Reducing dimensions of the histogram of oriented gradients (HOG) feature vector

Reducing dimensions of the histogram of oriented gradients (HOG) feature vector

S V Shidlovskiy*, A S Bondarchuk, S Poslavsky and M V Shikhman

[1]National Research Tomsk State University, 36 Lenin ave., 634050, Tomsk, Russia

*E-mail: ssv@mail.tsu.ru

Abstract. The article discusses the basic principles of a HOG-descriptor. A method is proposed for reducing dimensions of the feature vector obtained using the HOG-descriptor by applying the convolution operation and converting the resulting values of gradients and their directions. After using new feature vectors for training, the Support Vector Machine (SVM) classifier showed a 37-fold increase in performance when processing an image with a resolution of 282×159 pixels.

1. Introduction

Recently, the number of autonomous vehicles has grown significantly and computer vision systems allow autonomous vehicles to respond to the external environment and overcome various obstacles. The ability to respond to the environment and navigate in a given space is necessary for self-driving cars, especially in densely populated areas with a large number of road users.

Computer vision systems, which mainly use machine learning methods, allow driverless vehicles to recognize images of interest in the frames obtained from cameras. Pedestrian detection is important for driverless cars. Recognizing and tracking pedestrians allows you to react to them and avoid collisions, which is a necessary quality for any driverless car. In these types of recognition tasks, the speed of object detection is very important, since the reaction time allotted for necessary manoeuvres depends on it. To reduce the requirements for the computing system and improve the performance of systems, new approaches are needed in the field of classification algorithms.

This article discusses the use of directional gradients as a feature vector (HOG-descriptor) for classifying and detecting pedestrians in an image. A way to reduce the dimensions of the feature vector is outlined, to reduce the computational resources consumed while maintaining the recognition accuracy.

2. Histogram of Oriented Gradients

A descriptor is a special algorithm that extracts necessary information from an image and discards the rest. Descriptors are required for tasks such as pattern recognition and object detection in an image. The feature vector is a descriptor and consists of a certain number of consecutive values that describe the characteristic features of the image. Feature vectors are used to classify objects in an image using algorithms such as SVM.

HOG are special point descriptors that are used in computer vision and image processing for object recognition. This feature is based on calculating the number of gradient directions in local areas of the image [1-3].

The main idea of the algorithm is to assume that the appearance and shape of an object in an image can be described by the distribution of intensity gradients or the direction of the edges. Gradient values are calculated in the horizontal and / or vertical direction using a one-dimensional differentiating mask. This method requires filtering the colour or brightness component using the following filter cores: [-1,0,1] and $[-1,0,1]^T$.

To calculate the HOG-descriptor, the gradient value m and the direction of gradient θ of each pixel in the image are calculated using the following equations (1-4):

$$g_x = f(x+1, y) - f(x-1, y), \tag{1}$$

$$g_y = f(x, y+1) - f(x, y-1), \tag{2}$$

$$m(x, y) = \sqrt{g_x^2 + g_y^2}, \tag{3}$$

$$\theta(x, y) = \tan^{-1}\left(\frac{g_y}{g_x}\right). \tag{4}$$

where $f(x, y)$ is the pixel brightness value with coordinates (x, y), $\theta(x, y)$ is the direction of gradient, $m(x, y)$ is the pixel gradient value with coordinates (x, y).

For pixels on the image border, there is not enough information to calculate the magnitude and direction values, so these pixels are not used. Gradients are displayed on the image of the place where there is a clear change in the brightness of pixels. The value of gradients is greater at the edges and corners of the object, which contain much more information about the shape of the object than homogeneous areas. In this way, using gradients, highlights the contours of objects and discards unnecessary information, like a uniform coloured background. At each pixel, the gradient has a value and direction. For colour images, the gradients of three red-green-blue channels are evaluated. Then the gradient values are averaged.

