INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG2017/m40596
April 2017, Hobart, Australia

Source	**Poznań University of Technology**
Status	**Input**
Title	**Depth estimation from circular projection of 360 degree 3D video**

Author **Krzysztof Wegner, Olgierd Stankiewicz, Tomasz Grajek, Marek Domański**

1. Introduction

During the 115th meeting in Chengdu, a new MPEG-I project was established. Its goal is to explore and standardize technologies allowing support for 3DoF+ and 6DoF 3D video. The considered prospective technologies include depth image based rendering (DIBR). DIBR requires depth which has to be acquired or estimated.

The document presents mathematical foundations and a method for depth estimation for such a scenario. In particular, circular projection of 360 degree 3D video is considered. The paper addresses the relation of such depth estimation to a more explored case of parallel camera arrangement.

2. Circular projection of 360 degree 3D video

The 360-degree 3D video can be represented with a circular projection. In such an approach, the model of the video is composed of two cylindrical panoramas [1][2]. In each time instant (frame), each panorama represents an image captured by a rotating camera with a very narrow field of view (Fig. 1). An exemplary stereoscopic omnidirectional image in circular projection is presented in Figure 2.

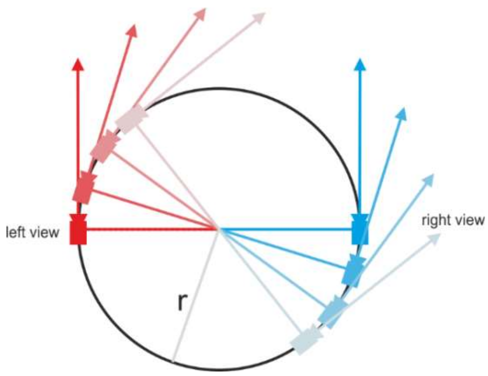

Figure 1. Circular projection obtained using a rotating pair of cameras.

Figure 2. The example of circular projection model (the left view on the top, the right view on bottom) [1]

Obviously, the widely discussed mentioned equirectangular projection differs from the cylindrical projection defined by Formula (1). The principle of cylindrical projection (Fig. 3) is different along vertical axis of the image, which can be expressed as follows:

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

where α and β are the azimuth and elevation, or the longitude and latitude, respectively (cf. Figure 3). Please note that the cylindrical projection was discussed in ages under the name of Mercator projection that is well-known as one of the basic cartographic projections.

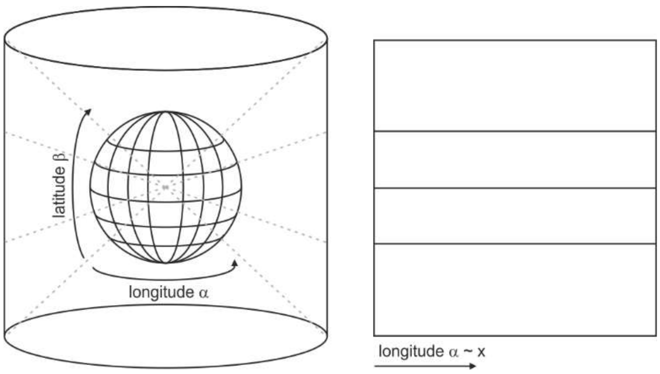

Figure 3. Cylindrical projection principle.

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 4 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.

