Automatic Determination of Cell Division Rate Using Microscope Images

K. V. Nekrasov, D. A. Laptev, and D. P. Vetrov

Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, 119991 Russia
e-mail: knek@list.ru, laptev.d.a@gmail.com, vetrovd@yandex.ru

Abstract—Estimating the dynamics of cell culture development using fluorescent microscopy images is of great interest in modern cellular and molecular biology. In large-scale studies of cell populations involving a great number of images taken within a certain interval, manual analysis has proved to be time consuming and inaccurate; therefore, various computer-aided tools are required. This paper presents an efficient algorithm for recognizing budding cells and determining their division rate using microscope images. Image binarization using the Laplacian of the Gaussian; classifications of objects on cells, buds, and noise; and the determination of the cell division rate based on the data obtained are considered.

Keywords: binarization, Laplacian of the Gaussian, median filter, image skeleton, maximal matching, Euclidean Distance Transform.

DOI: 10.1134/S1054661813010094

1. INTRODUCTION

Estimating the reproductive rate of the cell culture given various factors is of great importance in microbiology, molecular biology, and genetic studies. The division rate is determined as the ratio between the number of dividing cells and the overall number of cells. Typically, the division index is determined manually.

The dividing cell is marked by a small circular appendix known as a bud (see red squares in Fig. 1d).

It is obvious that manual calculation reduces the reliability of the results and requires massive efforts from experts. This paper presents an algorithm that processes images of the budding cells' culture and calculates the division rate automatically. A similar problem is discussed in [1, 2]. Figure 1 illustrates the main steps of the algorithm. The first step involves transformation of the source image into binary form in which every connected component is an identifiable group of cells. The binarization uses the Laplacian of the Gaussian [3, 4] with subsequent processing. The filtration and filling of regions are also performed [5, 6]. The second step involves the determination of segments with cells and buds in each connected region of the binarized image. The third step involves the calculation of the budding index using data obtained at the second step. At the end of the paper, a comparison of the results of the proposed algorithm and those of manual processing by biologists is presented.

Received March 14, 2011

2. BINARIZATION

The variation of the Laplacian of the Gaussian [3, 4] with subsequent closure of discontinuous contours is used for image binarization; this is one of widely known methods of boundary detection. The method stems from the fact that points in which the Laplacian is zero and around which the Laplacian has different signs are considered to be boundary points. Preliminarily, the image is modified by Gaussian smoothing to reduce noise sensitivity. The Gaussian smoothing and Laplacian search can be performed simultaneously by using a filter called the Laplacian of the Gaussian (LoG), which is calculated according to the following formulas (here, C is the normalizing constant):

$$G_\sigma(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \quad (1)$$

$$\Delta G_\sigma(x, y) = C\frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \quad (2)$$

In order to reduce still further the noise sensitivity of the algorithm, points with the gradient below the threshold are eliminated from the set of boundary points.

Let B be the result of filtering the source image of size $B_{width} \times B_{height}$ in gray gradations using the LoG core with $\sigma = 2$ and width $w = 13$. In this paper, the threshold is chosen to be