The gradient direction is calculated in radians and takes values in $(-\pi, \pi)$, which is called a "signed gradient". Then radians are converted to degrees and get values from 0° to 360°. It is better to use “unsigned” gradients to improve performance. Object detection in an image is mainly based on edge detection, which does not require knowledge of the gradient sign since the pixel colours are not as important. For example, a gradient direction of 90° and the opposite direction of 270° are considered the same. Since it is not so much important where the gradient is directed, but the general trend of changing pixel intensity is important to determine the border of objects. In other words, gradients must be converted so that their values are in the range from 0° to 180°. The solution to this transformation is to add the number π to the negative directions.

After finding the value and direction of the gradient for each pixel in the cropped part of the image, the image is divided into small spatial areas called cells. For example, the image was divided into 8-by-8-pixel cells. Figure 1 shows a single image cell overlaid with HOG vectors. Arrows indicate the direction of the gradient, and their lengths depend on its value.

For each cell, a histogram of gradient directions is calculated, in which each bin represents the sum of the values of gradients inside the cell in a certain range of directions [4]. The histogram consists of 9 bins, and each bin represents a certain angle of the gradient direction. The range of the histogram is from 0 to 180°, with the range of each bin is 20°.

The distribution of values over the histogram bin is as follows. First, bins are defined for each pixel (there can be two or one of them), for values to be entered. This value is then calculated using linear interpolation relative to the centre of these bins. The values entered in the bin are added together to create a histogram. If the gradient direction angle is greater than 170°, the value of this gradient is divided proportionally between bin 170 and 10. Figure 2 shows an example of calculating the HOG histogram distribution.

Here, for an angle of 177.33° and a value of 43.04, the value entered in bin 10 is calculated as:

$$((177.33 - 170) \cdot 43.04) / 20 = 15.77.$$

The value entered in the bin 170 is calculated as:

43.04 - 15.77 = 27.27.

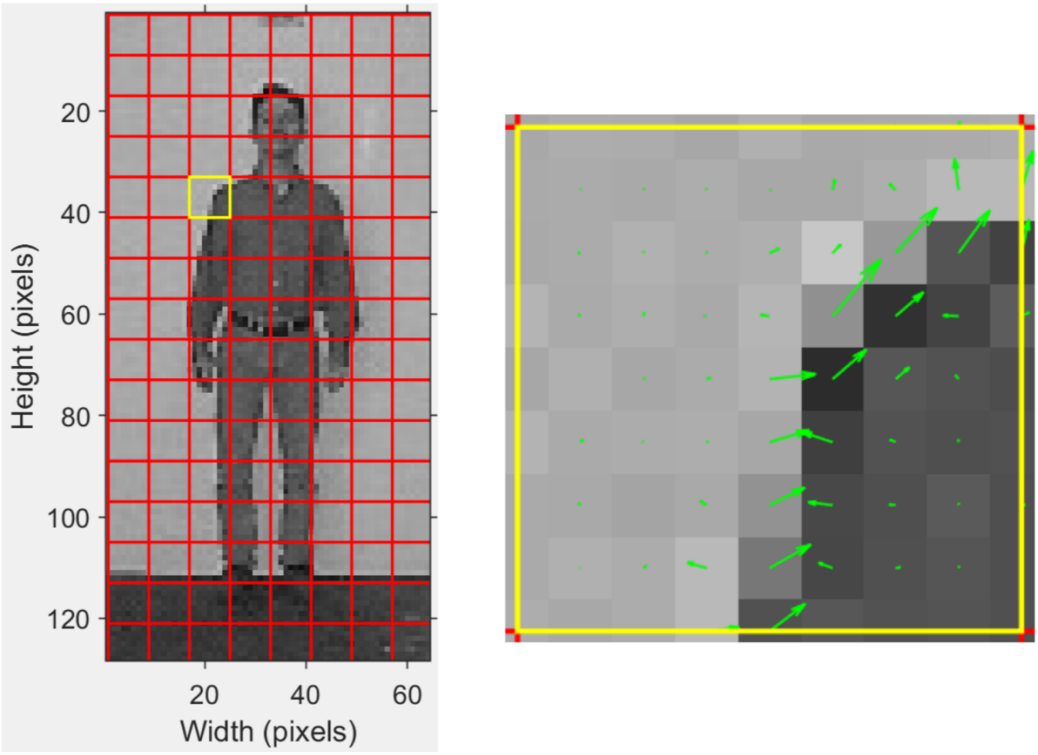

Figure 1. Visualization of the gradients.

