IEEE Second International Conference on Data Stream Mining & Processing
August 21-25, 2018, Lviv, Ukraine

Bioinspired Approaches to the Selection and Processing of Video Information

Vitaliy Boyun
Department of Intelligent Real Time Video Systems
V.M.Glushkov Institute of Cybernetics NASU
Kyiv, Ukraine
vboyun@gmail.com

***Abstract*— Significant expansion of the range of applications of real-time video systems requires further improvement in their productivity, efficiency and intelligence. Therefore, researchers are increasingly turning to the human eye analyzer as a prototype to create more sophisticated systems of technical vision. The paper proposes a number of approaches and methods for the selection and processing of video information inspired by the human visual analyzer. In particular: the method of hierarchical selective perception of video information; dynamic models of processes for finding objects, tracking them, panning the scene and the mechanisms of attention and allocation of the essence, which, by managing the parameters of reading information from a video sensor, provide reading of a part of the image relevant to the task; information measure of the dynamic image (δ-entropy), which characterizes its spatial frequencies and is an effective information feature for the search and recognition of objects; methods of expanding the dynamic range of perception of brightness; the principles of circular organization neurons of the central fovea, which provide increased contrast and the allocation of informative features; circular organization of the retinal neurons with a sumation of signal sticks that contribute to increased sensitivity in conditions of insufficient lighting; specialization of neurons and organization multilayer neural network.**

Keywords*— *video system, human visual analyzer, bio-inspired approaches, relevant information, neural network

I. INTRODUCTION

The tasks that arise in transport, industry, robotics, medical-biological research, defense-military sphere, etc., require further increase of productivity, efficiency and level of intelligence of computer video systems. Therefore, researchers are increasingly turning to the visual system of man, as the most perfect prototype for the construction of computer vision systems.

The human eye system has been improving for millions of years and has reached an extremely high level of organization. A generalized model of the human visual system is multifunctional and consists of several dozen or even hundreds of local models that describe a range of structural, physical, geometric and psychophysical mechanisms and processes. The process of perceiving visual information by a person is dynamic, with many parameters that change in the process of perception, with many feedback bonds. We not only see, but we react, that is, such process is active. Therefore, the phenomenon of vision provides a lot of versatile elegant solutions for computer vision systems.

Given the perfection of the human visual analyzer, it is advisable to study and distinguish its elements for use in modern technical systems. It is not necessary to exactly copy them, but, conversely, understanding their functioning, implement them taking into account the statement of a specific technical problem and the capabilities of the level of technology.

II. ORGANIZATION OF THE HUMAN VISUAL SYSTEM

Approaches to the analysis of perception and processing of information are based on ascending and descending processes. Ascending processes, or processes for the transfer of information, begin with simple (basic) elements - discrete sensory, derived from sensory receptors. These basic features include: the difference in the luminosity of fragments, spatial frequencies, or the position of elements in space. Incoming touch information is transmitted from the base (lower) level to higher and integrative levels. In this case, the visual system carries out the design and creation of identifiable patterns by combining the basic elements under the influence of the reflex mechanisms of the visual system and the brain, not controlled by human.

Descending processes, or processes for the conceptualization of information data, use global, abstract, and higher levels of analysis for implementing processes at lower levels. These processes of perception of the form are based on the knowledge previously obtained by the observer, his previous experience, comprehension and interpretation, as well as his expectations.

Both descending and ascending processes are a form of manifestation of the activity of the visual system and the brain, and in most cases they are performed jointly, complementing each other.

Sticks and retinal cones perceive the image of the surrounding world as a set of individual points, although this world consists of separate, differentiated objects and surfaces that have a definite shape and outlines. For such a generalized perception of the world, the efforts of the multi-level neuronal network of the retina and higher levels of the visual system are being applied. Despite the huge streams of video information, the human visual-analyzing system copes with them due to its extremely high selectivity, which ensures the selection of only relevant information for specific conditions with the help of a multi-level neural network, which from level to level increases the abstract presentation of information.

