Proceeding of the 2015 IEEE
International Conference on Information and Automation
Lijiang, China, August 2015

A novel template matching algorithm based on the contextual semantic information

Shuoyan LIU, Kai FANG and Li JIANG
Institute of Computer Technology Department
China Academy of Railway Sciences
Beijing,100081,P.R.China
06112062 @bjtu.edu.cn

***Abstract* - This paper presents a novel template matching algorithm that copes with the occlusion on the target image. Most of the previous methods are sensitive to the occlusion because it induces some patches with similar semantic concepts but distinct appearances. Thus the template matching algorithm hardly accomplishes the task based on the appearance similarity only. To overcome this limitation, we integrate the contextual semantic information into the template matching algorithm. To this end, we first segment the template image into 9 patches. The center patch is used to compute the appearance similarity and its neighborhood patches are adopted to construct the contextual semantic constraint. And then we obtain the integrated distance by introducing the pseudo-likelihood to combine the feature appearance similarity and contextual semantic information together. Finally, the arbitrary regions of a target image are matched with the template image via integrated distance. The experimental results demonstrate the proposed method is more robust to occlusion than previous template matching techniques.**

Index Terms – template matching, contextual semantic information, pseudo-likelihood function

I. INTRODUCTION

This paper investigates template matching problem. Once a template image and a target image are given, it determines the matched region of the target image that is similar to the template image. Template matching is an important problem for computer vision, and has received considerable attention in the recent past. Most previous studies have focused on the template matching algorithms invariant to rotation and illumination changes. Normalized cross correlation (NCC)-based method is one of the most popular methods for illumination-invariant matching, and various schemes for efficient computation of NCC have been proposed [1–4]. For rotation-invariance, Ullah [5] introduced the concept of the orientation codes similar to the histogram of oriented gradients (HOG). Recently several algorithms invariant to both rotation and illumination changes have been proposed using circular projection and Zernike moments [6].

On the other hand, template matching methods that are robust to occlusion problem have rarely been considered. It is common that the features extracted from the template and target images are difference because of the change in the image acquisition conditions and environment, including the illumination conditions, the image quality change caused by using different cameras, etc. This is similar to partial occlusion on the target image. In addition, users have intended to cover something on the target image. For case, a protective covers for the front of a machine or device (as a door lock or computer component). As a result, the need for developing template matching method with robustness to occlusions has increased. Even though several proposed algorithms consider the appearance difference [7,8], they are unable to handle occlusion on the target image. Local descriptor-based image matching methods, such as SIFT, could be a solution to matching images with appearance difference [9]. However, as shown in Fig.1-(a), the template matching can be inaccurate when the protective covers for the wheel. The drawback stems from the dependence on the appearance similarity only. The SIFT-based matching can suffer when faced two regions with similar semantic concepts but distinct appearances. Distinct from them, we propose the novel template matching algorithm by introducing contextual information from the neighborhood regions. The introduced contextual information provides useful information or cue about matching, which can reduce the appearance ambiguity. The improved template matching is capable of enhancing the matching performance, as shown in Fig.1-(b).

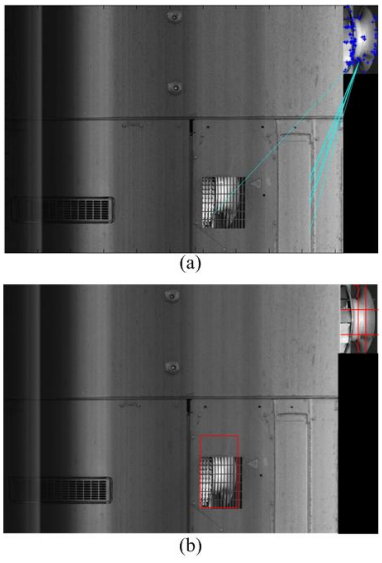


Fig. 1 An example of failure cases of local descriptor-based image matching. The image pairs (a) are the result of SIFT-based image matching and the (b) are the result of the proposed method, where the left image is a target image and the right image is a template image.

