The article touches upon the issue of the edge detection of digital images in robotics and in other fields. Algorithms of the edge detection of the most often met images in static pictures are viewed as the target of research. In this paper, HOG-based characteristics are used to search for images of a human figure. When searching for faces in images, the most popular algorithm called Viola-Jones is used. To identify fingerprints, a correlation comparison and a special point comparison is used. It should be also noted that today the algorithms under discussion are used very often. The review of these algorithms has showed that they are sufficiently effective. The conclusions were drawn on the basis of the obtained results, and the branches of application of these algorithms, which are supported by the examples of processing various images, have been indicated.

**Keywords:** HOG, Viola-Jones, identification of fingerprints, detection of a human figure, detection of faces, edge, robotics, vision system, pattern recognition, boundary detection.

### PATTERN RECOGNITION IN THE VISION SYSTEM

**E. E. Pelevin, S. V. Balyasny**
Tula State University
92 Lenin Ave., 300012, Tula, Russia

**E-mail:** defqonstyles@gmail.com

Different information technologies have been implemented in all fields of human activity over the past few years. They allow automating various processes. Such an approach to changing one of the aspects of an activity allows achieving the greatest efficiency of a department or an enterprise in general. Paying attention to the value of modern information in robotics, much attention has been paid to identifying object forms [2,3]. A special attention is also paid to the identification of a person on the basis of certain biological signs such as forms of a human body, certain facial features, or fingerprints, etc. These kinds of tasks of image detection are often used in vision systems. The problem of recognizing such patterns was being solved by scientists of different countries, and various methods were used. Many of them did not find any application because of their inefficiency, or they demanded too many resources for their functioning. Despite the high competition in this area, today there are some algorithms that are sufficiently effective. The conclusions were drawn on the basis of the obtained results, and the branches of application of these algorithms, which are supported by the examples of processing various images, have been indicated.

### РАСПОЗНАВАНИЕ ОБРАЗОВ В СИСТЕМЕ ТЕХНИЧЕСКОГО ЗРЕНИЯ

**E. Е. Пелевин, С. В. Балясный**
Тульский государственный университет
Россия, 300012 г. Тула, пр. Ленина, 92

**E-mail:** defqonstyles@gmail.com

В статье поднимается вопрос выделения контуров образов на изображениях в робототехнике и других сферах деятельности. В качестве объектов исследования выступают алгоритмы выделения наиболее часто встречающихся образов на статических изображениях. В данной работе для поиска образов фигур человека осуществляется анализ по HOG признакам. При поиске лиц на изображении применяется наиболее популярный алгоритм Viola-Jones. Для идентификации отпечатков пальцев используется корреляционное сравнение и сравнение по особым точкам. Также следует отметить, что в настоящее время рассматриваемые алгоритмы применяются очень часто. Обзор данных алгоритмов показал, что они являются достаточно эффективными. На основе полученных результатов были сделаны выводы и указаны отрасли применения данных алгоритмов, которые подкрепляются примерами обработки различных изображений.

**Ключевые слова:** HOG, Viola-Jones, идентификация отпечатков пальцев, выделение фигуры человека, обнаружение лиц, контур, робототехника, система технического зрения, распознавание образов, выделение границ объекта.
Here, the term $CC(T,I)$ is added to the equation (1). A mutual correlation or, in other words, a simple correlation is a measure of the similarity of the two images. Considering the fact that when comparing two images, they can be rotated or displaced relatively to each other, it would be irrational to obtain a positive result with the help of overlapping the images $T$ and $I$ on each other and applying the equation (2).

If the image is shifted from the center, the following expression is used:

$$S(T,I) = \max(\Delta x, \Delta y) \cdot CC(T,I(\Delta x, \Delta y))$$

(3)

In this expression, $I(\Delta x, \Delta y)$ is the rotation of the original image ($I$) for an angle $\theta$ with the regard to the center $\Delta x, \Delta y$.

When using this expression, the computational complexity is quite high. To simplify the calculation, we can use the correlation theorem, which states that the calculation of the correlation in the space domain (the operator $\otimes$) is the equivalent of doing pointwise multiplication in the Fourier distribution, particularly:

$$T \otimes I = F^{-1}(F \times (T) \times F(I))$$

(4)

In this formula, $F$ is Fourier transform in the image. The sign $\cdot$ points at complex conjugation, and the sign $\times$ means pointwise multiplication of two vectors.

From this perspective, the result of the calculation according to the formula above is the correlation image, whose point value $[x,y]$ corresponds to the correlation between $T$ and $I$, and the displacement will be $\Delta x = x$ and $\Delta y = y$.

The test is run on the test image of the fingerprint (Figure 1).

![Figure 1. A fingerprint template](image1.png)

If the template is compared successfully with the existing one, the fingerprint will coincide and will be highlighted in green (Figure 2).

