

ESTIMATION OF SPATIAL COORDINATES OF 3D OBJECTS BY STEREOSCOPIC SCANNING

Ondrej KOVÁČ, Ján MIHALÍK

Laboratory of Digital Image Processing and Videocommunications
Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic,
tel. +421 55 602 4309, +421 55 602 2854, e-mail: ondrej.kovac@tuke.sk, jan.mihalik@tuke.sk

ABSTRACT

In the article, mathematical model of simple camera by the perspective projection is described. Furthermore, two basic techniques of stereoscopic scanning of 3D visual scene are described. The first technique is a stereoscopic system with parallel optical axes and the second technique is stereoscopic system where optical axes of cameras are crossed in defined point. From these mathematical constructions equations for spatial coordinates estimation of 3D objects in 3D visual scene are derived. Afterwards, a simulation for different parameters of individual stereoscopic camera systems is performed. Based on the results of simulations, features of both stereoscopic camera systems are analysed.

Keywords: Estimation of spatial coordinates, stereoscopic scanning, perspective projection, anaglyph, disparity

1. INTRODUCTION

Stereoscopic scanning of 3D visual scenes is based on properties of human visual system, because the human visual system in principle is a stereoscopic system. Human brain then evaluates this projection as a 3D visual perception. Currently, record and reconstruction of 3D visual scene require except for 2D projection also record of depth. There are several options how to obtain this information. For example there is a method based on laser scanning or method which uses a projector which projects known pattern on 3D visual scene and objects situated in it. Based on its deformation it is possible to determine depth coordinates [1]. In our article we introduce a method of depth estimation based on stereoscopic scanning of 3D visual scene. Spatial coordinate estimation is necessary in robotics where the robot must be able to determine independently its position in space. We can also use this estimation in 3D modelling of objects. Stereoscopic estimation of spatial coordinates found its application also in automobile factory, where the car automatically evaluates a potentially dangerous situation on the road and warns the driver.

The analysed stereoscopic system of estimation of spatial coordinates is applicable in the situations in which active scanning of 3D visual scene by the sonar, radar or already mentioned techniques are not possible. Further advantage is that storage of depth map is not necessary. Depth map can be computed directly from stereo images. In the 3D visual scene reconstruction it is not necessary to capture textures by using another camera system since one of the stereo images can be used as texture.

Following chapters describe basic mathematical model of the camera system called perspective projection. Based on this mathematical model, two basic arrangements of stereoscopic camera system are described. At the end of the article these stereoscopic systems are compared according to the spatial coordinate estimation accuracy.

2. PERSPECTIVE PROJECTION

Fig.1 shows a simplified arrangement of coordinate system of camera and 3D visual scene (VS). The real camera can be described with mathematical model of perspective projection. In this projection [2] all points of 3DVS are projected to the image plane across the focal point of system.

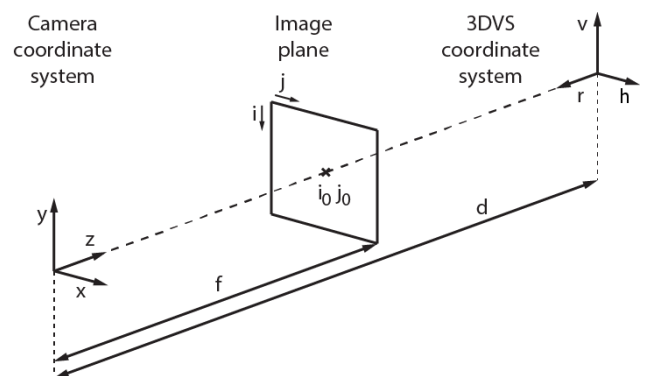


Fig. 1 Simplified model of the camera

Before the calculation of projection of the point of 3DVS into the image plane it is necessary to transform coordinate system of 3DVS into the coordinate system of camera [3, 4]. Thus

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} h \\ v \\ d - r \end{bmatrix} \quad (1)$$

where “d” is the distance between beginnings of coordinate systems. Perspective projection of the point with (x, y, z) coordinates of 3DVS into the point of image with (i, j) coordinates is defined as

$$i = -f_y \frac{y}{z} + i_0 \quad (2)$$

$$j = f_x \frac{x}{z} + j_0 \quad (3)$$

where $f_x = m_x f$ and $f_y = m_y f$ represent the camera focal length linearly adjusted by the scale of m_x and m_y scaling. Point with coordinates (i_0, j_0) is representing principal point of image.

