

Received February 2, 2022, accepted March 7, 2022, date of publication March 14, 2022, date of current version April 15, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3159650

Study of Keypoints Detectors and Descriptors Performance on X-Ray Images Compared to the Visible Light Spectrum Images

MIKHAIL CHEKANOV¹, OLEG SHIPITKO^{ID}¹, AND NATALIA SKORYUKINA^{2,3}

¹Institute for Information Transmission Problems (IITP), Russian Academy of Sciences (RAS), 127051 Moscow, Russia

²Federal Research Center Computer Science and Control, Russian Academy of Sciences, 119333 Moscow, Russia

³Smart Engines Service, 121205 Moscow, Russia

Corresponding author: Mikhail Chekanov (chekanov@visillect.com)

This work was supported in part by the Russian Foundation for Basic Research under Project 18-29-26036 and Project 18-29-26037.

ABSTRACT In this work, we study the performance of wide-used keypoints detection and description algorithms: Scale-Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), Oriented FAST and Rotated BRIEF (ORB), Binary Robust Invariant Scalable Keypoints (BRISK), Accelerated KAZE (AKAZE), which were originally developed for images taken in visible light but widely applied in the fields where images are taken in a different spectrum. We compare the quality of algorithms and their robustness to various image transformations. The algorithms' performance is tested on two image sets in the different spectra: digital X-Ray images and images taken in the visible spectrum. Each dataset captures complex scenes with many objects and partial occlusions. Geometrical transformations (rotation, shearing, scaling), linear color transformations, Gaussian blur are applied to the images. Then the detection and description algorithms are tested on the original and transformed images. The repeatability and number of corresponding points are calculated to assess detection algorithms. The ratio of correctly matched descriptors together with the ratio of the distances between the query descriptor, the nearest descriptor, and the second matched descriptor is computed to evaluate descriptors' quality. The algorithms showed different behavior on different spectra. SURF demonstrated to be the best X-ray keypoint detector and for the visible spectrum, it shares first place with AKAZE detector. SIFT is the best descriptor in both spectra. The strong and weak points of each algorithm are discussed in the paper.

INDEX TERMS Keypoints, repeatability, robustness, digital X-ray images, computed tomography, CT, detectors, descriptors.

I. INTRODUCTION

In computer vision systems a keypoint is a projection of a point of a three-dimensional scene onto the image plane, which meets the following requirements [1]:

- 1) *distinctness* - a specific point must clearly stand out from the background and be distinguishable (unique) in its vicinity;
- 2) *invariance* - the definition of a singular point must be resistant to affine transformations;
- 3) *stability* - The definition of a special point must be robust to noise and errors;

- 4) *uniqueness* - In addition to being locally distinct, a feature must be globally unique to improve the discernibility of repeating patterns;
- 5) *interpretability* - Feature points should be defined so that they can be used to analyze correspondences and identify interpretable information from the image.

Keypoints are used to solve problems such as the body pose estimation [2], object classification [3], three-dimensional scenes reconstruction [4], visual odometry and navigation [5], image registration, radiometric correction and many others. In addition to the task of keypoints detection, there is the task of keypoint description which is performed based on the detected point's vicinity to compare it with keypoints in other images. The algorithm that solves the first problem is called a *detector*, and the second one – a *descriptor*.

The associate editor coordinating the review of this manuscript and approving it for publication was Hengyong Yu ^{ID}.

Due to the practical importance of keypoints, a variety of detection and description algorithms have been proposed [6]–[9]. These algorithms are based on classical image processing techniques. In addition, detectors and descriptors based on neural networks are actively developed [10]. In this paper, only classical algorithms for the detection and description of keypoints are considered. They all use different approaches to define informative image areas. For this reason, these algorithms show different behavior and quality depending on the image structure. Many works are devoted to the keypoints comparison. Among them, the works of Song and Klette [11] and Mikolajczyk *et al.* [12], which compare different detection algorithms on geometrically and photometrically transformed images. In the work of Hu *et al.* [13] the description algorithms are compared on a set of images obtained from different viewpoints. All mentioned studies investigate the quality of detectors and descriptors only on images in the visible spectrum. At the same time, keypoints are applied in many problems where images have a different nature [14]–[16] and, in particular, images obtained in the X-ray spectrum. Thus, in [17], keypoints are used to detect vertebrae on X-ray images. In the article [18], keypoints are used to restore the trajectory of the circular motion of the tomograph from the tomographic projections made by it. The work [19] considers the problem of prohibited items detection on X-ray images of baggage and uses keypoints to detect an object.

When X-rays pass through the material, they are attenuated. Digital X-ray images are generated by recording this attenuation along the beam corresponding to each pixel of the detector matrix. Thus, the principle of formation of X-ray images differs significantly from the formation of images in the optical spectrum, which records information about the intensity of electromagnetic signal reflected from an object. Unlike images made in the visible spectrum, X-ray images are characterized by the translucency of the recorded object, the absence of textures, and low image sharpness. Therefore the applicability and robustness of detectors and descriptors designed to work with images of other properties on digital X-ray images need to be studied. The authors are not aware of the works in which this issue was investigated. The purpose of this work is to study the difference in performance and robustness of wide-used keypoints detection and description algorithms for digital X-ray and the visible spectra images and for various photometric.

For the study, we use two datasets: in the visible (HPatches [20]) and X-ray spectra (GDXray [21]). Various affine transformations (rotation, scaling, bevel), color correction (changes in brightness and contrast), and Gaussian blur were applied to the images. On pairs of original and transformed images, we tested five detection and description algorithms: SIFT, SURF, ORB, BRISK, AKAZE, and compared quality metrics separately for X-ray and visible spectrum images.