Much of the pre-processing of visual information is already at the retina level. The most important role in our perception of form, edges and boundaries of the regions is the contours that appear when the adjacent surfaces are illuminated differently (that is, taking into account the intensity and color). The contrast of the surfaces, acting on the visual receptors, causes their interaction and creates conditions for perception of contours by the visual system. Ganglion cells of the retina with their lateral bonds and mechanisms of excitation and lateral inhibition are of great importance in the perception of contours.

Changes in the brightness (color) of the image on the retina sticks cause the excitation of the neurons (the mechanism of attention), which controls the rapid movements of the eye (saccades). Excited areas of the image are consistently (with priority) reflexively transmitted to the central fovea for detailed consideration and allocation of local informative features (ascending processes of control). The central fovea and periphery of the retina are organized according to the ring principle. However, in the central fossa, each cone has a gateway to the ganglion cell, and on the peripheral retina the sticks are grouped together, summing up signals from larger regions of the receptor field and providing increased sensitivity in low light (in exchange for a decrease in spatial resolution).

Besides to the roughly-accurate perception of visual information on space (i.e., in X, Y coordinates), the human eye reacts not to the amount of luminance or chromaticity in the image, but to changes between the luminance values of neighboring receptors, or the luminance values of a given receptor in time, that is, on dynamics of this parameter.

An arbitrary part field of view with contrasting light and dark areas can be analyzed and transformed into its spatial frequency - the number of variations of luminosity in a certain area of space, or the number of cycles of alternation of dark and light bands in a given field of view. It is proved that in the visual system there are detectors of spatial frequency - specialized cells, are most sensitive to certain spatial frequencies.

Thus, spatial frequencies are a simple and reliable way to describe and generalize the structural details of various visual objects.

The human eye system provides an extremely high range of light perception (10^{10}), which is ensured by the logarithmic perception of brightness, high sensitivity of the sticks and the lower sensitivity of the cones, as well as a number of organizational principles of the neural network of the retina. In particular, this is facilitated by: summing signals from the sticks of the peripheral retina to increase sensitivity in conditions of insufficient illumination, circular organization of neurons of peripheral retina and of central fovea , and the like.

From the ganglion cells of the central fossa and the peripheral retina, a whole series of dynamic images is sent to the corpus geniculatum laterale, each of which displays only one aspect of the overall visual picture. Each video stream is transmitted over its group of optic nerve fibers. In particular, streams of local features (orientational, colors, movement, etc.) are transmitted to the corpus geniculatum laterale.

The organization of the corpus geniculatum laterale is similar to the ring organization of the retina, but it covers and analyzes large portions of the receptor field.

From the corpus geniculatum laterale, signs of a higher level enter the visual cortex, which has a line organization (lines, bands, rectangles are allocated, their length, width, orientation is determined), that is, signs of a higher level of abstraction are formed - signs of the essence of the image. These signs in the human brain are compared with models that were acquired from human experience. They control the conscious movements of the eyes (downward control processes), determined by the cognitive process of perception of information. In this case, the brain, which has a complete retina model and acquired experience in perceiving images of objects, receiving information from the retina and comparing it with the models of objects, in accordance with the goal and characteristics of objects determines high-frequency automatic modes of controlling eye movements and directs low-frequency regimes [1–4].

Thus, the use of functions, principles of construction and adaptive mechanisms of the human visual analyzer will contribute to the creation of computer vision systems of the new generation.

III. Use of Elements of the Visual Analyzer in the System of Computer Vision

The most important features of the human visual analyzer for computer vision systems are high selectivity of perception of video information and wide parallelization of information processing on layers of neurons of the retina and higher levels of the brain.

Taking into account the knowledge of high-level control of the movements of the eye of the visual system, the Institute of Cybernetics of the National Academy of Sciences of Ukraine developed dynamic models of processes for finding objects in the image, tracking them, panning, etc., which contribute to the selection of relevant information, greatly reducing the redundancy of its presentation. In particular, unlike pyramidal perception Burt [5], the method of ***hierarchical selective perception*** [6-8] is proposed, which, by analogy with the human visual analyzer, is based on reading a scene with a low resolution (an analogue of the peripheral retina) for a quick search of object of the given features (the mechanism of attention) and the next reading (analog of the saccades) with high resolution (analogue of the central fovea) of the part of the image with the object for further consideration, measurement, recognition. This can significantly reduce the amount of information processed, increase the efficiency and effectiveness of systems of technical vision.

