Conformed estimates of histograms of oriented gradients

Kirill Pugachev
Samara National Research University
Samara, Russia
pugachev_k.g@mail.ru

***Abstract*—In this paper, we propose a new image matching algorithm based on the scene feature points matching. The algorithm is based on the well-known HOG descriptor, which divides the image into small areas and calculates histogram for each gradient direction. The idea is to use a proximity measure based on the conformity principle as a proximity criterion for the feature descriptors instead of the traditionally used Euclidean distance. The efficiency of this measure of proximity in image matching problem based on image fragments' intensity was researched in our previous papers. We use the ratio of correctly found feature points' number after cross-checking to the total number of correctly found feature points as a quality criterion of image matching results.**

Keywords—image matching, conformity principle, proximity measures, descriptors, histograms of oriented gradients

I. Introduction

The image matching task is to find the corresponding point for every point of the first image in the second image of the same scene. Image matching process consists of three main stages. Firstly, image preprocessing is performed, for example, using methods [1]. Then, images are matched. Finally, matching errors are corrected. Both direct and indirect matching methods are used.

Direct methods use values of image pixel intensity or its gradients directly. For example, area-based image matching methods are direct methods. In this case, image fragments with a center at this point are formed for every point of both images. To find corresponding points, search area in the second image is specified for every point of the first image. The corresponding point is the point in the search area, where the value of fragment proximity criterion is maximum or minimum.

Indirect methods [2–5] are based on the matching of the scene features, for example, feature points, object edges and etc.

One of the most commonly used features are SIFT-features [2]. These features have good invariance on rotation, scale scaling, brightness, and a certain degree of stability on changes in perspective, affine transformation, and noise.

There are many variations of this method. In paper [6] SIFT-method was evaluated in various color spaces. Evaluation of SIFT-like algorithm's performance was presented in paper [7].

Development of the SIFT-method is continued even in our days. So, new SIFT algorithm based on the adaptive contrast threshold was presented in paper [8]. The proposed algorithm has higher efficiency and accuracy. It realized the efficient control of a feature point number in multi-scene matching.

A new high-resolution optical-to-SAR image registration algorithm was presented in paper [9]. The experimental results of experiments show that the proposed OS-SIFT algorithm gives a robust registration result for optical-to-SAR images and outperforms other state-of-the-art algorithms in terms of registration accuracy.

The performance of image detection and matching using SIFT-method and RANSAC with Euclidean metric was researched in paper [10]. It was shown that the best results are archived using threshold value of 0.6. Evaluation of the SIFT-method in image recognition task using various proximity measures was presented in paper [11].

Improved SIFT-BRISK algorithm was proposed in paper [12]. This algorithm allows using the advantages of both algorithms. A distinctive and robust weighted local intensity binary SIFT descriptor (WLIB-SIFT) was proposed in paper [13].

To present these features, the descriptors are formed, and then image matching is performed using some measure of fragments proximity. Sum of absolute differences [14, 15], a sum of squared differences [16, 17], value normalized cross-correlation of fragments [18, 19] and etc. are widely used in both approaches as a criterion of proximity.

We have presented and implemented an area-based matching method, based on the principle of fragments conformity in papers [20, 21]. We have shown that conformity criterion allows performing more accurate image fragments matching in these papers. This is due to the fact that the method of conformed estimates performs many more comparisons than other proximity measures.

In this paper, we develop this line of research. Some scenes can contain similar objects. In this case, uniqueness of features can be very important. So, to improve it, we propose a new image matching criterion based on the conformity principle as a measure of proximity of the histogram of oriented gradients.

II. Problem Statement

Let us state that F_1 and F_2 are two images, which are obtained by multi-view shooting of the same scene. Let us also state that some features of the scene were detected on both images, for example, feature points.

To describe these features, we form descriptor vectors $\mathbf{f}_i^1$ and $\mathbf{f}_j^2$, $i=\overline{1,N}$, $j=\overline{1,M}$. N and M are count of feature points, which were found for first and second images correspondingly. Every descriptor contains K elements. The task is to find the corresponding point with descriptor $\mathbf{f}_j^2$ of the second image for every point with descriptor $\mathbf{f}_i^1$ of the first image.

As stated above, the sum of absolute differences or sum of squared differences of vector components are most often used during descriptor matching. These proximity measures are commonly used to compare elements, which present values of image pixel intensity or values of intensity gradient. Hamming distance is also used to compare binary descriptors, which are used, for example, to represent ORB-features.

Let us take a look at the problem of descriptor vectors matching using a proximity measure based on the principle of conformity. An important feature of this measure of proximity is that not only the comparison of descriptor vectors elements, but also all its possible combinations are performed. Wherein a big number of values is formed even with small fragment size. It allows increasing informativity of data and the probability of correct feature points matching.

