Method for Detecting False Responses of Localization and Identification Algorithms Using Global Features

N. S. Skoryukina[I,II], E. A. Shalnova[I], V. V. Arlazarov[I,II]

[I] Smart Engines Service LLC, Moscow, Russia
[II] Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

Abstract. The paper presents a method for detecting false responses of localization and identification algorithms. The method considers matching image characteristics that cannot be described by local features stably and completely. It is proposed to use image zones containing such features, describe them and use them to assess the validity of the algorithm response. In the work we demonstrate how the algorithm works on ID documents. Possible features are images of the coats of arms and flags of countries, background filling and text unique to the considered document type. To illustrate the proposed algorithm, the MIDV-500 and MIDV-LAIT datasets were taken. The first is used to show that the rejector does not reject correct system responses, the second - that it rejects the incorrect ones. We test several methods of zone description. The experimental results show that false type selection decreases with the use of any description type and the local CNN-descriptor shows the best performance. The increase of classes with marked zones is shown to improve the filtration of false responses. The experiments show the improvement from by 13% with one type with zones to by 4 times with 10 types.

Keywords: candidate rejection (rejector), image features, localization, identification

DOI 10.14357/20718632230403 **EDN** CGRAFY

Introduction

Recognition systems are an important component of RPA solutions with data entry automation. Various organizations, both government establishments and private companies, use such solutions to improve the processes efficiency, incl. document flow. The use of recognition systems for document images not only speeds the input up, but also provides the ability to check documents for authenticity. Although the problem of document recognition has been studied for more than 50 years, it remains relevant due to development of mobile devices, remote servers and anti-fraud methods.

Document localization and identification in an image in the coarse-to-fine approach [1] is the first stage of the document recognition process. The task can be formulated as follows:

- The input is a query image that may contain a document and a list of document classes that the system must identify;
- The output is a set of candidates consisting of document position and document type.

There are different approaches to solving this problem. For example, the works [2, 3] present solutions to this problem using neural networks. State-of-the-art solutions for documents are obtained using a common approach using image matching based on local features. According to works [4-6], the following sequential stages are distinguished:

0. Template images are set in accordance with document classes;

1. local features selection and description in the query image and template images;
2. feature matching between the query image and template images;
3. the determination of the most suitable document class and its outer border in the query image.

There are lots of different approaches for each stage. For example, one of the most common algorithms for keypoints detection and description are SURF [7] and SIFT [8]. The brute force search can be considered as a matching algorithm. Based on the found matches, a homography is determined using the RANSAC method [9].

Number of papers are devoted to the improvement of one of the stages of the algorithm. For example, RFD [10], HOG [11] and BEBLID [12] are examples of different approaches to keypoints description step. A common example of extracting and descripting non-point features is segment detection, as described in the article [13]. As an example of improving the matching stage, we can cite the algorithm for searching the nearest neighbors described in [14]. The RANSAC method also has improvements. For example, the authors of [15] propose the USAC method (Universal RANSAC) combining many SAC-related optimizations and improvements that have been proposed over the years. In the paper [16], the authors focus on geometric constraints and improve the method for the document recognition problem.

There are also articles that improve the entire approach to solving the recognition problem. For example, the article [17] proposes a multiple hypothesis approach, which consists of running several document localization strategies in parallel, and the selection of the most competitive one based on a visual similarity score. In addition, some works present new and/or improved algorithms, there are rejections of candidates by the geometric properties of the quadrangle that defines the boundaries of the document according to the found homography, and by the number of inliers – the matched keypoints inside the quadrangle [5, 6, 18].

Despite all the modifications and improvements described, the approach to document localization and identification by matching keypoint descriptors presented above actually solves the problem by analyzing individual local neighborhoods. At the same time, the document template itself is not considered as a whole, in a global sense, leading to possible localization and identification mistakes (Fig. 1). Such mistakes can be avoided if not only local, but also global features are taken into account during localization and identification. In the work [19] global features are taken into account at the time of identification providing an algorithm for classifying documents by the presence of a logo detected without segmentation.

Fig. 1. Example of a false positive response, segments are estimated inliers, the quad is the hypothesis of document boundaries

Thus, the analysis of the global features of the document template at the time of document localization and identification is an unsolved problem. In this work, we focus on the problem of rejecting system response candidates by matching unique global static zones.

