

# Depth and Perspective Perception of Flat Images in Static and Dynamic Visual Scenes

Marsel Fazlyyyakhmatov<sup>1,2,\*</sup>, Yana Nurieva<sup>1</sup>, Ildar Khafizov<sup>1</sup>, Alexander Zhegallo<sup>3,4</sup> and Vladimir Antipov<sup>1</sup>

<sup>1</sup>Kazan Federal University, Russia

<sup>2</sup>Kazan National Research Technical University named after A.N. Tupolev–KAI, Russia

<sup>3</sup>Institute of Psychology, Russian Academy of Sciences, Russia

<sup>4</sup>Moscow State University of Psychology and Education, Russia

**Abstract:** The paper shows that a sense of depth can arise from two-dimensional (2D) scenes without the presence of a stereoscopic depth signal. Experimental information was obtained on three-dimensional (3D) visual perception of 2D static and dynamic scenes. The technique is based on fixing the conditions of eye movement during the perception of two-dimensional stimulus scenes. To obtain registration of the depth perception effects, they used volume and spatial perspective of 2D images (3D phenomenon), and a binocular eye tracker. The 3D phenomenon is identified using 3D raster images. It is assumed that the comparison of eye movements during a 3D raster image viewing allows you to identify uniquely the effects of the 3D phenomenon of stimulus planar scenes displayed on the monitor screen. The first part of the work shows the conditions for the emergence of a 3D phenomenon on two plots of dynamic and static scenes. The second part demonstrates the three-dimensional attributes of dynamic scenes with the highlighting of various video components. We emphasize that dynamic and static scenes are obtained directly from TV programs. The proposed graphical and mathematical method of analysis made it possible to show qualitatively the perception of the 3D phenomenon by KFU students and revealed the features of volume observation for planar images without the occurrence of binocular disparity.

**Keywords:** Eye movement, binocular eye tracker, flat images, binocular depth perception.

## INTRODUCTION

Over the last decade, various technologies for visualizing three-dimensional (3D) scenes on displays have been technologically demonstrated and refined, among them such of stereoscopic, multi-view, integral imaging, volumetric, or holographic type. In our work, we use a binocular eye tracker to register the ability to perceive static and dynamic scenes with the effects of depth, volume, and spatial perspective based on eye movement (Fazlyyyakhmatov & Antipov, 2019).

In (Shurupova, & Krasnoperov, & Tereshchenko Latanov 2015), the authors believe that the studies of eye movements and their interaction with neurophysiological mechanisms remain relevant, and continue to develop. (Hasse, & Bruder, 2015).

Since 1838, it has been known (Wheatstone, (1838) that binocular disparity is enough to cause the perception of three-dimensional scene depth. The optical systems of the eyes direct two slightly displaced projections of three-dimensional objects onto the retinal structures. Subsequently, they are recognized as three-dimensional images in the neural networks of the brain

(Howard, & Rogers, 1995). Howard, 2012). If two identical projections (for example, from a painting) fall on the retina, then visual perception uses monocular signs of perspective. But they do not create all the elements inherent in the perception of 3D scenes. (Rauschenbach, 2001). It offers a way out - stereograms or holography.

Reviews (Kowler, 2011). (Schütz, Braun, & Gegenfurtner, 2011). Reflect the interest and various fields of application, and eye movement features. The appearance of binocular eye trackers greatly simplified the methodology for studying the processes of eye movement. They are used to study binocular coordination (Bucci & Kapoula 2006), (Liversedge, Rayner, White, Findlay, & McSorley, 2006). The duration of fixation (Kliegl, Nuthmann, & Engbert, 2006), etc.

It is known that eye movements are controlled by both visual stimulation and “cognitive” factors such as attention, expectation, memory and voluntary control (Ringach, Hawken, & Shapley, 1996). . It was argued in (Iijima, Komagata, Kiryu, Bando, & Hasegawa, 2012). that a sense of depth can arise from two-dimensional (2D) films without a stereoscopic depth signal. The depth of perception in three-dimensional (3D) space depends on stereoscopic view stability along the

\*Address correspondence to this author at the Kazan Federal University, Russia; E-mail: mfazlyjy@kpfu.ru

vergence, compensating for the displacement of the retinal image from two eyes (i.e., binocular disparity) (Angelaki, & Hess, 2005). On the other hand, the oculomotor mechanisms that stabilize the stereoscopic look and depth perception of 2D space in the cinema remain unclear. (Howard, 2008).

In this paper, we studied the three-dimensional perception of two-dimensional scenes recorded from a TV screen.

## MATERIAL AND METHODS

Information processing is carried out upon receipt of the difference  $\Delta X = X(R) - X(L)$  and the development of histograms of numerical value difference during the record time  $\Delta t$ . By the sign  $\Delta X$ , they determine the focal planes of stimulus image perception relative to the location of the monitor. The principle of calculating the distance to the focusing planes is presented in (Antipov, Zhegallo, Galimullin, & Fazlyyyakhmatov, 2018), (Antipov, Popov, Fazlyyyakhmatov, & Ustin, 2019).

The purpose of the article is to determine the results of three-dimensional perception of static and dynamic scenes obtained during eye movement record of first-year students at the KFU Engineering Institute. To compare the eye movements of students during the perception of stimulus images, the comparison was performed with similar characteristics of an experienced subject (An-2, the author of this work). We used the TheEyeTribe binocular portable eye tracker (Denmark) (record frequency 60 Hz). Eye movement record time  $\Delta t = 60$  s.

Eye movement recordings were conducted for three dynamic scenes selected during watching conventional TV programs by An-2. The criterion for program selection is his perception of depth and volume of subjective effects.

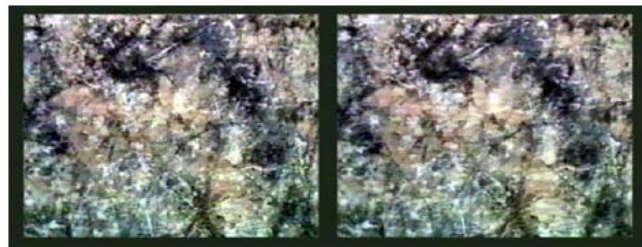
The first plot was chosen from the picture "The Bodyguard" (1992, directed by Mick Jackson): the foreground shows the movement of the water surface, which creates the effect of perspective and volume (Figure 1).



