l}_1\,, \tag{3.1}$$

$$\mathbf{H} = [\mathbf{e}_2]_\times\,\mathbf{F} + \mathbf{e}_2^\top\,\mathbf{a}\,, \tag{3.2}$$

where $\mathbf{a}$ is an arbitrary vector for which the determinat of $\mathbf{H}$ is nonzero. Let $\mathbf{m}_1$, $\mathbf{m}_2$ be point correspondences in respective images. We may use these correspondences to determine whether the directional vectors of $\mathbf{l}_1$ and $\mathbf{l}_2$ aim towards the images of the same scene points. To do this we compute the dot products $f_{1,2} = \mathbf{l}_{1,2}^\top\,\mathbf{m}_{1,2}$, if they match in sign, the epipolar line transfer by homography $\mathbf{H}$ is oriented, i.e. the directional vectors of half-epipolar lines aim towards the respective images of the same scene point, if not we simply change the sign of $\mathbf{H}$.

One epipole inside and one outside.

Let us assume that the reference image has an epipole inside it. Then the common region is determined only by the epipolar lines passing through the outer corners $\mathbf{c}^{3,4}$ of the target image. We then get matching epipolar lines in the reference view $\mathbf{l}_1^{3,4} = \mathbf{F}^\top\,\mathbf{c}^{3,4}$ and apply concept of oriented epipolar geometry described in previous part to decide on corresponding half-epipolar lines.

Having determined the lines bounding the common region we will now denote one pair of corresponding half-epipolar lines $\mathbf{l}^B_{1,2}$ and the other $\mathbf{l}^E_{1,2}$, where B, E stand for begin and end respectively. These lines intersect the image boundary in points $\mathbf{b}^B_{1,2}$ and $\mathbf{b}^E_{1,2}$, Figure 3.3. We will refer to them as the points where the sweep begins and ends.

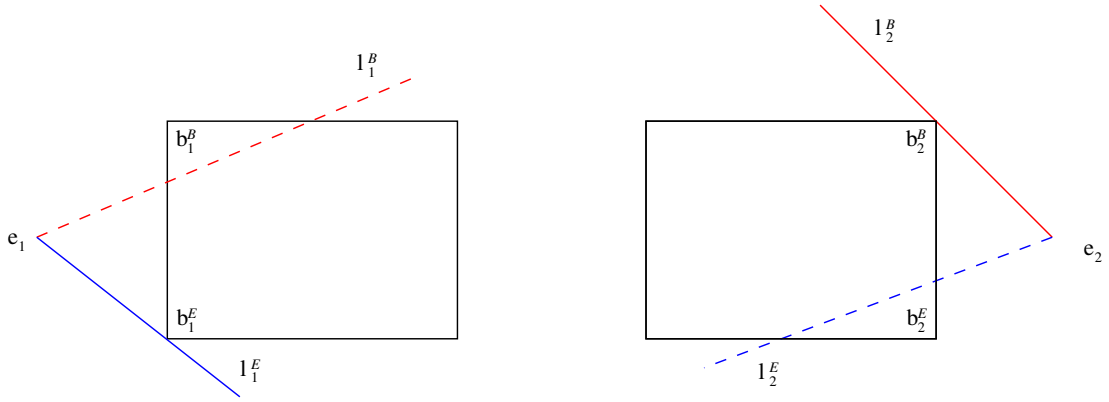


Figure 3.3 Common region is defined once the extremal epipolar lines are detected.

Sweeping the image

Once the common region is determined the rectified images are built up step by step, mapping corresponding epipolar lines to the same row in respective rectified images. To determine the distance between subsequent epipolar lines, simple condition of no pixel loss is imposed, therefore the perpendicular distances $\mathbf{d}_{1,2}$ between two consecutive epipolar lines in respective images must be less than or equal to 1 pixel, Figure 3.4. Starting from a point $\mathbf{p}_1^{\text{old}}$and setting the distance $\mathbf{d}_1^1$ to 1 we compute the candidate for a next consecutive epipolar line by first getting $\mathbf{p}_1^{\text{new}}$ in the reference image. We then compute a corresponding epipolar line in the target view and determine the perpendicular distance $\mathbf{d}_2^1$. If it is less than or equal to 1 pixel, we have found the epipolar lines in respective views. If it is greater than 1 pixel, we determine an epipolar line with perpendicular distance 1 pixel from a point $\mathbf{p}_2^{\text{old}}$ and transfer it to the reference view. Without further checking we have found the epipolar lines in question. This way the algorithm sweeps the common region saving the coordinates of sampled boundary in each step for both images.

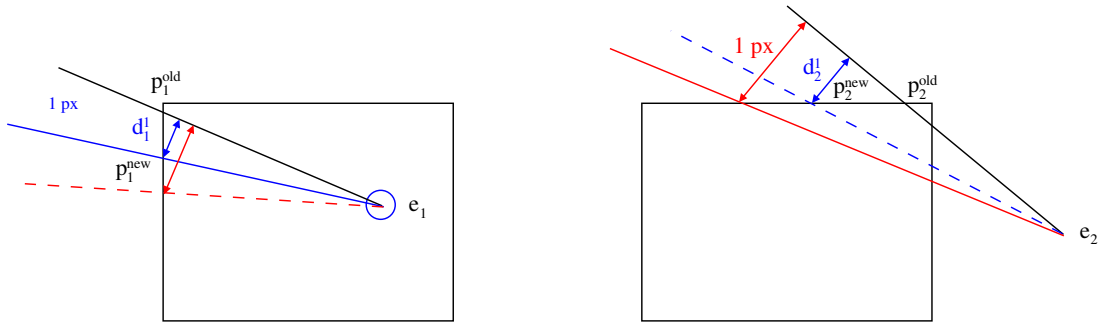


Figure 3.4 Consecutive epipolar lines distance detection.

End of sweep

To ensure that the algorithm recognizes the end of common region, i.e. all the image boundary of common region has been sampled a dot product of vectors $\mathbf{v}_{\text{stop}}$ and $\mathbf{v}_{\text{current}}$ is employed. A dot product can be seen as a projection of one vector into a direction of another. Its value varies from maximum negative value, when vectors face oposite directions, and maximum positive value when vectors coincide.

Let us consider a case with an epipole outside the image. Here, as we sweep the image towards the terminal point, the dot product grows at the end reaching a maximum. Thus the algorithm will terminate the sweep when it detects a change in monotony.

When both epipoles are inside the image, we wish to sweep across whole area of image and terminate at the starting point. At first the dot product decreases reaching a maximum negative value oposite the starting point, from then on it increases up to a maximum positive value at the starting point. Thus we only need to stop the sweep when second monotony change is detected.

In a case when one epipole is inside and the other outside the image, we may rely on detecting the end of sweep in the latter image, though the same reasoning would apply in the first case image, as the extremal epipolar lines must contain an angle lower than 180°. See Figure 3.5 for concept illustration.

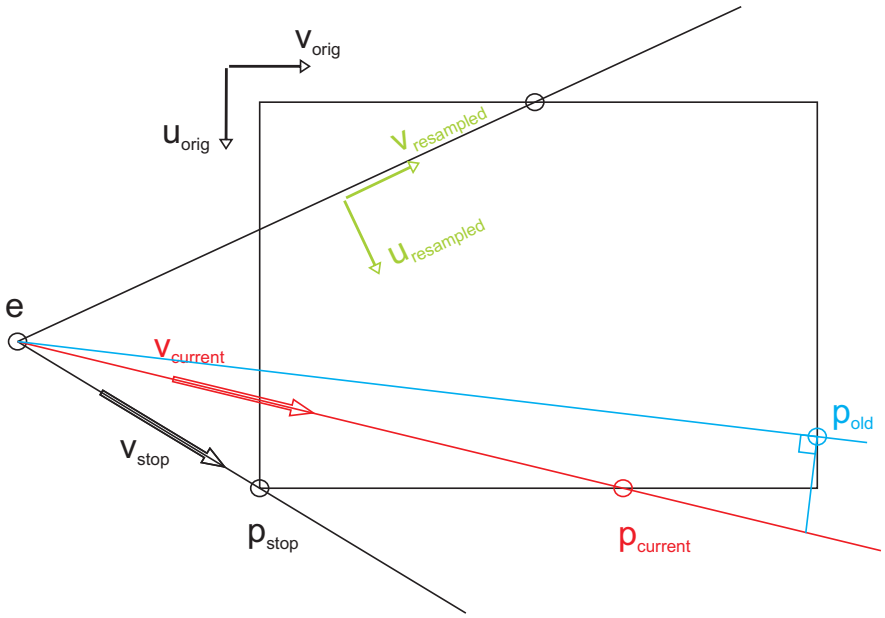


Figure 3.5 End of sweep detection.

Epipolar line read-out

At this point we have obtained a list of points $\mathbf{p}_1^i$, $\mathbf{p}_2^i$, samples' coordinates of the image boundary of respective images, and may proceed with reading out the half-epipolar lines defined by $\mathbf{l}_{1,2}^i = [\mathbf{e}_{1,2}]_\times \mathbf{p}_{1,2}^i$.

The direction of epipolar line read-out is uniquely determined by the coordinate system of original image and by the sense of rotation while sweeping the image. The original image uses a coordinate system $\mathbf{u}_{\text{orig}}$, standing for row coordinate, and $\mathbf{v}_{\text{orig}}$, standing for column coordinate. In order for the rectified image not to be mirrored we need to preserve the handedness of the original image, thus coming to a coordinate system of the rectified image $\mathbf{u}_{\text{resampled}}$ denoting row coordinate, and $\mathbf{v}_{\text{resampled}}$ denoting column coordinate. Figure 3.5 demonstrates this concept on one epipolar line.

Let us consider one image only. We sample the epipolar line $\mathbf{l}_i$ with one pixel being the distance between two consecutive samples to acquire coordinates $\mathbf{s}_{i,j} = [u^{i,j}_{\text{orig}}, v^{i,j}_{\text{orig}}]$ in the original image, where i denotes a row index and j a column index of the sample in the rectified image, Figure 3.6. Applying the above process in both images yields a pair of look-up tables, where i^{th} rows are corresponding sampled epipolar lines $\mathbf{l}^i_{1,2}$. To get functional values of rectified images the original images are interpolated at points specified in look-up tables.

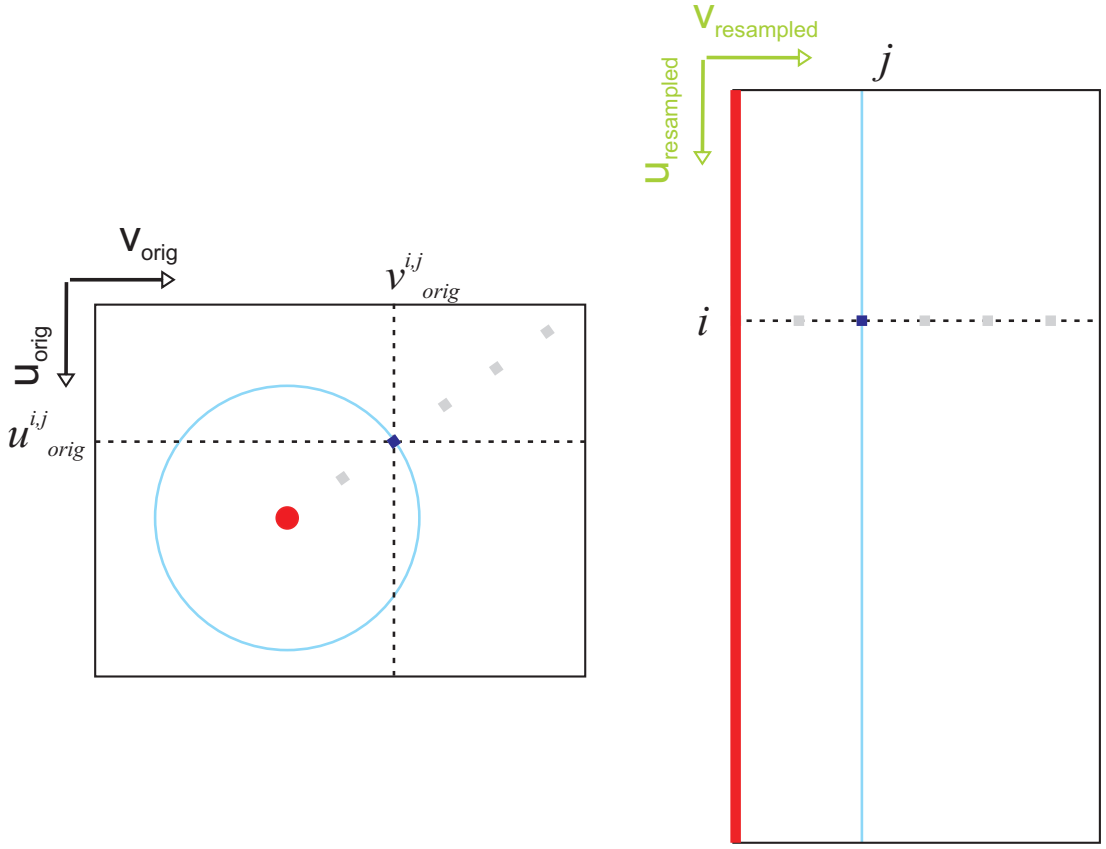


Figure 3.6 Sampling of the epipolar line and construction of the rectified image. Note that polar parameterization around the epipole is used to obtain the coordinates look-up table for resampled image. A circle centered at the epipole is mapped onto a column in the rectified image.

