DEPTH ESTIMATION FROM STEREOSCOPIC 360-DEGREE VIDEO

Krzysztof Wegner, Olgierd Stankiewicz, Tomasz Grajek, Marek Domański

Chair of Multimedia Telecommunications and Microelectronics,
Poznań University of Technology,
Polanka 3, 61-131 Poznań, Poland,
email: kwegner@multimedia.edu.pl

ABSTRACT

In this paper we present a novel method for depth estimation from stereoscopic 360-degree video. As for now, no such methods have been presented in the literature. First we show mathematical foundations for correspondence matching polar coordinates. Then, the attained formulas are simplified to show relationship to depth estimation in rectified pairs of rectangular images. The proposed method has been implemented in Depth Estimation Reference Software (DERS) developed by Motion Picture Experts Group (MPEG) of International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). Finally, we present experimental results in form of depth maps estimated for *"Dancing 360"* stereoscopic omnidirectional sequence.

Index Terms — Depth estimation, 360 degree video, omnidirectional video, circular projection, 360 degree 3D video, stereoscopic omnidirectional video

1. INTRODUCTION

Highly immersive media grow very rapidly nowadays. Among them, virtual reality and omnidirectional video are worth mentioning, that can be watched on head mounted displays (HMD), vastly deployed on the market.

Referring to this area, further research stages are currently under consideration in scientific and standardization committees. One of them is new standardization project launched spring last year by International Organization for Standardization called MPEG-I Immersive aiming to support not only omnidirectional video (360 degree) with 3 degrees of freedom (yaw, pitch and roll), but with 6 degrees of freedom allowing user to move freely. It is anticipated, that in the first generation of such systems, only small movements around the central position will be possible, in order to accommodate head movements.

Figure 1. Circular projection obtained using a rotating pair of cameras.

There are 360 degree cameras or rigs of cameras available on the market for some time, e.g. Nokia Oz or GoPro Odyssey (Fig. 1). They allow recording of omnidirectional video, but noticeably they are also able to produce stereoscopic omnidirectional video called 360-degree 3D video (Fig. 2). When watched on HMD, such a video provide binocular experience with some depth impression.

However, the technology is not perfect. From physical reasons 360-degree 3D video provides the best depth experience along equator of omnidirectional video, and the depth impression vanishes gradually with increasing elevation to finally vanish completely in zenith.

New MPEG activity aims not only at providing some motion parallax but also true full binocular experience everywhere over entire sphere of 360 video. The current state of works show some indication that this can be provided with a technology that has been recently investigated in the context of free viewpoint television: Depth Image Based Rendering (DIBR).

DIBR has been studied extensively for over 20 years now. Depth estimation along with view synthesis allows creation of additional viewpoints of a scene. It means that off-center positioned view (to provide motion parallax), as well as additional binocular view (for true stereoscopic perception) can be generated based on captured image and the depth data.

In particular, stereoscopic videos, where a scene is captured by a pair of parallel cameras, have been thoroughly studied. For such case many depth estimation algorithms are known, some of them are collected by the Middlebury benchmark webpage [1].

In this paper we introduce a 360-degree 3D video model, formulate a depth estimation problem from such a video, and finally show how to apply well known parallel view stereoscopic depth estimation algorithms. As for now, no depth estimation method for 360-degree video has been presented in the literature.

2. CIRCULAR PROJECTION OF 360 DEGREE 3D VIDEO

The 360-degree 3D video can be represented with a circular projection. In such an approach, stereoscopic omnidirectional video is composed of two cylindrical panoramas [2], [3]. An exemplary stereoscopic omnidirectional image in circular projection is presented in Fig. 2.

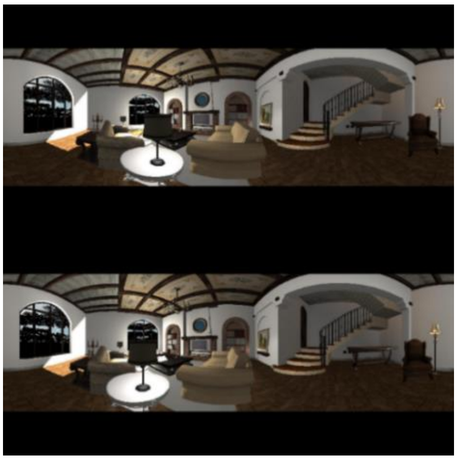


Figure 2. The example of circular projection model (the left view on the top, the right view on bottom) [2].

