

ROBUST FEATURE CORRESPONDENCES FROM A LARGE SET OF UNSORTED WIDE BASELINE IMAGES

Yanpeng Cao and John McDonald

Department of Computer Science
National University of Ireland, Maynooth, Ireland
{ycao, johnmcd}@cs.nuim.ie

ABSTRACT

Given a set of unordered images taken in a wide area, an effective solution is proposed for establishing robust feature correspondences among them. Two major improvements are made in our work as follows: firstly, a robust technique is proposed for the self-organization of a large number of images without spatial orderings; secondly, a novel wide-baseline matching approach is developed to obtain good correspondences over images taken from substantially different viewpoints. The output consists of many sets of reliable pair-wise feature correspondences which are essential in various computer vision applications. Realistic experiments were carried out to evaluate the performances of the proposed method by using a large amount of images captured from our university's campus.

Index Terms— feature correspondence, wide baseline matching, image self-organization

1. INTRODUCTION

Finding feature correspondences among different views is a fundamental problem in many computer vision tasks, such as structure from motion, camera pose estimation, object recognition, and so on. For correspondence results to be robust and accurate, two spatially connected images are needed. In a time-indexed video, enough spatial consistency can be guaranteed due to the short time interval between frames (e.g. 40 ms). Unfortunately, this is not the case for a sparse set of images taken from arbitrary viewpoints in a wide area. Even after two overlapped images are correctly determined, it is also a challenging task to obtain high-quality correspondences between them in the presence of substantial viewpoint and illumination changes.

In this paper, we present an effective method to establish high-quality feature correspondences for a set of images without spatial orderings. The first contribution is the introduction of an effective solution for image self-organization via the use of Speeded-Up Robust Feature (SURF) [1]. After a set of SURF descriptors are extracted in an image, a global representation is defined for robust image similarity evaluation. A

good strategy is proposed to use the fewer SURF features detected in a coarser level to improve computation efficiency. A number of query images are automatically selected to improve the results of image self-organization. The second contribution of this paper is that we propose a novel method to support wide baseline feature matching. It produces a large amount of robust feature correspondences which significantly enhance the connectedness of input images.

The remainder of the paper is organized as follows. Firstly, the image self-organization technique is presented in Section 2. Then, Section 3 explains the method for wide baseline matching in details. Experimental evaluations are reported in Section 4. Finally, the conclusion and future research are discussed in Section 5.

2. IMAGE SELF-ORGANIZATION

Given a large set of images taken from various unknown viewpoints, we applied the Speeded-Up Robust Features (SURF) for robust feature detection in each frame. Then a sparse set of well-conditioned key frames are selected to set up a global reference. In [2], a good guideline was provided for key frame selection. Starting from a randomly captured image, the challenge is how to relate it to the selected key frames. More specifically, we need to identify which key frames it overlaps with. We proposed to solve this problem using a two-step approach. Firstly a global representation is defined over each image, and then the images are related to the key frames based on the similarity evaluation of their global representations. It has been demonstrated that the extracted 64-dimensional SURF descriptor set is a good choice to represent an image [1]. However the similarity evaluation based on SURF descriptor sets has two major drawbacks. Firstly, it's very time consuming to search all the possible matches in two large sets of descriptors. Secondly, it's difficult to find a proper ratio check threshold to determine correct matches.

In this work an improved representation is defined over each image by following the works of Grauman and Darrell [3]. Here we consider two sets of SURF descriptors $D_1 =$

$\{d_1, \dots, d_{n_1}\}$ and $D_2 = \{d_1, \dots, d_{n_2}\}$ derived from two images (where d_i is a 64-dimensional descriptor), a global representation is defined over each descriptor set as:

$$\Psi(D_i) = [H_0(D_i), H_1(D_i), \dots, H_L(D_i)] \quad (1)$$

where H_i is a histogram vector that records the number of descriptors that fall into 64-dimensional bins (corresponding to the 64-dimensional SURF descriptor) of side length 2^i . The bins in the finest level 0 are small enough that each feature descriptor falls into its own bin, while all descriptors fall into one single bin at the coarsest level L . Then the similarity between two images is measured by comparing their corresponding histograms at different levels as:

$$S(D_1, D_2) = \sum_{i=1}^L \omega^i (\chi_{i+1}(D_1, D_2) - \chi_i(D_1, D_2)) \quad (2)$$

where χ_i is the overlap of histograms at level i , and ω^i is the weight coefficient which gives more credits to the overlap found in a smaller size bin.

