ORB-based Template Matching Through Convolutional Features Map

Dong Zhou[1], Yingxin Tian[2], Xiang Li[1] and Jiefei Wu[2]

[1]School of Aeronautics, Harbin Institute of Techology, Harbin, China
[2]School of computer science and technology, Harbin Institute of Techology, Harbin, China

***Abstract*—Template matching is an important part of computer vision, but most of common methods do not work well in some cases, such as complicated background clutter, deformation and partial occlusion. Therefore, we present a novel method for image matching, which is useful, robust and fast. Its essence is the ORB-based Convolutional Features Map (CFM), which is used to measure the similarity between template and target image. We study its properties and apply it to a real-world dataset in complex environment. The result of experiments demonstrates that our algorithm outperforms other commonly used algorithm.**

Keywords—Object Tracking, template matching, ORB-based convolutional feature map

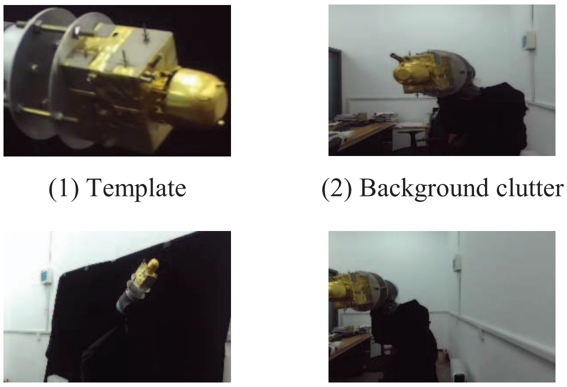
I. INTRODUCTION

Image template matching is significant to computer vision application, such as object detection, object tracking, motion estimation and remote sensing. In general, template is the region of interest in source image that contains an object of interest.

The object detection and track have been applied successfully to target image for decades by template matching methods, however, there are still some problems need to be solved. Specifically, these methods often perform simple arithmetic operation on corresponding pixels, when measuring the similarity between template and candidate image window. Hence, The matching result of these method is unsatisfied in some case, for example, large differences of background between template and target image, illumination changes, object with spatial motion, non-rigid deformation, and even partial occlusions (see Fig. 1).

In addition, It is necessary for some of template matching method with a specific parametric deformation model between template and target image, which degrade the robustness of matching method. Meanwhile, a number of parameters should be confirmed when complex deformation are considered.

To solve these problems above, paper [1] proposes a novel template matching method, and introduces a useful, robust, non-parameter similarity measurement—Best-Buddies Similarity (BBS). The BBS method has been applied to image matching by representing all the template and candidate image patches as point-sets in a $xyRGB$ feature space and considering the best-buddies pairs (BBPs) as the similarity measurement.



(1) Template (2) Background clutter

(3) Deformation (4) Occlusion

Fig. 1 **Some difficult problems for template matching:** (a), The template, arbitrary fetched from source image, only contains an object of interest. (b), Background clutter, the background behind object of interest often have a significant changes. (c), Deformation, the object in target image may have spatial motion and non-rigid deformation. (d), Occlusion, it often happens in object tracking.

Only the color and position information between template and target image have been taken into account in [1], hence, the tracking result of BBS method applied to object with spatial motion in complex background is not so well. And that the matching algorithm of best-buddies pairs is complicated and time consuming, which fails to satisfy the requires of real-time object tracking.

In this paper, we present a template matching algorithm and introduce a similarity measure method termed ORB-based Convolutional Features Map (ORB-based CFM). At first, we obtain the good matched features map based on ORB algorithm under L2 distance. Next, we apply the ORB-based CFM method to the good matched features map. Finally, the position of maximum of convolutional features map is considered as best matched region

We also employ our method to different specific cases and compare its performance to other commonly used template matching algorithm on a challenging dataset, which demonstrates the robustness and effectiveness of our method.