Dynamic models of object tracking processes are based on the sequential reading of the location of an object, taking into account the direction and speed of its movement, and possibly also changes in overall dimensions (analogue - follow-up eye movements). Dynamic models of panning processes (analogous to processes of eye movement or head rotation) are based on reading an additional image that appears due to these movements, and "pasting" it with the previous image.

To determine the amount of information in the video sequence, a potential estimate based on the amplitude-spatial and temporal resolution is usually used

$$C_{в.п.} = \frac{X}{\Delta x} \cdot \frac{Y}{\Delta y} \cdot \log_2 \left(\frac{Z}{\delta z} + 1 \right) \frac{1}{\Delta t},$$

where X and Y are the size of the image field; Z is the brightness coordinate of the image; Δx, Δy, δz, Δt – the discreteness of representing the corresponding coordinates of the image.

The values of X, Y, and Z in the formula are usually taken to be fixed and equal to the maximum value; the values of Dx, Dy, dz, Dt are also fixed, so this approach gives an upper estimate of the amount of information that is very overestimated.

The considered dynamic models, by changing the parameters in the formula, allow you to allocate useful (dynamic) information relevant to the mode of perception, greatly reducing the redundancy of the representation of the image and the cost of its transmission and processing. These dynamic models have become the basis for the development of methods for dynamically managing the parameters of reading information from a video sensor. Modern K-MON-video sensors have in their composition, in addition to the sensor matrix, several hundred registers for adjusting the readout parameters of the sensor matrix and up to ten specialized processors for the preliminary technological preparation of the image prior to use. Effective realization of these possibilities ensures time alignment of the processes of input and processing of information, as a result of which there is an opportunity to obtain parameters for managing the reading of the next frame with minimal delay information after the processing of the current frame [9].

Since the human eye reacts not to the amount of luminance or chromaticity in the image, but to changes between the luminance values of neighboring receptors, or the luminance values of a given receptor in time, that is, to the dynamics of this parameter, as the dynamics of the image are proposed, it is proposed to select the concepts of δ-entropy, the average value of the derivative of the rows and columns of the image [6,10].

In the discrete form δ-entropy the image is defined as

$$H_\delta = \frac{1}{MN} \sum_{j=1}^{M} \sum_{i=1}^{N} \left(\left| \frac{\Delta Z_{ij}}{\delta z} \right| \right),$$

where M, N - the size of the image field,

ΔZ – the differences between the brightness of the pixels in rows or columns.

In contrast to the method of averaging the transverse sections of the image brightness profile by rows and columns [11] to determine the location of the object with the background in the image, it is suggested to use δ-entropy. That is, averaging modulo transverse slices of the image with brightness differences between adjacent pixels in rows and columns [9]:

$$H\delta_i = \frac{1}{M} \sum_{j=1}^{M} \left| \frac{\Delta Z_{ij}}{\delta z} \right|; \qquad H\delta_j = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{\Delta Z_{ij}}{\delta z} \right|.$$

Such a dynamic amount of information can be effectively used to segment the image on a high and low dynamic domain, segmentation of textual information in the image, search and classification of textures (Fig.1), search of car numbers (Fig.2), bar codes, DMX codes (Fig.3), fingerprints, character recognition (Fig.4), control of the shooting frequency of the camcorder, and the like.

