Visual Saccades for Object Recognition

Janusz A. Starzyk[1, 2]

[1]Ohio University, Athens, OH, USA
[2]University of Information Technology and Management, Rzeszow, Poland.
starzykj@ohio.edu

Abstract. This paper describes a method for rapid location and characterization of objects in 2D images. It derives optimizing parameters of a normalized Gaussian that best approximates the observed object, simultaneously finding the object location in the observed scene. A similarity measure to this optimized Gaussian is used to characterize the object. Optimization process has global and exponentially fast convergence, thus it can be used to implement saccadic motion for object recognition and scene analysis. This method was inspired by Perlovsky's work on neural dynamic logic used for fast location, characterization, and identification of objects. Developed method was tested and illustrated with an example of an object location and characterization.

Keywords: Object recognition, visual saccades, dynamic logic.

1 Introduction

This paper addresses an important problem of visual object recognition. Koch argues that while it is possible to imagine the algorithms that could perform object identification and recognition at a specific location, it is difficult to think of such parallel algorithm in which object recognition takes place in the entire field of vision [1]. Such algorithm would face tremendous computational complexity. Koch's work shows that the visual information is processed in a sequence of operations; each one identifies an object at a specific location, while early preprocessing stage may be performed in parallel on the entire image.

Poggio developed computational vision approach to image recognition and understanding [2]. A computational vision system assembles scene description from the input image in which early vision modules extract various features of the observed image like edges, color, texture, size, direction of motion etc. Poggio's computational vision system strived to match performance of human visual system in fast feedforward object categorization. He assumed that such system will perform mostly feedforward parallel operations for high speed processing.

Subsequently, Poggio and his coworkers introduced a set of higher level features for object recognition [3]. These features were obtained from simple features (like edge detectors) combined into complex cells capable of detecting edges in the cell neighborhood, independently on their location and orientation. In [4] authors show that using a hierarchy of feature extraction circuits involving simple and complex cells (similar to those observed by Hubel and Wiesel in research on ganglion cells in early vision) they can match human performance in a simple animal vs. non-animal categorization task. In a similar work, authors developed a tool for visual recognition of complex scenes [5].

Hierarchical organization of an object features and their assemblies is a foundation of the visual saccades based object representation, memory organization and recognition presented in this work. The work reduces tremendous computational complexity indicated by Koch while preserving flexibility in resolution level, scale and rotation of the original object image. The system is compatible with visual memory organization that can use higher order visual features extracted from simple or complex cells obtained from self-organizing feedforward neural networks in the visual input.

Section 2 presents the concept of visual saccades and visual attention. Section 3 describes finding characteristic features that represent an object or its parts using the most similar Gaussian distribution. Section 4 shows an example of object location and characterization. Section 5 contains conclusions and future work.

2 Visual Perception

2.1 Visual saccades

A visual **saccade** is a fast movement of an eye, head or of an optical device. It is initiated either consciously or subconsciously, and serves as a mechanism for focusing the visual attention on an object or its part [6]. Saccades are used to locate interesting parts of the observed scene in order to recognize the observed objects and build a representation of the observed scene. Saccades are very rapid and end with the gaze fixated on a selected spot. Additional advantage of the saccadic movement is to apply full resolution of the central part of the retina to the observed scene fragment in order to help recognize the observed object. This leads to a better use of the computational resources, improves the processing speed, and increases the recognition accuracy.

Saccadic movements are used to repeatedly revisit the same locations with a high saliency while reconstructing the whole scene. This is particularly useful at the object recognition stage. Once a mental image of the familiar object was made, it may serve as a reference for object recognition. Gradually the observed object and its features are inspected and compared to the internal image model with the best matching model selected for object recognition. Focus of attention associated with visual saccade improves recognition, and provides accurate information about the object location [7].

Features that were used to build such mental image are recalled and are used as guidelines for conscious saccades, to either confirm individual expected details of the model, or are basis for rejection of the inspected image if the observed image does not match with the expectations [8]. In this later case the observed object is considered unknown and a new model may be introduced and stored in the semantic memory.