The main contributions of this paper are:

- The article for the first time carefully studies the effect which different nature of X-ray images compared to the visible spectra images takes on the quality and robustness of keypoints detection algorithms.
- The robustness of the algorithms to affine transformations, as well as to changes in lighting conditions, was investigated in both spectra.
- Best detector and descriptor algorithms were chosen: AKAZE detector, SIFT descriptor for the visible spectrum, SURF detector, SIFT descriptor for the digital X-ray images.

The findings of the article might be used to choose the proper algorithm for applications in which the X-ray images are analyzed.

II. COMPARED ALGORITHMS

We start with the description of the keypoints detection and description algorithms compared in this study.

SIFT (Scale Invariant Feature Transform) was proposed in [6]. At the detection stage, the algorithm builds an image pyramid – an ordered sequence containing the original image and its reduced copies, calculates the difference of Gaussians of its layers, and then finds local extrema among the calculated values. The SIFT descriptor is a histogram of the orientations of the image gradients in the area around the detected feature point. The area around the keypoint is divided into 4×4 sub-areas, in each, a histogram of the gradient directions is built (each histogram has 8 intervals). The histograms are combined and normalized. The resulting vector is a keypoint descriptor.

SURF (Speeded-Up Robust Features), described in [9], was designed based on SIFT. The main goal of the authors was to offer a less computationally expensive algorithm and, at the same time, to obtain the quality comparable to SIFT. Like SIFT, SURF calculates the difference of Gaussians of the image pyramid layers, but unlike SIFT, instead of calculating the difference of Gaussians, it approximates the Laplacian of Gaussian using a set of wavelets. During the description, the area around the keypoint is divided into 4×4 sub-areas. The horizontal and vertical wavelets values are then calculated and weighted for the grids of pixels 5×5 within each sub-region. After that, the sums of the horizontal and vertical wavelet values, as well as the sums of their modules, are calculated. The values of these sums of each sub-region are combined into a vector that describes a keypoint.

ORB (Oriented FAST and Rotated BRIEF) was proposed in [7]. The ORB detector considers a point to be a keypoint if, at a certain image scale, on a circle of fixed radius, namely 9 pixels, centered at the point, there is a sequence of pixels that are darker or lighter than the central pixel. At the stage of description, the ORB compares 256 pairs of pixels of the smoothed image patch rotated according to keypoint's orientation. Pairs are obtained from a fixed sample of the Gaussian distribution around the keypoint's center. If the first pixel of the i -th pair is darker than the second, then 1 is written to the i -th bit of the descriptor, otherwise 0.

BRISK (Binary Robust Invariant Scalable Keypoints) [22] was created as an algorithm comparable in quality to SURF, but computationally more efficient. The BRISK detection stage is similar to ORB, but it discards the points at which the FAST-score (the sum of the absolute values of the difference between the pixels of the circle with the central pixel) is not a local maximum in the image pyramid. The BRISK descriptor is similar to ORB but uses a different pattern of matched pixel pairs.

AKAZE (Accelerated-KAZE) is presented in [8].

$$L_{\text{Hessian}}^i = \sigma_{i,norm}^2 (L_{xx}^i L_{yy}^i - L_{xy}^2),$$

where $L_{xx}^i, L_{yy}^i, L_{xy}^i$ are the second-order derivatives of the i -th image, $\sigma_{i,norm}^2$ is the coefficient determined by the scale of the image on which the function value is calculated. The description stage of AKAZE algorithm is similar to ORB and BRISK, with the difference that the intensities are not compared for individual pixels, but the average value of some image areas.

III. MATERIALS AND METHODS

This section describes data used for experiments and algorithms evaluation methodology.

Ideally, to compare keypoints algorithms in different spectra we would need a set of images of the same scenes taken in both spectrums. Unfortunately, there is no known to authors public dataset meeting this requirement. In our previous work [23] we presented such a dataset, however, it contains images of a single object and therefore is not extensive enough to make a reasonable conclusion. Thus, in this work algorithms were tested on two common datasets with a large number of scenes in each spectrum (see Fig. 1). For visible spectrum Homography-patches dataset (HPatches [20]) was used. It is widely used to study keypoints detection and description algorithms. For X-ray spectrum GDXray dataset [21] was used. To perform fair comparison of spectrum influence on algorithms quality all images were scaled with saving ratio so the large side is 407 pixels. Note that we do not specify whether the height or width of the image is the larger side, since the datasets we use have both vertically and horizontally oriented images. Also, only part of GDXray dataset was used – sequences 46-48 with baggage, as they are rich for features, so all the algorithms can detect enough keypoints for a fair comparison.

A. DETECTORS EVALUATION

Let there are two images: I_1 and its transformed version I_2 . H – transformation converting I_1 to I_2 . Fig. 2 shows examples of images transformations and corresponding transformation expressions. To compare the performance of detectors on a pair of images I_1, I_2 , the so-called *repeatability* was calculated [12]:

$$r = \frac{N_{\text{correspond}}}{\min(K_1, K_2)}, \tag{1}$$

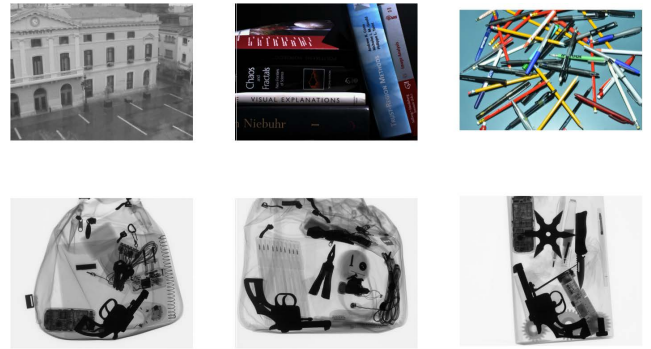


FIGURE 1. Samples of used images from HPatches [20] (top) and GDXray [21] (bottom) datasets.

