Immersion using stereo cameras and 3d reconstruction

Rafael Gonalves Colares
Universidade Federal de Minas Gerais
rcolares@dcc.ufmg.br

Abstract

This paper presents a pipeline for processing a stereo video for teleimmersion. It is the first stage of a bigger project of a system for power line maintenance. In this paper is discussed the capture of stereo videos, as well as, all the process necessary to display the video in a teleimmersion device. It presents a discussion about camera calibration, image rectification, disparity maps and 3D reconstruction. In the end, it shows some previous results obtained from the pipeline.

1. Introduction

This paper presents a system for power line maintenance in its first stages. The system is composed a video capture system, assembled on a agent, remotely controlled. The capture system send to a computer images of the power lines to a computer, which manipulates the images and present then to the user through a immersion device. The user, then, can teleoperate the agent sending commands to it. In this paper, will be discussed the capture system, as well as, part of the process involved on the images manipulation.

Immersion systems, such as stereoscopic devices and projection devices (CAVE), are powerful tools for observing 3D scenarios. These tools are capable of providing a semi-real interaction with the environment, even when the real scene is far away from the user. Thus, using these devices to teleoperate equipments from long distances, seems like a natural option. Stereoscopic devices, combined with robotic applications, creating teleoperation equipments, are already used in various areas, such as medical surgeries[8], equipments maintenance[3] exploration of unknown locations[10].

All the applications listed above use a camera, or cameras, to capture the scene, and then displays it to the user. Although, it seems like a simple task, it requires more than it shows. A single camera does not provide the necessary precision of the environment to the user, it lacks of information like depthness. For this reason, it is necessary to try other approaches to extract the information required by the user.

One good approach is to use stereo cameras. It is formed by placing two cameras, side-by-side, or up and down, and capturing the images at the same time. By doing so, one can have two images, from the same scene, from different angles, and can infer the depthness of the objects by the triangulation of the image points, presenting for the user a better precision of the scene.

The use of stereo cameras carries various issues that have to be taken care. First, synchronization. The two cameras must start capturing the videos at the same time, or else, the system must should have a way to synchronize both cameras. This is necessary because the same frame from each video has to be displayed to the user at the same time, or else the reconstruction will be carried wrongly.

Another issue is the camera calibration. As the cameras are in stereo formation, it is necessary to compute both intrinsic ad extrinsic parameters from the cameras and calculate the base line, which is the physical distance between the center image of the left camera to the center image of the right camera. With the calibration and the synchronization, it is possible to start the reconstruction of the scene by looking for correspondences from the left frame to the right. All this process, from the capture system to the 3D reconstruction is describe in details on chapter 3.

2. Related Work

A lot of work has been done on the stereo reconstruction field. In [7] is listed various stereos configurations, as well as, different strategies for using it. They also try to categorize the equipments, by number of cameras, or the visualization system. They discuss advantages and disadvantages from each one and give theirs opinion about each one.

At [3] they present a system for maintenance of power lines, with stereo cameras and a teleimmersion device, so that the user can control the agent from far away. The main difference from this paper and [3] is the nature of the agent and the processing of the videos.

As said before, teleimmersion is a powerful technique for manipulating devices from long distances. [8] brings a study about the use of this technology in medical surgeries. [10] analyzes the use of various immersion systems in the problem of the space exploration. And [6] discuss various insights about stereoscopic images for robots that are teleoperated.

3. Methodology

The project can be described as a circular pipeline that goes from the agent to the user, passing through several stages of image processing, and back to the agent again. Figure 1 shows a image of the pipeline.

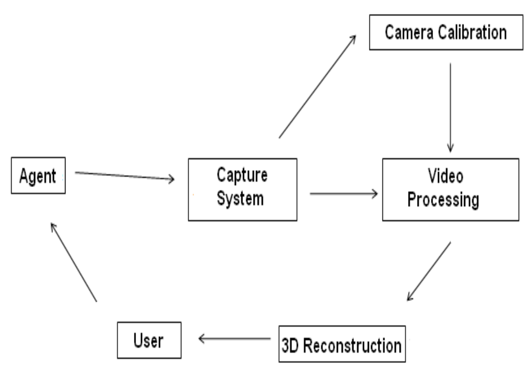

Figure 1. Methodology.