Corresponding Author: Yinxin Tian

Email: tyx@hit.edu.cn

978-1-7281-4094-0/19/$31.00 ©2019 IEEE

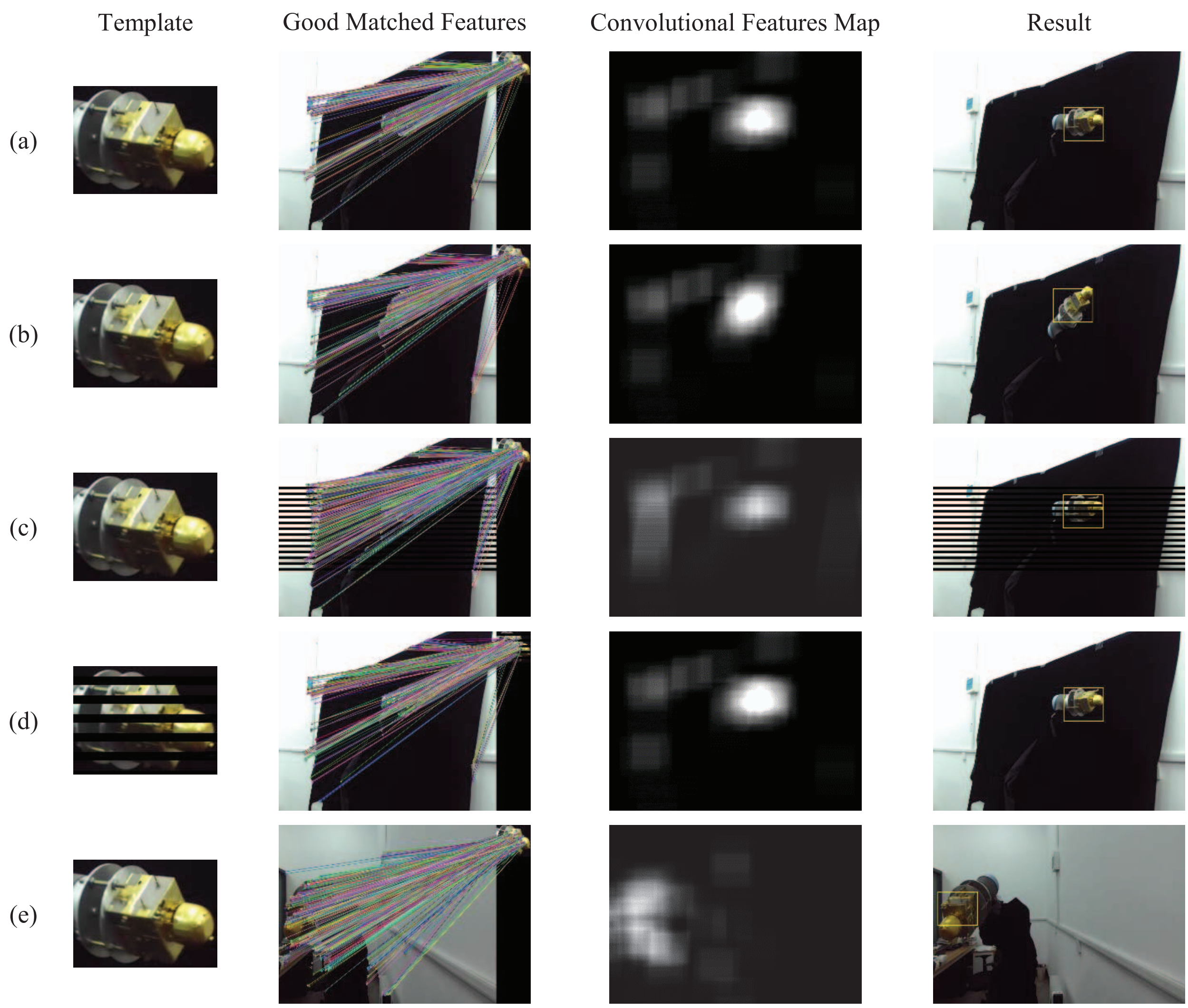


Fig. 2 **The template matching result of our methods.** Five different cases are shown: (a) Normal, (b) Deformation, the object of interest has a rigid geometric transformation between template and target image, because of spatial motion. (c) Target image with occlusion, there is a black belt covering the object of interest in target image. (d) Template with occlusion, which is partially covered. (e) Arbitrarily complex environment.