3. STEREOSCOPIC SCANNING

Stereoscopic scanning is one of the techniques of estimation of depth coordinates. There are two configurations. In first of them, optical axes of cameras of the stereoscopic camera system (SCS) are parallel. This system is called coaxial SCS or parallel stereoscopic camera system (PSCS). The other configuration is where the optical axes of cameras are crossed in defined point. This system is called abaxial SCS or toed-in stereoscopic camera system (TSCS) [5, 6].

3.1. Parallel stereoscopic camera system

As shown in Fig.2 the PSCS consists of two identical cameras which are in horizontal direction separated at distance "a". Center of the system is in the middle of this distance. In scanning of 3DVS this point is on the common straight line with beginning of its coordinate system. This straight line is perpendicular to identical planes made up with axes x and y of coordinate systems of cameras.

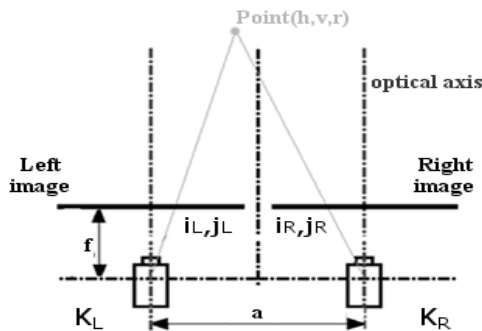


Fig. 2 The camera arrangement of the parallel stereoscopic camera system

Left (KL) and right (KR) camera can be replaced with their perspective projection models. Then for camera KL we can write

$$i_L = -f_y \frac{v}{d-r} + i_0 \quad (4)$$

$$j_L = f_x \frac{h + \frac{a}{2}}{d-r} + j_0 \quad (5)$$

and similarly for the camera K_R

$$i_R = i_L \quad (6)$$

$$j_R = f_x \frac{h - \frac{a}{2}}{d-r} + j_0 \quad (7)$$

Depth coordinate "r" is possible to obtain by solving of the system of equations (5) and (7).

$$j_L = f_x \frac{h + \frac{a}{2}}{d-r} + j_0 \Rightarrow h = (j_L - j_0) \frac{d-r}{f_x} - \frac{a}{2}$$

By substituting h in equation (7) we get

$$j_R = \frac{(j_L - j_0)(d-r) - 2f_x \frac{a}{2}}{d-r} + j_0 \quad (8)$$

After the expression of unknown r from the equation (8) we get

$$r = \frac{d(j_L - j_R) - 2f_x \frac{a}{2}}{(j_L - j_R)} \quad (9)$$

And by the substitution $D = (j_L - j_R)$ we get final equation for depth coordinate.

$$r = \frac{dD - f_x a}{D} = d - \frac{f_x a}{D} \quad (10)$$

where D is horizontal disparity [7]. Horizontal disparity represents difference of horizontal coordinates of projections of equal points on the left and right image. As it is clear from the equation (10), the value of depth coordinates grows with growth of disparity. For estimation of horizontal and vertical coordinates it is possible to deduce next equations. As it is clear from the equation (10), for estimation of depth coordinate it is necessary to know the distance between the SCS and 3DVS and the baseline separation "a" of cameras

$$h = \frac{a}{2D} (j_L + j_R - 2j_0) \quad (11)$$

$$v = \frac{f_x a}{-2D f_y} (i_L + i_R - 2i_0) \quad (12)$$

3.2. Toed-in stereoscopic camera system

Unlike the previous stereoscopic camera system, in this system optical axes of cameras are not parallel but they are crossed in a defined point [9]. As it is clear from Fig. 3, cameras are rotated. In point of crossing between the optical axes there is an angle 2β . Distance "zp" between the beginning of SCS and the point of intersection of optical axis is called the zero parallax and it is defined by equation

$$z_p = \frac{a}{2 \tan(\beta)} \quad (13)$$

where "a" is baseline separation of cameras and β is an angle of rotation of cameras.