Rectification capabilities of our implementation

A generator of camera configurations was developed to thoroughly test the functionality and performance of our implementation. Given the two epipoles and preset camera internal parameters it computes the camera matrices and fundamental matrix for such

configuration. The camera matrices are used to obtain synthetic images of virtual scene. An illustration of the algorithm performance using synthetic input data are shown in Figure 3.7.

In this chapter we have discussed the fundamentals of epipolar rectification. Considering all the relevant information we have chosen rectification methods suitable for reversing camera and talked about concrete implementation decisions taken.

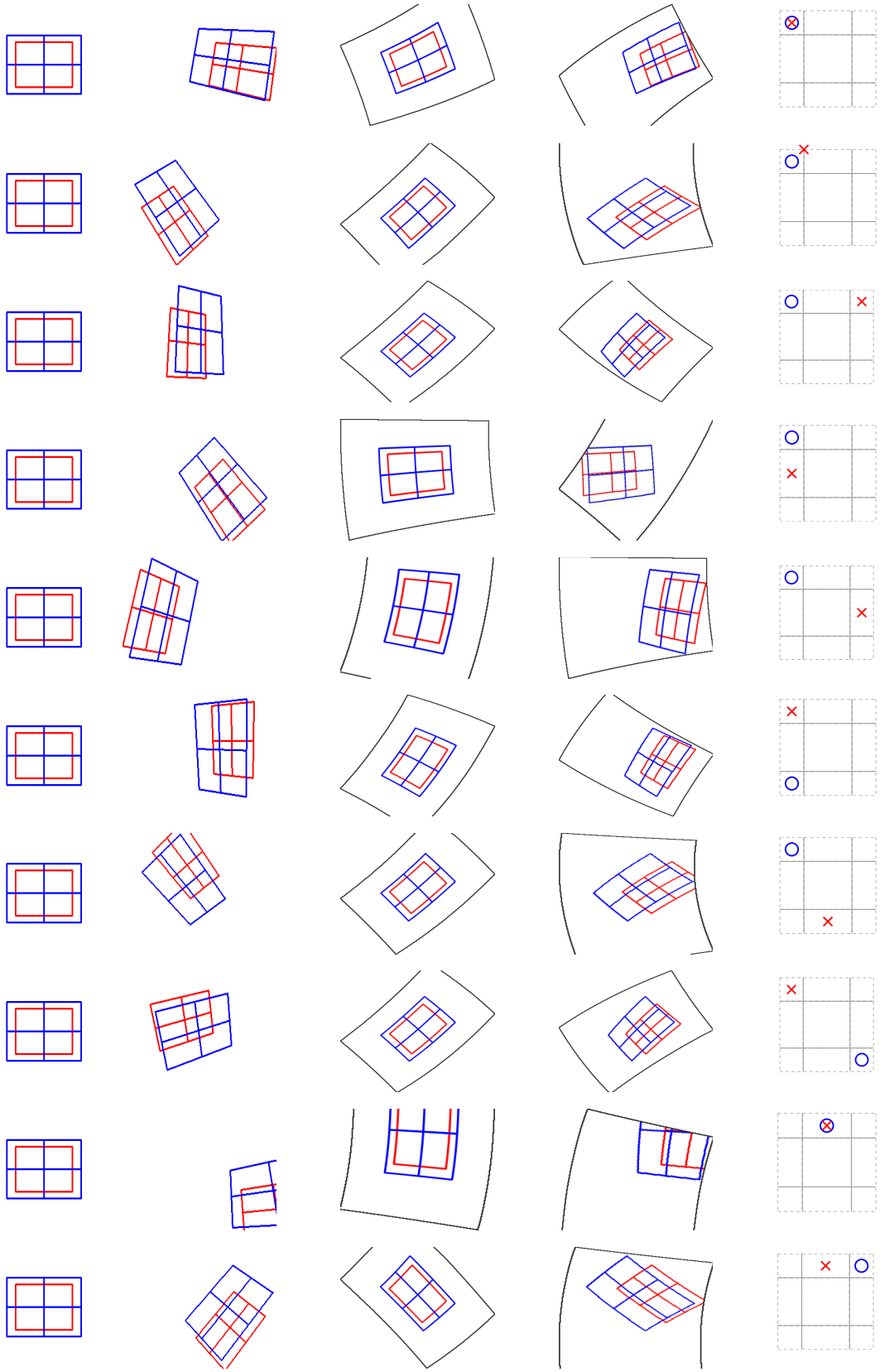


Figure 3.7 Epipolar configurations arising from a general motion of the camera. Synthetic refrence and target images are shown in the first two columns. Second two columns present the respective rectified images. The fifth column marks the positions of respective epipoles in one of the nine possible segments of the image plane. A circle denotes the position of reference image epipole with respect to the image, a cross the position of target image epipole.

4 Optimized sampling

The motivation behind the idea of optimized sampling was to design such a sampling scheme, that would either speed up the stereo matching algorithm without significantly distorting the resulting disparity maps. This may be achived by reducing the area of the rectified images. Now we face a question: How do we know which part of the rectified image may be reduced without affecting the disparity map?

4.1 Properties of an image

Typical image has some areas that carry significant amount of information and other areas that lack it. Figure 4.1 illustrates the idea. Looking at the image we see that sky area does not contain any valuable information from scene reconstruction point of view, because the stereo algorithm will be hardly able to find any matches. On the other hand we can assume that processing chimneys will result in many detected correspondences.

Figure 4.1 Different areas of an image contain different amount of information.

If we now consider an amplitude spectrum of such an image computed locally in window of certain size, we see that areas of interest produce peaks located on higher frequencies, whereas significant peaks of areas with little information are cumulated around the zero-frequency component.

Before further exploration of the image amplitide spectrum, we need to consider, what in fact happens with a given area of an image in the process of rectification.

4.2 Rectification transform details

Let us first take a more detailed look on the rectification transformation. Consider an image I, an oriented half-epipolar line $\mathbf{l}^\beta$ containing an angle β with the horizontal image boundary $\mathbf{l}^{hb} = [\mathbf{c}^1]_\times \mathbf{c}^2$, where $\mathbf{c}^{1,2}$ are corners of the image. Let $\frac{a}{2}$ be a perpendicular distance between $\mathbf{l}^\beta$ and $\mathbf{l}^1$ measured at pixel center and $\frac{A}{2}$ be this distance measured on the image boundary, Figure 4.2. A pixel is subject to rectification transformation $\mathbf{T}$, which may be locally aproximated by rotation $\mathbf{R}$ and scaling $\mathbf{S}$, such that $\mathbf{T} = \mathbf{S}\,\mathbf{R}$. The transformation matrices $\mathbf{R}$,$\mathbf{S}$ are determined by the angle β and the distance a, so that

$$\mathbf{R} = \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix}, \tag{4.1}$$

$$\mathbf{S} = \begin{bmatrix} \frac{1}{a} & 0 \\ 0 & 1 \end{bmatrix}. \tag{4.2}$$

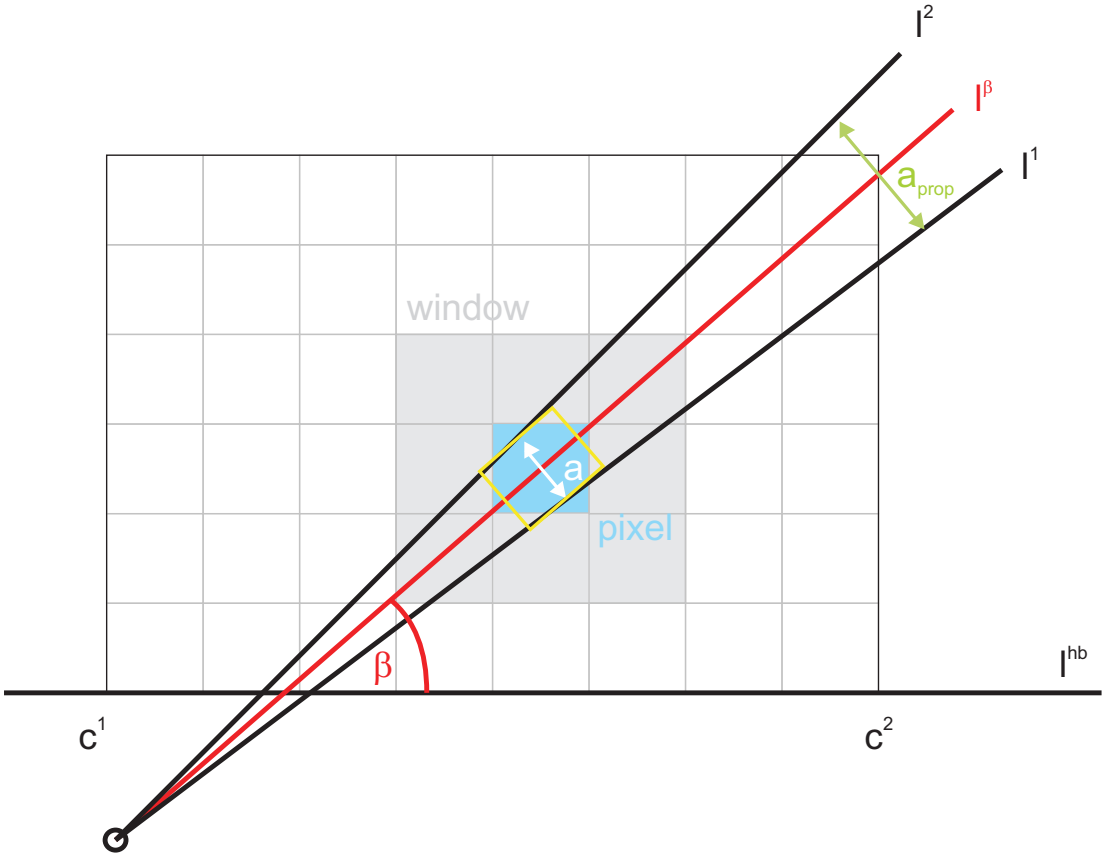


Figure 4.2 Details of pixel transformation in the rectification process.

We will now look at the transformation as a sequence of four steps, when firstly 2D image function is reconstructed in continuous domain, secondly the continuous image function is transformed, thirdly the transformed image function is filtered by a low-pass filter, before producing resulting image by sampling. We see that filter application prior to sampling according to Nyquist sampling theorem is a part of the transformation, that may result in spectrum losses. If we could measure these losses, it would give us an idea of how severely the spectrum of rectified image is distorted [10].

We will employ Bracewell's affine theorem [1], according to which, if we apply a linear tranformation expressed by matrix $\mathbf{T}$ on the image I to obtain an image $I^{'}$, then the amplitude spectrum of $I^{'}$ will undergo a linear transformation of $\mathbf{T}^{-\top}$. This can be rewritten as $\mathbf{T}^{-\top} = (\mathbf{S}\,\mathbf{R})^{-\top} = \mathbf{R}^{-\top}\,\mathbf{S}^{-\top} = \mathbf{R}\,\mathbf{S}^{-\top}$. Note that while in image domain the scaling factor is $\frac{1}{a}$ in frequency domain it becomes an a.