Each column of each panorama represents a light coming to the camera from a given angular direction, covering entire 360 degree. 360 degree 3D video can be understood (modelled) as a panorama captured by a rotating pair of cameras with a very narrow field of view (Fig. 3).

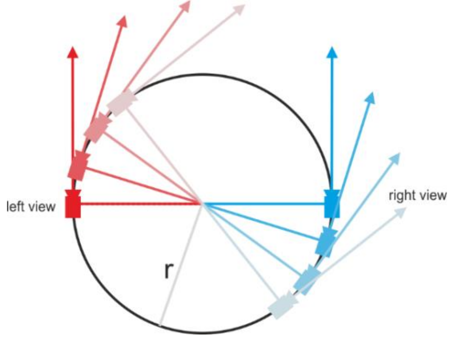


Figure 3. Circular projection obtained using a rotating pair of cameras [3].

In such a concept, each camera from the pair creates a separate circular view. Therefore, two circular views (Fig. 4) are created. Each such circular view is composed of columns shot by the given camera which is rotating and thus the entire panorama is created. It is called cylindrical projection (Fig. 4). Relationship between the direction of the ray of light and pixel position is formally defined by Formula (1),

$$x = R \cdot (\alpha - \alpha_0)\,, \quad y = R \cdot Tan(\beta)\,, \tag{1}$$

where α and β are the azimuth and elevation, or the longitude and latitude, respectively (cf. Figure 4). Please note that the cylindrical projection was discussed ages ago under the name of Mercator projection that is well-known as one of the basic cartographic projections.

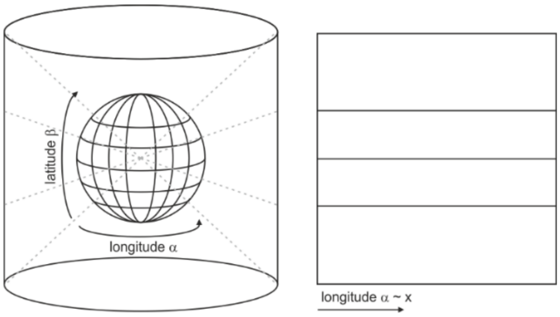


Figure 4. Cylindrical projection principle

3. DEPTH ESTIMATION FROM CIRCULAR PROJECTION

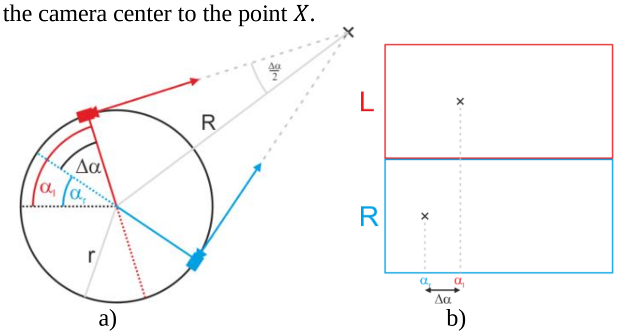
Let us assume that a point X is observed by omnidirectional stereoscopic camera that produces 360 degree 3D video in circular projection format. For simplicity, we start with consideration of a point observed at the equator (Figure 5 b).

3.1 Distance estimation

Point X is visible in the left view at angle α_l and in the right view at angle α_r. Let us denote the difference of position in both views as $\Delta\alpha$. With the use of simple trigonometric relationship we get:

$$Sin\left(\frac{\Delta\alpha}{2}\right) = \frac{r}{R}\ , \tag{2}$$

where r is radius of acquisition system, and R is distance from the camera center to the point X.



a) b)

Figure 5. Considered case of matching in circular projection.

Therefore, based on the position difference $\Delta\alpha = \alpha_l - \alpha_r$ of a given point in the left and in the right view we can get distance R:

$$R = \frac{r}{Sin\left(\frac{\Delta\alpha}{2}\right)}\ . \tag{3}$$

3.2 Pixel position

Position of a point in 360 degree 3D video can be expressed (Fig. 6) as angular position (α, β) but it is more convenient to express it in pixel coordinates (x, y).

