Epipolar Rectification for Autostereoscopic Camera Setup

Fabrice Boutarel, Vincent Nozick

The 8th France-Japan and
6th Europe-Asia Congress on Mechatronics
November 22-24, 2010, Yokohama, Japan

Epipolar Rectification for Autostereoscopic Camera Setup

Fabrice Boutarel
Gaspard Monge Institute, UMR 8049
Paris-Est Marne-la-Vallee University, France
Email: fabrice.boutarel@univ-mlv.fr

Vincent Nozick
Gaspard Monge Institute, UMR 8049
Paris-Est Marne-la-Vallee University, France
Email: vincent.nozick@univ-mlv.fr

***Abstract*—This paper presents a method to perform geometric transformations on stereoscopic images. Our paper mainly focuses on epipolar image rectification for more than two views. Indeed, this rectification is well suited to remove the vertical parallax that is a significant cause of headache related to stereoscopic perception. We show that this rectification should satisfy some constraints to provide a 3D restitution in the correct geometric proportions.**

I. Introduction

The development of 3D-TV and autostereoscopic displays makes numerous researchers and commercial companies focus on stereoscopic rendering. Indeed, stereoscopic rendering involves a variety of issues such as data compression, 3D displays, colorimetric rectifications, etc. Our paper is about stereoscopic image rectification. Image rectification is a geometric operation that makes a rough stereoscopic pair comfortable to see in 3D.

Concerning stereoscopic issues, most of the current research is done for two views, but not for $n > 2$ views required for most of the autostereoscopic displays. In this paper, we will show that the extension of the image rectification for n views is not straightforward.

The paper is organized as follows: we will first remind the context of stereoscopic rendering and especially on how to generate a comfortable stereoscopic image pair that provide a 3D restitution in the correct geometric proportions. Then, we will overview the camera calibrations methods well suited for a stereoscopic camera rig. The next part presents the state of the art of epipolar image rectification commonly used for computer vision problem. Finally, we introduce our method to perform this rectification on a set of more than 2 cameras.

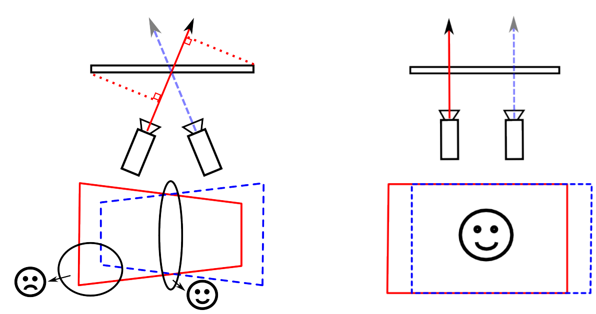
II. Stereoscopic rendering

As presented by Meesters et al. [8] and Lambooij et al. [6], stereoscopic rendering involves numerous constraints and rules to follow. Indeed, in order to compute a stereoscopic image, like an anaglyph, we need to superimpose two images of a scene, taken from a pair of cameras. The choice of the method used for doing this superimposition has strong consequences on the 3D perception of the represented scene.

A. Vertical Parallax

When doing this superimposition, the position of an object in one image will probably not be the same in the other image. This displacement, called parallax, has two components: the horizontal parallax which is what makes possible for us to see in 3D, and the vertical parallax that causes headaches and that we want to minimize. As depicted on Figure 1-(a), the vertical parallax can come from a difference of perspective between the two eyes due to the convergence of the Human vision. This effect, presented in Woods et al. [11], is negligible around the convergence zone but not in the rest of the view. Fortunately, in the every day life, it does not disturb our vision thanks to our very small field of view. However, in the stereoscopic case, only a small area of the stereoscopic image corresponding to the focus point will not be affected by the vertical parallax, but seeing the rest of the image will be uncomfortable.

The usual way to avoid this vertical parallax is to use a camera setup such as camera optical axis are parallel. This technique removes the perspective effect that induces the vertical parallax, as shown in Figure 1-(b). Satisfying this constraint is not optional in order to generate comfortable stereoscopic images.



(a) Toe-in configuration (b) Parallel configuration

Fig. 1: (a) The camera optical axis convergence involves a perspective effect that generates vertical parallax. (b) Parallel camera optical axis removes the vertical parallax.

Moreover one have to be careful of camera rotation around the optical axis and radial distortion [9], which also generates vertical parallax.