**Figure 1:** Stereopair - plot 1.

To observe the depth in Figure 1, it is necessary to focus the eyes on the point in front of the stereo pair and get three projections. The state of stereoscopic depth observation occurs on the middle projection.

The second plot is the fragments of a documentary about the American artist Jackson Pollock, which shows his painting "Number 1, 1950 (Lavender Mist)" (<http://www.ibiblio.org/wm/paint/auth/pollock/lavender-mist/pollock.lavender-mist.jpg>) (Figure 2). The screen had volume and depth effects for An-2.



**Figure 2:** The stereo pair based on the painting by D. Pollock "Lavender Fog" - plot 2.

The third plot is selected from the programs of the channel "Culture", the plot is about the weather forecast according to the picture "The Starry Night" (1889) by Vincent van Gogh (<https://www.moma.org/collection/works/79802>) (Figure 3).



**Figure 3:** The stereo pair based on "The Starry Night" by Van Gogh - plot 3.

According to plot 3, the eye movement of the survey participants was obtained for three modifications (fragments) of the video. The first fragment is the record demonstration during the program. The features of the fragment consisted of the fact that the movement of the images on the screen occurred with horizontal and vertical movements. The second fragment - the records were highlighted in which the images moved only horizontally. And the third modification - the part of the weather forecast was used with the demonstration of three-dimensional modeling of the painting "The Starry Night".

Three versions of eye movement records were obtained for plots 1 and 2. The first is viewing a

dynamic scene video on the monitor screen. Further the record during one static frame selection. Then a 3D raster image installation in front of the monitor and eye movement record when it is viewed. Negative  $\Delta X$  values indicate the perception of depth between the eyes and the location of the stimulus image. Positive values of  $\Delta X$  are the readings of depth perception beyond the stimulus image plane.

Verbal consent for interviewing was obtained from all subjects.

## RESULTS AND DISCUSSION

Analysis of the studies was carried out according to the following features:

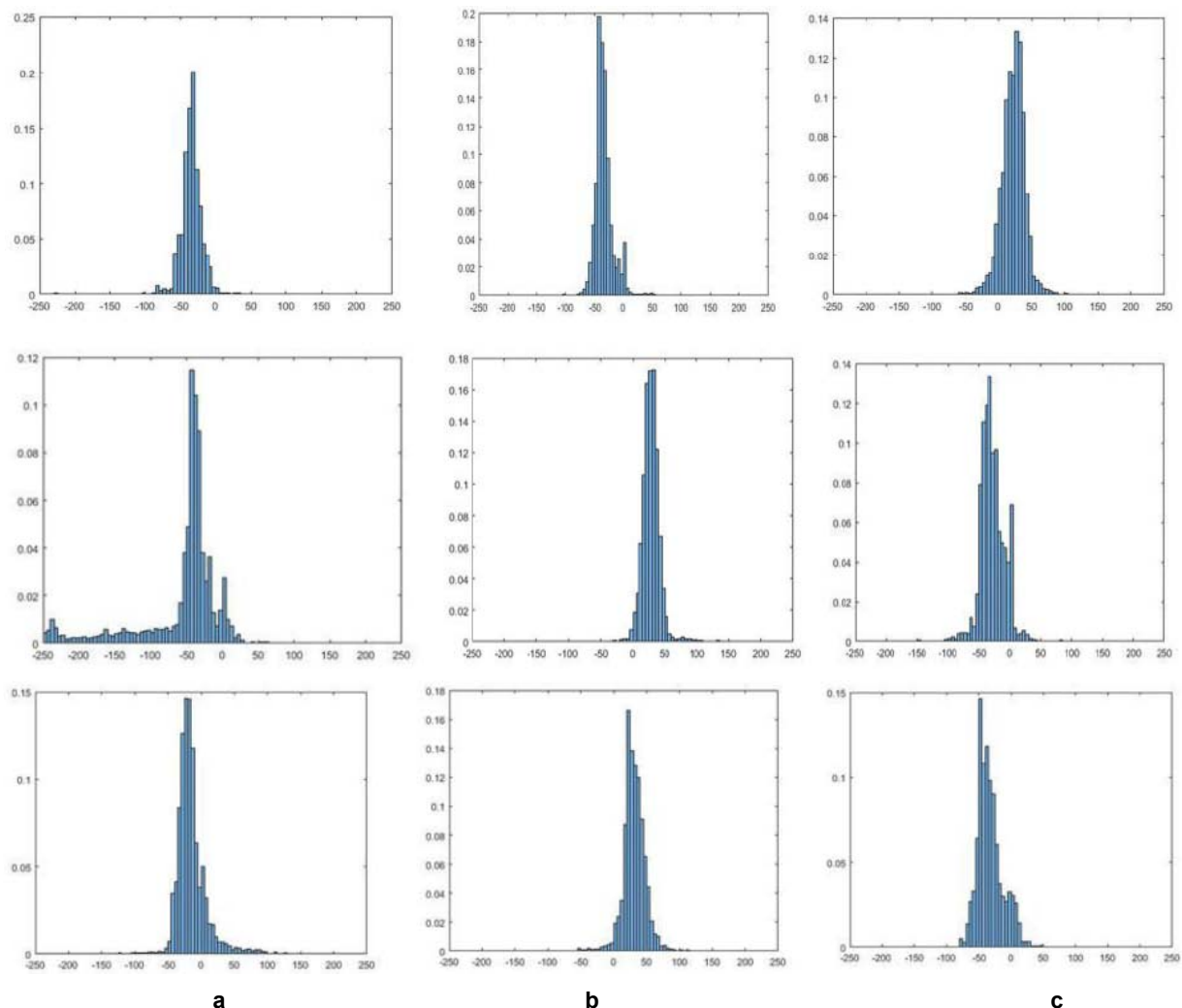
1. Determination of the conditions for matching eye movement records obtained during registration under dynamic and static conditions and during observation of a 3D raster image depth;

comparison of the difference histogram contours provided that dynamic and static scenes are shown.