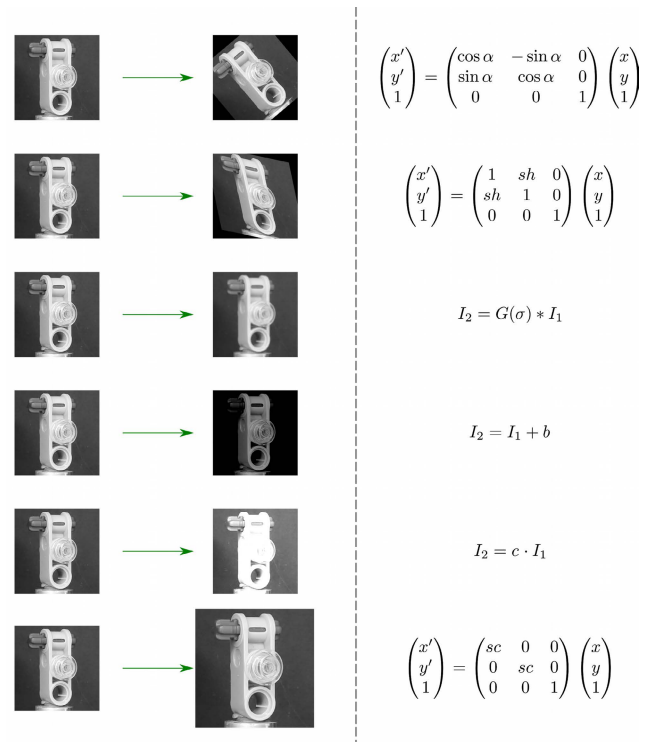


FIGURE 2. Applied transformations in detector evaluation experiments.

where $N_{\text{correspond}}$ is the number of keypoints matched between two images, K_1, K_2 are the number of points detected in the first and second images, respectively.

Let R_{μ_a} be a circle area around a keypoint on I_1 . The diameter of the region is equal to the size of the keypoint's region used to compute it's descriptor. Let also $R_{H^T \mu_b H}$ be the elliptic area around the keypoint on I_2 projected onto I_1 by the inverse mapping H^{-1} . Points are considered matched if the ratio of the intersection area of two elliptic areas to their union is greater than a certain threshold $1 - \varepsilon$ (see Fig. 3):

$$1 - \frac{R_{\mu_a} \cap R_{H^T \mu_b H}}{R_{\mu_a} \cup R_{H^T \mu_b H}} < \varepsilon. \tag{2}$$

In all experiments ε was equal to 0.4 – the default value used in OpenCV implementation.

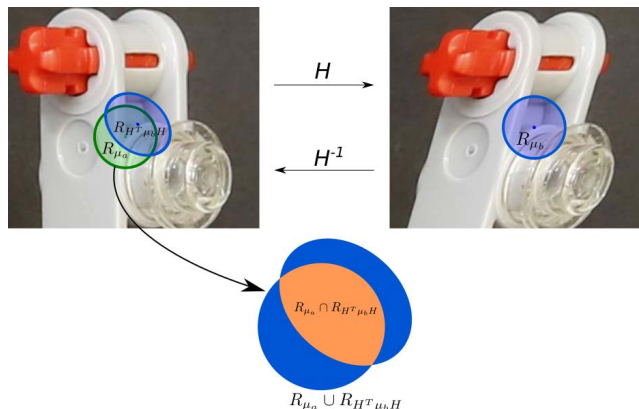


FIGURE 3. Areas of keypoints used to evaluate repeatability.

Note that when performing turn and shearing transforms some image points go beyond the fixed image boundaries. Points that are lost in the transformed image are not taken into account when calculating the *repeatability*.

The experiments to compare detectors were carried out as follows:

- 1) Image I_1 was randomly selected from the dataset; For optical images, the original image I_1 was converted to single-channel grayscale image;
- 2) To obtain I_2 , an H transforms (one of the list: rotation, shift, scaling, additive brightness changes, contrast changes, or Gaussian blur) was applied to I_1 ;
- 3) The repeatability r was calculated for the pair I_1, I_2 .

For each fixed transform H , a described experiment was performed 100 times, after which the average *repeatability* was calculated. The experiments were carried out separately for the images of the visible and X-ray spectra.

Transformations H were determined by the following parameters:

- image rotation angle α for rotation (from -15 to 15 degrees with step 0.5 degree);
- shearing sh along X and Y image axes (from 0 to 0.5 with step 0.02);
- image scaling factor s (from 0.25 to 2 with step 0.05);
- intensity b added to the intensity of all pixels to convert brightness (from -255 to 255 with a minimum intensity value of 0 and a maximum value of 255 and with step 5);
- image intensity multiplier c for contrasting (from 0 to 4 with step 0.1);
- standard deviation σ for Gaussian blur (from 0.04 to 8 pixels with step 0.04);

The dependencies of the *repeatability* r and the number of matched keypoints $N_{\text{correspond}}$ on the transformation are shown in Fig. 4.

In all experiments, software implementations of detectors and descriptors from the OpenCV library were used [24]. The parameters proposed by the authors of the original detectors and descriptors were used [6]–[9], [22].