1. Problem Formulation

Given an input image I_q and the image classes $C = \{C_0 \dots C_N\}$ is defined for N object* types:

- C_i - class of images with i-th type, for $i \in [1, N]$
- C_0 - class of other images ("trash")

(* - identity documents in this paper)

For each class C_i an ideal image - template - T_i is defined. The template is an image of a document cropped to its border (see Fig. 2, a).

A common response of a localization and identification system can be formalized as $<C_i, H>$, where H is the transformation: $H: I_q \rightarrow T_i$, so H translates pixels of a planar object in I_q to pixels of T_i for $i \in [1,N]$. In this a family of projective transforms is considered, which can be caused by a pinhole camera and which can be expressed as a 3×3 matrix operator.

Template images can contain variable data zones. In case of ID this kind of zone contains personal data. Following works similar to [5], variable data zones are selected on each template image. Image features like feature points inside selected zones are masked and get ignored during feature matching (Fig. 2, b).

There are few symptoms of ill-conditioned solution in obtained response. In SOTA approach, the result pair $<C_i, H>$ is ill-conditioned if the lack of inliers, or correspondent points are densely placed on a small part of T_i, or the difference (numbers of inliers) between result and 2-nd rate candidate is insignificant. Still and all, these symptoms can be caused by the template image specificity, among other things. So all of these are not enough to reject the result hypothesis.

Ill-conditioned solution can be rejected by detection of invalid geometric transformation (impossible camera pose) and requires additional validation otherwise.

Let us note that:

- sometimes we can match toneless regions instead of or together with salient regions
- auto distinguish between visual-similar classes are confined due local features have small view area; it needs more complex context for compute difference.

So in this paper we propose to introduce the special zones - rectangle areas which are manually selected on template images. Zone matching will be used for acceptance or rejection of hypotheses. Next part we will consider the principles of zone selection and matching in detail.

2. Suggested Method

Let us consider a specific implementation of the general document match algorithm. At first the keypoints are extracted for the template images and the query image and described using the SURF method. After that, the query points are matched with each set of template points with brute force search. Based

a) b)

Fig. 2. Template example

a) image of document, cropped by the border; b) image of document with a mask on the obtained comparisons, the RANSAC method is used to find a homography, which is used to determine the quadrangle of the document. Next, from all matching candidates, the candidate with the largest number of inliers. The block diagram of the described algorithm is presented in Fig. 3.

The described algorithm may incorrectly identify a document in images having similar local features (see Figure 1). In order to reduce the number of such responses, we propose to use global features of documents as follows. Each document contains unique features, such as, for example, images of coats of arms and flags, specific words, pieces of background, etc. (Fig. 4). Uniqueness lies not only in the content of the feature, but also in its location on the document. It is proposed to highlight such features in a template image with a quadrangle, further called a rejection zone, and divide it into blocks. Each block is proposed to be described and used for comparison.

It is proposed to do the same with the system response candidate. Based on the considered quadrilateral, being the position of the document, the query image is transformed to a projectively normalized image of the document, the type of which the system has identified. In such an image, rejection zones corresponding to the document are distinguished and described (Fig. 5). Next, the descriptors corresponding to the same block and computed on the template image and the query image are compared and their similarity is assessed.

For the previously described localization and identification algorithm, it is proposed to use rejection zones in the following way. Candidates coming

