

# Towards Measuring of Depth Perception from Monocular Shadow Technique with Application in a Classical Painting

Wei Wen\*, Siamak Khatibi

Department of Communication Blekinge Tekniska Högskola, Karlskrona, Sweden.

\* Corresponding author. Tel.: +46-455385860; email: wei.wen@bth.se

Manuscript submitted July 22, 2015; accepted October 28, 2015.

doi: 10.17706/jcp.11.4.310-319

---

**Abstract:** Depth perception is one of important abilities of the human visual system to perceive the three dimensional world. Shadow technique that offers different depth information from different viewing points, known as Da Vinci stereopsis, has been used in classical paintings. In this paper, we report a method towards measuring the relative depth information stimulated by Da Vinci stereopsis in a classical painting. We set up a positioning array of cameras for capturing images from the portrait using a high resolution camera, where the changes of shadow areas are measured by featuring the effects as point and line changes. The result shows that 3D effects of the classical painting are not only a perceptual phenomenon but they are also physically tangible and can be measured. We confirm validity of the method by its implementation even on a typical single image and comparison of results between the single image and the portrait.

**Key words:** Depth perception, Da Vinci stereopsis, pictorial cues, classic painting, shadow technique.

---

## 1. Introduction

Our 3D world is perceived by our binocular vision. However the 3D perception is not limited to have two views of a scene. We have also ability to grasp the 3D content from certain cues of a single image. These cues, known as pictorial cues, stimulate a depth perception [1], which have been understood for centuries and are used in classical paintings since Renaissance time, such as Raphael's 'School of Athens' (1510), Leonardo Da Vinci's Mona Lisa's Smile (1517) and so on.

The modern 3D display technology has brought a new method in perceiving 3D content where an interlaced image is generated by fusion of several images captured from different points of view of a scene. When the interlaced image is shown by optical means to viewers, two different images can be seen in several zones. Although a 3D perception can be stimulated by the interlace method or by the pictorial cues in a single image, the interlaced method can offer different 3D content information from different points of view which is not feasible with pictorial cues in single images.

Some of paintings, mostly classical ones, make different impression of 3D content when they are observed from different points of view. The stimulation of 3D perception in these paintings can be explained by the concept of Da Vinci stereopsis. As far as they are considered as single images and are not viewed by any optical means, some justified questions possibly can be arisen. Is it possible to experience the same viewers' impression by capturing images? Is it possible to quantify such impression? In this paper we present a quantification method which will help us to analyze and answer such mentioned questions. To the best of our knowledge, our approach is the first work that tries arising such questions and answering

them.

This paper is organized as follows. In the Section 2, the pictorial cues and Da Vinci stereopsis phenomena are discussed which can explain the stimulation of the depth perception from single images. A depth perception model is proposed in Section 3. Section 4 presents the measurement setup. The results and measurement analysis are presented in Sections 5 and 6. Finally, we summarized our work in this paper in Section 7.

## 2. Depth Perception in Single Images

The depth perception is the visual ability to perceive the world in three-dimensions and distinguish the distance of an object [2]. In single images the stimulation of depth perception can be explained by pictorial cues or Da Vinci stereopsis concept. In everyday perception of our surrounding world we get help of different cues which seems to be processed together in the brain. Some of these cues are related to our depth perception and are used by artists to create a sense of depth. Three major cue classes (binocular, Da Vinci stereopsis and monocular) contribute to the depth perception and within each classes there are several subclasses of cues. The cues involved in the depth perception are shown in Fig. 1.

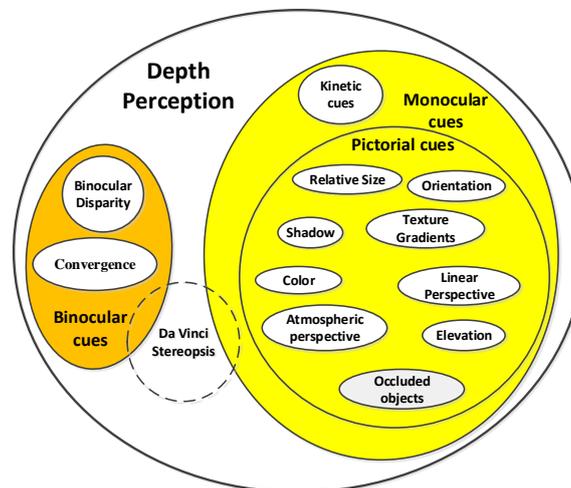


Fig. 1. The cues involved in the depth perception.