B. Orthostereoscopic rendering

As shown by Jones et al. [5], the way to superimpose the two input images will also impact on the proportion perception of the 3D scene. When we see stereoscopic images with all objects keeping the same 3D proportion as in the original scene, we talk about orthostereoscopic rendering. This might not be the case if an input image is wrongly slid along the other. In order to obtain this orthostereoscopic rendering, the pair of camera must be assimilated to Human eyes (except for the fact that their optical should be parallels). Other parameters such as relative position of the cameras must be chosen such as they match Human eyes as well as possible, up to a scale factor.

We can note that the orthostereoscopic rendering is designed for an unique position, where the original scene can be seen as if the camera were your eyes. This position is called the orthostereoscopic point of view and will conserve all the proportions of the objects in the scene. Moreover, watching a stereoscopic image around this position will also provide a comfortable 3D restitution.

C. Constraints on the cameras setup

In a standard stereoscopic case with two views as well as for autostereoscopic displays involving more than two cameras, vertical parallax minimization and orthostereoscopic rendering lead to the following constraints on the cameras :

- the cameras should have the same focal length.
- the optical axis of the camera should be parallel
- the camera should share the same "vertical".
- the camera center should be aligned along a "horizontal" line.(if more than 2 views)
- the camera center must be equidistant from one to the next, since all adjacent pair of camera represent the same eyes pair.
- the radial distortion should be corrected on every view.

In practice, the camera rigs are often set up manually and the above constraints will never be perfectly satisfied. However, most of the setup imperfection can be partially or totally corrected by computer vision tools. More precisely, focal length or camera orientation problem can be perfectly corrected up to a worsening of image resolution, but camera center position error cannot be corrected without a full 3D reconstruction of the scene. For the first class of problem, the computer vision solutions involve that the cameras are calibrated i.e. the optics and geometrical relation between the cameras are known.

III. Camera calibration for autostereoscopic setup

The set of cameras can be calibrated with very well known techniques using 2D markers like Zhang [12] or 3D markers [4] (gold standard algorithm). However these methods usually involve a manual set up that is better to avoid.

In the case of stereoscopic rendering, some automatic methods may take advantages of the constraints on the cameras such as the relative pose or the similarity between each camera internal parameters. These automatic methods require an initial estimation, usually assuming that the camera internal parameters are known (i.e. focal length, principal point [4]). If it is not the case, a usual initial value for the focal length can be chosen as the image diagonal distance (in pixel units) and the principal point as the center of the image. Moreover, if the cameras have all the same optics, we can consider that they have the same internal parameters. The automatic methods also require an estimation of the camera external parameters (i.e. position and orientation). Again, the stereoscopic constraints can be easily used to set as an initial value the rotation matrix to identity and the camera center position as $(i, 0, 0)^\top$ where i denotes the index of the camera.

Then, some pixel correspondences between the camera images should be determined using image descriptors like SIFT [7]. These correspondences are use to compute an initial least square 3D reconstruction. Finally, the internal and external parameters of the camera, as well as a 3D reconstruction of the selected points, can be refined using a bundle adjustment technique [10]. For this purpose, a good guideline of the implementation of Levenberg-Marquardt algorithm can be found in [4]. We can note that the radial distortion can also be corrected during the bundle adjustment.

IV. Image rectification for autostereoscopic setup

In a computer vision framework, epipolar image rectification is a process that transform a pair of stereoscopic image such as their conjugate epipolar line become collinear (see Figure 2). In a stereoscopic framework, rectifying two images completely removes the vertical parallax. Indeed, image rectification can be considered as a camera rotation around its center and a variation of the image focal such the two cameras share the same image plane (i.e. the camera optical axis are parallel). In practice this transformation corresponds to the computation of a 2D homography for each images.

Fig. 2: Rectifyed image pair using Hartley [3] : the epipolar lines are horizontal and aligned from an image to the other. The rectified images are suited for computer vision processing, but not for stereoscopic rendering.

A. Standard image rectification methods

First, we can notice that the epipolar rectification can be performed without a full camera calibration and also that this problem has an infinity of solution that are not equivalent in term of stereoscopic perception. The methods based on epipolar geometry, like Hartley [3], select the solution that minimize the image distortion. These approaches can be extended to the multi-camera case, however the selected solution might not be optimal for stereoscopic purposes. Moreover, there is no guaranty to obtain a stereoscopic pair that reaches orthostereoscopic rendering. Fusiello et al. [1], [2] present an epipolar rectification method for calibrated images designed for computer vision purpose, however his method presents no guaranty on the orthostereoscopic rendering.

In the next part, we propose a rectification method that takes advantage of the camera calibration using a geometric interpretation of the camera setup, and that guaranties an ortostereoscopic rendering.

