Scene reconstruction and geometrical rectification from stereo images

Antonio Javier GALLEGO SÁNCHEZ, Rafael MOLINA CARMONA, Carlos VILLAGRÁ ARNEDO
Grupo de Informática Industrial e Inteligencia Artificial (I3A)
Universidad de Alicante
Ctra. San Vicente del Raspeig, S/N
03080 – Alicante (Spain)
Tel. +34 965 90 3900
Fax. +34 965 90 3902
{ajgallego, rmolina, villagra}@dccia.ua.es

ABSTRACT

A system to reconstruct three-dimensional scenes from stereo images is presented. The reconstruction is based on a dense disparity image obtained by a process of window correlation, applying a geometrical rectification before generating a three-dimensional matrix which stores the spatial occupation. The geometrical rectification is essential to correct the conical perspective of the camera and to obtain real scenes. For the geometrical rectification, three approaches are proposed, based on one linear and two logarithmic functions. As a result, the rectification allows the system to adequately correct the reconstruction, showing that the best choice depends on the features of the original images.

Keywords: *stereoscopic vision, disparity images, geometrical rectification and three-dimensional scene reconstruction.*

1. INTRODUCTION

A central aspect nowadays in artificial intelligence research is the perception of the environment by artificial systems. Specifically, stereoscopic vision opens new paths that in future will allow these systems to capture the three-dimensional structure of the environments without any physical contact. For instance, a first solution to three-dimensional reconstruction with stereo technology is developed in Carnegie Mellon University. The possibility of composing several three-dimensional views from the camera transforms is set out, to build the so-called "3D evidence grid" [1]

Most proposals in this field are based on disparity maps obtained by the extraction of scene characteristics, such as corners, borders, and so on, or this extraction can be done once the disparity map is obtained [2][3]. There are some solutions which use a background image to obtain objects and silhouettes, making a difference operation with the image to reconstruct the scene [4][5]. It is also habitual to use a general model or *a priori* knowledge to compare the result with, so that the reconstruction is built using this knowledge as a model. This is the case with the reconstruction of faces or known objects [6]. For instance, in [2] the author reconstructs some basic objects from a stereo image using primitives, such as cubes or boxes, and the objects are displayed in 3D. Moreover, several views of a scene or a sequence of them can be used to make the reconstruction [7]; or it can be based on other type of sensors, such as laser range sensors.

All these algorithms cannot be applied in a general manner and their field of application is limited. Nevertheless, our proposal has some advantages, due to the fact that it does not make assumptions about the scene nor the object structure, no characteristics are extracted, it does not segment objects trying to find a known shape and a stereo pair only is used, not a sequence.

The method that we propose reconstructs a three-dimensional scene from a dense disparity map (the map contains depth information for every pixel in the image) obtained from a binocular camera. It makes a geometrical rectification to show the same aspect as the real scene, removing the conical perspective (see section 3) that the images from a camera show. For instance, if the image of a corridor is reconstructed with no rectification, the walls, the floor and the ceiling would appear with a slope, seeming to converge at a point.

In the second section a more detailed description of the problem is given, with special focus on stereo vision and disparity maps. The proposed model is explained in the third section, including several solutions. Experiments are shown in the fourth section and, finally, some conclusions and future works to be developed are presented.

2. PROBLEM DESCRIPTION

Stereo vision techniques are based on the possibility of extracting three-dimensional information from a scene, using two or more images taken from different view points. We will focus on the basic case of two images belonging to a scene. The function of corresponding pixels from each image must be obtained. Let us consider that the camera objectives are parallel, so that the search area for each pixel in the image from the left camera (reference image) is reduced to the same row as the image from the right camera. Every pixel in this row, named epipolar line (figure 1), has a corresponding pixel in the other image, placed in a different column, due to the separation of cameras. The difference between the columns in absolute value is called disparity. The further the object is from the camera, the smaller the disparity, and vice versa.

Figure 1. A stereo par showing an epipolar line (above). Image of disparity (below)

Due to the fact that the disparity is a numerical value, it can be displayed as an image, named depth or disparity image (figure 1), assigning a grey level to every pixel. Bright colours represent high disparities (near objects) and dark ones low disparities (distant objects). There is an inverse relationship between the disparity value and the distance to the object. Depending on the camera geometry, the distance can be transformed to coordinates in an Euclidean space, where the centre is placed in the camera position. [8] [9] [2] [10] [11]

The quality of the three-dimensional reconstruction depends on the quality of the disparity map. Strange objects (or false objects) can cause mistaken shapes and depth values, so errors will be translated to the reconstruction. Another important consideration is the characteristics of the environment where the images are taken. Most techniques are developed to work in indoor environments, such as office spaces or similar. This is due to the fact that the capture systems usually have a limited reach and modelling of the environment is conducted using geometrical primitives (walls, ceilings, floors and, even, guiding marks). All these constraints avoid the problems to be solved in open spaces. Therefore, some unlimited and general solutions to non-structured environments are set out in this paper.