![Figure 2. Highlighting of the matched fingerprints](image2.png)

This method has only one drawback, which is related to the demand towards computing power. Nowadays, computers have fallen sufficiently in price. So, now this drawback can be considered as insignificant.

A frequent task is to find human figures in images or in computer generated images. This allows interactive robotics systems to detect a person, to carry on dialogues or other operations. There are many algorithms which are capable to detect similar images, but the detection with the help of HOG-based characteristics is considered to be effective. To determine these characteristics, several stages are used. The first is the gradient calculation. The most common method is the use of a one-dimensional differentiating mask in the horizontal and/or in the vertical direction. This method requires filtering the color or brightness component with the help of the following nuclear filtering:

$$[-1,0,1]$$

and $$[-1,0,1]^T$$

(5)

Scientists used more complex masks during the development. One of them is the $3 \times 3$ Sobel operator as well as various diagonal masks. Despite their effectiveness in solving other tasks, they show very low productivity here, which has generally an adverse impact on the vision system. To improve the final results, a Gaussian blur was also used in the studies, but, practically, this was not leading to the improvement of results.

The next step in the algorithm process is the calculation of histograms of cells. The basis of a histogram is the value of gradients. At the same time, each pixel of the source image participates in the weighted voting for each of the available channels of a histogram. The cells of the source image may be only of two shapes – rectangular and square, and the channels of a histogram may be within the limits of 0 and 180 or 0 and 360 degrees. The values in these ranges are distributed equally. The ranges of degree measures depend on whether a signed or unsigned gradient is used. Empirical research of many scientists shows that an unsigned gradient with 9 channels has achieved the best results in the recognition of humans.

In order to make the results more accurate, brightness and contrast of the points should be taken into account. For this, the gradient must be locally normalized. For this purpose, cells are combined into blocks of a larger size. The HOG descriptor is a vector of normalized cell histograms forming a block. As a rule, in this case, one cell can be used not in one block, but in several ones.

When applied, two types of such blocks are used: R-HOG (having a rectangular shape) and C-HOG (of a round shape). In most cases the first type of blocks represents a grid, which is described by three basic parameters – the number of cells included in one block, the number of pixels of an image belonging to one block, and the number of channels in each histogram of a cell.

R-HOG blocks are similar to the SIFT descriptors. Despite the similarity of the structure, R-HOG blocks are computed in grids at a fixed scale. However, the direction is not taken into account. In its turn, the SIFT descriptors are computed in image grids that are not insensitive towards the scale. In this case, rotation is used for changing the direction. It should be also noted that R-HOG blocks are used together for information coding about the form of objects, while the SIFT descriptors are used separately. C-HOG blocks can be found in two variants: those with a single, central cell and those with an angularly divided central cell. They are characterized by four basic parameters: by the number of sectors and rings, by the radius of the central ring and the expansion coefficient for the
radii of the rest of the rings.

Using these characteristics, the task of detecting the image of a person in an image is confined to the usage of the Run method. Let it be required to search all the objects of the size \( w \times h \) in the image. In this regard, rectangular regions of an image that measure \( w \times h \) with the upper left corner having the coordinates \( (dx,dy), i=[0,m], j=[0,n] \) are considered with the help of a little vertical \((dy)\) and a horizontal step \((dx)\). A classification is made for each of them. Thus, a mask (a window) with the size \( w \times h \) seems to run on the image. Further generalization of the method in order to search objects of different sizes is possible due to multiple scaling of an image and realization of detection by the method described above [1].

The results of the test (Figure 3) have showed that this method can effectively detect human contours in images of a sufficiently low quality, which reduces a computational load on the final device.

![Figure 3. The result of person detection by a dome camera](image)

The test image shows that the algorithm identifies only, if there is a large part of a person’s figure and if his/her contours are clear enough.

As a rule, in a simpler task we may not require to detect the whole person, but only his/her face as, for example, to identify a person by unique features. In this case, the Viola-Jones algorithm is applied. It was named after the scientists who had developed it. This algorithm uses a pre-generated base of features that are typical for the detection of a particular class of objects. A similar base can be obtained by searching for all combinations of Haar-like attributes and by classifiers that provide an error, i.e. mistakenly do not find an object in the image. It is not necessary to generate a database of features in order to search for faces in the images, but to use the ready one. For example, the file haarcascade_frontalface_alt.xml of the OpenCV project contains the set of required features, which is considered to be one of the most popular for image processing.

The cascade of the used features represents a set of several stages. The cascade stage is characterized by a set of specific features. For example, the cascade from the OpenCV library includes features that are rectangles (rects), and each of them is endowed with a positive or a negative weight. During the algorithm execution, a window measuring \( Wh \times Ww \) pixels, first linearly moves along the image horizontally and then, vertically. The initial dimensions of the moved window have to be chosen exactly the same as the window sizes in the cascade. After passing the next step, the window sizes increase. In order to increase, two methods are used. The first method consists in the calculating of a scalable coefficient, in the correction of coordinates and rectangles that are inside the features. The second method assumes the usage of a scaling of the original image. In this work, the window size is multiplied by the scaling factor \( \text{ITER}_\text{SCALE} \) after passing the next iteration. At this point, the window is shifted from the current coordinates horizontally (by \( w\text{.step}_x \) pixels) and vertically (by \( w\text{.step}_y \) pixels).

