Additional Section. (Trends of modern computer systems. 15.06.2020.)

UDC 004.8

Pavlo Serhiienko, Mariia Orlova, Heorhii Loutskii.

OVERVIEW OF ALGORITHMS FOR GRAPHIC PATTERNS RECOGNITION IN IMAGES

The article considers the features of technical vision as a process of pattern recognition in images. The review and comparison of characteristic features of actual recognition algorithms, and conclusions about their possibility of realization on FPGA of the average size are carried out. A conclusion is made that the methods based on the scalable feature extraction in the image are perspective ones.

Key words: FPGA, technical vision, SIFT.

Actuality of theme. At present, the Advanced Driver Assistance System (ADAS) technology is widely expanded over the world. One of the most difficult tasks in ADAS is the timely recognition of obstacles and critical situations in the road environment. A similar challenge is facing modern weapon control systems, which must recognize the targets reliably. Therefore, the new technical vision architectures are developed by a number of companies. Some of them like AMD, Aldec, Lattice, Intel, Huawei are corporated in the Embedded Vision Alliance.

The technical problem is that existing systems for the object recognition have high hardware costs, high power consumption, insufficient speed, or insufficient probability of correct recognition. There is an urgent need to create reliable autonomous vehicle control systems and target recognition systems, which have a reasonable costs and ensure high reliability. The problem that needs to be addressed is to reduce hardware costs, increase performance, and increase the likelihood of object recognition.

Pattern recognition basics. The digital image processing is the processing of a two-dimensional representation of a spatial object or scene using various computer tools. The image recognition techniques are based on procedures for image enhancement, restoration, reconstruction, analysis, and pattern recognition.

The pattern recognition is a component of machine learning theory. There are two approaches to learning: inductive and deductive ones. The inductive learning, or learning by precedent, is based on identifying the general properties of objects based on empirical information. The deductive learning involves the formalization of expert knowledge in the form of expert systems. The graphic image recognition can be divided into a number of subtasks: feature generation, feature selection, classifier

construction, and classifier estimation. The most important steps are to generate the features that characterize the object, and to select the features. These two steps are the most time consuming because they must compress the information in the image from megabytes to kilobytes. In this case, the extracted features must effectively characterize the object to ensure reliable recognition in adverse conditions.

Currently, methods of recognizing graphic images based on artificial neural networks and methods based on the detection of characteristic points have become widespread. These methods are discussed below in detail.

Pattern recognition in artificial neural networks. Artificial neural networks have been recognized as a powerful tool for studying and reproducing systems in various applications. Neural networks resemble the behavior of the brain and consist of one or more layers of neurons connected to each other. Neural networks have the ability to learn from input data. The learning process is achieved by changing the architecture (connections) of the networks and the connection weights according to this information. The network topology and associated weights can be derived from both training data and application knowledge, or a combination of the two. Such systems have advantages in ease of construction by users who are not able to create a functional structure of the model, but this will be a significant disadvantage if such a structure can be found, for example, analytically [1].

The functional networks [1] are an alternative to the neural networks that can combine information from both input and application. The functional networks require knowledge of the scope to obtain the functional equations and to make assumptions about the form that the unknown functions should perform. It was found that the functional equations significantly reduce the degree of freedom of the initial neural functions, which leads to the significant simplifications of the network.

The convolutional neural net or CNN has become widespread for the recognition of graphic images. In essence, they belong to the functional networks. They have primary and secondary processing units. The first unit converts the image with a certain set of filter cores, resulting in a series of images that are constructed from certain local features, such as edges, spots, angles. The second unit is implemented, in fact, as a neural network that performs the final pattern recognition.

A special type of the convolutional neural network are the deep learning networks (deep learning CNN). Such a network consists of several layers of CNN, each of them performs the compression of information and recognition of more generalized images. To obtain one pixel of an image that consists of local features, CNN must perform from 100 to 100 thousand arithmetic and logic operations. In addition, each node of the neural network performs hundreds and thousands of arithmetic operations when processing such pixels that come to it from neighboring

nodes. And these operations are usually performed with the floating-point data. Due to the complexity of the objects being recognized, the size of the images, and the need to perform the task in real-time, CNN requires extremely high computational costs to implement.

Therefore, the successful modern CNN image recognition projects are usually based on large systems composed of high-performance graphics accelerators (GPUs), the total performance of which is many teraflops. A typical technical solution is to run CNN in a large cloud server.