In this paper, we present a novel template matching

978-1-4673-9104-7/15/$31.00 ©2015 IEEE

algorithm that copes with the occlusion on the target image. Specifically, we first segment the template image into 9 patches and the target image is partitioned into patches with the same size. And then the histograms of the oriented gradients (HoG) features [10] are extracted from each regular segmentation patch. Thereafter we compute the appearance similarity between the arbitrary patches of a target image and centering patch of the template image. The contextual semantic information is the relationship of neighborhood regions for the arbitrary patch and the template image. The integrated distance is obtained by introducing the pseudo-likelihood to combine the feature appearance similarity and contextual semantic information together. Finally, the arbitrary regions of a target image are matched with the template image via integrated distance. The experimental results demonstrate the proposed method is more robust to occlusion than previous template matching techniques.

The rest of the paper is organized as follow. Section 2 gives the proposed framework in detail. We show the performance of the proposed method on two challenging datasets in Section 3. Finally, Section 4 concludes the paper.

II. A NOVEL TEMPLATE MATCHING ALGORITHM BASED ON THE CONTEXTUAL SEMANTIC INFORMATION

In this paper, we present a novel template matching method based on the contextual information with robustness to the occlusion for the target images. The key idea is to integrate contextual features with local features can find the matched region in the target image is most similar to the template image. To this end, we first segment the template image into 9 patches. The center patch is used to compute the appearance similarity and its neighborhood patches are adopted to construct the contextual semantic constraint. And then we obtain the integrated distance by introducing the pseudo-likelihood to combine the feature appearance similarity and contextual semantic information together. Finally, the arbitrary regions of a target image are matched with the template image via integrated distance. An overview of the architecture is shown in Fig.2.

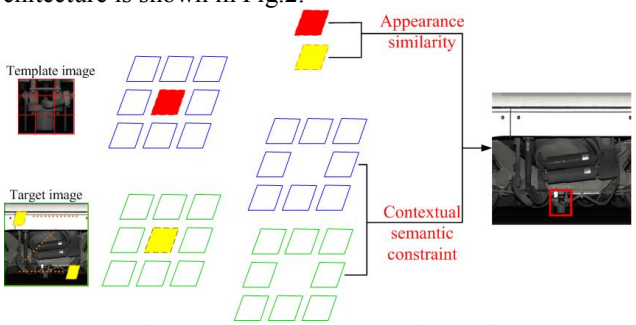


Fig. 2 Work flow of the proposed approach

A. The appearance similarity

The proposed template matching method utilizes the histogram of oriented gradients (HOG) [10] feature. The template image is divided into 3×3 patches. And then the target image is segment into $l\times l$ rectangular patches with the same patch size. We extract the histograms of the oriented gradients (HoG) features from each regular segmentation patch.

The HOG is a histogram that is constructed from the magnitude and orientation information of gradients. In special, we first divide the image patch into 2×2 cells, for each cell accumulating a histogram of gradient directions over the pixels of the cell. And then it is also useful to contrast-normalize the local responses before using them. This can be done by accumulating a measure of local histogram over patches and using the results to normalize all of the cells in the patch.

After obtaining the HoG features, we compute the appearance similarity between the arbitrary patches of a target image and centering patch of the template image according to the histogram intersection:

$$Sim(H_1,H_2)=\frac{\sum_{i=1}^{N}\min(H_1(i),H_2(i))}{\sum_{i=1}^{N}H_1(i)} \quad (1)$$

where H_1, H_2 define the HoG features of image patch in the template and target image, respectively.

B. The contextual semantic constraint

The introduced contextual information provides useful information or cue about the centering patch matching. Two patches differed from each other when analyzed appearance similarity independently, might be matched with the help of context knowledge [11]. In this paper, we assume that most of the contextual information can be captured from the neighborhood regions of the center patch at the same scale level. Although, the regions with large distance from the center and the regions at other scale levels can provide some useful contextual information, it is not easy to integrate this information in a totally unsupervised manner because of the large searching space for finding the appropriate way to combine all the contextual information. Thus, in current stage, we choose neglect the contextual information from longer distance and at other scale levels [12].