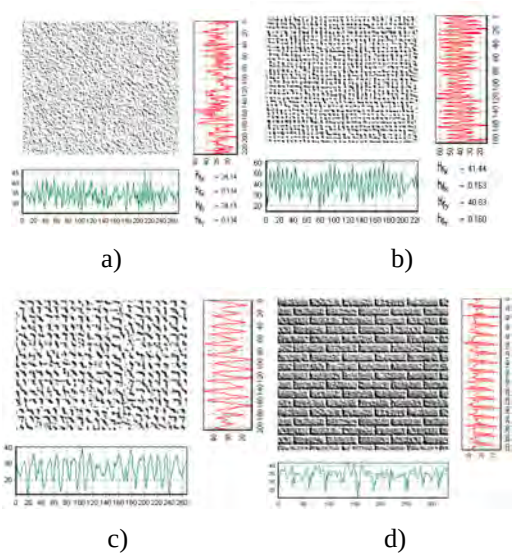


Fig. 1. Evaluation of texture parameters: a) sand, b) linen, c) sacking, d) brickwork.

The δ-entropy makes it easier to compare of the texture, the size of the grain with respect to spatial frequencies, the contrast of the amplitude of the differences, to reveal defects in texture images (violation of regularity, Fig. 1,c).

Fig. 2. Searching of the car number in the car image.

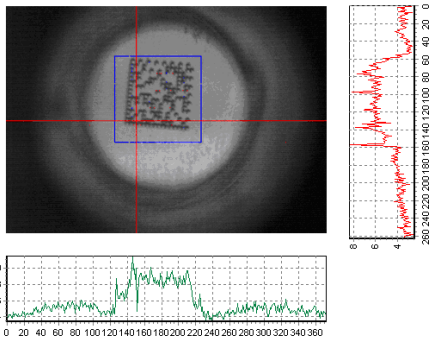
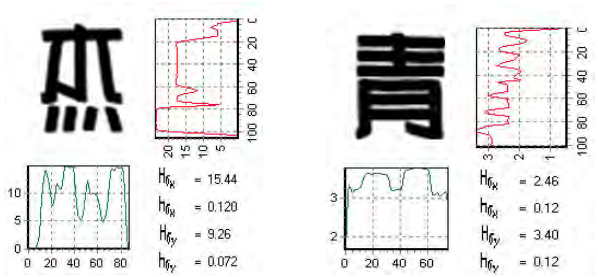
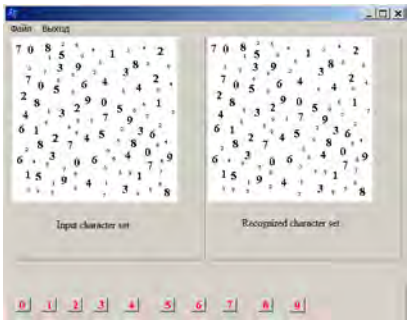


Fig. 3. Searching of the DMX-code in the micro image.

For the recognition of symbols (Fig. 4), a modification of the concept of δ-entropy is used, in particular, the number of swings from "0" to "1" and vice versa. In this case, for noisy images, it is advisable to enter a check on the threshold of the values of the differences or to carry out a low-frequency filtration of the image. It will be effective to use the integral value of δ-entropy in rows and columns as a characteristic vector, or even their sum, that is, one value of the characteristic.



a)



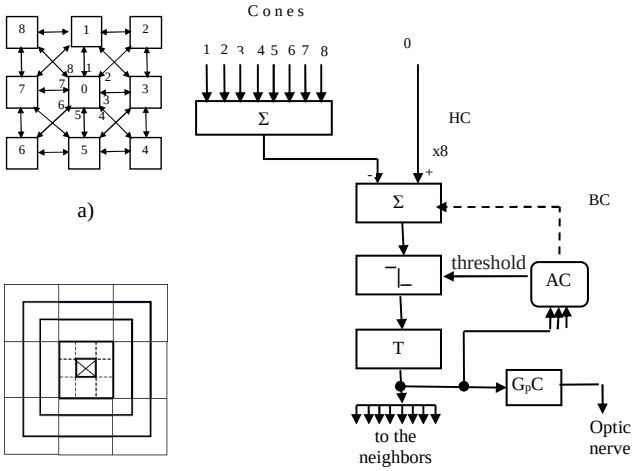
b)

Fig. 4. Character recognition based on δ-entropy: a) hieroglyphs; b) characters with different scales.