The CNN learning process is very time consuming, as it requires up to hundreds of thousands of images to be submitted to CNN. But after such a training, the structure of CNN is stable, and the number of interconnections between the nodes is reduced by hundreds of times. Therefore, there is a technology by which CNN after training in the practical stage, is performed in another, simplified computer system.

Recently, there is a tendency to implement such a system in FPGA. However, there is still the problem of translating the arithmetic operations and truncating the coefficients to the short integers. In addition, it is almost impossible to organize the additional training or retraining of such a network. Primary the processing units belong to the field of digital signal processing. Therefore, they are often performed with a fixed point arithmetic. Moreover, there is a tendency to perform this part of CNN in ASIC. For example, the Google company has developed and implemented an ASIC that performs the convolution using 65,000 multiplication blocks and an adder tree with 65,000 inputs, and is CNN is the part of huge programable CNN.

Thus, CNN shows the great advances in image recognition. But they require large amounts of floating-point computations, which also entails large dimensions and high power consumption of the image recognition systems. This factor significantly limits their application and efficiency in their implementation in the portable devices and devices that perform processing of the moving images.

Methods based on feature point detecting. The approach of recognizing graphic images by identifying features is based on abstracting the information of the image at each of its points and deciding that at this particular point there is a certain characteristic feature of the image. Thus, the detected image characteristics are presented in the form of connected areas, isolated points and continuous curves. Further, by analyzing the detection and formed characteristic points (feature descriptors) and their mutual position, the system is able to recognize images that have similar points and their positions.

Unlike CNN, here image recognition is performed purposefully using existing image recognition methods based on dividing the image space into clusters and finding the image by distance to the center of the cluster. Because feature descriptors take into

account rotations and image scale, the learning phase is much faster. In addition, when calculating and comparing descriptors, it is not necessary to perform such complex calculations as in neural network nodes.

The detection of characteristics varies according to the type of application, the properties of the image and therefore can not be associated with an accurate or universal definition. A characteristic feature of the image can be considered a function of the local area of the image. Many computer vision algorithms use such a function as a starting point, so the success of these algorithms depends on how well the property detector is performed. Feature repeatability is a necessary property of a feature detector that looks for similarities between two or more different images of the same type.

Processing with a characteristic feature is the first operation performed on an image. In this case, each pixel is checked whether this function takes an extreme value on it or not. Thus, such detection is considered a low-level operation in image processing.

The feature detector does not process the entire image, but only certain regions. In many methods, images with different scales are processed by a two-dimensional low-pass filter with a Gaussian core. And this is the initial stage of performing the functions of characteristic features. Often such a function is calculated as the brightness derivative in the vicinity of a given pixel.

Property detection is the finding of a key pixel that is the center of a local area that bears the characteristics of that property. Once a detector has found a property, the same property is searched in another image by the same detector. Moreover, the properties of the same name in both images must be defined uniformly. That is, the property must be described in some standard way, and the process of such a description is called a feature description.

Types of image features. Determination of image properties by the stage of primary processing of various machine vision algorithms. The methods of determining the features differ significantly in repeatability, computational complexity and types of such features. But it is constant for most methods of finding such features as edge, corner, spot, blob.

A blob is not only a blurred point, but also a small area with a certain local structure (region of interest, ROI). The spot may be an angle, but the blurred angle may not be recognized as an angle, but detected as a spot.

Ridge is the name for elongated objects. Often such an object has an axis of symmetry. The road in the aerial photograph, the blood vessel in the medical image are the most common examples of ridges.

Image feature searching methods. The Scale Invariant Feature Detection (SIFT) method is the most famous one. It has 4 stages:

1. Detection of extreme points. The image is presented in several scales and contains points of local brightness gradient maximum and minimum.

2. Localization of feature points. Stable points are chosen among the extreme points and noise.

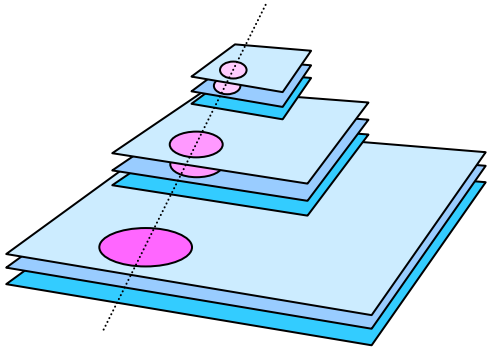

Fig.1. Pyramid of processed images

3. Determining the direction. For each feature point, the brightness gradient direction vectors in their vicinity are determined. Therefore, the feature points have parameters of coordinates, orientation, scale which are invariant to the image transformations.

