SRI
Small Vision System

Calibration Supplement
to the User's Manual
Software version 4.4d
May 2007

©Kurt Konolige and David Beymer
SRI International
konolige@ai.sri.com
http://www.ai.sri.com/~konolige

1 Using the Calibration Procedure of SVS

The SVS Calibration Procedure for stereo cameras is an efficient, simple method for finding the internal (lens distortion and decentering) and external (camera spatial offset) parameters of a two-camera stereo rig. The output is a set of parameters that can be used with the SVS to warp the stereo images from the cameras into images that would be produced by ideal pinhole cameras whose imagers are in the same plane.

The calibration procedure is briefly described in the Calibration section of the SVS User's Manual. This document gives an expanded treatment of the calibration procedure, as well as helpful hints to those experiencing difficulties with the calibration procedure. It also gives technical details of the calibration and its results.

In an ideal setting, the cameras of a stereo rig are perfect pinhole cameras, with exactly the same focal length, and parallel optical axes. In practice, cameras are imperfect, suffering from lens distortion and differing focal lengths, and misaligned optical axes. The goal of calibration is to determine two sets of parameters, *intrinsic* and *extrinsic*, that compensate for the imperfections in the stereo rig. The intrinsic parameters correct lens distortion and uneven focal length; the extrinsic parameters determine the spatial offset of the two cameras, including the stereo baseline and any deviation from parallel optical axes. These parameters can then be used to warp the camera images into *standard position*, that is, images that would be seen by pinhole cameras with parallel optical axes.

The calibration procedure in SVS uses a simple planar target that can be printed out on a standard PC printer. When shown the target in 5 or more different views, the calibration procedure can determine the calibration parameters using an efficient nonlinear optimization procedure.

1.1 Getting Started

The calibration procedure is part of the *smallvcal* application. The necessary calibration routines are compiled into the application. To run the calibration procedure, select the menu item *Calibration* at the top of the *smallvcal* window. This will bring up a calibration dialog window, as shown in

Figure 1-1 below.

> NOTE: The smallvcal program is a standalone application; it cannot be compiled into other applications. The sources for this program are not part of the SVS system.

Calibration requires at least 5 different image pairs of the checkerboard calibration object, which can either be input live, or from stored files. To familiarize yourself with the procedure, it is helpful to run through the procedure using stored images available in the *data/calN-X.bmp* files. The following steps show a typical calibration sequence.

1. Click on the *Load All* button at the top right of the dialog to bring up a load dialog, and choose the *data/cal1-L.bmp* image. The stereo images will be loaded into the windows, starting with *cal1-L/R.bmp*.
2. Set the parameters. The data sequence was taken using a MEGA-D camera, and the panel square size was 108 mm. All of the parameters are correct (the defaults for the MEGA-D), except for the panel square size. Click in the number output area, and enter "108".
3. Choose the *Features* button on the **bottom** of the windows to find the corners of the checkerboard

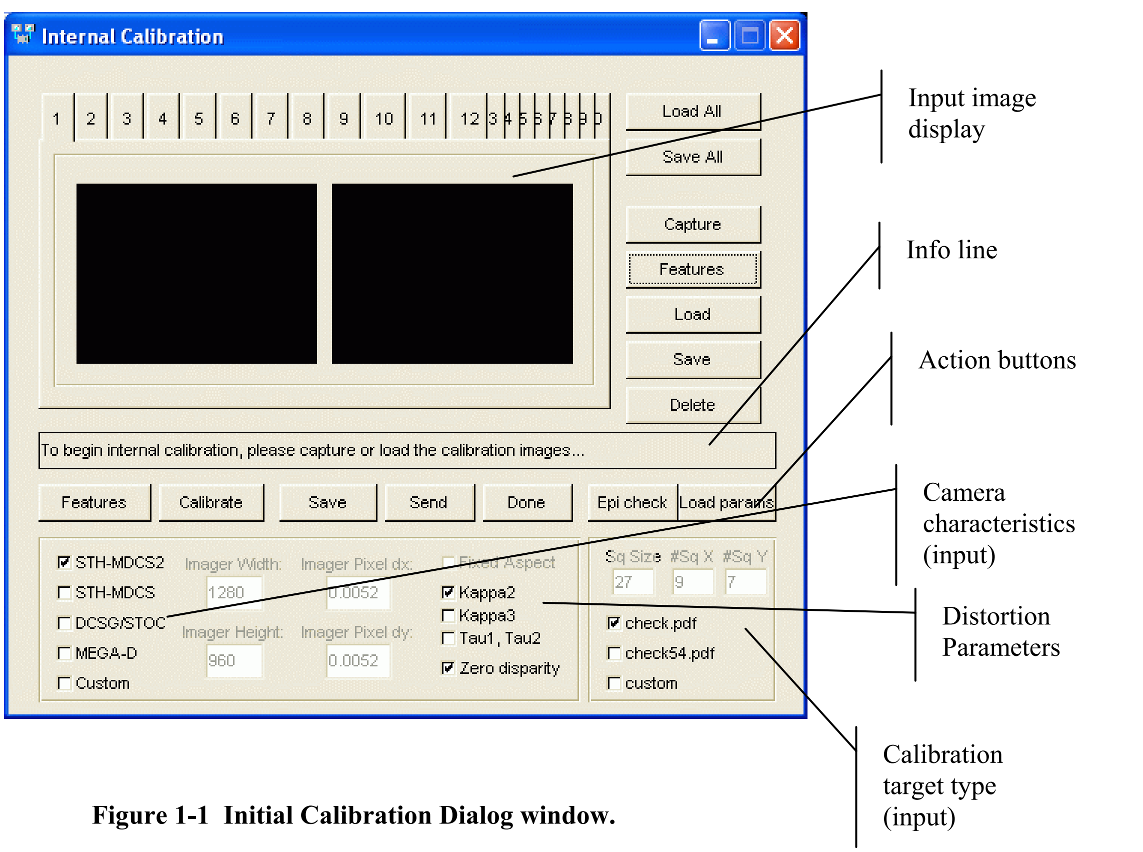

Figure 1-1 Initial Calibration Dialog window.

squares for all images (the one to the right of the image display finds features for just the displayed image pair). After 20 seconds or so, the images will show green crosses where the feature-finder detected the corners.

4. Once all the features have been detected, choose *Calibrate* from the action buttons. At this point, the calibration routines will attempt to find a good set of parameters for the images you have input, using a nonlinear optimization technique. The debug output box will show the progress of the routines. First, all the left images are calibrated together; then, all the right images; and finally, the two are put together to determine the extrinsic parameters. If the calibration is successful, then the final estimated parameters are displayed in the *Debug* window (*Left intrinsics, Right intrinsics,* and *Extrinsics*).
5. You can save the parameters to a file, which is recommended, and send the parameters to *smallvcal* application with the *Send* button. The calibration dialog will remain open so that you can perform other calibrations if you choose. To close the calibration dialog and send the parameters to *smallvcal* , use the *Done* button. To exit the dialog without sending parameters, just close the window.
6. To test the calibration, input the *data/wall* images using the *File* menu. The images will be rectified according to the calibration parameters, and stereo processing will work correctly.

