Problems of identification in information systems

UDC 004.93 **doi:** https://doi.org/10.20998/2522-9052.2024.2.01

Oleksii Gorokhovatskyi[1], Olena Yakovleva[2,3]

[1] Simon Kuznets Kharkiv National University of Economics, Kharkiv, Ukraine
[2] Bratislava University of Economics and Management, Bratislava, Slovakia
[3] Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

MEDOIDS AS A PACKING OF ORB IMAGE DESCRIPTORS

Abstract. The aim of the research. The paper presents the research about the feasibility to use matching medoids obtained from the set of ORB descriptors instead matching the full set of binary descriptors for image classification problem. **Research results.** Different methods that include direct brute force medoids matching, grouping of medoids for separate classes, and grouping of descriptors followed by calculation of medoids amongst them were proposed. Numerical experiments were performed for all these methods in order to compare the classification accuracy and inference time. It has been shown that using of medoids allowed us to redistribute processing time in order to perform more calculations during preprocessing rather than during classification. According to modelling performed on the Leeds Butterly dataset matching images based on medoids could have the same accuracy as matching of descriptors (0.69–0.88 for different number of features). Medoids require additional time for the calculation during preprocessing stage but classification time becomes faster: in our experiments we have obtained about 9–10 times faster classification and same 9–10 times increasing preprocessing time for the models that have comparable accuracies. Finally, the efficiency of the proposed ideas was compared to the CNN trained and evaluated on the same data. As expected, CNN required much more preprocessing (training) time but the result is worth it: this approach provides the best classification accuracy and inference time. **Conclusion**. Medoids matching could have the same accuracy as direct descriptors matching, but the usage of medoids allows us to redistribute the overall modeling time with the increasing preprocessing time and making inference faster.

Keywords: Image features; Keypoints; Medoids; Classification; ORB; Binary descriptors; Features matching; Grouping; Bag of features; Repeatability; Classification accuracy.

Introduction

The solution for various image classification problems last decade shifted significantly from the methods based on the analysis of explicit structural features (descriptors) of the image to the application of artificial neural networks (ANN), particularly, convolutional (CNN) ones. Despite obvious advantages of CNN approaches that include classification accuracy and inference time the artificial neural networks still require a lot of labeled training data, which can be a limitation for the problems with specific datasets. Additionally, ANN have black-box nature with a lack of interpretability and transparency, which makes the reasoning of classification results complex. ANN can also overfit and fail to generalize to unseen data.

On the other hand, the structural descriptors do not require a training phase and a lot of training labeled data, but the implementation of descriptor-based methods often requires some fast storage/retrieving/filtering of descriptors being compared. One of the questions arising here is about the way to select the best representatives of descriptors instead of comparing all of them to make method work faster but preserving good accuracy.

This experiment-driven research sheds some light on the possible improvement of ORB image descriptors comparison with the usage of medoids instead of descriptors, that could achieve better classification time with worser accuracy.

This paper is organized as follows. Section 1 presents the literature review of similar approaches and methods. Section 2 describes the motivation to use ORB descriptors for this paper. Section 3 presents the descriptors matching procedure. Section 4 shows the basic properties of medoids, including calculation, the dataset used in the paper and the repeatability of medoids matching compared to descriptors matching. Section 5 describes different approaches based on medoids for image classification problem. Finally, the last section is the conclusions.

1. Related work

The usage of feature descriptors for solving image classification problems is well-known. To classify images based on descriptors matching, we need to extract features (which can be thousands per image), store them, and perform matching. These stages typically require a lot of computation time and/or memory resources [1]. To overcome these issues, various solutions have been proposed, including the usage of binary descriptors (e.g., BRISK, ORB) instead of float ones (SURF, SIFT) and packing (aggregation) the entire set of image descriptors into smaller quantities. Different aggregation approaches have been shown to be effective under various conditions [1-4]. It is worth noting that a simple k-means clustering procedure could be used to quantize descriptors, build a vocabulary of visual words (Bag of Words, BOW), or create some cluster representations of the descriptor sets to compare them more effectively, thus reducing computational efforts.

The usage of ORB descriptors for image classification BOW model with k-majority clustering was proposed in [5]. It is mentioned that k-medoids algorithm requires the computation of the entire distance matrix that is very slow (including that BOW is often built for the thousands of words).

Various implementations of descriptors aggregation following artificial neural networking classification are also known. The approach based on the bag of features build from image SIFT keypoint descriptors following by k-means clustering and SVM classification was introduced in [6]. The numerical results in this research showed that selection of quantity of words in the dictionary for BOW model should be a responsible choice as its increasing could make results worser. Classification of ORB descriptors without any aggregation using SVM and KNN directly was presented in [7].