The depth cues in the single images can be pictorial cues [2] or Da Vinci Stereopsis cues [3] which are observed by two eyes. The binocular cues stimulate a depth perception by observing two images where each image is only seen by one eye [4]-[6]. The binocular cues as present of parallax disparities in the two observed retinal images cause the depth sensation. The parallax disparities refer to the objects which are in overlapping region of a scene seen by the two eyes. They also refer to points within objects that are well defined, such as edge points. The binocular cues in absence of parallax disparities but present of shadows in the two observed retinal images may also cause the depth perception [7]. The shadows can be defined as regions in an image which have abrupt luminance and/or color changes but are neither edges nor boundaries. The da Vinci stereopsis cues in single images, as a phenomenon, have a close relation to binocular cues of shadows stereopsis in two images [3]. In binocular observation of the da Vinci Stereopsis cues, the entire observation region is not seen by two eyes. Assuming the observed region has a distinguishable background and foreground objects, when part of the background is occluded by the foreground objects, some regions are seen only by one eye which are called monocular regions [8], [9]. The monocular regions left out the use of parallax disparity cues in the related regions on the observed two images. However the monocular regions can contribute to stimulation of a depth perception if they are

adjacent to e.g. a shading region, a texture area or a color differences area [3]. The perceived depth is quantitative in nature [9]. The perceived depth of a monocular point increases with increasing its separation from occluding edge.

Fig. 2 elaborates the four basic types of Da Vinci stereopsis cues [10]. The black thick line shown in each sub-Figures of a-d in the figure represents the occluder and the non-circle color area is the representative marker for visibility area either for the left eye (LE) as blue or right eye (RE) as red. Each colored area is seen only by one eye as the monocular region. In three types of Da Vinci stereopsis cues, shown in Fig. 2; occlusion; aperture; and camouflage, the boundaries of the background and foreground are rigid and independent of any angle of scene view. However in the phantom type of such cues shown in Fig. 2d, shadows as non-apparent luminance edges cause phantom surfaces to generate monocular regions which in their turn they stimulate a depth perception. Apparently the boundaries of the background and foreground become non-rigid and dependable to angle of scene view [11], [12].

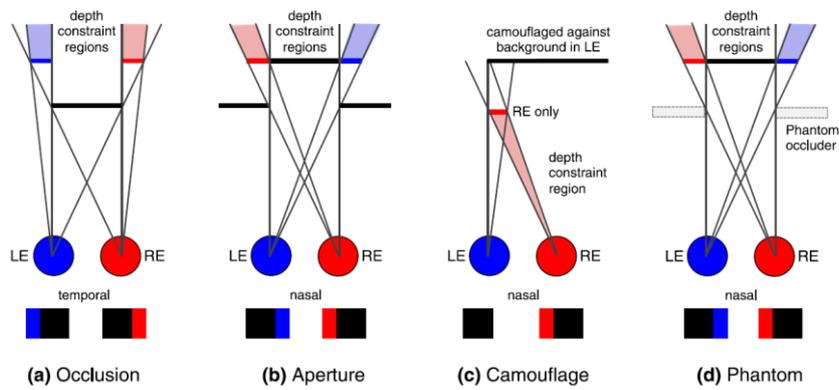


Fig. 2. Four basic types of Da Vinci stereopsis cues [8].