Let's $\mathbf{f}_i^1$ and $\mathbf{f}_j^2$, $i=\overline{1,N},\ j=\overline{1,M}$, are two descriptor vectors of points of first and second images correspondingly. If every descriptor is a vector with K elements, conformed measure of proximity is defined as:

$$W_{i,j}=\sum_{\substack{s=1\\p=s+1}}^{K}(\Delta f_s^{i,j}-\Delta f_p^{i,j})^2, \qquad (1)$$

where $\Delta f_s^{i,j}$ and $\Delta f_p^{i,j}$ are s^th^ and p^th^ elements of difference vector $\mathbf{\Delta f}^{i,j}$:

$$\mathbf{\Delta f}^{i,j}=\mathbf{f}_i^1-\mathbf{f}_j^2.$$

Values of the conformity function (1) of feature points obtained from the first image are calculated for every feature point of the second image. j^{th} feature point of the second image is chosen as the corresponding point for i^{th} point of the first image if its value of the function $W_{i,j}$ is minimum. The task is to construct an image matching algorithm by feature points, which uses a conformity function value as a measure of proximity for descriptors based on histograms of oriented gradients (HOG).

III. Conforming Matching of HOG-Descriptors

To construct the algorithm, we based on the popular implementation of HOG-descriptors in SIFT method [3]. We form descriptors using image fragments with a size of 16×16 pixels. According to paper [3], we divide fragment into 16 subfragments with a size of 4×4 pixels. For every subfragment we form a histogram of oriented gradients with 8 blocks, which are related to 8 descriptor elements. Thus, every descriptor contains $4\times4\times8=128$ elements.

The use of a conformed measure of proximity for descriptor vectors with the size of 128×1 need a lot of computational resources. To reduce calculational efforts, we propose a new algorithm calculating measures of proximity for every of 16 descriptor subareas based on the values of its histograms of oriented gradients. A modified conformed measure of proximity is defined as:

$$W_{i,j}=\sum_{\substack{q=1\\r=1}}^{4}W_{i,j}^{q,r}, \qquad (2)$$

where

$$W_{i,j}^{q,r}=\sum_{\substack{s=1\\p=s+1}}^{8}(\Delta f_s^{i,j,q,r}-\Delta f_p^{i,j,q,r})^2.$$

$W_{i,j}^{q,r}$ is the value of proximity measure of q^th^,r^th^ part of descriptor vectors for i^th^ and j^th^ feature points correspondingly, $\Delta f_s^{i,j,q,r}$ and $\Delta f_p^{i,j,q,r}$ are s^th^ and p^th^ elements of difference vector $\mathbf{\Delta f}^{i,j,q,r}=\mathbf{f}_{i,q,r}^1-\mathbf{f}_{j,q,r}^2$, $\mathbf{f}_{i,q,r}^1$ and $\mathbf{f}_{j,q,r}^2$ are q^th^,r^th^ histograms of oriented gradients of i^th^ and j^th^ feature points of first and second images correspondingly. In this case $K=8$, because histograms contain 8 elements.

It can be easily checked, that in case of this size of descriptor vectors using of component proximity measure (2) allows reducing computational efforts in more than 36 times. Of course, due to reliability reduction, question about changes in the accuracy of this measure is stated. In the next section we show results of experimental research of accuracy for this feature points matching method.

IV. Results of Experiments

The efficiency of the proposed algorithm, which is based on using of conformity function as proximity measure of descriptor vectors, was tested on images [22] presented on Figure 1.

Fig. 1. Test images.

Feature points were detected for both images, and descriptors were formed using functions from computer vision library OpenCV [23]. Then, corresponding points were found for every point of the first image using two methods: component proximity measure (2) and Euclidean metric.

We performed the same operations for every point of the second image: feature point detection and matching. Then, to exclude incorrectly found points, cross-check of obtained points was performed.

The results of experiments are presented in Table 1. The number of correctly found feature points (first row) and the number of correctly found feature points after cross-check (second row) were counted. We use the ratio of number of correctly found feature points before and after cross-check as measure of accuracy. In the third row of Table 1 relative values of accuracy for both methods are presented.

Table 1 shows that the low number of points is found correctly when a conformed measure of proximity (3) was used. At the same time, the relative value of correctly found feature points is higher for this measure. In many applications of computer vision, this factor may be critical because even the low number of incorrect points can disrupt the work of technologies based on image matching using.

TABLE I. RESULTS OF RESEARCH

Measure of proximity	**Proximity measure**	
	Euclidean distance	***Conformity function***
Total number of correctly found feature points	571	535
Number of correctly found feature points after cross-check	396	377
Relative number of correctly found feature points	0.693	0.70

V. CONCLUSION

The proposed modification of the image matching algorithm using component conformity function as a measure of proximity allows reducing computational efforts more than 36 times. It was determined that the constructed algorithm has a high accuracy. Although the number of correctly found feature points was slightly lower, when criterion based on component conformity function was used, the relative number of correctly found feature points was higher. This can be important in some applications of computer vision.

ACKNOWLEDGMENT

This work was financially supported by the Russian Foundation for Basic Research under grant #18-07-01390, and by the Ministry of Science and Higher Education (project).

REFERENCES

[1] V. Fursov, "Constructing a quadratic-exponential FIR-filter with an extended frequency response midrange," Computer Optics, vol. 42, no. 2, pp. 297-305, 2018. DOI: 10.18287/2412-6179-2018-42-2-297-305.

