H-033

Fast and Robust Geometric Verification for Local Descriptor based Image Matching

Ruihan Bao Kyota Higa Kota Iwamoto
Information and Media Processing Laboratories, NEC Corporation

1. Introduction

Image matching is of great importance for applications such as image retrieval and object identification. For image matching, methods using local descriptors such as SIFT and SURF receive the most attentions due to its robustness and relative low computational cost compared to template matching based methods. Conventional methods using local descriptors contain two major steps called feature matching and geometric verification, in which geometric verification is usually carried out using RANSAC.

Recently for real-time applications, binary descriptors such as BRIEF [4] and BRIGHT [2] are proposed which significantly improve the speed of feature matching. This improvement, however, signifies the necessity to improve the speed of geometric verification methods using RANSAC because the time consumed by RANSAC has dominated the total processing time. This is because RANASAC is an iterative method that performs badly when the inlier ratio is below 50%, which usually happens when two images are unrelated.

In this paper, we propose a fast and robust geometric verification method for image matching. Different from RANSAC in which only the coordinates of feature matches are used, we utilize scale and orientation in addition to the coordinates, which can be readily obtained from local descriptors, to check the geometric consistency between feature matches.

2. Proposed Method

In order to perform fast geometric verification, we apply a non-iterative 4D bin voting method on the transformation parameters computed from feature matches using keypoint scale, orientation and coordinates. Then we use the maximum votes in a 4D bin as the score to represent geometric consistency between a query image and database image. Although originally we focused on the improvement of the speed of the geometric verification, experiments showed that the proposed method not only work much faster than RANSAC, but also achieve better performance in identification accuracy.

Figure 1 shows our proposed method for image matching. When a query image comes, keypoints are detected around the images and local descriptors are extracted around those keypoints. Once local descriptors are extracted, they are matched with those saved in the database based on the vector distance in the feature space. Furthermore, ratio test are carried out so that only keypoints with distinct features (descriptors) pass the process.

After feature matching, each matched pair votes for a hypothetical image (or object) transformation model based on the parameters computed from keypoint coordinates, scale and rotation. This is based on the fact that all keypoints of an image (object) experience the same scaling, rotation and translation. Therefore, if the transformation parameters for each keypoint are obtained, it is expected that the transformation parameters estimated from keypoints of correct feature matches form a cluster (corresponding to correct transformation model) in the parameter space.

The hypothetical transformation parameters (i.e. scaling, rotation and translation) estimated from feature matches *(p,q)* can be computed by the following equations,,

$$\rho = s_p / s_q \ , \tag{1}$$

$$\sigma = \theta_p - \theta_q \ , \tag{2}$$

$$[x_\tau \ y_\tau]^T = [x_p \ y_p]^T - \rho R(\sigma)\left[x_q \ y_q\right]^T \quad . \tag{3}$$

Here, ρ, σ and $[x_\tau \ y_\tau]^T$ are parameters corresponding to scaling, rotation and translation between two keypoints. $R(*)$ is the rotation matrix, s_p, s_q are the scale of keypoint p and q, θ_p,θ_q are the orientation and $[x_p, y_{p,}]^T$, $[x_q, y_q]^T$ are the coordinates of the keypoints.

Once associated transformation parameters are computed, we perform clustering on those parameters. Similar to [1], where a fast grid voting method is carried out in 2D space to cluster object center, we apply a 4D bin voting method that collects transformation votes not only from translation (equivalent to object centers in [1]), but also from both scaling and rotation computed by (1) and (2). Specifically, when a set of hypothetical transformation parameters are computed from a feature match using equation (1), (2) and (3), the feature match immediately votes for a 4D bin by quantizing the transformation parameters. In practice, the translation bin is equally spaced between image height and width, the rotation bin is equally spaced between $[0, 2\pi]$, and scaling bin is equally partitioned in

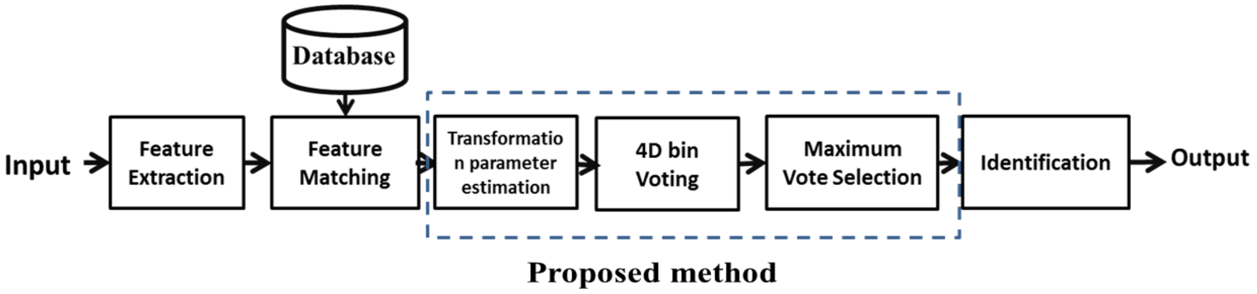


Figure 1 Overview of the proposed method

Copyright © 2014 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.

logarithm space.