The are also examples of other practical applications that include Bag of Visual Words using KAZE feature descriptors [8] following SVM to classify proteins (it was shown numerically that the accuracy with increasing the quantity of words leads plateau at some point), and the usage of Harris and KAZE features with SVM classification approach to perform human identification by ears [9].

The aggregation of image descriptors could be considered as a more common selection of representative features problem. The application of principal component analysis (PCA) to ORB and other features to reduce the dimensionality of descriptors and improve their quality has been proposed in [10]. It was also confirmed that ORB-PCA descriptors can provide better classification quality [11]. The other way is to improve the quality of descriptors, e.g., the using of the neural networks were proposed to improve the discrimination ability of different descriptor types [12].

Our previous researches about the clustering of descriptors with further classification over the cluster distributions has been proposed in [13, 14], the clustering following hashing procedure was presented in [15]. No medoids as cluster center were involved there, only few (4-5) clusters were used. The current paper follows the same common pattern for the matching the request image with train (etalon) images when the major calculations are performed at once during data preprocessing stage that allows us to get performance improvement. The main drawback of previous research is the lack of experiments, the main idea and application prototype were tested only, making it difficult to understand whether the common idea is scalable. Our other research [16] about using image feature descriptors for the detection of near-duplicate images helped us to understand some nuances for the usage of descriptors in practical applications.

As a summary it could be concluded that the usage of clustering (and medoids) is a common choice for the processing of descriptors with BOW models. The key idea for BOW approach is the association of each descriptor with some cluster that allows us to build a sort of image signature in such a way. The usage of medoids just instead of descriptors preserving further typical descriptors matching pipeline seems to be not investigated good enough. It is commonly clear that comparing less quantity of medoids instead of more descriptors should be not so good in terms of quality but the value of this difference is unknown. Is this quality good enough to implement classification models in practice? What is the difference for classification time using medoids and comparing all descriptors? Are the medoids robust enough for such classification applications? Finally, how this approach compares (in terms of quality firstly) to the usage of CNN?

The contribution of the paper includes the research about the feasibility to use less quantity of medoids instead of the full set of descriptors, redistribute the preprocessing and classification time to make image classification process faster, and the measurement of the possible decrease of the performance and comparison to CNN.

2. Image descriptors

The searching for the image features as a form of descriptors usually includes two stages: keypoint detection, and the construction of descriptors for them. Keypoints are points of interest in the image that are likely to be stable under changes in illumination, viewpoint, and other image transformations. The descriptor is a vector of numbers that represents the local features of an image around a keypoint. Descriptors are used to match keypoints between different images. Common keypoint detection methods are Harris corner detector and FAST algorithm. Famous methods for the descriptors detections are SIFT, SIRF, BRIIF, ORB, BRISK, FREAK and others.

In this research we use ORB (Oriented FAST and Rotated BRIEF) image descriptors, that was proposed in [17] as an alternative to previous known SIFT and SURF methods. This feature detector builds descriptor in binary form. It is a combination of the FAST keypoint detector and the BRIEF descriptor, with some added features to improve performance. ORB is robust to noise (but is sensitive to blur), illumination changes, and viewpoint changes. It is also invariant to rotation, which makes it useful for tasks such as image matching and object recognition. The length of ORB descriptor is 256 bits. The descriptors are usually compared using Hamming distance and some quantity of mismatched bits in successfully matched descriptors is possible to exist.

3. Matching (comparison) algorithm

Descriptors in this paper are compared using the brute force approach with cross check option. Such matching is still very popular in practice and requires almost no parameters to be set up despite more modern comparison methods exist. The main drawback of this approach is the processing time.

Distance (Hamming) matrix between all combinations of binary ORB descriptors is used following by filtering to leave only those columns and rows (possible good matches) having distance less that threshold (31 mismatching bits between two descriptors are allowed at most). All descriptors left in the matrix are filtered once more to find good matches: the match between i-th and j-th descriptors is considered to be good only when the distance between i-th and j-th is minimal and vice versa.

4. Basics properties of medoids

4.1. Medoids calculation

A medoid is a representative object within a data set or cluster that is centrally located to all other members of the set. It is similar to a centroid or mean, but unlike those, a medoid is always an actual data point from the set itself, rather than a calculated value.

Let $X = \{x_1, x_2, ..., x_n\}$ be the set of ORB descriptors found in the image. Medoid is such a representative of the set X that:

$$x_m \in X : x_m = \arg\min_{y \in X} \sum_{i=1}^{n} d(x_i, y)$$

where d is the Hamming distance function used to compare two ORB descriptors. In this paper we use M medoids – those are the first M representatives of set X having the minimal distance to all other set items $x_i, i = \overline{1,n}$. The effective quantity of medoids is the subject of our interest in this paper as well.