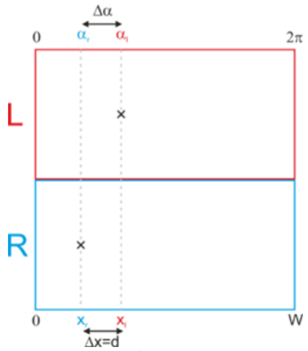


Figure 6. Angular and pixel position of a point in 3D degree 3D video.

If an acquired image has a resolution of W by H (which typically are in the proportion of $\frac{W}{H} = \frac{2}{1}$) then angular position can be expressed as:

$$\alpha = x \cdot \frac{2\pi}{W}. \tag{4}$$

Equation (4) can be used for (3):

$$R = \frac{r}{Sin\left(\frac{\Delta\alpha}{2}\right)} = \frac{r}{Sin\left(\frac{\alpha_l - \alpha_r}{2}\right)} = \frac{r}{Sin\left(\frac{(x_l - y_r)\cdot\frac{2\pi}{W}}{2}\right)} = \frac{r}{Sin\left((x_l - y_r)\cdot\frac{\pi}{W}\right)} \tag{5}$$

where x_l and x_r are pixel coordinates of a considered point in the left and in the right view, respectively. In depth estimation, a difference in pixel coordinates of a given point in the left and the right view is called disparity, and denoted as d. Therefore, (5) can be simplified to:

$$R = \frac{r}{Sin\left(d \cdot \frac{\pi}{W}\right)}\ . \tag{6}$$

3.3 Relationship to depth estimation from rectified pairs of images

Equation (6) is somehow similar to the expression that defines a distance estimated from a rectified video pair (7).

$$Z = \frac{f \cdot b}{d} \quad , \tag{7}$$

where f is the focal length of a camera, b is the baseline of camera pair, and d is the disparity (defined exactly as in this document).

The main difference between (6) and (7) is the sine function in the denominator. This Sin operator in the denominator can be expanded with Taylor series which yields:

$$R = \frac{r}{Sin\left(d \cdot \frac{\pi}{W}\right)} = \frac{r}{d \cdot \frac{\pi}{W} - \frac{\left(d \cdot \frac{\pi}{W}\right)^3}{3!} + \frac{\left(d \cdot \frac{\pi}{W}\right)^5}{5!} + \cdots} \tag{8}$$

Let us assume that $d \cdot \frac{\pi}{W} \ll 1$ is very small, so that we can omit higher order components in Taylor series in (8):

$$R = \frac{r}{Sin\left(d \cdot \frac{\pi}{W}\right)} \approx \frac{r}{d \cdot \frac{\pi}{W}} = \frac{r \cdot \frac{W}{\pi}}{d} \quad . \tag{9}$$

As we can see, the resultant approximated distance R value attained in (9) is the same as (7) but just with different constants. Therefore, it can be stated that for small values of disparity, the calculation (6) of distance R used in depth estimation in circular projection can be approximated with formula used in stereoscopic planar depth estimation (7). Small values of disparity occur in far objects. Therefore, the mentioned approximation can be used directly for objects that are far from the camera. Of course, one cannot control application of depth estimation for far objects prior to determining their distance, therefore the only practical approach is to assume that all objects are far in the analyzed scene.

Therefore, it can be summarized that the presented approximation can be used for scenes that are far from the camera set.

3.4 Far objects

Now we will analyze the case when approximation (9) is an accurate enough approximation of (6).

Let us define an approximation error δR as a difference of distance R calculated from (6) and (9) in proportion to the distance R from (6):

$$\delta R = \frac{\frac{r}{Sin\left(d \cdot \frac{\pi}{W}\right)} - \frac{r}{d \cdot \frac{\pi}{W}}}{\frac{r}{Sin\left(d \cdot \frac{\pi}{W}\right)}} = 1 - \frac{Sin\left(d \cdot \frac{\pi}{W}\right)}{d \cdot \frac{\pi}{W}} \quad . \tag{10}$$

Now, (10) expresses relative estimation error δR as a function of the observed disparity. Even more interesting is an expression of relative estimation error δR in a function of the distance R itself. For (6) we can derive formula for disparity d as:

$$d = \frac{W}{\pi} \cdot ArcSin\left(\frac{r}{R}\right) \quad . \tag{11}$$

After substitution of (11) to (10) and simplification we get:

$$\delta R = 1 - \frac{Sin\left(d \cdot \frac{\pi}{W}\right)}{d \cdot \frac{\pi}{W}} = 1 - \frac{\frac{r}{R}}{ArcSin\left(\frac{r}{R}\right)} \quad , \tag{12}$$

which is a direct formula for relative estimation error δR as a function of the distance R. It is presented in Fig. 7.

