SERHIIENKO P.,
SERGIYENKO A.,
ORLOVA M.

LOCAL FEATURE EXTRACTION IN IMAGES

The methods of the local feature point extraction are analyzed. The analysis shows that the most effective detectors are based on the brightness gradient determination. They usually use the Harris angle detector, which is complex in calculations. The algorithm complexity minimization contradicts both the detector effectiveness and to the high dynamic range of the analyzed image. As a result, the high-speed methods could not recognize the feature points in the heavy luminance conditions.

The modification of the high dynamic range (HDR) image compression algorithm based on the Retinex method is proposed. It contains an adaptive filter, which preserves the image edges. The filter is based on a set of feature detectors performing the Harris-Laplace transform which is much simpler than the Harris angle detector. A prototype of the HDR video camera is designed which provides sharp images. Its structure simplifies the design of the artificial intelligence engine, which is implemented in FPGA of medium or large size.

Keywords: FPGA, feature extraction, HDR, pattern recognition, artificial intelligence.

1.Introduction

The local feature points play an important role in computer vision and pattern recognition, including image matching, object recognition or clustering, image construction, object tracking, face recognition, image registration. The need for rapid detection of the feature points is manifested in applications of computer vision to obtain content-based image retrieval in robots and autonomous driving [1].

Modern approaches to identifying the local feature points of the object provide taking into account the scale, so the object can be recognized regardless of its apparent size. An object can also be characterized by a set of feature points. This allows the system to recognize the object viewed from a different angle and distance. Using the relative position of the points also provides to recognize complex objects in conditions of noise and uneven lighting.

The definition of the feature points is usually performed in two stages. At the first stage, a point is detected using the appropriate algorithm named the detector. At the second stage, the found feature point information is encoded to the descriptor.

If the image moves, rotates, scales along two axes, and the detection result remains unchanged, it means that the recognition of the points is stable, and the found points are feature points. Fig. 1 shows an example of finding a set of feature points which helps to recognize the object.

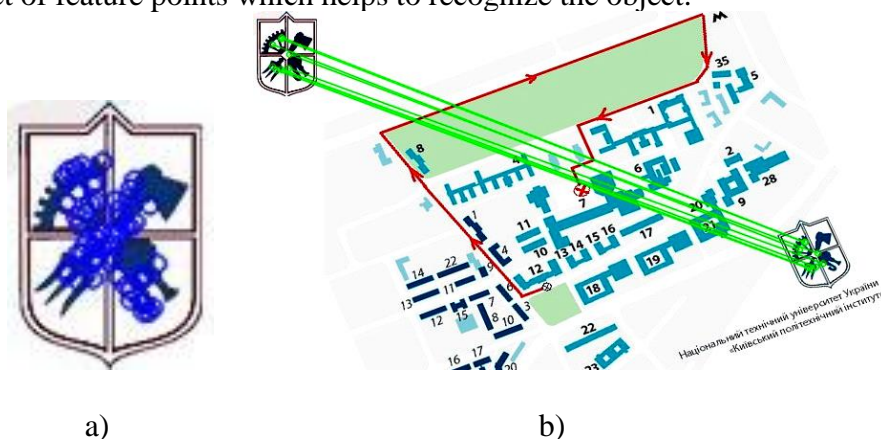a)   b)

Fig.1. Extracted feature points (a) and their use for the pattern recognition (b)

The feature point descriptors should describe the key characteristics of the image with repeatability, compactness, accuracy, and efficient representation, which are consistent and reliable in terms of scaling, rotation, display, occlusion, and lighting [2]. In this article, the methods of the local feature extraction are reviewed to select the best-fitted method for the pattern recognition in the heavy lightness conditions.

2.Image feature detection methods

Algorithms or methods for detecting and constructing a feature descriptor are divided into six categories:
— low-level detectors;
— methods based on the analysis of the brightness gradient;
— methods based on contrast analysis;
— methods based on the analysis of spatial frequency;
— learning-based methods;
— convolutional neural networks.

2.1 Low-level detectors

The low-level features are the features of the image that can be identified in it without any shape information about the object. Such features can be detected by the threshold processing, the functions sensitive to the edge, corner. The Laplace detector calculates the second-order derivative and where the result intersects the zero level, there is the edge of the form.

The Sobel detectors are based on the first derivatives of the pixel $I(x, y)$ to be considered in the coordinates (x, y). The Canny detector uses the Gaussian filtration

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y),$$

where $G(x, y, \sigma)$ is the Gauss kernel, σ is the filtering scale, and $*$ is the convolution operator. Then, the image with the highlighted edge features is calculated as

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma). \qquad (1)$$

This filter is an approximation of the Laplace filter but is simpler in calculations. This advantage is utilized in many complex detectors.

Because the derivative function is sensitive to noise, the rejection of weak results, median filter, low-pass filter, and Gaussian filter are often used for the resulting improvement. And the limit of the rejection is usually selected manually.

The comparison of edge detectors highlights their shortcomings: incomplete contours of the closed figures, the need for selective thresholds, sensitivity to noise. In addition, the threshold should be selected separately for regions with different local lighting [3].

The detector with phase congruence operates on the principle that at the edge point the sinusoidal components of the signal obtained by the Fourier transform or wavelet transform have the same phase. But this detector is also sensitive to noise and has low localization accuracy [4].

If the edge of an object is derived, then the angle and curvature detectors can find the feature points in it. The curvature is defined as the rate of change of the direction of a flat curve that describes a figure. But in all cases, finding a closed curve and its analytical description is a difficult task [5].