Calculation of medoids is based on computing of distance matrix for each pair of descriptors following by summation of distances over columns and selection of required quantity of descriptors with minimal sum of distances.

The medoids in the paper are used as the way to choose robust representatives of the descriptor set only, without following the full clustering procedure.

4.2. Dataset

All experiments in this paper were performed using Leeds Butterfly dataset [18, 19] that contains ten classes of butterflies having from 55 to 100 images per class (832 images in total). Each image has segmentation mask that allows us to distinguish butterfly from the background. We performed modeling using the foreground part of the image only to be sure that we compare descriptors of the butterfly not the scene around it.

4.3. Repeatability

The measure to evaluate the quality of medoids representation instead of descriptors we used is the repeatability r [20-23]. It is the quantity of successfully matched descriptors/medoids for the original image I and its geometrically transformed version I^t:

$$r = \frac{\#\text{of matches between descriptors in } I \text{ and } I^t}{\min(\#of \text{ descriptors in } I,\ \#\text{of descriptors in } I^t)}$$

Rotation transformations were used in experiments as ORB descriptors are stable to them [17]. Repeatability values were averaged by rotation angle for all tuples of original and transformed image of the dataset.

Our intuition behind using of medoids instead of descriptors comes from Fig. 1. There are 5 curves for different descriptor sets and their mean repeatability values under rotation angles (the value is r=1 for 0 and 360 degrees as expected). The base (reference) curve here is the repeatability for 500 ORB descriptors, that shows the best repeatability. Corresponding values values are less for 120 medoids which were selected from these 500 descriptors, and even less for 20 medoids only. There is a significant difference between curves built for 20 medoids and first 20 descriptors that confirms that usage of medoids instead of descriptors matters. Finally, the subset from 20 medoids uniformly distributed across all 500 original descriptors has the worst repeatability values. This gives the idea to use few medoids which are found sequentially one by one instead of skipping some of them in order to cover the entire set of descriptors uniformly.

The time required to compare descriptors (including the time for calculating medoids) is presented if Fig. 2. As can be seen the matching medoids is approximately 5 times slower compared to matching descriptors, but there is no essential difference between calculation of 20 and 120 medoids.

5. Classification using medoids: accuracy and performance

During the classification based on medoids we are interested primarily in the level of accuracy decreasing comparing usage of medoids and descriptors, as well as the classification (inference) time.