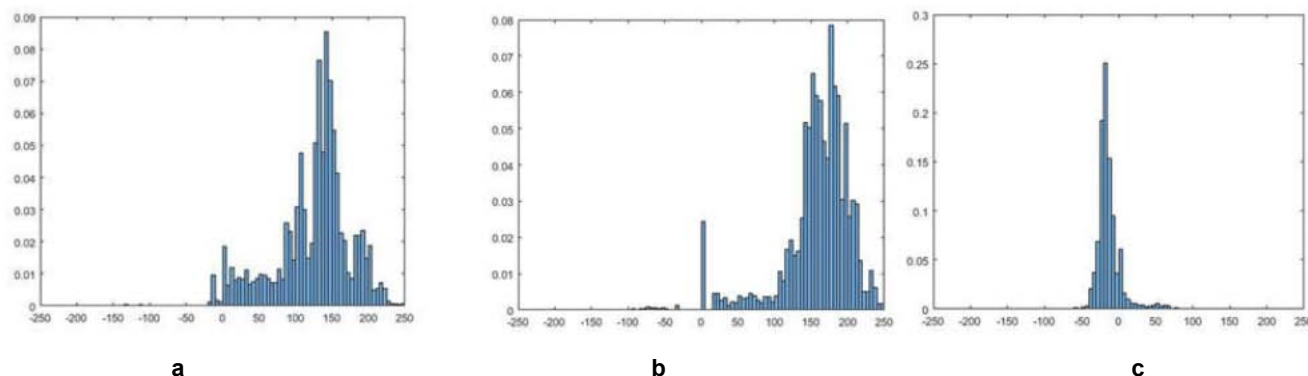
2. Determination of the difference histogram contour range on the  $\Delta X$  axis in dynamic and static scenes. Thus, we obtain the depth perception relative to the displayed stimulus scene.
3. Registration of the range of values of the difference histogram contour width.
4. Determination of obtained value range of the difference histograms during the perception of the raster 3D image depth.

### Plot 1

The survey involved 17 students. Perception (with the movement of the water surface) during a dynamic



**Figure 4:** Histograms: a), b) - two students; c) An-2.



**Figure 5:** Histograms of students perceiving depth in different positions relative to the raster image depth.

plot display corresponds to the 3D raster projection perception. Figure 4 shows histograms. The vertical rows of Figure 4 are the difference of histograms of three people. On horizontal rows: the top row is a dynamic plot; medium - statics; lower - the raster perception.

Some students perceive all three record options either in front of the screen (Figure 4-a) or behind the screen plane (Figure 4-b). At the same time, the histogram location of the 3D raster image depth perception (lower horizontal row) identifies the depth perception.

The histograms of perceptual difference in front of stimulus images were obtained (Figure 4-a) for seven

Students, - this is 41%. Similar histograms were recorded for an experienced researcher (Figure 4-c). 3 students (or  $\approx 18\%$ ) perceive behind the monitor plane (Figure 4-b). The comparison of the difference histograms in the conditions of static and dynamic plots with the perception of the raster image depth proves the general perception of depth for 59% of the survey participants. For the rest of the students, histograms were obtained with the center either in the region of zero readings of  $\Delta X$  or on opposite sides of zero values.

There are the records when the dynamic and static plots do not coincide with the location of the difference histogram relative to  $\Delta X \approx 0$  axis. For example, a student perceives depth behind the monitor plane under dynamic (Figure 5-a) and static display (Figure 5-b). And he perceives the depth of the raster image in front of the screen (Figure 5-c). Note that 11 students observe the depth similar to the histograms of Figure 5. In other words, some students perceive the plot depth, but it does not coincide with the depth of the displayed raster image.

A student perceives the perspective of spatial construction of images. And the depth perception is much greater for this student than the applied disparity effects for a 3D raster image creation (Figure 5).

## Plot 2

The survey involved 34 students. First, they showed the dynamic version of the plot. Then the last frame is static. In the end, a 3D raster was installed in front of the screen.

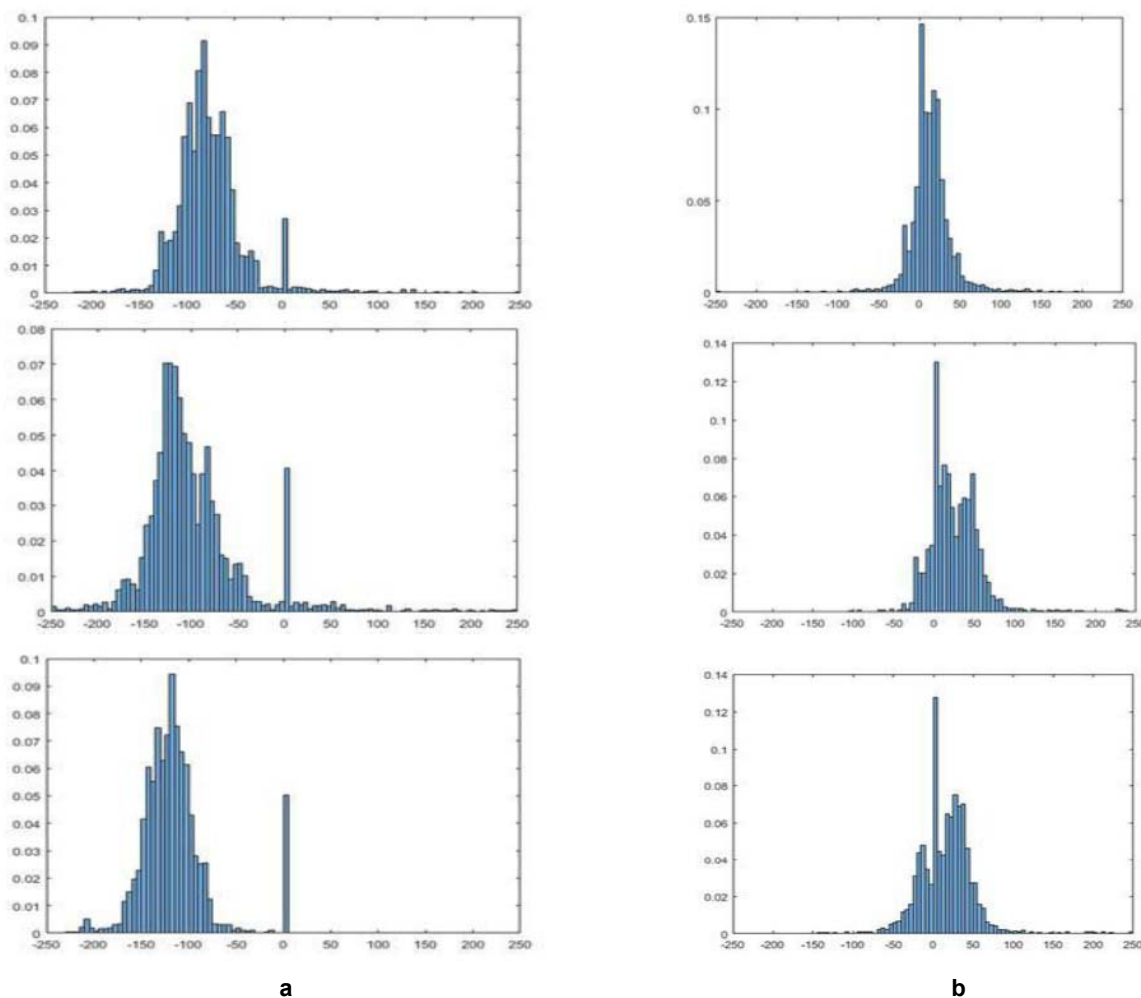
The different histograms of all students show that they record the effects of depth. There is no unequivocal location of the difference histograms relative to the location plane of the stimulus images.