The degree of the curvature can be obtained by considering changes in brightness along with alternate directions in the image. This is the basic idea of the Moravec angle detection operator. This operator calculates the average change in image intensity when the analysis window is shifted in several directions. That is, for a pixel $I(x, y)$ which stays in the centrum of a window $w(x,y)$ we have a detector output, which is a measure of belonging to the line

$$E_{uv}(x,y) = \sum_{j \in w} (I(i, j) - I(u+i, v+j))^2.$$

The shifts (u,v) are performed in four directions: $(1,0)$, $(0, -1)$, $(0,1)$ and $(-1,0)$. This equation is an approximation of the brightness autocorrelation function in the (u,v) direction. If the pixel belongs to a straight line, the value of $E_{uv}(x,y)$ is small for the offset along the line and large for the offset perpendicular to the line. The disadvantage of the detector is that it takes into account only a small set of possible directions [6].

This problem is solved in the Harris angle detector [7]. The autocorrelation of the luminance gradient along the horizontal $A(x,y)$, the vertical $B(x,y)$, and the mutual correlation $C(x,y)$ are calculated. And the measure of belonging to the line in the direction (u,v) is calculated as

$$E_{uv}(x,y) = A(x,y)u^2 + 2C(x,y)uv + B(x,y)\ v^2.$$

To normalize this function, it is rotated to the angle determined by the eigenvector (α,β):

$$F_{uv}(x,y) = \alpha^2 u^2 + \beta^2 v^2.$$

According to the properties of (α,β), if the point (x,y) defines the boundary of a straight line, then one value is large and the other is small. If a point defines a boundary with high curvature, both values are large. Then, the curvature degree is defined as

$$c(x,y) = \alpha\beta - k(\alpha + \beta)^2. \tag{2}$$

This method is used in many methods described below for its effective detecting of the lines and corners. However, it has rather high complexity and needs the floating point computations.

2.2 Methods based on the analysis of the brightness gradient

The brightness gradient is measured as the intensity of the derivative in some direction. Each feature point is characterized by a set of such gradients in some of its neighborhoods. Therefore, many methods are based on deriving these gradients as orientation histograms.

The most famous and effective method is the scale-invariant feature transform (SIFT) one [8]. The method is based on detecting the characteristic points based on the concept of the scale space, ie., the relative position of the feature points should be preserved in images with different scales and rotations. The scale space is most often determined by images, which are successively smoothed by a Gaussian filter. A set of difference images is formed using the formula (1), and then, a pyramid of images is formed at different scales σ.

The SIFT method includes four stages. At the extremum detection stage, a pyramid of difference-scale images is built. In these images, the local maxima/minima as centers of characteristic points are looked for. The pixel found must be larger/smaller than the neighboring pixels in this layer and in the adjacent layers of the pyramid as is shown in Fig.2.
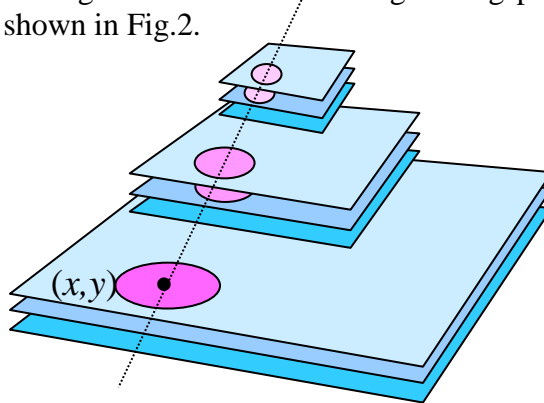

Fig. 2. The pyramid of images at different scales σ and the feature point in it

At the stage of the key point localization, the unstable feature points are filtered out, ie., the points of low contrast and those that are not angular or spots. The angles are detected by the Harris detector (2).

At the orientation search stage, the circular histogram of luminance gradients around the found point is constructed. The normalizing angle is determined, to which the local image around the center (x,y) is returned. It is possible to find 2 or 3 alternative angles, then 2 or 3 descriptors are calculated for this point.

At the last stage, the feature point descriptor is calculated and coded. A set of orientation histograms is created for the windows near the point (x,y) with 8 bins in each, resulting in a descriptor vector of 128 elements.

The SIFT method is the most reliable of many methods for determining the characteristic points, but also the most complex one. A large number of similar but much simpler methods have been developed, which are based on it and mentioned below.

The method of the speeded-up robust features (SURF), in contrast to SIFT, uses simple (coefficients only 0, ±1, ±2), but effective Gaussian filters that calculate edge detection (1) at different scales. But the

large-scale images are not decimated. The central pixel (x,y) of the characteristic point is found as the maximum of the determinant of the Hess matrix. The elements of this 2·2 matrix are the responses of various edge filters with the Haar wavelets. Due to the simplicity of calculating the filters, a high speed of finding the feature points is provided. Although the number of detected points is less than in the SIFT method [9].

The method of the speeded-up robust features (SURF) for the formation of these features uses Haar-like filters that are easy to calculate. But the method does not use the normalization, as in SIFT, which creates the incorrect descriptors [9].

The descriptive expressive robust features (DERF) method uses a model with the spatial distribution of pixels and their processing, which simulates the properties of the ganglion cells and the retina of primates [10].

The algorithm for constructing the SIFT descriptor, which is invariant to affine transformations (ASIFT), is an improvement of the original SIFT algorithm due to its complexity. To simplify the algorithm, it combines the best parts of the SURF and SIFT algorithms [11].

