MULTI-VIEWPOINT SYNTHESIS FROM UNCALIBRATED STEREO CAMERAS

Marcelo M. Perez and Carla L. Pagliari

Department of Electrical Engineering – Military Institute of Engineering, Brazil

ABSTRACT

View synthesis from arbitrary viewpoints is an important component of virtual reality. This work addresses the problem of synthesizing a novel view from an arbitrary position given two images only (stereo pair). The proposed algorithm creates new viewpoints allowing six degrees of freedom to the virtual camera, i.e., translations along the three coordinate axis, pan, tilt and rotation around its own optical axis. The method generates new views for a totally non-calibrated stereo system. The novelty of the method relies in the placement of the virtual camera and its effect on the view synthesis. The synthesized views showed good results when compared with views produced by well-known viewpoint synthesis methods.

Index Terms— stereo, view synthesis, virtual view

1. INTRODUCTION

View synthesis, intermediate-view synthesis, viewpoint synthesis, multi-viewpoint synthesis, virtual view generation or view transfer treats visual reconstruction as a point-to-point image mapping that creates new views of a scene given two or more real images of it. The novel viewpoints can be useful in practical applications such as area surveillance, 3-D video conferencing with limited network bandwidth, interactive services and 3-D television, among others.

Several papers present different approaches on generating intermediate views from two or more cameras. One class of method needs 3-D scene information, while other class requires pixel correspondence. In [1] a fast method uses disparity gradient, Delaunay triangulation of the stereo disparities and image warping (with the corresponding disparities) information to generate intermediate-views. Fan and Ngan [2] proposed a coding method of disparity map based on adaptive triangular surface modeling where the algorithm finds a smooth disparity map using block-based hierarchical disparity estimation and models the acquired disparity map by Delaunay triangulation on a set of nodes. In [3], a compression method of multi-view image sequences using a generalized quadtree is proposed. Wang and Wang [4] proposed mesh-based analysis and coding of multi-view video sequences.

We present a method that generates natural synthesized views. Using an efficient mathematical formulation, the various movements of a virtual camera are concatenated. For maximum realism, six degrees of freedom are allowed for the virtual camera, and the method was designed for uncalibrated systems. Very realistic results are obtained even for large movements of the virtual camera.

2. GENERATING MULTI-VIEWPOINT IMAGES

The proposed method intends to offer as much realism as can be obtained from a stereo image pair. Of course, the maximum of information that can be included in the new views to be synthesized is limited by the amount of information initially contained in the two starting stereo views.

In the remainder of this Section, the arbitrary final position of the virtual camera is established using six extrinsic parameters, and the resulting pixel mapping for the generated virtual view is developed.

2.1. Coordinate systems and the imaging process

In the following development, an approach is adopted in order to deduce the complete equations for the virtual camera movements in a way that simplify the concatenation of these movements. Initially, the 3-D spatial coordinate system and the 2-D camera image coordinate system are positioned with a common origin and coincident horizontal and vertical axis. The used coordinate axis orientations are presented in Figure 1.

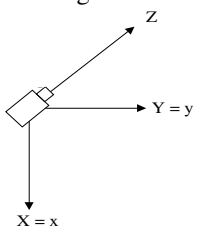

Figure 1. Coordinate axis orientations.

Although the orientations used in this work do not obey the classical canonical convention, they are more in accordance with the directions observed in the imaging process. The first coordinate, 'x' (in the image plane) or 'X' (in the 3-D space), is oriented in the direction of increase of the image scanline numbers. The same occurs with the 'y' or 'Y' coordinate, oriented according to the image columns numbering. The depth or 'Z' coordinate, that obviously only exists in the 3-D reference system, is also oriented in the usual way, with depth increasing as the imaged point moves away from the camera. Also, the origins of both coordinate systems are coincident with the camera principal point, which is defined as the interception of the camera optical axis with its image plane. Consequently, the common origin of the 2-D and 3-D coordinate systems will be placed in the center of the camera image plane.

Already including the effects of the image uninversion performed by the camera in relation to the space-plane projection that occurs in the 'pin-hole' model (assumed for the camera), the relations between the 2-D image plane coordinates (x, y) and the

3-D spatial coordinates (X, Y, Z) for a general scene point P become [5]:

$$x = -\frac{fX}{f - Z} \quad (1)$$

$$y = -\frac{fY}{f - Z} \quad (2)$$

where 'f' is the focal length of the camera.