The first stage is the capture system. It will be assembled on the agent, and will be responsible for capturing videos in stereo configuration. There is now an extra stage, the camera calibration stage. It is done off-line, and its outputs are important for all the pipeline. Next comes the processing stage. This stage is divided by three steps:

- Finding the epipolar line
- Image Rectification
- Finding Disparity Map

And last, the 3D reconstruction, which uses the output from all other stages. It is done by processing a triangulation with the points on the disparity map and a projection matrix, to express depthness on the scene.

The pipeline was implemented in C language with the Open Source Computer Vision(OpenCV)[1] library.

All this stages are going to be explained in details next.

3.1. Capture System

The capture system is responsible for providing to the pipeline the videos in stereo. The configuration is side-by-side frontal-parallel, which means that the cameras are arranged on a left-right orientation, and the normal vectors of the cameras planes are parallel.

This configuration is good to find correspondences between the two images. As its geometry is simple, it is easy to find correlated points from the left image on the right. Figure 2 shows the geometry involved on the system.

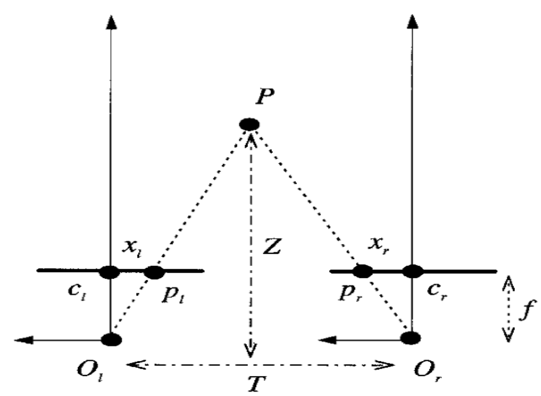

Figure 2. Parallel Stereo Configuration.

O is the center of the cameras planes, f is the focal distance, c is the center point of the images. P is a scene point, and pl and pr are the projections of point P on the image plane. T is the base line, the distance between Ol and Or. All this parameters can be discovered by running a camera calibration.

The cameras used were two Dragonflies mounted at a tripe on a frontal-parallel configuration.

3.2. Stereo Camera Calibration

This is a important stage of the pipeline. Despite of the fact that it is an extra stage, because it is done off-line, this stage produces outputs that will be used throughout all the pipeline.

It is in this stage that the intrinsic and extrinsic parameters of the camera are calculated. This is done by running the Bouguet's[2] algorithm implemented on the OpenCV library. The main goal of this calibration is to find the Essential and Fundamental matrices. The first one contains informations about rotation and translation between both cameras in the physical world, involving the extrinsic parameters of the cameras. The second, contains practically the same information as the essential matrix in addition to the intrinsic parameters from both cameras, relating the two cameras in pixel coordinates.

3.3. Image Processing

This stage can be divided in three steps: the first is to find the epipolar line, second, rectify both left and right images, and third, find the disparity map. As said before, the calibration stage plays a important role in all three steps, as the geometry of the stereo configuration is exhaustively used.

3.3.1. Finding Epipolar Lines Given a point from the left image, the epipolar line is the line in which the correspondent point should appear on the right image. Finding this line is important because it reduces the search area to look for correlated points on both images.

Finding the epipolar line is a matter of geometry, and since the camera calibration has already been executed at this time, it is now possible to calculate the lines. The epipolar geometry is represented in figure 3. Using the Fundamental matrix, one can have the points correspondences in pixels from the left image to the right, so it is possible to draw the epipolar lines.

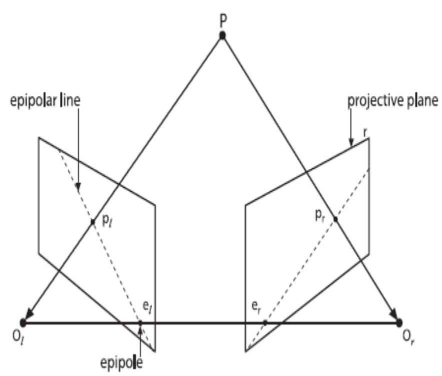

Figure 3. Epipolar Geometry.

The main problem about the epipolar lines is that not always the lines are equal for both images. Usually, the epipolar line on the right image is not horizontal, but mostly it has an angle with the x-axis. This issue can be of great importance if the epipoles are calculated more than one time, or even on large images. For this reason, there is another process that rectify the images.

3.3.2. Image Rectification As explained before, this step is important, so that the search area for the stereo correspondence is held over a horizontal line through the right image. Actually, the aim of the rectification process is that every row in both images to be aligned, so that every correspondence points have the same y coordinate as on the left image as on the right. To achieve rectification, the epipoles tend to be located at infinity.