In the DAISY method, eight orientation maps $Go(x,y)$ are calculated, one for each quantized direction, where $Go(x,y)$ is equal to the value of the positive image gradient in the pixel $I(x,y)$ for the direction o. Then, the orientation maps are formed with different scales by filtering with Gaussian filters of different sizes as in the SIFT method. This improves speed and reliability, as well as insensitivity to affine transformation and brightness changes. In this case, the convolution with a large Gaussian core is simplified by building a pyramid of images. The feature point is described as a combination of normalized vectors with a large amplitude of all scales. [12].

The BOLD algorithm first finds the line segments, then generates geometric primitives as the angle between pairs of adjacent segments. The set of such primitives provide the consistency of the method in terms of rotation, mapping, scale, and noise with high informativeness [13].

2.3 Methods based on the contrast analysis

The methods based on the contrast analysis are applied to compare the intensities of pixels sampled at the different locations.

The features from the accelerated segment test (FAST) approach [14] are based on simple tests of 16 points around a pixel with x, y coordinates. The points are located along the circle, and if the brightness of the pixels in a particular pattern of points exceeds the brightness of $I(x,y)$, the pixel is denoted as a feature point. The method is quite fast. But the dependence on the threshold and not taking into account the orientation reduces its reliability. Machine learning is used to improve it.

The binary robust invariant scalable keypoints (BRISK) method uses the AGAST angle detector [15], which is an accelerated version of the FAST detector. The detector is implemented as a logical decision tree with a root corresponding to the central pixel $I(x,y)$ and branches leading to the vertices of adjacent pixels. For scale invariance, a pyramid of images with different scales obtained by the Gaussian blur is used. The symmetrical patterns with sampling points in concentric circles are used to describe the features. Then, up to 512 comparisons of pixels that are spaced from the center pixel are made. For each such comparison, the difference of the vectors leading to these pixels is calculated. These vectors are then averaged to find the dominant direction of the gradient, local gradients which represent some shape inside the analyzed window. Finally, this template is scaled and rotated. BRISK requires much more computation and more memory [16].

In the method of rapid determination of the feature points on the properties of the retina (Fast Retina Keypoint, FREAK) [17], an analogy with the work of the retina in the human visual system was used. The image is smoothed using a Gaussian filter. A window with near and far fields is created for the point under consideration. The binary descriptor is formed as a set of thresholds of differences in brightness of field pairs. Among the many options for comparing fields, the authors selected 512 of the most effective, which are used to build a descriptor. These pairs provide a highly structured pattern that mimics the jumping search of human eyes when looking at an object. The orientation of the feature point is calculated using local gradients among the selected pairs. As a result, the method decreases the computing time by two degrees of magnitude comparing to SIFT.

A similar approach, which is based on a set of binary robust independent elementary features (BRIEF), uses a brightness comparison detector and has a high speed [18].

2.4 Methods based on the analysis of the spatial frequency

At the two-dimensional transformation of space where the feature point is presented, it is possible to receive its orthogonal parameters. The Fourier descriptor is a set of two-dimensional Fourier transform coefficients of a shape region. The generalized Fourier descriptor (GFD) is obtained by pre-normalizing the image and converting it into polar coordinates [19]. GFD is invariant to rotation, scale, and mapping. This is because the Fourier coefficients are intrinsic functions that are invariant to the offset, which reflect the details of the feature point with different scales. But they have no information about its spatial position.

The wavelet transform is the result of decomposing an object into a pyramid of images with different scales and low-frequency and high-frequency components [20]. This pyramid has several necessary functions, such as simple computation, representation with different scales, constancy, spatial localization and stability, the ability to reproduce the image. The disadvantage is the sensitivity to shifting, to rotation, the lack of phase information. A survey on shape correspondence in computer vision, pattern recognition is represented in [21].

Thus, the descriptor using the spatial frequencies represents the figure of the feature point as a whole, however, without specifying its structural features. The required accuracy of the representation is achieved by choosing the number of conversion factors, which can be quite large.

2.5 Learning-based methods

The concept of the feature point is really not clear. To select the effective feature things in the image around the point under consideration, different learning-based methods are used effectively.

The supervised learning of low-dimensional feature descriptors is proposed in [22], which is applied for boosting to obtain a non-linear mapping of the input to a high-dimensional feature space. In the work [23] a learning local feature descriptor as a convex optimization problem by applying the sparsity is suggested. The proposed method can decrease the dimensionality as well as improve the descriptor effectiveness by applying the Mahalanobis norm regularization.

An evolutionary learning method is used to automatically generate the domain adaptive feature descriptor [24]. Here, the multi-objective genetic programming is applied to evolve the robust feature trees for the domain specific images with the fitness criterion taking into account the classification error rate and tree complexity.

The learning is usually used as a part of the feature extraction method. The oriented FAST and rotated BRIEF (ORB) detector detects the feature points based on the FAST algorithm. Then, a pyramid of images at different scales is created. A 9·9 window is analyzed around the feature point. Its center and orientation are calculated using the first-order moments, which are measures of some shape. The learning method is then used to increase the independence of the descriptor elements with respect to rotation. The execution of the ORB algorithm is an order of magnitude faster than the SIFT algorithm with almost the same recognition efficiency [25].