4. Feature descriptor formation. The local image around the feature point is rotated to the found direction angle. The gradients in it are measured and are converted into the distribution diagrams. Such an image representation is invariant to scale and is actually called SIFT.

As a result, a feature point is represented by a multidimensional tag (more than a hundred coordinates), which represents it in a wide range of coordinates and scales. But such point information is much smaller than the pixel number: for a 500x500 pixel image, there are about 2000 stable points (although their number depends on the scene). The feature points are convenient to use for the reliable recognition of images. For example, to identify a small object, it is enough to recognize 3 adjacent points. Such sets of points can be stored in an image database.

To recognize the image in the new image, it is transformed into a set of feature points. The feature points of the object being searched are then compared with the points in the image by calculating the Euclidean distance. The SIFT method provides stable object recognition with different scales even when it is rotated up to 30°. But the method did not achieve wide propagation because of high complexity and that it is patented [1]. To minimize these disadvantages a set of this method improvements was developed.

The Histogram of Oriented Gradients (HOG) method was adapted to recognize pedestrians, animals and cars in static images. According to the method, the image is

first divided into small square cells, which calculate the gradients and their distribution. The latter is normalized by a function that is specific to a particular object [2].

The Dense Scale Invariant Feature Transform (DSIFT) method is a more carefully developed version of SIFT. However, its descriptors are not able to represent different scale and rotation, due to which they began to be processed much faster. [3].

The Speeded up Robust Features (SURF) method, like the SIFT method, is based on the detection of local properties, but is faster and uses an algorithm that protects against errors. The speed-up is derived by the use of more simpler Haar and wavelet functions, integer approximation of functions [4].

The Binary Robust Independent Elementary Features (BRIEF) method is based on a sample of templates consisting of about 256 image point comparisons that are randomly selected by a Gaussian distribution from around the feature point. In terms of recognition efficiency, the method is close to SIFT, but is much faster [5].

The Oriented FAST and Rotated BRIEF (ORB) method calculates the center of gravity of the feature point and the direction to this center. The method is a fast and effective alternative to the SIFT method. [6].

The Binary Robust Invariant Scalable Keypoints (BRISK) method defines a binary image of blurred spots scattered around a feature point. Then such points are grouped into pairs for which the distance and gradient are calculated. Actually, BRISK is a 512-bit binary descriptor, which is calculated through the Gaussian distribution of peaks around a key point. Due to this simplification, the recognition is accelerated at times [7].

The Fast Retina Key point (FREAK) method, unlike BRISK, improves the selection of key point pairs and image search. The descriptor is calculated taking into account the patterns of functioning of the retina. In this case, the points involved in the calculation of the Gaussian distribution are concentrated around a feature point. This helps to speed up the recognition by an order of magnitude [8].

HDR Adaptive filter is an edge-preserving filter, it is a method that works in two stages: Logarithmic-characteristic transformation (LHP) and low-pass filter (LPF). In the first stage, all pixels are classified as one of the colors, depending on the feature, for example, all vertical contours are marked in red (see Fig. 2). The output brightness is put in logarithmic scale and normalized. In this way it is possible to obtain a simple and concise descriptor of the features. In the second stage, the image noise is removed and key points such as line intersection, angle, or something else are recognized. To ensure resilience to scaling, the algorithm is run several times, for the same neighborhood with different scales of decimation [9].