Let us now consider an image $I^{''}$, which is obtained from $I^{'}$ by backward transformation $\mathbf{T^{-1}}$. The spectrum of such an image is in fact a clipped spectrum of original image. The clipped area of spectrum is determined by Nyquist bounding box transformed by $\mathbf{T}^{\top}$ according to Bracewell's affine theorem. Inversely, looking at the spectrum of the image I, we may directly infer, what part of the spectrum will be clipped off during the rectification transformation.

4.3 Spectral loss function

We have seen, that it is possible to establish a measure of spectrum losses introduced by rectification transformation. Now it is at hand to compute such measure with reagrds to the epipolar geometry, so that it would tell us what sampling should be employed in specific image areas.

Let $\mathbf{p}$ be an image point lying on $\mathbf{l}^{\beta}$. Let β and A be the parameters of the transformation as defined previously. Let L be a distance from the epipole to a furthermost point on the image boundary measured on a line $\mathbf{l}^{\beta}$ and l be a distance from the epipole to the point $\mathbf{p}$. Then the parameter a at point $\mathbf{p}$ is given by $a = \frac{l}{L}\,A$.

Let W be a square window of size $w \times w$, where w is an odd positive integer. Then $I^{W}_{\mathbf{p}}$ is a part of the image I cut out by window W centered at $\mathbf{p}$ and $\mathcal{F}\{f^{W}_{\mathbf{p}}\}$ is a Fourier transform of continuous function $f^{W}_{\mathbf{p}}$ reconstructed from $I^{W}_{\mathbf{p}}$. According to Matoušek [10] we may now compute a local spectrum loss at a point $\mathbf{p}$ as follows:

$$\lambda(\mathbf{p} \mid \mathbf{T}) = \iint_{\overline{S}_{(\mathbf{T},\mathbf{p})}} h\Big(\,|\,\mathcal{F}\{f^{W}_{\mathbf{p}}\}(u)\,|\,\Big)du, \tag{4.3}$$

where $\overline{S}_{(\mathbf{T},\mathbf{p})}$ denotes an area of lost frequencies at a point $\mathbf{p}$ given a transformation $\mathbf{T}$ and $h(\cdot)$ is a frequency weight kernel.

Let us now consider a set of equally spaced points $\mathbf{p}_i$ on the line $\mathbf{l}^{\beta}$. A point $\mathbf{p}_i$ and its near surroundings are subject to transformation $\mathbf{T}_i$, which may be expressed as $\mathbf{T}_i = \mathbf{S}_i\,\mathbf{R}$. Since rotation remains the same for all the line $\mathbf{l}^{\beta}$, the only difference is the scaling factor $a_i = \frac{l_i}{L}\,A$, where l_i stands for a distance between the epipole and a point $\mathbf{p}_i$. We now define a spectrum loss Λ^{β} along the epipolar line $\mathbf{l}^{\beta}$.

$$\Lambda^{\beta} = \sum_{i} \frac{l_i}{L} \lambda(\mathbf{p}_i \mid \mathbf{T}_i)\,, \tag{4.4}$$

where factor $\frac{l_i}{L}$ is a weight, saying that the closer the area is to the epipole the lower significance it has. This comes from a fact that the density of epipolar lines grows inversely to the distance from the epipole, thus resulting in larger overlaps. To compensate for these large overlaps we introduce the above mentioned weight.

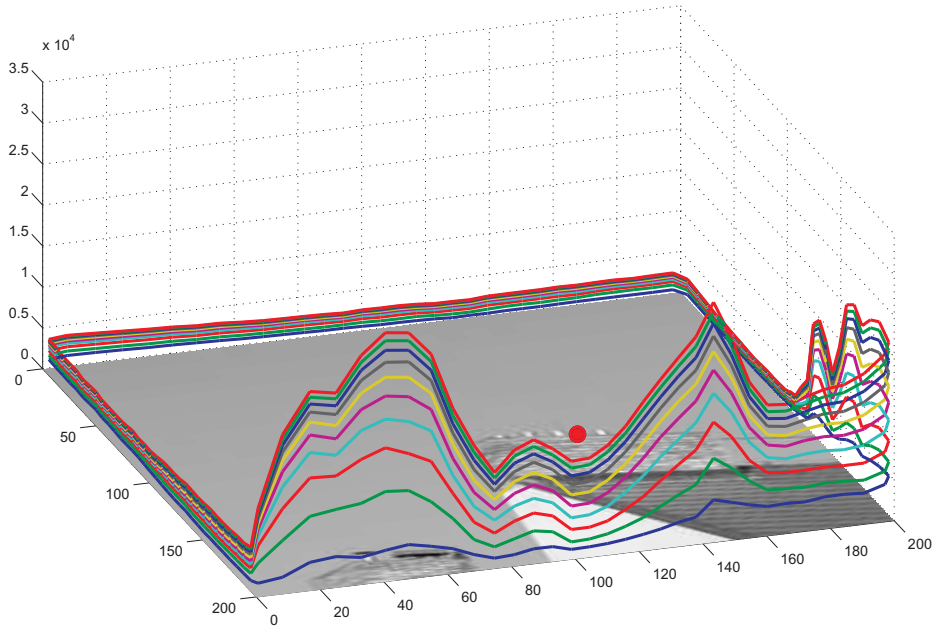


Figure 4.3 Loss function look-up detail. Different colors stand for a different distance between two consecutive epipolar lines, i.e. the parameter A_k.

To define a spectrum loss function for all the image, Figure 4.3, we start with a set of half-epipolar lines $\mathbf{l}^{\beta_j}$, where consecutive lines are separated by a constant perpendicular distance on the image boundary. Each such line defines an angle β_j and a point on the boundary $\mathbf{b}^j$. We may want the perpendicular distance to be less than or equal to a half of the window size w in order to provide a sufficient overlap between two consecutive windowed image parts. We now compute a spectrum loss along each epipolar line $\mathbf{l}^{\beta_j}$ according to (4.5),

$$\Lambda^{\beta_j} = \sum_i \frac{l_{i,j}}{L_j} \lambda(\mathbf{p}_{i,j} \mid \mathbf{T}_{i,j}), \tag{4.5}$$

where $l_{i,j}$ is a distance measured from the epipole to the point $\mathbf{p}_{i,j}$, L_j is a distance from the epipole to the point on the boundary and $\mathbf{T}_{i,j}$ is a transformation defined by scaling factor a_i and an angle of rotation β_j.

The values Λ^{β_j} are in fact samples of some continuous function. This function is computed for certain value of parameter A_k, i.e. for certain perpendicular distance between consecutive epipolar lines. We will denote such function as $\Lambda(\beta \mid A_k)$.

Total spectrum as a function of epipolar lines

Let us now define a local spectrum total as

$$\tau(\mathbf{p}_i) = \iint_S h\Big(\mid \mathcal{F}\{f_{\mathbf{p}}^W\}(u) \mid \Big) du, \tag{4.6}$$

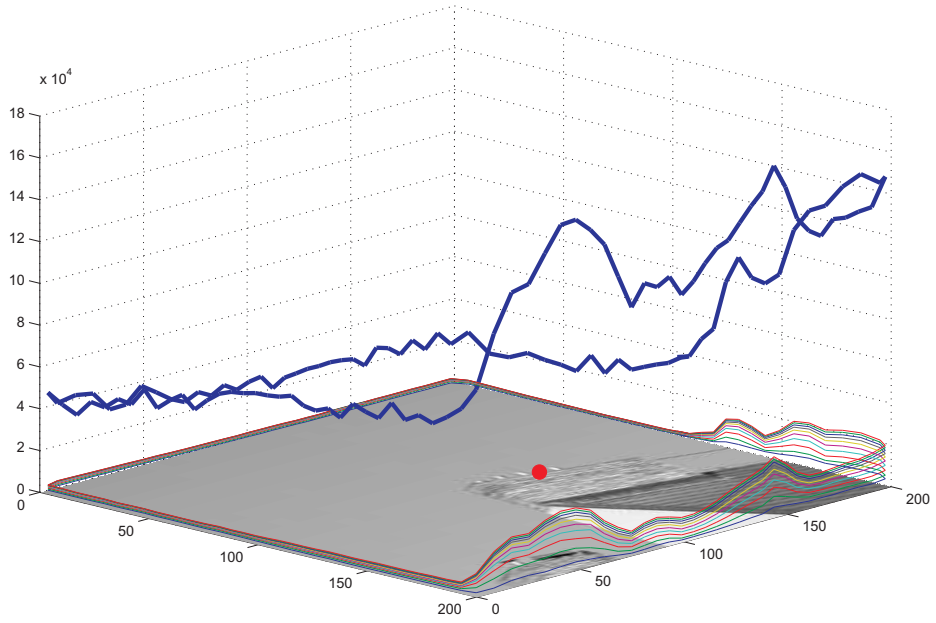


Figure 4.4 Total spectrum (thick blue curve) accumulated along the epipolar lines along with lost frequencies for different parameter A. The value computed for a particular wedge segment of epipolar lines shows a hypothetical loss for a case when whole segment is left out during the rectification transformation. The red spot marks the position of the epipole.

where S is an area of whole amplitude spectrum $\mathcal{F}\{f^W_{\mathbf{p}}\}$. Similarly to spectrum loss along the epipolar line we obtain spectrum total along the epipolar line

$$\mathrm{T}^{\beta} = \sum_i \frac{l_i}{L} \tau(\mathbf{p}_i). \tag{4.7}$$

Finally we compute the spectrum totals T^{β_j} for each epipolar line $\mathbf{l}^{\beta_j}$. Summing these values yields a total spectrum T computed along the epipolar lines. Figure 4.4 shows both total spectrum and spectrum loss function computed in the manner described above.

4.4 Sweeping the image according to the loss function

In order to rectify the images with regards to their local spectral content, we first need to retrieve these spectra and set other important parameters of the transformation and then we move on to a decision scheme working on the basis of these data and parameters.

Parameters

Let B be a total length of the boundary between the two points $\mathbf{b}^B$ and $\mathbf{b}^E$ defined by common region and let b^j be a distance between a point $\mathbf{b}^j$ on the boundary and a point where the sweep begins $\mathbf{b}^B$ measured along the boundary, Figure 4.5.

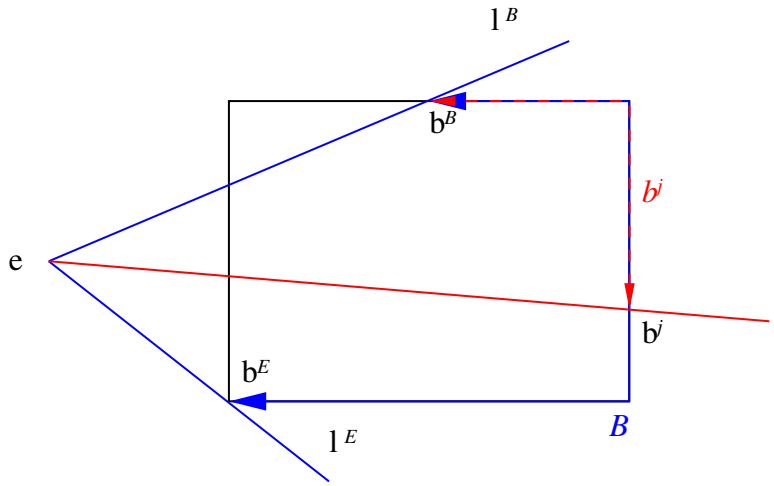


Figure 4.5 Loss function parameters illustration.

We may now wish to limit a total amount of lost spectrum over all the image by a certain number computed as $\Delta = \eta\,\mathrm{T}$, where $\eta \in \langle 0, 1\rangle$. We define a maximum allowed loss per unit length as $\delta = \frac{\Delta}{B}$. We compute a total amount of lost spectrum cummulated up to a point $\mathbf{b}^j$ as $\Gamma^{b^j} = b^j\,\delta$.

Spectral content retrieval

Before we proceed further, we need to employ the concept of spectral loss function, which offers a measure of information significance of considered image area. Given arbitrary half-epipolar line $\mathbf{l}^{\beta_a}$ we would like to find such a scheme, that would tell us, with regards to the local spectral content of an image, what is the maximum angular step θ such that the following half-epipolar line is $\mathbf{l}^{\beta_a+\theta}$. An approach, that is based on retrieval of spectral loss function samples, was chosen.