The learning-based methods don't need manually labeled features of a point and therefore, they are more flexible than the conventional handcrafted features described above. Moreover, the trained deep learning network can select the features which capture effectively the complex morphological patterns in the image [26]. But the deep learning can only provide a limited amount of data about the feature points which are not normed and are hard to be used in the feature point descriptors.

2.6 Convolutional neural networks

The main functions of computer vision are the detection, localization, and classification of characteristic points, which can be achieved in convolutional neural networks. The Overfeat method proposes to use a multi-scale approach with adjustable windows in the convolutional neural network (CNN) [27]. In this case, the search for the feature points, ie., low-level recognition operators are performed in the first convolutional layers.

The CNN models by themselves are used for feature extraction in tools like VGGNet [28], GoogleNet [29]. They are viewed as a set of non-linear functions which are composed of a number of layers including convolution, pooling, non-linearity [30].

The calculation of the feature point search operators allows the neural work to significantly improve the image before its processing in its middle and end layers. The research in [31] offers a method of

cleaning the image from noise to improve deep learning. A similar goal was achieved in [32], in which the CNN network was modified.

One of the advantages of the traditional image recognition methods is that the user can easily find evidence of how and why the methods work successfully. In CNN, this is a difficult task, because it is impossible to reasonably answer the question: what exactly has the system learned? This can be critical in some areas, such as financial, medicine image analysis, military and security systems. For reliable evidence, it is necessary to know the probable error of assessing the significance of any result. The models of CNN and machine learning are usually used as a black box, ie., they do not provide information about what exactly makes them come to their predictions [33].

2.7 Image feature detection methods comparison

The feature detector effectiveness can be tested and compared objectively only in the pattern recognition system when the proper descriptor is composed of the detector results. The descriptor is often evaluated by two tests: image matching and homography evaluation.

The image matching test is performed as follows [34]. The image collection has groups of images in each of them a reference image is and several images are distorted by scale, rotation, noise with increasing distortion (cut-off level). For all images in the collection, all descriptors are searched and calculated, which are entered into the database together with their coordinates in the image. Upon receipt of the image for recognition, the descriptors are calculated for it, and for them there are descriptors with the closest match in the database. If the distance between such descriptors is less than the threshold, then a match is recorded.

For image collections, a classification is made according to the number of matches and accuracy (average distance between descriptors). And for each algorithm of calculation of the descriptor the threshold of distance is chosen. The distance is calculated as the Euclidean distance for non-binary descriptors and as the Hamming distance for binary descriptors.

The test of the homography estimation is to calculate the coincidence of features between two images, followed by the calculation of homography. In fact, homography is the mapping of image points into a similar image, but after an affine transformation (rotation, change of perspective, etc.).

The nearest neighbor search algorithm is used to find a match to find the best match in the second image for each feature in the first image. When searching for the nearest neighbor, the Lowe rule [35] is often used, according to which the match is fixed only if the ratio of the distances from the nearest to the second nearest descriptor is less than 0.8.

So, the feature point detector is better if it provides the highest number of the feature points found in the testing image by the first method or the number of point pairs in two similar images by the second method.

The accuracy of matching is determined for the database of the image collection. The images in it are grouped by $k \leq 5$ or more images of one scene but at different scales, angles, and noise levels with the increasing inability to recognize, ie., with different ranks (cut-off). Then, the accuracy is calculated as

$$P@k = \textbf{Error!} .$$

The frequency of the image recognition (recall) is calculated as

$$R@k = \textbf{Error!} .$$

The accuracy-frequency curves are constructed for different feature detectors. The accuracy-frequency curve of recognition tends to decrease from left to right, ie., with increasing of the cut-off rank. The method for which the curve remains higher is considered the best [36].

Due to these curves, the methods SURF, SIFT, ORB, and BRIEF have approximately equal recognition effectiveness. But the method BRIEF has the minimum complexity among them which is in two orders of magnitude less than one of SURF and SIFT [36].

The mentioned above methods analysis shows the following conclusions.

To find the feature points the whole image is scanned sequentially by the feature point detector. In many cases, the traversing is performed in many layers of the pyramid of the images with different scales. To minimize the operation number, both the pyramid forming and the second-order derivative approximation are calculated using the Gauss filtering.

The most effective detectors are based on the analysis of the brightness gradient. They provide the scale-, mapping-, and rotation-invariant detecting. They use the Harris angle detector frequently, which is complex in calculations but it finds the corners and estimates their parameters effectively. Therefore, these detectors are usually performed in software using the floating point calculations intensively.

To minimize the feature extraction, and further pattern recognition complexity, the amount of data generated by the detector is minimized preserving the information about the feature point. The methods are mostly different in the approaches which select this information. In most detectors, the luminance gradient histograms are calculated taking the informative results. The evolutionary algorithms in the learning-based methods are used for this purpose to select the effective patterns around the point centrum.

The algorithm complexity minimization contradicts both the detector effectiveness and to the high dynamic range (HDR) of the analyzed image. Therefore, the frequent methods could not recognize the feature points in the heavy luminance conditions.

3.Method for the HDR feature point detection

The HDR image pixels are distinguished in large bit width which is much higher than the eight-bit pixels of the images used in the detectors mentioned above. Such an image is effectively used in the security, military systems because it perceives the information in the difficult luminance conditions. But with the HDR image processing, the problem of the dynamic range compression of the signal has to be solved to preserve the loss of the readability of the image both in illuminated and in darkened areas.