3. PROPOSED MODEL

Three-dimensional reconstruction

The reconstruction is based on a dense disparity image obtained through a process of window correlation (the correspondence between pixels from both images is done using a window correlation criterion, in order to identify similar areas in both images) [10] [11]. This depth image V_{depth} contains the disparity which is associated to each pixel in the reference image (left image). Therefore, for every pixel *[i,j]* in the original image, we can find the disparity value in $V_{depth}[i][j]$. Horizontal and vertical components for each point are directly obtained from the row and the column in which the point is located in the image. [12] [13]

To perform a three-dimensional reconstruction process four basic steps are taken: firstly the disparity image is stored in a two-dimensional matrix, then some smooth filters are applied (average and/or median filters), then the geometrical rectification is done (see next section) and, finally, a three-dimensional matrix is generated to store the spatial occupation.

$$\begin{aligned} &M2D[x][y] = V_{depth}[x][y] \qquad \forall\, x, y \in V_{depth} \\ &M2D = ApplySmoothFilters(\, M2D\,) \\ &M3D = ApplyRectification(\, M2D\,) \\ &\text{Display}(\, \text{ObtainRealUnits}(\, M3D[x][y]\,)\,) \qquad \forall\, x, y \in M3D \end{aligned} \tag{1}$$

where:

- V_{depth}*: disparity matrix*
- *M2D: two-dimensional matrix*
- *M3D: three-dimensional matrix*
- *ApplySmoothFilters(): applies smooth filters on a 2D matrix, and leaves the result in a new 2D matrix.*
- *ApplyRectification(): applies geometrical rectification on a 2D matrix, and returns a 3D matrix containing the result.*
- *Display(): displays a given point on the screen.*
- *ObtainRealUnits(): returns a value in real units (metres) given a value in pixels.*

Geometrical rectification

In order to correct the perspective in the images a rectification is needed. An image taken with a camera is in conical perspective, such that all parallel lines converge at a point (see figure 2). So, to perform the reconstruction correctly the perspective must be rectified.

Figure 2. Conical perspective

A rectification on each point coordinates x and y is performed, maintaining the same value for z. The rectification depends on depth and on coordinates x and y. The higher the value of z, the higher the rectification; the more central the point appears in the image, the higher the rectification. Figure 3 shows the scheme for the rectification process: figure 3(a) shows a non-rectified scene in 3D, in figure 3(b) the scene is seen from the top (only x and z coordinates are shown), and figure 3(c) shows the scene after being rectified.

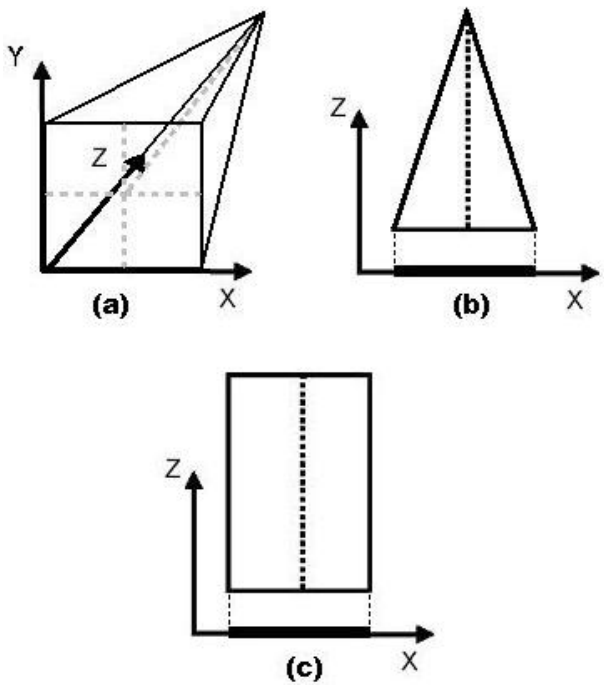


Figure 3: rectification scheme

After the analysis of variables implied in geometrical rectification, we can conclude that the process of rectification depends on depth and difference between x, y and centre coordinates. As a result, the so called **Lineal Rectification** is obtained. The coordinates are lineally corrected, so that the rectification directly depends on grey level (that is, z coordinate) and position (x and y coordinates). In an ideal situation, the Linear Rectification rectifys the scene to obtain the result given in figure 3(c), but in real images some problems arise. The most important drawback is the fact that this method does not distinguish whether the figure is very close, and so some errors occur with certain images, especially if the main object is too far from the camera. In fact, pixels corresponding to a distant object are split, leaving a hole whose dimensions increase as the distance to the object increases. So, important far non-centred objects can have holes and be dispersed. The rectification equation is:

$$newX = x + (V_{depth}[x][y] / kMax) * factorX$$
$$newY = y + (V_{depth}[x][y] / kMax) * factorY$$
$$V_{depth}[newX][newY] = V_{depth}[x][y] \quad (2)$$

where:

