A Real Time 2D to 3D Image Conversion Techniques
Miroslav Nikolov Galabov

Abstract—This article describes methods developed for 2D-3D conversion of images based on motion parallax, depth cues in still pictures and gray shade and luminance setting for multiview autostereoscopic displays. Detailed exposed a new 2D-to-3D image conversion technique with the modified time difference (MTD) and the computed image depth (CID) realizes to convert any type of visual resources into the 3D images. Proposed a method for conversion from 2D to 3D based on gray scale and luminance setting, which does not require a complex motion analysis.

INDEX TERMS— 2D TO 3D IMAGE CONVERSION, MOTION PARALLAX, IMAGE DEPTH, STEREOSCOPIC IMAGE.

I. INTRODUCTION

Depending on the number of input images, we can categorize the existing 2D to 3D conversion algorithms into two groups: algorithms based on two or more images and algorithms based on a single still image. In the first case, the two or more input images could be taken either by multiple fixed cameras located at different viewing angles or by a single camera with moving objects in the scenes. We call the depth cues used by the first group the multi-ocular depth cues. The second group of depth cues operates on a single still image, and they are referred to as the monocular depth cues. According to the depth cues on which the algorithms reply, the algorithms are classified into the following 12 categories: binocular disparity [1,2,32,42], motion [1,7,8,30,36], defocus [3,9,10], focus [4], silhouette [5], atmosphere scattering [11], shading [12], linear perspective [6], patterned texture [13], symmetric patterns [14], occlusion (curvature, simple transform)[15] and statistical patterns [6].

The conversion of 2D content into 3D content involves creating missing information [33,35,37,40,43]. The process involves an automatic aspect, where parallax is created from other depth cues present in the scene, and an aspect carried out by human operators, adding a creative dimension to the procedure.

Methods developed for 2D–3D conversion may also be used for parallax correction in existing, but unsatisfactory, stereoscopic content.

While the domain has been explored in detail, the generation of a depth map from a single image is a problem, which presents an infinite number of solutions, and the proposed methods cannot, therefore, claim to offer universally acceptable solutions [24,25,26,27,34,38,39,41].

II. CONVERSION OF 2D TO 3D IMAGES BASED ON MOTION PARALLAX

The relative motion between the viewing camera and the observed scene provides an important cue to depth perception: near objects move faster across the retina than far objects do. The extraction of 3D structures and the camera motion from image sequences is termed as structure from motion. The motion may be seen as a form of “disparity over time”, represented by the concept of motion field. The motion field is the 2D velocity vectors of the image points, induced by the relative motion between the viewing camera and the observed scene. The basic assumptions for structure-from-motion are that the objects do not deform and their movements are linear. Suppose that there is only one rigid relative motion, denoted by $V$, between the camera and scenes. Let $P = [X,Y,Z]^T$ be a 3D point in the conventional camera reference frame. The relative motion $V$ between $P$ and the camera can be described as:

$$V = -T - \omega \times P,$$

(1)

Where $T$ and $\omega$ are the translational velocity vector and the angular velocity of the camera respectively. The connection between the depth of 3D points and its 2D motion field is incorporated in the basic equations of the motion field, which combines equation (1) and the knowledge of perspective projection:
\[
\nu_x = \frac{T_x - T_s f}{Z} - \omega_x f + \omega_z y + \frac{\omega_{xy}}{f} - \frac{\omega_x x^2}{f}, \quad (2)
\]
\[
\nu_y = \frac{T_y - T_s f}{Z} + \omega_y f - \omega_z y - \frac{\omega_{xy}}{f} + \frac{\omega_y y^2}{f}, \quad (3)
\]
where \( \nu_x \) and \( \nu_y \) are the components of motion field in the x and y direction respectively; \( Z \) is the depth of the corresponding 3D point; and the subscripts \( x \), \( y \) and \( z \) indicate the component of the x-axis, y-axis and z-axis directions. In order to solve this basic equation for depth values, various constraints and simplifications have been developed to lower the degree of freedom of the equation, which leads to the different algorithms for depth estimation, each suitable for solving problem in a specific domain. Some of them compute the motion field explicitly before recovering the depth information; others estimate the 3D structure directly with motion field integrated in the estimation process.

It is worth to note that the sufficiently small average spatial disparity of corresponding points in consecutive frames is beneficial to the stability and robustness for the 3D reconstruction from the time integration of long sequences of frames. On the other hand, when the average disparity between frames is large, the depth reconstruction can be done in a way as that of binocular disparity. The motion field becomes equal to the stereo disparity map only if the spatial and temporal variances between frames are sufficiently small.