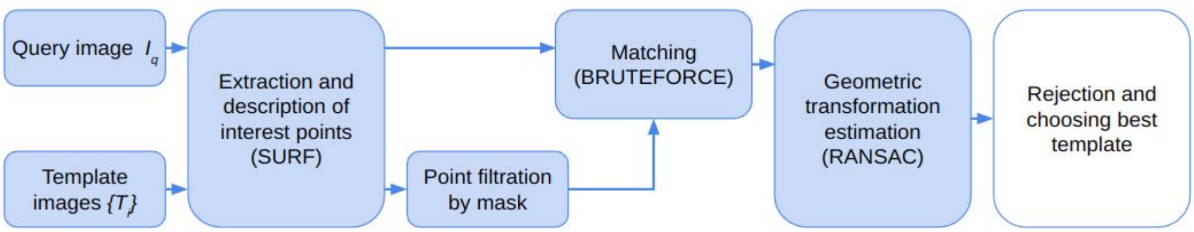

Fig. 3. Diagram of the general recognition algorithm with the proposed stage

Fig. 4. Example of unique global features on documents of various types

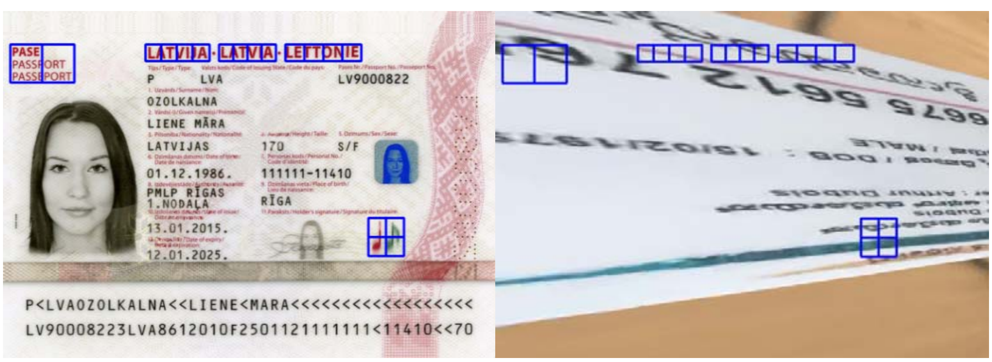

Fig. 5. Example of rejection zones on a template and a projectively corrected query image

from the keypoint matching stage are sorted by the number of inliers. In the process of counting the number of inliers, the quadrangle corresponding to the homography is checked for validity based on its geometric properties [6]. Starting with the candidate with the highest value, a check is performed across the template zones. If, after the homography, a candidate zone corresponding to a template has descriptors the distance from which to the corresponding template descriptors is greater than a given threshold for more than half of the blocks, then this zone is considered dissimilar. If the number of dissimilar zones is more than half of the total number of zones in the template, then this candidate is not considered in choosing the best one (see Algorithm 1).

3. Experimental Data

MIDV-500 [20] and MIDV-LAIT [21] were taken as sets of query images. Images in which the document is completely present are considered. Each image corresponds to a file with markup including the document type.

MIDV-500 presents sets of images of 50 document types and their markup containing the document type presented in the image. The dataset also contains a set of 50 types of templates with markup. In the markup containing a mask of ten types, a description of the rejection zones was added, represented by a rectangle and the number of divisions along the x-axis and y-axis.

From MIDV-LAIT, sets of images containing documents not from the 50 MIDV-500 types were taken. These images were marked as images with no document.

This choice of datasets is dictated by the following. In order to show that the proposed algorithm works in cases where the image contains a document similar to one of the given document sets, the MIDV-LAIT dataset was taken. On it, the described localization and identification algorithm produces wrong answers. Ideally, this set of images should not contain any documents. MIDV-500 is used to evaluate the rejector for false positives. Ideally, none of the correct answers produced by the system should be rejected by the rejector.

INPUT projectively corrected query image I'_q, template zones Z_q, template zone descriptors D_q
OUTPUT $reject$

$$reject \coloneqq false$$
$$counter\_rejected\_zone \coloneqq 0$$

for each $z_i \in Z_q$ **do**

$$counter\_rejected\_block \coloneqq 0$$

  Let B be the set of blocks that are contained in z_i
  for each $b_j \in B$ **do**
    Let d'_j be descriptor of b_j computed on the image I'_q
    Let d_j be descriptor of b_j from D_q
    if $dist(d_j, d'_j) > BLOCK\_THRESHOLD$ **then**
      $counter\_rejected\_block \coloneqq counter\_rejected\_block + 1$
    end if
  end for
  if $counter\_rejected\_block > \frac{|B|}{2}$ **then**
    $counter\_rejected\_zone \coloneqq counter\_rejected\_zone + 1$
  end if
end for
if $counter\_rejected\_zone > \frac{|Z_q|}{2}$ **then**
  $reject \coloneqq true$
end if
return $reject$

Algorithm 1. Zone rejection algorithm

4. Results

The experiment was conducted as follows. In order to show that the method works regardless of the block description algorithm (there is a dependence only in quality), the SURF, BINBOOST [22] and neural network [23] methods were taken for comparison as a block description algorithm. Measurements were carried out without zones, with one, with five and with ten document types with zones on the MIDV-500 and MIDV-LAIT datasets. The limitation of ten types of documents is due to the fact that in the examples of the MIDV-LAIT dataset without a zone there were clearly ten types of documents that were falsely localized by the system.