$$T_{LoG} = \frac{0.75}{B_{width}B_{height}} \sum_{x, y} B(x, y). \quad (3)$$

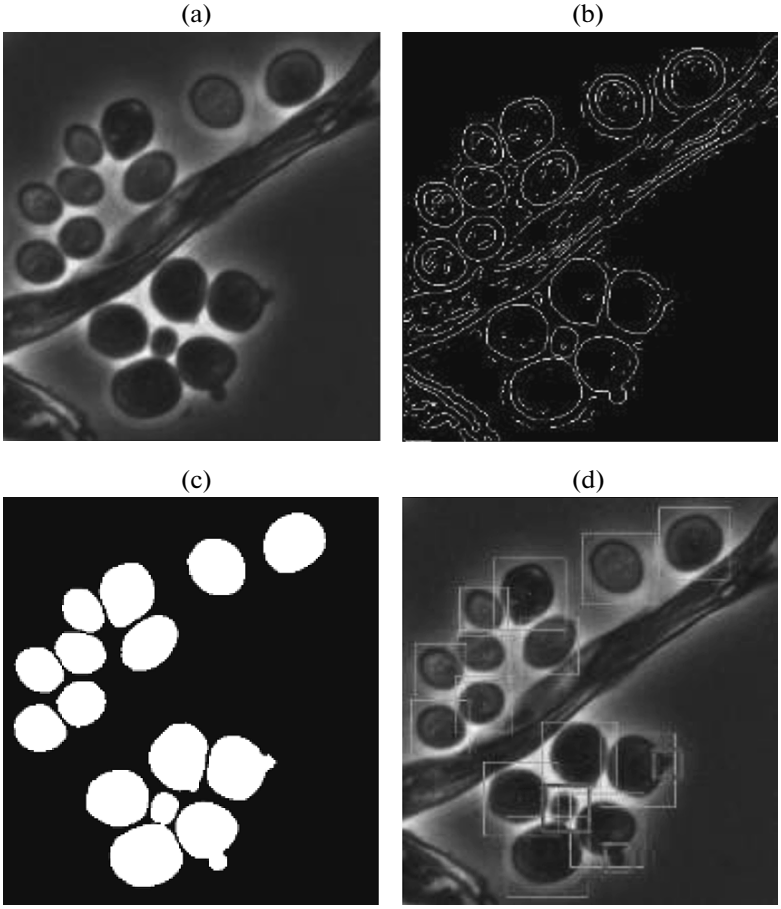

Fig. 1. Main steps of the algorithm: (a) part of the original image; (b) results of applying the boundary detection method with the Laplacian of the Gaussian; (c) image after the closure of discontinuous contours, filling, and filtration; (d) result image with marked cells (green squares) and buds (red squares).

Figure 1b shows the result of applying the LoG method (the original image is presented in Fig. 1a).

Contours of some cells are proved to be discontinuous, which hinders determination of single cells in the subsequent steps. Moreover, most of the cells have luminescent halos as a result of highlighting when taking the images; this leads to the double border effect (borders around the cells and halo borders) on the binarized image. This effect can be seen in Fig. 1b.

In contrast to the cell borders, the halo borders are characterized by a high density of endpoints [7]. The following simple heuristic algorithm was used to eliminate irrelevant borders and discontinuities:

(1) Find the map E of all endpoints on the binary image B.

(2) Choose any endpoint $p \in E$ and remove it from E.

(3) Find the set $M_r(p)$ of all the points from E that are in the vicinity of the radius r around the point p.

(4) If $|M_r| \le C_{\text{vr}}$ (low density of points), then connect the point p with every point from M_r on the image B.

(5) Repeat steps 2–4 until the set of points E is empty.

Parameters C_{vr} and r are set manually. Thresholds $C_{\text{vr}} = 3$ and $r = 6$ were chosen heuristically.

Once the algorithm is applied on the binary image B, each closed contour will be the boundary of the connected cells' group. Next, all closed regions on B are filled and the median filter of size 3×3 is used to eliminate all remaining noise on the image.

The filling of regions is described in [6, 11]. The main idea of the median filter is that each pixel of the filtered image contains the median value from the neighborhood 3×3 around the corresponding pixel on the input binary image [5]. Figure 1c shows the result of the filling and filtration.

3. CLASSIFICATION OF REGIONS

Now let us determine the location of cells and buds on the binary image. Let us consider an arbitrary connected region D on the obtained binary image B. This