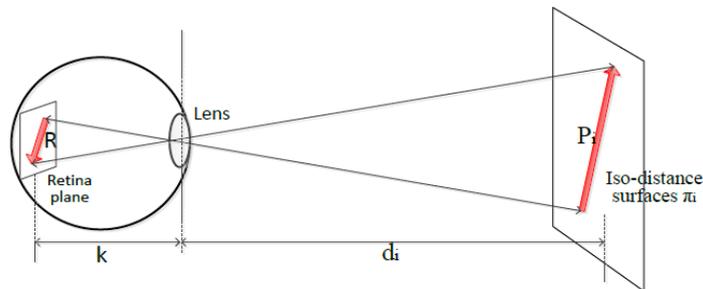


Fig. 3. Modeling of human eye.

In some of classical paintings, shadow technique in form of the phantom type of Da Vinci stereopsis cues has been used to stimulate the depth perception of the viewers. The pictorial cues are widely used in painting and photographing to stimulate or to enhance the human's 3D perception. However, when the depth information is associated to a real scene the information is changing with respect to viewers' position. This expectation is arisen from the observation of real physical 3D objects where different information from the objects can be revealed by changing the observation position. A painted portrait has strong limitation to evoke any depth perception caused by its 2D physical shape. The paintings with monocular pictorial cues do not stimulate real depth information and the information remains the same by change of view point. On the other hand, implementation of the Da Vinci stereopsis cues, such as phantom ones, can stimulate real depth information. This has been seen in some of classical paintings and it causes an impressive reaction on the viewers.

### 3. Depth Perception Modeling

Let us assume an object is seen by an eye in finite number of iso-distance surfaces ( $\pi_{1...N}$ ) where each iso-distance surface is normal to the eye's optical axis and have a certain distance to the eye. The size between any two points on the contour of the object in an iso-distance surface, e.g.  $\pi_i$ , can be seen on the eye's retinal plane by  $R = P_i k / d_i$ , where  $R, P_i, d_i, k$  are size of the retinal image, size between the two points on the contour of the object in  $\pi_i$ , the distance of the iso-distance surface of  $\pi_i$  from the eye's lens and the eye's focus length, respectively as it is shown in the Fig. 3.

By change of eye position in parallel to the iso-distance surfaces, the size of the retinal image  $R$  of certain points on the contour of the object in a certain iso-distance surface is not changed and the retinal image undergoes only a translation in the retinal plane. On the other hand if the size of the retinal image  $R$  remains the same and we consider size between certain two points on the contour of the object in different iso-distance surfaces, e.g.  $\pi_i$  and  $\pi_j$ , then it yields  $P_i / P_j = d_i / d_j$  where  $P_i, P_j, d_i, d_j$  are size between certain two points on the contour of the object in the iso-distance surfaces of  $\pi_i$  and  $\pi_j$  and the distance of the iso-distance surfaces  $\pi_i$  and  $\pi_j$  from the eye's lens. Our arguments in the above yields the same conclusion when we replace the eye with a camera where the eye's focus length  $k$  is changed to the camera focal length  $f$ .

### 4. Measurement Setup

The youngest Madonna "Nativity" portrait is a classical painting in Kulenovic collection in Karlskrona Sweden. The naked eye observation of the portrait reveals real depth information by majority of observers which makes the portrait interesting for experimental measurement. The experimental setup is shown in Fig. 4. A camera, Canon 450D with resolution of 4272 by 2848 was placed in six different positions parallel to the portrait along a straight line. The interval of camera positions was 5.5 centimeters, and the camera positions were aligned with a normal distance of 1.2 meters to the portrait plane to capture the whole portrait in each position. Each positioning of the camera was done using a laser distance measurement instrument, Bosch DLE 40, to obtain parallel positioning of the image plane to the portrait surface. The camera height in each position was arranged in a way that the middle of portrait and center of the image plane were aligned vertically. The experiment setup assures that the image planes in different camera positions are parallel to the portrait plane thus the distance of the object plane from the retinal image plane,  $d_j$  in Section 3, is constant in each camera positioning.

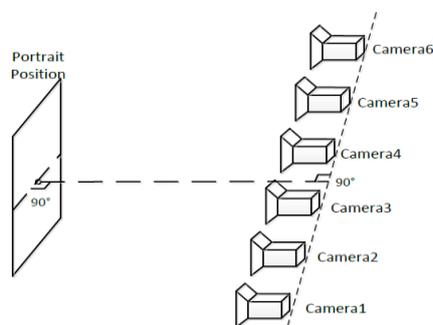


Fig. 4. The measurement setup.