In this work, it is assumed that the scene is imaged by a parallel stereo camera system, with the left camera placed at the origin of the coordinate systems and, as usual, the baseline B measured along the horizontal direction. It is important to remember that, in the particular case of a parallel stereo system, corresponding epipolar lines coincide with two homologous scanlines on the left and right image planes. In this case, the stereo disparity associated with any spatial point in the 3-D scene has only its horizontal component, thus reducing to a scalar value since its vertical component is zero. Then, from Equation (2), we can obtain the left-view based disparity (here named as 'd' for simplicity) for the point in space as:

$$d = y_r - y_l = -\frac{f(Y - B)}{f - Z} - \left(-\frac{fY}{f - Z}\right) = \frac{fB}{f - Z} \quad (3)$$

2.2. Virtual camera movements

The virtual camera that defines the new viewpoints to be generated can undergo translations along the three main axes. Additionally, it can also suffer rotation around the vertical axis and along the horizontal plane (pan), around the horizontal axis and along the vertical plane (tilt) and around its own camera optical axis (with this movement here being called simply 'rotation'). Figure 2 shows the six kinds of movements that can be applied to the 'virtual' camera.

The initial reference position adopted for the virtual camera will be the position of the left camera of the stereo system used to acquire the stereo image pair being used. So, the set of translations and rotations in space specified for the virtual camera takes the virtual camera from the left camera position to its final position in space. Nevertheless, the same results would be obtained by adopting the position of the right camera of the stereo pair as being the reference, and inverting the direction of the horizontal axis.

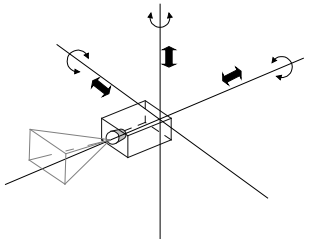

Figure 2. The six degrees of freedom of the virtual camera.

A very important point is that this method is developed for uncalibrated systems, to make it more general. If the stereo system parameters are known, then it is possible to determine the precise spatial coordinates of each point for each new generated viewpoint. However, the generated views themselves are exactly the same for both calibrated and non-calibrated case. As the method is based on uncalibrated stereo systems, the translational displacements are specified as percentages of reference units derived from the stereo system parameters. The rotational movements, on the other hand, can be directly given in degrees or any other angular units.

Initially, the left camera of the stereo system (assumed as the reference position for the virtual camera) and the virtual camera are considered coincident in space. The strategy is to determine, for each image point, its 3-D position from its left image coordinates and left-view disparity. It is important to stress that as every pixel in the image has to be mapped to a new position, the disparity map needs to be sufficiently accurate in order to avoid artifacts. The disparity maps used in this work were obtained using the method presented in [6]. Then, for each virtual camera movement, it is considered that all the virtual cameras and the 2-D and 3-D coordinate systems remain static, and that only the point in space undergoes the opposite movement in space than that experimented by the virtual camera. Knowing the 'new' 3-D position of the point, its new coordinates in the virtual camera image plane can be obtained using Equations (1) to (3) that characterize the imaging process.

Let's consider that a generic point P in space has initial 3-D space coordinates (X_0, Y_0, Z_0), 2-D left stereo view image coordinates (x_0, y_0) and left view disparity d_0. The next step is to concatenate the six movements of the virtual camera, with each new movement being added to the output of the previous (and supposing, as said, that the virtual camera remained static and that the point in space experimented just the opposite displacement than that of the virtual camera).

- HORIZONTAL TRANSLATION ($l_y=k_yB$): as the stereo system is considered uncalibrated, the horizontal translation l_y along the Y axis is expressed as a percentage k_y of the baseline length B.

-VERTICAL TRANSLATION ($l_x=k_xB$): as in the case of the horizontal translation, the vertical translation l_x also needs to be specified as a percentage of the baseline length B.

- ZOOM ($l_z=k_zfB$): here it is called simply 'zoom', just for abbreviation, the virtual camera translational movement along the 'Z' axis, i.e., along the coordinate axis associated with the depth information. In order for the final equations to assume adequate forms, this 'zoom' displacement l_z is given here as a percentage of the product between the stereo cameras focal length 'f' and the stereo system baseline 'B'.