In this initial stage of the project, two rectification algorithms were tested, Bouguet's and Hartley's[9]. Both algorithm were tested with the same inputs, and the results are discussed on chapter 4.

The main difference between the two algorithms is that Hartley's does not require previous camera calibration while Bouguet's does. Hartley's attempts to find homographies that map the epipoles to the infinity while minimizing the disparity between the stereo images. Actually, this algorithm only needs the fundamental matrix, which con be obtained by any 7 or more matched points from both images. Bouguet's algorithm just uses the rotation and translation matrix to reproject one image as the other.

3.3.3. Finding Disparity Map Finding the disparity map give us information about depth measurements, which is important for the 3D reconstruction of the scene. This is done by sliding a "sum of absolute difference" (SAD) windows to find matching points on the left and right rectified images. As this can be an exhaustive process to be held on every pixel from both images, here comes the importance of the first two steps.

The rectification align every row from both images. This implies that the epipolar line is align as well. The idea now is to copy the epipolar line, and distribute it through all the images. By doing so, one can limit the sliding area of the SAD windows for only the lines created, making it less computer exhaustive.

3.4. 3D Reconstruction

All the stages discussed until now, prepares the system for this stage. This is where the points calculated are reconstructed to represent the depthness of the scene.

The disparity map have only the information of proximity of the camera lenses, like if one object is closer to the camera than other. For processing the reconstruction, is necessary a reprojection matrix, with information of scale. This matrix can be defined arbitrary or can be calculated on the camera calibration stage. OpenCV's calibration function already calculates this matrix using Bouguet's algorithm.

With the disparity map and the reprojection matrix, it is just a matter of geometry to calculate the 3D reconstruction.

4. Tests and Results

There were held two tests: first with synthetic images, and another with videos taken from the two Dragonflies cameras.

4.1. Synthetic Images

The synthetic images were chosen from the Matlab Camera Calibration Toolbox website [2]. These images are from a chess board, taken from fourteen different angles. Figure 4 illustrates these images.

Figure 4. Chess boards images.

For these set of images, both Bouguet's and Hartley's rectification algorithm were tested. Both algorithm presented a satisfying result, as can be seen in figures 5 and 6. They represent the left and right images, already rectified and with the epipolar line distributed through all the images. It is easy to note that the pixels the are on the same line on both images represents a stereo correspondence. For synthetic images the results are similar.

For the disparity maps, using the rectified images cited above, Bouguet's and Hartley's algorithms presented different results. The first tries to find the disparity map through all the points on the image, and the second tries to find homography between the two images. For this reason each algorithm outputs a different disparity map. Figure 7 and 8 illustrate it.

As the disparities maps are different, the reconstruction will be different as well. Despite this fact, both results represents the depthness of the scene as they found on the disparity map. The images can be seen on figures 9 and 10.

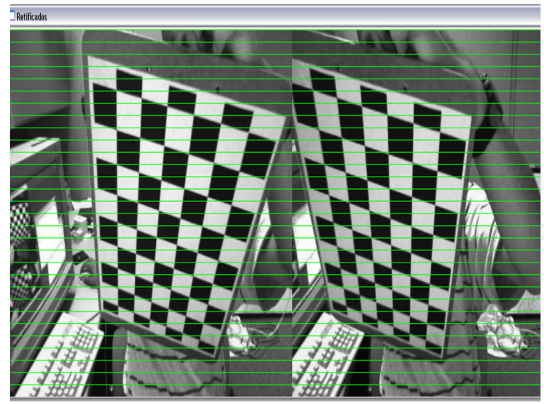

Figure 5. Rectified Images using Bouguet's algorithm.

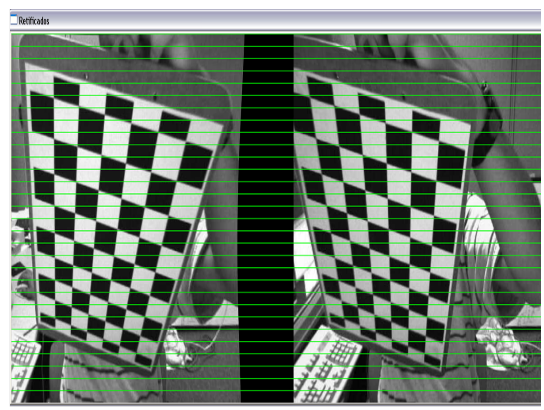

Figure 6. Rectified Images using Hartley's algorithm.

