

Towards a Robust Scale Invariant Feature Correspondence

Shady Y. El-Mashad^(✉) and Amin Shoukry

Computer Science and Engineering Department, Egypt-Japan University
for Science and Technology (E-JUST), Alexandria, Egypt
{shady.elmashad,amin.shoukry}@ejust.edu.eg

Abstract. In this paper, we introduce an improved scale invariant feature correspondence algorithm which depends on the Similarity-Topology Matching algorithm. It pays attention not only to the similarity between features but also to the spatial layout of every matched feature and its neighbours. The features are represented as an undirected graph where every node represents a local feature and every edge represents adjacency between them. The topology of the resulting graph can be considered as a robust global feature of the represented object. The matching process is modeled as a graph matching problem; which in turn is formulated as a variation of the quadratic assignment problem. The Similarity-Topology Matching algorithm achieves superior performance in almost all the experiments except when the image has been exposed to scaling deformations. An amendment has been done to the algorithm in order to cope with this limitation. In this work, we depend not only on the distance between the two interest points but also on the scale at which the interest points are detected to decide the neighbourhood relations between every pair of features. A set of challenging experiments conducted using 50 images (contain repeated structure) representing 5 objects from COIL-100 data-set with extra synthetic deformations reveal that the modified version of the Similarity-Topology Matching algorithm has better performance. It is considered more robust especially under the scale deformations.

Keywords: Features matching · Features extraction · Topological Relations · Graph matching · Performance evaluation

1 Introduction

Image matching or in other words, comparing images in order to obtain a measure of their similarity, is an important computer vision task. It is involved in many different applications, such as object detection and recognition, image classification, content based image retrieval, video data mining, image stitching, stereo vision, and 3D object modeling. A general solution for identifying similarities between objects and scenes within a database of images is still a faraway goal. There are a lot of challenges to overcome such as viewpoint or lighting

variations, deformations, and partial occlusions that may exist across different examples.

Furthermore, image matching as well as many other vision applications rely on representing images with sparse number of distinct keypoints. A real challenge is to efficiently detect and describe keypoints, with robust representations invariant against scale, rotation, view point change, noise, as well as combinations of them [1].

Keypoint detection and matching pipeline has three distinct stages which are feature detection, feature description and feature matching. In the feature detection stage, every pixel in the image is checked to see if there is a unique feature at this pixel or not. Subsequently, during the feature description stage, each region (patch) around the selected keypoints is described with a robust and invariant descriptor which can be used to match against other descriptors. Finally, at the feature matching stage, an efficient search for prospective matching descriptors in other images is made [2].

In the context of matching, a lot of studies have been used to evaluate interest point detectors as in [3,4]. On the other hand, little efficient work has been done on the evaluation of local descriptors. K. Mikolajczyk and C. Schmid [5], proposed and compared different feature detectors and descriptors as well as different matching approaches in their study. Although this work proposed an exhaustive evaluation of feature descriptors, it is still unclear which descriptors are more appropriate in general and how their performance depends on the interest point detector.

D.G. Lowe [6], proposed a new matching technique using distinctive invariant features for object recognition. Interest points are matched independently via a fast nearest-neighbour algorithm to the whole set of interest points extracted from the database images. Therefore, a Hough transform to identify clusters belonging to a single object has been applied. Finally, least-squares solution for consistent pose parameters has been used for the verification.

Another technique to find correspondences is RANSAC. The most beneficial side of RANSAC is the ability of jointly estimating the largest set of mutual compatible correspondences between two views. Zhang and Kosecka [7] demonstrate the shortcomings of RANSAC when dealing with images containing repetitive structures. The failure of RANSAC in these cases is due to the fact that similarity measure is used to find matching based only on feature descriptor and, with repetitive structures, the chosen descriptors can change dramatically. Therefore, the nearest neighbor strategy is not an appropriate solution.

There are two levels to measure the images similarity which are patch and image levels. In the patch level, the distance between any two patches is measured based on their descriptors. In the image level, the overall similarity between any two images is calculated which in most cases contain many patches.

The Minkowski-type metric has been used to measure the distance between patches in most of researches. The Minkowski metric is defined as in (1):

$$D(X, Y) = \left(\sum_{i=1}^P |X_i - Y_i|^r \right)^{1/r} \quad (1)$$

when $r = 2$, it is the Euclidean distance (L2 distance), and it is the Manhattan distance (L1 distance) when $r = 1$ [8].