We will present some approaches to extract disparity information from a 2D image and use it for the construction of a 3D image. The description of these approaches is intended to familiarize us with physiological depth cues, such as, for example, cues based on the Pulfrich effect\[16\]. This effect is associated with motion parallax.

\[\text{Fig. 1. Determination of the left and right eye images from a 2D object moving to the right.}\]

Five temporal sequences shows a bird flying to the right in front of mountains as the original images and, above, the same images delayed by two time slots (Figure 1). The original image in time slot 4 is chosen as the left eye image and the delayed image in time slot 2 as the right eye image as depicted below. The eyes are rotated until their axes intersect at the present location of the bird. So the locations of the bird provide a sensation of depth. However, this is an illusionary depth because the speed of the bird has no relation at all to its depth. This is further elucidated by the next observation. If the bird flies slower it would be located further to left in Figure 1 as indicated by the dashed line from the left eye, while the starting position for the right eye remains the same. In this case the intersection of the axes of the eyes is of course further to the left but also higher up closer to the mountains. This indicates a larger depth even though the bird has the same depth as before. This again is an illusionary depth, which we have to cope with in the next section. In the present case it requires a correction that we do not have to deal with now. This method of depth generation \[18\] is based on a so-called modified time difference (MTD).
If the object, such as the car in Figure 2, moves in the opposite direction to the left, the axis of the left eye is directed toward the earlier position of the car, while the axis of the right eye follows the car to its later position. This is the reverse of the movement to the right. Also, here a correction according to the speed of the car has to be done.

Fig. 2. Determination of the left and right eye images from a 2D object moving to the left.

The above described activities of the eyes serve only to explain the construction of the left and right eye images for the successful generation of 3D images. It is not assumed that the eyes react that way in reality.

Fig. 3. Block diagram for the 2D/3D conversion according to the MTD process.

Signal processing for the MTD process is shown in Figure 3. The ADC provides the digital form of the analog signal, which is again converted back to analog form by the DAC at the output. The movement detector provides the direction and the speed of the movement, whereas the delay time controller provides the speed-dependent correction of the depth. The delay direction controller guides the starting position to the right eye for a movement to the right and to the left eye for a movement to the left.

III. CONVERSION FROM 2D TO 3D BASED ON DEPTH CUES IN STILL PICTURES

The MTD method works only for moving objects. For still images it has to include a disparity extraction based on contrast, sharpness, and chrominance. Contrast and sharpness are associated with luminance. Sharpness correlates with high spatial frequencies, while contrast is related to medium spatial frequencies. Chrominance is associated with the hue and the tint of the color. The approach based on these features is called the computed image depth (CID) method [19,20,31].

CID is proposed for converting from still 2D images into 3D images. When we watch a 2D picture, we generally recognize the far-and-near positional relationship between the objects in the picture by some information in it. This information is supposed to be useful for the 2D-to-3D image conversion. So we use the sharpness and the contrast of the input images for computing the far-and-near positional relationship of the objects in the CID.
The CID consists of the following two processes. One is the image depth computation process that computes the image depth parameters with the contrast, the sharpness and the chrominance of the input images.

The other is the 3D image generation process that generates the 3D images according to the image depth parameters. Figure 4 shows the basic principle of the CID.

At first, each sharpness, contrast and chrominance values of the separated areas in the input images is detected respectively. The sharpness means the high frequency element of the luminance signal of the input images. The contrast means the middle frequency element of the luminance signal of the input images. The chrominance means the hue and the tint of the colour signal of the input images.

Furthermore, the adjacent areas that have close color are grouped according to the chrominance values. The image depth computation works to be based on these grouped areas.

Fig.4. Determination process for classification of depth as near–middle–far based on contrast, sharpness, and composition.

The image depth computation process uses the contrast values and the sharpness values. Near objects exhibit a higher contrast and a higher sharpness than objects positioned farther away. So contrast and sharpness are inversely proportional to depth. Adjacent areas exhibit close chrominance values, thus indicating that they have the same depth. Chrominance is a measure for the composition of the 2D image. The features contrast, sharpness, and chrominance allow the depth classification far–mid–near as depicted in Figure 4.

Therefore, these contrast and sharpness values are inversely proportional to the distance from the camera to the objects. If only these values are used for the image depth computation, it often occurs that the center of the images become nearer than both sides, top and bottom of the images. This cause is that the focused object is generally positioned at the center of the images, and the ground or the floor is positioned the bottom of the images that has flat surface generally and few contrast and sharpness values are taken from the bottom areas. So, it is adopted to compensate these values by the image’s composition. The composition has the tendency that the center or the bottom side of the images is nearer than the upper side in the general images. So, each image depth parameter is decided by the average of each area’s sharpness and contrast value that is weighted by the image’s composition. This compensation would be better way to get good 3D effect, but it should be changed according to the applications.