2.2 Visual attention

Visually salient features (like brightness, movement, color, etc) are used to attract the visual attention. The visual attention supports efficient management of computing resources, reducing time cost and performing different visual tasks in a normal, cluttered and dynamic environment. It is used in the object recognition in coordination with the object model stored in the semantic memory.

Attention was considered as a mechanism for binding of distributed activations in response to presented stimuli [9]. It is a sequential mechanism used to select one of possibly many elements of the scene. Attention is used to temporarily suppress stimuli that do not belong to the object in attention focus, activating those that are correlated with the object of attention. The mechanism of temporal binding provided by attention may be used to quickly provide connections necessary to bind the activated groups of neurons forming long lasting memories.

Selective visual attention, linked to visual saccades, is needed to recognize objects and to understand a complex scene. The question is which part of the observed scene should be focused on, or how can we know where to look for objects we want to recognize? This task can be accomplished by evaluating saliency of various parts of the observed scene, concentrating attention on the most salient regions. After the winning region is inhibited, the next most prominent salient location is automatically selected through the same mechanism. However, it was determined that saliency is not the only factor for visual attention focus and that a significant portion of visual saccades is affected by associations between the observed artifacts [10].

For fast location and better identification of observed objects, attentional selection of objects was used in [11]. An interesting part of the image was selected using bottom-up attention based on salient features, to provide object location, and subsequently, an object was recognized using grouping based on segmentation. Thus both salient and homogeneous areas were used to locate and identify the object. Attentional modulation of neural activity helps to recognize object in a clattered scene [12].

Since computers process images performing sequential operations, then bottom-up attention algorithms limit the processing effort to analyze selected locations in the image. Koch and Itti have built a complex model of saliency-based spatial attention [13]. In their model a Winner-Take-All (WTA) neural network selects a location based on the saliency map to shift the visual attention to the selected spot, and to examine the image in the selected location.

Tsotsos et al. [14] used inhibition of the examined areas in order to perform attentional based selection and to obtain a selective tuning model of the observed object or scene. He used a top-down WTA attentional selection, with inhibition used for switching attention to the next salient feature.

Clark et al. [15] proposed a model where each task-specific feature detector is associated with a weight representing the relative importance of the particular feature to the current task. Also in his model, WTA operations are used on the saliency map to direct and switch spatial attention (triggering visual saccades). He used color and stereo vision to for attention focus and figure/ground separation.

Grossberg developed adaptive resonance theory (ART) to perform attention based perceptual grouping [16]. He proposed how a machine can learn new objects and events without forgetting those that were previously learned. He also suggested how bottom-up and top-down pathways can be used to focus attention on expected combinations of input features. ART also determines the level of mismatch between bottom-up feature patterns and top-down expectations to trigger memory search, or hypothesis testing, for recognition of objects and categories.

In [17] a hierarchical object-based computational model of visual attention was presented. This model combines object-based with visual saliency based model of visual attention [18] and uses bottom-up and top-down interaction [19,20]. The model integrates object and location based attention with visual representations of features. Top down attention provides priming to search for the expected features.

3 Feature selection

Fitting models to data typically requires selecting parameters corresponding to various models. However, the number of useful subsets of model parameters is combinatorially large. Thus model-based approaches encountered computational complexity and required treatments of NP complete algorithms. This problem was addressed by Perlovsky [21,22] where he described computational mechanism of going "from vague-fuzzy to crisp," that he called the dynamic logic.

In fast location, characterization and identification of objects using neural dynamic logic an important aspect is fast alignment of the object and its model. However, direct application of the dynamic logic needs a quick estimation of log-similarity between the image and the model. While this can be easily done for special cases (like matching two Gaussian functions), in general no constructive algorithm was proposed, and computational complexity similarity estimation is unknown.

An object may have a complex hierarchical structure of its visual features. Such features can be extracted and recognized using various methods. Representation of objects in the memory, which allow for recognition of objects irrespectively of the different viewing distance, direction, and other conditions, can be obtained using visual features descriptors generated with the SURF (Speeded Up Robust Feature) method [23].

In this paper I introduce a new approach which combines Perlovsky's concept of finding proper parameters to represent the object with the idea of saccading movement and attention switching for fast alignment of the object and its model. This will yield object representation and will help recognition of objects and visual scenes.