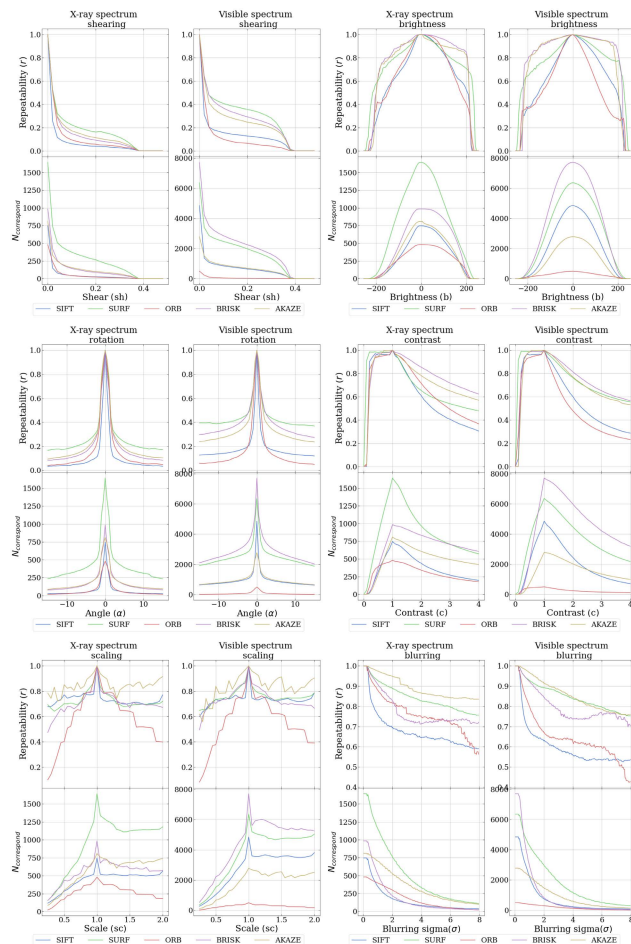


FIGURE 4. Results of detectors' comparison experiment.

B. DESCRIPTORS EVALUATION

While comparing *descriptors* one has to mitigate the influence of the *detectors* quality, since *detection* stage precedes the *description*. To evaluate descriptors and to completely exclude the influence of the detectors on the quality of the description, the identical points were chosen for description in all images instead of automatically detected keypoints. Description was performed in the nodes of the regular grid superimposed on the image and transformed with it (see Fig. 5). The grid size was 200×200 nodes. The nodes for which any algorithm could not compute a descriptor (here the descriptor is understood as the mathematical description that is produced by a description algorithm, and not an algorithm itself) did not participate in the calculation of the quality metrics described below. The scale and orientation of the grid points were also fixed. In contrast to the experiment with detectors, the transformed images were not cropped (see Fig. 5).

AKAZE was excluded from consideration in this experiment, since the software implementation of its descriptor in OpenCV requires using the points detected by its detector as input.

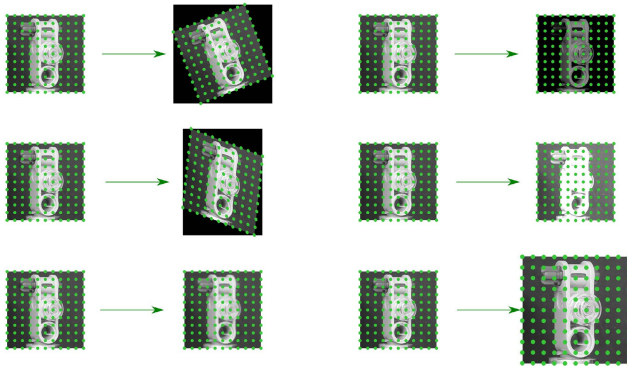


FIGURE 5. Applied transformations in descriptor evaluation experiments.

Let l_1 be the shortest distance among all distances from some fixed keypoint's descriptor of the original image to all descriptors of the transformed one, and l_2 be the second shortest distance as well. Here by distance we mean L_2 norm for SURF and SIFT descriptors and Hamming distance for BRISK, ORB. To compare descriptors we measured l_1/l_2 value. It was shown in [6] that for the SIFT descriptor for correctly matched points, this ratio most likely does not exceed ≈ 0.75 . However, this value is sometimes used when choosing the correct matches for other descriptors [25].

It is possible that a pair of the nearest descriptors for the original and transformed images are not descriptors of the same point. To take this into account, we additionally measured the proportion of correctly matched points, that is, points for which the descriptor of the transformed point is closest to the descriptor of the original point. When this fraction becomes less than 0.5 – the algorithm becomes unreliable – obtaining an incorrect match is more probable than a correct one.

Descriptors comparison experiments were carried out as follows:

- 1) Image I_1 was randomly selected from the dataset; For optical images, the original image I_1 was converted to single-channel grayscale image;
- 2) To obtain I_2 , an H transforms (one of the list: rotation, shift, scaling, additive brightness changes, contrast changes, or Gaussian blur) was applied to I_1 ;
- 3) The ratio of the distances from descriptor of each point of the original grid to the first and second nearest descriptor among the points of the transformed grid was calculated. Then the calculated values for all grid nodes were averaged.

As in the study of descriptors for each transformation H , measurements were made on 100 randomly selected images and then averaged. The experimental results are shown in Fig. 6. Based on the results obtained, it can be concluded that often the ratio of correctly matched points falls below 50% earlier, or, on the contrary, much later than the ratio of the two nearest descriptors reaches 0.75. This, however, is not an indicator of the inconsistency of the selected quality metrics, since the position of the points for which the