B. Proposed method

Let's consider a set of n fully calibrated cameras, i.e. the camera internal and external parameters are known (cf. III). Let's define for each camera a 3D referential (Figure 3) composed of three orthogonal axis:

- The optical axis of the camera, or roll axis.
- The yaw axis (the up vector in GPU programming).
- The pitch axis (parallel to image baseline).

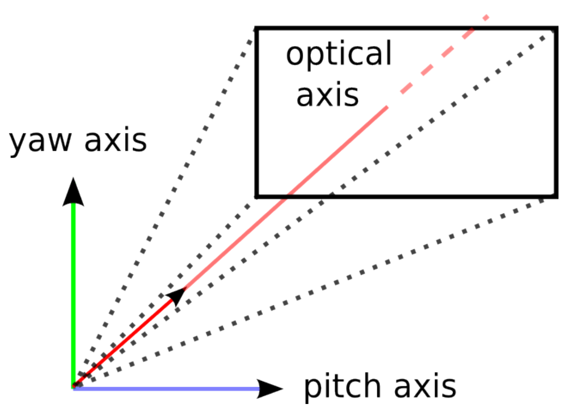


Fig. 3: Camera 3D referential.

The epipolar rectification is equivalent to a 2D homography on each images (cf. IV). In the case of a setup with $n = 2$ cameras, these homographies can be considered as two operations which transform the two calibrated input cameras into a pair of new cameras, that are identical up to a translation on the pitch axis. These cameras share the same focal length. Moreover they have parallel optical axis and yaw axis. As explained in section II-C these constraints guaranty a setup of rectified cameras.

For an autostereoscopic setup, we want to correct $n > 2$ images. Even if image rectification for 2 images works fine, it will not be applicable for more that 2 views, since it does not ensure any global coherence between all rectified images.

We propose a method to transform a set of n fully calibrated calibrated cameras in n rectified cameras (Figure 4). As defined in section II-C, the rectified cameras should satisfy some constraints. First, all the rectified cameras center have to share the same horizontal axis. Moreover each center must be equidistant with its 2 neighbours camera center. This horizontal axis is chosen such as the sum of distances between each input camera center and this line is minimal (least square method). All rectified cameras must also have parallel optical axis, in order to put all rectified camera pairs epipoles at infinity. This common optical axis is defined as the average of the input cameras optical axis such as it also verifies the orthogonality with the previously defined horizontal axis. Respecting these constraints is mandatory to remove vertical parallax and maximize the stereoscopic perception comfort.

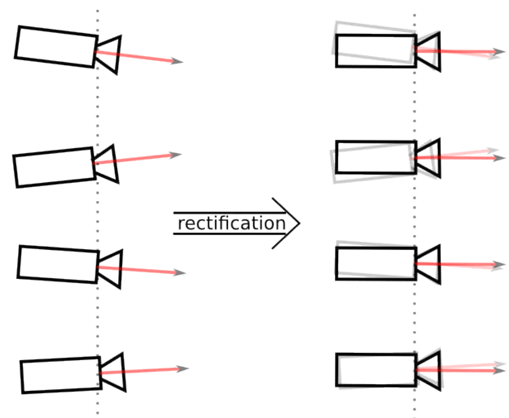


Fig. 4: The rectification process transforms a set of cameras in a new set of rectified cameras.

As presented in section 2, the rectification of each input camera is performed by an homography. In our method, we express each homography as an image projection from the calibrated camera on a common projection plane, followed by a back-projection on the associated rectified camera. In order to have some coherence between all the rectified images, the projection plane should be common for every input camera to be rectified. We can note that any random plane will give a set of rectified images, however these images are more or less distorted according to the orientation of this plane. Indeed, the more the projection plane is parallel to the optical axis of the input cameras, the more more the projected images will be distorted. This will involve a bad 3D restitution and cause some resolution loss during the back-projection. The projection plane orientation is defined such as it faces as much as possible all cameras. For this purpose, the normal of the projection plane needs to be parallel to the rectified cameras optical axis.

The next step is to ensure orthostereoscopic conditions. In addition to the constraints presented in part II-C, we have to define the distance between the rectified cameras and the projection plane. For this purpose, we have to take into consideration the restitution device. Indeed, the distance between the rectified cameras and the projection plane should be chosen in consideration to an analogy between the viewer if front of the stereoscopic display, and the cameras positioned in front of this plane. The following parameters should be equals, up to a common scale :

- The distance between the eyes of the viewer, called Intra Ocular Distance ($\sim$ 65 mm for Humans) $\leftrightarrow$ the distance between two neighbouring cameras.
- The distance between the viewer and the screen $\leftrightarrow$ the distance between the cameras and the projection plane.
- The size and resolution of the screen $\leftrightarrow$ the size and resolution of the projection plane.