Then, different from [1], after all feature matches vote in the parameter space, the bin containing maximum votes are selected and the number of votes for that bin is served as the score for geometric consistency. In order to illustrate the logic of this method, we define a joint probability $p_i(I_k,T)$ representing the probability of matching database image I_k with query image I,

$$p_i(I_k,T) = n_{s,k}/N_{i,k}V_s \ , \qquad (4)$$

Where I_k is the k-th database image, $n_{s,k}$ is the number of vote for s-th bin given k-th database image, T is the transformation model, $N_{i,k}$ is the number of feature matches between image i and k, V_s is the bin size of the s-th bin. Since V_s is a the same for all the bin, taking maximum votes $n_{s,k}$ through all 4D bins (i.e. among all s) is equal to find the maximum possible transformation model T_s with database image I_k (i.e. $\max_s p_i(I_k,T_s)$).

3. Experiment Results

We evaluated the proposed method using three test dataset from MPEG evaluation [3] (Fig.2), including (a) Graphics and texts (e.g. CDs/books/documents), (b) common objects and (c) buildings. Dataset (a) consists of planar objects while datasets (b) and (c) consist of 3D objects. Each dataset contains matching pairs which depict the same objects, and non-matching pairs which depict different objects. For each dataset, we changed the threshold of matching score and evaluated TPR (true positive rate) and FPR (false positive rate) under the threshold.

We compare the proposed method with methods using RANSAC with homography (denoted by RANSAC_HOMOG) and fundamental matrix (denoted by RANSAC_FUND) model. Moreover, in order to better show the effectiveness of the proposed method over RANSAC, we implemented a binary descriptor called BRIGHT [2] in the experiment because the time spent on geometric verification for such method plays an important role in total processing time.

Fig.3 shows the identification results for all three dataset. For the Graphics and Text dataset (a), it shows that the proposed method slightly outperform RANSAC with homography and fundamental matrix. This is because that RANSAC works well with 2D objects. For the common object (3D objects) dataset (b) and building dataset (c), it shows that the proposed method outperforms the conventional methods by a significant margin. For instance, at 5% FRR rate, the proposed method improves the TPR by 7.69% for common object dataset (b) and 7.41% for building dataset (c).

Fig.4 shows the computational time between the proposed and conventional methods. We plot matching time and geometric verification time, respectively. It shows that the proposed method achieved 43.6x speed up for matching pairs in geometric verification (2.99x in total processing time combined with feature matching), 535.3x speed up for non-matching pairs (5.73x speed up in total processing time).

4. Conclusion

We proposed a fast and robust geometric verification method for local descriptor based image (object) matching. We evaluated the proposed method on public MPEG dataset containing planar and 3D objects taken from different viewpoints, and experiment results show that our method not only reduces the processing time (2.99x-5.7x) but also achieves higher matching accuracy (in both recall and precision) compared with conventional RANSAC based methods.

[1] K. Higa, K. Iwamoto, and T. Nomura, "Multiple Object Identification using Grid Voting of Object Center Estimated from Keypoint Matches," Proceedings of ICIP, pp. 2973-2977, 2013.

[2] K. Iwamoto, R. Mase, and T. Nomura, "BRIGHT: A Scalable and Compact Binary Descriptor for Low-Latency and High Accuracy Object Identification" Proceedings of ICIP, pp. 2915-2919, 2013.

[3] MPEG, "Call for Proposals for Compact Descriptors for VisualSearch", ISO/IEC JTC1/SC29/WG11 N12201, July 2011.

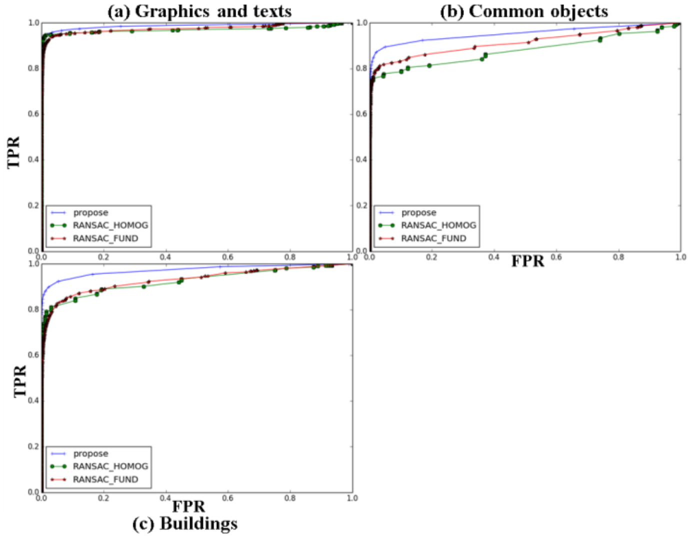
[4] Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., & Fua, P. (2012). BRIEF: Computing a local binary descriptor very fast. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34(7), 1281-1298.

(a) graphics&texts (3,000 matching, 30,000 non-matching)

(b) common objects (2,550 matching, 25,500 non-matching)

(c) buildings (4,005 matching, 48,675 non-matching)

Figure 2 Examples of the dataset



(c) Buildings

Figure 3 Performance on MPEG datasets

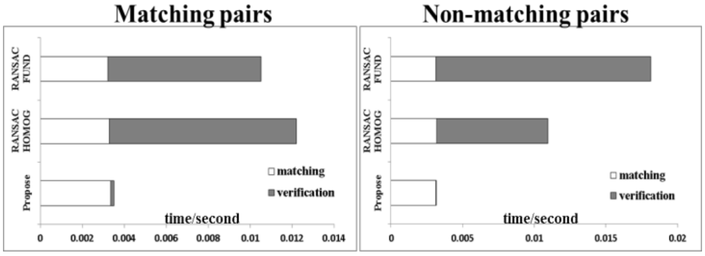


Figure 4 Computation time

Copyright © 2014 by
The Institute of Electronics, Information and Communication Engineers and
Information Processing Society of Japan All rights reserved.