In the approach proposed in the present paper both local and global features are considered simultaneously. We try to retain the locality of the features advantages in addition to preserving the overall layout of the objects. The similarity between the local features has been used in conjunction with the topological relations between them as a global feature of the object.

In this paper, the approach presented in [9,10] is modified to be scale invariant. In addition, intensive experiments are conducted mainly focused on images with different resolutions as the objective of the modified algorithm to be more scale invariant. The images contain a duplication of the same object which reflects the scope of work (dealing with repeated structure).

This paper is organized as follows: The proposed scale invariant feature correspondence algorithm is introduced in Sect. 2. Section 3, presents the conducted experiments to evaluate the performance of the modified matching approach. Finally, the conclusions of this work and the recommendations for future work are presented in Sects. 4 and 5, respectively.

2 Proposed Matching Approach

Conventional matching approaches reduce the matching problem to a metric problem. Therefore, the choice of a metric is substantial for the matching of local features. Most approaches depend mainly on finding the minimum distance between features (descriptors) in feature space as shown in (2), where D_{ij} is the distance measure between feature i from the first image and feature j from the second image. X_{ij} is a matching indicator between feature i and feature j , i.e. $X_{ij} = 1$ if feature i in the 1st image is mapped to feature j in the 2nd image and $X_{ij} = 0$ otherwise. Note that $X_{ij} \in \{0, 1\}$.

$$\text{Min } F = \sum_{\forall i,j} D_{ij} X_{ij} \quad (2)$$

Limitations: The similarity measure between features deals with each feature individually rather than a group of features. Consequently, the minimum distance between features can be misleading in some cases and as a result the performance of the algorithm deteriorates. In other words, the minimum distance criterion has no objection for a feature to be wrongly matched as long as it successfully achieves the minimum distance objective.

2.1 Similarity-Topology Matching Algorithm

In [9], a new matching algorithm called ‘‘Similarity-Topology Matching’’ has been proposed. This algorithm pays attention not only to the similarity between features but also to the spatial layout of every matched feature and its neighbors. A new term, describing the neighbourhood/ topological relations between every pair of features has been added $\alpha \sum_{\forall i,j,k,l} X_{ij} X_{kl} P_{ij,kl}$. In addition, another

term has been added to relax the constraints $\beta (Min(m, n) - \sum_{\forall i,j} X_{ij})$ as shown below in (3).

$$Min F = \sum_{\forall i,j} D_{ij} X_{ij} + \alpha \sum_{\forall i,j,k,l} X_{ij} X_{kl} P_{ij,kl} + \beta (Min(m, n) - \sum_{\forall i,j} X_{ij}) \quad (3)$$

Subject to:

$$\sum_{j=1}^n X_{ij} \leq 1 \quad (a)$$

$$\sum_{i=1}^m X_{ij} \leq 1 \quad (b)$$

The second term in (3) represents a penalty term over all pairs of features. $P_{ij,kl}$ is called a penalty matrix. It is used to penalize matching pairs of features i and k in one image with corresponding pairs j and l in the other image if they have different topologies. It is binary and of $(m \times n, m \times n)$ dimension; where m, n are the number of features in the first and the second images respectively. $P_{ij,kl} = 1$ if the features j, l in the second image have different topology when compared to features i, k in the first image. Accordingly, the penalty matrix is calculated by applying the XOR logical operation to the adjacency matrices (AM1, AM2) of the two images as in (4). In XOR, the output is true whenever both inputs are different from each other. For example, if one input is true and the other is false. The output is false whenever both inputs are similar to each other, i.e., both inputs are true or false.

$$P(i, j, k, l) = XOR (AM1(i, k), AM2(j, l)) \quad (4)$$

(α) is called a topology coefficient. It indicates how much the matching algorithm depends on the topology between images. In the experiments, (α) is chosen in a range from 0 to 0.1. The topology coefficient is effective and has a great impact when the interest points are similar to each other. On the contrary, it has almost no impact when the difference of similarities between the interest points is high. (β) is called a threshold coefficient. It indicates how much the matching algorithm depends on the features matching threshold. In the experiments, (β) is chosen in a range from 0 to 0.5. These parameters are determined by cross validations.

Constraints Interpretation: Constraint (a): There exists at most one '1' in every column of x . Constraint (b): There exists at most one '1' in every row of x . The two constraints ensure that every feature in the first image should match to at most one feature in the second image.