We adopt the MRF model to describe the contextual semantic constraint. The MRF model has shown that the state of i^{th} grid depends on the state of the neighborhood grid $N(i)$ according to $P(z_i \mid z_{S/\{i\}}) = P(z_i \mid z_{N(i)})$, given the state of the i^{th} grid. Hence we describe the contextual information as following:

$$P(z_i \mid z_{N(i)})=\frac{1}{8}\sum_{j\in N(i)}Sim(H_i,H_j)=\frac{1}{8}\sum_{j\in N(i)}\frac{\sum_{k=1}^{N}\min(H_i(k),H_j(k))}{\sum_{k=1}^{N}H_i(k)} \quad (2)$$

C. The template matching based contextual semantic information

The proposed template matching algorithm is integrating contextual features with local features to find the matched

region in the target image. Take center patch x_k of template image as an example to illustrate the proposed approach, when the neighborhood patches are matched into the location l_i in the target image, even if the low-level features of patch x_k exclude the possibility, our algorithm trades off the two effects and makes the x_k be labeled to l_i, and vice versa.

The template matching algorithm based contextual semantic information is presented in Table 1. Let x_k is the center image patch in the template image and t_k is arbitrary patch in the target image. In the second step, we first compute the appearance similarity $d^2(x_k, t_k)$. And then the neighborhood contextual information is obtained by the $P(z_i \mid z_{N(i)})$. In the third step, we calculate the integrated distance according to the pseudo-likelihood function using the following equation:

$$d^2_{Integrated}(x_k, t_k) = d^2(x_k, t_k) / P_G(z_i = \mathrm{k} \mid z_{N(i)}) \quad (3)$$

To match template and target images, in the target image, we find the minimization integrated distance between the matched region and template image. The target image is scanned from the top-left to the bottom-right with rectangular sliding windows, which have the same sizes of the template image. Both the rectangular sliding window and the template image can be segmented into 9 patches. Their center patches are used to compute the appearance similarity and their neighborhood patches are adopted to construct the contextual semantic constraint. And then we obtain the integrated distance by introducing the pseudo-likelihood to combine the feature appearance similarity and contextual semantic information together. The matched region is located according to minimization of the integrated distance $d_i^2(x_k, t_k)$.

TABLE I PSEUDO CODE OF THE TEMPLATE MATCHING ALGORITHM BASED ON CONTEXTUAL SEMANTIC INFORMATION

Algorithm: Template matching algorithm based on contextual semantic information

Input: template image x, target image t

Output: the location l of the matched region in the target image

Process：

The target image is scanned from the top-left to the bottom-right with rectangular sliding windows, which have the same sizes of the template image. Both the rectangular sliding window and the template image can be segmented into 9 patches.

Initializing step：

1.The histograms of the oriented gradients (HoG) features are extracted from each regular segmentation patch.

2. x_k is the center image patch of the template image;

3. t_k is center patch of the rectangular sliding windows.

Matching Step：

4.Compute the appearance similarity $d^2(x_k, t_k)$;

5.Obtain the neighborhood contextual information $P(z_i \mid z_{N(i)})$;

6.The integrated distance is obtained using $d^2_{integrated}(x_k, t_k) = d^2(x_k, t_k) / P_G(z_i = \mathrm{k} \mid z_{N(i)})$

End for

The matched region is located according to minimization of the integrated distance $d^2_{integrated}(x_k, t_k)$.

III. EXPERIMENTS

In this section, we evaluate the robustness to occlusion of the proposed algorithm. The test dataset is the image with some occlusion, which collected by the trouble of moving EMU detection system (TEDS)[14]. TEDS acquires the moving EMU images using the line scan cameras which fixed around the rail. It is common that the features extracted from current and former images are difference because of the change in the image acquisition conditions and environment. In addition, the troubles have been easily occurred in the train for a long time. Hence, these images are suitable for the testing the occlusion-invariance of the proposed approach.