Saccadic movement is obtained through rapid finding of object location, orientation and scale on 2D image plane. In this work a two-step approach is used. First, each object in the semantic memory is characterized by its best matching Gaussian model, and then Gaussian functions are used to quickly locate and characterize the image objects. Memory objects are compared with the observed images after proper rotation and scale. In such approach, derivative information needed to find an optimum alignment is easily obtained by combining the observed image with properties of Gaussian function.

3.1 Finding the most similar Gaussian distribution

We will characterize and locate objects in 2D image plane using square root of the normalized Gaussian (based on multivariate normal distribution)

$$f(x,\mu,\Sigma) = \frac{1}{\sqrt{2\pi\sqrt{\det(\Sigma)}}} e^{-\frac{(x-\mu)\,\Sigma^{-1}(x-\mu)^T}{4}}. \tag{1}$$

where $x = [x_1, x_2]$, $\mu = [\mu_1, \mu_2]$, and Σ is 2x2 covariance matrix. Parameters $x, \mu,$ and Σ are chosen to maximize similarity S to the target function computed as the normalized inner product between Gaussian and the target function

$$\mathrm{S}(\mu, \Sigma, \phi) = \int_{-\infty}^{\infty} f(x, \mu, \sigma)\ \phi(x)\ \mathrm{dx} = \frac{f(x,\mu,\sigma) * \phi(x)}{\|f(x,\mu,\sigma)\| \|\phi(x)\|}. \quad (2)$$

To find the optimum values of μ and Σ we need to calculate derivatives $\frac{\partial \mathrm{S}(\mu,\Sigma,\bar{\mathrm{y}})}{\partial \mu}$ and $\frac{\partial \mathrm{S}(\mu,\Sigma,\bar{\mathrm{y}})}{\partial \Sigma}$ and set them to zero. First let us find

$$\frac{\partial \mathrm{f}(x, \mu, \Sigma)}{\partial \mu} = \mathrm{f}(x, \mu, \Sigma) * \left(-\frac{1}{4}\frac{\partial}{\partial \mu}\left(\mathrm{tr}\left(\Sigma^{-1} * (x - \mu)^{\mathrm{T}} * (x - \mu)\right)\right)\right) =$$

$$= \mathrm{f}(x, \mu, \Sigma) * \left(-\frac{1}{4}\left(x - \mu\right) * \Sigma^{-1}\right) \quad (3)$$

where tr(A) is trace of a square matrix $\mathrm{A} = \left[\mathrm{a_{ij}}\right]$, $\mathrm{tr(A)} = \sum_{\mathrm{i}} \mathrm{a_{ii}}$. After normalizing both Gaussian and target functions we have:

$$\mathrm{S}(\mu, \Sigma, \phi) = \bar{f}(x, \mu, \Sigma) * \bar{\mathrm{y}}(x) = \sum_i \bar{f}(x_i, \mu, \Sigma) * \bar{\mathrm{y}}(x_i) \quad (4)$$

and derivative of the similarity function is set to 0 to find the optimum value μ.

$$\frac{\partial \mathrm{S}(\mu,\Sigma,\bar{\mathrm{y}})}{\partial \mu} = \frac{\partial \bar{f}(x,\mu,\Sigma) * \bar{\mathrm{y}}}{\partial \mu} = \sum_i \bar{f}(x_i) * \bar{\mathrm{y}}_i * \left(-\frac{1}{4}\left(x_i - \mu\right) * \Sigma^{-1}\right) = 0. \quad (5)$$

Solving for the optimum values of μ we get:

$$\mu = \frac{\sum_i \bar{f}(x_i,\mu_0,\Sigma_0) * \bar{\mathrm{y}}_i * x_i}{\sum_i \bar{f}(x_i,\mu_0,\Sigma_0) * \bar{\mathrm{y}}_i}, \quad (6)$$

where x_i is a 2D coordinate vector $x_i = [x_{i1}, x_{i2}]$, and $\bar{f}(x_i)$, and $\bar{\mathrm{y}}_i$ are scalar Gaussian and target function values at x_i. In addition from

$$\frac{\partial \mathrm{f}(x, \mu, \Sigma)}{\partial \Sigma^{-1}} = \mathrm{f}(x, \mu, \Sigma) * \left(\frac{1}{4 * \det(\Sigma^{-1})} * \frac{\partial \det(\Sigma^{-1})}{\partial \Sigma^{-1}}\right) +$$