Hence, after the virtual camera translations, the 3-D coordinates (X_1, Y_1, Z_1) of the point in space become:

$$X_1 = X_0 - k_xB \ , \quad Y_1 = Y_0 - k_yB \ , \quad Z_1 = Z_0 - k_zfB \quad (4)$$

Next, for the deduction of the point's new 3-D coordinates after each one of the rotations applied to the virtual camera, it is necessary the use of the trigonometric expressions for the rotation of 2-D coordinate systems.

- PAN (ϕ): the pan movement is a rotation around the vertical axis and along the horizontal plane. The movement is quantified in angular units, with degrees being used in this work. In order to be consistent with the orientations of Figure 1, the positive direction for the pan rotation will be leftward. After the pan movement, the coordinates of the point in space turn out to be:

$$X_2 = X_1$$
$$Y_2 = Y_1 \cos\phi + Z_1 \sin\phi \quad (5)$$
$$Z_2 = -Y_1 \sin\phi + Z_1 \cos\phi$$

- TILT (θ): tilt is the camera angular rotation around the horizontal coordinate axis and along the vertical plane. For consistency with the orientations of Figure 1, the tilt positive direction corresponds to an upward rotation. So, after the tilt, the 3-D coordinates of the spatial point becomes:

$$X_3 = X_2 \cos\theta + Z_2 \sin\theta = X_1 \cos\theta - \\ - Y_1 \sin\phi \sin\theta + Z_1 \cos\phi \sin\theta$$
$$Y_3 = Y_2 = Y_1 \cos\phi + Z_1 \sin\phi \quad (6)$$
$$Z_3 = -X_2 \sin\theta + Z_2 \cos\theta = -X_1 \sin\theta - \\ - Y_1 \sin\phi \cos\theta + Z_1 \cos\phi \cos\theta$$

- ROTATION (φ): in this work, it is called simply 'rotation' the rotation movement of the virtual camera around its own optical axis. For consistency with the orientations adopted here, the positive direction for the camera rotation will correspond to an anti-clockwise movement. This way, the new 3-D position of the point in space after the 'rotation' movement finally becomes:

$$X_4 = X_3 \cos\varphi + Y_3 \sin\varphi = X_1 \cos\theta \cos\varphi - \\ - Y_1 \sin\phi \sin\theta \cos\varphi + Z_1 \cos\phi \sin\theta \cos\varphi + \\ + Y_1 \cos\phi \sin\varphi + Z_1 \sin\phi \sin\varphi$$
$$Y_4 = -X_3 \sin\varphi + Y_3 \cos\varphi = -X_1 \cos\theta \sin\varphi + \\ + Y_1 \sin\phi \sin\theta \sin\varphi - Z_1 \cos\phi \sin\theta \sin\varphi + \\ + Y_1 \cos\phi \cos\varphi + Z_1 \sin\phi \cos\varphi \quad (7)$$
$$Z_4 = Z_3 = -X_1 \sin\theta - Y_1 \sin\phi \cos\theta + Z_1 \cos\phi \cos\theta$$

Suppose a point in space with coordinates (X_i, Y_i, Z_i) projects over the left-view with coordinates (x_i, y_i) and has a left-view disparity d_i. Assuming the generally valid condition of magnification ratio sufficiently large, i.e., considering that $Z_i/f >> 1$, from Equations (1), (2) and (3) we have the following approximations:

$$x_i \approx \frac{fX_i}{Z_i} \quad y_i \approx \frac{fY_i}{Z_i} \quad Z_i \approx -\frac{fB}{d_i} \quad (8)$$

It should be remembered that, while both spatial and image plane coordinates were centered over the principal points in the image plane, the usual notation is to reference image lines and columns starting from the superior left corner of the image. Therefore, substituting X_4, Y_4 and Z_4 into Equations (1) and (2), using the approximate relations of Equation (8) and remembering that X_0, Y_0, Z_0, x_0, y_0 and d_0 satisfy Equations (1) to (3), it follows that (where L and C represent the total number of lines and columns of the image):

$$\left(\hat{x} - \frac{L}{2}\right) = f\frac{N_1}{D} \quad , \quad \left(\hat{y} - \frac{C}{2}\right) = f\frac{N_2}{D} \quad (9)$$