Secondly, the 3D image generation process generates the left and the right eye images according to the image depth parameter of each grouped area. If the parameter of an area indicates near, the left images are made by shifting the input images to the right, and the right images are made by shifting to the left. If the parameter of an area indicates far, both images are made by shifting to each opposite direction. The horizontal shift value of each separated area is proportional to the 3D effect. Furthermore, when the image depth parameters are changed quickly or frequently, the converted images become hard to be watching. Therefore, each shift value is adjusted to decrease the quick
changes of the image depth parameters between the adjacent areas. As a result of these processes, the 3D images that are easy to watch can be generated.

The CID is especially suitable for converting from still images, because it does not need any motions of the objects in the images. Of course, the CID can be also adapted for the images with moving objects.

IV. CONVERSION FROM 2D TO 3D BASED ON GRAY SCALE AND LUMINANCE SETTING
In [21] three attractive and successful features for the determination of depth in 2D images are investigated: namely, gray-scale analysis, relative spatial setting, and multiview 3D rendering. A color image is simply converted into one intensity value $I$ with a gray scale

$$I = \frac{(I_R + I_G + I_B)}{3}, \quad (4)$$

where the right side contains the intensities of the colors. In Figure 5 and in the block diagram in Figure 6 this is called gray-scale conversion. The gray scale $I$ is expanded into $I'$ with a range from 255 to 0 for an 8-bit word by the equation

$$I' = \frac{(I - I_{\text{min}}) \times 255}{(I_{\text{max}} - I_{\text{min}})} \quad (5)$$

This is called the dynamic contrast enhancement, which is followed by a narrowing down of the gray scale to the range 0–63. Figure 5 shows the appearance of the image after these individual steps. In the next step the luminance of the entire image is reset by assigning a smaller luminance to the upper portion which is gradually getting brighter toward the lower portion.

![Fig.5. The gray-scale conversions of a figure.](image)

![Fig.6. Block diagram for gray-scale conversions.](image)

![Fig.7. The pixel arrangement for four different views](image)
After application of the setting, the image with the increasing gray scale toward the bottom conveys a very impressive sensation of depth (even though the reproduction quality of the figure may be low). This reminds us of another depth-enhancing cue in brighter images, which is rendering objects slightly more bluish the farther away they are.

Counteracting this depth enhancement is a spot at any depth reflecting light. This effect induces the sensation of a shorter depth. A 1D median smooth filter [22, 28, 29] is used to suppress this effect. After this filtering the eye looks free of reflection.

The last step is multiview rendering for a presentation through a slanted array of lenticular lenses. The pixel arrangement for four views is also applied in the present case and is shown in Figure 7. The four views are paired into two views according to different depths assigned to each pair as provided by the depth map. For the image on the left in Figure 7 the depth map is shown in Figure 8 on the right with brighter areas for a smaller depth. The four viewing directions are shown in the second line.

IV. CONCLUSION

A single solution to convert the entire class of 2D images to 3D models does not exist. Combing depth cues enhances the accuracy of the results. Most 2D to 3D conversion algorithms for generating stereoscopic images and ad-hoc standards are based on the generation of a depth map. However, a depth map has a disadvantage that it needs to be fairly dense and accurate. Otherwise local deformations in the derived stereo pairs are easy to happen. And it is also helpful to explore the alternatives than to confine ourselves only in the conventional methods based on depth maps.

The new 2D-to-3D image conversion technique with the MTD and the CID realizes to convert any type of visual resources into the 3D images. Proposed method for conversion from 2D to 3D based on gray scale and luminance setting does not require a complex motion analysis. Certain commercial solutions offer fully automated 2D–3D conversion, but the results are generally unsatisfactory, with the exception of very specific cases where the geometry of the scene is subject to strong constraints, movements are linear and predictable and segmentation is simple. Not all content is equally suited to 2D–3D conversion.

ACKNOWLEDGMENT

The presented article is part of research work carriedout in the “Analysis, research and creation of multimedia tools and scenarios for e-learning” project - Contract No: RD - 09-590-12/10.04.2013, which is financially supported by the St. Cyril and St. Methodius University of Veliko Turnovo, Bulgaria.
REFERENCES

Miroslav GALABOV was born in Veliko Turnovo. He received his M.S.E degree in Radio Television Engineering from the Higher Naval School N. Vapcarov, Varna, Bulgaria, in 1989. After that he worked as a design engineer for the Institute of Radio Electronics, Veliko Turnovo. From 1992 to 2001 he was an assistant professor at the Higher Military University, Veliko Turnovo. He received his Ph.D. degree in Automation Systems for Processing of Information and Control from the Higher Military University, in 1999. Since 2002 he has been an assistant professor and from 2005 he has been an associate professor in the Computer Systems and Technologies Department, St. Cyril and St. Methodius University of Veliko Turnovo. He is the author of ten textbooks, and over 40 papers. His current interests are in signal processing, 3D technologies and multimedia.