Because the calibration procedure is a nonlinear optimization problem, on rare occasions it can get stuck and not find a good calibration. When this occurs, one or more of the input images will be rejected as inconsistent with the ongoing optimization. It helps to bring up the *Debug* window to watch the progress of the calculation, and assure yourself that something is happening; the progress is also indicated in the info line in the calibration dialog. Note that the calibration dialog and *smallvcal* windows are unresponsive during the calculation, except for a brief moment when the next image pair is added.

In some cases, even when the procedure terminates successfully, the parameters that are output will not be good ones. In most of the problematic cases, a good calibration can be found by simply replacing some of the image pairs.

The purpose of this document is to give tips for the efficient use of the calibration procedure.

1.2 Checking the Current Calibration

The calibration system has a feature, called *Epi Check*, that will check the current calibration of the device. The Epi Check will help to determine if you need to re-calibrate the device, by computing statistics for the epipolar alignment of the cameras. The epipolar alignment means that a scene point should project onto the same horizontal line in both cameras.

To run the Epi Check, you must have a target checkerboard. See Section 1.4.2 for information about making a target. Try to get the largest target you can (which will also be useful for calibration), although the size is not critical.

- Capture several images of the target, making sure to cover all the different areas of the image, especially the corners (see Section 1.4.3). The captured images should be *rectified*, since the rectification correctness is being check. Make sure the device calibration parameters are loaded into the main `smallvcal` window, and the `Warp` button is checked.
- After the images have been captured, click on the `Features` button under the info line, on the left. This will find features in all the images. If there are errors finding features in some images, capture new images in their place, and find features again.
- All the camera and target characteristics can be ignored, they are not used in the Epi Check.
- Press the `Epi Check` button. The epipolar alignment will be computed and printed on the Debug window:

```
[EpiCheck] Running epipolar check
[EpiCheck] Images should be rectified

[EpiCheck] 48 points in image 1
[EpiCheck] 48 points in image 2
[EpiCheck] 48 points in image 3
[EpiCheck] 48 points in image 4
[EpiCheck] 48 points in image 5

*** RMS Error from epipolar line is 0.14 pixels
*** Average Bias from epipolar line is -0.03 pixels
*** STD of epipolar error is 0.14 pixels
```

Here the epipolar check used 5 images.

The Bias is a measure of systematic bias between the images. For example, if the right image were shifted lower relative to the left image, there would be a high bias. The bias should be very small, typically less than 0.05 pixels. This indicates that most of the epipolar error is caused by uncertainty in the feature localization, and the calibration is very good.

The RMS error is the average deviation of the features from the ideal epipolar placement. In this case, the RMS error is 0.14 pixels, which is good – typical errors for a good calibration are 0.10 to 0.15 pixels. The uncertainty in localizing features means that the RMS error seldom gets under 0.10 pixels.

The STD is the standard deviation of the epipolar error, and is usually a measure of the uncertainty of feature localization (RMS error minus the bias).

1.3 Setting the Camera Characteristics

The calibration procedure must have knowledge of the cameras that produced the input images. This is true whether you use stored or live images. Each image pair must come from the *same* stereo rig: you can't mix image sets taken from different stereo rigs. Also, each camera in the stereo rig must have approximately the same characteristics, i.e., the same lens focal length and imager type.

There are preset values for the stereo heads sold by Videre Design: the STH-V2 and STH-V3 analog heads, and the MEGA-D and MDCS megapixel digital heads. Choosing one of these stereo head types, in the camera characteristics section of the dialog, sets most of the required values. The only choice remaining is whether to use lens distortion parameters *Kappa2*, *Kappa3*, *Tau1*, and *Tau2*, and whether to have zero disparity at infinity. These choices are discussed below.

If you want to calibrate a stereo rig other than the Videre Design stereo heads, then you must choose the *Custom* setting and input the camera parameters. Usually these parameters can be found from the specifications of the cameras.

> **Note:** the camera characteristics are not used by the calibration procedure. Their only purpose is to transform the focal length shown to the user from pixels to millimeters. The camera characteristics can be left in their default setting, if it is not necessary to know the metric focal length.

The *Kappa* and *Tau* values are parameters to correct for lens distortion. Allowing the calibration routines to search for a good value for *Kappa2* and *Kappa3* can help with lenses that have a large amount of radial distortion ("barrel" distortion), which is most noticeable on wide-angle lenses (> 50 degree field of view). The *Tau* parameters can correct for tangential distortion. Most modern lenses have negligible tangential distortion, and the *Tau* parameters should almost never be used.

> **Note:** Unless you have a verged (pointed-in) set of cameras, please leave the *Zero disparity* box in its checked condition.

The distortion parameters are used to rectify the input images. If your stereo rig has cameras that are approximately parallel (the normal case), then the disparity between left and right images for objects at infinity will be approximately zero. The rectification step will convert the input images into ideal parallel camera images, and then the disparity at infinity will be exactly zero.

In cases where the cameras are not parallel, for example, in some machine vision settings where the objects are very small and close, the zero disparity will occur at a finite distance from the cameras. In this case, setting the "zero disparity at infinity" parameter will cause the calibration to move the right image out of the field of view of the camera, because the cameras do not have any common view of objects at infinity, or just a small overlap. Unchecking the *Zero disparity* box will let the calibration move the rectified images so that they cover as much as possible of the original input images. Note that all of the stereo functions will work just as they would otherwise, including the correct calculation of the 3D XYZ coordinates.

1.4 Input Images

The quality of the calibration depends mostly on the input images that are given to the calibration procedure. Here we give some general tips for the characteristics of good input images.

1.4.1 Image size

A large image size helps the feature finder to accurately locate the corners of the target checkerboard, and leads to a more accurate calibration. However, larger sizes also make feature-finding more difficult, because there is more image area to search. A good compromise is to use 320x240 images, if they are available from the framegrabber or digital stereo rig. Table 1-1 gives the recommended input image sizes for Videre Design stereo heads.

> **Note:** The results of a calibration are independent of the input image size. Thus, it is possible to use a calibration done with images of size 320x240, on input images of size 640x480. The only requirement is that the calibration images cover the same portion of the imager as the images being rectified. Usually, the area is the full field of the imager. When using *subwindows* of the imager (e.g., with the MEGA-D), then the calibration must cover the same subwindow as the desired input images.

Some cameras and framegrabbers let you input a *subwindow* of the full image produced by the camera. If you are going to work with just this subwindow, then you should use the subwindow for your input images. In most cases, however, you will work with the full image produced by the cameras, and the input images for calibration should encompass the full image size. For example, with the MEGA-D megapizel digital imager, the full image size is 1280 x 960 pixels. This is rather large for the feature finder, so for inputting images, it is best to put the MEGA-D into a mode where the output image is only 320 x 240, but encompasses the full space of the original 1280 x 960 image. The MEGA-D has a *subsampling* mode that effectively reduces the output image resolution, while maintaining the same spatial extent of the image. The preferred Subsampling mode is to decimate the image by a factor of 2, and to further average (bin) the image by a factor of 2. This gives a total reduction of x4, producing a full-field 320x240 image.

Analog framegrabbers also allow image reduction of the output image by scaling the resolution of the image. For the STH-V2 and V3 stereo heads, in line interlace mode, it is best to ask for output images of size 160x120. These images are scaled by the framegrabber to encompass the full original field of the imager.