On each of six captured images, hundred points were selected in one hundred specific areas manually by five subjects, where to each point a label number was assigned as shown in Fig. 5. Each correspondent point in the images received the same label number. For selection of correspondent points in six images, each image was zoomed and the same perceptual corner area was chosen where the same interpolation

algorithm for zooming was applied to reduce the effect of interpolation on the marked positions. Fig. 6 shows the corner selecting procedure of one interesting perceptual corner area, shown in the left image. In each six images, shown in the right image of Fig. 6, the corner was marked manually by five subjects, shown as blue circles. The geometrical average of manual marking is then calculated, shown as red circles. Two corner detection algorithms, Harris [13] and Chord-to-Point Distance Accumulation (CPDA) [14] are used for selecting corner points automatically in each interesting perceptual corner area and by considering a rectangle area generated from marked positions by subjects, shown in Fig. 6 as the green and yellow circles. All connections of each two points in each image generated 4950 line segments. Each line segment represented a size of object as described in the Section 3. A certain label number was assigned to each line segment in conjunction to the connection points in a way to generate 4950 correspondent lines in each image.



Fig. 5. 100 selected points on the image of the portrait by human volunteers and corner detection algorithms.

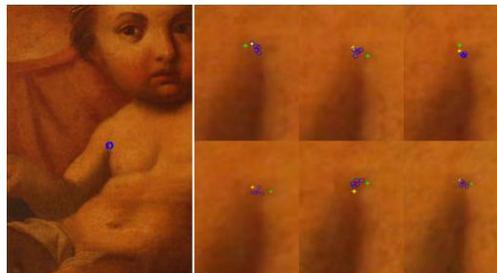


Fig. 6. A typical marking of the correspondent points on each of six portrait images is demonstrated.

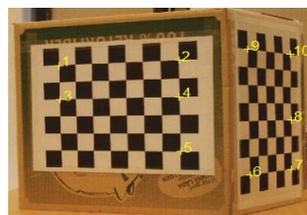


Fig. 7. The ten selected points on one image of six captured images from a simple 3D scene.

For verification of the model presented in the Section 3 and comparison of the portrait to single images, a reference 2D single image with simple 3D scene content was captured from six points of view, as shown in Fig. 4. Ten points were selected manually at each of the six captured images with the same zooming procedure as for the portrait images by one subject. The marked ten points are shown in Fig. 7. Then for the

ten chosen points the same procedure was carried out as it was done for the portrait. In each image forty five line segments were generated by connecting of each two points. A certain label number was assigned to each line segment in conjunction to the connection points in a way to generate forty five correspondent lines in each image.

### 5. Measurement Results

The histograms of each six captured images from both the portrait and reference image are computed. The Chi-square and Bhattacharyya distance are used for measuring the difference of histograms [15] and [16], shown in Table 1. The small values in Table 1 reveal that the histograms are heavily correlated and indicate that the images have been captured in highly correlated light conditions as far as the intensity information of each two captured images of the portrait and reference image are almost remained the same. The Table 1 also shows that the images from the portrait has been captured even in better light correlated condition than captured images from the reference image due to their lower values in the histogram comparisons. The amount of position changes, in pixel unit, for the selected points in each two subsequent images of the reference image is as shown in Fig. 8. In portrait case, Fig. 9 shows the amount of point position changes, in pixel unit, for the geometrical average selected points in each two subsequent captured images where the positions of all selected points are changing significantly compared to the reference image despite of less light condition changes in capturing images from the portrait in comparison to the reference image.