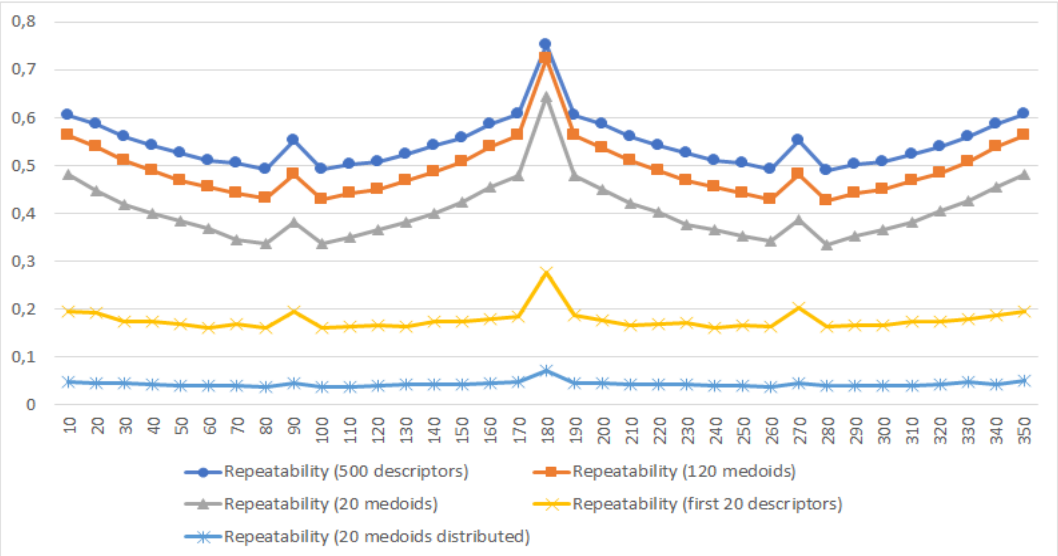

Fig. 1. The average repeatability for sets of descriptors under rotations

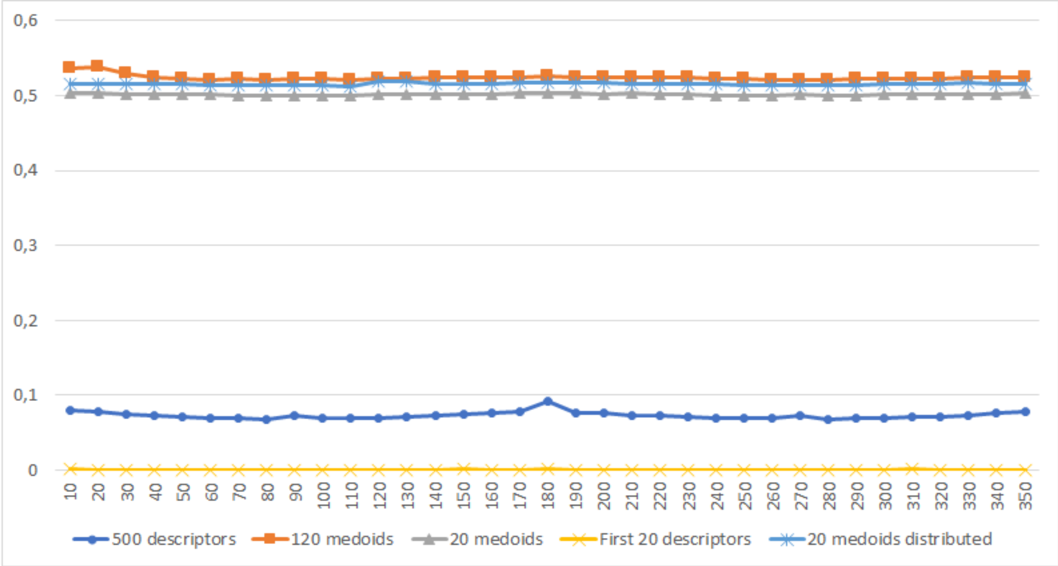

Fig. 2. The average sets of descriptors comparison time

We used 80% of the dataset (65 images for class 001, 75 for class 002, 48 for 003, 72 for 004, 70 for 005, 80 for 006, 71 for 007, 44 for 008, 72 for 009, and 67 for 010, 663 images in total) as training set in all classification experiments grabbing first 80% of images for each existing class, other 20% were used as test set. All images were used without background, i.e., after applying ground truth segmentation mask to consider the part of the image containing butterfly only.

5.1. Direct comparison of descriptors

This is the baseline method to compare other with: ORB descriptors for each image in the training set are gathered and stored separately per image, the classification stage is the iteration over all sets of descriptors in the training dataset comparing each with the set of descriptors for the image being classified. The class of the training item having the maximum quantity of matched descriptors is considered as output (Fig. 3).

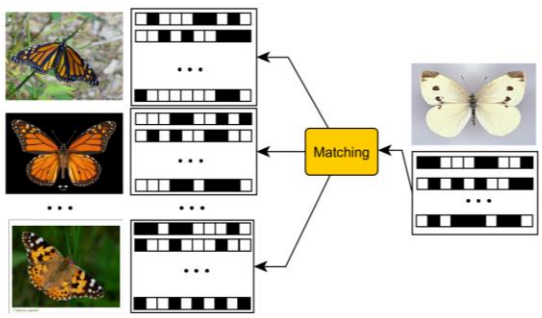

Fig. 3. Matching descriptors

The full comparison between each pair of descriptors is performed (brute force) with the cross check of minimal distance and the maximum deviation between descriptors was allowed to be 31 bits.

The matching 500 ORB descriptors in such a way allows us to reach 0.8757 accuracy, comparing 120 descriptors – 0.7633, 80 descriptors – 0.6923. We have tried to achieve better accuracy with more descriptors and got 0.8521 accuracy for 600 ones, 0.8698 – for 750, 0.8402 – for 1000 descriptors, so it seems the possibilities of direct descriptors matching are limited. The quantity of descriptors to be find in images are not guaranteed so we controlled the average factual number of descriptors: 476 for 500 requested, 567 – for 600, 701 – for 750, and 915 – for 1000 though there were few images with limited number of descriptors. So, we refer to 500 ORB descriptors as baseline experiment further.

5.2. Matching medoids

The second classification approach is based on the simple usage of medoids instead of descriptors according to the same matching procedure (Fig. 4).