$$+\mathrm{f}(x, \mu, \Sigma) * \frac{\partial\left(-\frac{1}{4}\mathrm{tr}\left(\Sigma^{-1} * (x - \mu)^{\mathrm{T}} * (x - \mu)\right)\right)}{\partial \Sigma^{-1}} =$$

$$= \mathrm{f}(x, \mu, \Sigma) * \left(\frac{\Sigma^{\mathrm{T}}}{4} - \frac{1}{4}\left(x - \mu\right)^{\mathrm{T}} * \left(x - \mu\right)\right) \quad (7)$$

where we used

$$\frac{\partial \det(\Sigma^{-1})}{\partial \Sigma^{-1}} = \det(\Sigma^{-1}) * \Sigma^{\mathrm{T}},$$

we have

$$\frac{\partial \mathrm{S}(\mu, \Sigma, \bar{\mathrm{y}})}{\partial \Sigma^{-1}} = \frac{\partial \bar{f}(x, \mu, \Sigma) * \bar{\mathrm{y}}}{\partial \Sigma^{-1}} =$$

$$= \frac{1}{4}\sum_i \bar{f}(x_i) * \bar{\mathrm{y}}_i * \left(\Sigma^{\mathrm{T}} - \left(x - \mu\right)^{\mathrm{T}} * \left(x - \mu\right)\right) = 0 \quad (8)$$

Using (6) and (8) μ αnd Σ can be iteratively updated and the convergence is very fast. We will obtain

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \frac{\sum_i \bar{f}(x_i,\mu_0,\Sigma_0)*\bar{y}_i*x_{i1}}{\sum_i \bar{f}(x_i,\mu_0,\Sigma_0)*\bar{y}_i} \\ \frac{\sum_i \bar{f}(x_i,\mu_0,\Sigma_0)*\bar{y}_i*x_{i2}}{\sum_i \bar{f}(x_i,\mu_0,\Sigma_0)*\bar{y}_i} \end{bmatrix}, \tag{9}$$

and

$$\Sigma = \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix} =$$

$$= \frac{1}{\sum_i \bar{f}_i * \bar{y}_i} \begin{bmatrix} \sum_i \bar{f}_i * (x_{i1} - \mu_1)^2 * \bar{y}_i & \sum_i \bar{f}_i * (x_{i1} - \mu_1) * (x_{i2} - \mu_2) * \bar{y}_i \\ \sum_i \bar{f}_i * (x_{i1} - \mu_1) * (x_{i2} - \mu_2) * \bar{y}_i & \sum_i \bar{f}_i * (x_{i2} - \mu_2)^2 * \bar{y}_i \end{bmatrix} \tag{10}$$

where for simplicity

$$f_i = \bar{f}(x_i, \mu_0, \Sigma_0) \tag{11}$$

Assume that the set of points that represent a 2D object were rotated by ϑ. In addition, if we scale the object in x and y directions by scaling factors λ_1 and λ_2 and translate it by shifting all the object points on the plane by vector a, we can represent a linear transformation of all points as:

$$\mathrm{Ax} + \mathrm{a} = \mathrm{R} * \Lambda * \mathrm{x} + \mathrm{a} = \begin{bmatrix} \cos(\vartheta) & -\sin(\vartheta) \\ \sin(\vartheta) & \cos(\vartheta) \end{bmatrix} * \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} * \mathrm{x} + \mathrm{a} \tag{12}$$

Using properties of the covariance matrix we have for a matrix A and a vector a:

$$\Sigma(\mathrm{Ax} + \mathrm{a}) = cov(Ax + a) = A * \Sigma(\mathrm{x}) * \mathrm{A}^{\mathrm{T}} \tag{13}$$

Using (13) it is easy to obtain covariance of the translated, rotated and scaled set of points.

To characterize and object we will first need to find Σ and μ of the transformed Gaussian, that maximizes similarity S described by (2). This corresponds to finding the rotation matrix R, scale matrix Λ and mean values μ.