Expansion of the dynamic range of computer vision systems can be achieved by non-linear perception of brightness in an analog-to-digital conversion, using as a prototype a visual analyzer of a person.

The neural network organization of calculations is extremely effective, there are already dozens of variants for solving various problems, but they are extremely complicated with hardware implementation and require a complex adjustment to the task. As a prototype for processing information in the video system at the lower level, it is suggested to use the principles of short-range interaction with the ring organization of the on- and off-centers of the human eye, specialization of the layers of neurons, feedback between cells for controlling perception, adapting the sizes and forms of receptor fields (so-called plasticity of neurons).

The central fossa of the retina is specialized for clear vision and is organized on cones, horizontal (HC), bipolar (BC) and ganglionic P-type cells (G_PC) on a ring basis ("on" - and "off" centers) (Fig.5) Horizontal cells are inhibitory. Feedback through the amakrinovi (AC) cells control the perception of contrast by changing the threshold or by building up layers of neurons around the central. Such organization of the neural network is in good agreement with the arsenal of methods for distinguishing various informative features using the masks of Laplace, Sobel, Previti, Roberts and others [2]. Thanks to this, it is possible to effectively emphasize the contours, to highlight the features of the edge, to identify informative points, lines, their orientation, calculate gradients and etc. The ring organization of the neurons of the central fovea is considered to increase the contrast, which increases with the buildup of layers of neurons around the central element. The organization of connections between neurons is quite universal and enables the implementation of a convolution with matrices 3x3, 5x5, ... due to an increase in computation time on such a structure. All calculations on the structure are performed in parallel. It should also be noted that the calculation of the sums from the outputs of ring neurons and the calculation of the difference between the exciting and inhibitory neurons is carried out by a sequential code, it makes it possible to implement on this circuit not only the bit codes after the threshold limitation, but also the continuous codes. In this case, the computation time increases almost in proportion to the full-bit codes. The coefficients of the matrices can almost always be taken as numbers proportional to the power of the two (0, 1, 2, 4, 8, ...), which eliminates the need for multiplication.



a)

c)

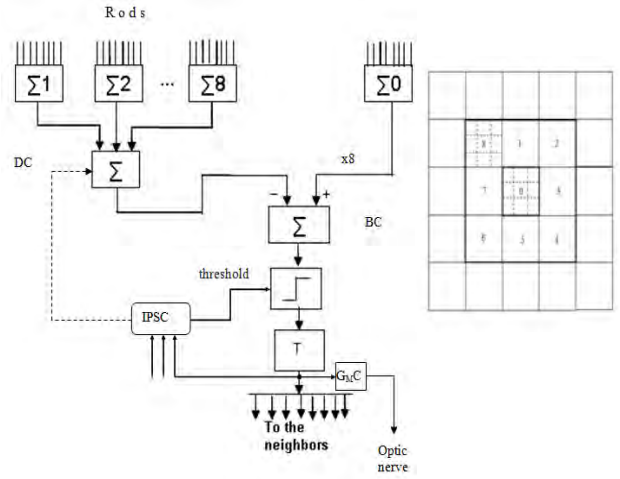
b)

Fig. 5. Horizontal and bipolar cells with increasing of rings: a) circular organization of neurons in the central fovea ("on" - center): element "0" in the center - excitatory, elements "1-8" around - inhibitory; b) implementation of the "on" - centre of 3x3; c) rings around the central element to increase sensitivity to the perception of contrast.

The periphery of the retina is a specialized on high sensitivity and is organized on the sticks of the retina, diffuse (DC), bipolar (BC) and ganglionic M-type cells (G_MC) also on the ring principle (Fig. 6). Increased sensitivity in conditions of insufficient illumination is ensured by the summation of signals from a large number of rods and the action of inhibitory diffuse cells. At the same time, accordingly, the spatial resolution is reduced. Interphase-shaped (IPSC) -linked cells control thresholds or receptor field sizes.