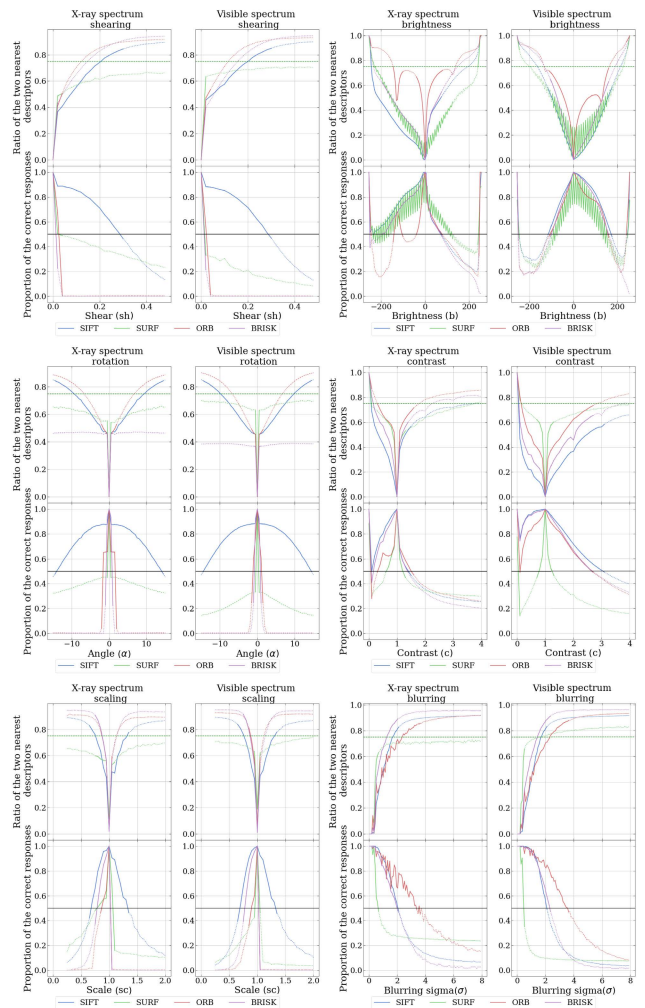


FIGURE 6. Results of descriptors' comparison experiment.

descriptors were calculated was initially set without taking into account the informativeness of their surrounding areas. The obtained result should be considered as a demonstration of the fact that this metric should be used with caution when analyzing different pairs of detectors and descriptors, since the detector can detect keypoints that are not informative for the descriptor.

IV. RESULTS

A. DETECTORS

Table 1 lists the best and worst detection algorithms for each type of image and applied transformation. The algorithms were ranked based on the repeatability value (the first two columns of the graphs in Fig. 4). We found it difficult to choose the worst algorithm in experiments to change the contrast in the X-ray spectra, as different detectors exhibited the lowest repeatability for different values of the contrast parameter. We chose SIFT as the worst detector under contrasting, because when the contrast is increased, it is definitely inferior to the others, and when decreased it is comparable to them.

A sharp drop in the repeatability of all detectors is observed when the image is rotated. All algorithms are robust to

TABLE 1. The best/worst detectors for each type of transformation.

Transform	X-Ray spectrum		Visible spectrum	
	best detector	worst detector	best detector	worst detector
Blurring	AKAZE	SIFT	AKAZE	SIFT
Brightness	BRISK	ORB/SIFT	BRISK/AKAZE	ORB
Contrast	BRISK/SURF	SIFT	BRISK/SURF	ORB
Rotation	SURF	SIFT	SURF	ORB
Shearing	SURF	SIFT	SURF	ORB
Scaling	AKAZE	ORB	AKAZE	ORB

brightness changes. All algorithms find significantly fewer points in the X-ray spectrum than in the visible one, which is most likely due to the absence of textures. Table 2 lists the types of transformations in which the algorithm showed the best or worst quality of work, as well as the average number of detected points. And table 3 contains ratios of area under the actual detector's plot to the "perfect" one's, i.e. constant equal to the actual plot's max value. On X-ray data, SURF is often the most repeatable. AKAZE, in turn, outperforms other algorithms on images in the visible spectrum. SIFT showed itself worst of all.

B. DESCRIPTORS

On Fig. 6) the proportion of correct matches, a sharp jump is observed with changes in brightness and contrast. This is due to the fact that the images become almost uniformly black or white during conversion, which leads to almost identical descriptors for different image patches. Table 4 provides an analysis of the best and worst descriptors for each type of transformation. The main criterion for choosing algorithms was the proportion of correct matches (the third and fourth columns of the graphs in Fig. 6). In disputable situations, an algorithm was chosen with a smaller ratio of the distance to the two nearest descriptors. Table 5 contains ratios of area under the descriptor's plot for all algorithms, transforms and spectra (the x-axis is normalized, i.e. the maximal possible area under the graph is 1). Table 6 shows the types of transformations in which the description algorithm showed the best or worst quality of work. The SIFT descriptor showed definitely the best quality, SURF showed itself the worst. We found it difficult to choose the best algorithm in experiments on changing the brightness of X-ray images, since the three algorithms demonstrated similar performance. The quality of the description algorithms, as in the case of detectors, is most sensitive to image rotation. In general, the quality indicators for the description algorithms differ less in different spectra than the quality indicators of the detection algorithms. Let's analyze in more detail all detection and description algorithms.

C. SIFT

1) DETECTOR

The algorithm demonstrates the worst performance in the X-ray spectrum and comparatively poor in the visible one. Differences in the behavior of the algorithm on images in different spectra are observed for rotation and shearing: the

repeatability for the X-ray spectrum decreases significantly faster. The algorithm turned out to be the most stable to a change in scale; on both spectra, after a certain drop, the repeatability remained generally constant. The other advantages of the algorithm also include a fairly large number of detected points (about 750 for the X-ray spectrum and about 5000 for the visible one).

2) DESCRIPTOR

In terms of the description quality, SIFT demonstrates the best quality for many transformations. Specifically, for shearing and rotation, the proportion of correct matches decreases much slower than for other algorithms. For scaling, brightness, and contrast changes, the algorithm also shows the best quality, albeit with a smaller advantage over other descriptors. The only case where the algorithm does not show superior quality is blurring. It is interesting to note the similarity of the behavior of correct matches for SIFT and BRISK for blurring.

In general, both in the visible and X-ray spectra, the algorithm shows the same behavior. The only exception is the behavior when changing brightness and contrast. With an increase in the brightness in the visible spectrum, the quality of the algorithm decreases somewhat slowly (although comparable) than with a decrease. In the X-ray spectrum, the opposite is observed: the quality decreases noticeably slowly with decreasing brightness than with increasing. It can be assumed that this is due to the nature of the images, for example, that the areas in the X-ray image well described by the algorithm contain mainly black or light pixels, therefore, with increasing brightness, light pixels of the area become simply white. As the brightness decreases, the pixels remain distinguishable. Similarly, a sharp drop in quality can be explained when the contrast changes.

