Pattern Recognition Letters 68 (2015) 76–82

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Feature description using local neighborhoods☆

Man Hee Lee [a], Minsu Cho [b], In Kyu Park [c,*]

[a] Computer Graphics Research Laboratory, Electronics and Telecommunications Research Institute, Daejeon 305-700, Republic of Korea
[b] WILLOW Project-Team, Département d'Informatique de l'Ecole Normale Supérieure, ENS/Inria/CNRS UMR 8548, Paris, France
[c] Department of Information and Communication Engineering, Inha University, Incheon 402-751, Republic of Korea

ARTICLE INFO

Article history:
Received 13 September 2014
Available online 3 September 2015

Keywords:
Feature description
Local neighborhoods
Similarity function
Graph matching
MRF

ABSTRACT

Feature description and matching is an essential part of many computer vision applications. Numerous feature description algorithms have been developed to achieve reliable performance in image matching, *e.g.* SIFT, SURF, ORB, and BRISK. However, their descriptors usually fail when the images have undergone large viewpoint changes or shape deformation. To remedy the problem, we propose a novel feature description and similarity measure based on local neighborhoods. The proposed descriptor and similarity is useful for a wide range of matching methods including nearest neighbor matching methods and popular graph matching algorithms. Experimental results show that the proposed method detects reliable matches for image matching, and performs robustly to viewpoint changes and shape deformation.

1. Introduction

Many computer vision techniques such as structure from motion [1], visual SLAM (simultaneous localization and mapping) [17] use multiple images or video frames for their input. To estimate camera motion or generate 3D maps in those applications, they usually extract local features, and then match/track them over a sequence of multiple images. Consequently, the performance is highly affected by the accuracy of feature extraction and matching algorithms used. Furthermore, a computationally efficient feature extraction is required to process a large number of high resolution images. During the last decade, excellent methods such as SIFT (scale-invariant feature transform) [21] and SURF (speeded up robust features) [5] have been developed and widely used for feature detection and description. Although they show robust performance when images have undergone in-plane rotation, scale variation, and illumination changes, they fail when the images involve severe variations caused by out-of-plane rotation of cameras, large viewpoint changes, or non-rigid object deformation. Note that those variations are often observed in real-world scenes, and thus should be handled by further research. In this paper, we propose a novel feature description method based on local neighborhoods, which effectively handles significant viewpoint changes and deformation of non-rigid objects. In the proposed approach, initial keypoints are first detected by existing feature detectors such as SURF [5], and then local graphs are constructed for each feature using its neighboring features. For robust matching, similarity between two features is evaluated based on their local graphs. Unlike previous patch-based descriptors [5,21], the proposed method provides a robust similarity function by taking into account neighboring features altogether. It is significantly more efficient than previous high-order graph matching [12] and progressive matching [10] techniques. The preliminary version of this paper was published in [19].

The main contributions of this paper are summarized as follows.

- We introduce a novel feature description based on local neighborhoods, and propose a similarity function that is robust to viewpoint changes and shape deformation.
- We demonstrate effective graph-based feature matching using the proposed similarity function.
- The proposed method is compared with recent popular feature descriptors, and shows superior performance in matching accuracy and efficiency.

This paper is organized as follows. In Section 2, we introduce related work. Section 3 describes the proposed feature description in detail. In Section 4, we describe feature matching based on a MRF formulation. Experimental results are presented in Section 5. We conclude with final remarks in Section 6.

2. Related work

2.1. Feature detection and feature description

Numerous feature detectors have been proposed in the literature. Mikolajczyk and Schmid [23] proposed Harris–Affine feature detector that is invariant to scale and affine transformations. They detect

☆ This paper has been recommended for acceptance by J.K. Kämäräinen.

* Corresponding author:. Tel.: +82 32 860 9190; fax: +82 32 873 8970.
E-mail address: pik@inha.ac.kr, pik63479834@gmail.com (I.K. Park).

http://dx.doi.org/10.1016/j.patrec.2015.08.016

interest points with a Harris–Laplace operator, that is adapted to affine transformations based on the second moment matrix. Agrawal et al. [2] proposed a scale-invariant detector (*i.e.* CenSurE) that is computed over multiple scales on the original image. They utilized the simplified bi-level kernels to achieve fast computation using a slanted integral image. Rosten et al. [25] proposed a fast and high quality corner detector (*i.e.* FAST) based on a segment-test algorithm. They applied a machine learning approach to implement the real-time algorithm, and generalized the detector with a ternary decision tree. Lowe [21] proposed a scale-invariant feature detector based on a multiscale extrema of DoG (difference-of-Gaussians) function. Alcantarilla et al. [4] proposed KAZE feature detection algorithm in a nonlinear scale space using nonlinear diffusion filtering rather than using Gaussian filtering. Dragon et al. [11] proposed NF-Features which represent object regions in the non-textured area.