Table 1. Histogram Comparison of Images from the Portrait and the Reference Image

Images for Comparison	Chi-square		Bhattacharyya	
	Portrait	2D image	Portrait	2D image
1st vs 2nd	3.29	12.27	0.11	0.19
1st vs 3rd	3.29	12.27	0.11	0.19
1st vs 4nd	2.06	12.72	0.12	0.21
1st vs 5nd	3.29	12.27	0.11	0.19
1st vs 6nd	3.29	12.27	0.11	0.19

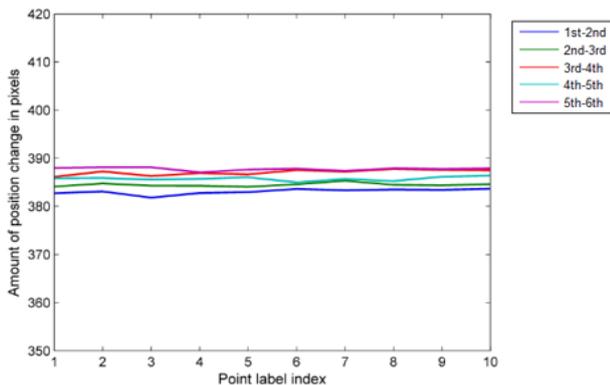


Fig. 8. Position changing of selected points in each two subsequent images of the reference image.

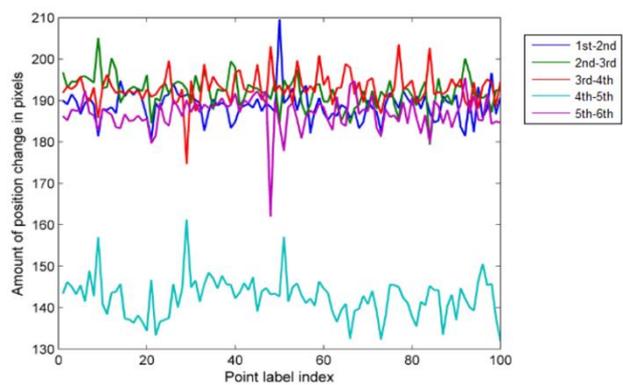


Fig. 9. Position changing of geometrical average selected points in each two subsequent images of the portrait.

In the captured images from the reference image, the results of length ratio of correspondent lines on each two subsequent images, as shown in Fig. 10, indicate that there are no depth perceptual changes in each two subsequent images of the reference image. On the contrary, Fig. 11 shows significant changes in length ratio of correspondent lines in each two subsequent images of the portrait, indicating the changes in the relative depth perception in each two subsequent images; according to Section 3. Also the relative depth

perception is not changing homogenously in each two subsequent images. Fig. 12 and Fig. 13 are the histogram results of the position change and length ratio for the portrait images. From top to the bottom, there are results of geometrical average of subjects' selected points, CPDA corner points and Harris corner points respectively. In either Fig. 12 or Fig. 13, the histogram results by automatic corner detection algorithms verify the subjective results; particularly the CPDA results are much closer to the subjective one. To visualize the most significant correspondent lines, from each two points of the geometrical average selected points, those which had a length ratio of less than 0.9 or bigger than 1.1 were selected in each two subsequent images; the result is shown in Fig. 14.

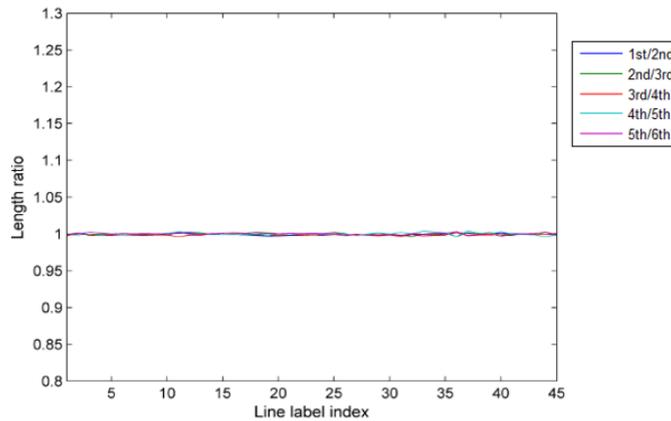


Fig. 10. The length ratio of correspondent lines in each two subsequent images of the reference image.

