Using descriptors for image recognition

K A Bekmuratov¹, O A Mamaraufov¹, M O Shamiyev¹

¹Samarkand branch of Tashkent University of Information Technologies, Shokhrukha Mirzo st., 41, Samarkand, Uzbekistan.

E-mail: nurshod86@mail.ru

Abstract. Methods for finding singular points and forming their descriptors are considered. The aim is to study the existing search methods and determine the feature point descriptors to select the best match between feature point detectors and their descriptors for different types of images. In this paper, we carry out a comparative analysis of the PCA-SIFT, ORB, BRISK, AKAZE methods, which detect singular points and describe their descriptors in the image. An algorithm has been developed that, based on the work of these methods, groups photographs according to the degree of similarity. A comparative analysis of the methods is carried out on several collections of photographs of different quality at different compression rates and at different values of the descriptor matching parameter.

1. Introduction

There are many methods for recognizing objects in an image. The choice of specific methods is due to the characteristics of the object that you want to recognize. It often happens that the recognition problem is posed in an informal way - the properties of the sought object are set without strict mathematical parameters. To solve such a problem, it is necessary to formulate the properties of the desired object and create a stable method for detecting objects that correspond to the specified parameters. To solve the problem, it is necessary to find, generalize and formulate empirical observations in mathematical terms. That is, formalize the parameters of the desired object. The main difficulty is that it is almost impossible to describe all the properties, and these properties may not correspond to all objects. Therefore, in the process of mathematical formalization, simplifications are allowed, which, as a result, reduce the quality of the algorithm and lower the accuracy. As a result, we can say that when solving the recognition problem, it is necessary to find the optimal ratio of the computational complexity and the desired accuracy [1].

The input data of the descriptor is an image and a set of special points selected on a given image. The output of the descriptor is the set of feature vectors for the original set of feature points. It should be noted that some descriptors simultaneously solve two problems - the search for singular points and the construction of descriptors for these points [2].

An important stage in solving a wide class of problems of processing, analysis, understanding and search of digital images is the choice of features, often referred to in the modern literature as descriptors [3, 4].

There are various classifications of features, reflecting the specifics of approaches to their obtaining. For example, in monographs [6, 7] features are subdivided into geometric, topological, probabilistic and spectral.

In papers [5, 8], local descriptors are subdivided into probabilistic, spectral, spectral-frequency and...
differential. Regardless of the classification used, there are signs that can combine properties characteristic of different categories.

Such descriptors include a model-oriented descriptor proposed and developed by the author [9 - 10]. Its difference from existing works is the presence of a priori given (or predetermined from the training set of images) probability distribution of the gradient field characterizing the model of the analyzed image and / or the problem being solved.

The descriptor attribute is the normalized value of the probability density with the argument in the form of a specific gradient field. It turns out to be dependent both on the implementation (a specific image) and on the model (probability distribution), which made it possible to characterize this descriptor as model-oriented. Characterizing the proposed approach, it should be noted its features and differences from other known solutions. The gradient is widely used in the tasks of recognition and image search. The most well-known gradient-based descriptors are HOG and SIFT descriptors [6].

However, these and most other known descriptors and features do not use the image gradient as a vector field (the interpretation used below is a complex image / signal) fully. Namely, the phase or amplitude component of the gradient in the known solutions is completely or partially ignored. The most striking examples of this approach are works [7]. In particular, if the phase component of the gradient is completely ignored, then in works, on the contrary [8], the amplitude is ignored. In general, the article is devoted to methods and algorithms for processing a special subclass of images, characterized only by the direction of change in the brightness function, the so-called field of directions [12].

2. Statement and method of solving the problem

Comparison of descriptors among themselves will consist in comparing bits in strings. In particular, to compare descriptors, early methods used measures such as the sum of differences modulo $\sum|f_a - f_b|$ or the sum of the squares of the differences $\sum(f_a - f_b)^2$. In the case of binary descriptors, it is possible to use the Hamming measure $\sum X OR(f_a, f_b)$, which is calculated quickly - with one command [11].