1.4.2 Planar Target

The calibration procedure requires a set of at least 5 image pairs of a planar checkerboard target. The standard target is in *data/check.pdf*. Print out the target on a standard 8.5 x 11 inch sheet of paper (or A4 paper), making sure to specify 1:1 scaling in the printer options (do *not* specify "fit to page," which will stretch or shrink the image). Paste the printed pattern to a flat object, such as a cutting board or a clipboard. Figure 1-2 shows an image of the checkerboard.

Most stereo rigs can be calibrated with the standard target, which is convenient because it fits on a

Stereo Head	Recommended image size
STH-V2, STH-V3	160 x 120
MEGA-D	640x480, dec x2
STH-DCAM (-VAR)	320 x 240, bin x2
STH-MDCS(-VAR)	640x480
STH-DCSG(-VAR), STOC	640x480

Table 1-1 Recommended input image sizes for Videre Design stereo heads.

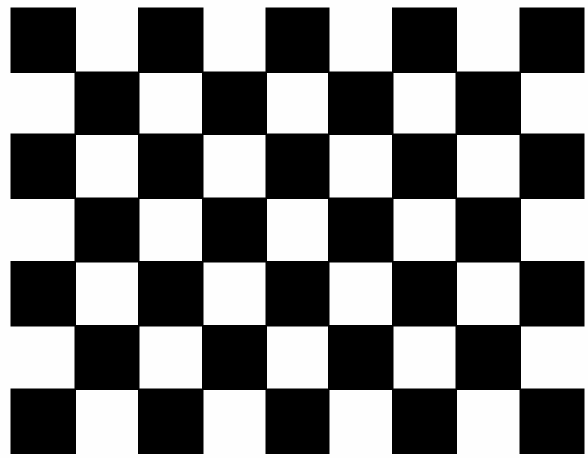

Figure 1-2 Checkerboard calibration object.

single sheet of paper. In some cases, the standard target is not sufficient: it is impossible to view the target in both images at a large enough size. This situation typically occurs with long focal length lenses and wide baselines, where the target must be placed far from the camera in order to appear in both images. For those stereo heads sold by Videre Design, only the wide baseline MEGA-D should require this target, or the variable baseline stereo devices with a wide separation.

In these situations, the file *data/check4.pdf* should be used. This file contains a set of 4 images, which comprise a target with a larger square size (54 mm vs. 27 mm for the standard target). Print out the four sheets, and paste them together on a flat board; some overlap is required. Be sure to check the *check4.pdf* button in the panel square size region. If you have access to a large-format printer, you can also print out a single sheet containing the 54 mm pattern (*data/check-54.pdf*) or even a 108 mm pattern (*data/check-108.pdf*).

For some custom stereo rigs, even the larger target may not be sufficient. The calibration dialog allows you to use any size target you care to make; just input the square size as a *custom* target size.

1.4.3 Planar Target Placement

The placement of the planar target in a calibration image set is a critical factor in getting good calibration results. The planar target itself is in *data/check.pdf*. Here are some hints on how to capture a good set of images of the target. You can also check the calibration images in *data/calN-X.bmp* to see an example of a good image set.

1. Try to have the target fill up a large portion of the image in each camera. This will give corner features that are spaced far apart, which is best for the feature finder and the calibration routines.
2. Have at least one image pair where the target is almost parallel to the cameras, and fills up a good portion of the image. This will help the calibration routines find the correct image distortion parameters.
3. For most of the images, show the target slanted away from or towards the cameras, so that the squares appear at different depths. Vary the slant in the image pairs, so some slant vertically, others horizontally.
4. In addition to (3), have at least one image at a very different depth from some of the other images. Note that this can conflict with (1), since you may have to put the image farther away from the cameras.
5. For several of the images, have the target come close to one of the edges of the images.

Most problems with convergence and calibration arise because of the input images. Typically, a good set of images will cover the image field, and adhere to the guidelines above. It is especially important not to place the target too far or too close. The images in Figure 1-3 show a good set of target placements for calibration; these can be used as guidelines for how to capture a calibration set.

The calibration routines will take up to 10 images. Usually 5 is adequate; but if there is some problem with getting a good calibration, then inputting more images may help, especially for very wide-angle lenses with high distortion.

Figure 1-3 Typical calibration image set (left images only). Note the good coverage of the image field, and the different slants.

1.4.4 Saving Calibration Images

It's a good idea to save image pairs that are captured from a live video source. In this way, you can re-input the images if the program crashes, or if for some reason there is a problem producing a good set of calibration parameters. The stored images can also be sent to SRI International for debugging. Please use the *Save All* button for this purpose, since it makes it easier to input them.

1.5 Finding Features

The calibration procedure first must find the image *features* in each of the image pairs. These features are the corners where 2 black and 2 white squares just touch. You can invoke the feature-finder for each image pair as it is acquired at the top of the dialog, or find features in all images by using the button at the bottom of the dialog. When the feature-finder is finished, it will print an information message about the results. Normally, the feature-finder is very robust, and it will succeed; in this case, the found features are displayed as red crosses in the images. You can verify that the features are correct by looking at the placement of the crosses: if any of them are obviously incorrect, you should delete the image pair and get a new one (Figure 1-4).

There are four conditions under which the feature-finder may fail:

1. The image lighting is poor. If the checkerboard image is washed-out or obscured in some way by lighting, the feature-finder won't be able to work correctly.
2. The checkerboard image is too small. The target must be a reasonable size in the image, so that the corners will be detectable.
3. The checkerboard image is too large. The feature-finder will have trouble if the full checkerboard is very close and fills the image.
4. Some parts of the checkerboard are not visible. The feature-finder must see all of the checkerboard junctions. Note that the squares on the edges of the board can be partly visible – the feature-finder cares about the junctions, not the squares themselves.

Again, if these conditions cause the feature-finder to fail, just delete the image pair and get a new one.

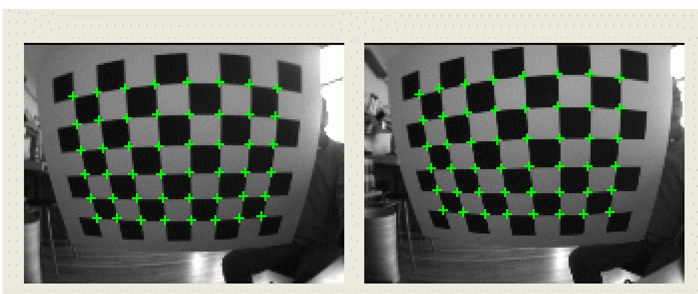

Figure 1-4 Features found in a typical calibration image.

1.6 Image Scaling

The calibration parameters tell how to warp an input image to a rectified output image. The input and output images both have the same size, for example, a 320x240 input image is warped to a 320x240 output image. In the warping process, the output image can be *scaled* to contain just a portion of the input image. Generally, there are two extremes for the scaling:

1. Every output pixel is mapped to some input pixel (i.e., there are no blank areas in the output image, put some portion of the input image may not appear).
2. Every input pixel is mapped to some output pixel (i.e., the full scene appears in the output image, but some pixels in the output image may be blank).