Compared to the descriptor set D_i , Ψ_i is a better global representation. Two sets of descriptor can be inserted in parallel into some pre-structured (not pre-created) multi-size bins, and their similarity is immediately obtained by counting the number of descriptors falling into the same bins. In this way, computational complexity is largely reduced from a set-to-set matching (polynomial dependence on the descriptor number) to two set-to-bin matching (linear dependence on the descriptor number). Also the pyramid matching procedure provides a good understanding of the similarity evaluation results between two sets of descriptors (e.g. number of matches found under different selection criteria).

A coarse-to-fine searching scheme is also implemented in our work for better efficiency. Image is downsized to the lowest resolution (1/8 of the original size) and features are detected. In this level, similarity evaluations are performed between the query image and all the key frames. If enough evidences are gathered at this resolution, the self-organization will stop there to save computation cost. Otherwise, we increase the image resolution and repeat the similarity evaluation procedures until the full resolution is reached. If the similarity variation is still too low, the new image is defined as an outlier (image covers sky, trees, or road). Due to the spatial continuity, an image should always have high similarity with two adjacent key frames. This constraint is imposed to eliminate false results.

Over time a number of query images are selected to update the self-organization scheme. If a new image cannot be correctly organized at its lowest resolution, its global representation defined over the detected SURF descriptor set is recorded. If lots of overlaps start to appear in bigger size bins (it means enough similarity can be only found when we loose the selection criteria), we also take that image in account. When a large amount of images are recorded in the

scheme, we will increase the sample rate based on image distribution for quick image matching and self-organization.

3. WIDE BASELINE MATCHING

After two spatially related images are correctly determined, a number of putative SURF feature correspondences are computed through a distance ratio test scheme [4]. More specifically, a pair of SURF features is considered as a correspondence if the distance ratio between the closest match and the second closest one is below some predefined threshold. However in a wide baseline situation, this scheme fails to produce good correspondences between two views due to the large viewpoint and illumination changes. In this paper we present a novel approach to establish a large number of robust feature correspondences under wide baseline circumstances. The key idea is that for each SURF features detected in the first frame, we restrict the correspondence search in the second frame so that it's easier to identify a distinctive match there. Our approach includes two major improvements. Firstly, the use of SURF detector allows robust and efficient feature detection and matching in the presences of large viewpoint and illumination changes. Also, this local feature detector brings more robustness to view occlusions compared to the methods based on invariant region extraction [5]. Secondly, we impose two general constraints to replace the more restrictive homography constraint [6]. Therefore, the applicability is wider.

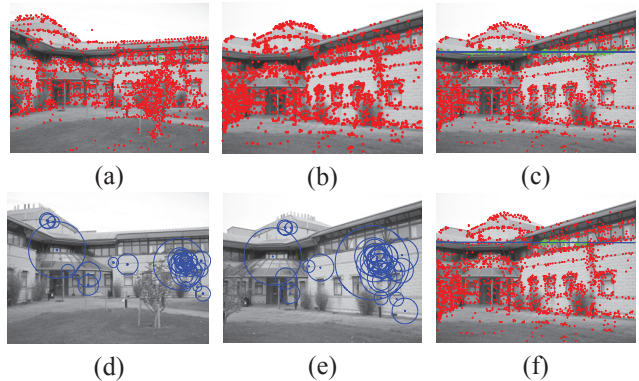


Fig. 1. An example of restricted feature correspondences search. (a)-(b) Detected SURF features in the first and second frames respectively (the highlighted point is a query feature); (c) survived candidates after considering the epipolar constraint; (d)-(e) topological relationships between SURF features in the first and second views respectively; (f) final candidates after considering the topological constraints.