To calculate these displacements, the following expressions are used [4]:

\[
W_{\text{step}} = \max \left(1, \min \left(4 \frac{Wh}{10} \right) \right) \tag{6}
\]

\[
W_{\text{step}} = \max \left(1, \min \left(4 \frac{Ww}{10} \right) \right) \tag{7}
\]

To increase the accuracy for the window, it is necessary to calculate the normalization coefficient with the help of the following expression:

\[
\text{win}_\text{norm} = 1.0Wh \times Ww \tag{8}
\]

The sum of brightness of all pixels \((w\text{.sum})\) is calculated inside the window, and their squares \((w\text{.ssq})\) are summarized with the usage of the integral representation of the images. Then, the mathematical expectation of a pixel \((\text{mean})\), variance \((\text{var})\) and standard deviation is calculated:

\[
\text{mean} = \frac{W_{\text{sum}} \times \text{win}_\text{norm}}{ \text{win}_\text{norm} } \tag{9}
\]

\[
\text{var} = \left( \frac{W_{\text{sum}} \times \text{win}_\text{norm}}{ \text{win}_\text{norm} } \right) - \text{mean}^2 \tag{10}
\]

\[
\text{stddev} = \sqrt{\text{var}} \text{ if } \text{var} \geq 0 \text{ otherwise, } 0,1 \tag{11}
\]

For a more optimized work of the algorithm, it is necessary to take into account the thresholding constant \( \text{STDDEV}_\text{MIN} \). In addition, if the standard deviation \( \text{stddev} \) is less than this value, such window is missed, then. This approach allows to skip areas with low brightness, which indicates the absence of image boundaries in them.

The cascade used consists of a set of stages, where the coefficient \( \text{sum}_\text{stage} \) is summed on each stage. Then, the obtained value is compared with the critical value that has been calculated in advance. If the sum obtained is less than the critical value, then, the stage is not considered to be passed through, and the processing is interrupted. If all stages of the cascade have been passed through, then, it may be assumed that a sought pattern is found in the image (for this cascade it is a face). In order to regulate the operating threshold of the algorithm, the threshold of the mean-square deviation \( \text{STDDEV}_\text{FACE} \) is introduced. For all attributes included in the cascade, the features \( \text{sum}_\text{feature} \times \text{win}_\text{norm} \) are considered as normalized. These attributes are compared with the critical value that has been normalized before. In the OpenCV library, it is written in the cascade \( \text{feature} \to \text{threshold} \times \text{stddev} \). In cases when the obtained value is less than the critical one, the value of the left attribute sub-tree is summarized with its stage value (in other situations, it is done with the right sub-tree). The parameter \( \text{sum}_\text{feature} \) is the sum of brightness of all pixels that compose the rectangular area of the feature. This parameter is multiplied by the weight of the area where it is located. In order to take into account the scale, the coordinates of the region are additionally multiplied by the coefficient \( \text{win}_\text{scale} \).

Nowadays, a number of methods have been developed that allow to solve the problem of strict requirements for the
image of a face being in the frame with the help of building 3D models. This method supposes the installation of several cameras which must be previously synchronized between each other. In the case when a person appears in the frame, a camera takes synchronously a series of pictures from all angles. After that, a three-dimensional model of the person is created, and the analysis is made point by point. Further comparison of the obtained model with the template is carried out by the same algorithm as in the case of a 2D space.

On the basis of the obtained results of the work and having considered the algorithms of pattern recognition in images, it can be concluded that modern algorithms detect objects in the original images quite effectively, which indicates their high suitability in the vision system. They can be also used in security-related systems where person identification is required.

REFERENCES
1. Matveyev D.V. Razrabotka algoritmov analiza auditorii dlya sistem prikladnogo televizirovaniya [Algorithm Development of Audience Analysis for Applied Television Systems]. 2015. p. 58-61.
2. Pelevin E.E., Balyasnyi S.V. Optimal algorithm of edge detection within the system of computer vision // Juvenis scientia. 2016. № 6. p. 6-8.
3. Pelevin E.E., Balyasnyi S.V. The usage of adaptive threshold method in the system of computer vision // Juvenis scientia. 2017. № 1. p. 4-7.
4. Tarasikov A.S. Programma obnaruzheniya lica na izobrazhenii na osnovе karakteristicheskikh priznakov [The Program of Face Detection in an Image on the Basis of the Characteristic Features]. 2013. p.7-10.
5. D. Maltoni, D. Maio, A. K. Jain, S. Prabhakar Handbook of Fingerprint Recognition. 2005. p. 147.

Received 04.04.2017