Figure 1-5 shows the difference between the two. On the left, the original input image contains a large amount of lens distortion. The middle image is rectified, and slightly enlarged so that it comes from the middle portion of the input image. Note that the doorway on the right is not fully present in the middle image. Every pixel in the output image comes from a pixel in the input image.

In contrast, the rectified image on the right is slightly reduced from the original image, so that every original image pixel is mapped to the output image. Note the blank areas on the sides of the image. The dotted rectangle shows how the middle image fits – it is a smaller rectified image, expanded to fill the 320x240 rectangle.

There is no option for scaling in the calibration dialog window, since scaling is performed after the calibration is produced, when rectifying images. A default scale factor of 1.0 is output when a calibration is saved. This scale factor ensures condition (1) above. Smaller scale factors (e.g., 0.9) will reduce compress the rectified image.

If the lens distortion is not extreme, the default output image is generally preferred, since there are no blank areas to confuse the stereo processing. But if the lens distortion is high, the reduced output image may be better, since no information is lost.

The scale factor can be changed by modifying the *frame* parameter of the saved *.ini* calibration file.

Figure 1-5 Effect of scaling during the rectification process. All images are 320x240. Left image is the orginal. Middle image is rectified, and enlarged slightly so that all output pixels are mapped. Right image is reduced so that all input pixels are mapped.

1.7 Calibration Calculation

After finding features in at least 5 images, the calibration procedure can be invoked by pressing the *Calibrate* button. The progress of the calibration calculation can be observed in the Debug window. If there is a problem with the procedure, it can be diagnosed here.

Figure 1-6 shows typical output in the Debug window, from a successful calibration. Five images are input, of size 640x480. The panel square size is 108 mm, which is four times the size of the standard panel – the *custom* box in the calibration window is used to input the square size. For a very accurate calibration, and for long focal length lenses, it helps to use a larger panel.

48 points are found in each panel. Then, the procedure calibrates the left images together, and prints out reprojection error statistics (the amount by which the calculated 3D position of the panel projects onto the image, and differs from the found features). Typically this error will be on the order of 5 to 10 hundreths of a pixel.

Next, the right images are calibrated, and again error statistics are printed. Finally, the left and right images are calibrated together, and a final calibration is produced and displayed in the calibration window.

Figure 1-7 and Figure 1-8 show the result of the calibration procedure applied to the calibration images in the *data* directory. For the standard Videre Design devices (STH-MDCS, MEGA-D), a good calibration will produce small values for *Ty*, *Tz*, *Rx*, *Ry*, and *Rz*. The distance *Tx* is the baseline of the stereo device.

In the case of cameras that exhibit *vergence*, that is, they are rotated towards each other a bit, the *Rx* parameter should be noticeably higher. The calibration procedure can correctly register stereo rigs with a high degree of vergence.

At the end of the calibration, an epipolar alignment check is performed to verify the calibration. See Section 1.2 for an explanation of the Epi Check results.

A detailed explanation of the parameters is in Section 2.

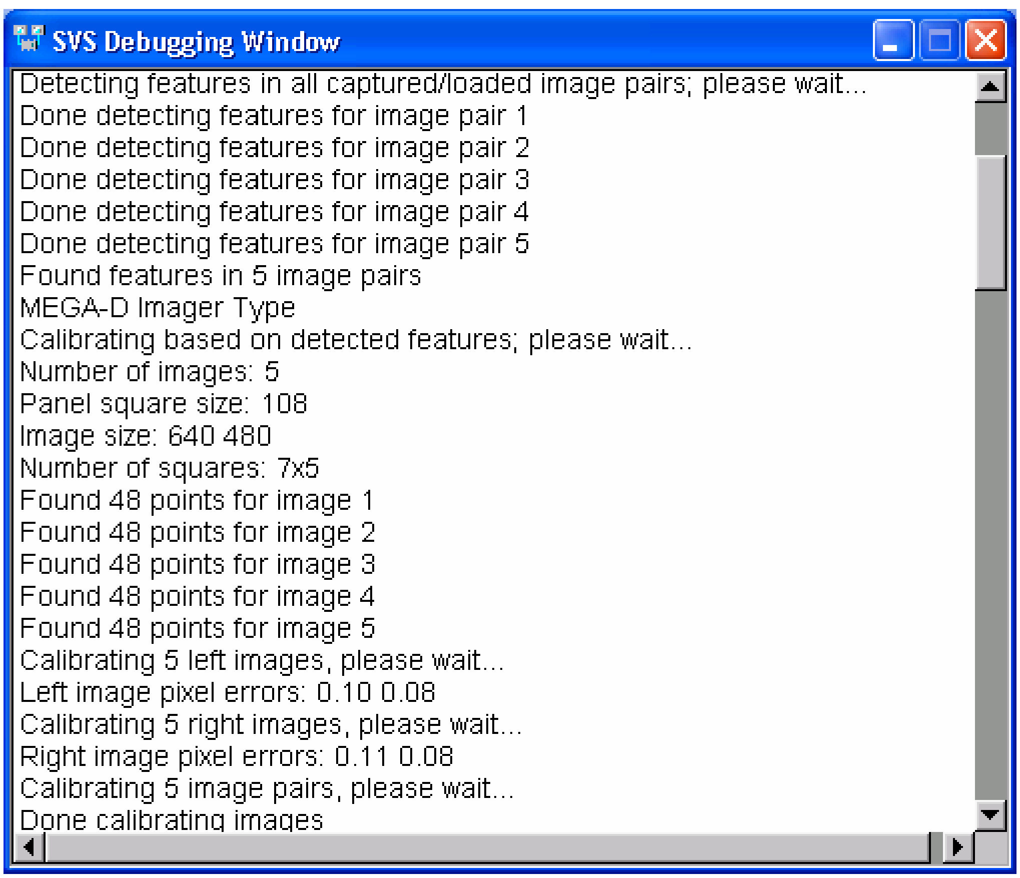

Figure 1-6 Debugging output during the calibration process.