$$\hat{Z} = \frac{Z_4}{B} = \left(\frac{x_0 - L/2}{d_0} + k_x\right)\sin\theta + \left(\frac{y_0 - C/2}{d_0} + k_y\right)\sin\phi\cos\theta - \\ \left(\frac{1}{d_0} + k_z\right) f \cos\phi\cos\theta \quad (10)$$

where $\hat{x}$ and $\hat{y}$ are the new positions in the virtual view and

$$N_1 = \frac{(x_0 - L/2) + k_x d_0}{f(1 + k_z d_0)}\cos\theta\cos\varphi + \frac{(y_0 - C/2) + k_y d_0}{f(1 + k_z d_0)}(\cos\phi\sin\varphi - \\ - \sin\phi\sin\theta\cos\varphi) + (\cos\phi\sin\theta\cos\varphi + \sin\phi\sin\varphi)$$
$$N_2 = -\frac{(x_0 - L/2) + k_x d_0}{f(1 + k_z d_0)}\cos\theta\sin\varphi + \\ + \frac{(y_0 - C/2) + k_y d_0}{f(1 + k_z d_0)}(\sin\phi\sin\theta\sin\varphi + \cos\phi\cos\varphi) + \\ + (\sin\phi\cos\varphi - \cos\phi\sin\theta\sin\varphi) \quad (11)$$
$$D = -\frac{(x_0 - L/2) + k_x d_0}{f(1 + k_z d_0)}\sin\theta - \frac{(y_0 - C/2) + k_y d_0}{f(1 + k_z d_0)}\sin\phi\cos\theta + \\ + \cos\phi\cos\theta$$

So, the procedure for generation of the new viewpoint is:

_ Determine the parameters $k_x, k_y, k_z, \phi, \theta$ and φ that define the new viewpoint in relation to the left stereo view.

_ For each point (x_0, y_0) with disparity d_0 in the left-view, compute its new position ($\hat{x}$, $\hat{y}$) in the virtual view and its new 'normalized depth' (true depth divided by B) $\hat{Z}$ using Equations (9) to (11).

_ If another point of the initial left-view is mapped to a new position already taken, this point will only replace the point already occupied if its normalized depth is smaller than that of the occupant point (visibility test).

_ After all the points of the original left-view have been mapped to their new positions, the remaining unfilled positions must be interpolated (in this work, a simple linear interpolation was used).

If pan and tilt movements are performed, an estimate of the focal length of the cameras is needed, and that's why the term 'f' is included in Equations (9) to (11). From geometrical considerations over the cameras' fields of view, this work uses half the number of image columns as a rough estimate of the focal length, if a more precise guess for the particular stereo camera system under consideration is not available.

3. RESULTS

In this section we show selected results of virtual camera movements and perform a comparison of the intermediate multi-viewpoint synthesis method proposed in this article with other well-known methods. The quantitative evaluations are conducted using the PSNR (peak signal to noise ratio) as the objective error measure. Is important to emphasize that we use two stereo views only to synthesize all the virtual views.

Figure 4 depicts the first left and right frames of stereo sequence MAN [7]. Figure 5 pictures the stereo pair CORRIDOR [8].

Figure 6(a) portrays a synthesized view of stereo pair MAN with pan (ϕ = -10.0°), and tilt (θ= -5.0 °) movements. The new synthesized pixels look natural and considering that the strip in the left side of the view belongs to the background area (which has no disparity associated), the view is perfectly created.

Figure 6(b) depicts the central view (k_y=0.5) of stereo pair MAN. One minor error appears in the junction of the background with the left shoulder. As the background possesses no disparity, the algorithm has to infer a disparity value to the background. Consequently, the transition from the background to the man is not perfect. Even so, the central view is truthfully generated.

Figure 4. MAN stereo pair.

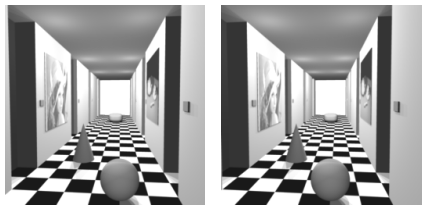

Figure 5. CORRIDOR stereo pair.

Figure 6. MAN. (a) virtual view, with pan and tilt; (b) synthesized central view.