[2] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004. DOI: 10.1023/b:visi.0000029664.99615.94.

[3] H. Bay, T. Tuytelaars and L. Van Gool, "SURF: Speeded Up Robust Features," Computer Vision – ECCV, pp. 404-417, 2006. DOI: 10.1007/11744023_32.

[4] E. Rublee, V. Rabaud, K. Konolige and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," International Conference on Computer Vision, Barcelona, pp. 2564-2571, 2011. DOI: 10.1109/ICCV.2011.6126544.

[5] P. Alcantarilla, J. Nuevo and A. Bartoli, "Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces," Procedings of the British Machine Vision Conference, 2013. DOI: 10.5244/c.27.13.

[6] K. van de Sande, T. Gevers and C. Snoek, "Evaluating Color Descriptors for Object and Scene Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1582-1596, 2010. DOI: 10.1109/TPAMI.2009.154.

[7] I. Abadi, A. Moussa and N. El-Sheimy, "Evaluation of feature points descriptors' performance for visual finger printing localization of smartphones," IEEE/ION Position, Location and Navigation Symposium (PLANS), Monterey, CA, pp. 1002-1008, 2018. DOI: 10.1109/PLANS.2018.8373478.

[8] J. Xu, S. Lin and A. Teng, "Improved SIFT algorithm based on adaptive contrast threshold," 4th International Conference on Computer and Technology Applications (ICCTA), Istanbul, pp. 151-155, 2018. DOI: 10.1109/CATA.2018.8398673.

[9] Y. Xiang, F. Wang and H. You, "OS-SIFT: A Robust SIFT-Like Algorithm for High-Resolution Optical-to-SAR Image Registration in Suburban Areas," IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 6, pp. 3078-3090, 2018. DOI: 10.1109/TGRS.2018.2790483.

[10] F. Guo, J. Yang, Y. Chen and B. Yao, "Research on image detection and matching based on SIFT features," 3rd International Conference on Control and Robotics Engineering (ICCRE), Nagoya, pp. 130-134, 2018. DOI: 10.1109/ICCRE.2018.8376448.

[11] L. Păvăloi and C. D. Niţă, "Iris recognition using SIFT descriptors with different distance measures," 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Iasi, Romania, pp. 1-4, 2018. DOI: 10.1109/ECAI.2018.8678987.

[12] B. Zhong and Y. Li, "Image Feature Point Matching Based on Improved SIFT Algorithm," IEEE 4th International Conference on Image, Vision and Computing (ICIVC), Xiamen, China, pp. 489-493, 2019. DOI: 10.1109/ICIVC47709.2019.8981329.

[13] M. Wei and P. Xiwei, "WLIB-SIFT: A Distinctive Local Image Feature Descriptor," IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP), Weihai, China, pp. 379-383, 2019. DOI: 10.1109/ICICSP48821.2019.8958587.

[14] S.H. Lee and S. Sharma, "Real-time disparity estimation algorithm for stereo camera systems," IEEE Transactions on Consumer Electronics, vol. 57, no. 3, pp. 1018-1026, 2011. DOI: 10.1109/TCE.2011.6018850.

[15] R.K. Gupta and S. Cho, "Window-based approach for fast stereo correspondence," IET Computer Vision, vol. 7, no. 2, pp. 123-134, 2013. DOI: 10.1049/iet-cvi.2011.0077.

[16] A. Fusiello, U. Castellani and V. Murino, "Relaxing Symmetric Multiple Windows Stereo Using Markov Random Fields," Lecture Notes in Computer Science, pp. 91-105, 2001. DOI: 10.1007/3-540-44745-8_7.

[17] R. Yang and M. Pollefeys, "Multi-resolution real-time stereo on commodity graphics hardware," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, pp. I-I, 2003. DOI: 10.1109/CVPR.2003.1211356.

[18] S. Satoh, "Simple low-dimensional features approximating NCC-based image matching," Pattern Recognition Letters, vol. 32, no. 14, pp. 1902-1911, 2011. DOI: 10.1016/j.patrec.2011.07.027.

[19] F. Cheng, H. Zhang, D. Yuan and M. Sun, "Stereo matching by using the global edge constraint," Neurocomputing, vol. 131, pp. 217-226, 2014. DOI: 10.1016/j.neucom.2013.10.022.

[20] V. Fursov, A. Gavrilov, Y. Goshin and K. Pugachev, "The technology of image matching by the criterion of conformity of image fragments samples," Journal of Physics: Conference Series, vol. 1096, 012084, 2018. DOI: 10.1088/1742-6596/1096/1/012084.

[21] V. Fursov, Y. Goshin and K. Pugachev, "Adaptive algorithm of conforming image matching," CEUR Workshop Proceedings, vol. 2416, pp. 26-33, 2019. DOI: 10.18287/1613-0073-2019-2416-26-33.

[22] "Middlebury Stereo Datasets," Vision.middlebury.edu, 2020 [Online]. URL: https://vision.middlebury.edu/stereo/.

[23] "OpenCV," 2020 [Online]. URL: https://opencv.org/.