As it can be seen from Figure 6, relative estimation error δR decreases very rapidly with increasing distance R. At the distance $R > 4 \cdot r$ relative error is below 1%. I.e. for a 10 cm radius 360 degree 3D camera, objects that are further than 40 cm away can be considered to be far enough. Condition of $4 \cdot r$ is not very strong requirement and it can be easily fulfilled for most of the omnidirectional cameras used nowadays.

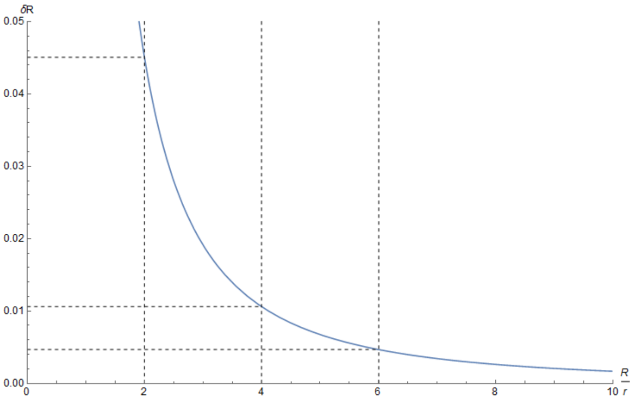


Figure 7. Relative estimation error δR in function of $\frac{R}{r}$.

3.5 Understanding camera parameters used in approximation

As it was mentioned, the equation (9) has similar form to (7) but is expressed with different constants. Now we will try to understand relation of these constants and thus the effective baseline b and focal length f of an approximation (9).

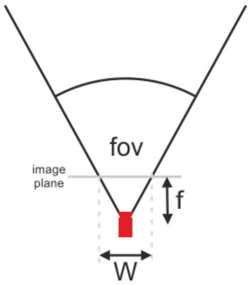


Figure 8. Relationship between focal length, field of view angle and resolution of the camera.

First let us analyze the remaining basic relationship between focal length f (expressed in pixel units), field of view angle fov and horizontal resolution W of the camera (Figure 8):

$$Tan\left(\frac{fov}{2}\right) = \frac{W/2}{f} \quad . \tag{13}$$

From (13) we can obtain focal length expressed as field of view angle:

$$f = \frac{W/2}{Tan\left(\frac{fov}{2}\right)} \quad . \tag{14}$$

By expanding it with Taylor series we get:

$$f = \frac{\frac{W}{2}}{Tan\left(\frac{fov}{2}\right)} = \frac{\frac{W}{2}}{\frac{fov}{2} + \frac{1}{3}\left(\frac{fov}{2}\right)^3 + \cdots} \approx \frac{\frac{W}{2}}{\frac{fov}{2}} = \frac{W}{fov} \quad . \tag{15}$$

For an omnidirectional camera $fov = 2\pi$, and baseline between pair of images is $b = 2r$. We can substitute this result into (7) and get:

$$Z = \frac{f \cdot b}{d} = \frac{\frac{W}{2\pi} \cdot 2r}{d} = \frac{\frac{W}{\pi} \cdot r}{d} = R \ , \qquad (16)$$

which gives equation for our derived approximation (9).

We conclude that camera parameters used for the approximation would be:

$$f = \frac{W}{2\pi} \quad b = 2 \cdot r \ . \qquad (17)$$

For far objects, this conclusion appears well-understood. Such camera parameters can be used in any existing stereo-pair-based depth estimation algorithms.

4. EXEMPLARY RESULTS

As it has been shown, the proposed method can be applied to any stereo-pair-based depth estimation software. One of such algorithms, widely used in literature, is Depth Estimation Reference Software (DERS) [4] developed by Motion Picture Experts Group (MPEG) of International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) during works on free viewpoint television (FTV). We have implemented the proposed method in DERS in order to allow experimental verification.