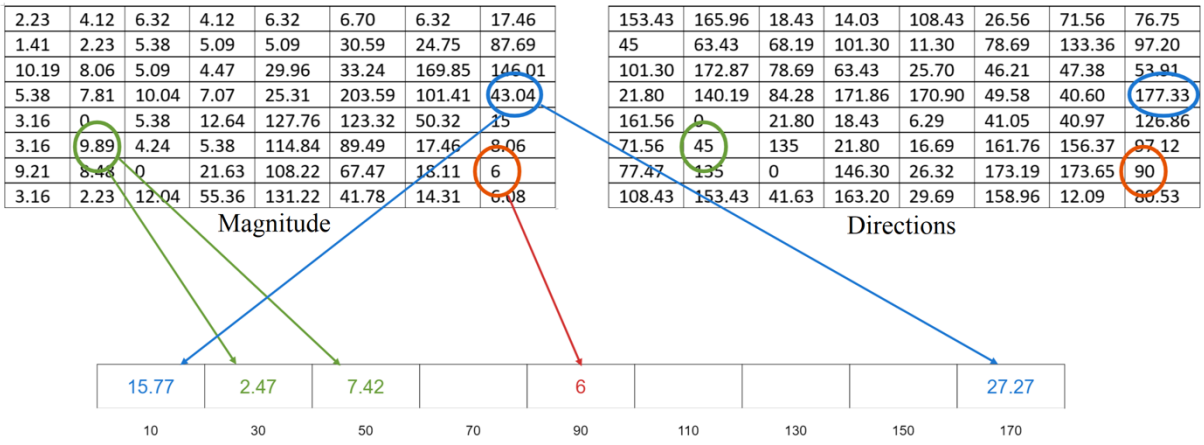

Figure 2. Distribution of values in the HOG histogram.

After building a histogram for each cell, the cells are combined into blocks. The blocks represent the overlapping areas in the image which contain few cells. For example, let's take a block with a size of 2 by 2 cells. The block is used to define cells whose histograms should be normalized together. The point of this is to make the feature vector resistant to changes in pixel value, for example, due to changes in lighting. Figure 3 shows the diagram for building a block histogram.

Since each block contains four cells, it combines four histograms. Therefore, the block histogram consists of 36 bins whose values are normalized by the L2-norm.

Mathematical representation of the L2-norm:

$$H_{L2} = \frac{H}{\sqrt{H_1^2 + H_2^2 + ... + H_n^2}}.$$

Where,
H_{L2} is the normalized histogram,
H is the block histogram,
n is the number of bins in the histogram,
H_n is the value of the n-th histogram bin.