Let us consider a growing sequence of parameters A_k. Each such parameter defines a continuous spectral loss function $\Lambda(\beta \mid A_k)$ in a given image. These functions were sampled at points $\mathbf{b}^j$ to obtain a look-up table, where to each point $\mathbf{b}^j$ and distance b^j corresponds a set of samples $\Lambda^{\beta_j}_{A_k}$. From now on by LUT we will understand such look-up table and by LUT^{b^j} a set of samples at b^j.

Decision scheme

Let us now consider a pair of corresponding half-epipolar lines $\mathbf{l}^p_{1,2}$ defining points on the boundary $\mathbf{b}^p_{1,2}$ in both images. In a reference image we have LUT_1 given at points $\mathbf{b}^m_1$, spectral loss per unit length δ_1 and cummulated lost spectrum $\Gamma^{b^p}_1$. For target image we have LUT_2 at $\mathbf{b}^m_2$, δ_2 and $\Gamma^{b^p}_2$. To obtain a pair of following half-epipolar lines $\mathbf{l}^q_{1,2}$ we will first decide on the maximum angular step in each image separetely and then choose such step, that would comply with the condition of maximum allowed spectrum loss in both images.

Maximum angular step in one image

We project the perpendicular distances A_k onto the image boundary to obtain $A_{k,proj}$. Then a linear estimate of cummulated spectral loss is computed for each $A_{k,proj}$ such that $\Gamma_L^{c_k} = c_k\,\delta$, where $c_k = b^p + A_{k,proj}$. We now use the look up table to find out function values LUT^{c_k} at distances c_k and compute a functional estimate of cummulated spectral loss as $\Gamma_F^{c_k} = \Gamma^{b_p} + LUT^{c_k}$. We may now compare the values of linear $\Gamma_L^{c_k}$ and functional $\Gamma_F^{c_k}$ estimate and look for such k for which $\Gamma_F^{c_k} \leq \Gamma_L^{c_k}$. Choosing a maximum k from the sequence obtained and a corresponding value of parameter A_k ensures a maximum angular step θ, while limiting the spectral loss to desired value.

Spectral loss with regards to epipolar geometry

Following the above pattern we compute the values of parameter A for both images and we will denote them A_1 and A_2 for reference and target image respectively. From these parameters we obtain candidate half-epipolar lines $\mathbf{l}_1^h$ and $\mathbf{l}_2^g$ and their correspondences in the other image $\mathbf{l}_2^h$ and $\mathbf{l}_1^g$. These lines are separated by a perpendicular distance A_2^h and A_1^g from $\mathbf{l}_2^p$ and $\mathbf{l}_1^p$ respectively. We now choose either A_1 or A_2 depending on whether $A_1 \leq A_1^g$ or $A_2 \leq A_2^h$ holds. This way we know that, while in one image coming very close to a maximum allowed spectral loss, we will certainly stay under the limit in the other image. This way we have determined a pair of corresponding epipolar lines $\mathbf{l}_1^q$ and $\mathbf{l}_2^q$, thus we are now able to sweep whole area of common region.

4.5 Rectified images obtained considering spectral criterion

Here we present the results of spectral criterion application to rectification. The pair of input images, Figure 4.6, was obtained by a camera forward motion. In Figure 4.7 we may observe the effects of allowing different spectrum loss in rectification transformation. Note the difference between leftmost and middle-left image. In the latter the chimney and roof areas are preserved while the sky area is reduced. The number of rows is reduced by 10%. The sky area is reduced even more in the middle-right image, but here we observe some degradation of possibly significant areas, though it mainly occurs on chimney tubes. The image size was reduced by 36%. The reduction of 55% in size was achived in fourth image. Here the sky area is reduced to 10% of the image size, while in the image rectified by non-spectral Pollefeys algorithm it is 28%. We see that setting the allowed spectrum loss to higher numbers brings a significant reduction of insignificant image area, this is summarized in Table 4.1.

Spectral loss %	0	1	3	5
Image area %	100	90	64	45

Table 4.1 Image size reduction as a function of allowed spectral loss.

For illustration of how the algorithm acts upon having the information about local spectrum see Figure 4.8. The algorithm needs to comply with spectral criterion in both images, while choosing corresponding half-epipolar lines. These two constraints on the maximum perpendicular step become obvious around pixel 200. Here the reference image loss function drops to a valley even with the little peak in the middle. This would lead to a much larger perpendicular step, if it were not for relatively higher loss function

Figure 4.6 Input images which are considered in this section.

values in the target image. But as there is local trough in the function, the algorithm decides for a larger step in comparison to previously swept parts of the image. Similarly we may look at the area starting at pixel 480, where peaky target loss function produces peaky course of perpendicular steps. Note the plateau of maximum steps at the end of reference function, leading to sloping sequence of perpendicular steps in target image. This again is due to the fact,that corresponding epipolar lines must be chosen.

Finally we mention that further reduction of rectified image size may be achieved by low-pass filtering of the image prior to rectification. The proposed spectral-conscious rectification algorithm reacts to noise present in the image. Assuming that the data are corrupted by noise with known or estimated characteristics, we may apply such image filtration which would reduce the amount of noise while not harming the data significantly. To illustrate this idea we show Figure 4.9. Here the size of rectified image was reduced by 22%. More interesting is the reduction of sky and chimney tubes area, where in filtered image it is 22%, while in non-filtered it is 32%.

We have shown that allowing different amount of spectral loss results in dramaticly different size of rectified images. We presented the comparison of perpendicular steps taken and spectrum loss function to prove that areas with little information tend to be sampled coarsely, while areas with high amount of information are preserved. The image size reduction is desirable as the stereo matching algorithm searches for correspondences in smaller area. The question we need to answer is, whether the results obtained are comparable with results obtained from images rectified by Pollefeys' algorithm. This question is addressed in the following chapter along with experimental results.

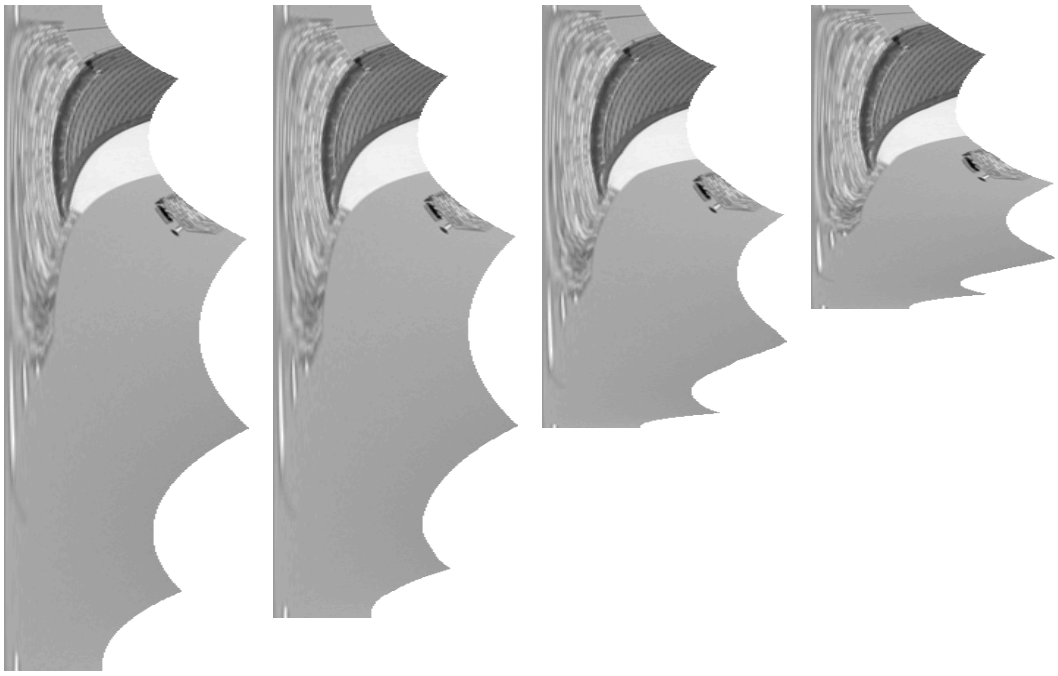


Figure 4.7 Rectified reference image for different values of allowed spectral loss. The images are shown preserving the relative ratios of their heights. Leftmost image: original Pollefeys algorithm, middle-left image: spectral criterion allowing 1% loss of total spectrum, middle-right image: spectral criterion allowing 3% loss of total spectrum, rightmost image 5% loss of total spectrum.

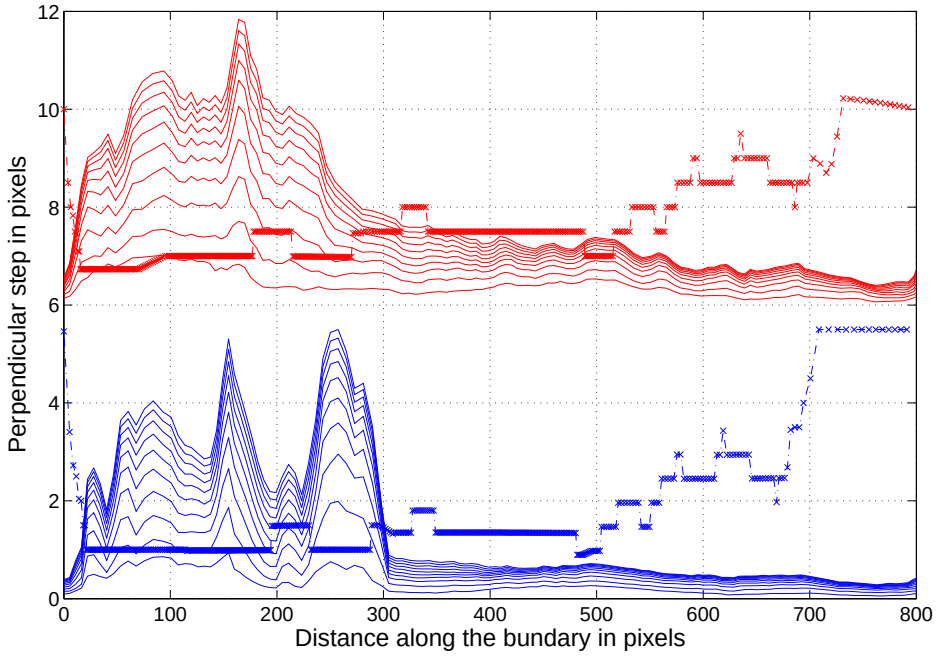


Figure 4.8 Illustration of image sweep and values of perpendicular steps separating consecutive epipolar lines. The loss functions for different values of parameter A are scaled down so that they are comparable with perpendicular steps. These are marked by $\times$ mark and connected by dash-dot line. The data pertaining to reference image are in blue, data pertaining to target image in red. The latter are plotted with 6 pixel offset for better intelligibility.

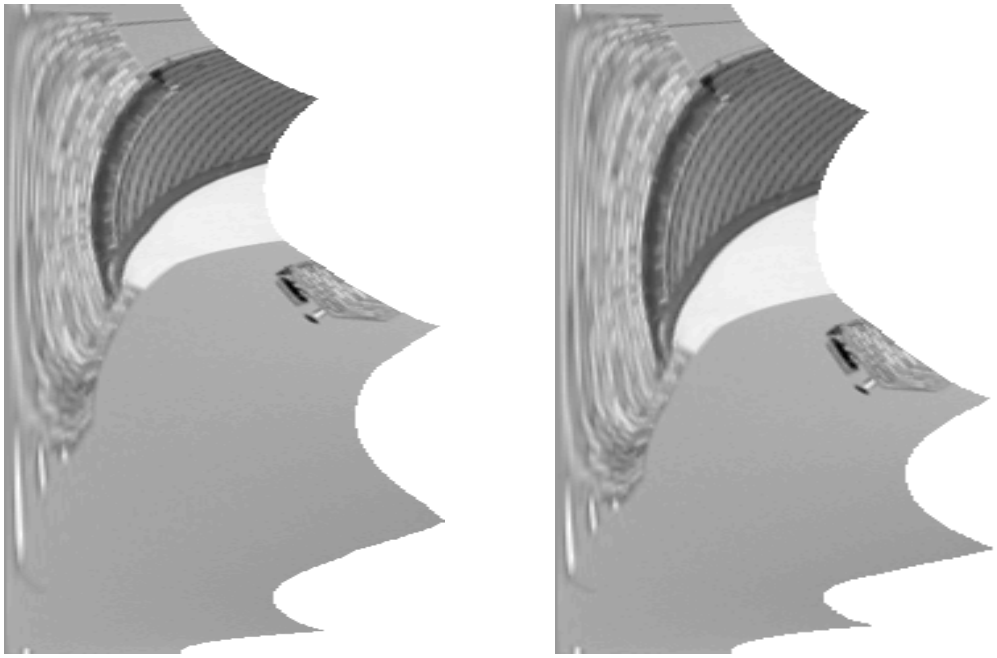


Figure 4.9 A comparison of spectral criterion rectification with and without filtering. Left: input image was not filtered, right: input image filtered with gaussian rotationally symmetric kernel of size 5×5 pixels and standard deviation $\sigma = 0.6$. Both images were subject to rectification transformation allowing a total spectrum loss of 3%. The images are scaled so that it becomes obvious which areas become compressed.