4.2. Videos

The videos were taken with two Dragonfly cameras, assembled on a frontal-parallel configuration. The results obtained from this test is still experimental and has to be better analyzed. A sample of the images used for this test can be seen at figure 11.

The rectification from Bouguet's method did not returned a useful output. One reason for this result is that for this set of test only 3 different images were used, against 14 from the synthetic images. As Hartley's algorithm does not need to do a camera calibration, it returned better results, but they have to be improved to start being accept-

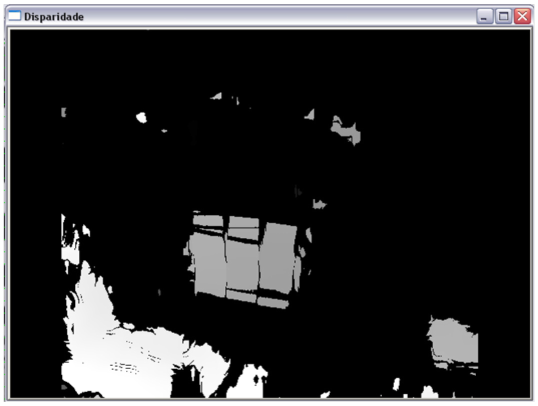

Figure 7. Disparity map using Bouguet's rectification.

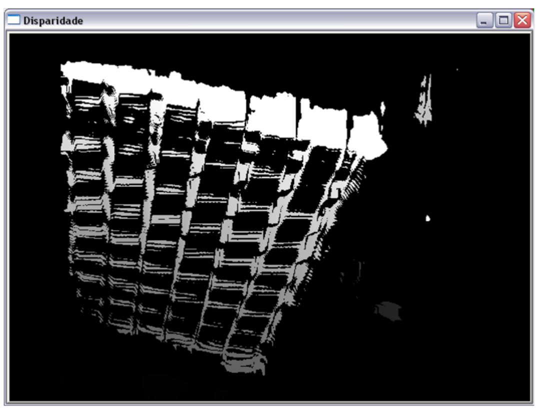

Figure 8. Disparity map using Hartley's rectification.

able. The Hartley's rectification image can be seen on figure 12

The disparity map were only calculated for Hartley's algorithm. It returned a expected result for the rectified image calculated, but still, not a good result. The disparity map is shown on figure 13

The 3D reconstruction was not able to be processed, because the camera calibration does not provide a valid reprojection matrix to be used with the reconstruction.

5. Conclusion an Future Works

The system is on its initial state, so different strategies are being tested to decide the best one. This paper presented the video processing stage, determining what has to be done, and presenting the first results obtained.

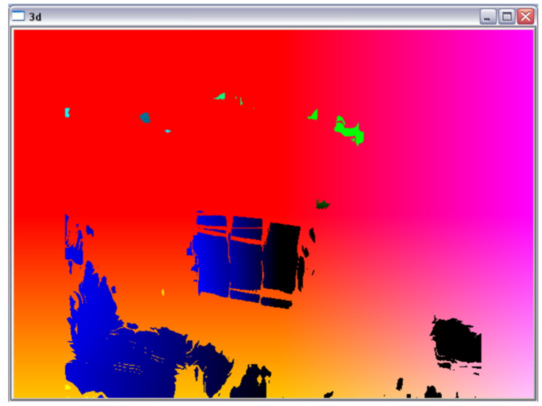

Figure 9. 3D Reconstruction using Bouguet's rectification.

Figure 10. 3D Reconstruction using Hartley's rectification.

As for the Calibration system, the Bouguet's algorithm is a good choice, since it returned good results. For the rectification system, so far, Hartley algorithm showed better results. It is, in part, because it does not depend on the calibration stage, which can have numerical errors. And since the system will need to run in real-time, and Hartley does not need a previous calibration, it is a possible choice.

The reconstruction stage is returning good results for the inputs it is given. The main problem about it, is that the points chosen for the reconstruction does not, necessarily, are points of interest. For a better scene reconstruction, a object identify algorithm should be implemented.

References

[1] http://sourceforge.net/projects/opencvlibrary/.

[2] http://www.vision.caltech.edu/bouguetj/calib_doc/.