- *Vdepth[x][y] contains the grey level corresponding to the pixel of coordinates 'x' and 'y'.*
- *kMax is the maximum value of depth.*
- $factorX = \begin{cases} -x & if\ x < width/2 \\ width - x & if\ x \geq width/2 \end{cases}$

  $factorY = \begin{cases} -y & if\ y < height/2 \\ height - y & if\ y \geq height/2 \end{cases}$

Observe that in the previous equation the new coordinates of a pixel are obtained from the former value of the coordinates, the depth value and the x, y position of the pixel. Variables *factorX* and *factorY* contain the highest displacement that can be performed to place one pixel at the borders of the scene, that is, if the pixel is placed at the centre, the highest displacement is half the image.

The value of $V_{depth}[x][y] / kMax$ is in the range [0, 1], and it depends on the depth: it is 0 if the pixel is in the foreground, and it is 1 if it is in the background (in this case, if the pixel is also in the centre of the image, the rectification is maximum, so the pixel is moved to the border of the image).

In order to improve the rectification and avoid the holes in the main objects, we propose a **Natural Logarithmic Rectification**. In this case, a logarithmic function is applied to the depth value. The logarithm has the property of reducing the rectification when the object is close to the camera, and of magnifying the rectification when the object is far away. So, objects in the background suffer a higher correction than those in the foreground.

$$newX = x + \mathbf{ln}\,(V_{depth}[x][y] / kMax) * factorX$$
$$newY = y + \mathbf{ln}\,(V_{depth}[x][y] / kMax) * factorY$$
$$V_{depth}[newX][newY] = V_{depth}[x][y] \quad (3)$$

Finally, we also propose to use the **Base-10 Logarithmic Rectification**, whose function has a smoother slope than the natural logarithm. In some images, it is a better option due to the fact that small details do not disappear and holes are smaller.

$$newX = x + \mathbf{log_{10}}\,(V_{depth}[x][y] / kMax) * factorX$$
$$newY = y + \mathbf{log_{10}}\,(V_{depth}[x][y] / kMax) * factorY$$
$$V_{depth}[newX][newY] = V_{depth}[x][y] \quad (4)$$

4. EXPERIMENTS

Some experiments have been done to prove the validity of every kind of proposed geometrical rectification. Figure 4 shows four cases: the first image corresponds to the reconstruction of the disparity image of figure 1, without any kind of rectification. The result of linear and both types of logarithmic rectifications are shown in images (b), (c) and (d) of figure 4. In the non-rectified image, the floor shows several steps, corresponding to the conical perspective effect.

(a)

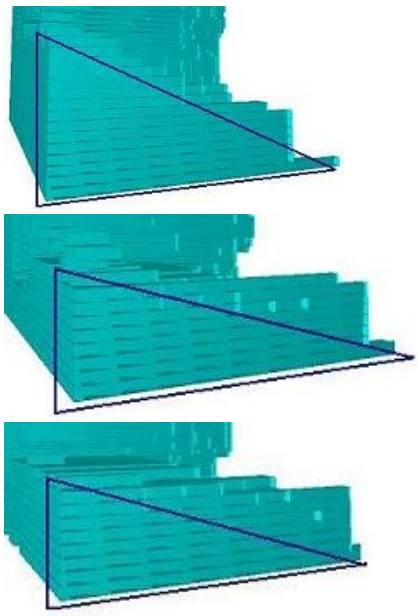
(b)

(c)

(d)

Figure 4. Reconstruction and rectification: (a) no rectification, (b) lineal rectification, (c) natural logarithmic rectification and (d) base-10 logarithmic rectification

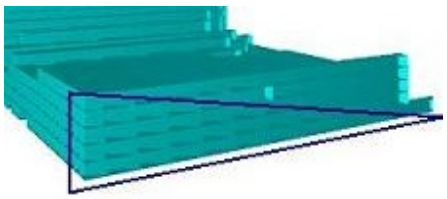
The images in figure 4 show that the rectification reduces the slope of the floor, especially in logarithmic rectifications. However, as the rectification increases, some details tend to disappear. Figure 5 show the effect of rectification on the floor.