5 Experiments and their results

In this chapter the experiments are desribed and their results are presented. We first look for a measure of accuracy of a disparity map obtained via stereo matching algorithm from images rectified by our implementation of Pollefeys' algorithm. Second we focus on comparison of disparity maps resulting from Pollefeys' algorithm and rectification algorithm using proposed spectral criterion. Third the data obtained by reversing camera used in cars are processed and resulting disparity maps are shown. All disparity maps were obtained by stereo matching algorithm 'GCS' introduced by Čech and Šára [2].

5.1 Disparity map accuracy

The following experiment was carried out on data from *.enpeda..*[1] project [13], which are publically available. The data consisted of image sequences, ego-motion data and time stamps. The image sequnces were obtained by a pair of horizontally aligned cameras, such that the resulting images were almost completely rectified. Precisely rectified images were obtained by standard projective epipolar rectification. This camera setup provided us with ground truth image pair for our experiment. Another pair of images was obtained from the left camera moving in forward direction, Figure 5.1.

The latter pair of images was rectified by our implementation of polar rectification described in previous chapter. We used the rectified horizontal pair of images as a reference set and tested the rectified images obtained from forward cameras against it. Once we obtain the rectified images 5.2 we apply the stereo matching. The result of matching is represented by a disparity map, where to a position of found correspondence in one image we assign a scalar saying where in the target image the matching point is located, specifically:

$$D(u, v_1) = v_1 - v_2\,, \tag{5.1}$$

where $D(u, v_1)$ is a value of disparity at point (u, v_1) computed from column coordinates v_1, v_2 of matching points located on row u. Note that only a scalar is needed, since the images are rectified, i.e. vertical disparity component is always zero.

To see the impact of rectification on disparity map accuracy we employed a ground truth measurement, where as a reference set we considered a pair of images obtained by horizontally aligned cameras $\mathbf{P}_1^l$, $\mathbf{P}_2^l$. The evaluation scheme was:

1. The point correspondences found in forward pair rectified images were triangulated to obtain a cloud of scene points $\mathbf{X}$.
2. Scene points $\mathbf{X}$ were then projected into the cameras $\mathbf{P}_1^l$, $\mathbf{P}_2^l$ of stereo pair. We denote these projections by $\mathbf{u}_1^l p$, $\mathbf{u}_2^l p$.
3. Disparity of projected points $\mathbf{u}_1^{lp}$, $\mathbf{u}_2^{lp}$ was computed as $D(u^{lp}, v_1^{lp}) = v_1^{lp} - v_2^{lp}$.
4. Applying a stereo matching algorithm to the images from cameras $\mathbf{P}_1^l$, $\mathbf{P}_2^l$ produced a disparity map D^G.
5. The disparity map D^G was interpolated at points $\mathbf{u}_1^{lp}$ to get values $D^G(u^{lp}, v_1^{lp})$.

[1] http://www.mi.auckland.ac.nz/DATA/6D/Examples01.htm

Figure 5.1 Cameras' setup used in this experiment. Bottow row: a pair of horizontal images. Left column: two consecutive images from left camera are used to obtain a pair of forward motion images.

6. Values of disparities $D(u^{lp}, v_1^{lp})$ and $D^G(u^{lp}, v_1^{lp})$ were subtracted to obtain an error of disparity $E^D(\mathbf{u}_1^{lp})$ at points $\mathbf{u}_1^{lp}$

The results of this experiment are shown in Figures 5.3 and 5.4. Vast number of error realization lie within the range $< -5, +5 >$ pixels. In relation to the total number of disparities considered the percent occurence moves around 80% depending on the frame considered.

Analysis of disparities' density in reference image prove that disparity content of similar density in relation to the horizontal pair of cameras may be obtained from forward motion cameras with typical value being 80%.

In both camera setups we can observe areas where stereo matching algorithm performance is impaired. In horizontal pair these areas are characterized by little or null texture in the direction of cameras baseline. The problem in forward pair is around the epipole, where very small image area is stretched to a large area in the rectified image and thus the texture is reduced. This may result in many ill estimated disparities in this region and thus become a source of large error of disparites transfered to the horizontal pair.

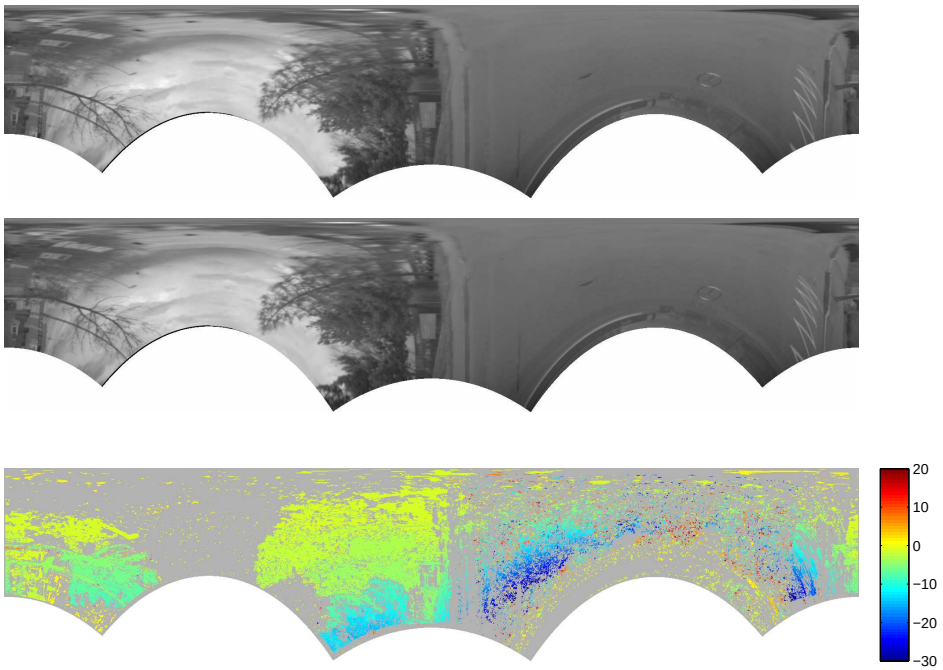


Figure 5.2 Rectified images along with their disparity map. From top: reference rectified image, target rectified image, disparity map. The colorbar shows the mapping of disparity value in pixels to colors. The images are rotated by 90°.

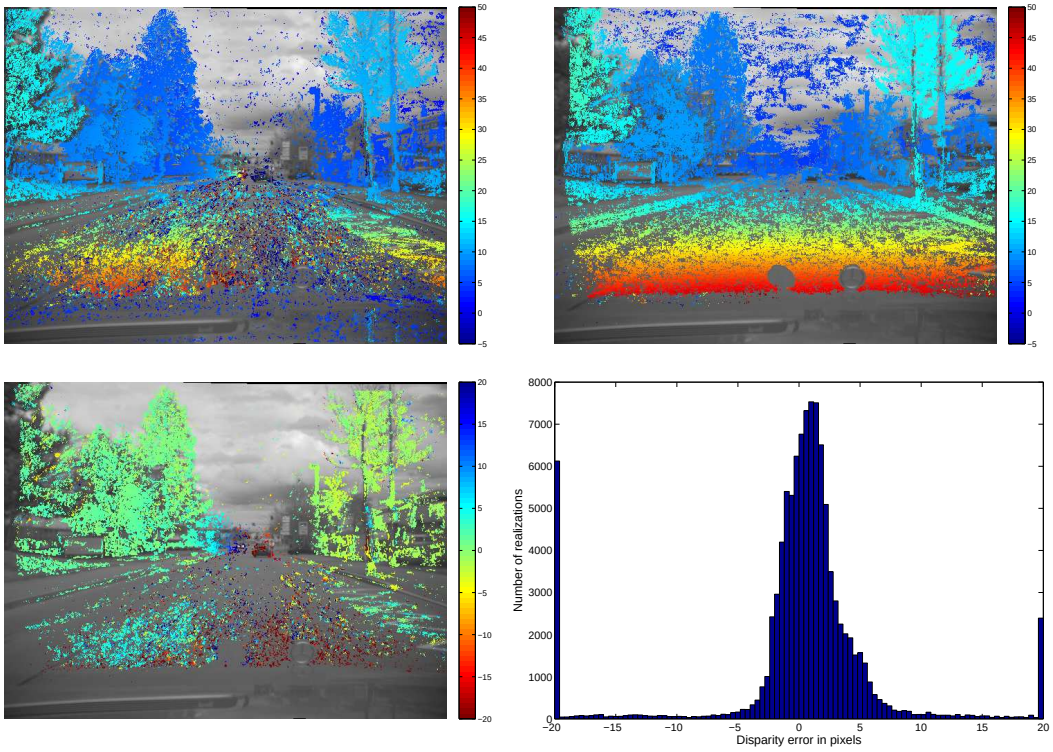


Figure 5.3 Results of disparity map comparison. Here the disparity values are plotted to the left camera of the stereo pair. Top left: disparities resulting from forward motion transfered to the horizontal pair. Note the area of mixed extremal disparities on the ground. Consdidering that the ground is static and is very near to the camera, we should get large positive values of disparity in this region, that is to say that negative values of disparity are wrong. Top right: disparities from stereo pair. The bottom left picture shows the error E at corresponding points. Vast majority of realization lie within $< -5, +5 >$ pixel interval. We see that the area of mixed extremal disparities becomes populated with large errors. Bottom right picture shows the histogram of error E.

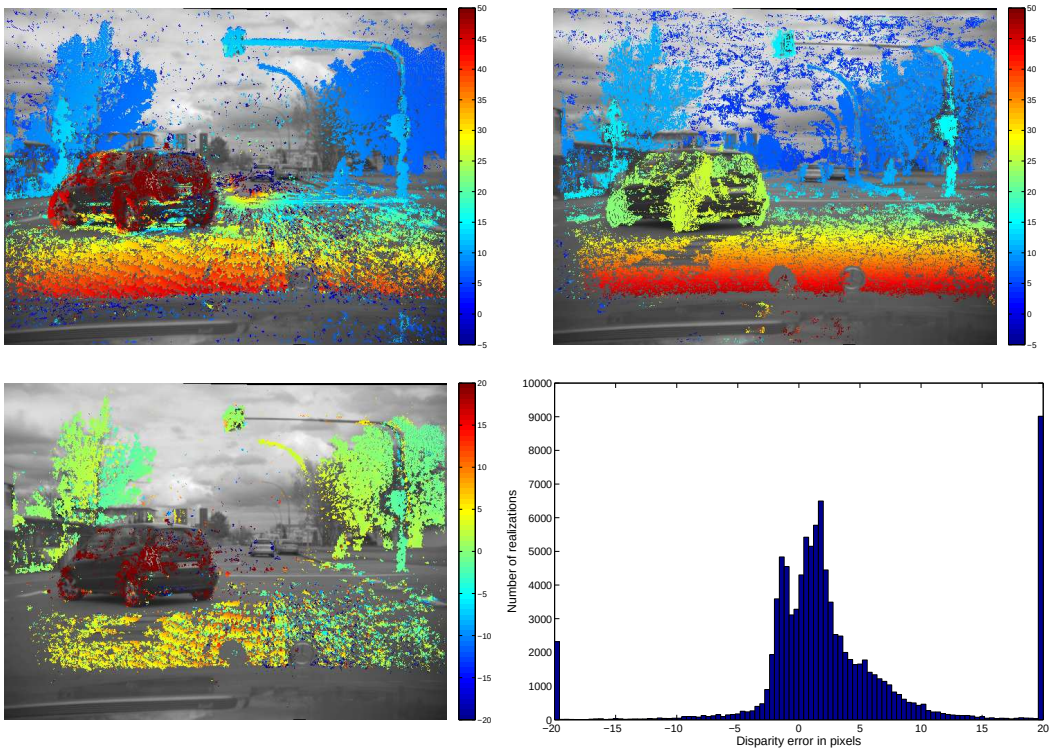


Figure 5.4 Results of disparity map comparison. Same positions of images as in Table 5.3. Note the large number of mathes found on the horizontal bar of traffic light in top right corner of the image in forward motion case and almost none in case of stereo pair. This may be explained by almost no texture along the epipolar line in case of stereo pair, whereas forward motion produces epipolar lines that cut across the horizontal bar at some angle. Interesting atrifact may be seen in error picture. Here the matches on the car produce a very large positive error values. This is caused by the motion of the car in forward camera pictures. An increased number of large positive errors may be accounted to this fact.