Auxiliary descriptors. More generally, descriptors are computed in several sliding steps. Let the descriptors $D$ obtained at the intermediate stage of sliding be transformed into the matrix $D'$ by some operator $F$ [13]:

$$\forall D \in R^{(m+n)\times h} \exists D' \in R^{(m+n)\times h'}, D' = F(D).$$

The rows of the matrix $D'$ are composed to form matrices $D''$, along which the independent sliding is performed:

$$D'' \in R^{m \times h}, D'' = R^{-1}_{m\times h}(D'(\cdot, i)), i = \{1, \ldots, h'\}.$$

When all the sliding stages are passed, these descriptor matrices are combined into one final matrix.

Fast descriptors of special points include the following: PCA-SIFT (Principal Component Analysis, PCA) - L Juan, Oubong Gwun, ORB (Oriented FAST and Rotated BRIEF) - E. Rublee, V. Rabaud, K. Konolige, G. Bradski, BRISK (Binary Robust Invariant Scalable Keypoints) - S. Leutenegger, M. Chli, RY Siegwart, AKAZE (Accelerated-KAZE) - P. Alcantarilla, A. Bartoli, A. Davison.

Descriptor PCA-SIFT. The PCA-SIFT descriptor [11] is essentially a modification of SIFT. At the initial stage, the values of the magnitude and orientation of the gradient are calculated in the same way. For each singular point only, a pixel-sized neighborhood with a center at a point that is singular is considered. In fact, a gradient map is built along the vertical and horizontal directions. As a consequence, a vector containing elements is obtained. Next, the SIFT descriptor is built according to the scheme described in the previous section. For the resulting set of SIFT descriptors, vectors are dimensioned down to elements through Principal Component Analysis (PCA).

ORB descriptor. The ORB method (Oriented FAST and Rotated BRIEF) [14] is designed to eliminate the above disadvantages of BRIEF.

In the ORB method, the coordinates of the center of gravity $C$ are used to calculate the orientation of the angle, calculated through the moments of the image $m_{pq}$.
Then the orientation of the angle will be given by a vector, the beginning of which will be at the center point and the end – at the center of gravity. And the angle will be equal to:

$$\theta = \arctg \frac{m_{01}}{m_{10}}$$

These ideas are embodied in the "steered" BRIEF method. For a set of binary tests of size n, with coordinates \((x_i, y_i)\), a 2x2 matrix \(S\) is constructed:

\[
S = \begin{pmatrix}
x_1 - x_n \\ y_1 - y_n
\end{pmatrix}
\]

Using the calculated angle \(\theta\), the rotation matrix \(R_\theta\) is constructed, and then the matrix \(S_\theta\), taking into account the rotation, is equal to \(S_\theta = R_\theta \ S\). Then the angle is discretized with an increment of \(2\pi / 30\), i.e. at \(12^0\), and the descriptor is searched and matched with \(S_\theta\).

BRISK descriptor. In order to achieve rotation invariance, in the BRISK (Binary Robust Invariant Scalable Keypoints) descriptor [15], points are selected in accordance with the pattern in Fig. 1.

To calculate the orientation of the key point, the local gradient is calculated between a pair of points \((p_i, p_j)\) among \(N (N - 1) / 2\) points of the area \(p\):

\[
g(p_i, p_j) = \frac{I(p_j, I) - I(p_i, I)}{\|p_i - p_j\|^2},
\]

where \(I(p_i, \sigma_i)\) – Gaussian smoothed intensities of points with standard deviation \(\sigma_i\). On the entire set of pairs of points A, "short pairs" - S and "long pairs" - L are defined. For long ones \(\|p_i - p_j\| < \delta_{\text{max}}\), while short ones \(\|p_i - p_j\| > \delta_{\text{min}}\), where \(\delta_{\text{max}}, \delta_{\text{min}}\) are threshold values.