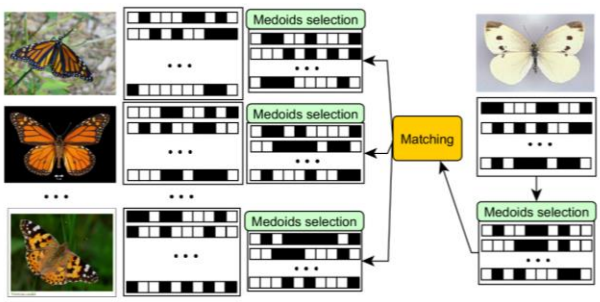

Fig. 4. Matching medoids

There were 500 ORB descriptors in these experiments combined with the selection of 80, 120 or 300 medoids amongst them. Using 80 medoids we achieved 0.7456 accuracy, for 120 medoids the accuracy was 0.7574, finally, the best result is 0.8402 using 300 medoids. As one can see, the matching 120 medoids selected from 500 initial descriptors allows us to achieve nearly the same quality as the usage of 120 initial ORB descriptors (0.7574 vs 0.7633). But the classification time is faster (more than 10 times) for this approach due to the ability to perform medoids selection at the preliminary stage during the processing of training set. On the other hand, the time required for this case is approximately 9 times longer compared to using direct descriptors matching.

The increasing of quantity of medoids does not necessarily improves the accuracy. We tested also different combinations of number of ORB descriptors (600, 750, 1000) and number of medoids calculated for them (200, 250, 300, 400). The best result we achieved is 0.8402 with 300 medoids selected from 500 descriptors.

We additionally verified that the choosing of medoids rather than just random descriptors matter for this experiment: five additional tests were performed with the selection of 80 random ORB descriptors (instead of medoids) from the initial 500 ones. The classification accuracy for these were in the range from 0.44 to 0.54 that is significantly less than for 80 medoids (approximately 0.75).

5.3. Matching medoids on the bag of descriptors

The next classification method includes the calculation of medoids for the bag of descriptors: all descriptors for train images of class 001 are grouped together, medoids are found amongst them, the same for class 002 and so on (Fig. 5). Hyperparameters that can be used here include the quantity of medoids to calculate from the entire set of descriptors and quantity of medoids from each image to use during classification.

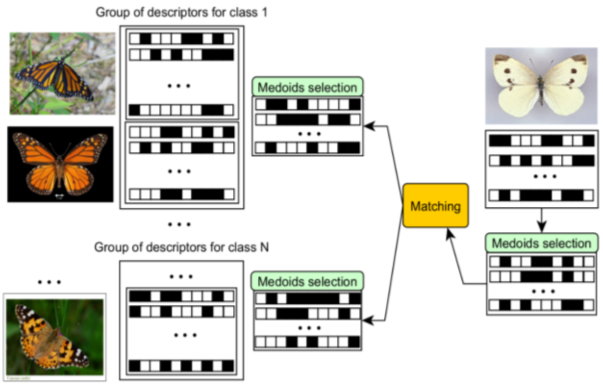

Fig. 5. Matching medoids on the bag of descriptors

The entire number of descriptors detected in all train images (assuming we search for 500 ORB features in each) is about 315 thousand. We tested 10, 20 and 30 thousand of medoids selected from them combined with 80, 120 and 300 medoids used from each test image at classification stage. The best accuracy we got is 0.8876 for 20 thousand of medoids and 300 medoids from each test image. The time required to calculate medoids from the bag of descriptors is almost the same for 10, 20 and 30 thousand, classification time increased linearly.

5.4. Matching bags of medoids

The last approach includes the building of bag of medoids: all medoids for the training images of the class 001 were grouped together, all medoids for training images representing class 002 are grouped also and so on (Fig. 6). Two hyperparameters could be considered here: the quantity of medoids used for train and test images. We tried different combinations of 150, 200 and 300 medoids for the training set and 80, 120 and 300 for the testing one preserving the initial number of 500 ORB descriptors.

It is possible to get classification accuracy 0.8817 using 300 medoids for each train and test image.

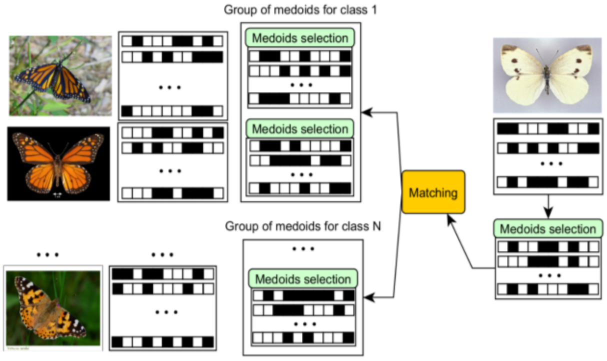

Fig. 6. Bag of medoids

Finally, all these methods were compared to the convolution neural network, that was trained on the same training set. The architecture of this network included convolutional, maxpooling and dense layers. The first three layers are a 2D convolutional ones all having 64 filters size of 3x3 with RELU activations and valid convolution type, 2D maximal pooling layer size of 2x2 follows after each convolutional (strides are always 1). There is another sequence of three convolution layers (32 filters) with maxpooling in between after that. Finally, two dense layers having 16 neurons (RELU activation) and 10 neurons (softmax activation) with the dropout layer (p=0.5) in between follow. We had to make an effort to build such an architecture as the networks with less quantity of layers reach only 0.83 – 0.9 classification accuracy. Training of the chosen structure of the CNN has been performed on the images rescaled to 256x256 pixels during 250 epochs with Adam optimizer.