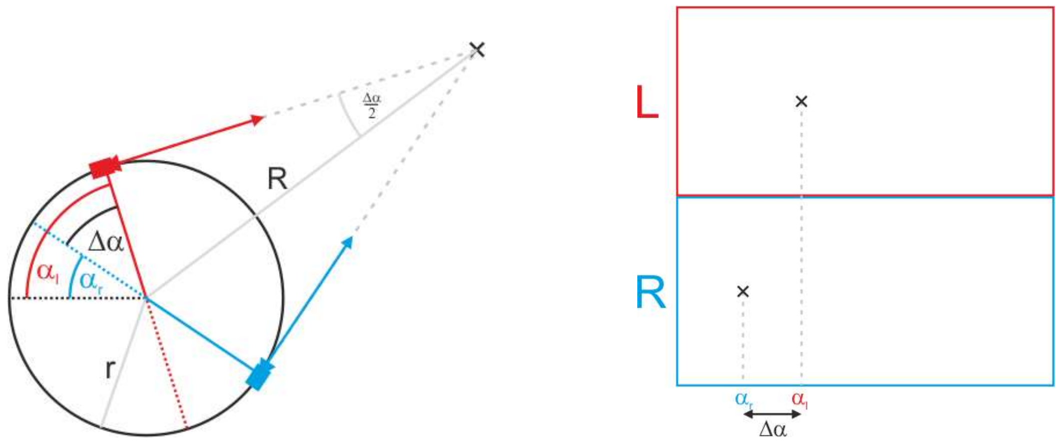

Figure 4. Considered case of matching in circular projection.

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. 5) as angular position (α, β) but it is more convenient to express it in pixel coordinates (x, y).

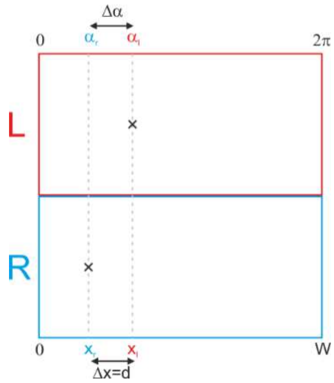

Figure 5. Angular and pixel position of a point in 3D degree 3D video.

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) where the base $b = r/2$:

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

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:

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

Les 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):

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

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 only 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:

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

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:

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

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

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

Which is a direct formula for relative estimation error δR as a function of the distance R. It is presented in Fig. 6.

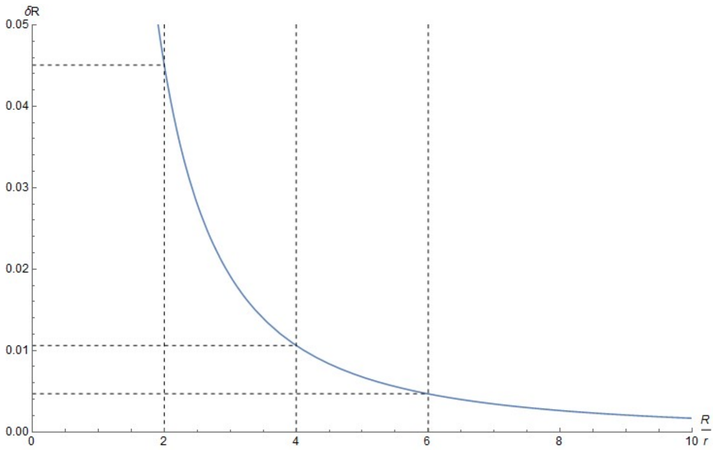

Figure 6. Relative estimation error δR in function of $\frac{R}{r}$

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.

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).

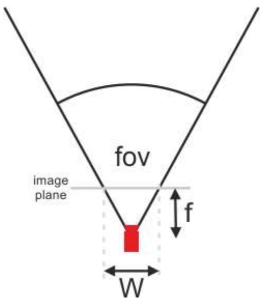

Figure 7. Relationship between focal length, field of view angle and resolution of the camera.

First let us analyze the remaining basic relationship between focal length f, field of view angle fov and horizontal resolution W of the camera (Figure 7):

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

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

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

By expanding it with Taylor series we get:

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

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:

$$Z=\frac{f\cdot b}{d}=\frac{\frac{W}{2\pi}\cdot 2r}{d}=\frac{\frac{W}{\pi}\cdot r}{d}=R \tag{16}$$

which gives equation for our derived approximation (9).

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

$$f=\frac{W}{2\pi} \quad b=2r\,. \tag{17}$$

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

4. Summary

A formula for distance measurement from 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 has been derived.

5. Acknowledgement

This work has been supported by the public funds as a DS research project.

6. References

[1] Ang Lu, Yule Sun, Bin Wang, Lu Yu, "Analysis on circular projection of 360 degree 3D video" ISO/IEC JTC1/SC29/WG11 MPEG doc. M40077, Geneva, Switzerland, January 2017.

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

[3] Krzysztof Wegner, Olgierd Stankiewicz, Marek Domański „Depth based view blending in View Synthesis Reference Software (VSRS)", ISO/IEC JTC1/SC29/WG11 MPEG2015, M37232, Geneva, Switzerland, 15-23 October 2015.