5.2 Spectrum motivated rectification and Pollefeys rectification disparity maps

Now we return to the concept of optimized sampling introduced in previous chapter. Having obtained the rectified images allowing for different amount of lost spectrum we compare them to the method not considering spectral information. The comparison is done on resulting disparity maps.

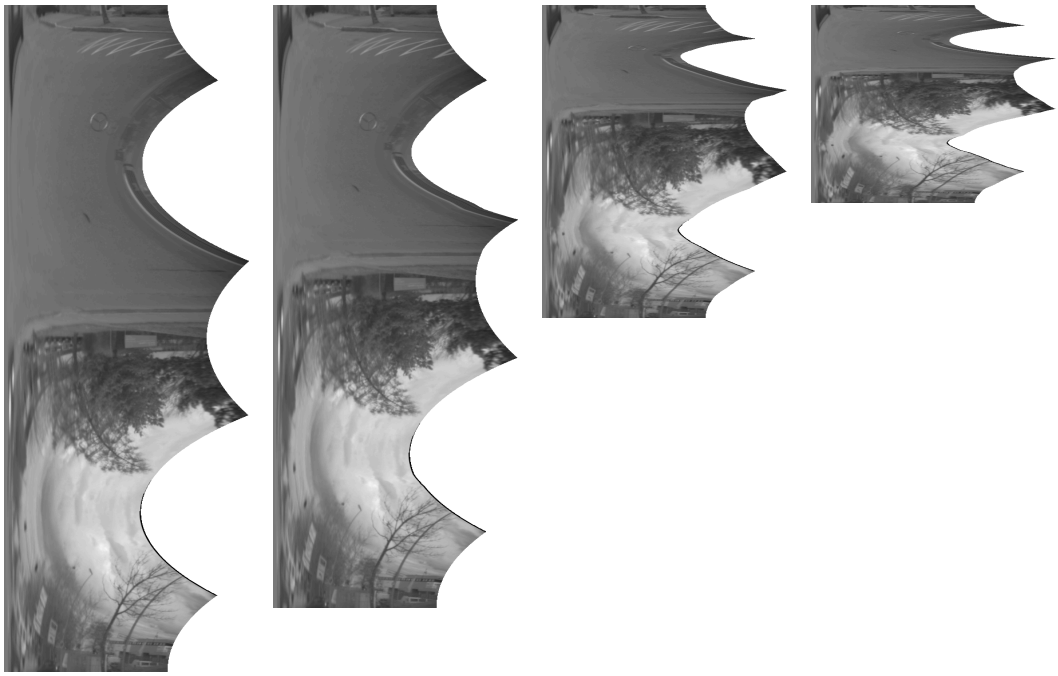


Figure 5.5 Rectified images: leftmost Pollefeys' algorithm, middle-left 1% of total spetrum loss allowed, middle-right 5% of total spetrum loss allowed, rightmost 10% of total spetrum loss allowed. The images are presented maintaining their relative height ratios.

Figure 5.5 shows the resulting rectified images. Allowing a maximum spectral loss of 1%, 5% and 10% reduced the size of images to 90.5%, 46.8% and 29.6% of an image resulting from Pollefeys' algorithm respectively. Note that areas with high texture are preserved. The most significant compression is done on the ground and the sky area. We would prefer the compression of the sky area to the ground compression. But as we may observe there are clouds in the sky that result in significant contrast change along the epipolar lines. This may also be seen on Figure 5.6 where higher loss function values are obtained for this area than for the ground area.

The results of stereo matching are shown in Figure 5.7. Here we may observe that the more dense the disparities are in an area the less likely the area is to be reduced. This certainly is what we tried to achive. Note the areas of sparse disparities in the leftmost image grow denser as we move to the right. This means that the algorithm preserves valuable information even in the areas that are compressed. Detailed versions of disparity maps are in Figure 5.8. Here the Pollefeys' disparity map was scaled down verticaly to match the height of a disparity map obtained allowing a 5% spectral loss.

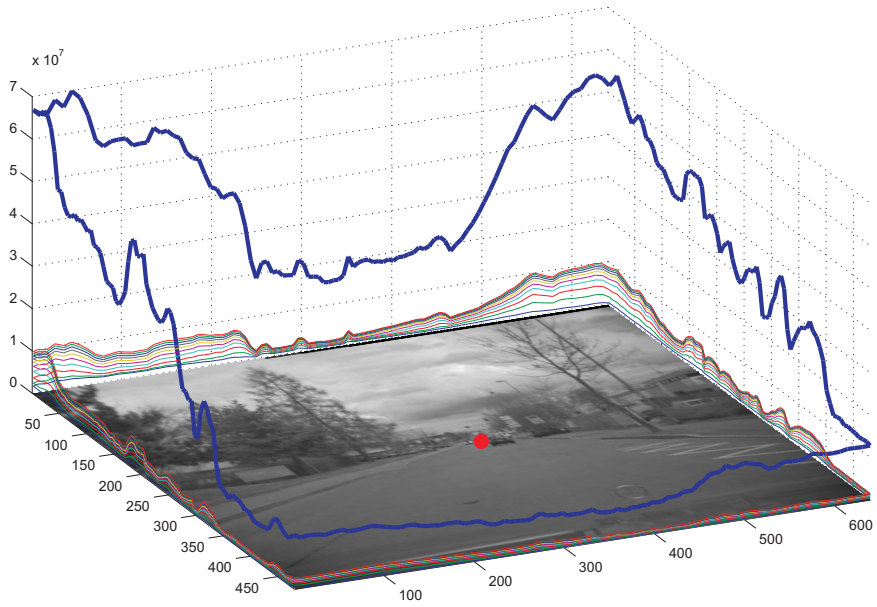


Figure 5.6 Total spectrum accumulated along the epipolar line along with computed loss function. Note the peaks on the left and on the right which in original image correspond to the respective sides of the road, e.g. trees and buildings.

We immediately see, that in the right image, relatively to its height, the disparity dense areas occupy larger parts of the map.

Disparities transfered to their proper positions in the original reference image are shown in Figure 5.9. The difference in top two images is impercievable, note the top-right image is obtained from images with height reduced to 90.5%. We may see that even the bottom-left image is comparable to the top images. The allowed 10% spectral loss lead to severe compression of the image and thus significantly lower density of transfered disparities. Table 5.2 compares the considered rectification processes numerically. The time consumption values say what portion of time unit is necessary for stereo matching and are computed as a ratio of given rectified image number of rows and number of rows in rectified image obtained with Pollefeys' algorithm. The values of time consumption are in relation to Pollefeys' algorithm time consumption which is a unit cost. This table also demonstrates how the disparity map density grows with growing spectral loss allowed. The density is computed as a sum of all found matches, i.e. defined disparity values, divided by total number of pixel in the image. Similarly we compute the density of disparities transfered to original image. Indeed we see that visual impression, that the top two images in Table 5.9 are the same, is supported by the disparity density values, which differ only by 1.2% in favor of Pollefys' algorithm.

Figure 5.10 shows a distribution of disparities in 100^{th} row of all images presented in Figure 5.9. Further visualization of the line is in Figure 5.11, where a detail of first 100 disparity line values are plotted with a relative offset. In both figures we may observe that disparity values of all lines closely match, but different density of disparity samples is achieved. Table 5.2 presents the number of computed disparities in chosen lines in all four images. Again note the similar scores for first two methods. Lines 75 and 100 exhibit an interesting fact, because the number of defined disparities is actually higher

in reduced image. This may be accounted to the fact, that in general each method chooses different half-epipolar lines. Note that one line of disparity values comes from many lines of disparity map, i.e. from many half epipolar lines. Therefore it is possible that the algorithm chose such lines, along which the contrast is better preserved.

	Size	Disparity map density			Time
		Absolute	Transfered	Relative	Relative
Pollefeys	1943 × 428	52.5%	37.6%	1	1
Spec. crit. 1%	1758 × 428	53.7%	36.4%	0.97	0.91
Spec. crit. 5%	910 × 428	57.8%	27.4%	0.73	0.47
Spec. crit. 10%	575 × 428	57.9%	20.1%	0.53	0.30

Table 5.1 Comparison of considered rectification processes. The size of original image was 481×640 pixels. Note large differences in computational cost.

	Row 75	Row 100	Row 200	Row 275	Row 300	Row 400
Pollefeys	284	322	364	353	240	124
Spec. crit. 1%	304	328	362	351	238	90
Spec. crit. 5%	260	293	312	248	171	25
Spec. crit. 10%	187	237	269	148	124	12

Table 5.2 Comparison of considered rectification processes. Here we show a number of detected matches in one row in images from Figure 5.9

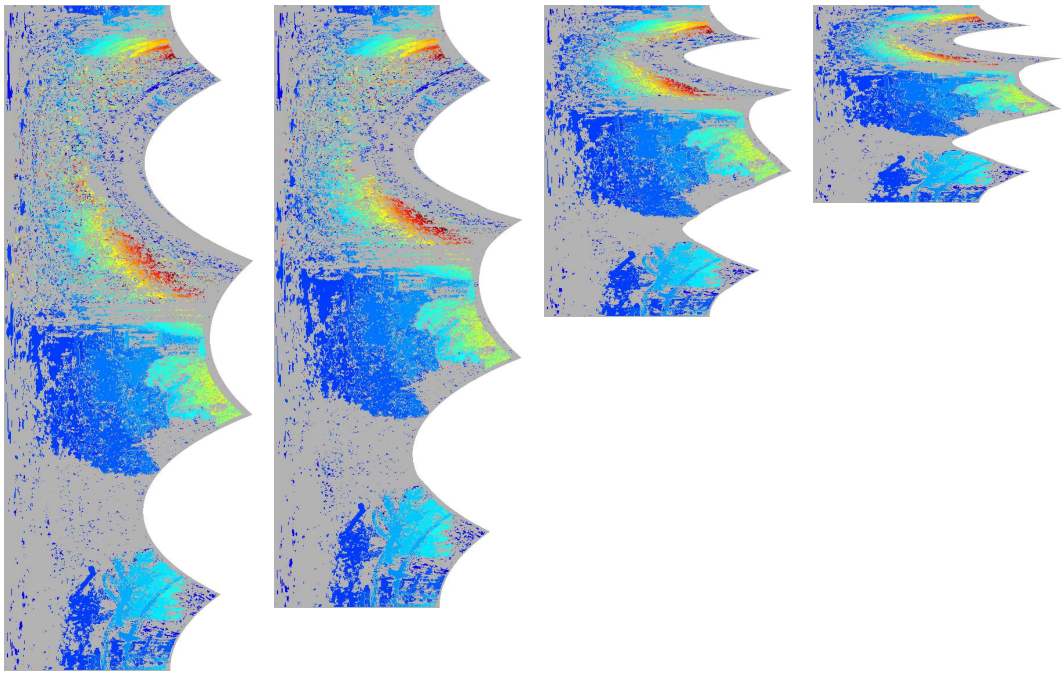


Figure 5.7 Disparity maps corresponding to the images shown in Figure 5.5, same order is followed.