II. Related Work

Template matching algorithm is very dependent on the similarity measure method, which is applied to match the template and candidate image patches. The common methods are the SQDIFF, CCORR and CCOEFF. Due to computational efficiency [2], there are different variants of these algorithms proposed to handle the problem of background clutter (BC) and illumination changes (IM) [3, 4].

Another type of similarity measurement consists of robust error function, for example, M-estimators [5, 6] or Hamming-based distance [7, 8]. The additive noise and "salt and pepper" have less influence on these methods compared to cross correlation related algorithm. But the whole algorithms mentioned above presume a rigid geometric transformation between template and the object in target image, and they only focus on the pixel-wise differences between template and candidate image patch.

To deal with those parametric problems, some improved template matching method have been proposed [9, 10]. Paper [11] provides a fast algorithm for approximate template matching under 2-dimension affine transformation that minimizes the Sum-of-Absolute-Differences (SAD) error measurement. A globally optimal estimation of non-rigid image distortions has been found as well in paper [12]. However, these methods mentioned above suppose that there is a one-to-one mapping between template and candidate image patch for the deformation, and it has been proved that the presence of outliers fails to handle, for example, those caused by complex BC, IM and occlusions. In addition, they presume a parametric model with deformation, which is unnecessary to our method.

Histogram Matching (HM) is one of most famous similarity measurement between color histograms, which provides a non-parametric method for solving deformation and is commonly used in visual tracking [13, 14]. However, HM completely ignores geometric information and all the pixels have been evenly processed. Other object tracking methods are provided in [15, 16] to handle complex background clutters and partial occlusions. But it is different that our tendency is to detect in a single image, of which the redundant time information is in lack.

(a)

(b)

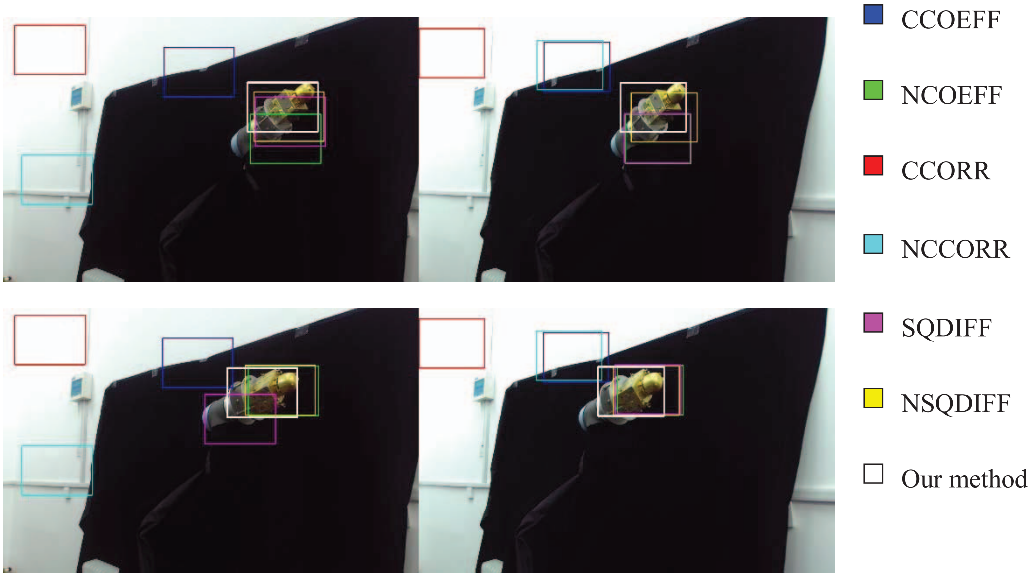


Fig. 3 **The result of different template matching algorithm:** Seven types of template matching algorithm including our method are applied to a real-world and simple dataset. Two of template matching results are shown in (a) and (b).