3.1 HDR adaptive filtering

The problem of the HDR image compression is solved using the theoretical Retinex model of the scene lighting. Then, the pixel brightness is a product

$$I(x, y) = L(x, y){\cdot}R(x, y),$$

where $L(x, y)$ is the illumination, and $R(x,y)$ is the reflected brightness of the object [37]. According to the Retinex approach, the image $I(x, y)$ is decomposed to the components $L(x, y)$, and $R(x, y)$. Then, the component $L(x, y)$ is processed with the dynamic range compression, and the contrast is improved in the component $R(x,y)$. These components are multiplied to obtain the resulting compressed image $I'(x,y)$. The brightness component is extracted using the function $F(I)$, which determines the illumination, so that

$$L(x,y) = F(I); \qquad R(x,y) = I(x,y) \ / \ F(I);$$
$$L'(x,y) = \square\square L(x,y)); \quad R'(x,y) = \square\square R(x,y)); \qquad (3)$$
$$I'(x,y) = L'(x,y){\cdot}R'(x,y).$$

Here, the compression function $\square(y)$ behaves as a logarithmical function, and the contrast improvement function $\square\square y)$ amplifies the signal near the level of 1.0. The illumination function $F(I)$ is a smart low-pass filter (LPF) that prevents the artifacts, appearing in the resulting image, preserving the edges in it. The bilateral filter is often used as the function $F(I)$ [38].

The bilateral filter is hard in computations. Therefore, the adaptive filter is proposed, which preserves the edges almost as well as the bilateral filter, but it is much simpler in the hardware implementation. Thanks to this, the processing of the HDR image is simplified and accelerated.

The adaptive filter structure, which preserves the edges of the image, is shown in Fig. 3. Such a filter consists of an image analyzer and an adjustable two-dimensional LPF. The idea of the Harris-Laplace detector [39] is used in the image analyzer. Such a detector is much easier to be implemented than detector (2) is and can be calculated using the fixed point computations. Its output signal is the eigenvector of the autocorrelation matrix of the neighborhood of the pixel under processing. There are five detectors $W_|$, W_{-}, $W_/$, $W_\backslash$, W_*, which are sensitive to the vertical, horizontal, inclined edges, or blobs in the image, respectively, and one LPF W_{LPF}, which estimates the local brightness of the image.

The output of such a detector is the signal in the logarithmic scale. The analyzer decision unit selects the maximum signal of the detectors and outputs it accompanying the detector number. The logarithm of the local brightness is subtracted from it providing the normalized signal $D(x, y)$ in the high dynamic

range. Fig. 4,a illustrates the results of the image analyzer for the image of the character **R**. The color of a pixel in it means the detector number.
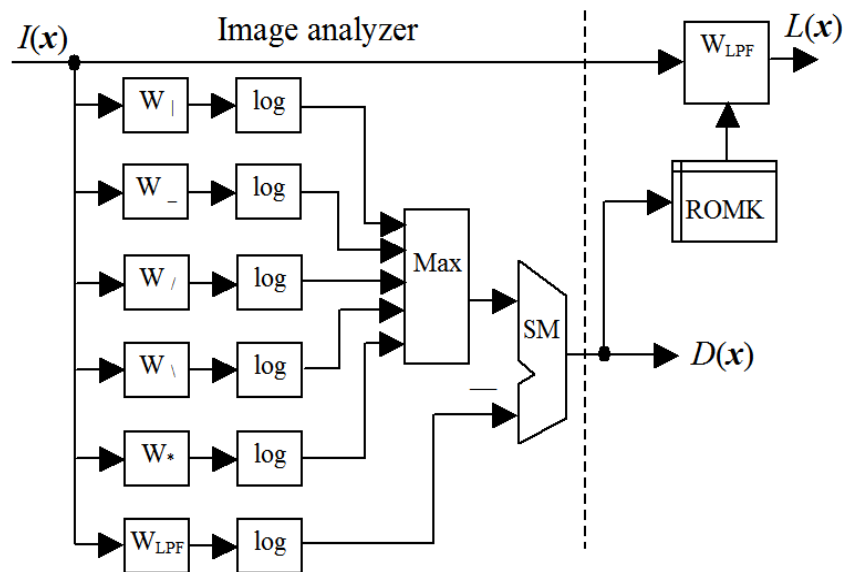

Fig. 3. Adaptive filter structure

The LPF filter contains a table ROMK of the filter kernels, which are distinguished depending on the local image type, i.e., if it is vertical, horizontal, inclined edge, and on its strength. The output signal of the image analyzer selects the proper kernel in the table for every pixel in the image. As a result, the image is filtered, providing the sharp edges in any luminance conditions. The resulting adaptive filter computes the illumination function $F(I)$ and is less complex than the bilateral filter. The work [40] describes this filter hardware implementation in detail.

3.2 Feature point extraction

To search for the feature points in the image, two additional computational steps are added to the adaptive HDR filter described above. The first one is the noise filtering step. Because of the logarithmical scale of the image $D(x, y)$, the usual linear filtering methods are not fitted. In this situation, the maximum homogeneity neighbor (MHN) filter is used [41]. The MHN filter is simplified here because the pixels $D(x, y)$ have the small bit width and are distinguished in five different colors, as in Fig. 4. The pixel $D(x, y)$ is smoothed in the filter if a stencil is found, which covers the pixels of the same color. Otherwise, this pixel gets the background color. The filtered image is shown in Fig. 4, b.