From (12) we have the covariance matrix:

$$\Sigma_{\mathrm{G}} = A * \Sigma_0 * \mathrm{A}^{\mathrm{T}} = \mathrm{R} * \Lambda * \Lambda^{\mathrm{T}} * \mathrm{R}^{\mathrm{T}} = \mathrm{R} * \begin{bmatrix} \lambda_{\mathrm{G1}}^2 & 0 \\ 0 & \lambda_{\mathrm{G2}}^2 \end{bmatrix} * \mathrm{R}^{\mathrm{T}} \tag{14}$$

Using diagonalization of a matrix Σ_{G} we can get

$$\Lambda_{\mathrm{G}} = \begin{bmatrix} \lambda_{\mathrm{G1}}^2 & 0 \\ 0 & \lambda_{\mathrm{G2}}^2 \end{bmatrix} = \mathrm{X}^{-1} * \Sigma_{\mathrm{G}} * \mathrm{X} \tag{15}$$

where X is the matrix with eigenvectors of Σ_{G} as column of X and $\lambda_{\mathrm{G1}}^2$and λ_{G2}^2 are equal to eigenvalues of Σ_{G}. Thus the scale factors in y and x directions λ_{G1} and λ_{G2} can be obtained from (15). In addition, the rotation matrix $\mathrm{R}_{\mathrm{G}} = \mathrm{X}^{-1}$.

3.2 Fitting Gaussian function:

Fitting Gaussian function is used to characterize the object and to find its location. This is a quickly convergent iterative process and it is a part of object characterization and location procedures. Fitting Gaussian function it is performed by the following algorithm.

Fitting Gaussian Function Algorithm:

1. Start with the initial value for $\mu = \left[\left|\frac{x_{1max}-x_{1min}}{2}\right| \quad \left|\frac{x_{2max}-x_{2min}}{2}\right|\right]$, where x_{1max}, x_{1min}, x_{2max}, and x_{2min} are respectively the maximum and minimum values of x and y coordinates in the observed scene. Notice, that for location and characterization the size of the observed scene is typically larger than the size of the object. Start with initial value for Σ equal to
$$\Sigma = \begin{bmatrix} (x_{1mxo} - x_{1mno})^2 & 0 \\ 0 & (x_{2mxo} - x_{2mno})^2 \end{bmatrix}$$
where x_{1mxo}, x_{1mno}, x_{2mxo}, and x_{2mno} are respectively the maximum and minimum values of x and y coordinates in the observed image.
2. Compute the square root of the Gaussian function
$$f(x, \mu, \Sigma) = \frac{1}{\sqrt{2\pi\sqrt{\det(\Sigma)}}} e^{-\frac{(x-\mu)\,\Sigma^{-1}(x-\mu)^T}{4}},$$
where $x = [x_1 \quad x_2]^T$, $\mu = [\mu_1 \quad \mu_2]^T$ and Σ is 2x2 covariance matrix.
3. Use (9) and (10) to compute new values for μ and Σ.
4. Since both μ and Σ influence computation of the function $f(x, \mu, \Sigma)$ values, we need to iterate repeating steps 2. and 3.

The algorithm convergence is very fast. Fig. 1 shows the convergence rates for both the covariance matrix and the mean values. As we can see, even after the first iteration the object is located within the distance of 2^3 pixels as the error of the mean value indicates. Considering that the object was located within a 960x740 pixels image this indicates less than 1% error for the first iteration and the error is reduced exponentially to less than 2^{-8} pixel distance after 20 iterations. Such accuracy is seldom required and a single iteration is sufficient to locate the object or its feature, implementing rapid saccading motion to the target area.

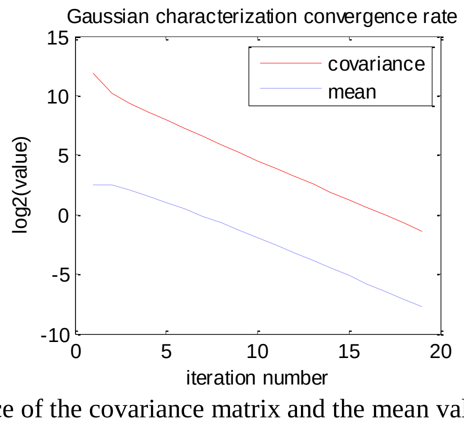

Fig. 1 Convergence of the covariance matrix and the mean values for the Gaussian characterization.