On Figure 6, the upper histograms are obtained under the conditions of a dynamic plot, the middle ones are the static states of perception. The lower histograms are the perception of the 3D raster image depth.

The difference histograms for 10 students show that the record perception (dynamics - static - raster) occurs between the eyes and stimulus images (Figure 6-a). These are negative  $\Delta X$  values. For one student, the histogram is in the region of positive  $\Delta X$  values (Figure 6-b).

The difference histograms in the combination of dynamics-statics were recorded for 9 students on one side of zero  $\Delta X$  values (negative levels).

The histograms of the first vertical row of Figure 6 demonstrate that depth perception occurs between stimulus images and eyes. In the second vertical row, the maximum of the difference histogram contour is located in the region of positive  $\Delta X$  values, the left-wing of which also covers negative  $\Delta X$  values. For this student, the perception of depth is observed behind the



**Figure 6:** Histograms of two students (vertical rows).

plane of the stimulus image, although partially in front of it. Vertical lines in the region of zero  $\Delta X$  values indicate that there are elements of planar components in the perception.

### Dynamic Scenes

Histograms showing the perception of depth between the monitor screen and the eyes were obtained for 22 students (Figure 7-a; b; c; d). The histograms of perception behind the screen plane were revealed for 6 students (Figure 7-f). The histograms covering the range including positive and negative  $\Delta X$  values are observed for 6 students (Figure 7-e).

### Static Scenes

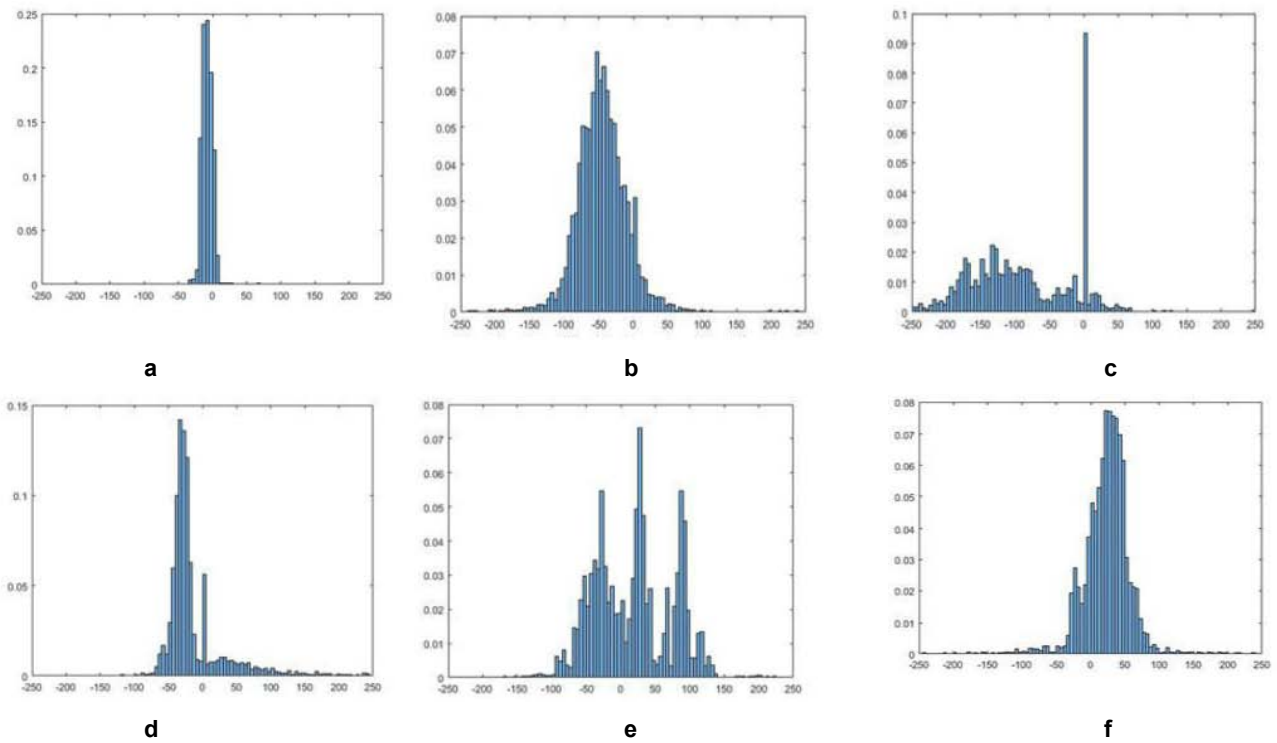
The histograms showing the perception of depth between the eyes and the monitor screen were obtained for 22 students (Figure 8-a; c). 6 students perceive the depth beyond the monitor plane (Figure 8-d). And the maximum of the difference histogram in the

region of zero  $\Delta X$  values was obtained for 6 students (Figure 8-b). The vertical lines  $\Delta X$  in Figure 8-c) and Figure 8-d) is the indicator of the scene planar perception elements.