The best performance and accuracy results for all approaches are shown in Table 1. First column shows the brief description of the method and its hyperparameters, the accuracy of the method is shown in the next column. We refer to preprocessing time as the time required to process train dataset, it includes calculation of descriptors, grouping of descriptors/medoids, calculation of medoids, CNN learning time. Ther last column includes the average inference time to classify one image from the test part of the dataset. We used the same hardware and conditions for all measurements.

As one can see it requires about 170 seconds to classify the input image using pure matching 500 ORB descriptors according to the policies and dataset described above having the total accuracy of this model to be 0.8757, we refer to this result as the baseline one. Matching less quantity of descriptors (120 and 80) is faster but has lower accuracy. On the other side, matching 300 medoids obtained from 500 initial ORB descriptors requires 14 seconds instead of 170 (with slightly worser quality) but calculation of medoids for the training set needs some time. Comparison of 120 medoids is close to comparison of 120 descriptors in terms of accuracy, and the comparison of 80 medoids reaches 0.7456 accuracy compared to 0.6923 for 80 descriptors preserving short inference time. The results confirm that the usage of medoids allows us to redistribute preprocessing and inference time to make classification faster. Bag models are able to achieve even slightly better accuracy (above

0.88) compared to the baseline experiment preserving the classification time at the same level as just comparison of medoids.

Finally, CNN requires much more time to train model but provides the best accuracy and instant inference as expected.

Table 1 – **Comparison of accuracies, preprocessing and the classification times**

Method	Accuracy	Prep. time, sec.	Inf. time, sec.
Matching 500 descriptors	0.8757	24	170
Matching 120 descriptors	0.7633	22	39
Matching 80 descriptors	0.6923	23	25
Matching 300 medoids (500 descriptors)	0.8402	200	14
Matching 120 medoids (500 descriptors)	0.7574	188	3.4
Matching 80 medoids (500 descriptors)	0.7456	188	2.2
Matching 20k medoids calculated on the bag of descriptors (500 descriptors) and 300 medoids from test image	0.8876	1907	14
Matching bag of medoids (300 medoids from each train image, 500 descriptors) and 300 medoids from test image	0.8817	196	14
Convolutional neural network	0.9527	10000	0,1

Conclusions

The paper presents the research about the feasibility to use medoids obtained from the set of ORB descriptors instead of the full set of descriptors to solve image classification problem. Using medoids allows us to redistribute processing time in order to perform more calculations during preprocessing rather than during classification. Different methods with grouping of medoids or without are considered: the direct comparison of medoids instead of descriptors, grouping of descriptors following by the calculation of medoids for the group, grouping of medoids. The measurement of classification accuracy and classification inference were performed.

Numerical experiments showed that matching medoids could have nearly the same accuracy as matching descriptors. At the same time, medoids require additional time for the calculation during preprocessing stage but classification time becomes faster. In our particular experiments with Leeds Butterfly dataset, we have obtained about 9-10 times faster classification and same 9-10 times increasing preprocessing time for the models that have comparable accuracies.

Finally, the efficiency of the proposed ideas was compared to the convolutional neural network trained and evaluated on the same data. As expected, CNN required much more preprocessing (training) time but the result is worth it: the CNN provides the best classification accuracy and inference time.

Acknowledgements

The work is funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under project No. 09I03-03-V01-00115.

We dedicate this paper to the Armed Forces of Ukraine and everyone who stands with Ukraine.