On the other hand, feature descriptors also have been developed along with feature detectors. Lowe [21] proposed SIFT feature descriptor based on a histogram of oriented gradients. Currently, it is still widely used as the state-of-the-art method with high repeatability and accuracy. However, it requires a high computation cost in matching as the dimensionality of the descriptor is relatively high. Mikolajczyk and Schmid [24] proposed GLOH descriptor using gradient location and orientation histogram by extending the SIFT descriptor, which improves the robustness and distinctiveness using a circular sampling pattern. Bay et al. [5] proposed an accelerated version of SIFT (*i.e.* SURF) using an integral image and an approximated integer Gaussian filter. They do not resize the original image for multi-scale but simply increase the size of the filter. If it is applied to an integral image, Hessian determinant is calculated in a constant time. It is more robust to image noise than SIFT because they also integrate the gradient information within a subpatch. Alahi et al. [3] proposed FREAK descriptor inspired by the retina. They utilized a series of DoG functions over a retinal sampling pattern to generate the binary descriptor. Calonder et al. [7] proposed a fast and accurate descriptor (*i.e.* BRIEF) for real-time matching using short binary strings. They directly computed binary strings from image patches using a Gaussian sampling pattern and evaluated the Hamming distance for fast matching. However, it is neither rotation- nor scale- invariant. Stanski and Hellwich [29] proposed Spider descriptor which is characterized by a constellation of surrounding features.

Rublee et al. [26] proposed a fast feature detection and description algorithm based on the oriented FAST detector and the rotated BRIEF (binary robust independent elementary features) descriptor. Since BRIEF is rotation-variant, they calculated the orientation of FAST features using an intensity centroid scheme. Leutenegger et al. [20] proposed a fast and efficient feature detector and descriptor. They used scale space keypoint detection based on FAST in scale pyramid, and used a circular sampling pattern to generate a binary descriptor by performing simple brightness comparison.

Although these algorithms are robust in challenging conditions such as in-plane rotation and scale variations, most of them fail when images have undergone considerable out-of-plane rotation or shape deformation.

2.2. Object recognition and image matching

Berg et al. [6] proposed an algorithm for deformable shape matching. They defined a cost function to measure the matching quality and the degree of geometric distortion. Duchenne et al. [12] proposed a hypergraph matching algorithm to establish correspondences between two sets of visual features using a tensor based high-order score function. Cho and Lee [10] proposed a progressive matching algorithm that combines graph progression and graph matching. They efficiently update the most plausible target graphs in a Bayesian manner that boosts the matching objective at the subsequent graph matching. Hundelshausen and Sukthankar [16] proposed Descriptor-Nets that significantly improves accuracy in image matching. They construct a network using conventional keypoints as its nodes, and propose an efficient matching algorithm using a hash table. Chen et al. [8] proposed a robust feature matching algorithm using Hough transform and inverted Hough transform. They assume that nearby features on the same object share similar homographs in a transformation space. From this assumption, they cluster features using BPLR (boundary preserving dense local regions) [18] and project corresponding points into Hough space. They cast the task of feature matching into a density estimation problem in the transformed space.

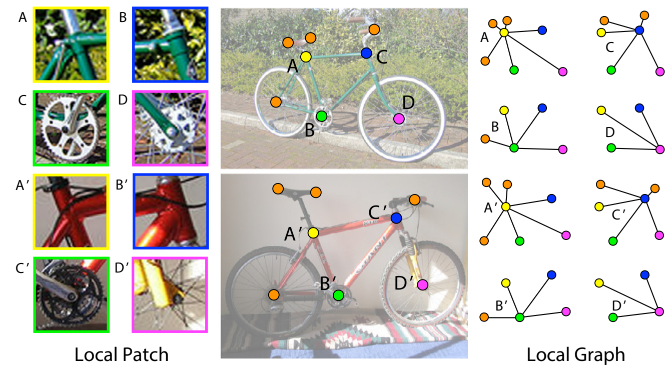

Fig. 1. An example of a patch based descriptor and the proposed local neighborhood representation. (Middle) Input images and initial keypoints. (Left) Local patches at the extracted keypoints. (Right) Local graphs constructed at the keypoint using neighboring features.

The algorithms use conventional feature detection and description algorithms to describe the object by a set of features and to find initial matches. However, such patch-based feature descriptors have inherent disadvantage in representing a feature under significant image variations. The proposed algorithm overcomes the difficulty by representing the feature using a local neighborhood structure of the image.

3. Feature description using local neighborhoods

3.1. Local graphs in image matching

Most image matching algorithms employ a patch-based feature descriptor to find initial correspondences between images. Some of them use a simple appearance-based matching algorithm as an initial step, and then apply geometric constraints to find reliable correspondences. Graph-based representation is widely used for object detection and recognition [14,15], where an object is modeled as deformable constellations of parts. Graph-based feature matching [9,10] usually construct graphs from detected features, and find correspondences using graph matching techniques to handle different objects in the same category or the same object with large view changes correctly. However, many of those algorithms require time-consuming cost minimization techniques for graph matching.

Unlike these previous methods, we propose a feature representation that encodes neighboring features relations. The local neighborhood has a larger supporting region than the patch-based descriptors, and preserves its structure not only for large view changes but also for deformable objects. These feature description can be used for robust matching.

Fig. 1 compares a conventional patch-based descriptor with the proposed local neighborhood representation, where initial keypoints are detected by SURF. The images on the left in Fig. 1 show local patches at the different keypoints. The local patches are inconsistent to each other because the target objects are different, even though they are from objects in the same category. It means that