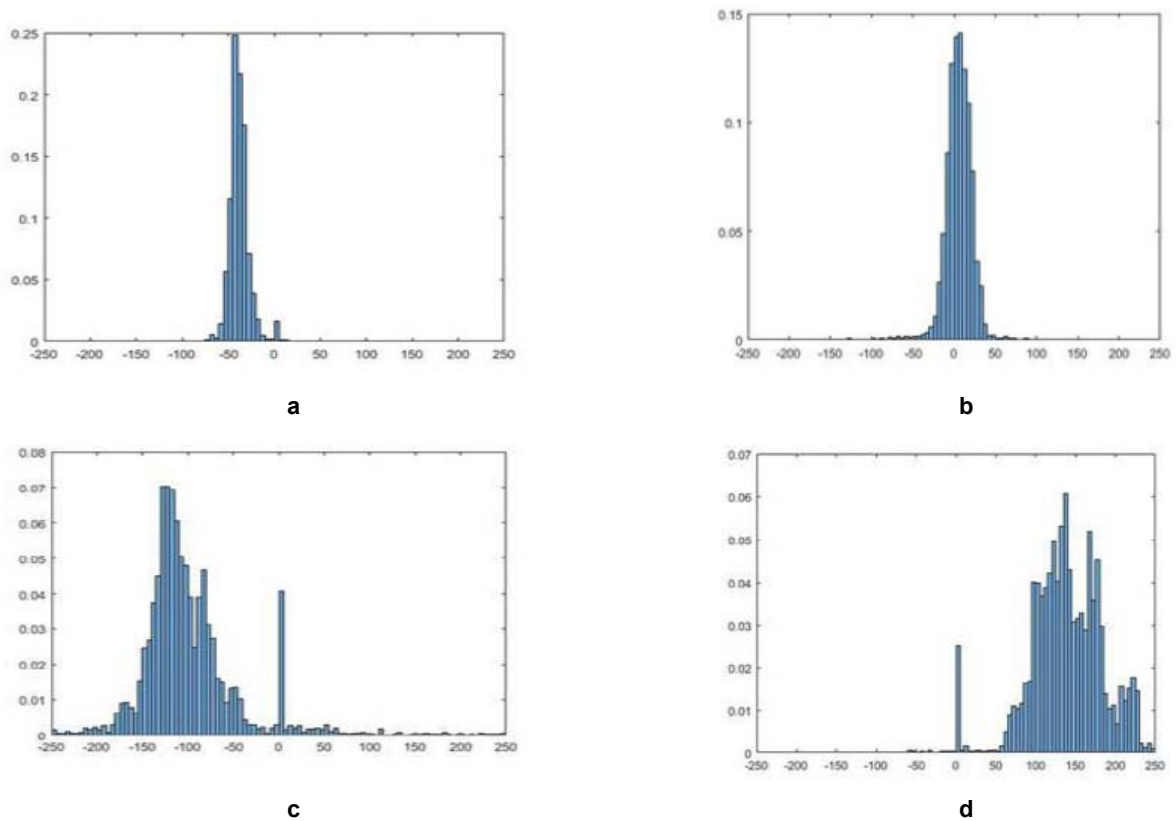
The minimum values of the difference histogram contour width are shown in Figure 8-a, the maximum values and other similar readings are shown in Figure 8-d.

The variants of eye movement during the perception of raster images show that 15 students perceive the depth between the eyes and the raster plate (Figure 9-a and 9-c). 13 students perceive the raster plane (Figure 9-b). The maximum of the difference histogram contour in the region of the zero values  $\Delta X$  range was recorded for 5 students (Figure 9-d). Note that the raster is mounted so that the effects of depth are observed as 15 students perceived it - Figure 9-a and 9-c. The perception of students according to Figure 9-b variant shows that they perceive the inverse state of depth.





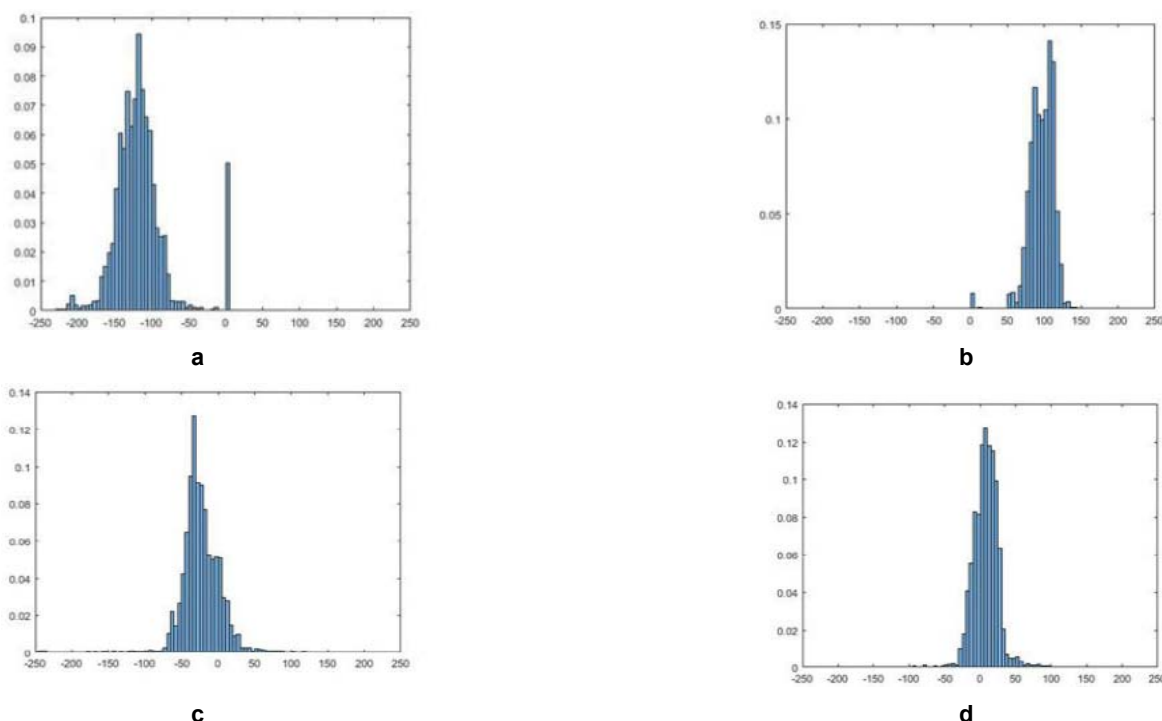
**Figure 7:** Histograms in the conditions of dynamic scene observation.