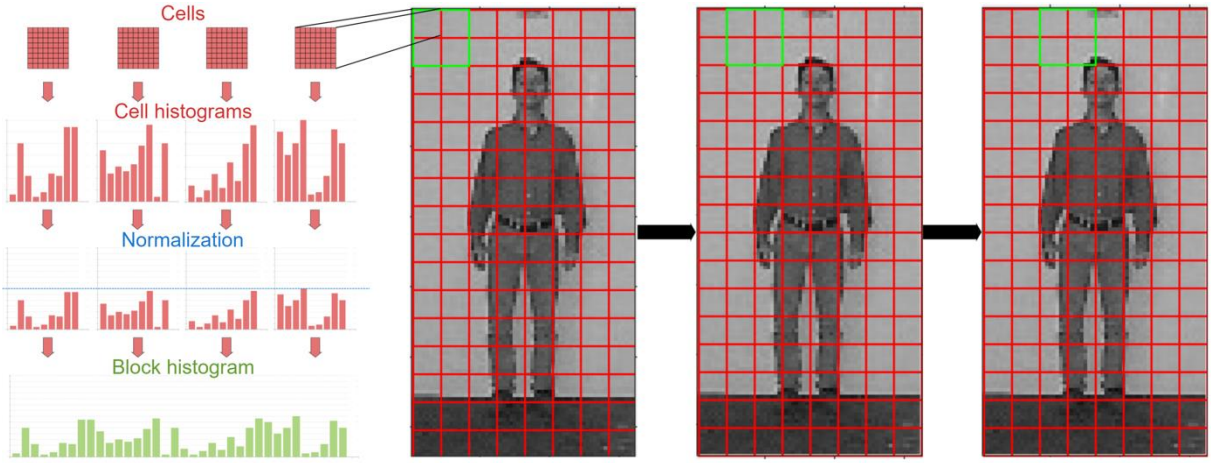

Figure 3. Construction of the histogram block.

Finally, the normalized histograms of all blocks are concatinated into one large histogram that characterizes the entire image. For a 64×128 image, the number of 2-by-2 cell blocks is 105. The histogram size of a single block is equal to 36 bins. Therefore, the histogram size of the entire image will be equal to 3,780.

Figure 4 shows a visualization of the HOG-descriptor. The image is overlaid with vectors of normalized histograms of 9 bin size in cells of 8×8 pixels. The dominant directions of histograms reflect the shape of a person.

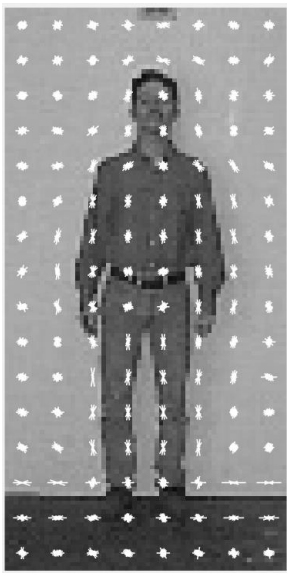

Figure 4. Visualization of HOG-descriptor.

3. Method for reducing the dimension of the HOG-descriptor feature vector

The classic HOG-descriptor results is a feature vector ***h***, which is then used to classify the source image using various machine learning methods (for example, the SVM classifier).

As a result of applying equations (1-4) to the original image, we obtain a matrix ***M*** of gradient values for each pixel of the original image and a matrix ***TETA*** of the directions of these gradients. Then the matrix data is converted to vectors ***m*** and ***θ*** using the following example:

$$\boldsymbol{M} = \begin{pmatrix} m_{11}\ m_{12}\ \ldots\ m_{1j} \\ m_{21}\ m_{22}\ \ldots\ m_{2j} \\ \ldots \\ m_{i1}\ m_{i2}\ \ldots\ m_{ij} \end{pmatrix} \rightarrow \boldsymbol{m} = \left(m_{11}\ \ldots\ m_{1j}\ m_{21}\ \ldots\ m_{2j}\ \ldots\ m_{i1}\ \ldots\ m_{ij}\right).$$

After that the vector ***h*** is formed by applying concatenation:

$$\boldsymbol{h} = \left(m_{11}\ \ldots\ m_{ij}\ \theta_{11}\ \ldots\ \theta_{ij}\right).$$

For the image considered above (Figure 4), which has a dimension of 64×128 pixels, the resulting feature vector ***h*** has a length equal to 15,624 (without taking into account the pixels on the image borders). To reduce the feature vectors, it is proposed to convert the resulting values of the pixel gradient and its direction by using the convolution operation (Figure 5). Convolution of the ***M*** and ***TETA*** matrices is performed using the sliding window principle with a vertical and horizontal strides equal to 2 (Figure 6). The length of the obtained feature vector is 3,720.