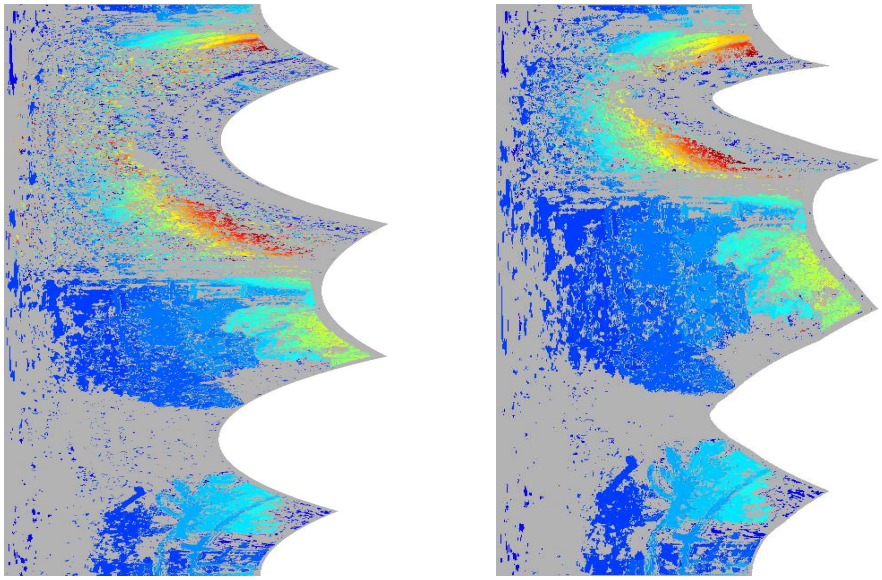


Figure 5.8 Illustration of reduced areas. On the left: a disparity map resulting from Pollefeyes' algorithm. The map was vertically scaled down to match the size of disparity map on the right. On the right: a disparity map resulting from images rectified considering spectral criterion allowing a loss of 5%. We may observe that areas with small number of disparaties in left image get reduced in right image.

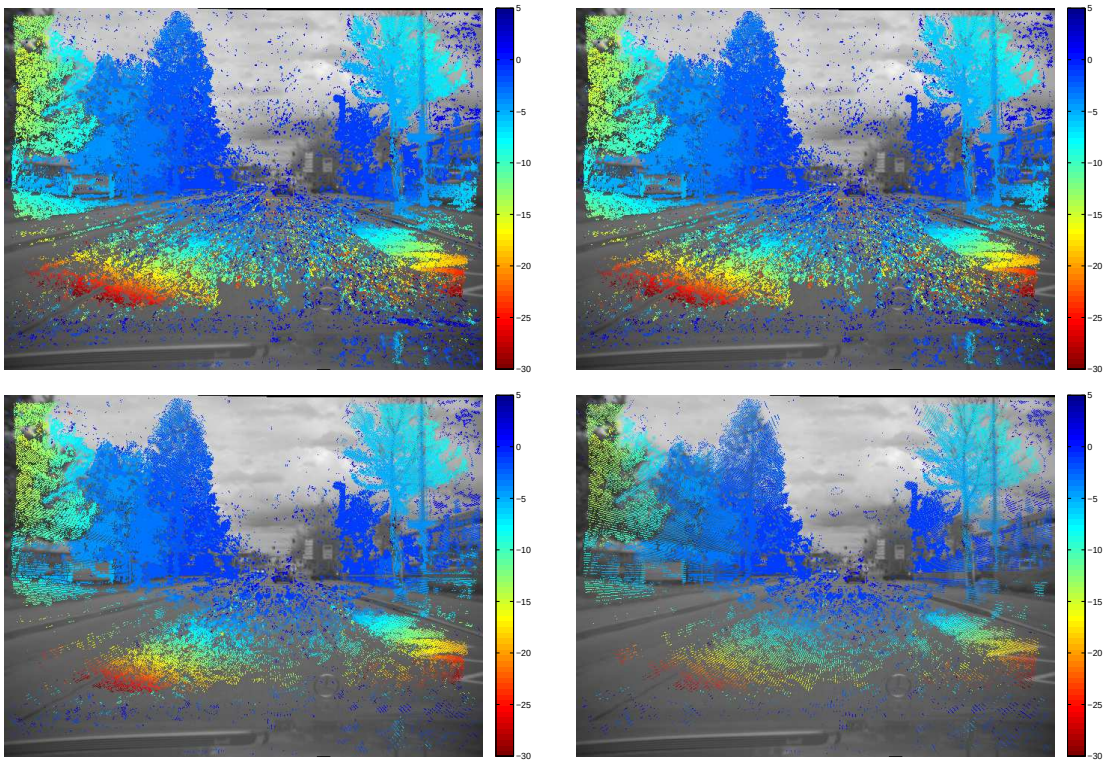


Figure 5.9 Disparities plotted to their proper coordinates in the original reference image. Disparity values originate from disparity maps shown in Figure 5.7. Top-left: Pollefeyes rectification, top-right: 1% spectral loss resulting in height reduction to 90.5%, bottom-right: 5% spectral loss resulting in height reduction to 46.8%, bottom-left: 10% spectral loss resulting in height reduction to 29.6%.

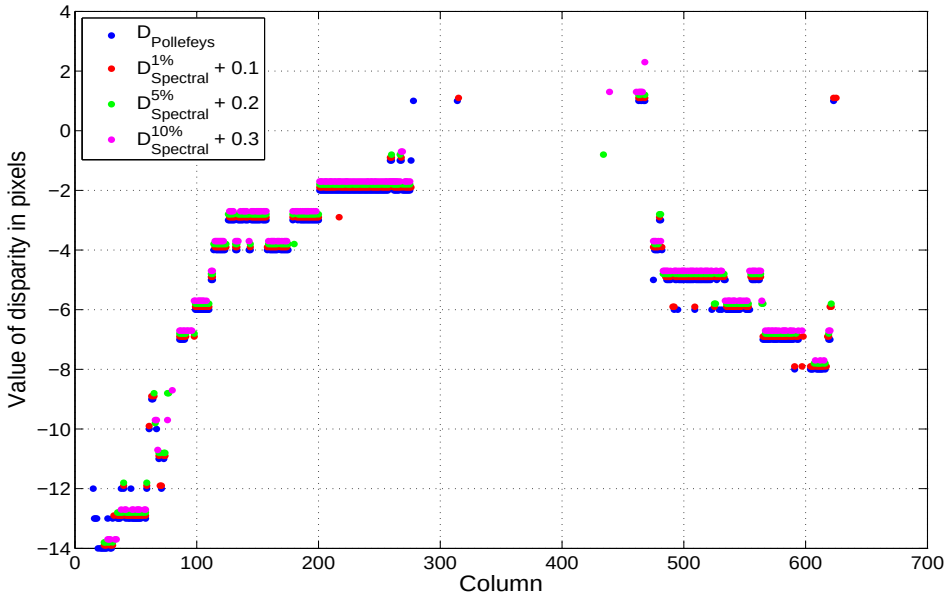


Figure 5.10 A profile of 100^{th} disparity row projected onto the original image for rectification methods: Pollefyes in blue, Spectral 1% loss in red, Spectral 5% loss in green, Spectral 10% in magenta.The values of disparities are plotted with a relative offset 0.1 pixel so that the visual comparison is possible, since the values coincide in most cases.

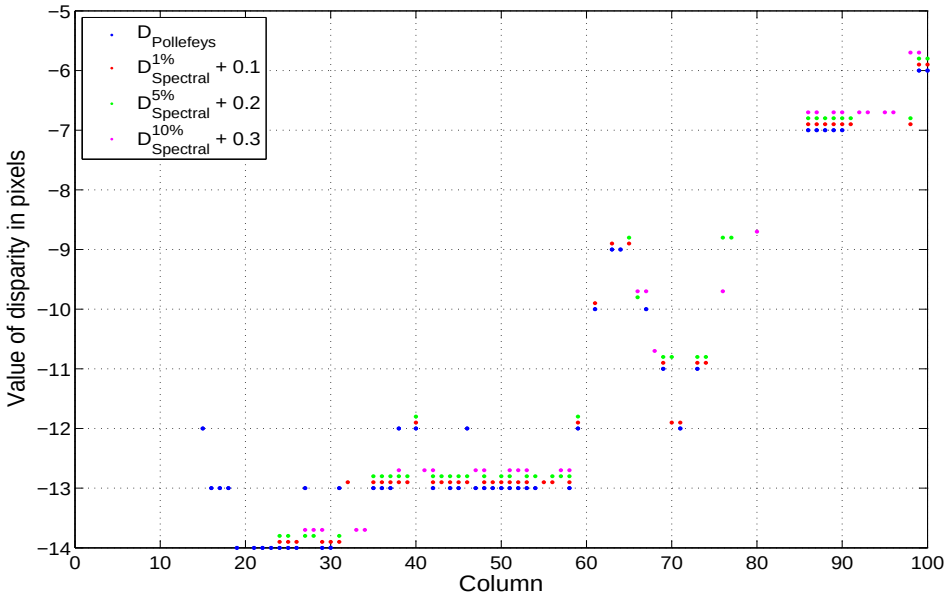


Figure 5.11 A profile of 100^{th} disparity row projected onto the original image for rectification methods: Pollefyes in blue, Spectral 1% loss in red, Spectral 5% loss in green, Spectral 10% in magenta. Detail of first 100 columns. The values of disparities are plotted with a relative offset 0.1 pixel so that the visual comparison is possible, since the values coincide in most cases.

5.3 Rectification of wide field of view cameras

In this section we present the data obtained with reversing camera. The camera was mounted on the rear part of the vehicle. Motion by this camera has a dominant forward (in this case backward) displacement component. Figure 5.12 shows two consecutive frames.

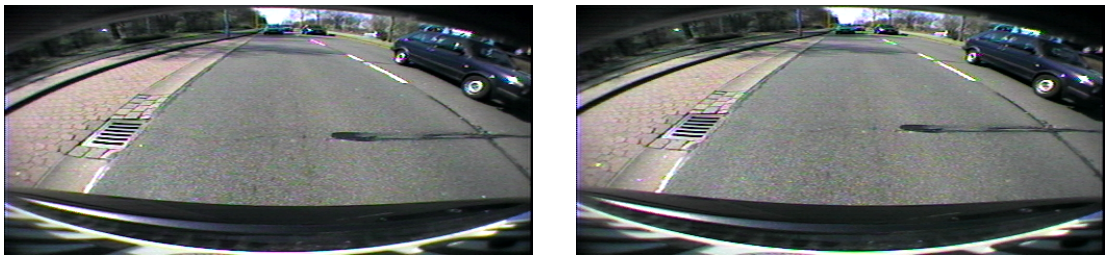


Figure 5.12 Original input images obtained by reversing camera. The camera was found to closely match a perspective camera model with radial distortion.

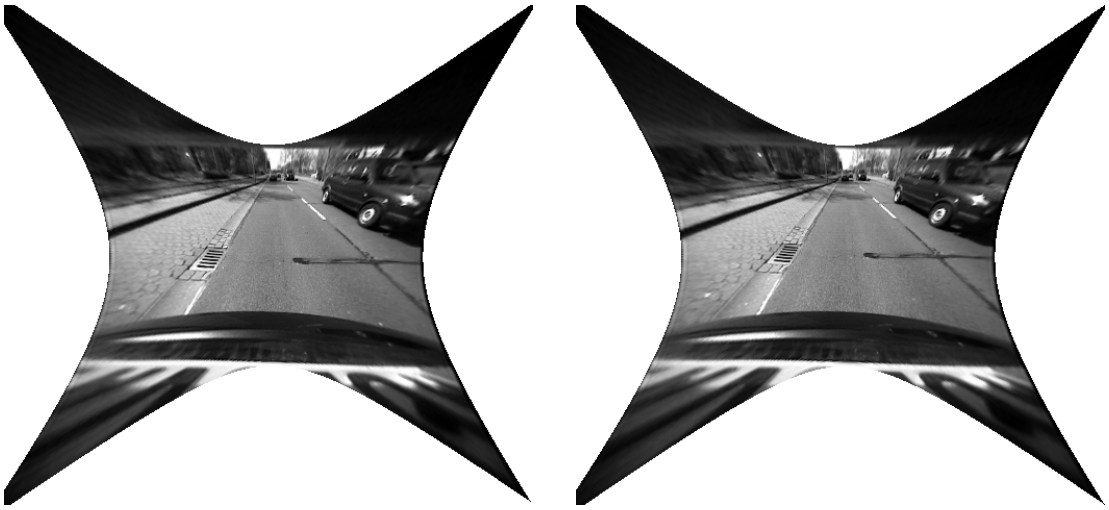


Figure 5.13 Images obtained by radial distortion removal from images in Figure 5.12. Note that large part of the view is occupied by insignificant data, e.g. license plate and hood of the car.