Fig.3 represents the overall information of our test data set, where (a)-(d) are the four different EMU images, respectively. The top image of each image pair is the template images, and the bottom one is the target images. In particular, the template image of (a) is covered by the engine hood in the target image. For (b), we can see the template image with the rubberized fabric in the target image. The template image region in (c) is partly covered and cut. Some objects are glued into the template image region in the target image (d).

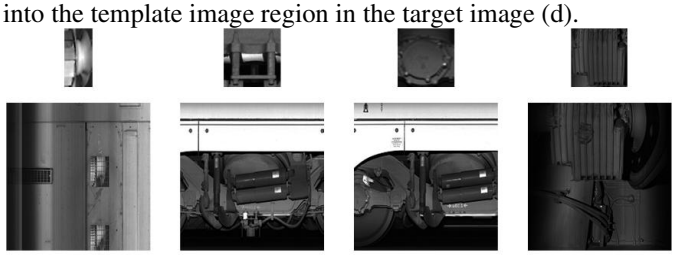


Fig. 3 Our test data set: (a)-(d) are the four different EMU images, the top image of each image pair is the template images, and the bottom one is the target images.

In order to evaluate the performance of our proposed method, we measure the occlusion -invariance using success rate $SR = [B_G \cap B_M] / [B_G \cup B_M]$, where B_G and B_M are the ground truth region and the matched region in the target image, respectively. More intuitive explanation for these measures is shown in Fig.4.

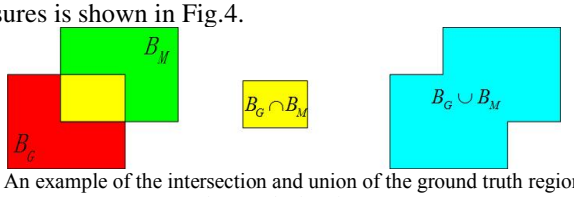


Fig. 4 An example of the intersection and union of the ground truth region and the matched region.

Table 5 shows the average success rate for the matched regions of the given template images, where $Image(a)$ - $Image(d)$ corresponds to a - d in the Fig3. In order to well investigate the proposed approach, we repeat 50 times with different target image acquired from different times. The average success rate as $ASR = \frac{1}{N} \sum_{i=1N} SR(\mathrm{i})$, N is the number of the target images. The contextual information can help match template into the target image with occlusion. A closer look at the table reveals that the lowest match rate occurs in the $Image(a)$. Since the engine hood covers all of target image, it decreases the usefulness of neighborhood patches. It is needed to consider the high-level region contextual information as the importance cues to match them. In addition,

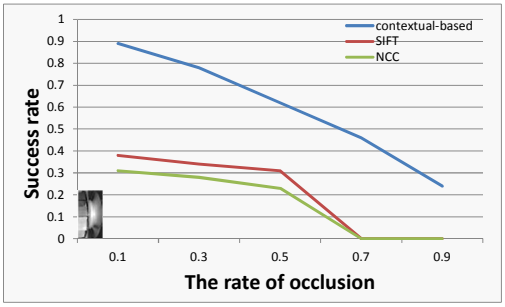
the best performance is achieved in both the *Image*(b) and *Image*(d). For target image, it is hardly matched the template image based on the appearance similarity of their center patches, their neighboring patches provide an informative cue. Thus, we can see that contextual semantic information can adversely affect template matching performance, which is also consistent with the results from Torralba [11].

TABLE II PERFORMANCE OF THE PROPOSED TEMPLATE MATCHING ALGORITHM, WHERE *Image*(a) - *Image*(d) CORRESPONDS TO a - d IN THE FIG3.

	Image(a)	*Image*(b)	*Image*(c)	*Image*(d)
Ave. Success rate	90.2%	94.8%	91.3%	95.6%

Thereafter, we investigate how the proposed algorithm is affected by the occlusion. For completeness, Fig. 5 lists the performance achieved according to the different extent of the occlusion, as well as the comparison performance with the template matching SIFT-based and NNC-based. For all three template matching alogirthms, results descrease dramatically as the rate of occlusion go from 0.1 to 0.9. In particular, it can be seen that the performance decreases progressively until the rate of occlusion is 0.5, and then drops off sharply. It shows that if the occlusion is too larger, the neighborhood region will sacrifice some useful cues, which decreases the matching performance.