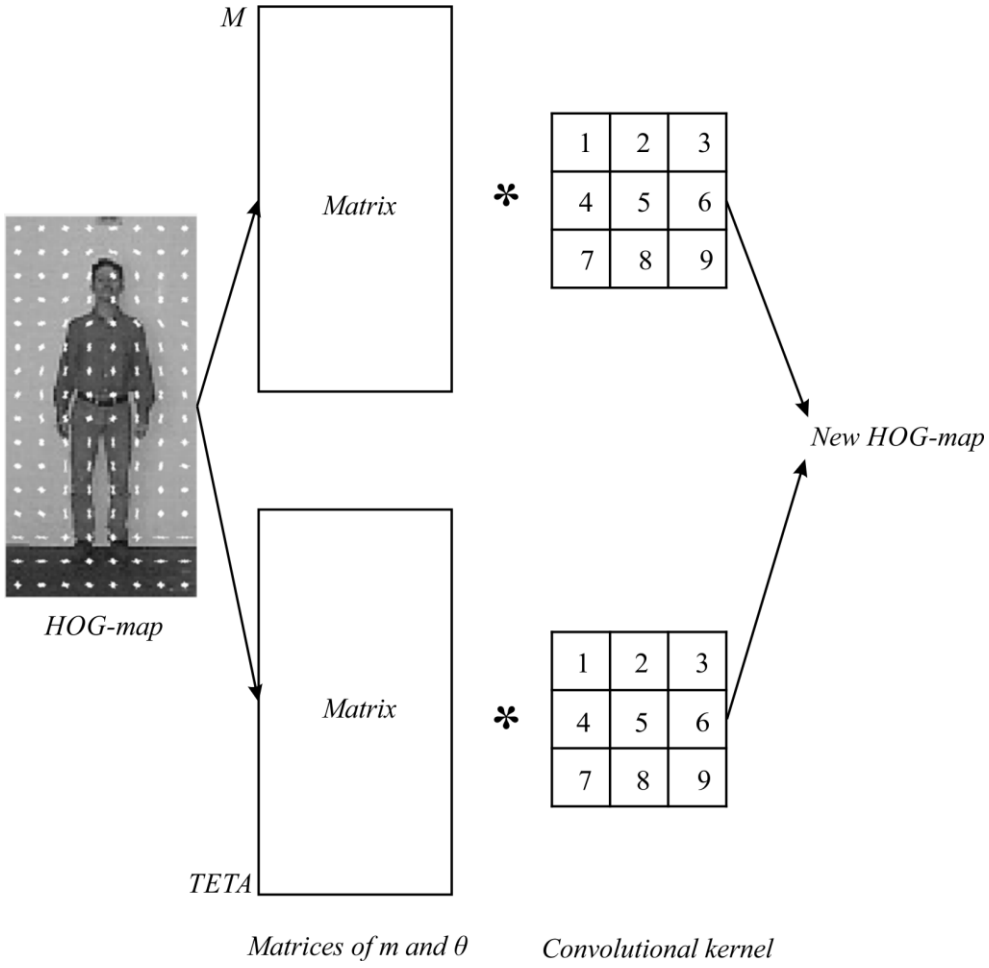

Figure 5. Proposed convolution.