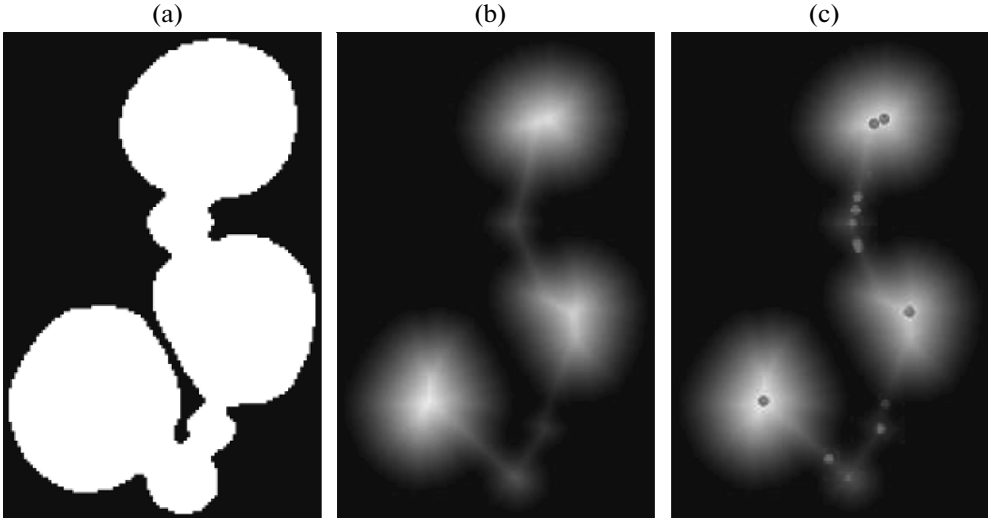

Fig. 2. Illustration of the region classification algorithm: (a) connected region D representing three stuck-together dividing cells; (b) Euclidean Distance Transform EDT on the region D: the brighter the light in the point, the greater the distance from the point to the background; (c) points on which EDT reaches its local maximums on the region D: assumed centers of buds are marked with red, assumed centers of cells are marked with green, and the noise or local maximums beyond the given range are marked with blue.

region can be the noise as well as the group of interconnected cells. We will consider the last case (an example of such a region is given in Fig. 2a).

Due to the form of cells and buds, their position on the binary region D can be approximated by circuits inscribed in D. Let us specify the tolerance range for radii of cells and buds in accordance with biologists' requirements. The cell radius varies from $R_{\min}$ to $R_{\max}$, while the bud ranges radius from $r_{\min}$ to $r_{\max}$.

We stipulate that the area of the considered region must exceed the area of the minimal cell ($S_D \le \pi R_{\min}^2$); otherwise, this is noise.

Let us find the EDT (Euclidian Distance Transform) that calculates Euclidean distances on the binary image D [8, 10]. For each pixel on the image D, the calculation result is determined by the distance between the current pixel and the nearest background (zero) pixel of the image D (Fig. 2b). Thus, the value of EDT in the arbitrary point $p \in D$ determines the radius of the circle centered at p and inscribed in D.

Let us find the set S of all points on which EDT reaches its local maximums on the region D. We paint the obtained points as follows: all the points on which EDT varies within the bud radius are marked with red, points on which EDT varies within the cell radius are marked with green, and all other points (noise) are marked with blue (Fig. 2c). Red and green points are the most likely centers of buds and cells, respectively.

Let us introduce the skeleton of the binary connected region as a geometric locus of points of centers inscribed in this region of circles [11] (Fig. 3a). Note that the skeleton of the region D contains points from the set S. The painting described above can be applied to the image skeleton as well (Fig. 3b). The set of points S could equally be sought on the skeleton of the region D.

Let us consider circuits centered at points from the set S, which are inscribed in D. We eliminate all the circuits that are completely covered by at least one circuit centered at the point from S (the case in which S is the skeleton of the region D is shown in Fig. 3b). Only centers of undeleted circuits are left in S. Unfortunately, this operation is insufficient for an accurate description of locations of centers of cells and buds, several cells or buds being characterized simultaneously by an arbitrary cloud of closely adjacent points. This can be seen in the right bottom of Fig. 3c where the red region represents two buds at a time.

Let us consider big and small circuits R and r, $R \ge r$; let l be the distance between centers of the circuits. If $\frac{r-l+R}{2r} \times 100\% \ge P\%$, the big circuit covers no less than P percent of the diameter of the small one, $x =$ const. Further, we will simply say that the big circuit covers the small one with the constant P.

Once again, let us consider the set of inscribed circuits centered at points S. Using the constant $P_{buds} = 50\%$, we remove all the circuits with bud radii that are covered by circuits with radius $R > r_{\max}$. Next, using the constant $P_{cells} = 25\%$, we remove all the circuits with cell radii that are covered by circuits with radius $R > R_{\max}$. Only centers of undeleted circuits with radii of cells and buds are left in S. The noise is eliminated.