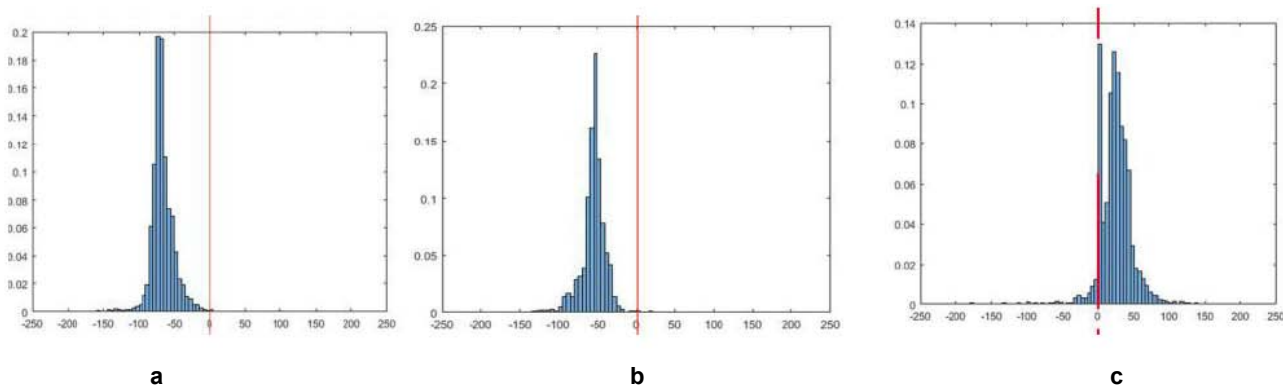
**Figure 8:** Histograms of perception in a static scene.

For example, how can they perceive the depth of a stereo pair if they focus the eye behind the plane of Figure 2? With this orientation of the eyes, inverse

conditions for the color gamut perception occur. The difference histograms "An-2" are shown in Figure 10.



**Figure 9:** Histograms during a 3D raster stimulus image perception.



**Figure 10:** An-2 histograms: **a)** in a dynamic scene; **b)** in a static frame; **c)** during a raster image perception.

Figure 10 shows that depth perception occurs in front of the screen of all stimulus images. Moreover, the location of the depth perception of the 3D raster image coincides with dynamic and static scenes. Similar locations of the difference histograms are registered for 10 students.