REFERENCES

1. Amato, G., Falchi, F. and Vadicamo, L. (2018), "Aggregating binary local descriptors for image retrieval", *Multimedia Tools and Applications*, vol. 77, pp. 5385–5415, doi: https://doi.org/10.1007/s11042-017-4450-2
2. Jégou, H., Perronnin, F., Douze, M. Sánchez, J., Pérez, P. and Schmid, C. (2012), "Aggregating Local Image Descriptors into Compact Codes", *IEEE Transactions on Pattern Analysis and Machine Intelligence"*, vol. 34(9), pp. 1704–1716, doi: https://doi.org/10.1109/TPAMI.2011.235
3. Jégou, H., Douze, M., Schmid, C. and Pérez, P. (2010), "Aggregating local descriptors into a compact image representation", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San *Francisco*, CA, USA, pp. 3304–3311, doi: https://doi.org/10.1109/CVPR.2010.5540039
4. Perronnin, F. and Dance, C. (2007), "Fisher kernels on visual vocabularies for image categorization," *Computer Vision and Pattern Recognition*, pp. 1–8, doi: https://doi.org/10.1109/CVPR.2007.383266
5. Grana, C., Borghesani, D., Manfredi, M. and Cucchiara, R. (2013), "A fast approach for integrating ORB descriptors in the bag of words model", *Multimedia Content and Mobile Devices*, vol. 8667, doi: https://doi.org/10.1117/12.2008460
6. Korytkowski, M., Scherer, R., Staszewski, P. and Woldan, P. (2015), "Bag-of-features image indexing and classification in microsoft SQL server relational database", *2015 IEEE 2nd International Conference on Cybernetics (CYBCONF)*, Gdynia, Poland, 2015, pp. 478–482, doi: https://doi.org/10.1109/CYBConf.2015.7175981
7. Rani, R., Kumar, R. and Singh, A. P. (2019), "Implementation of ORB and Object Classification using KNN and SVM Classifiers", *Int. Journal of Computer Sciences and Engineering*, vol. 7(3), pp. 280–285, doi: https://doi.org/10.26438/ijcse/v7i3.280285
8. Pražnikar, J. and Attygalle, N. (2022), "Quantitative analysis of visual codewords of a protein distance matrix", *PLoS One*, vol. 17(2), doi: https://doi.org/10.1371/journal.pone.0263566
9. Sawant, M. S. and Rawat, C. S. (2020), "Feature Detection using KAZE and Harris Detectors for Ear Biometrics", *Int. Journal on Engineering Research and Technology*, vol. 9(12), pp. 93–97, doi: https://doi.org/10.17577/IJERTV9IS120050
10. Zhu, J., Gong, C., M. Zhao, Wang, L. and Luo, Y. (2020), "Image mosaic algorithm based on PCA-ORB feature matching", *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-3-W10, pp. 83–89, doi: https://doi.org/10.5194/isprs-archives-XLII-3-W10-83-2020
11. Vinay, A., Kumar, Akshay C., Shenoy, Gaurav R., Murthy, K. N. Balasubramanaya and Natarajan, S. (2015), "ORB-PCA based feature extraction technique for face recognition", *Procedia Computer Science*, vol. 58, pp. 614–621, doi: https://doi.org/10.1016/j.procs.2015.08.080

12. Wang X., Liu Z., Hu Y., Y, Xi, W., Yu, W. and Zou D. (2023), “Feature Booster: Boosting Feature Descriptors with a Lightweight Neural Network”, *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7630–7639, doi: https://doi.org/10.48550/arXiv.2211.15069
13. Gorokhovatskyi, V., Tvoroshenko, I., Kobylin, O. and Vlasenko, N. (2023), “Search for Visual Objects by Request in the Form of a Cluster Representation for the Structural Image Description”, *Advances in Electrical and Electronic Engineering*, vol. 21(1), pp. 19–27, doi: https://doi.org/10.15598/aeee.v21i1.4661
14. Daradkeh, Y. I., Gorokhovatskyi, V. I., Tvoroshenko, I. and Zeghid, M. (2022), “Cluster representation of the structural description of images for effective classification”, *Computers, Materials & Continua*, vol. 73(3), 2022, pp. 6069–6084, doi: https://doi.org/10.32604/cmc.2022.030254
15. Daradkeh, Y. I., Gorokhovatskyi, V., Tvoroshenko, I. and Zeghid, M. (2022), “Tools for Fast Metric Data Search in Structural Methods for Image Classification”, *IEEE Access*, vol. 10, 2022, pp. 124738-124746, doi: https://doi.org/10.1109/ACCESS.2022.3225077
16. Gorokhovatskyi, O. and Peredrii, O. (2023), “Image Pair Comparison for Near-duplicates Detection”, *International Journal of Computing*, vol. 22(1), pp. 51–57, doi: https://doi.org/10.47839/ijc.22.1.2879
17. Rublee, E., Rabaud, V., Konolige, K. and Bradski, G. (2011), “ORB: An efficient alternative to SIFT or SURF”, *2011 Int. Conference on Computer Vision*, Barcelona, Spain, 2011, pp. 2564–2571, doi: https://doi.org/10.1109/ICCV.2011.6126544
18. Wang, J., Markert, K. and Everingham, M. (2009), “Learning Models for Object Recognition from Natural Language Descriptions”, *Proc. of the 20th British Machine Vision Conf. (BMVC2009)*, pp. 1–11, doi: https://doi.org/10.5244/C.23.2
19. Wang, J., Markert, K. and Everingham, M. (2023), *Leeds Butterfly DS*, available at: https://www.josiahwang.com/dataset/leedsbutterfly
20. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. and Van Gool, L. (2005), “A Comparison of Affine Region Detectors”, *IJCV*, vol. 65, pp. 43–72, doi: https://doi.org/10.1007/s11263-005-3848-x
21. Mouats, T., Aouf, N., Nam, D. and Vidas, S. (2018), “Performance Evaluation of Feature Detectors and Descriptors Beyond the Visible”, *Journal of Intelligent and Robotic Systems*, vol. 92, pp. 33–63, doi: https://doi.org/10.1007/s10846-017-0762-8
22. Schmid, C., Mohr, R. and Bauckhage, C. (2000), “Evaluation of interest point detectors”, *International Journal of Computer Vision*, vol. 37(2), pp. 151–172, doi: https://doi.org/10.1023/A:1008199403446
23. Yakovleva, O. and Nikolaieva, K. (2020), “Research Of Descriptor Based Image Normalization And Comparative Analysis Of SURF, SIFT, BRISK, ORB, KAZE, AKAZE Descriptors”, *Advanced Information Systems*, vol. 4(4), pp. 89–101, doi: https://doi.org/10.20998/2522-9052.2020.4.13