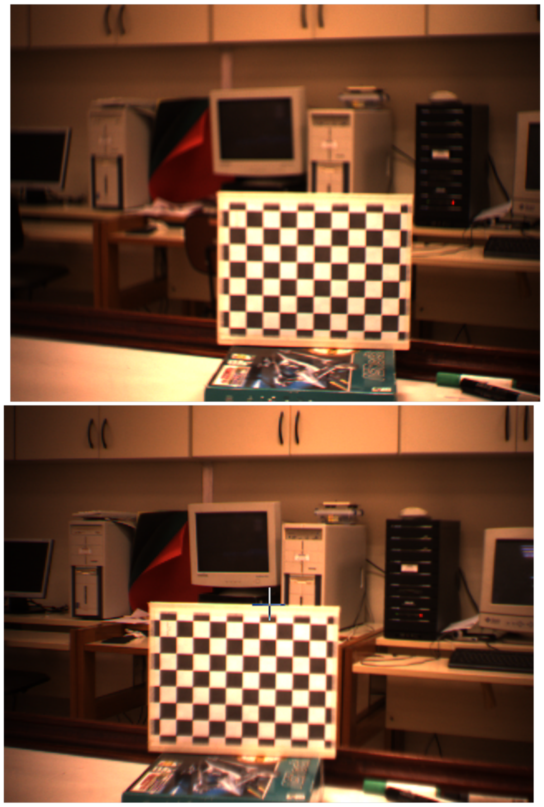

Figure 11. Chess boards Videos.

[3] R. Aracil, M. Ferre, M. Hernando, E. Pinto, and J. M. Sebastian. Telerobotic system for live-power line maintenance: Robtet. *Control Engineering Practice*, 10(11):1271–1281, November 2002.

[4] J. Baker. Generating images for a time-multiplexed stereoscopic computer graphics system. In *Int. Soc. for Optical Engineering, True 3D Imaging Techniques and Display Technologies*, pages 44–52, 1987.

[5] G. Bradsky and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV library*. O'Reilly, Sebastopol, CA, 2008.

[6] M. Ferre, R. Aracil, and M. Navas. Stereoscopic videos images for telerobotics applications. *J. Robot. Syst.*, 22(3):131–146, 2005.

[7] M. Ferre, R. Aracil, and M. Sanchez-Uran. Stereoscopic human interfaces. *Robotics and Automation Magazine, IEEE*, 15(4):50–57, December 2008.

[8] G. S. Guthart and J. K. Salisbury. The intuitive telesurgery system: Overview and application. In *IEEE Int. Conf. Robotics and Automation*, pages 618–621, 2000.

[9] R. I. Hartley. Theory and practice of projective rectification. *International Journal of Computer Vision*, (35):115–127, 1998.

[10] G. Podnar, J. Dolan, A. Elfes, M. Bergerman, H. B. Brown, and A. D. Guisewite. Human telesupervision of a fleet of autonomous robots for safe and efficient space exploration. In *ACM/IEEE 1st Annu. Conf. Human-Robot Interaction*, pages 325–326, 2006.

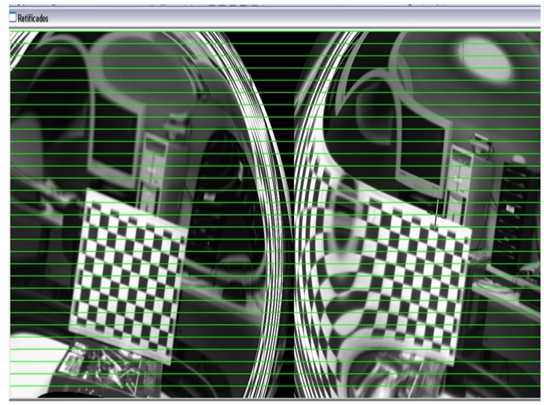

Figure 12. Rectified video using Hartley's algorithm.

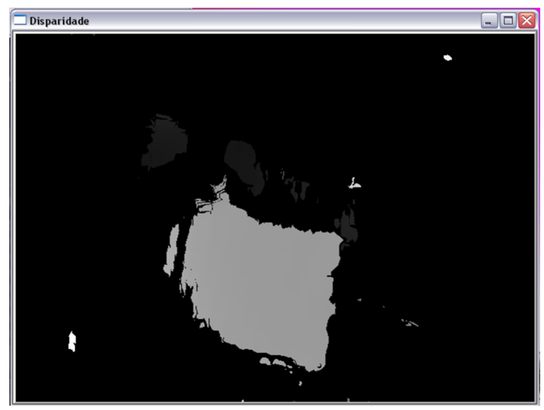

Figure 13. Disparity map from video using Hartley's algorithm.