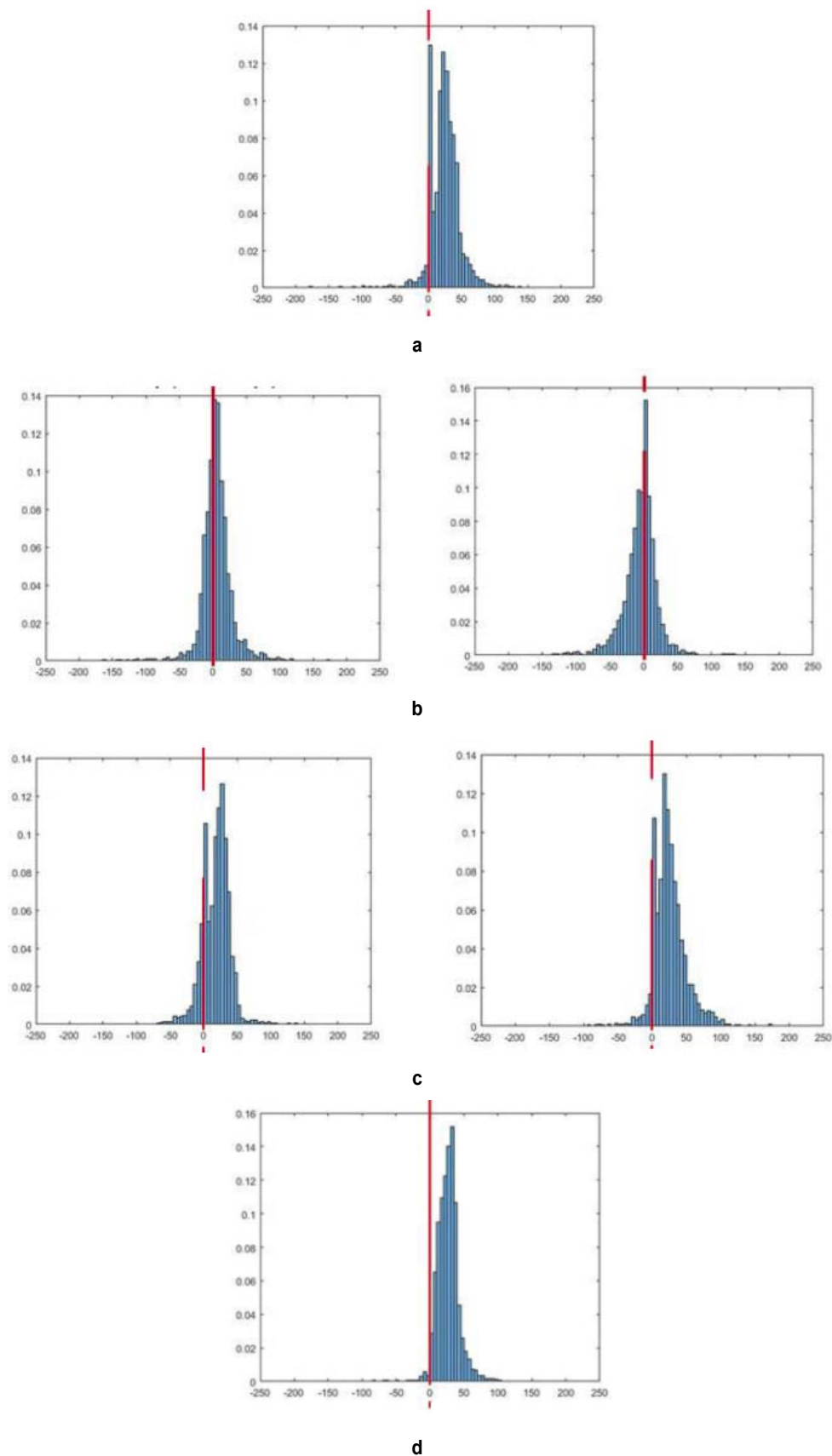
### Plot 3

The survey involved 12 students for two previously presented plots, the eye records of an experienced researcher (An-2) were similar  $\approx 40\%$  to student records. Therefore, first of all, we will study the various options of the "Weather" plot demonstration for him.

The eye movement was recorded for three dynamic plot options. The first is a direct record of painting

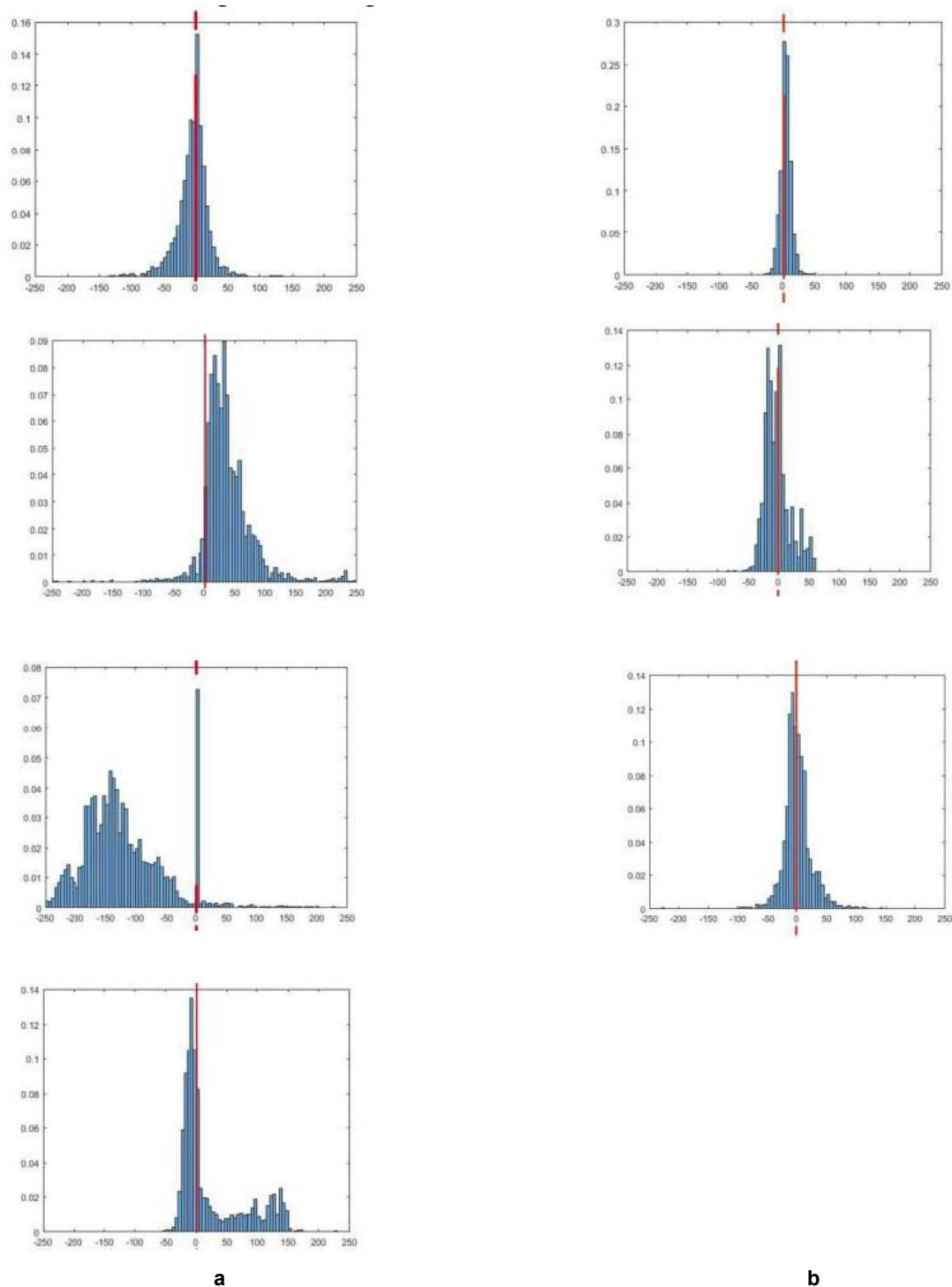
image movement demonstration. Second: fragments were excluded from the record in which the picture objects moved vertically. Only those images remained for which there was modeling in the horizontal direction. The third option - the picture was used in the demonstration (Figure 3) - the objects of the picture moved only horizontally in the video.

Figure 11 shows the difference histograms for An-2 and two students. The left vertical column is An-2 eye movement histogram: a) - dynamic option 1; b) dynamic option 2; c) - dynamic option 3; d) - a raster perceiving histogram. The histograms of eye movements of two students similar to the difference histograms of An-2 are on the right vertical column.



**Figure 11:** Perception histograms of three plot options: the left column - by the researcher An 2; the right column - by two students.





**Figure 12:** The variety of dynamic depth plot perception by students in the "format" of difference histograms: a) The plot "weather 2"; b) "weather 3".

Two plot options are shown for students: the second - b) and the third - c).