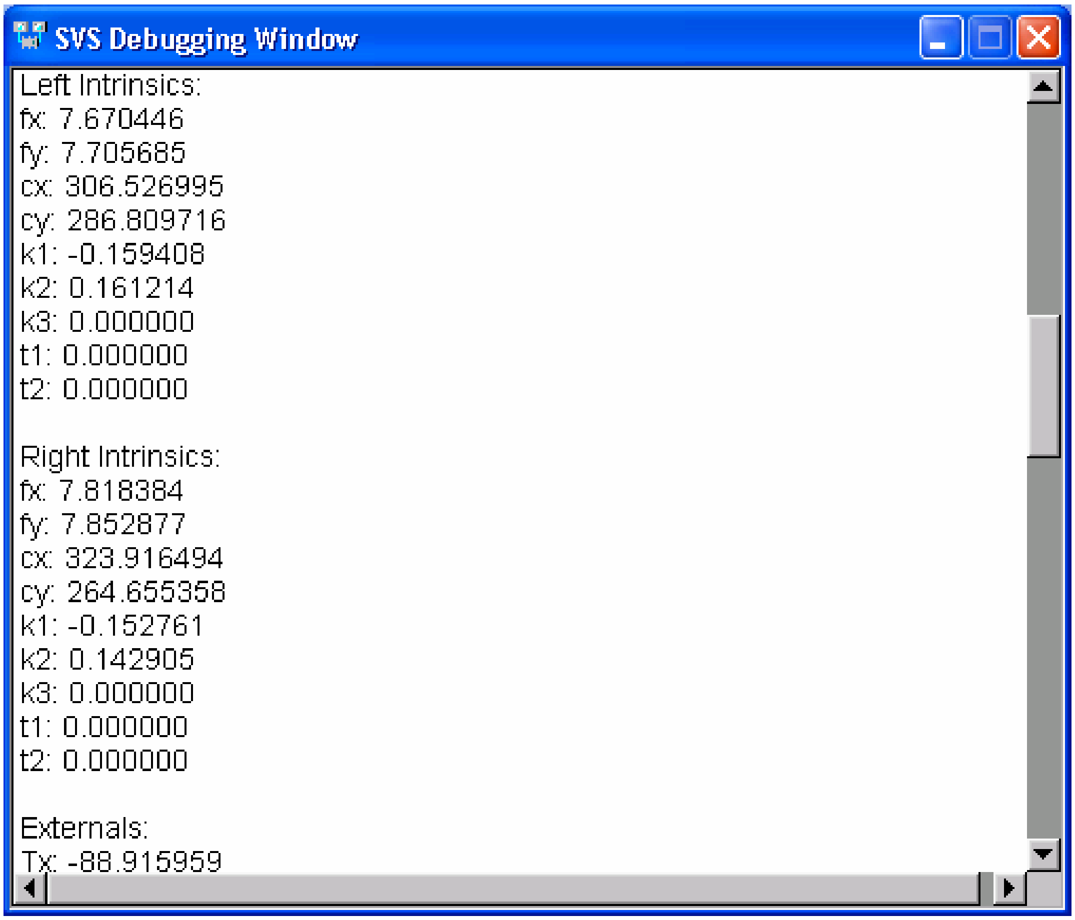

Figure 1-7 Results of the calibration procedure on the SVS calibration dataset – internal parameters. *cx* and *cy* are near the center of the image. Only the first two radial distortion parameters were used. Note the small values for Ty, Tz, Rx, Ry, and Rz.

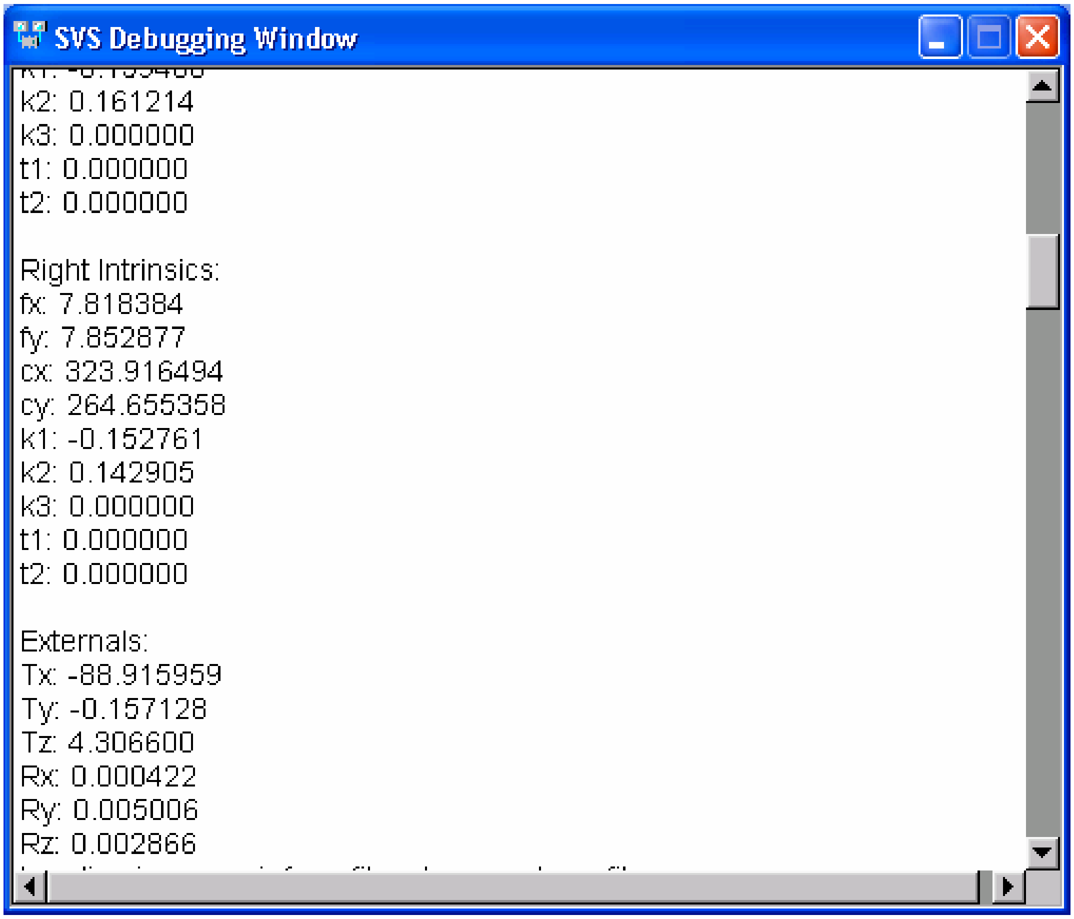

Figure 1-8 Results of the calibration procedure on the SVS calibration dataset – external parameters. Note the small values for Ty, Tz, Rx, Ry, and Rz.

1.7.1 Calibration Parameters

The parameters returned by a successful calibration are guaranteed to be correct, in the sense that they will transform the 5 or more input images into images as seen by pinhole cameras that have parallel optical axes. In almost all cases, they will give physically plausible values for the intrinsic and extrinsic parameters. For example, the cameras used in the images of Figure 1-7 were rated at 7.5 mm, and they were placed approximately 9 cm apart, with parallel optical axes. So, we expect the returned focal lengths to be about 7.5 mm, and they are about 7.7 mm (left) and 7.8 mm (right), which is close.

The remaining intrinsic parameters are concerned with lens distortion. *Lens decentering* is a deviation of the lens optical axis from the center of the imager. The estimated lens center is returned as the parameters *Cx,Cy*. These parameters should be close to the image array center, which in this case is (640,480). Finally, the lens *radial distortion* is the deviation of the lens convergence from a true pinhole projection. This distortion is modeled by five parameters. *Kappa1*, *Kappa2*, and *Kappa3* (not shown) determine the radial distortion -- how much the image stretches (positive) or shrinks (negative) away from the lens center. Most lenses have a small amount of negative radial distortion; lenses with very short focal lengths tend to have larger amounts. Lenses with smaller focal lengths will have larger Kappa values. **Error! Reference source not found.** shows the calibration parameters for a very small focal length lens (2.1 mm). You can see the extreme radial distortion in the images at the top of the calibration dialog.

The extrinsic parameters show how the cameras are situated relative to each other. All these numbers should be small, except for the value of *Tx*. This is the baseline between the cameras, and it should be a negative number of approximately the right size (in this case, 90 mm is the baseline estimate from the stereo rig design). All of the other extrinsic parameters in Figure 1-7 are small; for example, the value of *Tz* is just –0.16 mm, which indicates that the right imager is estimated to be –0.16 mm above the right imager. The parameter *Tz* is higher, about –4 mm, which indicates the right imager is this distance behind the plane of the left imager. Slightly higher values of *Tz* are typical, since this parameter is very sensitive to slight mechanical variations in imager position.