In the second step, the feature points are found. These points are blobs, corners in the edges, corners in the lines, intersections of the lines and edges. The examples of them are shown in Fig 5. These feature points are searched using the MHN method too. But the spatial stencils are adapted to the considered feature. For example, if the corner is considered, it has the proper angle and direction in the space which are coded by the colors. The selected feature point has the coordinates (x, y) of the respective feature filter with the maximum output magnitude in the considered image locality. For such behavior, this detector is rather similar to one in the FAST and BRISC methods.
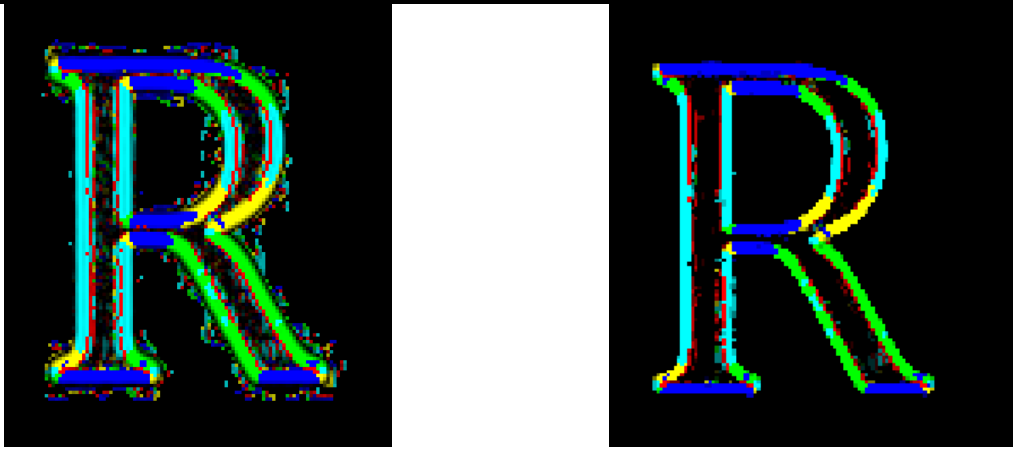
a)   b)
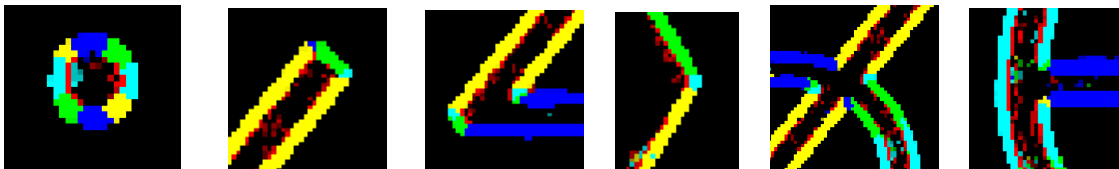Fig.4. The output of the adaptive filter (a) and after the noise filtering (b)

Fig.5. Examples of the feature point localizations

To do the scalable feature point searching, the derived logarithmic image has to be decimated many times. Such a decimation is performed using the MHN method except that it is performed to each even pixel. By this process, the feature point coordinates coincide with each other in the images with different scales.

The next step of the feature point descriptor forming consists in calculating the distribution diagrams as it is shown in [42].

3.3 Experimental results and future work

At present, the artificial intelligence engine is under development, which is implemented in FPGA of medium or large size. The input data is an HDR video camera data flow. To process the image with the HDR compression, the video camera was designed on the base of the Lattice HDR-60 board, which utilizes ECP3-70 FPGA. As a video sensor, the Aptina MT9M024 chip is used, which produces the 720·1280 HDR image stream at a rate of 60 frames per second with a dynamic range of 120 dB.

The colored image channel is split into the brightness and color channels. The brightness signal with the 20-bit pixels is compressed to 8-bit pixels due to the equations (3). Then, the colored image is restored using the color channel. The image analyzer filters (Fig. 3) are implemented as the pipelined multiplier-free adder networks. These networks provide both high-speed computations for the multi-bit data and small hardware volume comparing to the network of the hardware multipliers. The functions $\square(y)$ and $\square\square y)$ in (3) are implemented using the piecewise linear interpolation.

Fig. 6 illustrates the scene image with the heavy luminance conditions which is compressed by the video camera and the respective image for the feature extracton. It shows that the feature points can be selected robustly in such strong conditions.

Fig.6. HDR compressed image and results of its processing by the analyzer

This system was probed to analyze the blood image which was taken from the repository of the information on the biochemicals and cells in blood and body fluids [43]. Fig. 7 illustrates the results of the use of the developed system. Its analysis shows that the system can reliably select the cells and represent its features, which helps to recognize and make classification of the unnormal cells. Such cells have a set of feature points representing the angles which form a ring.
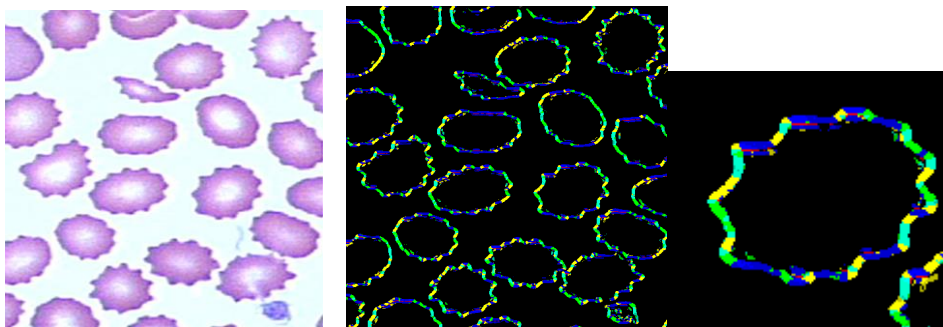

Fig.7. Initial image and results of its processing by the analyzer

The future work consists of development of the pattern recognition system. The pattern recognition process has two distinguished stages. In the first stage, the image is transformed into a pyramid of feature frames. For this process, an adaptive filter is used as in Fig. 3, which both compresses the pixel dynamic range and preserves the edges of the image. But for the noise filtering, and for the frame pyramid forming the MHN filter is used.