D. SURF

1) DETECTOR

SURF demonstrates consistently good quality at all transformations in both spectra. The algorithm performs best of all under shearing and rotating. Under blurring, the SURF quality is sometimes inferior comparable to AKAZE and superior to other algorithms. For scaling, brightness changing, and contrasting, the quality of the algorithm is not much different from most others. The difference in behavior on different spectra is observed for rotation and shearing – as for most algorithms, the quality in the X-ray spectrum drops faster. This is true for all of the detectors except ORB, which repeatability does not change. As an advantage, we can note the fact that SURF detects the largest average number of points in the X-ray spectrum – 1600, which is 1.5 times higher than the next best algorithm for this parameter - BRISK. For the visible spectrum, the number of points is inferior to BRISK but superior to other algorithms.

2) DESCRIPTOR

Unlike the detection algorithm, the SURF descriptor performed poorly. When rotating, changing contrast, blurring,

TABLE 2. The number of transformations in which the detection algorithms showed the best/worst quality and the average number of detected keypoints.

Detector	X-Ray spectrum			Visible spectrum			Summary	
	best	worst	average number of keypoints	best	worst	average number of keypoints	best	worst
SIFT	0	5	746	0	1	4847	0	6
SURF	3	0	1643	3	0	6358	6	0
ORB	0	2	481	0	5	497	0	7
BRISK	2	0	985	2	0	7707	4	0
AKAZE	2	0	810	3	0	2778	5	0

TABLE 3. Ratio of area under actual detector graph to area under “perfect” graph. Cell format is X-Ray/visible spectra values.

	Blurring		Brightness	
	Repeatability	$N_{\text{correspondences}}$	Repeatability	$N_{\text{correspondences}}$
SIFT	0.678/0.602	0.213/0.164	0.600/0.609	0.432/0.435
SURF	0.845/0.848	0.318/0.299	0.727/0.754	0.471/0.489
ORB	0.765/0.649	0.366/ 0.342	0.607/0.534	0.522/0.491
BRISK	0.778/0.800	0.197/0.136	0.739/0.796	0.488/0.489
AKAZE	0.893/0.855	0.437/0.338	0.717/0.795	0.442/0.437
	Contrast		Rotation	
	Repeatability	$N_{\text{correspondences}}$	Repeatability	$N_{\text{correspondences}}$
SIFT	0.602 / 0.588	0.489/0.425	0.106/0.185	0.100/0.196
SURF	0.696/ 0.758	0.531/0.578	0.271/0.443	0.261/0.398
ORB	0.658/0.527	0.622/0.500	0.154/0.153	0.148/0.147
BRISK	0.781/0.751	0.684/0.589	0.197/0.392	0.185/0.376
AKAZE	0.739/0.750	0.614/0.551	0.222/0.342	0.231/0.346
	Scaling		Shearing	
	Repeatability	$N_{\text{correspondences}}$	Repeatability	$N_{\text{correspondences}}$
SIFT	0.081/0.140	0.076/0.142	0.735/0.735	0.597/0.577
SURF	0.177/0.296	0.167/0.263	0.727/0.760	0.581/0.594
ORB	0.109/0.109	0.102/0.104	0.584/0.559	0.562/0.544
BRISK	0.134/0.266	0.127/0.259	0.705/0.740	0.569/0.576
AKAZE	0.147/0.236	0.150/0.232	0.848/0.835	0.697/0.659

TABLE 4. The best/worst descriptors for each type of transformation.

Transform	X-Ray spectrum		Visible spectrum	
	best descriptor	worst descriptor	best descriptor	worst descriptor
Blurring	ORB	SURF	ORB	SURF
Brightness	SURF	ORB	SIFT	SURF
Contrast	SIFT	SURF	SIFT	SURF
Rotation	SIFT	SURF	SIFT	SURF
Shearing	SIFT	BRISK	SIFT	SURF/BRISK
Scaling	SIFT	BRISK	SIFT	SURF

and scaling (in the visible spectrum), the algorithm works worst of all, because, before the rest of the algorithms, it experiences the quality drop when it correctly matches less than half of the described points. With shearing in the X-ray spectrum, the algorithm does not work much better than the worst descriptor – BRISK, but in the visible spectrum, they show similar performance. The algorithm works relatively well when changing the brightness, at some values it reaches the quality of the SIFT that showed itself best in this experiment. A significant change in behavior in different spectra can be noted only for the change in brightness. We assume that the explanation of this fact coincides with the explanation given for the SIFT.

E. ORB

1) DETECTOR

In most experiments, ORB performed worse than other algorithms and became the worst in the visible spectrum.

Especially poorly, the algorithm performs under scaling. When decreasing the scale, the repeatability of the algorithm quickly decreases to 0, while for other algorithms it remains above 0.5. The repeatability plot for this transformation has a stepped-like appearance which indicates poor ability to detect keypoints on intermediate scales between image pyramid layers. However, it can be noted that under blurring the algorithm shows moderate results comparable to BRISK and outperforming SIFT up to $\sigma \approx 7$. The ORB also detects the smallest number of keypoints. Changes in the algorithm behavior in different spectra can be noted in brightness, contrast, and blurring. For these transformations, repeatability decreases slightly more slowly in the X-ray spectrum. A possible explanation for this is that changing brightness and contrast leads to clipping pixel values and that changes ORB descriptors as it uses the pairwise comparison of pixel intensities. At the same time, X-ray images contain mostly middle range (grey) pixel values, so the clippings occur later for X-ray images than for visible ones. Also, visible-spectrum images are characterized by a lot of small details (e.g. in textures) that disappear when blurred. Thus blurring changes detector comparing pixels’ intensity more strongly for such images.

2) DESCRIPTOR

The ORB description algorithm is of ambiguous quality. In such transformations as changing brightness, rotation, scaling, its quality decreases rapidly. At the same time, when increasing the contrast for the visible spectrum,

TABLE 5. Area under actual descriptor graph. X-axis is normalized to 1. Cell format is X-Ray/visible spectra values. * – less is better.