Next, we compared our occulsion-invariant template matching with the template matching SIFT-based and NNC-based. The proposed template matching method outperforms other two template matching methods as shown in Fig.5 for the four objects. Occlusion induce the difference of appearance various between template and target image. Hence, the template matching NCC-based could not find the matched region in the target image dependence on the appearance similarity only. In addition, the template matching SIFT-based has shown better performance than NCC based method. The SIFT-based template matching is the local-descriptor matching and finds the optimal homography that maps the template image to the target image. Local descriptor-based image matching methods could be a solution to matching images with appearance difference. However, as shown in Fig.1-(a), the template matching can be inaccurate when the occlusion occurred in the target image.



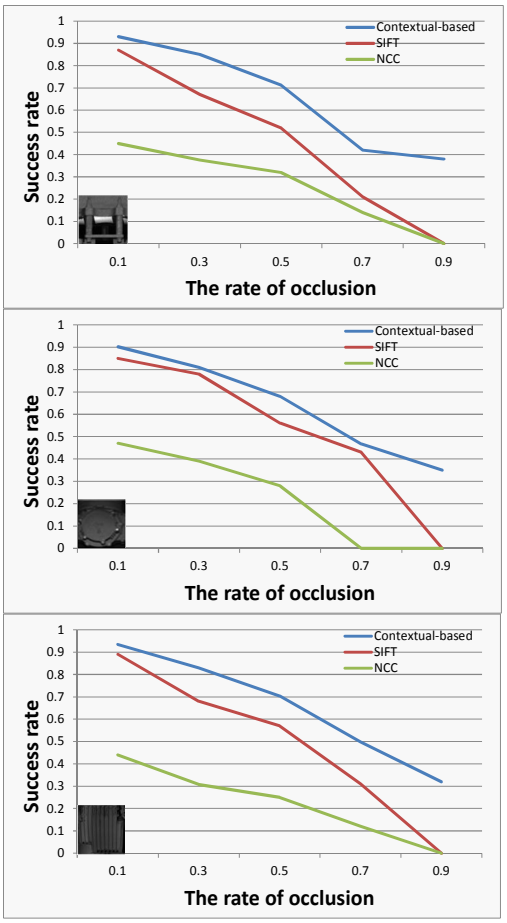


Fig. 5. Graphs for the success rate according to the rate of occlusion for each target image.

Finally, we apply the proposed template matching algorithm in the trouble of moving EMU detection system (TEDS). We adopt the on-line template matching process. The trouble detection is performed on a PC with Intel Core i7 with 3.4 GHz and 8 GB of RAM. The average processing time for the on-line process is 5min/train. The example images from the TEDS are shown in Fig.6.