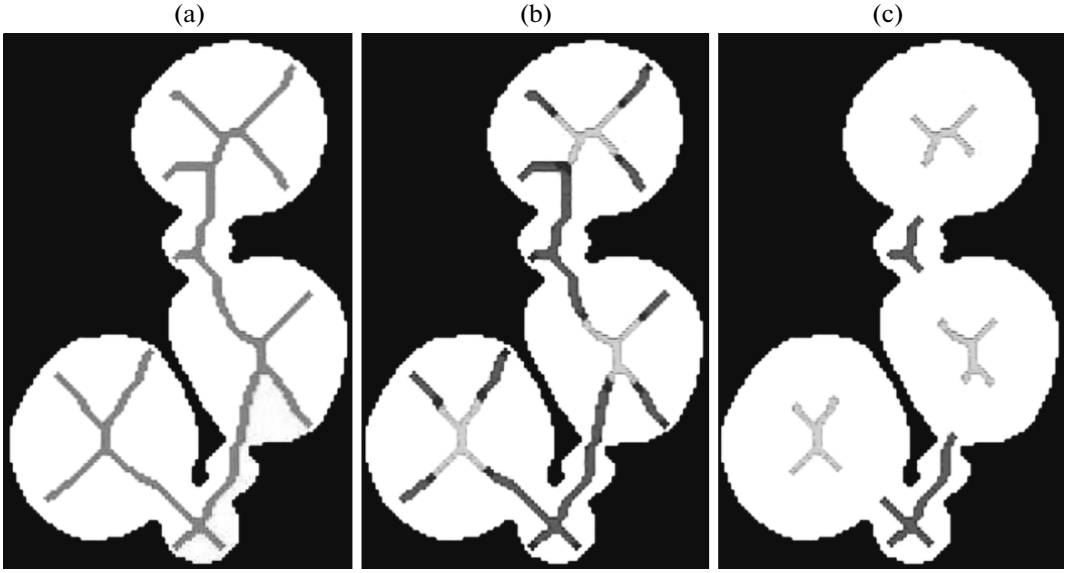

Fig. 3. Analogy to the bitmap image skeleton: (a) skeleton of the binary region; (b) skeleton points are colored as follows: red points are assumed centers of buds, green points are assumed centers of cells, and blue points represent noise; (c) centers of all circuits that are completely covered by at least one circuit with the center among the set of marked points are removed from the previous image.

It may well be that the entire cloud of assumed centers corresponds to a single cell or bud (as shown in Fig. 3c). Let us consider the set of assumed locations of cells S_{cells} and buds S_{buds} (Fig. 4a).

$$S_{cells} = \{(x, y, R) | (x, y) \in S, R = EDT(x, y), \quad R_{\min} \le R \le R_{\max}, \tag{4}$$

$$S_{buds} = \{(x, y, r) | (x, y) \in S, r = EDT(x, y), \quad r_{\min} \le r \le r_{\max}. \tag{5}$$

We break the sets of circuits S_{cells} and S_{buds} into disjoint groups in terms of proximity: two circuits belong to a group if the big circuit covers the diameter of the small one with the constant $P_{cells} = 25\%$ for cells from S_{cells} and with the constant $P_{buds} = 50\%$ for buds from S_{buds}. An efficient procedure of uniting disjoint sets is described in [9].

Let us assign the median circuit G_{mean} to each group $G = (X_i, Y_i, R_i)_{i=1}^{N}$:

$$G_{mean} = \left(\sum_{i=1}^{N} \frac{X_i}{N}, \sum_{i=1}^{N} \frac{Y_i}{N}, \sum_{i=1}^{N} \frac{R_i}{N} \right).$$

We denote the set of all median circuits with the cell radius by G_{cells}, while the set of all median circuits with the bud radius is denoted by G_{buds} (Fig. 4b). The circuit $G \in G_{cells}$ covers the region where the cell is most probably located; the circuit $G \in G_{buds}$ covers the region where the bud is most probably located.

Let us consider the bigraph that contains all the elements from G_{cells} in one part and all the elements from G_{buds} in another. If the segment from the center of the circuit *cell* to the center of the circuit *bud* lies completely in the region D, the cell $cell \in G_{cells}$ is connected by an edge with the bud $bud \in G_{buds}$. The maximal matching in the obtained bigraph provides a correct interconnection of cells and buds (various examples are presented in Fig. 4c). An efficient implementation of the procedure is described in [9].

Let us denote the maximal matching by $R_{pairs} = (cell_i, bud_i)_{i=1}^{M}$, where M is the number of pairs. All elements that are not included in the maximal matching R_{pairs} are removed from the set G_{buds}. All cells from G_{cells} that are not included in the maximal matching R_{pairs} are denoted by $R_{singles}$.

Result sets R_{pairs} and $R_{singles}$ contain coordinates and radii of dividing and nondividing cells. Each dividing cell is represented by two circuits: big cell and small bud.