2.2 Scale Invariant Similarity-Topology Matching Algorithm

Analysis and Modification. An analysis is done to determine why the algorithm isn't accurate enough in case of the scaling deformation. It is noticed that

the adjacency matrix (AM) of an image is constructed using the neighbourhood idea. In other words, if the distance between any two interest points in the same image is less than a threshold then they are called neighbours to each other. Consequently, the neighbourhood relation between each two interest points depends only on the distance between them, which is not valid specially when dealing with different scales.

The two interest points in Fig. 1 are the same. The algorithm considers them as neighbours to each other in the left image but not in the right image which is counter intuitive, as they are neighbours in both cases.



Fig. 1. Scaling problem example

An amendment is done to the algorithm in order to cope with this limitation. The modification makes the Neighbourhood Relation (NR) depend not only on the distance between the two interest points as in the Similarity-Topology but also on the scales at which the two interest points are detected. Hence, the Neighbourhood Relation (NR) between two interest points i and k in an image is defined as shown in (5).

$$NR = \frac{\text{Distance between two interest points}}{\text{Average scale of the two interest points}} = \frac{d_{ik}}{\text{Avg}(\sigma_i, \sigma_k)} \quad (5)$$

Accordingly, the adjacency matrix is modified and calculated as in (6):

$$AM(i, k) = \begin{cases} 1 & \text{if } NR < \text{Threshold} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where d_{ik} is the Euclidean distance between interest points i and k in the same image spatial domain. σ_i and σ_k are the scales at which the interest points i and k are detected respectively.

Scale Invariant Similarity-Topology Matching Algorithm. Algorithm (1) gives a summary of the modified version of the “Similarity-Topology Matching” approach. This new algorithm achieves superior performance in almost all the experiments specially when the images are exposed to scaling deformations.

Algorithm 1. Scale Invariant Similarity-Topology Matching

Input: A pair of images, topology coefficient (α), and threshold coefficient (β).

1. For every image:
 - (a) Detect interest points (select strongest n);
 - (b) Extract a descriptor for every interest point;
 - (c) Construct adjacency matrix using equations (5) and (6);
2. For every feature (descriptor) in the 1st image: compute the distance between it and all the features in the 2nd image using the Euclidean distance in feature space;
3. Penalize pairs of matched features taking into consideration their adjacency relations;

$$P(i, j, k, l) = XOR(AM1(i, k), AM2(j, l))$$

4. Solve the optimization problem using (7) (features similarity and topological constraints);

Output: List of features correspondences.

This investigated problem has a quadratic-objective function which is subject to linear constraints. It is called a binary (0-1) Quadratic Programming problem. Consequently, the objective function formulated in (3) can be rewritten as in (7):

$$Min F = \sum_{\forall i,j} X_{ij}(D_{ij} - \beta) + \alpha \sum_{\forall i,j,k,l} X_{ij} X_{kl} P_{ij,kl} \quad (7)$$

This optimization problem is solved using IBM ILOG CPLEX Optimization Studio (usually called just CPLEX for simplicity) which is an optimization software package.

3 Experimental Results

3.1 Data-Set

Columbia Object Image Library (COIL-100) has been used in the experiments [11]. COIL-100 is a database of color images which has 7200 images of 100 different objects (72 images per object). These collections of objects have a wide diversity of complex geometric and reflectance characteristics. Consequently, it is the most suitable data-set which can be helpful in the proof of concept of the proposed feature correspondence approach. Figure 2, depicts 10 objects from the Coil-100 data-set.



Fig. 2. Examples of objects from the COIL-100 data-set used for the evaluation

The Challenge. Fifty images representing five objects of the aforementioned data-set are chosen to perform the experiments. These objects with extra synthetic deformations such as rotation, scaling, partial occlusion and heavy noise are used for this purpose. In addition, a duplication of the same object is put in the same image with deformations, but one as a whole and one as parts to make the matching more challenging and to test the principle goal of the new matching strategy. In this case, a feature in the first image has almost two similar features in the second image. Figure 3, shows an example to illustrate the idea. The feature in the first image (left) has two similar features in the second image (right). This raises a question, which one should be matched. This challenge demonstrates the idea of the proposed approach, that rely on the similarity as well as the topological relations between the features as shown in the experiments in the next subsection.