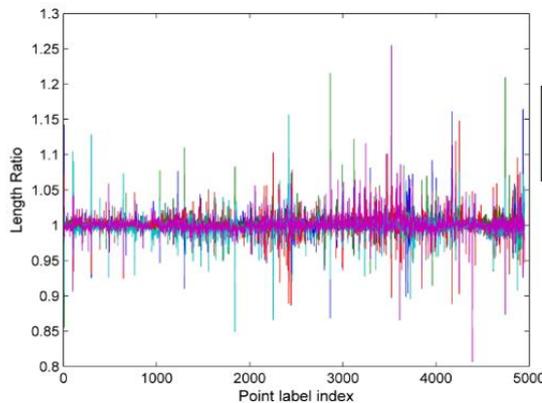


Fig. 11. The length ratio of correspondent lines by geometrical average selected points in each two subsequent images of the portrait.

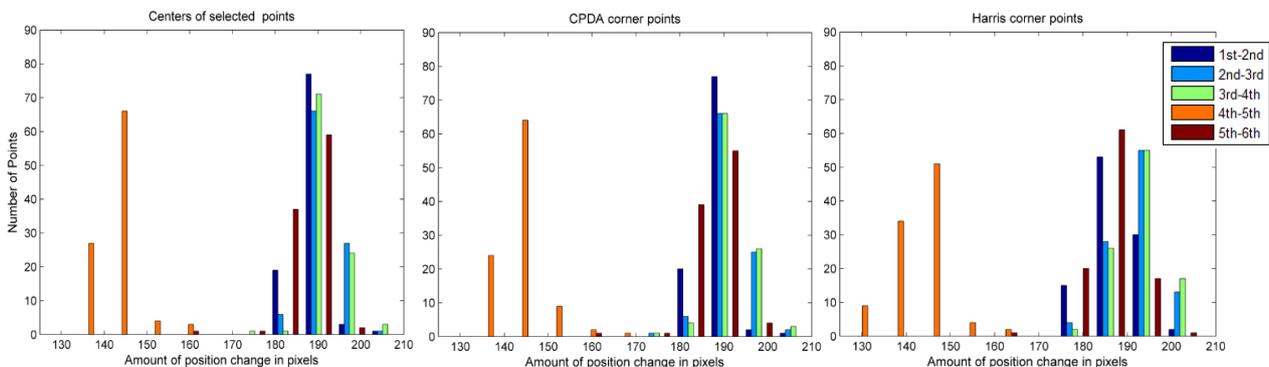


Fig. 12. Histograms of point position change in pixels for geometrical average selected points, CPDA corner points and Harris corner points (from left to right) in each two subsequent images of the portrait.

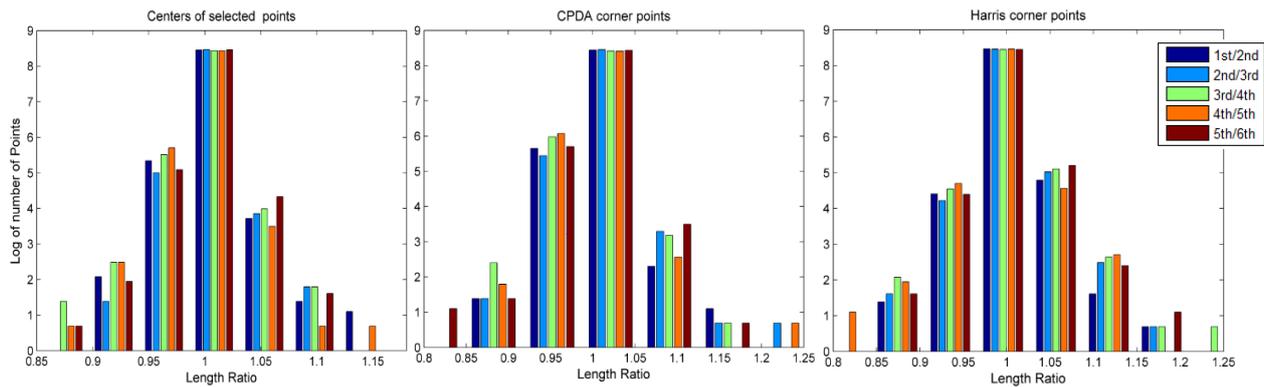


Fig. 13. Histograms of length ratio of correspondent lines by geometrical average selected points, CPDA corner points and Harris corner points (from left to right) in each two subsequent images of the portrait.