Pollefyes' polar rectification

As was noted previously the camera was found to closely follow the model of perspective camera with radial distortion, thus producing warped images. The radial distortion (2.6) was removed from the images, resulting pair of images is shown in Figure 5.13. As we may observe there are large areas in the image, which we do not wish to process, as they to not bear any significant information, therefore the images are further cropped to finally produce images which are suitable for rectification, Figure 5.14. The images were rectified and results are shown in Figure 5.15 along with disparity map computed by stereo matching. Again the disparities were transfered to the input reference image to produce 5.16.

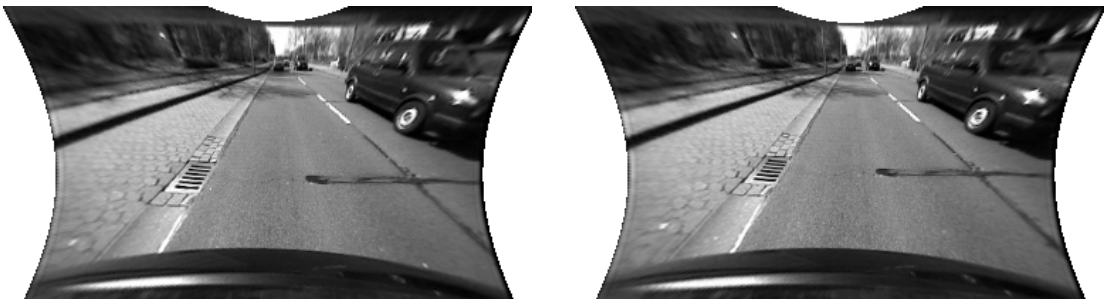


Figure 5.14 The images here were obtained by cropping images in Figure 5.13 with aim to remove areas of image that do not possess any significant information.

Geyer and Daniilidis rectification

Here we ought to present the results obtained using a chosen implementation of this conformal rectification. Unfortunatelly we were not able to match the calibration of reversing camera with camera model used in [8], although large amount of time was dedicated to resolve this problem. Therefore we have not obtained any results that would be comparable to outputs presented in previous sections.

We have presented the experiments and their results in this chapter. A ground truth measurement of disparity map accuracy was conducted and results obtained show that disparity maps from forward motion pair of images are comparable to disparity maps from a stereo pair of horizontal cameras both in density and accuracy. Employing spectral criterion we have provided a scheme reducing disparity map computation time, which is of interest with regards to potential real-time performance of considered applications. We have also shown that reversing camera may be used to obtain dense disparity content and commented on the set-back in experiments carried out using *OmniRect* toolbox [8].

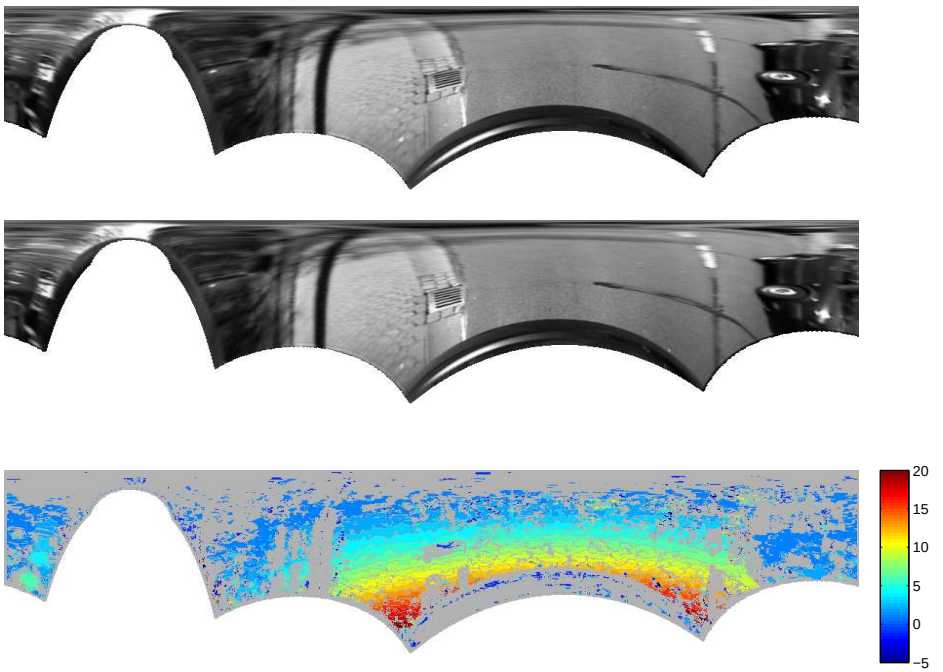


Figure 5.15 Images from Figure 5.14 were rectified and stereo matching was performed. The top two are the rectified images of reference and target image respectively and in the bottom a resulting disparity map is shown.

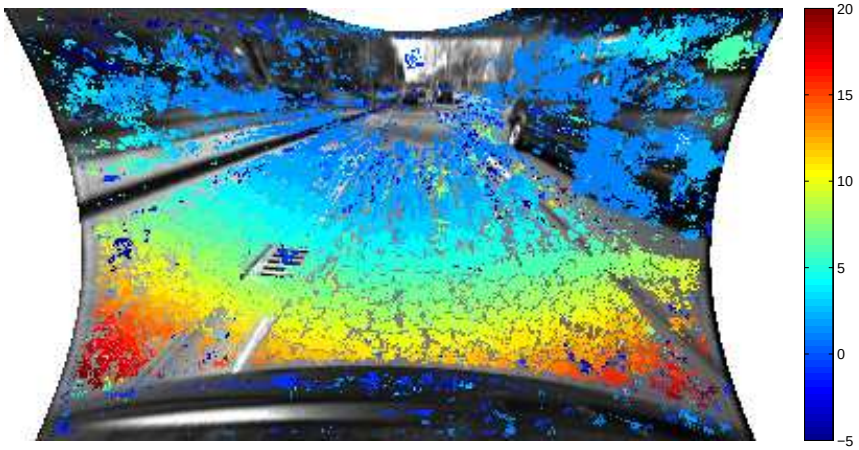


Figure 5.16 The disparities in disparity map Figure 5.15 were plotted to their proper position in the reference frame. Note that the disparities on the front part of the car to the right are all close to zero, although quite far away from the epipole. This is due to the fact that the car is moving in such way that relative motion of the car and the camera is close to zero.

6 Conclusion

In this thesis we studied the topic of epipolar rectification with regards to wide field of view cameras and dominant forward motion. This lead to a choice of rectification methods that were found most suitable for our application. The algorithm described by Pollefeys et al in [11] was implemented and throughly tested not only on the data provided by reversing camera but also on vast set of synthetic data, which were provided by data generator. An existing implementation of rectification method suitable for omnidirectional camera introduced by Geyer and Daniilidis in [6] was chosen to carry out the experiments but due to the fact, that calibration model of reversing camera could not be translated into a camera model used in *OmniRect* toolbox [8], it was not possible to obtain any results suitable for comparison with previously mentioned method.

The rectified images were used as input data for stereo matching algorithm which provided us with disparity maps. The accuracy of disparity maps obtained was measured by comparison with data from a pair of cameras aligned to produce rectified images. The results of this measurement showed that the two disparity maps are comparable both in density and accuracy. Relative density of 80% the density of disparity content from horizontal pair of images was a typical value. Disparity map accuracy measurement showed that around 80% of error realizations lie within $< -5, +5 >$ pixel range. We have seen that different information can be retrieved from horizontal pair of cameras and pair of forward cameras. While there are situations when stereo matching perfoms better in the horizontal pair, there are also situations when better results are obtained with forward pair. A topic to explore in the future work is the fusion of the two information channels. We could for example use them to identify moving objects, because these would stand out in the disparity error image.

The topic of optimized sampling was addressed and lead to a proposition of evaluating local spectral content of the image. Significant size reduction of resulting rectified images was accomplished. We have also demonstrated how the disparity maps are affected by the image size reduction. We used disparity maps obtained from images rectified by polar rectification as reference and proven that spectrum-conscious approach leads to a significantly higher reduction of rectified image size in relation to the density reduction of resulting disparity map. This results in lower time needed to compute a disparity content of specific density. One example for all: a disparity map density of 75% of reference density may be achieved in half the time needed for reference disparity map computation time.

We have seen that valuable content can be retrieved from a single moving camera and that it should be considered as a valid source of complementary information to that obtained from a pair of cameras.

...and the robot has moved.

Bibliography

[1] R. N. Bracewell, K.-Y. Chang, A. K. Jha, and Y. H. Wang. Affine theorem for two-dimensional Fourier transform. *Electronics Letters*, 29:304, February 1993.

[2] Jan Čech and Radim Šára. Efficient sampling of disparity space for fast and accurate matching. In *BenCOS 2007: CVPR Workshop Towards Benchmarking Automated Calibration, Orientation and Surface Reconstruction from Images*, page 8, Madison, USA, June 2007. IEEE, Omnipress.

[3] Andrew W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In Rangachar Kasturi and Gérard Medioni, editors, *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 125–132, Los Alamitos, California, USA, December 2001. IEEE Computer Society.

[4] Christopher Geyer and Kostas Daniilidis. Catadioptric projective geometry. *International Journal of Computer Vision*, 45(3):223–243, December 2001.

[5] Christopher Geyer and Kostas Daniilidis. Structure and motion from uncalibrated catadioptric views. In *International Conference on Computer Vision and Pattern Recognition*, 2001.

[6] Christopher Geyer and Kostas Daniilidis. Conformal rectification of omnidirectional stereo pairs. In *Proc. Omnivis 2003: Omnidirectional Vision and Camera Networks*, June 2003.

[7] Richard I. Hartley. Theory and practice of projective rectification. *Intl. J. on Computer Vision*, 35(2):115–127, November 1999.

[8] Jan Heller. Stereo reconstruction from wide-angle images. Master's thesis, Charles University in Prague, 2007.

[9] George Kamberov, Gerda Kamberova, Ondřej Chum, Štěpán Obdržálek, Daniel Martinec, Jana Kostková, Tomáš Pajdla, Jiří Matas, and Radim Šára. 3d geometry from uncalibrated images. In George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Paolo Remagnino, Ara V. Nefian, Meenakshisundaram Gopi, Valerio Pascucci, Jiri Zara, Jose Molineros, Holger Theisel, and Thomas Malzbender, editors, *ISVC 2006: Proc. 2nd Intl. Symposium on Visual Computing*, number 4292 in Lecture Notes in Computer Science, pages 802–813. Springer-Verlag, Berlin, Germany, November 2006.

[10] Martin Matoušek. *Epipolar Rectification Minimising Image Loss.* Phd thesis, Center for Machine Perception, K13133 FEE Czech Technical University, Prague, Czech Republic, October 2007.

[11] Marc Pollefeys, Reinhard Koch, and Luc van Gool. A simple and efficient rectification method for general motion. In John K. Tsotsos, Andrew Blake, Yuichi Ohta, and Steven W. Zucker, editors, *Proc. 7th IEEE Intl. Conf. on Computer Vision*,

volume 1, pages 496–501, Los Alamitos, California, USA, September 1999. IEEE Computer Society.

[12] Sébastien Roy, Jean Meunier, and Ingemar J. Cox. Cylindrical rectification to minimize epipolar distortion. In Ram Nevatia and Gérard Medioni, editors, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 393–399, Los Alamitos, California, USA, June 1997. IEEE Computer Society.

[13] Tobi Vaudrey, Clemens Rabe, Reinhard Klette, and James Milbrun. Differences between stereo and motion behaviour on synthetic and real-world stereo sequences. In *IVCNZ 2008: Proc. 23rd Intl. Conf. Image and Vision Computing New Zeland.* IEEE Xplore, November 2008.