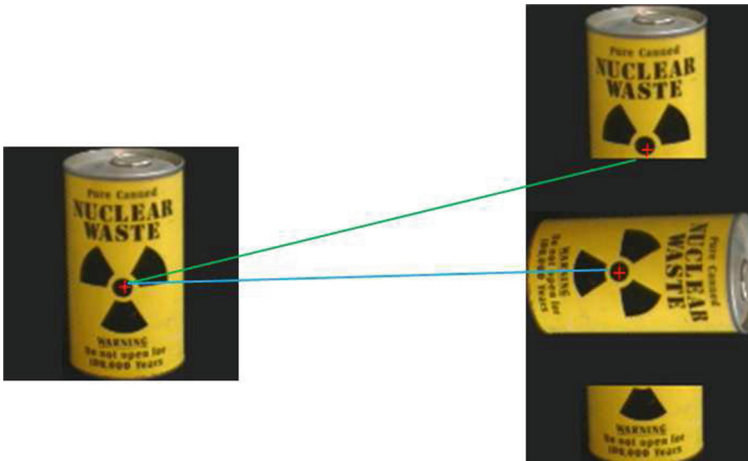


Fig. 3. An illustrative example of the duplication of the same feature

3.2 Experiments

Three different experiments are conducted to test the modification introduced in the “Similarity-Topology Matching” algorithm to make it scale invariant. All

of these tests are done on images having different resolutions. The first test is done between a pair of images with different scales only. The second test is done between a pair of images with different scales as well as a duplication of the same object as parts in the second image. The last test is done like the second experiment but with extra deformations such as rotation and view point changes. These tests are ranged in difficulty from easiest to hardest as shown Table 2.

Features Detection and extraction: the interest points are detected and extracted using SURF (Speeded Up Robust Features) [12]. We demonstrate in [10] that SURF algorithm can be used prior to the proposed matching approach to get more robust feature correspondence.

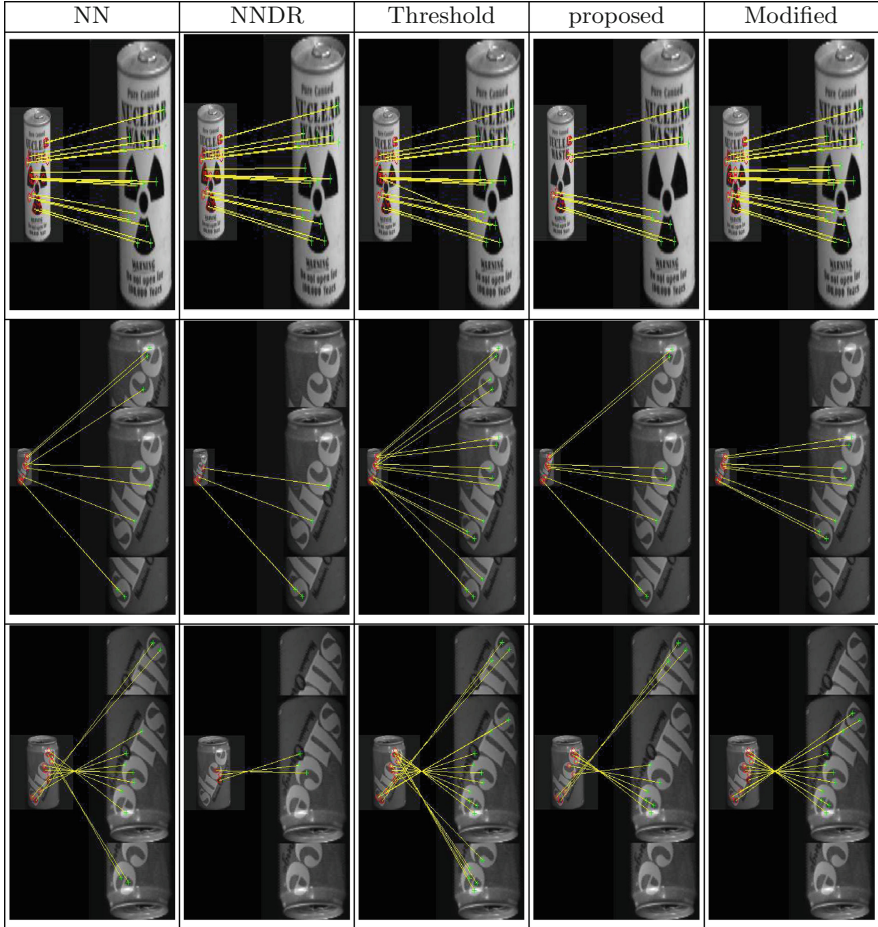
Evaluation criterion: For each pair of images, every interest point in image 1 is compared to all interest points in image 2 according to their descriptors. The detection Rate and the False Positive Rate (FPR) are calculated in order to evaluate the performance. The detection rate R is defined as the ratio between the number of correct matches and the number of all possible matches (number of correspondence). The target is to maximize the detection rate and to minimize the false positive rate.