3.3 2D Object Characterization

To characterize the object we perform the following operations:

2D Object Characterization Algorithm:

1. Normalize the object image and remove its background.
   a. Object image is reduced to black and white image.
   b. The black and white version of the image is normalized to have norm equal to 1.
2. Position the normalized object image in a bigger picture by shifting its lower left corner by a preset vector μ_C.
3. Fit Gaussian function $\bar{f}(x, \mu, \Sigma)$ to obtain relative location of the Gaussian $\mu_o = \begin{bmatrix} \mu_{o1} \\ \mu_{o2} \end{bmatrix}$ with respect to the lower left corner of the image. This is obtained by subtracting position of the left lower corner of the object image from the coordinates of the Gaussian $\mu_o = \mu_G - \mu_C$.
4. Obtain covariance matrix Σ_o of its best fitting Gaussian function and similarity $S(\mu_o, \Sigma_o, \phi_o) = \bar{f}(x, \mu, \Sigma) * \bar{t}(x)$ measure between the normalized image $\bar{t}(x)$ and this Gaussian.
5. The object image is characterized by $\mu_o, \Sigma_o, and\ S(\mu_o, \Sigma_o, \phi_o)$.

4 Example

To illustrate the discussed approach let us consider an image of an object and its gray scale version shown in Fig. 2. Vertical and horizontal axes correspond to y and x values.

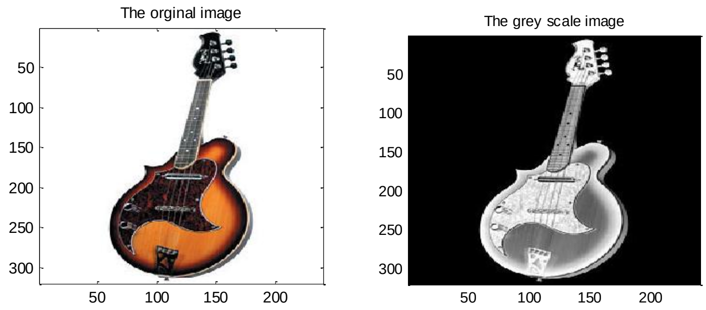

Fig. 2 An object and its grayscale version.

This image was shifted in such a way that its upper left corner was moved from the location (1,1) to the position $\mu_C = [320\ 240]^T$ as shown in Fig. 3.

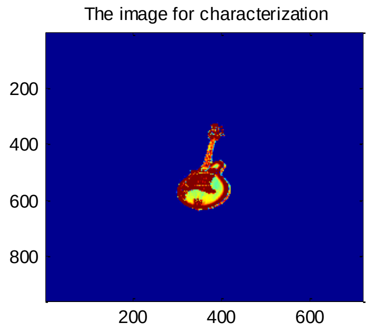

Fig. 3 The object placed in a larger area

The image was characterized using 2D object characterization algorithm by the Gaussian function located at $\mu_G = [\mu_{Gy} \quad \mu_{Gx}] = [545.58 \ 357.78]^T$ and the optimum Gaussian has its covariance matrix equal to:

$$\Sigma_o = \begin{bmatrix} 1719.7 & -245.2 \\ -245.2 & 677.5 \end{bmatrix}$$

After characterization this image similarity to square root of the optimum Gaussian was determined to be $S(\mu_o, \Sigma_o, \phi_o) = 0.9444$.

5 Conclusions and future work

Presented in this paper, quick characterization and location of the object image is a machine implementation of the visual saccades idea for object recognition. Using this approach the observed image is described based on the mean value and the covariance matrix of a 2D Gaussian function that is most similar to the observed image. Similarity measure between the best Gaussian fit and the observed object is used for object characterization. Subsequently, characterized and memorized objects that have specific Gaussian similarity are extracted from the memory, and after proper rotation and scale, are placed in the identified location for recognition.

Mathematical equations that solve the optimization problem to find the most similar Gaussian function are solved explicitly, yielding fast convergence of the iterative algorithm. One iteration of the algorithm typically suffices to find approximate location of the perceived object. More precise determination of the object location, its scale, and rotation require very few iterations and can be quickly computed.