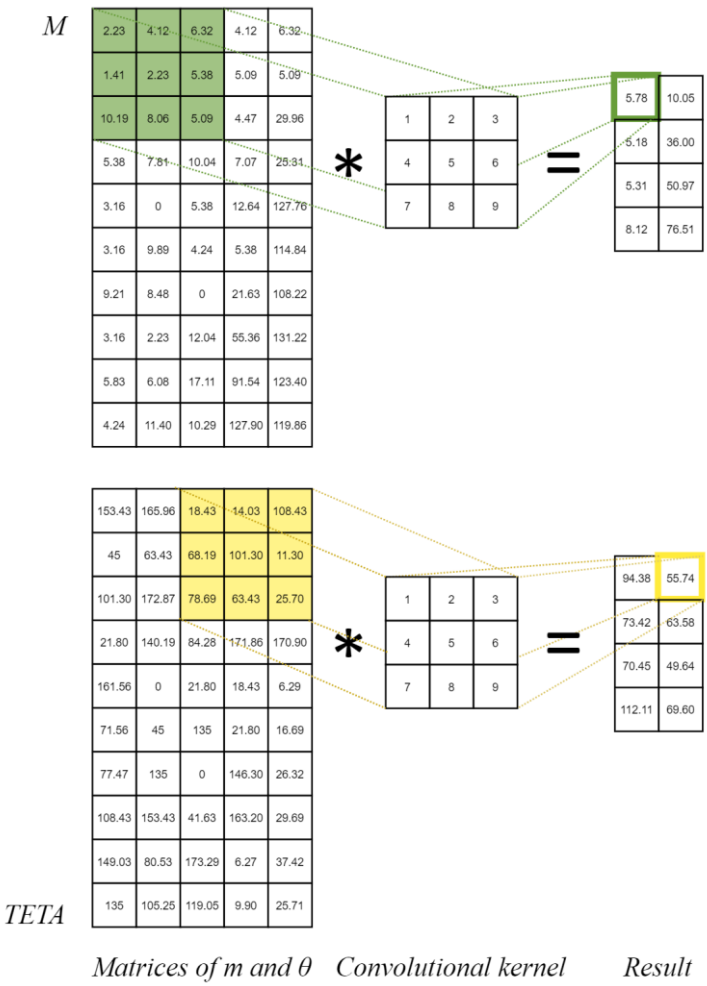

Figure 6. The example of using a convolution operation.

The resulting method for generating the HOG-descriptor feature vector was used to train the SVM classifier on the pedestrian database of Massachusetts Institute of Technology (MIT), the Center for Biological and Computational Learning (CBCL). After that, the time for processing and classifying pedestrians in an image with a dimension of 282×159 pixels was measured (table 1).

Table 1. Comparison of SVM speed with new feature vectors.

	SVM +Classical HOG	SVM + New HOG
Processing time, s	260	7

SVM analysis was performed in MATLAB R2020a using a computer with the following characteristics:

- Intel(R) Core(TM) i9-9880H CPU @ 2.30GHz;
- 32 Gb RAM;
- NVIDIA Geforce RTX 2080.

4. Conclusion

The paper discusses the principles of constructing the HOG-descriptor feature vector. To improve the efficiency of classifiers based on the HOG-descriptor, a method for reducing the resulting feature vector is proposed. Using computer simulation methods in MATLAB R2020a, it was shown that the SVM classifier trained on reduced feature vectors processes an image with a dimension of 282×159 pixels 37 times faster than the classic HOG. Further research will focus on implementing the descriptor and classifier in high-performance computing systems, including Field-Programmable Gate Arrays (FPGA) [5,6].

Acknowledgements

The reported study was funded by RFBR, project number 19-29-06078.

References

[1] Dalal N and Triggs B 2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (San Diego, USA: IEEE) pp. 886-893
[2] Zhang S, Benenson R and Schiele B 2015 *IEEE Conf. on Computer Vision and Pattern Recognition* (Boston: USA) pp. 1751-1760
[3] Pang Y, Yuan Y, Li X and Pan J 2011 *Signal Processing* **91** 773
[4] Vondrick C, Khosla A and Torralba A 2013 *IEEE International Conference on Computer Vision* (Sydney, NSW, Australia: IEEE) pp. 1-8
[5] The Ngueyn C and Shashev D 2018 *MATEC Web of Conf.* **155** 01016
[6] Shatravin V and Shashev D V 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **363** 012028