The first constraint to restrict correspondence search is the epipolar constraint [7]. Considering a feature detected in the first frame (see Fig. 1 (a)), its corresponding match must lie along the epipolar line in the second image. This point-to-line mapping relationship is fully encapsulated in the 3×3 fundamental matrix [7]. As a consequence, the possible matches

are effectively restricted within a small strip around the epipolar line, as shown in Fig. 1 (c). In this way, the number of candidate matches can be dramatically reduced from thousands to hundreds. However, there are still many false matches consistent with the estimated fundamental matrix. They cannot be removed by only checking the epipolar constraint. Here we introduce the second constraint, a topological constraint, to solve the problem. At each detected feature the SURF scheme gives three important components: position, descriptor, and scale. The scale component represents in which circle region the SURF feature is generated. Consider two pairs of SURF feature matches $\{P_i^1, P_i^2\}$ and $\{P_j^1, P_j^2\}$. If the corresponding circle region Ω_i^1 and Ω_j^1 overlapped in the first image, the circle region Ω_i^2 and Ω_j^2 should also overlap in the second view (see Fig. 1 (d)-(e)). This useful topological constraint can be effectively used to eliminate the false matches escaped the epipolar constraint checking (see Fig. 1 (f)). Unlike the homography constraint only works for planar mapping, these two constraints are more general and hold true in more cases.

The objective of our approach is twofold: firstly, to produce robust correspondences (qualitatively); secondly, to increase the number of correspondences (quantitatively). This is achieved through following procedures:

Initialization: Extract a set of putative feature correspondences and calculate the F-Matrix using RANSAC. In this step, a strict ratio test [4] is performed (we set distance ratio threshold at 0.65) to make sure the epipolar geometry can be correctly recovered. A number of feature matches $\{\bar{P}_k^1 \leftrightarrow \bar{P}_k^2\}$ ($k = 1 \dots K$) consistent with the F-Matrix are defined as the initial set of correct correspondences.

Growing correspondences:

(1) For a detected feature P_k^1 in the first view, an epipolar line is defined in the second image. Only the SURF features appeared within a small band around the line (± 15 pixels) are considered for matching. Let the survived candidates after checking the epipolar constraint be $\{P_n^2\}$ ($n = 1 \dots n$).

(2) Test the topological relationships between P_k^1 and $\{\bar{P}_k^1\}$ ($k = 1 \dots K$) in the first image and remove those candidates who violate these relationships in the second view. Let the survived candidates after checking the topological constraint be $\{P_m^2\}$ ($m = 1 \dots m, m \leq n$).

(3) Test the similarity level and distance ratio between the query feature P_k^1 and the candidate features $\{P_m^2\}$. If a distinctive match P_k^2 ($P_k^2 \in \{P_m^2\}$) can be found for P_k^1 , the resulting correspondence is added into the correct correspondences set. Otherwise, repeat above steps for the next detected SURF feature in the first image.

4. EXPERIMENTAL RESULTS

We firstly test the performance of the proposed methods for image self-organization. Over 1000 frames were recorded in both indoor and outdoor environments within the campus at National University of Ireland, Maynooth. The raw inputs

from camera are color images, while SURF implemented in our method only takes in gray-level images. Some representative images are demonstrated in Fig. 2.



Fig. 2. Some representative frames. Many challenging images were captured for testing our method (e.g. large illumination changes, building outlook got blocked, totally irrelevant images)

Full-size images (2272×1704 pixels) were downsized three times (1/2, 1/4, and 1/8) and SURF features were detected at each level. Three different experiments were carried out as follows:

(1) Accuracy test without update. 500 inlier frames (images covering campus buildings) were captured at different times during the day. Many images contained occlusions such as trees, vehicles, and pedestrians. Ground truth about where the images were taken is obtained through human observation. Another 30 key frames were manually selected to set up a global reference. Then image self-organization was performed following the method described in Section 2. In this first experiment we ran the test without any updating (no additional images were recorded during the test) and result is reported in Tab. 1.

	284×213	568×426	1136×852	2272×1704
Positive	56	278	87	44
Correct	50	253	75	40
Note: Accuracy rate = 83.6% , False positive = 35 (7%)				

Table 1. Image organization results without updating

(2) Accuracy test with update. This time we used same image dataset in the Test 1 for image organization, while allowing the method automatically store new images for updating. We divided 500 query frames into 2 subsets (250 frames each) and processed them sequentially to show the effect of updating. The result is shown in Tab. 2. Obvious improvements were noticed when more images were saved for updating.

(3) Robustness test in presence of outliers. 100 outlier frames were recorded, which contains images of trees (25 frames), road and vehicle (25 frames), indoor office and people (25 frames), other buildings (25 frames). The updated image self-organization method from the Test 2 was used for this evaluation. The result is reported in Tab. 3.