As a test material we have used *“Dancing360”* stereoscopic omnidirectional sequence [5] (Fig 9). The attained results of depth estimation are presented in Fig. 10. Unfortunately, there is no ground-truth data which would allow objective quality evaluation, nor there are other depth estimation algorithms for omnidirectional video which we could as reference in comparison.

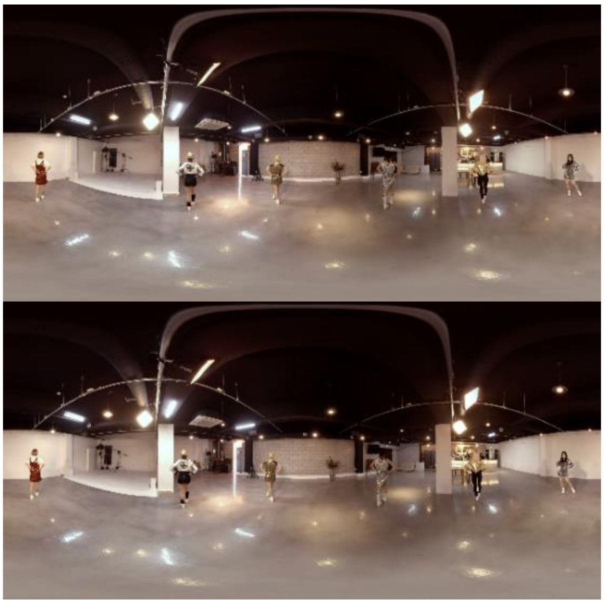


Figure 9. Left and right panorama from *“Dancing360”* test sequence.

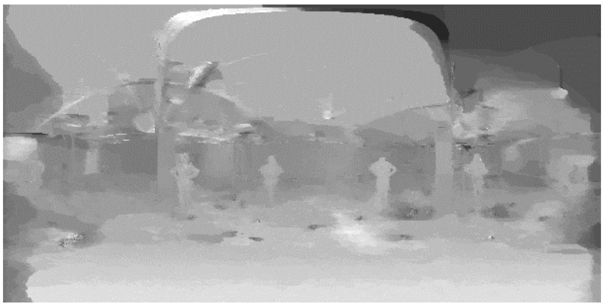


Figure 10. Estimated depth map for *“Dancing360” omnidirectional* test sequence.

5. SUMMARY

A formula for distance measurement in correspondence matching in 360 degree 3D video in circular projection has been presented. Further, similarities between the derived formula and the one describing depth estimation from a rectified pair of pictures have been shown. Then, an approximation of the derived formula is provided. The practical importance of this approximation is related to depth estimation using the currently available depth estimation software. Finally, a necessary condition for approximation has been given. As it was shown, reasonable formula of 4 times radius of omnidirectional camera allows less than 1% relative distance estimation error. Based on the presented analysis of an approximation, necessary camera parameters for depth estimation software have been derived.

The proposed method has been tested on *“Dancing360”* stereoscopic omnidirectional video. Only visual assessment of quality of the attained depth is possible, due to unavailability of ground truth datasets for the considered type of content.

6. ACKNOWLEDGEMENT

This work has been supported by the public funds as a DS research project for years 2017/2018.

7. REFERENCES

[1] vision.middlebury.edu/stereo – Middlebury stereo vision benchmark – [online access 10.02.2018]

[2] A. Lu, Y. Sun, Bin Wang, L. Yu, “*Analysis on circular projection of 360 degree 3D video*” ISO/IEC JTC1/SC29/WG11 MPEG doc. M40077, Geneva, Switzerland, January 2017.

[3] V. Couture V, M.S. Langer, S. Roy, “*Panoramic stereo video textures*”, IEEE International Conference on Computer Vision. 2011:1251-1258.

[4] K. Wegner, O. Stankiewicz, M. Domański „*Depth based view blending in View Synthesis Reference Software (VSRS)*”, ISO/IEC JTC1/SC29/WG11 MPEG2015, M37232, Geneva, Switzerland, 15-23 October 2015.

[5] G. Bang, G. S. Lee, N. Ho H., “Test materials for 360 3D video application discussion”, *ISO/IEC JTC1/SC29/WG11 MPEG2016/M37810*, San Diego, USA, February 2016.