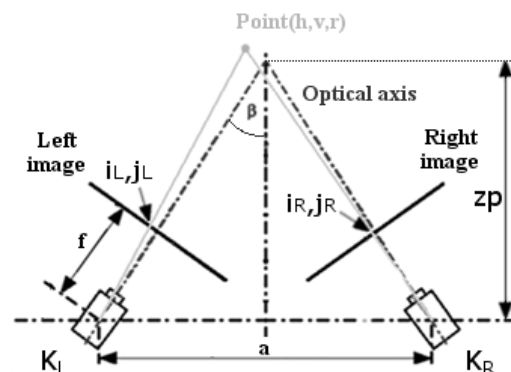


Fig. 3 The camera arrangement of the toed-in stereoscopic camera system

Transformation of coordinate system of 3DVS into the coordinate systems of cameras is defined as

$$\begin{bmatrix} x_L \\ y_L \\ z_L \end{bmatrix} = \begin{bmatrix} \left(h + \frac{a}{2}\right) \cos(\beta) - (d-r)\sin(\beta) \\ v \\ (d-r) \cos(\beta) + \left(h + \frac{a}{2}\right)\sin(\beta) \end{bmatrix} \quad (14)$$

$$\begin{bmatrix} x_R \\ y_R \\ z_R \end{bmatrix} = \begin{bmatrix} \left(h - \frac{a}{2}\right) \cos(\beta) - (d-r)\sin(\beta) \\ v \\ (d-r) \cos(\beta) + \left(h - \frac{a}{2}\right)\sin(\beta) \end{bmatrix} \quad (15)$$

After perspective projection, points of 3DVS are projected into the individual images by using these equations

$$\begin{bmatrix} i_L \\ j_L \end{bmatrix} = \begin{bmatrix} -f_y \frac{y_L}{z_L} + i_0 \\ f_x \frac{x_L}{z_L} + j_0 \end{bmatrix} \quad (16)$$

$$\begin{bmatrix} i_R \\ j_R \end{bmatrix} = \begin{bmatrix} -f_y \frac{y_R}{z_R} + i_0 \\ f_x \frac{x_R}{z_R} + j_0 \end{bmatrix} \quad (17)$$

Disadvantage of this method is the difficulty of computation of system of equations (16) and (17) for estimation of depth coordinate. Another disadvantage is that besides of the horizontal disparity, also the vertical disparity arises [8]. The vertical disparity represents the difference of vertical coordinates of projections of equal points of 3DVS in the left and right image. With the increase of coordinates in the horizontal direction, the vertical disparity is increasing too. This causes an error in estimation of the spatial coordinates. The estimation of depth coordinates is possible to get by equation (18)

$$r = d - \frac{af_x}{a \frac{f_x}{z_p} + D} = d - \frac{af_x}{2f_x \tan(\beta) + D} \quad (18)$$

As it is clear from the equation (18), in case if value of the angle β is equal to zero, the equation is changed onto the equation (10). This is confirmation that PSCS is only a special case of TSCS. For the estimation of horizontal and vertical coordinates it is possible to deduce following equations

$$h = \frac{a(j_L + j_R - 2j_0)}{2(a \frac{f_x}{z_p} + D)} = \frac{a(j_L + j_R - 2j_0)}{2(2f_x \tan(\beta) + D)} \quad (19)$$

$$v = \frac{af_x (i_L + i_R - 2i_0)}{-2f_y (a \frac{f_x}{z_p} + D)} = \frac{af_x (i_L + i_R - 2i_0)}{-2f_y (2f_x \tan(\beta) + D)} \quad (20)$$

As it is clear from the equation (18), for estimation of depth coordinate is necessary to know the distance between the SCS and 3DVS, the baseline separation “ a ” of cameras and the angle β . If $z_p = d$, then it can be shown that horizontal disparity is equal to zero for points which are from SCS in distance equal to “ d ”.

4. EXPERIMENTAL RESULTS

Fig. 4 shows 3DVS with three pyramids. Their distribution is the same for experiments with both of SCS. The main vertices of pyramids have depth $r_1 = 500$,

$r_2 = 400$, $r_3 = 250$ points and others vertices have zero value of depth in a common coordinate system of 3DVS.

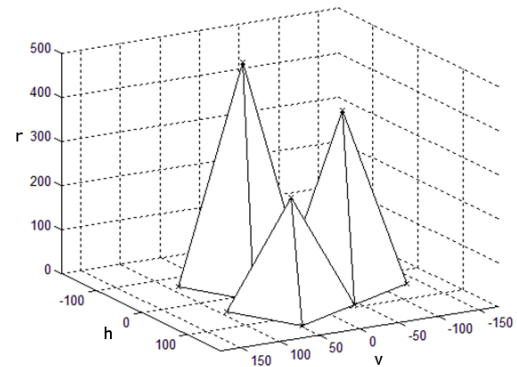


Fig. 4 The 3D visual scene with three pyramids

The anaglyph [10] acquired by the PSCS is shown in Fig.5.a and an anaglyph acquired by the TSCS is shown in Fig.5.b.