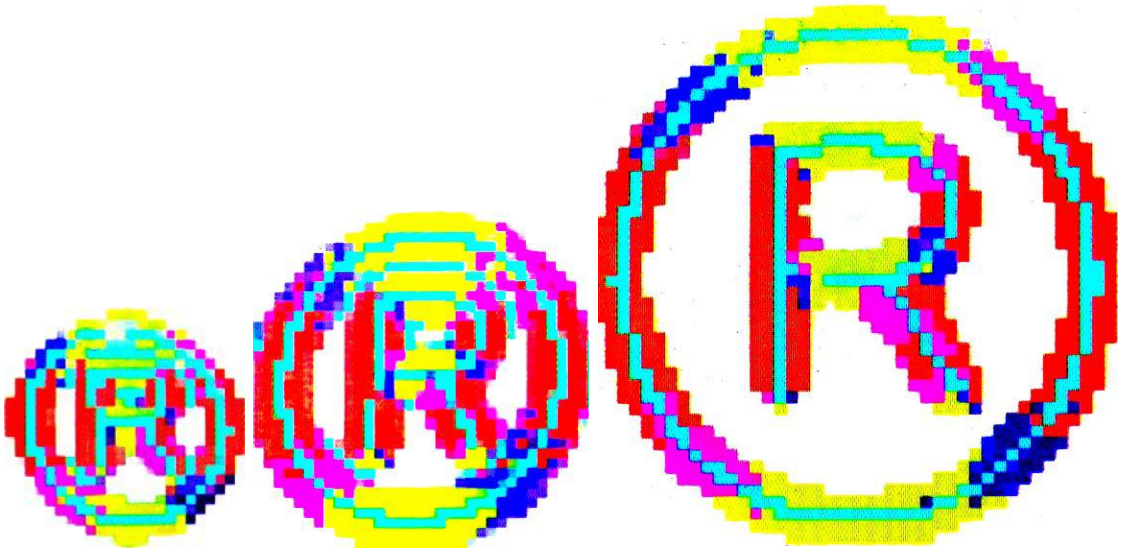

Fig. 2. Results of the image analyzer

Conclusions.

After analyzing common methods of recognizing graphic images, we can draw the following preliminary conclusions.

Common methods, which are based on the neural network, require excessive amounts of floating-point mathematical calculations and therefore they are not very suitable for implementation in FPGA.

Methods that are based on finding characteristic points and working with their descriptors are effective and allow their implementation in computing systems with moderate performance. Many of these methods work with integer and even bit data, so they require several times less hardware costs in hardware implementation and are therefore the best candidates for implementation in FPGA.

Among the considered characteristic point methods, the SIFT method is the oldest, but also the most reliable due to the fact that it provides recognition in poor lighting conditions, when returning images with a changed scale. This method is promising for implementation in FPGA, but it needs modernization. In particular, it is necessary to simplify the finding of characteristic points, the formation and processing of their descriptors.

References

1. Lowe D.G. Distinctive image features from scale-invariant keypoints / D.G. Lowe // International Journal of Computer Vision, 2004. —V. 60. — N2. — P. 91-110.

2. Speeded Up Robust Features : http://en.wikipedia.org/wiki/SURF Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary Bradski "ORB: an efficient alternative to SIFT or SURF", Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.

3. Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua, et al. “Brief: Binary robust independent elementary features.” Computer Vision–ECCV 2010. Springer Berlin Heidelberg, 2010. 778-792.

4. Lowe D.G. Object recognition from local scale-invariant features / D.G. Lowe / International Conference on Computer Vision, Corfu, Greece, 1999. — P. 1150-1157.

5. Mikolajczyk, K. An affine invariant interest point detector / Mikolajczyk, K., and Schmid, C // European Conference on Computer Vision (ECCV), Copenhagen, Denmark, 2002. — P. 128-142.

6. Brown, M. Invariant features from interest point groups / Brown, M. and Lowe, D.G. // British Machine Vision Conference, Cardiff, Wales, 2002. — P. 656-665.

7. Stefan Leutenegger, Margarita Chli and Roland Y. Siegwart. BRISK: Binary Robust Invariant Scalable Keypoints. ICCV 2011.

8. Alexandre Alahi, Raphael Ortiz, Pierre Vandergheynst. FREAK: Fast Retina Keypoint. In IEEE Conference on Computer Vision and Pattern Recognition, 2012.

9. Anatoliy Sergiyenko, Pavlo Serhiienko, and Juriy Zorin. High Dynamic Range Video Camera with Elements of the Pattern Recognition // ELNANO-2018, Lviw, Ukraine, 2018. – P. 435-438.

AUTHORS

Serhiienko Pavlo - PhD student, Department of System Programming and Specialized Computer Systems, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

paulsrgnk002@gmail.com

Mariia Orliva (supervisor) – associate professor, Department of System Programming and Specialized Computer Systems, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

Heorhii Loutskii (supervisor) – professor, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.