Paper [17] proposes a template matching probabilistic model based on maximum-likelihood estimation method used for both edge matching and gray-level image matching, where an image is represented in a 3D space. Furthermore, paper [18] uses *xyRGB* space and reduces template matching times in EMD measurement [19] between two point-sets. But our method is in a rBRIEF space and differs from EMD, which requires one-to-one matching and cannot distinguish between inliers and outliers.

As for image matching method, another famous measurement is the Hausdorff distance [20]. A fractional Hausdorff distance have been proposed in [20], where the Kth farthest points have been considered, rather than the farthest one. Therefore, it relies heavily on the parameter K that should be tuned.

A novel template matching algorithm termed Best-Buddies similarity(BBS) is proposed [1]. The Best Buddies was presented by Pomeranz *et al.* [21] in the context of solving jigsaw puzzles. In paper [1], the template and sliding window are represented as point sets in 2-dimension spatial space. The source image have been broken into patches with $k \times k$ size. Each of candidate image patches is transferred to a k^2 dimension feature vector, includes color and location information. And then the nearest-neighbor matching algorithm is applied to calculate all the BBPs between template and each of image patches. The best matched region is the candidate patch with the most BPPs. The BBS method can robustly match the template to arbitrary target image, even in the presence of some outliers (i.e., BC, IM and occlusions) and non-rigid object deformation.

However, there are two drawbacks in the BBS algorithm. (1) The number of feature vectors to be matched is large and the BBPs matching needs to undergo Nearest-Neighbor algorithm twice, which cause the low efficiency of BBS method. (2) Representing the template and each of image patches in *xyRGB* space is too weak, which may fail to track the object with arbitrarily spatial motion in an unconstrained environment.

III. Method

In this section, we describe the general ideal of our method, some definitions and the details of ORB-based CFM algorithm are also mentioned.

Good Matched Features Map: we assume the size of target image is $S(W,H)$ and the size of template is $R(w,h)$. Two sets of key points $P=\{p_i\}_{i=1}^{N}$ and $T=\{t_i\}_{i=1}^{M}$ are extracted from target image and template based on ORB algorithm, where $p_i, t_i \in R^d$ and $N \geq M$ .The good matched set of key points $G=\{g_i\}_{i=1}^{N}$

$$g_i = \hat{p} = NN(t_i, P) \tag{1}$$

where $NN(t_i,P)=\arg\min\{d(t_i,p_j)\}$, $j=1,2,...,N$ is a nearest neighbor algorithm and $d(t_i,p_j)$ denotes a distance measure, such as L1, L2 and etc. Meanwhile, we add a threshold selection to the good matched set *G*. We presume that the corresponding distance of g_i is d_i, the maximum distance of features set G is $d_{\max}$ and the threshold value $TH=0.6*d_{\max}$. If $d_i < TH$, this good matched features can be retained, otherwise, it would be deleted. The key points set after threshold selection is defined as $\bar{G}=\{\bar{g}_i\}_{i=1}^{K}$. The size of good matched features map is $M(W,H)$ and

$$M(x,y)=\begin{cases}255 & ,(x,y)=\bar{g}_i \\ 0 & ,else\end{cases} \tag{2}$$

where $(x,y) \in R^{W\times H}$.

The good matched set of features *G* between template and target image are shown in the second column of Fig. 3. As we can see, the density of good matches in the region of interest is much higher than background. Hence, we introduce the convolutional features map to describe the density of good matched features.