The distance between the descriptors for SURF and BINBOOST was calculated as follows. Two sets of image requests were taken: containing a document and correctly defined by the system, and those not containing the document, but the system detected the document. For such sets, the average distances between blocks of zones were calculated and the average between them was taken. The threshold for the neural network method was taken from the network structure.

Below, Tables 1 and 2 present the results of the experiment. The tables are divided into experiments with 1, 5, and 10 document types on MIDV-500 and MIDV-LAIT, respectively. For each experiment, data sets without a zone are measured as a control measurement.

Each table contains two rows: correct - the percentage of correct responses (MIDV-500 - correct doctype, MIDV-LAIT - response without doctype) from the total number of requests, incorrect - the percentage of responses with incorrect doctype from the total number of requests. Table 1 is missing a line - the percentage of responses without doctypes from the total number of requests.

As can be seen from the tables, adding zones reduces the number of incorrect system responses. This means that the rejector rejects incorrect answers, that is, it works as expected. The MIDV-500 shows a decrease in the percentage of correct answers, a decrease in incorrect answers and an increase in missing. This is due to the fact that part (less than 2%) of correct answers was filtered out along with the incorrect ones. From the tables with measurements with 10 types, it can be seen that adding zones reduces the number of false positives by more than 2 times in all three descriptor examples and reduces the percentage of correct answers by no more than 2%. The latter decrease is due to the fact that among the correct answers there are responses with the correct type, but incorrect localization. Such responses can be regarded as false, and getting rid of them leads to an increase in the quality of further stages of recognition.

Adding zones significantly reduces the number of incorrect responses, and at the same time slightly reduces the number of correct answers. In this

Table 1. Measurement without zones and with 1, 5 and 10 types of documents with zones on the MIDV-500

		1 zoned doctype			5 zoned doctypes			10 zoned doctypes		
	No zone	SURF	BINBOOST	CNN	SURF	BINBOOST	CNN	SURF	BINBOOST	CNN
correct	0.8999	0.9000	0.8997	0.9000	0.8965	0.8874	0.8930	0.8953	0.8840	0.8879
incorrect	0.0081	0.0073	0.0077	0.0073	0.0059	0.0088	0.0065	0.0047	0.0073	0.0052
missing	0.0920	0.0927	0.0926	0.0927	0.0975	0.1038	0.1005	0.1000	0.1088	0.1069

Table 2. Measurement without zones and with 1, 5 and 10 types of documents with zones on the MIDV-LAIT

		1 zoned doctype			5 zoned doctypes			10 zoned doctypes		
	No zone	SURF	BINBOOST	CNN	SURF	BINBOOST	CNN	SURF	BINBOOST	CNN
correct	0.9250	0.9338	0.9306	0.9341	0.9553	0.9459	0.9559	0.9797	0.9662	0.9812
incorrect	0.0750	0.0662	0.0694	0.0659	0.0447	0.0541	0.0441	0.0203	0.0338	0.0188

regard, it is proposed to introduce zones not on all documents, but only on those where local features tend to be mismatched to the features of other documents.

Conclusion

This paper presents a method for validating responses from localization and identification systems, based on rejection zones that represent unique global features of objects. The method realization requests access to template images. The method is based on the comparison of zone descriptors calculated on the query image and the template image. It is shown that the method improves the quality of document identification regardless of the zone description method used, by reducing system responses with wrong type.

This method can be used after any localization and identification algorithm that produces a response in the form of a document type and the position of the document in the query image. At the same time, it is important to note that for the method to work, it does not require the presence of zones on every type. It is enough to process only those that are more likely to get confused, following empirical or theoretical knowledge.