In the second stage, the feature points are found. These points are blobs, corners in the edges, corners in the lines, intersections of the lines, and edges. These feature points are searched using the MHN method as well. Then, the feature descriptors are formed as like as in the SIFT method.

At the period of learning, the computed feature descriptors are stored in the database, and at the working period, the found feature descriptors are compared to ones in the database.

To design the pipelined datapath in a short time, a framework is designed which helps to compile the data path algorithm representation into the FPGA hardware described in VHDL. The algorithm can be represented in the TensorFlow language as well.

4.Conclusions

A set of different methods of feature point extraction is analyzed. This analysis shows that the most effective detectors are based on the analysis of the brightness gradient. They usually use the Harris angle detector, which is complex in calculations. The algorithm complexity minimization contradicts both the

detector effectiveness and the high dynamic range of the analyzed image. Therefore, the high-peed methods could not recognize the feature points in the heavy luminance conditions.

The modification of the HDR image compression algorithm based on the Retinex method is proposed. It contains an adaptive filter, which preserves the image edges. The filter is based on a set of feature detectors performing the Harris-Laplace transform which is much simpler than the Harris angle detector. Such a filter reduces the compression complexity. A prototype of the HDR video camera is designed which provides sharp images both in illuminated and in darkened areas of the scene. The resulting signal of this camera simplifies the design of the pattern recognition system. The next step will be the design of the artificial intelligence engine, which is implemented in FPGA of medium or large size.

References

1. Nixon M. S., Aguado A. S. Feature Extraction and Image Processing for Computer Vision. 4-th ed. London: Academic Press, 2020. 650 p.
2. Tuytelaars T., Mikolajczyk K. Local Invariant Feature Detectors: A Survey. Foundations and Trends in Computer Graphics and Vision. Vol. 3, No 3. Hanover, MA, United States: Now Publishers Inc. 2007. P. 177–280.
3. Krig S. Interest Point Detector and Feature Descriptor Survey. Computer Vision Metrics. Springer, 2016. P. 187–246.
4. Kovesi, P. Image Features from Phase Congruency, Videre: Journal of Computer Vision Research, Vol. 1, No 3, MIT Press Journals, 1999. P. 1-27.
5. Kass, M., Witkin, A., Terzopoulos, D. Snakes: Active Contour Models International Journal of Computer Vision, Vol. 1, No 4, KIuwer Academic Publishers 1988. P 321-331.
6. Moravec H. Towards Automatic Visual Obstacle Avoidance Proceedings of the 5th International Joint Conference on Artificial Intelligence, Cambridge, 22-25 August, 1977. P. 584.
7. Harris, C., Stephens, M. A Combined Corner and Edge Detector. Proceedings of Fourth Alvey Vision Conference, Manchester. 1988. P. 147-151.
8. Lowe D. G. Distinctive Image Features from Scale-Invariant Key Points. Int. J. of Computer Vision, Vol. 60, No 2, Springer, 2004, pp. 91-110.
9. Bay H., Ess A., Tuytelaars T., Gool L. V. Speeded-up robust features (SURF), Computer Vision and Image Understanding, Vol. 110, No. 3, New York: Elsevier Science, 2008. P. 346-359.
10. Weng D. W., Wang Y. H., Gong M. M., Tao D. C., Wei H., Huang D. DERF: Distinctive efficient robust features from the biological modeling of the P ganglion cells. IEEE Transactions on Image Processing, Vol. 24, No. 8. IEEE 2015. P. 2287-2302.
11. Morel J. M., Yu G. ASIFT: a new framework for fully affine invariant image comparison. SIAM Journal on Imaging Sciences. Vol. 2, No. 2, 2009. P. 438–469.
12. Tola, E., Lepetit, V., Fua, P. DAISY: An efficient dense descriptor applied to wide baseline stereo. IEEE Transactions on Image Processing, Vol. 32, No. 2. IEEE 2010. P. 815–830.
13. Tombari, F., Franchi, A., Di Stefano L. BOLD features to detect texture-less objects. IEEE International Conference on Computer Vision : conf. proc. Dec. 1-8 2013. Sydney, Australia. IEEE 2013. P. 1265–1272.
14. Rosten, E., and Drummond, T. Machine Learning for High-Speed Corner Detection, Proceedings of ECCV: conf. proc. 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Springer, 2006. P. 430-443.
15. Mair, E., Hager, G.D., Burschka, D., Suppa, M., Hirzinger, G. Adaptive and generic corner detection based on the accelerated segment test. Proceedings of the 11th European Conference on Computer Vision: Part II: Heraklion, Crete, Greece September 5 - 11, 2010. Springler, 2010. P. 183–196.
16. Leutenegger, S., Chli, M., Siegwart, R.Y. BRISK: binary robust invariant scalable keypoints. IEEE International Conference on Computer Vision: conf. proc. Barcelona Nov. 6-13 2011. IEEE 2011. P. 2548–2555
17. Alahi, A., Ortiz, R., and Vandergheynst, P. Freak: Fast Retina Keypoint. IEEE Conf. on Computer Vision and Pattern Recognition: conf. proc. Providence, RI, USA, June 16-21, 2012. IEEE Computer Society 2012. P. 510-517.