One good way to check the parameters is to apply them to the calibration images themselves. Transfer the parameters to the *smallvcal* main window by pressing the OK button (but first save the parameters!). Then, in *smallvcal*, load one of the calibration image pairs, and press the Continuous button to see the effects of the warping. Check that the straight lines of the calibration target all appear straight in the rectified images. Also, pick a particular square, and check that it appears at the same height in each of the images.

1.8 Troubleshooting

There are two problems that can occur during the calibration calculation.

1. The calibration procedure does not converge.
2. The calibration procedure converges, but converges to values that are physically incorrect.
3. The calibration procedure converges and the values are physically correct, but the rectification results are not good, especially around the edges of the images.

1.8.1 No Convergence

In rare cases, the calibration procedure will find an image or set of images is incompatible with the rest of the images. In this case, the system will print warnings on the Debug window about the offending images, and the procedure will not converge. Input new images to replace the incompatible ones.

Non-convergence will generally only happen with very highly distorted lense.

1.8.2 Bad Calibration Parameters

The calibration procedure may converge to parameter values which are not physically correct, although they are mathematically correct. The problem is that the input images do not give enough information to determine a unique solution, and the procedure converged on the wrong one. Generally, this will happen only if there are not enough images with depth variation (the planar target slanted towards or away from the camera). In this case, there are two possible remedies. The first is to use only one *Kappa* parameter in the camera characteristics field of the calibration window. Allowing a second and third radial distortion parameter increases the chances for ambiguity, and restricting it can often lead to a good solution. The second remedy is to simply capture another set of calibration images, paying attention to the guidelines in Section 1.4.3.

The focal length parameter is not a concern. Internally, the calibration procedure represents the focal length in terms of *pixels*. For display purposes, the focal length is translated to mm, based on the provided camera pixel sizes. If the pixel sizes are wrong, the displayed focal length will be incorrect. However, an incorrectly displayed focal length will not affect the action of the calibration parameters in rectifying images.

1.8.3 Poor Rectification

After transferring the calibration parameters to *smallv*, you can check them by rectifying some images. Generally, the calibration images themselves will rectify correctly, because they were used in the rectification process. If the target in the calibration images did not cover the image field well, there may be problems with rectification in the areas not covered by the target. The only remedy here is to re-take some of the calibration images, getting more coverage of the image field.

2 Technical Aspects of Calibration

The full calibration procedure uses a number of views of a calibration object to estimate a model for the stereo head cameras. It utilizes a fairly complete model with up to 24 parameters. We now describe the camera model and its use in calibration.

2.1 Stereo Geometry

The relationship of the two rectified images is shown in Figure 2-1. The reference image is the left image, and the camera coordinate system is centered on its focal point. The baseline between the two cameras is in the positive X direction.

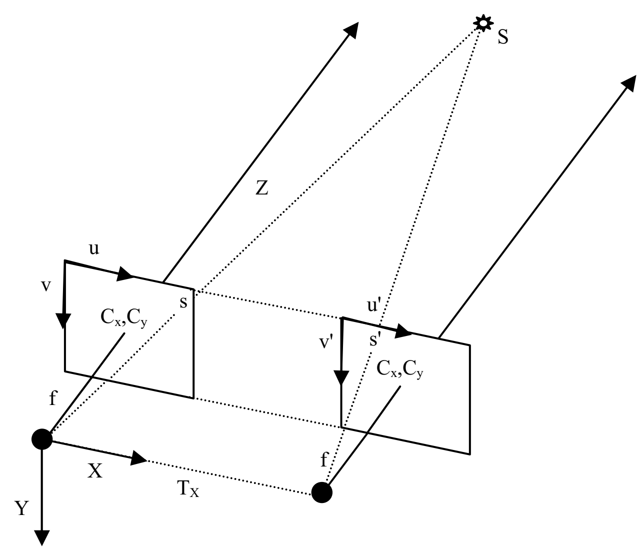

Figure 2-1 Basic stereo geometry. This figure shows the relationship of two *ideal* stereo cameras. The global coordinate system is centered on the focal point (camera center) of the left camera. It is a right-handed system, with positive Z in front of the camera, and positive X to the right. The camera principal ray pierces the image plane at C_x, C_y, which is the same in both cameras. The focal length is also the same. The images are lined up, with $v=v'$ for the coordinates of any scene point projected into the images. The distance between the focal points is aligned with the X axis.

2.2 Calibration File Parameters

This section discusses the parameters found in the calibration (*.ini*) files used by SVS. Technical details of the rectification process can be found in following sections.

With version 2.4 of SVS, the focal length parameters have changed. There are now two focal length parameters for each camera, the horizontal and vertical focal lengths, with units of pixels. In version 2.3 and below, there was a single focal length with units of mm, plus an aspect ratio parameter. Since the pixel size is not used in rectification and stereo calculations, it is included only as a point of information, so that the focal lengths can be translated to mm, if desired. The easiest way to distinguish a 2.3 from 2.4 calibration file is to check the *fy* parameter. If it is absent or 0, then it is a 2.3 calibration; otherwise it is a 2.4 calibration.

The calibration parameters characterize the transformation between *input images* and *rectified* or *ideal images*. During stereo processing, input images from a stereo camera device are first transformed into rectified images, and then stereo processing proceeds. Rectified stereo images have the property that every point in one image has its corresponding point on the same horizontal line in the other image.

The calibration process determines both the parameters of the actual cameras that are inputting images, and the ideal images that are used in the stereo process. Both these sets of parameters are included in the parameter file, as well as the *rectification matrix*, which translates between them, and the *frame* parameter, which introduces a scale factor. Information on how these parameters are used to transform the input image into the rectified image can be found in Sections 2.3 and 2.4.

2.2.1 Projection Matrix

The projection matrix transforms 3D coordinates into image coordinates. The 3D coordinates are in the frame of the left camera (see Section 2.4). There is a projection matrix for the left camera, and one for the right camera. The form of the 3x4 projection matrix P is shown in

$$\begin{bmatrix} F_x & 0 & C_x & -F_xT_x \\ 0 & F_y & C_y & 0 \\ 0 & 0 & 1 & 0 \\ & & & \end{bmatrix}$$

Table 2-1. A point in 3D $(X,Y,Z)^T$ is represented by homogeneous coordinates $(X,Y,Z,1)^T$ and the projection is performed using a matrix multiply

$$\begin{bmatrix} F_x & 0 & C_x & -F_xT_x \\ 0 & F_y & C_y & 0 \\ 0 & 0 & 1 & 0 \\ & & & \end{bmatrix}$$

Table 2-1 Projection matrix for a single camera. *Fx, Fy* is the focal length of the rectified image (pixels), and *Cx,Cy* is the optical center (pixels). *Tx* is the translation of the camera relative to the left (reference) camera. For the left camera, it is 0; for the right camra, it is the baseline in mm. Note that the focal lengths are for the *rectified* images, and thus will be the same; these are *not* the focal lengths *f,fy* given explicitly in the parameter file, which are for the original images.

$$\mathbf{P}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \cong \begin{bmatrix} x \\ y \\ w \end{bmatrix}$$

where $(x/w, y/w)$ are the screen coordinates. Note that this equation holds for *rectified* images, that is, the coordinates *(x,y)* are in the rectified image.