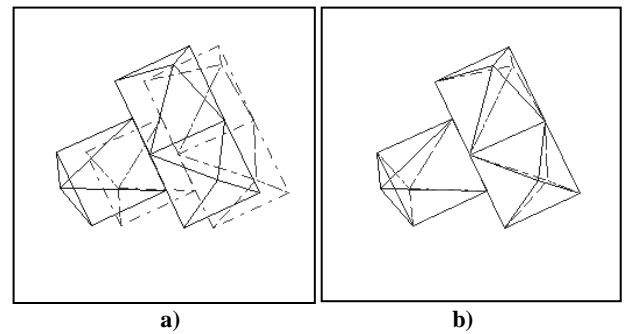


Fig. 5 The anaglyph acquired by the a) PSCS b) TSCS

Accuracy of depth estimation is evaluated in decibels according to the equation (21)

$$\text{SNR} = 10 \log \frac{\sigma_r^2}{\sigma_q^2} \quad (21)$$

where

$$\sigma_r^2 = \frac{1}{n} \sum_{i=1}^n r_i^2 \quad (22)$$

is the mean square value of depth coordinates of pyramids vertices and

$$\sigma_q^2 = \frac{1}{n} \sum_{i=1}^n (r_i - \bar{r}_i)^2 \quad (23)$$

is mean square value of error between the real and estimated values of depth coordinates of pyramids vertices. Similarly, it is possible to evaluate the accuracy of estimation of other coordinates. From the equations for the estimation of horizontal and vertical coordinates it is clear that these equations as well as the equation for the depth estimation depend only on values of coordinates of the point in raster in left and right images. Also disparity depends on these coordinates and other variables are known. Therefore for purposes of evaluation of our experiments, the evaluation of depth accuracy is sufficient. Accuracy is evaluated for all vertices of pyramids. Subsequently, it is evaluated influence of the camera baseline separation and the distance between SCS

and 3DVS on the accuracy of depth estimation. In our experiments it is supposed that the scaling factors of cameras in horizontal and vertical direction are equal, thus $m_x = m_y = 10$. Also it is supposed that the $z_p = d$. Generally, the coordinate system of 3DVS can be given in scene units [su]. For simplicity, all variables are given in points. In our case one point represents 0.25 mm. Conversion between distances given in points and millimeters is achieved by the equation (24).

$$d[mm] = \frac{d[point]}{m_p} \tag{24}$$

where m_p is the point scale and in our case it is $m_p = 0.25 \frac{point}{mm}$.

The image is of a resolution 512x512 points. In Tables 1 - 4 the parameters of stereoscopic camera systems are shown and considering them the accuracy of estimation is evaluated. The distance between SCS and 3DVS is evaluated for values 1622, 2433, 3244 and 4055 points. The influence of the baseline separation on the accuracy of estimation is evaluated for the distance from the interval <15, 55> mm with step 5mm. These are converted in to points by equation (24). Also in the Table 1 mean square of the depth and the estimation errors are shown.

Table 1 Dependence of SNR given the baseline

d=1622, f=85, m _x =10, m _y =10					
PSCS			TSCS		
a	σ_r^2	σ_q^2	SNR[dB]	σ_q^2	SNR[dB]
61	31500	42,80	28,6687	37,0667	29,2933
81	31500	259,67	20,8389	27,2	30,6374
101	31500	124,53	24,0302	20,8667	31,7886
122	31500	14,80	33,2805	12,0667	34,1672
142	31500	79,53	25,9776	6,73333	36,7008
162	31500	50,13	27,9818	6,9333	36,5737
182	31500	77,20	26,1069	3	40,2119
203	31500	38,07	29,1777	16,9333	32,6957
223	31500	22,20	31,5196	17,2	32,6278

separation of the cameras for distance 1622 points

Table 2 Dependence of SNR given the baseline separation of the cameras for distance 2433 points

d=2433, f=85, m _x =10, m _y =10					
PSCS			TSCS		
a	σ_r^2	σ_q^2	SNR[dB]	σ_q^2	SNR[dB]
61	31500	3584,73	9,43854	464,6	18,3123
81	31500	869,2	15,5919	27,667	20,5635
101	31500	1554,8	13,0664	3,26667	39,8421
122	31500	996,467	14,9985	40,1333	28,9481
142	31500	371,6	19,2823	58	27,3488
162	31500	524,467	17,7859	27	20,6695
182	31500	244,333	21,1033	27,9333	30,5219
203	31500	213,267	21,6939	48,9333	28,0871
223	31500	38,3333	29,1473	4,1333	38,8201