	Blurring		Brightness	
	Ratio*	$N_{\text{correspond}}$	Ratio*	$N_{\text{correspond}}$
SIFT	0.785/0.777	0.322/0.314	0.467/0.489	0.585/0.621
SURF	0.669/0.755	0.310/0.142	0.443/0.500	0.603/0.521
ORB	0.759/0.764	0.489/0.484	0.744/0.670	0.446/0.553
BRISK	0.850/0.844	0.552/0.311	0.552/0.568	0.519/0.534

	Blurring		Brightness	
	Ratio*	$N_{\text{correspond}}$	Ratio*	$N_{\text{correspond}}$
SIFT	0.603/0.411	0.475/0.712	0.671/0.659	0.734/0.749
SURF	0.644/0.668	0.402/0.321	0.602/0.664	0.399/0.240
ORB	0.747/0.623	0.440/0.634	0.725/0.728	0.084/0.073
BRISK	0.679/0.518	0.438/0.672	0.458/0.375	0.054/0.057

	Blurring		Brightness	
	Ratio*	$N_{\text{correspond}}$	Ratio*	$N_{\text{correspond}}$
SIFT	0.707/0.724	0.575/0.573	0.752/0.755	0.429/0.418
SURF	0.592/0.659	0.387/0.215	0.622/0.678	0.265/0.149
ORB	0.781/0.797	0.070/0.064	0.833/0.852	0.094/0.087
BRISK	0.776/0.768	0.049/0.048	0.861/0.860	0.138/0.142

TABLE 6. The number of transformations in which the description algorithms showed the best/worst quality.

Descriptor	X-Ray spectrum		Visible spectrum		Total	
	Best	Worst	Best	Worst	Best	Worst
SIFT	4	0	5	0	9	0
SURF	1	3	0	6	1	9
ORB	1	1	1	0	2	1
BRISK	0	2	0	1	0	3

the proportion of correct matches of the algorithm is comparable to the best algorithms for this transformation (SIFT, BRISK), and for blurring, the algorithm is superior to the rest. The behavior when changing the brightness is noteworthy, since there is a temporary decrease in the ratio of two nearest descriptors.

As with other algorithms, a change in behavior for different spectra is observed when changing the brightness and contrast. In addition, upon rotation in the X-ray spectrum, a small plateau is observed in the vicinity of 0 for the fraction of correct matches (if we exclude the identical transformation). This plateau is absent in the visible spectrum.

F. BRISK

1) DETECTOR

BRISK outperforms other algorithms under increasing contrast. As well as AKAZE it also demonstrates the best repeatability in experiments with changing brightness and the best one after SURF in visible spectrum under rotation. For other transformations, quality varies but never becomes the worst. It should be noted that the algorithm detects the largest average number of detected keypoints in the visible spectrum – 7700. There are no significant changes in the behavior of the detector when changing the spectrum, except for the previously mentioned shearing and rotation.

2) DESCRIPTOR

The algorithm demonstrates moderate quality for all transformations. Under shearing BRISK works worse than other

algorithms (comparable to ORB), however, under the shearing, the quality of all algorithms, except SIFT, is poor. Most significant behavior changes in different spectra can be seen under changes in brightness, contrast, and blurring. A possible explanation for this was given in the discussion of the results for the ORB detector.

G. AKAZE

1) DETECTOR

The algorithm has demonstrated good quality of work in most experiments. For scaling and blurring experiments AKAZE has become the leader among the compared algorithms. When changing the contrast, the algorithm lags a little behind BRISK in some places, but still performs well. For the rest of the transformations, it shows moderate results. When changing spectrum the behavior changes for brightness transformation: repeatability drops for X-ray images faster.

2) DESCRIPTOR

As mentioned before, AKAZE was not involved in the study of descriptors. In the experiments, we used OpenCV implementations of the algorithms and AKAZE implementation requires descriptors computation only for points detected by the same detector. This is not consistent with the design of the experiment.

V. CONCLUSION

In this work, we studied the performance of wide-used keypoints detection and description algorithms (SIFT, SURF, ORB, BRISK, AKAZE) applied to images taken in visible and X-ray spectra. We also studied the robustness of the algorithms to various transformations (rotation, shearing, scaling, brightness and contrast changes, Gaussian blur).

We confirmed the assumption about the difference in the behavior of algorithms when working with images in X-ray and visible spectra. First of all, in the images of the X-ray spectrum, all algorithms detect a significantly smaller number of keypoints, which can adversely affect the quality of the computer vision systems that use keypoints. SURF showed the best quality among detection algorithms on X-ray images, and on visible ones, it shares the primacy with AKAZE. The worst detection quality was observed for SIFT and ORB for X-ray and visible spectra respectively. It is shown that in the X-ray spectrum all detection algorithms are less robust for rotation and shearing. For the rest of the studied transformations, in general, there is no significant change in the quality of both keypoints detection and description.

SIFT became the best descriptor in our experiments, and SURF became the worst one. Note that the AKAZE descriptor, due to the limitations of its software implementation and the specifics of the experiment, did not participate in the comparison. Taking into account the fact that SURF uses similar principles of keypoints detections as SIFT, one should expect that a SURF detector paired with a SIFT descriptor will give the best performance in the majority of applications. We plan to investigate this issue in more detail in future work.

Note that the performed experiment for comparing descriptors has a drawback: it calculates descriptors, at the nodes of a regular grid (see Fig.5), which makes the comparison fair, but does not guarantee that selected pixels and their vicinity are informative. It can lead to poor quality of keypoints matching. In the future, we plan to conduct additional experiments to take described drawback into account. In addition, we plan to add neural network detectors and descriptors into the study.

In addition, the speed of the algorithms was not investigated, which is certainly important for real-time application. Although it is beyond the scope of this paper, we are planning to include relevant studies in further works.