4. FINAL CALCULATION

Let K connected regions be located on the binary image B obtained at the end of the first step. We determine the coordinates of dividing $R_{pairs}(D_i)$ and nondividing $R_{singles}(D_i)$ cells for each connected region D_i

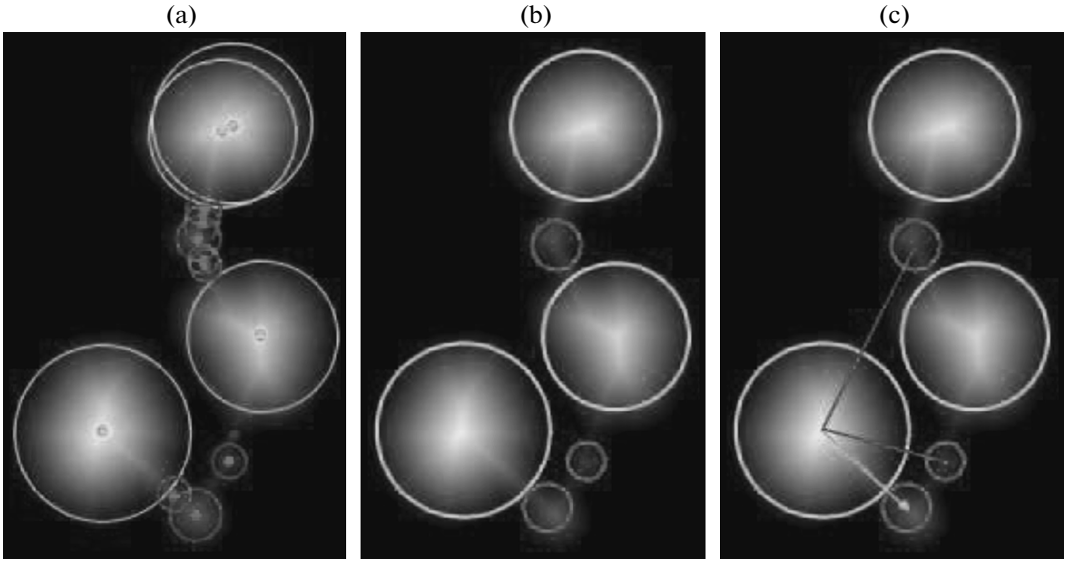

Fig. 4. Illustration of the region classification algorithm: (a) sets of inscribed circuits S_{cells} and S_{buds} are marked with green and red, respectively; (b) sets of median circuits G_{cells} and G_{buds} are marked with green and red, respectively; (c) various alternatives of connecting a cell with buds: the green arrow corresponds to the correct variant, red arrows represent incorrect variants.

($i = 1...K$). Thus, the division index (*index*) is calculated according to the following formulas:

$$budding = \sum_{i=1}^{K} R_{pairs}(D_i),$$

$$single = \sum_{i=1}^{K} R_{singles}(D_i),$$

$$index = \frac{budding}{budding + single} \times 100\%.$$

5. RESULTS

We analyzed a set of images provided by biologists (Fig. 5). It can be seen that results of the proposed algorithm agree well with those obtained by manual processing.

The analysis of three cell populations kept in different biological conditions shows that the results of the proposed method differ from the expert marking by less than 6%, the difference being less than 4% in 22 out of 24 images.

The algorithm execution time in MathLab is about 15 seconds when processing one image with a high cell density and resolution 2560 × 1520. Manual processing takes about 5 minutes per image due to the fact that a typical image contains up to several hundred cells.

6. CONCLUSIONS

The algorithm for automatic determination of the cell division rate using microscope images is implemented. The main steps of the algorithm—binarization using the Laplacian of the Gaussian with subsequent closure of discontinuous contours, the determination of cell and bud locations by constructing the set of circuits on the binary region and removing part of them, and the final calculation—are described in detail. The proposed algorithm provides a high-speed and high-accuracy determination of the cells' location