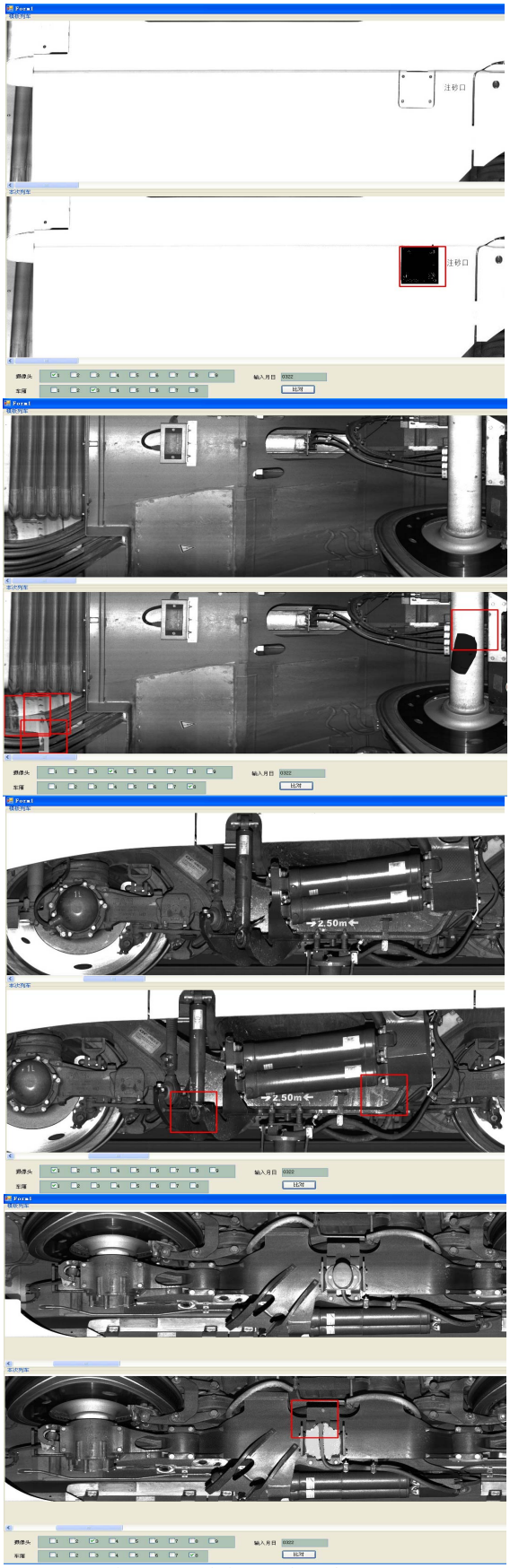


Fig. 6 Example images from the TEDS

IV. Conclusion

We present a novel template matching algorithm that copes with the occlusion on the target image. The experimental result has shown that the proposed method is more robust to occlusion than previous template matching techniques. The proposed template matching method has been a useful tool for TEDS (Trouble of moving EMU Detection System) and other trouble detection applications. Further efforts plan to take other cognitive knowledge related with contextual semantic information into consideration to improve the matching performance.

References

[1] J.P. Lewis, "Fast template matching," In: Proc. of *Vision Interface*, Canada, 1995:120-123.

[2] S.D. Wei, S.H. Lai, "Fast template matching based on normalized cross correlation with adaptive multilevel winner update," *IEEE Trans. Image Process*.17 (11) (2008)2227–2235.

[3] A. Mahmood, S.K. Han, "Exploiting transitivity of correlation for fast template matching," *IEEE Trans. Image Process.*19(8)(2010)2190–2200.

[4] A. Mahmood, S. Khan, "Correlation-coefficient-based fast template matching through partial elimination," *IEEE Trans. Image Process.* 21(4)(2012) 2099–2108.

[5] F. Ullah, S. Kaneko, "Using orientation codes for rotation-invariant template matching," *Pattern Recognition* 37(2004)201–209.

[6] M.S. Choi, W.Y. Kim, "A novel two stage template matching method for rotation and illumination invariance," *Pattern Recognition*. 35(2002):119-129.

[7] G. Tzimiropoulos, V. Argyriou, "Robust FFT-based scale-invariant image registration," *IEEETrans.PatternAnal.Mach.Intell* .32(10)(2010) 1899–1906.

[8] J. Yoo, S.S. Hwang, S.D. Kim," Scale-invariant template matching using histogram of dominant gradients," *Pattern Recognition*, 2014,47(9):3006-3018

[9] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int.J. Comput. Vis.* 60(2)(2004)91–110.

[10]N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection". *In: proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2005:886-893.

[11]A. Torralba, A. Oliva, " Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search." *Psychological Review*, 113(4), 766 – 786.

[12]J. Qin, NHC. Yung, "Scene categorization via contextual visual words," *Pattern Recognition*,2010,43(1):1874-1888.

[13]J. Besag, "On the Statistical analysis of dirty pictures," *Journal of Royal Statistical Society*, 1986, B43 (3):259-302.

[14]L. JUN, "The Design and Implementation of TEDS System," *Beijing University of Posts and Telecommunications*, 2012