Using these methods, a number of specialized technical solutions protected by patents for inventions have been developed to combine the processes of perception of video information with its processing directly on the sensor array. In particular, these are sensor matrices with parallel binarization of the image and determination of the location and parameters of the object, with the calculation of the first and second moments of inertia binarized image for rows, columns and the whole image, with morphological processing of binarized images, parallel analog-to-digital

conversion and the possibility of nonlinear perception of the brightness. As an example, Fig.7 shows a generalized block diagram of one layer of a sensor array with image processing. Each element of the matrix has connections only with neighboring elements of the matrix (locality principle), that is, by analogy with the human visual analyzer, can realize “on”- and “off”-centers with a ring organization.



a) b)

Fig. 6. Diffuse and bipolar cells with increasing rings: a) the implementation of the ring organization of the peripheral neurons of the retina; b) rings for increased sensitivity in low light conditions.

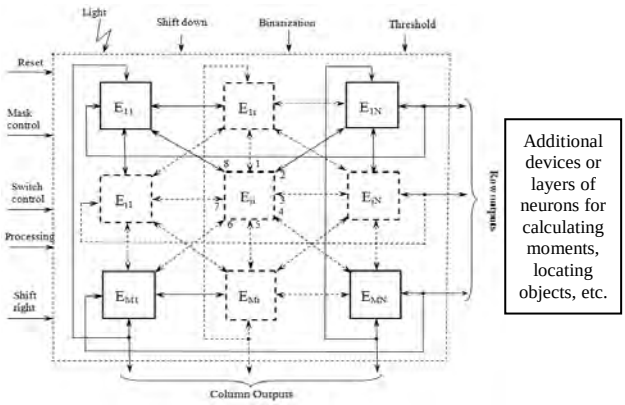


Fig. 7. Structure of the sensor matrix with image processing.

IV. Conclusions

Thus, the use of the principles of construction and operation of the human visual analyzer as a prototype of the computer vision system allows us to develop a whole arsenal of methods for increasing the efficiency of the processes of perception, processing and recognition of images.

References

[1] R. Schiffmann, Sensation and perception. 5-thed.Piter, SPb., Russia, 2003. (in Russian)

[2] R. Gonsales, and R. Woods, Digital image processing. Moscow, Russia,Technosphere, 2005. (inRussian)

[3] S. Shan, and M. D. Levine, “Visual Information Processing in Primate Cone Pathways - Part 1: A Model. Part 11: Experiments,” IEEE Trans. On Systems. Man, and Cybernetics – Part B: Cybernetics, vol.26, no.2, Apr. 1996.

[4] D. Anderson, Cognitive psychology. 5-th ed.Piter, SPb., Russia, 2002. (in Russian)

[5] P. J. Bur,t “Smart Sensing within a Pyramid Vision Machine,” IEEE, vol. 76, no. 8, pp. 175-185, 1988.

[6] V. Boyun, “Intelligent selective perception of visual information,” Informational aspects. Artificial intellect. no. 3. pp.16-24, 2011. (in Ukrainian)

[7] V. Boyun, “Intelligent Selective Perception of Visual Information in Vision Systems,” 6-th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Application. (IDAACS”2011). Prague, Czech Republic, vol.1, pp. 412-416, 2011.

[8] V. Boyun, “A human visual analyzer as a prototype forconstruction of the set of dedicated systems of machine vision,” International conference “Artificialintelligence. Intelligent systems. II-2010, vol. 1, pp. 21-26, 2010. (in Ukrainian)

[9] V. Boyun, “Directions of Development of Intelligent Real Time Video Systems,” Application and Theory of Computer Technology,[S.l.], vol. 2, no. 3, pp. 48-66, 2017. ISSN 2514-1694. Available at:<http://www.archyworld.com/journals/index.php/atct/article/view/65>. Date accessed: 26 sep. 2017. doi: https://doi.org/10.22496/atct.v2i3.65.

[10] V. Boyun, The dynamic theory of information. Fundamentals andapplications. Institute of Cybernetics of NASU, Kyiv, Ukraine, 2001, (in Russian)

[11] W. Prett, Digital Image processing. vol. 2, Mir, Moscow, USSR, 1982., (in Russian)