Convolutional Features Map: we assume that template size is $R(w,h)$ and the size of good matched features map is $M(W,H)$. Meanwhile, we design a convolutional kernel $K(w,h)$, of which all the elements is one. The convolutional features map $C(W,H)$ can be obtained by

$$C = K * M \tag{3}$$

The examples of convolutional features map can be seen in the third column of Fig. 3. It is clear to see that the region of interest corresponds to the brightest region in CFM.

IV. Experiments

We employ our method in different situations to show its robustness and accuracy. We compare ORB-based CFM with other six type of similarity measurement usually applied for template matching: 1) Square-Difference (SQDIFF), 2) Normalized-Square-Difference (NSQDIFF), 3) Cross-Correlation (CCORR), 4) Normalized-Cross-Correlation (NCCORR), 5) Correlation Coefficient (CCOEFF), 6) Normalized Correlation Coefficient (NCCOEFF).

A. *Implementation in different cases*

We apply ORB-based CFM method to five different cases, including normal, deformation, target with occlusion, template with occlusion and arbitrarily complex environment. More details of experimental results can be seen in Fig. 2.

The first column of Fig. 2 is template, which fetched randomly from source image. The template only contains one object of interest and its size is much smaller than target image. All the templates used in experiment are the same, excepted that the template (d) is covered by a black belt. The second and third column of Fig. 2 are the good matched features map and convolutional features map. It's easy to find where the object of interest is in these maps. The final template matching results are shown in the last column of Fig. 2, and all the objects of interest have been successfully highlighted by yellow box.

This experiment proves that our method works well in any complex case, even the template or target image is partially covered by something. It is worth mentioning, that the running time of one frame with ORB-based CFM method only needs 100 milliseconds under the windows 10 platform (intel i5-4210M CPU @2.60GHz), which can satisfy the requirement of real-time object tracking.

B. *Comparison with other altorithm*

We employ our method and other six types of template matching algorithm to a real-world annotated dataset. This dataset consists of 290 binocular images captured by a binocular camera, which describes arbitrarily spatial motion of a typical model. And the object of interest in each image of dataset has been bounded in box artificially. Two of comparing result are shown in the Fig. 3. Different template matching result have been bounded by different color box. The result of our method is in white.

We can found from Fig. 3 that the target is exactly bounded by white box, which means that our method can track the object of interest accurately. The bounding boxes of NCOEEF, SQDIFF and NSQDIFF merely contain part of object. Even worse, the CCOEFF, CCORR and NCCORR methods is completely failed to track object.

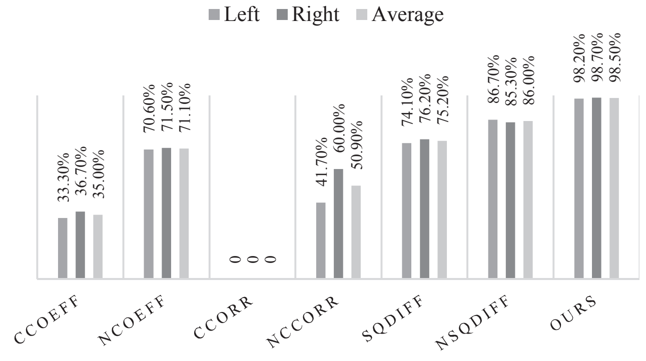


Fig. 4 **Accuracy** : The accuracy rates of template matching methods on a challenging dataset which includes left and right images are shown.

We evaluate the accuracy of seven methods by comparing each of template matching results in experiment to its corresponding ground truth (see Fig. 4). The evaluation proves that our method has higher accuracy than other algorithm.

V. Conclusion

We introduced a new measure, ORB-based CFM, for template matching in unconstrained environment. Through two groups of experiments, it has been proved that our method can handle different extremely cases successfully and outperform plenty of commonly used image matching algorithm, for example, CCORR, SQDIFF, CCOEFF and so on.

It is worth noticing that the algorithm proposed in our paper have been not really parametric-free, which is introduced by ORB algorithm and the threshold selection of good matches. Meanwhile, our method may not work well when the template is very small, because the ORB algorithm cannot extract sufficient key points.