REFERENCES

- [1] R. M. Haralick and L. G. Shapiro, "Connected components labeling," *Comput. Robot. Vis.*, vol. 1, pp. 28–48, Oct. 1992.
- [2] V. Lepetit and P. Fua, "Keypoint recognition using randomized trees," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1465–1479, Sep. 2006.
- [3] N. S. Skoryukina, A. N. Milovzorov, D. V. Polevoy, and V. V. Arlazarov, "Paintings recognition in uncontrolled conditions using one-shot learning," *Trudy ISA RAN*, vol. 68, pp. 5–14, May 2018, doi: [10.14357/20790279180501](https://doi.org/10.14357/20790279180501).
- [4] S. Agarwal, "Building Rome in a day," *Commun. ACM*, vol. 54, no. 10, pp. 105–112, Oct. 2011.
- [5] D. Scaramuzza, "Performance evaluation of 1-point-RANSAC visual odometry," *J. Field Robot.*, vol. 28, no. 5, pp. 792–811, Sep. 2011.
- [6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [7] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571.
- [8] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "Kaze features," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 214–227.
- [9] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2006, pp. 404–417.
- [10] M. Rodriguez, G. Facciolo, R. G. Von Gioi, P. Muse, J.-M. Morel, and J. Delon, "SIFT-AID: Boosting sift with an affine invariant descriptor based on convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 4225–4229.
- [11] Z. Song and R. Klette, "Robustness of point feature detection," in *Proc. Int. Conf. Comput. Anal. Images Patterns*. Berlin, Germany: Springer, 2013, pp. 91–99.
- [12] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1, pp. 43–72, 2005.
- [13] J. Hu, X. Peng, and C. Fu, "A comparison of feature description algorithms," *Optik*, vol. 126, no. 2, pp. 274–278, Jan. 2015.
- [14] S. Akhmedkhan, S. Gladilin, and E. A. Shvets, "Optical-to-SAR image registration using a combination of CNN descriptors and cross-correlation coefficient," in *Proc. 12th Int. Conf. Mach. Vis. (ICMV)*, Jan. 2020, Art. no. 114331.
- [15] M. Moradi, P. Abolmaesoumi, and P. Mousavi, "Deformable registration using scale space keypoints," *Proc. SPIE Med. Imag. Process.*, vol. 6144, Aug. 2006, Art. no. 61442G.
- [16] V. V. Volkov and E. A. Shvets, "Dataset and method for evaluating optical-to-sar image registration algorithms based on keypoints," *Inf. Tekhnol. Syst.*, vol. 2, pp. 44–57, Oct. 2021, doi: [10.14357/20718632210205](https://doi.org/10.14357/20718632210205).
- [17] F. Lecron, M. Benjelloun, and S. Mahmoudi, "Descriptive image feature for object detection in medical images," in *Proc. Int. Conf. Image Anal. Recognit.* Berlin, Germany: Springer, 2012, pp. 331–338.
- [18] M. O. Chekanov, O. C. Shipitko, and E. I. Ershov, "1-point ransac for axial rotation angle estimation by tomographic projections," *Sensory Syst.*, vol. 34, no. 1, pp. 72–86, 2020, doi: [10.31857/S0235009220010060](https://doi.org/10.31857/S0235009220010060).
- [19] D. Mery, E. Svec, M. Arias, V. Riffo, J. M. Saavedra, and S. Banerjee, "Modern computer vision techniques for X-ray testing in baggage inspection," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 4, pp. 682–692, Apr. 2017.
- [20] V. Balntas, K. Lenc, A. Vedaldi, and K. Mikolajczyk, "HPatches: A benchmark and evaluation of handcrafted and learned local descriptors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5173–5182.
- [21] D. Mery, V. Riffo, U. Zscherpel, G. Mondragón, I. Lillo, I. Zuccar, H. Lobel, and M. Carrasco, "GDxRay: The database of X-ray images for nondestructive testing," *J. Nondestruct. Eval.*, vol. 34, no. 4, pp. 1–12, 2015.
- [22] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2548–2555.
- [23] M. Chekanov and O. Shipitko, "X-ray and visible spectra circular motion images dataset," 2019, *arXiv:1909.13730*.
- [24] G. Bradski, "The OpenCV Library," *Softw. Tools Prof. Programmer*, vol. 25, no. 11, pp. 120–123, 2000.
- [25] O. Andersson and S. R. Marquez, "A comparison of object detection algorithms using unmanipulated testing images: Comparing SIFT, KAZE, AKAZE and ORB," B.S. Thesis, Dept. Comput. Sci., KTH Roy. Inst. Technol., Stockholm, Sweden, 2016.



MIKHAIL CHEKANOV was born in Russia, in 1998. He received the B.E. degree (Hons.) in applied mathematics and physics from the Moscow Institute of Physics and Technology (MIPT), National Research University, Moscow, Russia, in 2020. He is currently pursuing the master's degree with MIPT. He joined as a Research Intern with the 11th Visual Systems Laboratory, IITP, RAS, Moscow, in 2019. Since 2020, he has been a Software Engineer with Evocargo LLC, Moscow. His main research interests include robotics and classical computer vision.



OLEG SHIPITKO was born in Yeisk, Russia. He received the bachelor's degree in electrical engineering from Bauman Moscow State Technical University, in 2015, and the master's degree in computer science from the Skolkovo Institute of Science and Technology, in 2017. Since 2017, he has been with the Laboratory of Vision Systems, Institute for Information Transmission Problems, Russian Academy of Sciences (Kharkevich Institute). His research interest includes vision-based localization algorithms for robotics.



NATALIA SKORYUKINA received the graduate degree from the National University of Science and Technology «MISiS», in 2013, majoring in applied mathematics. She is a Computer Programmer (1st category) with the Federal Research Center of Computer Science and Control, Russian Academy of Sciences. Her research interests include image analysis and computer vision.

...