(a)

(b)

(c)



(d)

Figure 5. Effect of rectification on the floor: (a) no rectification, (b) lineal rectification, (c) natural logarithmic rectification and (d) base-10 logarithmic rectification

5. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a method of reconstructing three-dimensional scenes from stereo images. The reconstruction is corrected using three kinds of rectification. Although the results show that rectification improves the reconstruction, the final quality of the reconstructed image depends on the quality of the disparity map. In future, we will try to obtain better disparity images that will improve the final result.

It can also be concluded that each type of rectification is suitable for a different kind of images. If the main object in the image is central and close to the observer, the best choice is the logarithmic rectification, to avoid holes in the shapes. However, is the scene is made up of several objects, distributed throughout the image, the best choice is linear rectification, to conserve all details.

We are very interested in obtaining a method to rectify all kinds of images. So, a new line of research, the Continuous Geometrical Rectification, is set out. Due to the discreteness of disparity maps, rectifying each pixel produces holes, as we have already shown. To solve this drawback, we propose treating the disparity maps in a continuous way, so that given two points in the space, the number of points (or pixels, in the map) between them is not finite but infinite. Nevertheless, although from a theoretical point of view the real space is continuous, the representation we use (the disparity map) is a discrete sample of the real space. So a way is needed to fill the "gaps" between given points (the pixels of the map). In order to fill these gaps it is first necessary to detect which the objects in the scene are, so some kind of segmentation is needed. As a first approach, the objects are detected using segmentation by colour and then, starting with the first pixel detected, the object is explored to obtain its silhouette. Only the points in the silhouette are rectified and the holes are interpolated to obtain a filled object. In figure 6, the proposed scheme is shown.

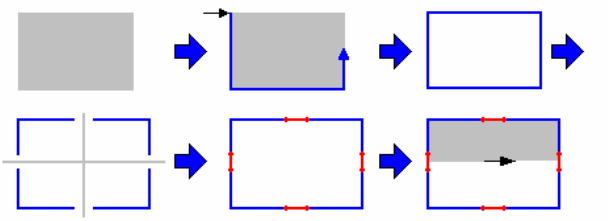


Figure 6. Scheme for continuous rectification

6. ACKNOWLEDGEMENTS

This work has been done with the support of the project "Scene reconstruction and integration of 3D visual information in an augmented reality system (GV04A-730)" given by the "Generalitat Valenciana" (regional government of Valencia, Spain).

7. REFERENCES

[1] Hans P. Moravec. "*Robot spatial perception by stereoscopic vision and 3D evidence grids*". The Robotics Institute Carnegie Mellon University. Pittsburgh, Pennsylvania, 1996.

[2] Amador González, J.A. "*Adquisición y procesamiento de imágenes estereoscópicas y modelado de mundos 3D para su implementación en exploración de ambientes*". Tesis. Universidad de las Américas-Puebla. 2004.

[3] Camillo J. Taylor. "*Surface Reconstruction from Feature Based Stereo*". Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003), pp. 184-190.

[4] Bastian Goldlücke, Marcus A. Magnor. "*Joint 3D-Reconstruction and Background Separation in Multiple Views using Graph Cuts*". Proceedings of CVPR 2003, pp. 683-694, IEEE Computer Society, Madison, USA, June 2003.

[5] K.M. Cheung, T. Kanade, J. Bouguet, and M. Holler. "*A real time system for robust 3D voxel reconstruction of human motions*". Proceedings of the 2000 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '00), Vol. 2, June, 2000, pp. 714 - 720.

[6] Richard Lengagne, Pascal Fua and Olivier Monga. "*3D Stereo Reconstruction of Human Faces driven by Differential Constraints*". Image and Vision Computing 18, pp.337-343, 2000.

[7] M. Li, H. Schirmacher, M. Magnor, and H.P. Seidel. "*Combining stereo and visual hull information for on-line reconstruction and rendering of dynamic scenes"*. In IEEE Workshop on MMSP, pages 9--12, 2002.

[8] E. Trucco and A. Verri. "*Introductory techniques for 3-D Computer Vision*". Prentice Hall, 1998

[9] I. Cox, S. Ignoran, and S. Rao. "*A maximum lilelihood stereo algorithm*". Computer Vision and Image Understanding, 63, 1996.

[10] Compañ Rosique, Patricia. "*Estimación de la disparidad en visión estereoscópica. Tratamiento del color*". Tesis doctoral. Universidad de Alicante. 2004

[11] Satorre Cuerda, Rosana. "*Visión estéreo, multirresolución y modelo integrado*": tesis doctoral. Universidad de Alicante. 2002

[12] M. Pollefeys, R. Koch and L. Van Gool. "*A simple and efficient rectification method for general motion*". Proc. International Conference on Computer Vision, pp.496-501, Corfu (Greece), 1999.

[13] M. Pollefeys. "*3D modelling from images*". Tutorial on conjunction with ECCV 2000, 2000.