2.2.2 Reprojection

A point *(x,y)* in the left camera can be re-projected to 3D coordinates, if its disparity is known. The SVS functions Calc3D and CalcPoint3D in the class svsStereoProcess are provided to perform this calculation (see the SVS User Manual). Here we give the equations that govern the transformation.

The reprojection transformation is influenced by the *frame* and scaling factors in the calibration (e.g., if the calibration was performed at one resolution, and the input images are at a different resolution). It also depends on any horopter offset. For the following calculation, we assume that the *frame* and scaling are unity, and the horopter offset is zero.

Define the *reprojection matrix* as follows:

$$\mathbf{Q} \equiv \begin{bmatrix} 1 & 0 & 0 & -C_x \\ 0 & 1 & 0 & -C_y \\ 0 & 0 & 0 & f \\ 0 & 0 & \frac{-1}{T_x} & \frac{C_x - C'_x}{T_x} \end{bmatrix}$$

where C'_x is from the right rectified image, and the other internal parameters are from the left rectified image. Normally, the calibration produced by SVS will set C'_x equal to C_x, so the last term is 0. Under this condition, the disparity at infinity will be 0. For verged cameras (pointing inwards rather than parallel), it may be useful to have C'_x different from C_x, in order to get the right rectified image to be less offset. In this case, the disparity at infinity will not be zero, it will be negative. The calculation of X,Y,Z coordinates will still be correct, though, using the equation below.

From an image point homogeneous coordinates $(x, y, d, 1)^T$, with *d* the disparity, the corresponding 3D point in homogenous coordinates is calculated as:

$$\mathbf{Q}\begin{bmatrix} x \\ y \\ d \\ 1 \end{bmatrix} \cong \begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix}$$

The actual 3D point is *(X/W, Y/W, Z/W)*.

2.2.3 Internal Parameters

The internal parameters specify the optical center of the image, any distortion caused by the lenses, the focal length along horizontal and vertical directions, and any *skew* from slanted pixels. These parameters are listed separately for each camera. The parameters for the input images are listed in the parameter file explicitly, as *parameter value* pairs. The parameters for the rectified images are implicitly

given by the projection matrix for the camera: see

$$\begin{bmatrix} F_x & 0 & C_x & -F_xT_x \\ 0 & F_y & C_y & 0 \\ 0 & 0 & 1 & 0 \\ \end{bmatrix}$$

Table 2-1.

Parameter	**Input Image**	**Rectified Image**
pwidth, pheight [pixels]	Size of the full-frame images used for calibration. This parameter is necessary in order to have calibrations work for different size input images.	Not applicable
dpx, dpy [mm]	Size of the pixels in the calibration images. These parameters are not used internally by SVS, but are for user information. The focal length in mm can be computed using these parameters and *f, fy*.	N. A.
sx [ratio]	Aspect ratio of the focal length in the horizontal direction to the focal length in the vertical direction. Only used with v2.3- calibrations; for v2.4+ calibrations, the aspect ratio is encoded by the ratio of *f* to *fy*.	1.0 Rectified images always have an aspect ratio of 1.0.
Cx, Cy [pixels]	Optical center of the image, where the principal ray intersects the image. Usually close to the actual center of the image.	Optical center of the rectified image. Usually close to the input image center, but can be displaced so that more of the input images are shown in the rectified image.
f, fy [pixels] v2.4+ [mm] v2.3-	Actual focal length of the camera. For v2.4+, the two different focal lengths can account for non-square pixels. For square pixels, they are usually very close. For v2.3-, only *f* is given (in mm). It is combined with the *sx* parameter to give the vertical focal length.	Focal lengths are the same (square pixels), and close to the input image focal length.
alpha [ratio]	Skew parameter. A zero value indicates no skew in the pixels (rectangular rather than trapezoidal). All calibrations assume zero skew on input images.	Always 0.
kappa1-3 tau1-2 [number]	Lens distortion parameters. The *kappa* parameters are radial distortion parameters, and the *tau* are tangential distortion. Typically only *kappa1* and *kappa2* are used, except for lenses with very high distortion.	N. A.

2.2.4 External Parameters

The external parameters describe the relation between the two cameras. These parameters are based on the left camera as the reference coordinate system. The right camera position is described by six parameters – three position coordinates and three rotation coordinates. For the input images, all parameters are necessary. For the rectified images, all the parameters are zero, except for translation along the X axis, which is the baseline separation of the cameras.

As with internal parameters, the external parameters for the input images are listed explicitly in the parameter file. The single external parameter, *Tx*, is part of the projection matrix for the right camera (see

$$\begin{bmatrix} F_x & 0 & C_x & -F_x T_x \\ 0 & F_y & C_y & 0 \\ 0 & 0 & 1 & 0 \\ & & & \end{bmatrix}$$

Table 2-1).

Parameter	**Input Image**	**Rectified Image**
Rx, Ry, Rz [radians]	Rotation of the right camera with reference to the coordinate system of the left camera. These values are a *rotation vector*. The angle of the vector is the rotation axis, and the magnitude of the vector is the amount of rotation (in radians).	Always 0.
Tx, Ty, Tz [mm]	Translation of the center of the right camera with respect to the coordinate system of the left camera. The *Tx* parameter is usually negative (the right camera is to the right of the left camera).	All zero except for *Tx*, which is the baseline between the cameras.

2.3 Internal Camera Model

Our model of the SVS stereo head consists of internal parameters of both the left and right cameras and the external parameters describing the rigid transformation between the two cameras.

First we present the internal parameters. These parameters describe the distortions introduced in each individual camera by imperfect lenses and lens placement. The biggest effects are from radial distortion, in which the image is compressed towards the edges; and lens decentering, in which the center of focus of the lens is not at the center of the imaging array. The effects are most prominent in wide-angle lenses; lenses with viewing angles of 50 degrees and less generally will not need any correction to help the quality of stereo matching. However, if range accuracy is important, then even narrow field lenses should be corrected. The correction is done digitally, by warping the nonideal image into the image that would be produced by an ideal camera.

(C_x, C_y)	Camera center, pixels
f_x, f_y	Focal length in horizontal and vertical directions, pixels
$\kappa_1, \kappa_2, \kappa_3$	Radial distortion parameters
τ_1, τ_2	Tangential distortion parameters

Table 2-2 Intrinsic camera parameters.