Let's compare the different histograms in Figure 11-c) and Figure 11-d) for An-2. The difference histograms show that, first of all, the histograms are located in the general range of  $\Delta X$  values. We conclude that the perception of the depth and image perspective is

reliable under the conditions of the dynamic mode of plot 3 demonstration. Secondly, the images are perceived behind the stimulus image location (Figure 11-a; c).

Some histograms of depth perception by students are shown in Figure 12. Let's pay attention to the histogram of Figure 12-a (in the lower projection). The

maximum is observed in front of the screen. All that is in the region of positive  $\Delta X$  values is the perception of the image perspective. The elements of perspective are also present on the second projection from above.

## CONCLUSION

Motion parallax, the perception of depth resulting from an observer's self-movement, has almost always been studied with random dot textures in simplified orthographic rendering. A visual comparison of the difference histograms in terms of dynamic and static plots does not allow one to determine the advantages of the image depth and location perception relative to the location of stimulus images.

The generalized qualitative analysis of the presented histograms about static and dynamic plots of spatial effect perception has no uniform laws for all students. The results obtained do not contradict the assumptions of the article authors (Ringach, Hawken, & Shapley, 1996).

The paper presents the graphical and mathematical method to study and identify the perception of depth, volume, and spatial perspective of planar images. Three plots recorded from TV shows showed that eye movement record on a binocular IT tracker allows you to explore three-dimensional features of planar image perception. The surveys conducted among students confirm the previously presented conditions for the possibility of planar projection conversion into three-dimensional plots obtained by An-2 (Antipov & images in a computerized environment", *Experimental Psychology*, 2014). It is shown that the effects of depth and spatial perspective do not have a clear perception for the entire group of surveyed students.

In the future, it is planned to draw up a program to compare the difference histograms, to quantify the location of the focus planes relative to the location of the stimulus images. It will reveal the difference in-depth perception in terms of static and dynamic scenes.

## ACKNOWLEDGEMENTS

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

## REFERENCES

- Angelaki, D. E., & Hess, B. J. (2005). Self-motion-induced eye movements: effects on visual acuity and navigation. *Nature Reviews Neuroscience*, 6(12), 966-976. <https://doi.org/10.1038/nrn1804>
- Antipov V.N & images in a computerized environment", *Experimental Psychology*, V.(2014) 7, No. 3, pp. 97–111.
- Antipov, V., Popov, L., Fazlyyyakhmatov, M., & Ustin, P. (2019, May). Cognitive Three-dimensional Perception of Planar Images of the Person in Its Educational Activity. In *2019 International Conference on Pedagogy, Communication and Sociology (ICPCS 2019)*. Atlantis Press. <https://doi.org/10.2991/icpcs-19.2019.6>
- Antipov, V., Zhegallo, A., Galimullin, D., & Fazlyyyakhmatov, M. (2018). The experimental study of the visual perception of depth the flat images, eye movement registration. *Psychology. Journal of Higher School of Economics*, 15(2), 384-399.
- Bucci M.P. & Kapoula Z.A.A. (2006), "Binocular coordination of saccades in 7 years old children in single word reading and target fixation", *Vision Research*, vol. 46, pp. 457–466. <https://doi.org/10.1016/j.visres.2005.05.004>
- Fazlyyyakhmatov M.G., Antipov V.N. (2019). "Creative and Cognitive Perception of the Planar Images", *International Journal of Educational Sciences*, vol. 27, № 1-3, pp. 70-77. <https://doi.org/10.31901/24566322.2019/27.1-3.1105>
- Hasse, C., & Bruder, C. (2015). Eye-tracking measurements and their link to a normative model of monitoring behaviour. *Ergonomics*, 58(3), 355-367. <https://doi.org/10.1080/00140139.2014.967310>
- Howard, I. (2008). Vergence modulation as a cue to movement in depth. *Spatial Vision*, 21(6), 581-592. <https://doi.org/10.1163/156856808786451417>
- Howard, I. P. (2012). *Perceiving in depth, volume 1: Basic mechanisms*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199764143.001.0001>
- Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. Oxford University Press, USA. <https://doi.org/10.1093/acprof:oso/9780195084764.001.0001>
- Iijima, A., Komagata, S., Kiryu, T., Bando, T., & Hasegawa, I. (2012). Vergence eye movements signifying 3D depth perception from 2D movies. *Displays*, 33(2), 91-97. <https://doi.org/10.1016/j.displa.2011.11.001>
- Kirkby, J. A., Blythe, H. I., Benson, V., & Liversedge, S. P. (2010). Binocular coordination during scanning of simple dot stimuli. *Vision Research*, 50(2), 171-180. <https://doi.org/10.1016/j.visres.2009.11.008>
- Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of experimental psychology: General*, 135(1), 12. <https://doi.org/10.1037/0096-3445.135.1.12>
- Kowler, E. (2011). Eye movements: The past 25 years. *Vision research*, 51(13), 1457-1483. <https://doi.org/10.1016/j.visres.2010.12.014>
- Liversedge, S. P., Rayner, K., White, S. J., Findlay, J. M., & McSorley, E. (2006). Binocular coordination of the eyes during reading. *Current Biology*, 16(17), 1726-1729. <https://doi.org/10.1016/j.cub.2006.07.051>
- Rauschenbach B. (2001). "Geometry of the picture and visual perception", St. Petersburg. ABC classic, 320 p.
- Ringach, D. L., Hawken, M. J., & Shapley, R. (1996). Binocular eye movements caused by the perception of three-dimensional structure from motion. *Vision research*, 36(10), 1479-1492. [https://doi.org/10.1016/0042-6989\(95\)00285-5](https://doi.org/10.1016/0042-6989(95)00285-5)
- Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of vision*, 11(5), 9-9. <https://doi.org/10.1167/11.5.9>
- Shurupova M.A. & Krasnoperov A.V. & Tereshchenko L.V., Latanov A.V. (2015) "The influence of cognitive task on the

parameters of eye movement during static and dynamic scene observation". Eye tracking in psychological science and practice. Ed. in charge: V.A. Barabanshikov. - M.: Kogito-Center, pp. 202-212.

Wheatstone C.(1838) , "Contributions to the physiology of vision: part

the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision", *Philos Trans R Soc Lond*, vol. 128, pp. 371–394.  
<https://doi.org/10.1098/rstl.1838.0019>