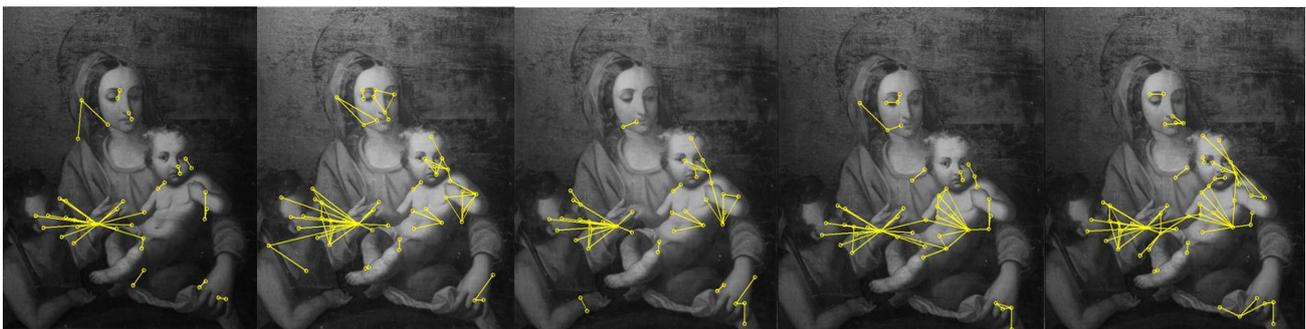


Fig. 14. The most significant correspondent lines, in yellow and with a length ratio of less than 0.9 or bigger than 1.1, in each two subsequent images are presented.

## 6. Measurement Analysis and Discussion

On the portrait images manually one hundred points were selected which were in perceptual correspondent corner areas in all six captured images. The selection of each correspondent corner point was done by zooming on the captured images and by careful inspection of corner area as shown in Fig. 6. The position changing of correspondent points varied differently and the small fluctuation in changing values can reasonably be caused by error in the procedure. However Fig. 12 and Fig. 13 show that the procedure of subjective selecting has no effect on the results due to the correlation between the automatic and subjective results. Thus the significant position changes of correspondent points, as shown in Fig. 9, and length ratio of correspondent lines shown in Fig. 11 can only be explained by the variation of iso-distance surfaces in captured images from the portrait. The histogram analysis proved that the pixel intensity values in the captured images have very small changes, according to Table 1 which shows that the histogram differences are very small. These argue that the shadows on the portrait, as non-apparent luminance of edges, had not affected the image capturing, sampling and quantization procedure, significantly due to the results in the Table 1 but still enough to generate significant perceptual effects. According to the above facts we hypothesize that the shadow size and color can regulate the size of the phantom occluder which was explained in Section 2. The blind implementation of correspondent lines showed beneficial because the visualization of most significant length ratio of them revealed unexpected results, shown in Fig. 14, which can verify reasonable depth information area on the portrait.

## 7. Conclusion

In this paper, a method was presented towards measuring of depth perception from monocular shadow

technique. The method was implemented in a classical painting, the youngest Madonna “Nativity” in Kulenovic collection in Karlskrona Sweden. A relative depth perception model was proposed and the portrait was captured from different points of view. We showed that the shadow technique was present in the portrait and its 3D effect can be explained by phantom type of Da Vinci stereoscopic phenomena. The 3D effects could be verified and measured by two features of position changes of correspondent points and length ratio of correspondent lines. The results show that 3D effect on such classical paintings is not only a perceptual phenomenon, but also thanks to their measurement it is possible to arrange a topological map. The fully understanding of the mechanism behind the topological map generation by implementation of shadow technique can open new era for new type of 3D display.