References

1. Arlazarov, V.L., Arlazarov, V.V., Bulatov, K.B., & et al. (2022). Mobile ID Document Recognition–Coarse-to-Fine Approach. Pattern Recognit, Image Anal, 32, 89–108 . doi: 10.1134/S1054661822010023.
2. Attivissimo, F., Giaquinto, N., Scarpetta, M., & Spadavecchia, M. (2019). An Automatic Reader of Identity Documents. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 3525-3530, doi: 10.1109/SMC.2019.8914438.
3. Das Neves, R. B., Felipe Vercosa, L., Macedo, D., Dantas Bezerra, B. L., & Zanchettin, C. (2020). A Fast Fully Octave Convolutional Neural Network for Document Image Segmentation. 2020 International Joint Conference on Neural Networks (IJCNN). doi: 10.1109/IJCNN48605.2020.9206711.
4. Augereau, O., Journet, N., & Domenger, J.-P. (2013). Semi-structured document image matching and recognition. Document Recognition and Retrieval XX. doi:10.1117/12.2003911.
5. Awal, A. M., Ghanmi, N., Sicre, R., & Furon, T. (2017). Complex Document Classification and Localization Application on Identity Document Images. 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). doi:10.1109/icdar.2017.77.
6. Skoryukina, N., Arlazarov, V., & Nikolaev, D. (2019). Fast Method of ID Documents Location and Type Identification for Mobile and Server Application. 2019 International Conference on Document Analysis and Recognition (ICDAR). doi:10.1109/icdar.2019.00141.
7. Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding, 110(3), 346–359. doi:10.1016/j.cviu.2007.09.014.
8. Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 60(2), 91–110. doi:10.1023/B:VISI.0000029664.99615.94.
9. Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), 381–395. doi:10.1145/358669.358692.
10. Bin Fan, Qingqun Kong, Trzcinski, T., Zhiheng Wang, Chunhong Pan, & Fua, P. (2014). Receptive Fields Selection for Binary Feature Description. IEEE Transactions on Image Processing, 23(6), 2583–2595. doi:10.1109/TIP.2014.2317981.
11. Dalal, N., & Triggs, B. (n.d.). Histograms of Oriented Gradients for Human Detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05). doi:10.1109/CVPR.2005.177
12. Suárez, I., Sfeir, G., Buenaposada, J. M., & Baumela, L. (2020). BEBLID: Boosted Efficient Binary Local Image Descriptor. Pattern Recognition Letters. doi:10.1016/j.patrec.2020.04.005.
13. Bay, H., Ferraris, V., & Van Gool, L.(2005). Wide-Baseline Stereo Matching with Line Segments. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1, 329-336. doi:10.1109/CVPR.2005.375.
14. Muja, M., & Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. VISAPP (1), 2(331-340), 2. doi:10.5220/0001787803310340
15. Raguram, R., Chum, O., Pollefeys, M., Matas, J., & Frahm, J.-M. (2013). USAC: A Universal Framework for Random Sample Consensus. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 2022–2038. doi:10.1109/TPAMI.2012.257
16. Skoryukina, N., Faradjev, I., Bulatov, K., & Arlazarov, V.V. (2020). Impact of geometrical restrictions in RANSAC sampling on the ID document classification. Twelfth International Conference on Machine Vision (ICMV 2019), 1143306. doi:10.1117/12.2559306
17. Chiron, G., Ghanmi, N., & Awal, A. M. (2021). ID documents matching and localization with multi-hypothesis constraints. 2020 25th International Conference on Pattern Recognition (ICPR). doi:10.1109/ICPR48806.2021.9412437.
18. Chiron, G., Ghanmi, N., & Awal, A. M. (2021). ID documents matching and localization with multi-hypothesis constraints. 2020 25th International Conference on Pattern Recognition (ICPR). doi:10.1109/ICPR48806.2021.9412437.
19. Rusiñol, M., & Lladós, J. (2009). Logo Spotting by a Bag-of-words Approach for Document Categorization. 2009 10th International Conference on Document Analysis and Recognition. doi:10.1109/ICDAR.2009.103.
20. Arlazarov, V. V., Bulatov, K., Chernov, T., & Arlazarov, V. L. (2019). MIDV-500: A Dataset for Identity Document Analysis and Recognition on Mobile Devices in Video

Stream. Computer Optics, 43(5), 818–824. doi:10.18287/2412-6179-2019-43-5-818-824.
21. Chernyshova, Y., Emelianova, E., Sheshkus, A., & Arlazarov, V.V. (2021). MIDV-LAIT: a challenging dataset for recognition of IDs with Perso-Arabic, Thai, and Indian scripts. The 16th International Conference on Document Analysis and Recognition (ICDAR). doi:10.1007/978-3-030-86331-9_17.
22. Trzcinski, T., Christoudias, M., Fua, P., & Lepetit, V. (2013). Boosting Binary Keypoint Descriptors. 2013 IEEE Conference on Computer Vision and Pattern Recognition. doi:10.1109/CVPR.2013.370.
23. Sheshkus, A., Chirvonaya, A., & Arlazarov, V.L.(2022). Tiny CNN for feature point description for document analysis: approach and dataset. Computer Optics, 46(3), 429-435. doi:10.18287/2412-6179-CO-1016.