Next we tested the proposed matching method on several wide baseline image pairs (see Fig. 3). Our method was compared with the matching scheme based on distance ratio test

(the distance ratio threshold was set at 0.8). The number of putative correspondences and correct ones were counted manually. The results and evaluations were reported in Fig. 3 and Tab. 4, in which we can see significant quality improvements of the resulting correspondences (much more correct correspondences were generated).

First 250 images

	284 × 213	568 × 426	1136 × 852	2272 × 1704
Positive	78	97	43	22
Correct	72	91	40	20

Note: 53 query images were recorded
Accuracy rate =89.2%, False positive = 10 (4%)

Second 250 images

	284 × 213	568 × 426	1136 × 852	2272 × 1704
Positive	103	90	37	17
Correct	99	87	35	17

Note: 26 query images were recorded
Accurate rate =95.2%, False positive = 3 (1.2%)

Table 2. Image organization results with updating

	Tree	Vehicle	Indoor	Building
Correct	25	25	23	20

Table 3. The result of robustness test in presence of outliers

Image pairs	Distance ratio test	Our proposed approach
Engineer. Bd (a)(b)	31(T)/76 (N)	223(T)/236(N)
Hulme Bd (c)(d)	34(T)/64(N)	245(T)/260(N)
Bio. Bd(e)(f)	47(T)/83(N)	227(T)/249(N)
Callan Bd (g)(h)	35(T)/62(N)	229(T)/244(N)

Table 4. The result of wide baseline matching and evaluation (N:number of generated correspondences, T:number of correct ones)

5. CONCLUSIONS AND FEATURE WORKS

In this paper, we propose an effective method to establish robust feature correspondences for a large set of unsorted images taken in a wide area. The key components include an effective method for wide baseline feature matching and a robust image self-organization solution where the unordered images are spatially related. In the future we plan to further evaluate the method in some more complex and large-scale environments. Eventually the method will be adapted for applications such as user navigation, augmented reality, and intelligent robotics.

6. ACKNOWLEDGE

Research presented in this paper was funded by a Strategic Research Cluster grant (07/SRC/I1168) by Science Foundation Ireland under the National Development Plan. The authors gratefully acknowledge this support.

7. REFERENCES

- [1] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, "Speeded-up robust features," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.
- [2] Marc Pollefeys, Luc Van Gool, Maarten Vergauwen, Frank Verbiest, Kurt Cornelis, Jan Tops, and Reinhard Koch, "Visual modeling with a hand-held camera," *Int. J. Comput. Vision*, vol. 59, no. 3, pp. 207–232, 2004.
- [3] Kristen Grauman and Trevor Darrell, "The pyramid match kernel: discriminative classification with sets of image features," *IEEE International Conf. on Computer Vision*, 2005.
- [4] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [5] Vittorio Ferrari, Tinne Tuytelaars, and Luc Van Gool, "Wide-baseline multiple-view correspondences," *IEEE Conf. on Computer Vision and Pattern Recognition*, 2003.
- [6] Philip Pritchett and Andrew Zisserman, "Wide baseline stereo matching," *IEEE International Conf. on Computer Vision*, 1998.
- [7] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, New York, NY, USA, 2003.

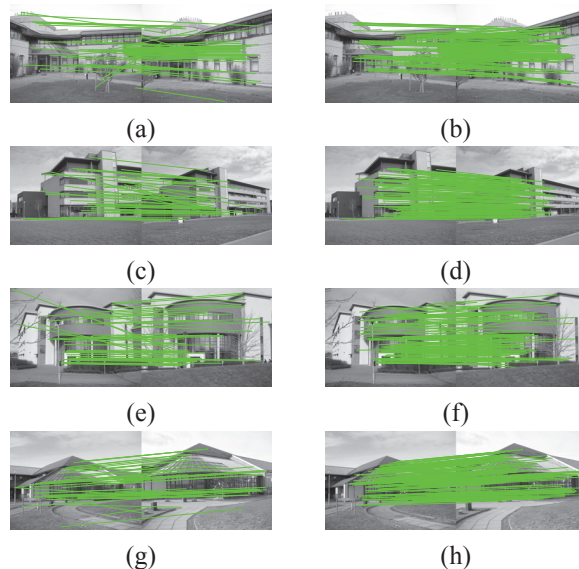


Fig. 3. (a), (c), (e), (g) results based on distance ratio test; (b), (d), (f), (h) results of our proposed wide baseline matching approach.