Long L pairs are used to calculate the direction of the singular point. To calculate the orientation, the sum of all "long gradients" is determined:

\[
g = \frac{g_x}{g_y} \sum_{(p_i, p_j) \in L} g(p_i, p_j)
\]

and \(\alpha = \arctg (g_x / g_y)\) is calculated. The binary descriptor itself is calculated as in BRIEF, but only for "short" pairs. The mapping of descriptors is similar to the previous methods.

Descriptor AKAZE. This method [15] describes the detection and description of singular points in nonlinear scale spaces. The idea behind this approach is to create a series of intermediate images at different scales (multiscale space) by applying various kinds of filtering to the original image.

Singular points are detected by calculating the determinant of the Hessian matrix for each filtered component of nonlinear scale representations \(L_i\) of the original image:

\[
L_{ij} = \sigma^2 (L_{xx} L_{yy} - L_{xy}^2),
\]

where \((L_{xx}, L_{yy})\) is the second horizontal and vertical derivative.
To calculate the descriptor, a version of the M-SURF descriptor, modified for nonlinear multiscale space, is used. The first derivative $L_x$ and $L_y$ in the rectangular area $24 \sigma \times 24 \sigma$ are calculated for the detected singular point at the scale $\sigma_i$. Based on the obtained general direction of the vicinity of the singular point, each area from the considered rectangular area is rotated according to the obtained general direction. In addition, derivatives are calculated in accordance with this direction. Finally, the resulting descriptor is normalized to a 64-bit vector of relative units to achieve invariance to image contrast.

3. Algorithm for solving the problem

The developed algorithm uses specially prepared matrices of binary patterns of symbols found in images. The use of such standards can improve performance and reduce the effect of noise.

The image matching algorithm using descriptor structures consists of the following steps.

Step 1. Matching using the PCA-SIFT, ORB, BRISK and AKAZE descriptors and the RANSAC method.

Step 2. Finding the first five correct matches. By sequential enumeration, the groups are analyzed, including five matches. If for each match in a group the element in the matrix is the maximum in the row and in the column, then all matches in this group are accepted as correct. The group for which this condition is met is considered basic and is used in step 3 to check the rest of the matches. After finding one base group, step 2 ends.

Step 3. Checking each match found at step 1 and not included in the base group obtained at step 2. Graphs are constructed based on any four correct matches obtained at step 2 and one checked match. If for each match in the group the element in the matrix $B$ is the maximum in the row and in the column, then the checked match is assumed to be true.

4. Experimental results

To implement the specified algorithm for image recognition processing, software was written with a graphical interface in the C++ language.

To develop the program, the Qt cross-platform library was used [17]. A number of stages of this algorithm were implemented using the library of computer vision algorithms, image processing and general-purpose numerical algorithms with open source OpenCV [16]. The graphical interface settings, a priori information, and auxiliary data of the image recognition process are stored in XML files, and are read by the program when it is loaded. The correct results of recognition of symbolic information contained in all images to be recognized familiarity are recorded in the SQL database in real time. Graphs of motion parameters contour analysis are built and displayed using a set of Qt widgets and Qwt helper classes [18] in the process of software operation. The program implements multithreaded processing of all familiarity, which allows you to recognize a large amount of symbolic information in parallel.

The algorithm was tested on a computer with an Intel (R) Core (TM) i5-2410M 2.30 GHz processor, 800 MHz DDR3 RAM, and a 64-bit Windows 10 operating system. The results of experimental studies have shown that the recognition time for one familiarity in an image containing five characters was 10 ms. Recognition of 300 test images with symbolic information was carried out. The accuracy of character information recognition in images was 85%.
Table 1. Analytical comparison of recognition results

| Descriptor type | Sliding time (ms) | Multiscale slip time (ms) | Recognition quality, % | Required memory, MB | Matching time, sec |
|-----------------|------------------|---------------------------|------------------------|---------------------|-------------------|
| PCA-SIFT        | 23.7             | 66.2                      | 63                     | 0.98                | 0.5-1             |
| ORB             | 44.2             | 119.0                     | 74                     | 0.05                | 0.5-1             |
| BRISK           | 11.3             | 27.5                      | 68                     | 6.1                 | 0.4-0.8           |
| AKAZE           | 55.1             | 147.3                     | 88                     | 18.3                | 0.8-1.2           |

5. Conclusions
In the work, the basic principles of object recognition in the image, the recognition methods used and the features of objects, according to which the recognition systems operate, were considered.