Natalya Sergeevna S. Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia. Researcher. Research field: image analysis, computer vision. E-mail: skleppy.inc@smartengines.com

Evgeniia Aleksandrovna S. Smart Engines Service LLC, Moscow, Russia. Researcher. Research field: image analysis, applied mathematics. E-mail: e.shalnova@smartengines.com

Vladimir Viktorovich A. Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia. PhD-postDoc. Head of the laboratory.Research field: Artificial intelligence, machine learning, recognition system, information technology, queueing theory. E-mail: vva777@gmail.com

Метод выявления ложных ответов алгоритмов локализации и идентификации с помощью глобальных особенностей

Н. С. Скорюкина[I,II], Е. А. Шальнова[I], В. В. Арлазаров[I,I]

[I] Смарт Энджинс Сервис, Москва, Россия
[II] Федеральный исследовательский центр «Информатика и управление» РАН, Москва, Россия

Аннотация. В данной работе представлен метод отсечения ложных ответов алгоритмов локализации и идентификации на изображениях. Метод основан на сопоставлении характерных особенностей изображений, которые неустойчиво либо неполно описываются локальными признаками. Предлагается выделять зоны изображений, содержащие такие особенности, рассчитывать их компактную форму (дескриптор) и использовать для оценки валидности ответа алгоритма. В работе демонстрируется работа алгоритма на примере ID документов. В качестве особенностей рассматриваем изображения гербов и флагов стран, фоновое заполнение и текст, присущие именно этому типу документа. Тестирование проведено на наборах данных MIDV-500 и MIDV-LAIT. MIDV-500 использован в качестве положительной выборки (реджектор не должен отклонять правильные ответы системы), MIDV-LAIT - в качестве негативной выборки. Протестированы различные методы дескрипции зон. Результаты эксперимента показывают, что число ложных типизаций снижается при любом способе дескрипции, а локальный CNN-дескриптор показывает лучший результат. Также показано, что увеличение количества классов с выделенными зонами улучшает фильтрацию ложных срабатываний. На экспериментальных данных показано улучшение от ~13% при 1 типе с зонами до 4 раз при 10 типах.

Ключевые слова: отклонение кандидатов (реджектор), ключевые точки, локализация, идентификация.

DOI 10.14357/20718632230403 **EDN** CGRAFY

Литература

1. Arlazarov, V.L., Arlazarov, V.V., Bulatov, K.B., & et al. (2022). Mobile ID Document Recognition–Coarse-to-Fine Approach. Pattern Recognit, Image Anal, 32, 89–108 . doi: 10.1134/S1054661822010023.
2. Attivissimo, F., Giaquinto, N., Scarpetta, M., & Spadavecchia, M. (2019). An Automatic Reader of Identity Documents. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 3525-3530, doi: 10.1109/SMC.2019.8914438.
3. Das Neves, R. B., Felipe Vercosa, L., Macedo, D., Dantas Bezerra, B. L., & Zanchettin, C. (2020). A Fast Fully Octave Convolutional Neural Network for Document Image Segmentation. 2020 International Joint Conference on Neural Networks (IJCNN). doi: 10.1109/IJCNN48605.2020.9206711.