Table 3 Dependence of SNR given the baseline separation of the cameras for distance 3244 points

d=3244, f=85, m _x =10, m _y =10					
PSCS			TSCS		
a	σ_r^2	σ_q^2	SNR[dB]	σ_q^2	SNR[dB]
61	31500	42,8	28,6687	37,0667	29,2933
81	31500	4952,07	8,03524	696,6	16,5533
101	31500	3124,33	10,0355	92,3333	25,3295
122	31500	539,6	17,6624	581,533	17,3374
142	31500	1787,87	12,4598	724	16,3857
162	31500	1366,4	13,6273	574	17,394
182	31500	1055,6	14,7481	423,733	18,7122
203	31500	961,533	15,1535	394	19,0281
223	31500	657	16,8075	229,933	21,3671

Table 4 Dependence of SNR given the baseline separation of the cameras for distance 4055 points

d=4055, f=85, m _x =10, m _y =10					
PSCS			TSCS		
a	σ_r^2	σ_q^2	SNR[dB]	σ_q^2	SNR[dB]
61	31500	13907,2	3,5507	1742,93	12,5703
81	31500	10487,9	4,77621	442,267	18,5263
101	31500	4630,67	8,32667	466,533	18,2943
122	31500	5253,73	7,77843	1031,73	14,8474
142	31500	29,79,1	10,2423	455	18,403
162	31500	149,133	23,2474	113,667	24,4268
182	31500	1425,13	13,4446	216,067	21,6372
203	31500	2202,33	11,5543	199,6	21,9815
223	31500	1468,87	13,3133	302	20,183

The table 5 shows dependence of mean value of SNR for all value of baseline on distance between SCS and 3DVS.

Table 5 Dependence of mean value of SNR given by the distance between SCS and 3DVS

d [points]	Mean value of SNR [dB]	
	PSCS	TSCS
1622	27,51	33,86
2433	18,01	28,12
3244	15,24	20,16
4055	10,69	18,99

From the comparison of graphs in Fig. 6, 7 and 8 it is clear that TSCS achieves better results than PSCS. As it is obvious from graph in Fig. 8 and the table 5 it can be also seen that with growing distance between SCS and 3DVS accuracy of depth estimation decreases in both systems.

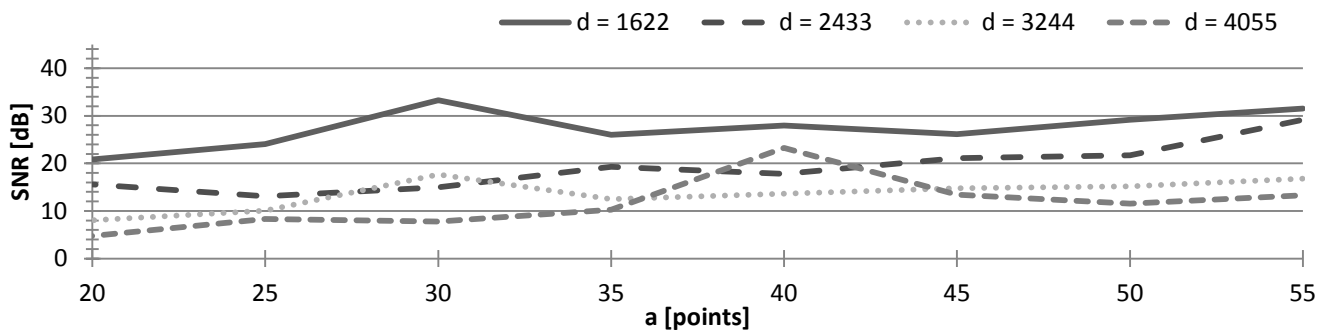


Fig. 6 Graphs of dependence of SNR on the given baseline separation of the cameras for PSCS