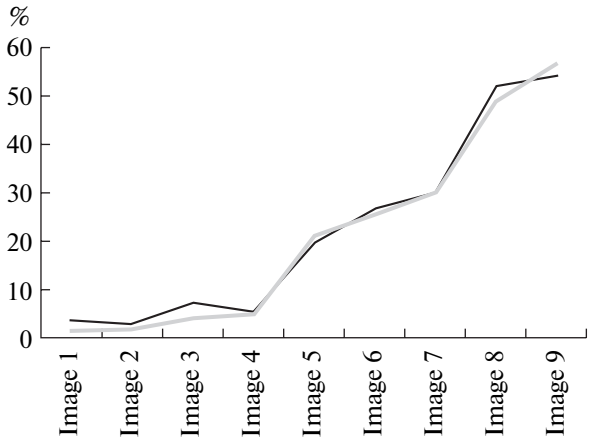

Fig. 5. Results of the automatic and manual processing of images of a cell population: the vertical axis represents the division rate, the black line corresponds to the automatic analysis, and the gray line corresponds to manual processing.

on the image, state of cells, and division rate of the entire cell population. Manual image processing is time consuming and requires a certain qualification of personnel. The proposed method standardizes image processing, makes no demands on the operator, and can thus be used in large-scale studies. The algorithm has been implemented and is currently used in the research laboratory of the Faculty of Biology.

REFERENCES

1. D. Laptev, "Automated Determination of Cells Fission Intensity by Using Microscope Photo," *Proc. Graphicon-2009* (Moscow, 2009).
2. M. Kvarnstrom, K. Logg, A. Diez, K. Bodvard, and M. Kall, "Image Analysis Algorithms for Cell Contour Recognition in Budding Yeast," Opt. Express **16** (17) (2008).
3. Om Pavithra Bonam and Sridhar Godavarthy, "Edge Detection," in *Computer Vision (CAP 6415: Project 2)* (University of South Florida, 2003).
4. T. Lindeberg, "Edge Detection and Ridge Detection with Automatic Scale Selection," Tech. Report ISRN KTH NA/P-96/06-SE (1996).
5. B. Weiss, "Fast Median and Bilateral Filtering," Tech. Rep. (Shell & Slate Software Corp., 2006).
6. P. Soille, *Morphological Image Analysis: Principles and Applications* (Springer-Verlag, 1999).
7. T. Y. Kong and A. Rosenfeld, *Topological Algorithms for Digital Image Processing* (Elsevier Sci., 1996).
8. S. Zhang and M. A. Karim, "Euclidean Distance Transform by Stack Filters", IEEE Signal Processing Lett. **6** (10) (1999).
9. T. Cormen, *Introduction to Algorithms* (MIT Press, McGraw-Hill, 2001).
10. H. Breu, J. Gil, D. Kirkpatrick, and M. Werman, "Linear Time Euclidean Distance Transform Algorithms," IEEE Trans. Pattern Anal. Mach. Intelligence **17** (5), 529–533 (1995).
11. L. M. Mestetskii, *Continuous Morphology of Binary Images: Figures, Skeletons, Circulars* (Fizmatlit, Moscow, 2009) [in Russian].

SPELL: 1. nondividing

Konstantin Viktorovich Nekrasov. Born 1991. Student of the 5th year of the Department of Mathematical Forecasting Methods of the Faculty of Computational Mathematics and Cybernetics of Moscow State University. Scientific interests: machine learning, graphic models, computer vision, optimization.

Scientific interests: Bayesian methods in machine learning, graphic models, computer vision, cognitive studies.

Dmitrii Petrovich Vetrov. Born 1981. Graduated from the Faculty of Computational Mathematics and Cybernetics of Moscow State University in 2003. Received candidate's degree in 2006. Researcher and Academic Secretary of the Department of Mathematical Forecasting Methods of the Faculty of Computational Mathematics and Cybernetics of Moscow State University, Head of the Laboratory of Neuroinformatics at the Kurchatov Scientific Center, Head of special seminars on Bayesian methods of machine learning, and Head of the group on Bayesian methods at Moscow State University. Author of more than 100 papers including 15 papers in scientific journals. Winner of the Presidential scholarship for graduate students (2005), Presidential grant for young Candidates of Science (2008–2009, 2010–2011), Moscow State University scholarships for talented young teachers and scientists (2010, 2011), and VMK scholarships for talented young teachers and scientists (2012). Head of two projects of the Russian Foundation for Basic Research.

Dmitrii Laptev. Born 1989. Graduated from the Faculty of Computational Mathematics and Cybernetics of Moscow State University and the Educational Scientific Center of Moscow State University. Postgraduate student of the ETH Zurich Department of Computer Science. Scientific interests: machine learning, graphic models, computer vision, optimization.