$$R = \frac{\text{Number of correct matches}}{\text{Number of possible matches within the full instance}}$$

The experiments have been done using three state-of-the-art strategies which are Threshold, Nearest Neighbour (NN) and Nearest Neighbour Distance Ratio (NNDR) in addition to the Similarity-Topology Matching as well as its modified version. In the proposed algorithm and its modified version as well, the values of the topology penalty coefficient and the threshold penalty coefficient are 0.05 and 0.3 respectively. The modified version of the ‘‘Similarity-Topology Matching’’ algorithm has better performance. It is considered more robust specially under the scale deformations. As shown in Table 1, the Modified-Version of the algorithm not only has higher detection rate (0.65), but also it almost eliminates the false matches (0.01) which is more important specially in the localization problem.

Table 1. The experimental results summary

Matching strategy	Detection rate	FPR
NNDR	0.40	0.04
NN	0.48	0.13
Threshold	0.55	0.28
Similarity-topology	0.46	0.08
Modified-version	0.65	0.01

Table 2. Scale invariant feature correspondence examples

4 Conclusions

In this paper, an improved scale invariant feature correspondence algorithm which depends on the “Similarity-Topology Matching” algorithm has been introduced. In this approach, both local and global features are considered simultaneously and a set of control parameters is employed to tune the performance by adjusting the significance of global vs. local features. The major contribution of this research is depending not only on the distance between the two interest points but also on the scale at which the interest points are detected to decide the neighbourhood relations between every pair of features. Three different tests focusing on scaling deformations have been conducted. From the experimental results, it is noticed that the number of correctly matched features is increased.

In conclusion, the modified version of the “Similarity-Topology Matching” algorithm has superior performance specially when the images have been exposed to scaling deformations.

5 Future Work

After the proof of concept of the aforementioned approach has been verified, a lot of work remains to be done in order to generalize the local features matching approach and achieve high degree of robustness and computational efficiency. First, a preprocessing step is required to automatically evaluate the parameters values (alpha, beta). Second, an optimization of the algorithm to be more computationally efficient should be made without any loss in the algorithm accuracy as this algorithm may be used in real-time applications. Finally, applying this approach in a particular robot application such as mobile robot localization. The proposed approach can be used in conjunction with other approach [13] which depends on wifi-signals to determine the location of a mobile robot (such as KheperaIII) in indoor limited areas.

Acknowledgments. This research has been supported by the Ministry of Higher Education (MoHE) of Egypt through a Ph.D. fellowship. Our sincere thanks to Egypt-Japan University for Science and Technology (E-JUST) for guidance and support. I wish to express an extended appreciation to Prof. Mohamed Hussein for his fruitful discussions and helpful suggestions.

References

1. Zitova, B., Flusser, J.: Image registration methods: a survey. *Image Vision Comput.* **21**(11), 977–1000 (2003)
2. Szeliski, R.: *Computer Vision: Algorithms and Applications*. Springer, London (2011)
3. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) *ECCV 2002, Part I. LNCS*, vol. 2350, pp. 128–142. Springer, Heidelberg (2002)
4. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A comparison of affine region detectors. *Int. J. Comput. Vision* **65**(1–2), 43–72 (2005)
5. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(10), 1615–1630 (2005)
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* **60**(2), 91–110 (2004)
7. Zhang, W., Kosecka, J.: Generalized RANSAC framework for relaxed correspondence problems. In: *Third International Symposium on 3D Data Processing, Visualization, and Transmission*, pp. 854–860. IEEE (2006)
8. Liu, Y., Zhang, D., Lu, G., Ma, W.Y.: A survey of content-based image retrieval with high-level semantics. *Pattern Recogn.* **40**(1), 262–282 (2007)

9. El-Mashad, S.Y., Shoukry, A.: A more robust feature correspondence for more accurate image recognition. In: 2014 International Conference on Computer and Robot Vision (CRV). IEEE (2014)
10. El-Mashad, S.Y., Shoukry, A.: Evaluating the robustness of feature correspondence using different feature extractors. In: 2014 International Conference on Methods and Models in Automation and Robotics. IEEE (2014)
11. Nayar, S., Nene, S., Murase, H.: Columbia object image library (coil 100). Technical Report CUCS-006-96, Department of Computer Science, Columbia University (1996)
12. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
13. Elbasiony, R., Gomaa, W.: WiFi localization for mobile robots based on random forests and GPLVM. In: 2014 13th International Conference on Machine Learning and Applications (ICMLA), pp. 225–230, December 2014