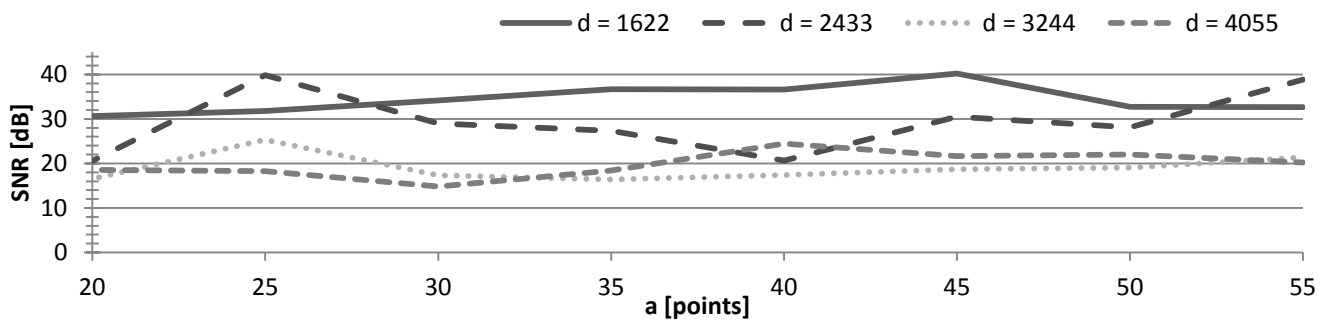


Fig. 7 Graphs of dependence of SNR on the given baseline separation of the cameras for TSCS

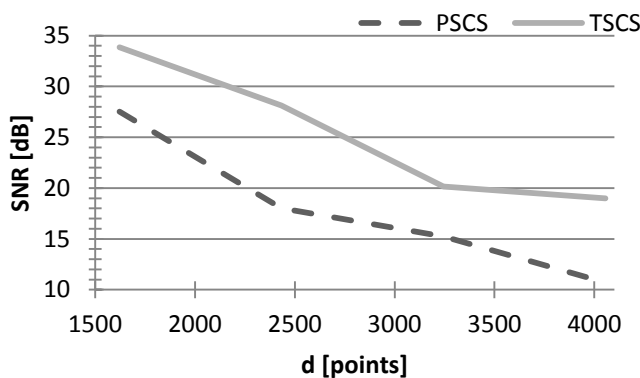


Fig. 8 Graphs of dependence mean value of SNR on distance between SCS and 3DVS

As we mentioned above in TSCS except for the horizontal disparity, the vertical disparity also arises. This causes an inaccurate estimation of depth coordinate of the points shifted in horizontal direction from the center of the 3DVS coordinate system. In the case of PSCS these kinds of errors caused by the vertical disparity are not occurred. Vertical disparity causes the curvature in depth coordinates of the reconstructed 3DVS as is shown on the Fig.9. Fig.9.a shows that the orthogonal grid is placed over the plane $h\nu$ and Fig.9.b shows its reconstruction by the $h\nu$ and r . The points of 3DVS shifted in the horizontal direction from the center of the 3DVS coordinate system are below the ideal depth plane. That is why it is necessary to put the objects of the 3DVS in the area around the center of the 3DVS coordinate system.

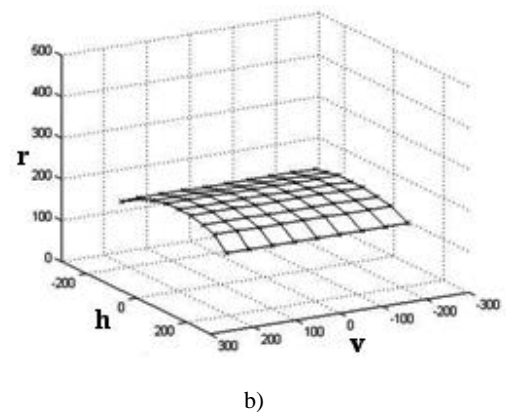
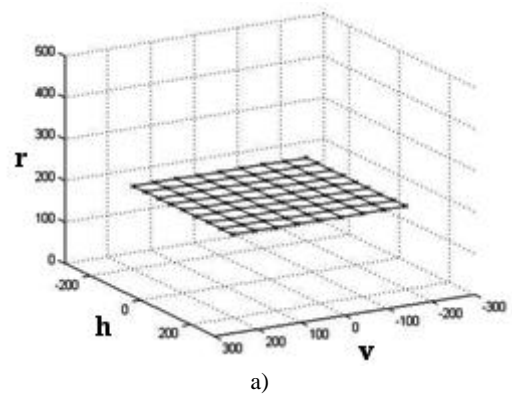