4. Augereau, O., Journet, N., & Domenger, J.-P. (2013). Semi-structured document image matching and recognition. Document Recognition and Retrieval XX. doi:10.1117/12.2003911.
5. Awal, A. M., Ghanmi, N., Sicre, R., & Furon, T. (2017). Complex Document Classification and Localization Application on Identity Document Images. 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). doi:10.1109/icdar.2017.77.
6. Skoryukina, N., Arlazarov, V., & Nikolaev, D. (2019). Fast Method of ID Documents Location and Type Identification for Mobile and Server Application. 2019 International Conference on Document Analysis and Recognition (ICDAR). doi:10.1109/icdar.2019.00141.
7. Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding, 110(3), 346–359. doi:10.1016/j.cviu.2007.09.014.
8. Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 60(2), 91–110. doi:10.1023/B:VISI.0000029664.99615.94.
9. Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), 381–395. doi:10.1145/358669.358692.
10. Bin Fan, Qingqun Kong, Trzcinski, T., Zhiheng Wang, Chunhong Pan, & Fua, P. (2014). Receptive Fields Selection for Binary Feature Description. IEEE Transactions on Image Processing, 23(6), 2583–2595. doi:10.1109/TIP.2014.2317981.
11. Dalal, N., & Triggs, B. (n.d.). Histograms of Oriented Gradients for Human Detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). doi:10.1109/CVPR.2005.177.
12. Suárez, I., Sfeir, G., Buenaposada, J. M., & Baumela, L. (2020). BEBLID: Boosted Efficient Binary Local Image Descriptor. Pattern Recognition Letters. doi:10.1016/j.patrec.2020.04.005.
13. Bay, H., Ferraris, V., & Van Gool, L.(2005). Wide-Baseline Stereo Matching with Line Segments. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1, 329-336. doi:10.1109/CVPR.2005.375.
14. Muja, M., & Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. VISAPP (1), 2(331-340), 2. doi:10.5220/0001787803310340
15. Raguram, R., Chum, O., Pollefeys, M., Matas, J., & Frahm, J.-M. (2013). USAC: A Universal Framework for Random Sample Consensus. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 2022–2038. doi:10.1109/TPAMI.2012.257.
16. Skoryukina, N., Faradjev, I., Bulatov, K., & Arlazarov, V.V. (2020). Impact of geometrical restrictions in RANSAC sampling on the ID document classification. Twelfth International Conference on Machine Vision (ICMV 2019), 1143306. doi:10.1117/12.2559306.
17. Chiron, G., Ghanmi, N., & Awal, A. M. (2021). ID documents matching and localization with multi-hypothesis constraints. 2020 25th International Conference on Pattern Recognition (ICPR). doi:10.1109/ICPR48806.2021.9412437.
18. Chiron, G., Ghanmi, N., & Awal, A. M. (2021). ID documents matching and localization with multi-hypothesis constraints. 2020 25th International Conference on Pattern Recognition (ICPR). doi:10.1109/ICPR48806.2021.9412437.
19. Rusiñol, M., & Lladós, J. (2009). Logo Spotting by a Bag-of-words Approach for Document Categorization. 2009 10th International Conference on Document Analysis and Recognition. doi:10.1109/ICDAR.2009.103.
20. Arlazarov, V. V., Bulatov, K., Chernov, T., & Arlazarov, V. L. (2019). MIDV-500: A Dataset for Identity Document Analysis and Recognition on Mobile Devices in Video Stream. Computer Optics, 43(5), 818–824. doi:10.18287/2412-6179-2019-43-5-818-824.
21. Chernyshova, Y., Emelianova, E., Sheshkus, A., & Arlazarov, V.V. (2021). MIDV-LAIT: a challenging dataset for recognition of IDs with Perso-Arabic, Thai, and Indian scripts. The 16th International Conference on Document Analysis and Recognition (ICDAR). doi:10.1007/978-3-030-86331-9_17.
22. Trzcinski, T., Christoudias, M., Fua, P., & Lepetit, V. (2013). Boosting Binary Keypoint Descriptors. 2013 IEEE Conference on Computer Vision and Pattern Recognition. doi:10.1109/CVPR.2013.370.
23. Sheshkus, A., Chirvonaya, A., & Arlazarov, V.L.(2022). Tiny CNN for feature point description for document analysis: approach and dataset. Computer Optics, 46(3), 429-435. doi:10.18287/2412-6179-CO-1016.

Наталья Сергеевна Скорюкина Федеральный исследовательский центр «Информатика и управление» РАН, Москва, Россия. Научный сотрудник. Область научных интересов: анализ изображений, компьютерное зрение.
E-mail: skleppy.inc@smartengines.com

Евгения Александровна Шальнова. "Смарт Энджинс Сервис", Москва, Россия. Научный сотрудник. Область научных интересов: анализ изображений, прикладная математика. E-mail: e.shalnova@smartengines.com

Владимир Викторович Арлазаров. Федеральный исследовательский центр «Информатика и управление» РАН, Москва, Россия. Д.т.н. Заведующий отделом. Область научных интересов: Искусственный интеллект, машинное обучение, системы распознавания, информационные технологии, теория массового обслуживания. E-mail: vva777@gmail.com