Received (Надійшла) 17.12.2023
Accepted for publication (Прийнята до друку) 20.03.2024

ABOUT THE AUTHORS / ВІДОМОСТІ ПРО АВТОРІВ

Гороховатський Олексій Володимирович – кандидат технічних наук, доцент кафедри інформатики та комп’ютерної техніки, Харківський національний економічний університет імені Семена Кузнеця, Харків, Україна;
Oleksii Gorokhovatskyi – PhD, Associate Professor, Department of Informatics and Computer Engineering, Simon Kuznets Kharkiv National University of Economics, Kharkiv, Ukraine;
e-mail: oleksii.gorokhovatskyi@gmail.com; ORCID: https://orcid.org/0000-0003-3477-2132;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=23099879900.

Яковлева Олена Володимирівна – кандидат технічних наук, кафедра економіки та фінансів, Братиславський університет економіки та менеджменту, Братислава, Словаччина; доцент кафедри інформатики, Харківський національний університет радіоелектроніки, Харків, Україна;
Olena Yakovleva – PhD, Department of Economics and Finance, Bratislava University of Economics and Management, Bratislava, Slovakia; Associate Professor of Department of Informatics, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: olena.yakovleva@vsemba.sk; ORCID: https://orcid.org/0000-0002-6129-6146;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57214222033.

Медоїди як спосіб упаковки ORB-дескрипторів зображення

О. В. Гороховатський, О. В. Яковлева

Анотація. Мета дослідження. У статті представлено дослідження можливості використання співставлення медоїдів, отриманих із набору дескрипторів ORB, замість порівняння повного набору бінарних дескрипторів для задачі класифікації зображень. **Результати дослідження.** Було запропоновано різні методи, які включають пряме зіставлення медоїдів повним перебором, групування медоїдів для окремих класів і групування дескрипторів з наступним обчисленням медоїдів серед них. Чисельні експерименти були проведені для всіх цих методів, щоб порівняти точність та час класифікації. Було показано, що використання медоїдів дозволило перерозподілити час обробки таким чином, щоб виконати більше обчислень під час попередньої обробки, а не під час класифікації. Відповідно до моделювання, виконаного на наборі даних Leeds Butterly, порівняння зображень на основі медоїдів може мати таку ж точність, як і порівняння дескрипторів (0.69–0.88 для різної кількості дескрипторів). Обчислення медоїдів вимагає додаткового часу на етапі попередньої обробки, але час класифікації стає швидшим: у наших експериментах ми отримали приблизно в 9-10 разів швидшу класифікацію та в 9-10 разів збільшили час попередньої обробки для моделей, які мають порівнянну точність. Нарешті, ефективність запропонованих ідей було порівняно з навченими CNN та оцінено на тих самих даних. Як і очікувалося, CNN вимагало набагато більше часу на попередню обробку (навчання), але результат був того вартий: цей підхід забезпечує найкращі точність та час класифікації. **Висновки.** Зіставлення медоїдів може досягти таку саму точність, як і пряме зіставлення дескрипторів, але використання медоїдів дозволяє перерозподіляти загальний час моделювання збільшивши час попередньої обробки та прискоривши класифікацію.

Ключові слова: ознаки зображення; ключові точки; медоїди; класифікація; ORB; бінарні дескриптори; зіставлення ознак; групування; мішок ознак; повторюваність; точність класифікації.