PCA-SIFT descriptors are not without their drawbacks. Not all received points and their descriptors will meet the requirements. Naturally, this will affect the further solution of the problem of image matching. In some cases, a solution may not be found, even if it exists.

The ORB method has the best speed in calculating special points and calculating their descriptors, which makes it possible to use it in tasks where image processing is required in real time. One of these tasks is tracking a moving object. But the high speed of work does not affect the accuracy of image matching for the better. The presence of digital noise or blurred images further degrades the results of the program.

BRISK differs from other methods in that it determines the largest number of singular points, but, unfortunately, digital noise also gets into them, while filtering the resulting false connections takes a significant amount of time, although the final accuracy is high. At the same time, a small number of features were identified on blurred images, which resulted in unsatisfactory performance of the classifier due to lack of data. The highest quality photographs show the best grouping accuracy.

The AKAZE method does not have the same speed as ORB and does not have as many singular points as BRISK. But at the same time, due to the peculiarities of its structure, it is finding singular points on a nonlinear multiscale pyramid and describing descriptors by three parameters, instead of one, as in ORB and BRISK. We get high accuracy when comparing images and their further distribution into groups, both with high and poor image quality.

References
[1] Fazilov Sh Kh, Mirzaev N M, Nurmukhamedov T R, Ibragimova K A, Ibragimova S N 2019 *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* **9**(1) 2278-3075
[2] Kamalov M M, Mirzaev N M, Radjabov S S 2009 "Knowledge - Ontology - Theories" (ZONT-09): Materials of the All-Russian conference with international participation 1 35-41
[3] Fazilov Sh X, Mahkamov A A, Jumayev T S 2018 *Informatics: problems, methodology, technologies: Materials of the XVII international scientific and methodological conference* 2 3-7
[4] Nishanov A H, Mamaraufov O A 2016 *Scientific journal Problems of Computational and Applied Mathematics. Tashkent: Centre for the development of software and hardware-program complex* 3 (5) 72-79
[5] Mahkamov A A, Bozarov A A, Jumaev T S 2014 *Eighth world conference on intelligent systems for industrial automation* 454-457
[6] Mikolajczyk K, Mikolajczyk C Schmid 2005 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(10) 1615-1630
[7] Gashnikov M V, Glumov N I, Ilyasova N Yu, Myasnikov V V, Popov S B, Sergeev V V, Soifer V A, Khramov A G, Chernov A V, Chernov V M, Chicheva M A, Fursov V A 2003 *Methods of computer image processing*; ed. V A Soyfer
[8] Duda R O, Hart P E 1973 (New York: Wiley) 512
[9] Myasnikov V V 2012 Computer Optics 36(4) 596-604
[10] Myasnikov V V 2012 Computer Optics 36(4) 429-438
[11] Kuznetsov A V, Myasnikov V V 2016 Pattern Recognition and Image Analysis 26(3) 593-596
[12] Shamiyev M O 2020 International Journal of Advanced Research in Science, Engineering and Technology 7(10) 2350-0328
[13] Rublee E V Rabaud, K Konolige, G Bradski 2011 IEEE International Conference on Computer Vision. 58(11)
[14] Shamiyev M O 2020 International Journal of Psychosocial Rehabilitation. 24(9)
[15] Alcantarilla P, Bartoli A, Davison 2012 KAZE Features European Conference on Computer Vision. 4 214-227
[16] OpenCV API Reference. Available at: http://docs.opencv.org/modules/refman.html.
[17] Qt Project Documentation. Available at: http://qt-project.org/doc/
[18] Qwt User’s Guide. Available at: http://qwt.sourceforge.net/