The convolutional features map can also be extended to other features detector such as SIFT, SURF and so on, which relies on the requirements of practical application.

Acknowledgment

This work was kindly supported by National Key R&D Program of China through grant SQ2019YFB130028.

References

[1] T. Dekel, S. Oron, M. Rubinstein, S. Avidan, and W. T. B. I. Freeman, "Best-Buddies Similarity for Robust Template Matching". in IEEE Conference on CVPR, 2015, pp. 2021-2029.

[2] W. Ouyang, F. Tombari, S. Mattoccia, L. Di Stefano, and W. K. Cham. "Performance evaluation of full search equivalent pattern matching algorithms". IEEE Trans. on PAMI, 2012. 2.

[3] E. Elboher and M. Werman. "Asymmetric correlation: a noise robust similarity measure for template matching". IEEE Trans. on Image Processing, 2013. 2.

[4] Y. Hel-Or, H. Hel-Or, and E. David. “Matching by tone mapping: Photometric invariant template matching”. IEEE Trans. on PAMI, 2014. 2. pp. 317-330.

[5] J. H. Chen, C. S. Chen, and Y. S. Chen. “Fast algorithm for robust template matching with m-estimators”. IEEE Trans. on Signal Processing, 2003. 2.

[6] A. sibiryakov. “Fast and high-performance template matching method”. In IEEE Conference on CVPR, 2011. 2.

[7] B. G. Shin, S. Y. Park, and J. J. Lee. “Fast and robust template matching algorithm in noisy image”. In International Conferece on Control, Automation, and Systems, 2007. 2.

[8] O. Pele and M. Werman. “Robust real-time pattern matching using bayesian sequential hypothesis testing”. IEEE Trans. on PAMI, 2008. 2

[9] H. Y. Kim and S. A. De Araujo. “Grayscale template matching invariant to rotation, scale, translation, brightness and contrast.” In AIVT. Springer, 2007. 2.

[10] D. M. Tsai and C. H. Chiang. Rotation-invariant pattern matching using wavelet decomposition. IEEE Trans. on PAMI, 2002. 2.

[11] S. Korman, D. Reichman, G. Tsur, and S. Avidan. “Fast match: Fast affine template matching”. In IEEE Conference on CVPR, 2013. 2.

[12] Y. Tian and S. G. Narasimhan. “Globally optimal estimation of nonrigid image distortion”. In International Journal of Computer Vision, 2012. 2.

[13] D. Comaniciu, V. Ramesh, and P. Meer. “Real-time tracking of non-rigid objects using mean shift”, In International Conference on CVPR, 2000. 2.

[14] P. Perez, C. Hue, J. Vermaak, and M. Gangnet. “Color-based probabilistic tracking”, In Europe Conference on Computer Vision, 2002, 2.

[15] C. Bao, Y. Wu, H. Ling, and H. Ji. “Real time robust l1 tracker using accelerated proximal gradient approch”. In International Conference on CVPR, 2012, 2.

[16] X. Jia, H. Lu and M. Yang. “Visual tracking via adaptive structural local sparse appearance model”. In International Conference on CVPR, 2012, 2.

[17] C. F. Olson. “Maximum-likelihood image matching”. IEEE Trans. on PAMI, 2002. 2. pp: 853-857.

[18] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan. “Locally orderless tracking”. International Journal of Computer Vision, 2014. 3.

[19] Y. Rubner, C. Tomasi, and L. Guibas. “The earth mover’s distance as a metric for image retrieval”. International Journal of Computer Vision, 2000. 3.

[20] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. “Comparing images using the hausdorff distance”. IEEE Trans. on PAMI, 1993. 3. pp: 850-863.

[21] D. Pomeranz, M. Shemesh, and O. Ben-Shahar. “A fully automated greedy square jigsaw puzzle solver”. In the 24th IEEE Conference on CVPR, 2011. 3. pp: 9 – 16.