## Acknowledgment

We would like to thank RizahKulenovic, the owner of the Kulenovic collection in Karlskrona Sweden, for providing us the opportunity to work on the youngest Madonna “Nativity”.

## References

- [1] Goldstein, E. B. (2001). Pictorial perception and art. In E. B. Goldstein (Ed.), *Blackwell Handbook of Perception* (pp. 344–378). Oxford, UK: Blackwell.
- [2] Howard, I. P., & Rogers, B. J. (2012). *Perceiving in Depth*, New York: Oxford University Press.
- [3] Nakayama, K., & Shimojo, S. (1990). Da Vinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision Research*, 30, 1811–1825.
- [4] Sternberg, R. J., & Sternberg, K. (2011). *Cognitive Psychology* (6th ed.). Wadsworth: Cengage Learning.
- [5] Goldstein, E. B. (2010). *Sensation and perception*. Wadsworth: Cengage Learning.
- [6] Willemsen, O. H., Zwart, S. T. de., Hiddink, M. G. H., Boer, D. K. G. de., & Krijin, M. P. C. M. (2007). Multi-view 3D Displays. *Proceedings of Society for Information Display* (pp. 1154–1157).
- [7] Cavanagh, P., & Leclerc, Y. G. (1989). Shape from shadows. *Journal of Experimental Psychology: Human Perception & Performance*, 15(1), 3-27.
- [8] Gillam, B., & Borsting, E. (1988). The role of monocular regions in stereoscopic displays. *Perception*, 17(5), 603–608.
- [9] Harris, J. M., & Wilcox, L. M. (2009). The role of monocularly visible regions in depth and surface perception. *Vision Research*, 49, 2666–2685.
- [10] Wardle, S. G., & Gillam, B. J. (2013). Phantom surfaces in da Vinci stereopsis. *Journal of Vision*, 13(2), 1–14.
- [11] Gillam, B., & Nakayama, K. (1999). Quantitative depth for a phantom surface can be based on cyclopean occlusion cues alone. *Vision Research*, 39, 109–112.
- [12] Liu, L., Stevenson, S. B., & Schor, C. M. (1994). Quantitative stereoscopic depth without binocular correspondence. *Nature*, 367, 66–69.
- [13] Harris, C., & Stephens, M. (1988). A combined corner and edge detector. *Proceedings of 4th Alvey Vision Conference* (pp. 147–151).
- [14] Awrangjeb, M., & Lu, G. (2008). Robust image cornerdetection based on the chord-to-point distance accumulation technique. *IEEE Trans. Multimedia*, 10(6), 1059–1072.
- [15] Brunelli, R., & Mich, O. (2001). Histograms analysis for image retrieval. *Pattern Recognition*, 34(8), 1625–1637.
- [16] Rahman, M. M., Desai, B. C., & Bhattacharya, P. (2006). A feature level fusion in similarity matching to content-based image retrieval. *Proceedings of the 9th International Conference on Information Fusion* (pp. 1–6).



**Wei Wen** was born in Jinan, China, on July 1983. He received the BS degree from the Department of Information Science and Engineering, Shandong University, China in 2006, and the MS in electronic engineering at Blekinge Tekniska Högskola, Sweden in 2009. Since 2012, he has been studying at the Department of Communication in Blekinge Tekniska Högskola as a Ph.D candidate. His current research interests include computer vision, cognitive vision, image processing, signal processing.



**Siamak khatibi** received the M.Sc. degree in electrical engineering from the Chalmers University of Technology, Gothenburg, Sweden, in 1993, and the Ph.D. degree in electrical engineering from Chalmers University of Technology, Gothenburg, Sweden, in 1999. He is currently an associate professor in computer science at Blekinge Institute of technology, Karlskrona, Sweden. He is the director of Image & Perception Lab (IPL). His research interests are in image/video processing with primary emphasis on 3D computer vision, 3D pose estimation, 3D object tracking, 3D model building, multiresolution texture analysis, image retrieval, image restoration, super resolution and high dynamic images, and image quality.