Since the first point is usually already fixed, it will indicate the common scale to compute the adequate size of the plane and its distance from the cameras.

Finally, the rectification can be computed for each input camera by projecting the input image on the common projection plane and back-projecting this image on the associated rectified camera. This method rectifies the input images, regardless of the input cameras optical axis orientation. However it can not correct wrong set up of the cameras center.

This method can also be used to generate the stereoscopic image build from the rectified images. In this case, the projection plane can be considered as the screen and the back projection is not necessary. The parallax between all the projected views will be automatically computed to reach orthostereoscopic rendering.

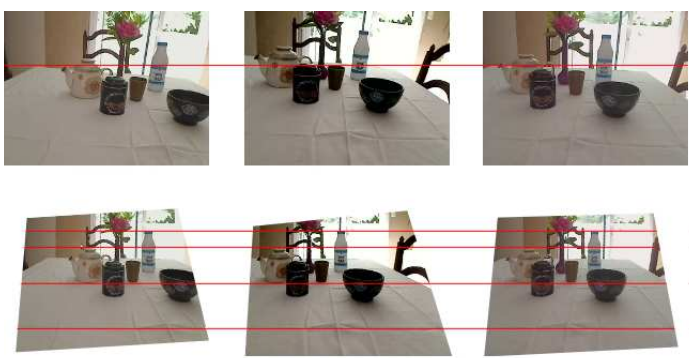


Fig. 5: Image rectification of a set of images.

V. Tests and results

We implemented our method in C++ language. Concerning the camera calibration, we tested both Zhang method [12] and a bundle adjustment from a set of 2D correspondences. These two methods give good results to use for the image rectification.

We implemented the image rectification on the GPU using GLSL. Indeed, this method is very well suited to be used with the GPU since every image can be rectified independently. We use projected textures to compute the image rectification and multi-texturing for the stereoscopic image computation. Since projected texture is a fully parallel process, this method is extremely fast and well suited for autostereoscopic displays that can involves to rectify more than height images simultaneously.

Fig. 6: Stereoscopic image computed with our method.

Figure 5 depicts some results of rectified images. Figure 6 shows an anaglyph image computed with our method.

VI. Conclusion

In this paper, we present an image rectification method designed for stereoscopic purpose, that can rectify more than two views simultaneously. Moreover, our method is easily set up to render the stereoscopic image in orthostereoscopic conditions, i.e. such that the 3D scene is perceived in the same proportion as the original scene. We show that this method is well suited to be implemented on the GPU and performs very fast image rectifications.

References

[1] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 1999. 3

[2] A. Fusiello, E. Trucco, A. Verri, and R. Verri. Quasi-euclidean epipolar rectification. *ICPR08*, pages 1–4, 2008. 3

[3] R. I. Hartley. Theory and practice of projective rectification. *International Journal of Computer Vision*, 35(2):115–127, 1999. 2, 3

[4] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. 2

[5] G. R. Jones, D. Lee, N. S. Holliman, and D. Ezra. Controlling perceived depth in stereoscopic images. *Stereoscopic Displays and Virtual Reality Systems VIII*, 4297(1):42 – 53, 2001. 1

[6] M. T. M. Lambooij, W. A. IJsselsteijn, and I. Heynderickx. Visual discomfort in stereoscopic displays: a review. *Stereoscopic Displays and Virtual Reality Systems XIV*, 6490(1), 2007. 1

[7] D. G. Lowe. Object recognition from local scale-invariant features. *International Journal of Computer Vision*, 2(1):1150, 1999. 2

[8] L. Meesters, W. IJsselsteijn, and P. Seuntiens. A survey of perceptual evaluations and requirements of three-dimensional tv. *CirSysVideo*, 14(3):381–391, 2004. 1

[9] G. Stein. Lens distortion calibration using point correspondences. *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, page 602, 1997. 1

[10] B. Triggs, P. McLauchlan, R. I. Hartley, and A. Fitzgibbon. Bundle adjustment – a modern synthesis. *Vision Algorithms: Theory and Practice*, pages 153–177, 2000. 2

[11] A. Woods, T. Docherty, and R. Koch. Image distortions in stereoscopic video systems. *Proc. SPIE*, 1915:36–48, 1993. 1

[12] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000. 2, 4