Fig. 9 a) the 2D orthogonal grid b) reconstruction of the 2D orthogonal grid

5. CONCLUSION

In the article, we mathematically modelled the camera and based on this we mathematically described two stereoscopic systems with different configuration. The first configuration is the system with two cameras with parallel optical axes and the second configuration is system where the optical axes of cameras are crossed in defined point. We showed that the PSCS is only special case of the TSCS. By the simulation of both systems we created the images for the varying distances between the SCS and the 3DVS and for the varying baseline separation of cameras. On basis of the images we performed estimation of depth coordinates of the simulated 3DVS with three pyramids. By the comparison of results it is clear that for purposes of depth estimation and the reconstruction of the 3DVS for both systems, comparable results are achieved. The main aim of the article is to mathematically describe two basic stereoscopic systems and on the basis of the achieved experimental results perform the analysis of their properties. For the real cameras, their parameters, like the focal length and the scales of scaling for the horizontal and vertical direction or also the distance between SCS and 3DVS are not always known. These intrinsic and extrinsic parameters are possible to get by the calibration. The real cameras also cause the additional distortion which is possible to compensate by the methods of correction.

REFERENCES

- [1] LANMAN, D. – TAUBIN, G.: Build your own 3D scanner: 3D photography for beginners. *ACM SIGGRAPH 2009 Courses*, New Orleans, Louisiana, p.1-94. 2009.
- [2] GONZALES, R.C. – WOODS, R.E.: Digital Image Processing, 2-nd Edition, *Prentice Hall*, 2002.
- [3] MIHALÍK, J. – MICHALČIN, V.: 3D Motion Estimation and Texturing of Human Head Model. *Radiengineering*, Vol.13, No.1, p.26- 3, 2004.
- [4] MARQUEZ, J.: Canonical 3D Geometric Transformations in Matrix Notation. *Ccadet unam*, 2011.
- [5] YAMANOUE H.: The differences between toed-in camera configurations and parallel camera configurations in shooting stereoscopic images. *Int. Conf. on Multimedia and Expo*, p. 1701-1704, 2006.
- [6] WOODS, A. – DOCHERTY, T. – KOCH, R.: Image Distortions in Stereoscopic Video Systems. *Proc. of the SPIE Vol. 1915, Stereoscopic Displays and Applications IV*, San Jose, California, February 1993.
- [7] LANG, M. – HORNUNG, A. – WANG O. – POU-LAKOS, S. – SMOLIC, A. – GROSS, M.: Nonlinear Disparity Mapping for Stereoscopic 3D. *Proc. of ACM SIGGRAPH Vol.29*, no. 3, 2010.
- [8] ALLISON, R.S.: An analysis of the influence of vertical disparities arising in toed-in stereoscopic cameras. *Journal of Imaging Science and Technology*, 51(4), p.317-327, 2007.
- [9] HEINZLE, S. – GREISEN, P. – GALLUP, D. – CHEN, CH. – SANER, D. – SMOLIC, A. – BURG, A. – MATUSIK, W. – GROSS, M.: Computational stereo camera system with programmable control loop. *ACM Trans. on Graphics*, Vol.30, Num. 4, p. 1-10, 2011.
- [10] KRŤŽENECKÝ, J.: Anaglyf aneb barevná separace obrazů, <http://stereofotograf.eu/navody/anaglyf/>

Received March 11, 2014, accepted November 24, 2014

BIOGRAPHIES

Ondrej Kováč was born on 8.7.1986. In 2011 he received the MSc. degree in Multimedia telecommunications from the Faculty of Electrical Engineering, Technical University of Košice. In this time he is PhD student at the Department of Electronics and Multimedia Telecommunications, Technical University of Kosice. His PhD. thesis topic is focused to texture generating, 3D modeling and coding of human head. He works in Laboratory of Digital Image Processing and Videocommunications

Ján Mihalík graduated from the Technical University of Bratislava in 1976. In 1979, he joined the Faculty of Electrical Engineering and Informatics of Technical University of Košice, where he received his PhD degree of radioelectronics in 1985. Currently, he is a full professor of electronics and telecommunications and the head of the Laboratory of Digital Image Processing and Videocommunications at the Department of Electronics and Multimedia Telecommunications. His research interests include information theory, image and video coding, digital image and video processing and multimedia videocommunications.