Our model of the camera intrinsics includes the parameters listed in Table 4-1 and follows the model of Tsai [1]. Given a 3D point $\boldsymbol{P} = (X, Y, Z)^T$ in left camera coordinates, we map to screen coordinates using the following steps. First, we project $\boldsymbol{P}$ to the normalized image plane using the standard perspective equations

$$X_u = \frac{X}{Z} \qquad Y_u = \frac{Y}{Z}, \tag{4.1}$$

where (X_u, Y_u) is the undistorted, normalized image coordinates. Next, radial lens distortion is modeled

$$\begin{aligned} X_d &= X_u(1+\kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6) + dx \\ Y_d &= Y_u(1+\kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6) + dy \end{aligned}, \tag{4.2}$$

by

where (X_d, Y_d) are the distorted image coordinates, $r^2 = X_d^2 + Y_d^2$, and the tangential distortions dx and dy are given by:

$$\begin{aligned} dx &= 2\tau_1 X_u Y_u + \tau_2(r^2 + 2X_u{}^2) \\ dy &= 2\tau_2 X_u Y_u + \tau_1(r^2 + 2Y_u{}^2) \end{aligned}, \tag{4.2b}$$

Finally, the distorted image coordinates are mapped to screen (pixel) coordinates (X_f, Y_f) using

$$X_f = f_x X_d + C_x \qquad Y_f = f_y Y_d + C_y, \tag{4.3}$$

[1] Tsai, Roger Y. "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the Shelf TV Cameras and Lenses", *IEEE Journal of Robotics and Automation*, vol. RA-3, no. 4, August 1987, pp. 323-344.

where f_x and f_y are the horizontal and vertical focal lengths of the lens, in pixels. It is also possible to account for *skew* in the pixels (non-rectangular pixels), but since modern imagers have negligible amounts of skew, we ignore this parameter.

This last equation can be written in a standard form using the camera matrix *K*:

$$K = \begin{bmatrix} f_x & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{X}_f = K\mathbf{X}_d \tag{4.4}$$

The above equations can be used to *undistort* an image, starting from a distorted input image, by resampling the image I_{left} :

$$I_{left}^{undist}(x, y) = I_{left}(x', y'),$$

where (x', y') is related to (x, y) through equations (4.2) and (4.3), as shown in Table 2-3

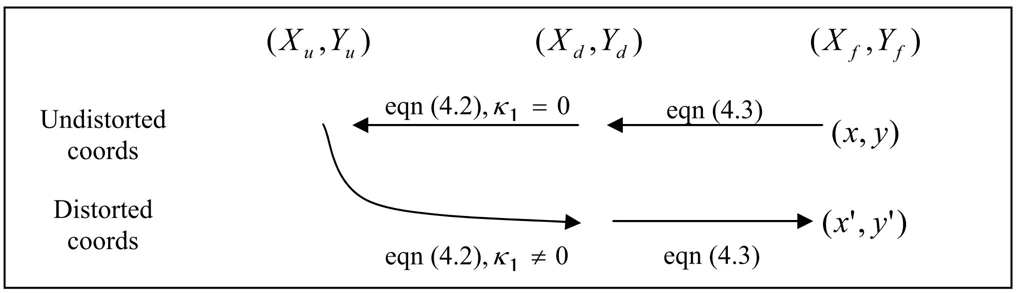

Table 2-3 Mapping from undistorted coordinates (x,y) to undistorted ones (x’,y’).

2.4 External Camera Model

For stereo matching to work well, the camera image planes must be co-planar, and the epipolar geometry adjusted so that corresponding scan lines match. As shown in Figure 2-2 and Table 2-4, there is a rigid transformation between the two cameras, with three translational and three rotational degrees of freedom. It is difficult to achieve this alignment mechanically unless a precise (and expensive) adjustment rig is incorporated into the stereo head. Fortunately, if the cameras are somewhat close to their ideal alignment, digital adjustment of the parameters is possible. Our calibration procedure estimates these external parameters and uses them to rectify the stereo images as described in the next subsection.

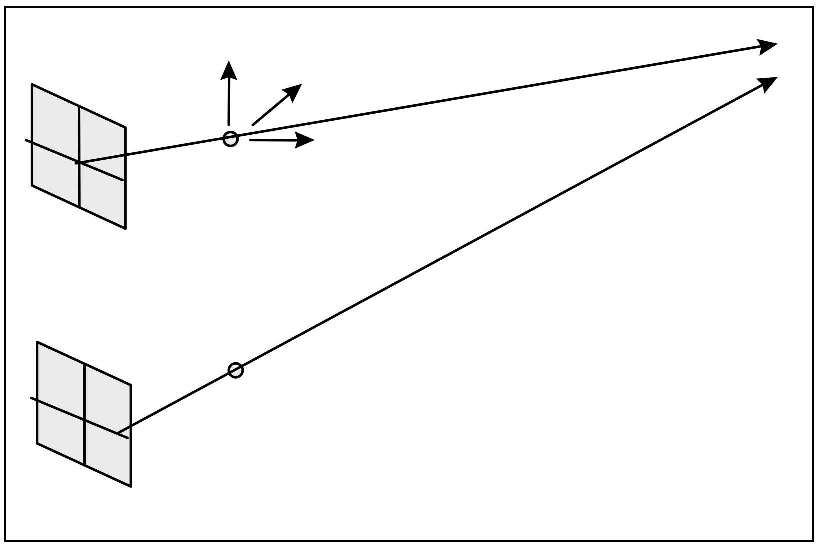

Figure 2-2 Stereo cameras and their spatial offset.

(T_x, T_y, T_z)	Location of right COP in left camera coordinates, mm
(R_x, R_y, R_z)	Rotation of right camera coordinate frame with respect to left camera, rad

Table 2-4 Stereo head external parameters

Stereo rectification compensates for the fact that the optical axes of the left and right cameras are not parallel. The output of the rectification procedure are image warps that can be used to effectively rotate the optical axes of the left and right cameras to create the ideal stereo setup. One can combine the rectification warping with the warp that corrects for distortion, so rectification adds no computational costs in the real-time warping code.

Our rectification code models projection using 3x4 projective matrices $\mathbf{P}$. A point in 3D $(X,Y,Z)^T$ is represented by homogeneous coordinates $(X,Y,Z,1)^T$ and the projection is modeled using a matrix multiply

$$s\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{P}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

where (x, y) are screen coordinates. This model assumes radial distortion has been eliminated as in Table 2-3.

The output of rectification is two projection matrices $\mathbf{P}_0$ and $\mathbf{P}_1$ of the rectified left and right cameras, respectively. The 3D coordinate system used by the projection matrices is the local left camera coordinate system. In addition, rectification produces two homographies $\mathbf{H}_0$ and $\mathbf{H}_1$ for warping the left and right images so that their optical axes are parallel. For instance, the following rectifies the left image:

$$I_{left}^{rect}(x, y) = I_{left}^{undist}(x', y'),$$

where

$$s\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \mathbf{H}_0\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

2.5 3D Reconstruction

The SVS API provides a function for reconstructing the 3D location of a point given $(x, y, \text{disparity})$ in the rectified left image. The (X, Y, Z) location of the point is computed in the coordinate frame of the left camera. See the API reference in the SVS User's Manual for more detail.

2.6 Setting the horopter

The rectification procedure sets up the horopter so that an X offset of zero places the plane of furthest match at infinity. This is often a good idea since it is usually possible to control how close objects get to the camera, but not how far away. On the other hand, if objects of interest are close to the camera and there are no distant objects, it would be better to move the horopter closer to the camera. By making the X offset, x_{off}, negative, we move the horopter to the following depth range

$$[z_{\min}, \quad z_{\max}] = \left[\frac{bf}{d_{\max} - x_{off}}, \quad \frac{bf}{-x_{off}}\right],$$

where $d_{\max}$ is the maximum pixel disparity. This may be useful for dynamically changing the horopter while tracking an object moving in depth.