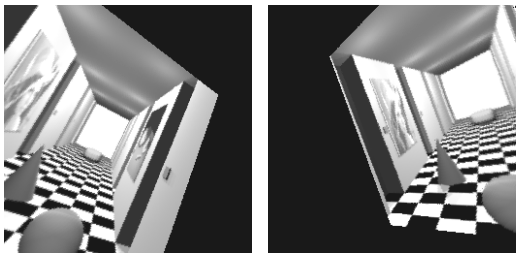

Figure 7. CORRIDOR virtual views. (a) with translation, pan, tilt, and rotation; (b). with translation, pan, and rotation

Figure 7(a) displays the synthetic view of image CORRIDOR after a horizontal translation (k_y=0.5), pan (θ = -15.0 ^{o}), tilt (ϕ = 10.0 ^{o}), and rotation (φ = 25.0 ^{o}). Figure 7(b) pictures other view of image CORRIDOR with a horizontal translation along the baseline of k_y=0.5, with a pan movement of θ = 30.0^o plus a rotation of φ = 20.0 ^{o}.

As in video coding, we use the PNSR measure to evaluate the quality of the reconstructed view. Table 1 presents the PNSR values between the original and synthesized right-views of frame 0 of stereo sequences MAN and AQUA [9], and stereo image CORRIDOR. The intermediate-view algorithm proposed in this work is used combined with the disparity estimator presented in [6]. Furthermore, Table 1 displays PNSR values obtained by intermediate view methods proposed in other four well-known works [1, 2, 3, 4].

The values presented in Table 1 show that the multi-viewpoint method proposed in this work in combination with the disparity estimator proposed in [6] shows the best PSNR values among all the other methods considered.

Table 1. PSNR between original and reconstructed views.

STEREO PAIR	METHOD				
	Proposed	[3]	[2]	[4]	[1]
	PSNR (dB)				
AQUA	34.1	24.3*	23.5*	24.0*	23.7*
CORRIDOR	40.4	24.5*	31.9*	31.0*	31.9*
MAN	34.1	N/A	N/A	32.0	N/A

*The PNSR values have been taken from the PSNR curves of Figures 16 and 17 from [1]. Hence, their values are not exact since they have to be interpolated.

4. CONCLUSIONS

The proposed method generates flexible virtual views of stereo pairs from arbitrarily defined new viewpoints, being able to cope with large virtual camera movements as well as their combination. Section 3 displays selected virtual views of stereo pairs, created from arbitrarily viewpoints, which look natural, revealing few hardly noticeable distortions.

The presented multi-viewpoint image synthesis algorithm provides a novel approach that only requires the disparity maps and the compulsory original image pairs. Experimental results were presented and verified using both objective (PSNR) and subjective comparisons. The quantitative results displayed in Table 1, as well as the subjective analysis of the pictures displayed in Section 3 suggest that the proposed method produces good quality virtual views when compared to other well-known methods [1,2,3,4].

5. ACKNOWLEDGMENTS

The authors would like to thank the Brazilian Funding Agency FINEP under contract no. 2645/06 FINEP/FAPEB, the Army Technological Center, and the Military Institute of Engineering-Brazil for their financial support.

6. REFERENCES

[1] J. H. Park and H. W. Park, "Fast View Interpolation of Stereo Images Using Image Gradient and Disparity Triangulation", *Signal Processing: Image Communication*, vol. 18, pp. 401-416, 2003.

[2]. H. Fan, K. N. Ngan, "Disparity Map Coding Based on Adaptive Triangular Surface Modeling", *Signal Processing: Image Communication*, vol. 14,no. 2, pp. 119-130, 1998.

[3]. S. Sethuraman, Stereoscopic Image Sequence Compression using Multiresolution and Quadtree Decomposition Based Disparity – and Motion-adaptive Segmentation, Ph.D. Dissertation, Carnegie Mellon University, 1996.

[4]. R.-S. Wang, Y. Wang, "Multiview Video Sequence Analysis, Compression, and Virtual Viewpoint Synthesis", *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 10, no. 3, pp. 397-410, 2000.

[5] R. J. Schalkoff, *Digital Image Processing and Computer Vision*, John Willey & Sons, 1989.

[6] M. M. Perez, "Stereo-Based Image Interpretation Using Cooperative Disparity and Segmentation Analysis", Ph.D. Dissertation, University of Essex, U.K., 2004.

[7] Heinrich-Hertz-Institut (HHI), Berlin, Germany.

[8] http://www-dbv.informatik.uni-bonn.de/stereo_data.

[9] CCETT (test sequences shot and distributed under RACE DISTIMA European Project), France.