This procedure can be applied either to the entire image, its parts, or to a complex scene. Inhibition of previously visited areas of the image will force saccadic searches to describe and recognize various objects in the observed scene.

Future work is to test this concept in realistic scenes with several objects, to build hierarchical object representations in the semantic memory, and to apply various resolution levels in order to minimize the processing time.

Acknowledgements: This research was supported by The National Science Centre, grant No. 2011/03/B/ST7/02518.

References

1. C. Koch, S. Ullman , "Shifts in selective visual attention: towards the underlying neural circuitry", Matters of Intelligence, 115-141, 1987.
2. T. Poggio, V. Torre, C. Koch , "Computational vision and regularization theory", Image understanding 3 (1-18), 111, 1989.
3. T. Serre, L. Wolf, T. Poggio , "Object recognition with features inspired by visual cortex", Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer , 2005.
4. T. Serre, A. Oliva, T. Poggio ,"A feedforward architecture accounts for rapid categorization", Proceedings of the National Academy of Sciences 104 (15), 6424-6429, 2007.
5. T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, T. Poggio ,"Robust object recognition with cortex-like mechanisms", IEEE Trans. on PAMI, 29 (3), 411-426, 2007.
6. H. Kirchner, S. J. Thorpe ,"Ultra rapid object detection with saccadic eye movements: Visual processing speed revisited", Vision Research, vol. 46, no. 11, pp. 1762-1776, 2006.
7. D. E. Irwin and R. D. Gordon, "Eye movements, attention, and trans-saccadic memory." Visual Cognition, 5, pp. 127-155, 1998.
8. Herwig and W. X. Schneider, "Predicting object features across saccades: evidence from object recognition and visual search.", J Exp Psychol Gen. 143(5): pp.1903-22, Oct. 2014.
9. J. H. Reynolds and R. Desimone R., The role of neural mechanisms of attention in solving the binding problem, Neuron. 1999 Sep; 24(1): pp. 19-29.
10. R.J. Peters, A. Iyer, L. Itti, C. Koch ,"Components of bottom-up gaze allocation in natural images", Vision Research 45 (18), pp.2397-2416, 2005.
11. D. Walther, U. Rutishauser, C. Koch, and P. Perona (2005) Selective visual attention enables learning and recognition of multiple objects in cluttered scenes, *Computer Vision and Image* Understanding, 100: 41-63.
12. D. Walther, L. Itti, M. Riesenhuber, T. Poggio, C. Koch , "Attentional selection for object recognition—a gentle way", Biologically Motivated Computer Vision, 472-479, 2002.
13. L. Itti, C. Koch, A saliency-based search mechanism for overt and covert shifts of visual attention, Vision Res. 40 (10–12) pp. 1489–1506, 2000.
14. J. K. Tsotsos, et al., Modelling visual attention via selective tuning, Artificial Intelligence 78 (1995) 507–545.
15. J. J. Clark, Spatial attention and latencies of saccadic eye movements, Vision Res. 39 (3) (1998) 583–600.
16. S. Grossberg, How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex, Spatial Vision 12 (2) (1999) 13–185.
17. Y. Sun and R. Fisher, "Object-based visual attention for computer vision, Artificial Intelligence, Vol. 146, no. 1, pp. 77–123, May 2003.
18. L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Machine Intel. 20 (11) (1998) 1254–1259.
19. J.W. Wolfe, Visual search, in: H. Pashler (Ed.), Attention, Psychology Press, 1998, pp. 13–73.
20. S. Yantis, Control of visual attention, in: H. Pashler (Ed.), Attention, Psychology Press, 1998, pp. 223–256.
21. L. I. Perlovsky, ""Vague-to-Crisp" Neural Mechanism of Perception," *IEEE Transactions on Neural Networks*, vol. 20, pp. 1363-1367, 2009.
22. L. I. Perlovsky, "Neural Mechanisms of the Mind: Aristotle, Zadeh and fMRI," *IEEE Transactions on Neural Networks*, vol. 21, pp. 718-733, 2010.
23. H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, SURF: Speeded Up Robust Features, Comput. Vis. Image Und., 110(3), 346-359 (2008).