18. Calonder M., Lepetit V., Özuysal M., Trzcinski T., Strecha C., Fua P. BRIEF: Computing a local binary descriptor very fast. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 34, No. 7, 2012. P. 1281-1298.

19. Zhang D., Lu G. Generic Fourier descriptor for shape-based image retrieval. IEEE International Conference on Multimedia and Expo. Vol.1: conf. proc. 26-29 Aug. 2002. IEEE 2002. P. 425-428.

20. Nabout A. A., Tibken B. Wavelet Descriptors for Object Recognition using Mexican Hat Function. IFAC Proceedings Volumes, Vol. 38, No. 1: conf. proc. Prague, July, 3-8 2005. Elsevier, 2005. P. 1107-1112.

21. Van Kaick O., Zhang H., Hamarneh G., Cohen-Or D. A survey on shape correspondence. Computer Graphics Forum, Vol. 30, No. 6, 2011. P. 1681-1707.

22. Trzcinski T., Christoudias M., Lepetit V. Learning image descriptors with boosting. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 37, No. 3. IEEE 2015. P. 597-610.

23. Simonyan K., Vedaldi A., Zisserman A. Learning local feature descriptors using convex optimisation. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 36, No. 8, IEEE 2014. P. 1573-1585.

24. Shao L., Liu L., Li X. L. Feature learning for image classification via multiobjective genetic programming. IEEE Transactions on Neural Networks and Learning Systems. Vol. 25. No. 7, IEEE 2014. P. 1359-1371.

25. Rublee E., Rabaud V., Konolige K., Bradski G. ORB: An efficient alternative to SIFT or SURF. IEEE International Conference on Computer Vision: conf. proc. Barcelona Nov. 6-13 2011. IEEE 2011. P. 2564–2571.

26. Wu G. R., Kim M. J., Wang Q., Munsell B. C., Shen D. G. S-calable high-performance image registration framework by unsupervised deep feature representations learning. IEEE Transactions on Biomedical Engineering. Vol. 63, No. 7, IEEE 2016. P. 1505-1516.

27. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y. Overfeat: Integrated Recognition, Localization and Detection Using Convolutional Networks, arXiv preprint arXiv:1312.6229, 2013.

28. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. Proceedings of the International Conference on Learning Representations: Conf. proc. San Diego, CA, USA, May 7-9, 2015. arXiv:1409.1556 2015.

29. Szegedy C., Liu W., Jia Y. Q., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going deeper with convolutions. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition: Boston, MA, USA, June 7-12, 2015. IEEE 2015. P. 1-9.

30. Zheng L., Yang Y., Tian Q. SIFT meets CNN: A decade survey of instance retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 40. No. 5, IEEE 2018. P. 1224-1244.

31. Gul, M. S. K., Gunturk, B. K. Spatial and Angular Resolution Enhancement of Light Fields Using Convolutional Neural Networks. IEEE Transactions on Image Processing Vol. 27, No. 5. IEEE 2018. P. 2146-2159.

32. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. IEEE Transactions on Image Processing. Vol. 26. No. 7, IEEE 2017. P. 3142-3155.

33. Zhang, Q. S., Zhu, S. C. Visual Interpretability for Deep Learning: a Survey. Frontiers of Information Technology & Electronic Engineering, Vol. 19, No 1, Springer, 2018. P.27-39.

34. Muja, M., Lowe, D.G. Scalable nearest neighbour algorithms for high dimensional data. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 36, No 11, IEEE 2014. P. 2227–2240.

35. Lowe, D.G. Distinctive image features from scale invariant keypoints. Int. J. of Computer Vision. Vol. 60, No. 2. Springer, 2004. P. 91–110.

36. Khan, N., McCane, Mills B. S. Better than SIFT? Machine Vision and Applications. Vol. 26, 2015, P.819–836.

37. McCann J. J., Land E. H. Lightness and retinex theory. Journal of the Optical Society of America, Vol. 61, No. 1, 1971. P. 1 – 11.

38. Paris S., Kornprobst P., Tumblin J., Durand F. Bilateral filtering: theory and applications. Foundations and Trends in Computer Graphics and Vision, Vol. 4, No 1, 2008. P. 1–73.

39. Hassaballah M., Abdelmgeid A. A., Alshazly H. A. Image features detection, description and matching. In: Image Feature Detectors and Descriptors. Foundations and Applications. A. I. Awad, M. Hassaballah, Eds. Springer, 2016, P. 11 − 46.

40. Sergiyenko A., Serhiienko P., Zorin Ju. High Dynamic Range Video Camera with Elements of the Pattern Recognition, IEEE 38th International Conference on Electronics and Nanotechnology, ELNANO'18, Kyiv, April 2018. IEEE 2018. P. 435-438.

41. Nagao M., Matsuyama T. Edge preserving smoothing. Computer Graphics and Image Processing. Vol. 9, No 4, 1979. P. 394–407.

42. Sergiyenko A., Serhiienko P., Orlova M., Molchanov O. System of Feature Extraction for Video Pattern Recognition on FPGA, 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON), 2019. P. 1175-1178.

43. Medical Laboratories Portal. http://www.medical-labs.net/wp-content/uploads/2014/01/Crenated-Cells.jpg