Information Sciences 442–443 (2018) 173–185

Contents lists available at ScienceDirect

Information Sciences

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

Local multi-feature hashing based fast matching for aerial images

Suting Chen[a,b,*], Yunjiao Shi[a,b], Yanyan Zhang[a,b], Jiaojiao Zhao[c], Chuang Zhang[a,b], Tao Pei[a,b]

[a] CICAEET, School of Nanjing University of Information Science & Technology, Nanjing 210044, China
[b] Jiangsu Key Laboratory of Meteorological Observation and Information Processing, Nanjing University of Information Science & Technology, Nanjing 210044, China
[c] Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, UK

ARTICLE INFO

Article history:
Received 15 May 2017
Revised 15 February 2018
Accepted 16 February 2018
Available online 21 February 2018

Keywords:
Local feature descriptors
Aerial image matching
Hashing
Multi-bit quantization

ABSTRACT

Despite their superior accuracy, existing floating-point feature descriptor based image matching algorithms are too computationally intensive for real world aerial imaging applications, where target images are often of extremely high resolutions. To improve the efficiency of these algorithms, this paper presents a novel low-complexity image matching approach, namely local multi-feature hashing (LMFH). Similar to conventional techniques, LMFH employs a floating-point local multi-feature descriptor for accurate matching, however, the proposed method does not compare descriptors directly. Instead, LMFH projects each high-dimensional descriptor to a compact binary code in Hamming space using a trained hash function and then performs the feature comparison in the mapped domain. Experimental results show that, in terms of the computational cost, LMFH is on pair with the most efficient modern binary descriptors. Furthermore, LMFH has a performance gain which is significantly higher than the state-of-the-art binary descriptors and is comparable with conventional floating-point descriptors in terms of accuracy.

1. Introduction

Image matching, the technique for locating the common regions of two or more images, is an important research topic in the field of computer vision and computer graphics. Image matching is a key step in many applications such as scene reconstruction, video surveillance, remote sensing and image classification [10,32,37]. For aerial imaging, image matching has been extensively used for stitching multiple overlapped aerial images into one high-resolution panoramic image. However, due to the ever increasing resolutions of aerial cameras, existing methods become problematic in handling such a large amount of image data collected in one flight.

Most image matching techniques are based on the technique of local feature matching. Generally speaking, a local feature matching algorithm contains two main steps [5]: feature point detection, and feature descriptors generation. A local feature descriptor is a vector that contains a few image statistics outlining the neighborhood of a feature point. For accurate image matching, feature descriptors should be discriminative and robust. Many local feature descriptors have been proposed in the

* Corresponding author at: Jiangsu Key Laboratory of Meteorological Observation and Information Processing, Nanjing University of Information Science & Technology, Nanjing 210044, China.
E-mail address: sutingchen@nuist.edu.cn (S. Chen).

https://doi.org/10.1016/j.ins.2018.02.036
0020-0255/

literature. SIFT [23] and SURF [6] are two popular examples, both of which are vectors of floating-point numbers. However, calculating and comparing floating-point numbers are relatively slow on a computer, making these descriptors unsuitable for large scale images.

In order to expedite feature descriptor matching, several binary feature descriptors have been proposed to replace floating-point feature descriptors. In addition to their speed advantages, binary feature descriptors require less memory and easier to be matched by measuring Hamming distance. Binary feature descriptors can be generally categorized into two groups [1]. Descriptors of the first group convert the floating-point features to binary form in order to reduce computational cost without significantly compromising their performance. Descriptors of the second group, such as BRIEF [8], ORB [25], BRISK [15], and FREAK [2], obtain the binary strings directly from image patches by measuring the intensity differences between predefined pixel locations. In comparison with floating-point descriptors, binary feature descriptors are faster but less discriminative and robust. Recent binary-based descriptors could outperform the SIFT and the SURF in terms of matching speed and the memory footprint, but the matching accuracy is usually significantly lower than that of the SIFT and the SURF algorithms. For aerial image matching tasks, the accuracy of binary feature descriptors is usually not high enough for application.

Recently, with the research boom of machine learning methods, emerged an interesting and promising idea namely learning to hash [14,18–20,26,31,35,38]. By training abundant samples and learning the hash function, the high dimensional data points are embedded into a similarity-preserved Hamming space with low-dimensional compact binary string, which brings a large improvement in both memory usage and time complexity [17]. In general, current hashing techniques can be divided into two groups, the unsupervised methods [19,20,31] and supervised methods [14,18,26]. Currently, both of the conventional unsupervised and supervised hashing methods are primarily designed for estimating the global features of the images [14,18,20], and are mainly applied in the fields such as image retrieval [18–20,26,31,35,38] and target recognition [14] and target recognition [14].

In this work, based on the idea of learning-to-hash, we develop a novel technique, called local multi-feature hashing (LMFH), for aerial image matching. Different from global representation based hashing, LMFH learns hashing function from local features. In order to make matching more accurate, similarity weight is assigned according to the pair wise labels of local features during the construction of the hash function. Furthermore, LMFH employs alternative iteration and multi-bit quantization strategy for better performance of hash learning.

2. Related work

The conventional local feature descriptors are usually described by a set of floating-point numerical values, these descriptors, such as SIFT and SURF are successfully applied in image matching and registration algorithms. However, these algorithms are usually high in computational complexity and memory requirement for implementing SIFT and SURF algorithms. Thus, it is generally difficult to use the primal SIFT algorithm in some applications, such as real-time tracking or matching with a sequence of images. In order to reduce the memory requirement and improve its efficiency, many advanced approaches have been proposed in the last few years. Combined with SIFT and SVM, a robust feature point matching method is proposed for dynamic aerial image registration [21]. Additionally, a Block-SIFT method is proposed for splitting large images into several blocks. And the method extracts features from the corresponding blocks using the SiftGPU for low memory requirement and speed up the extracting and image matching procedures [28]. Furthermore, an efficient image-matching approach is reported in [4]: Three different sets of local features are extracted by the SIFT algorithm and then k-means clustering approach is performed to achieve an accurate matching of images. Although the above proposed algorithm improved the original SIFT, they usually suffer from high complexity due to the high dimensionality of the descriptors.

According to the characteristics of flight acquisition frequency and the required overlap for remote sensing and aerial images, various improvements are presented [3,11,16,22,34,36]. In order to overcome the restrictions of exterior orientation parameters and improve the matching accuracy, an automatic matching technique for large remote sensing is proposed [22] based on online aerial images matching. To further improve the performance, a robust and effective multi-view dense matching algorithm based on a graph network for high-resolution aerial images has been proposed [34] which utilizes the overlap ratio and the intersection angle between the image pairs. A recognition model which mines the graphlets from aerial images, where the graphlets are small connected subgraphs reflecting both the geometric properties and color/texture distribution of an aerial image is proposed [36] to categories aerial images. Moreover, as the developments of deep learning methods, it has been effectively used to provide great improvement in various aerial fields such as aerial image retrieval and aerial scene classification [3,11,16]. Data-driven deep learning-based approaches have also been proposed [3] by reforming the problem as a classification task. With the pre-trained CNN models as universal feature extractors, and using a support vector machine (SVM), deep convolutional neural network (CNN) for scene classification is proposed [11]. Considering the limited training data, a fusion strategy for integrating multi-layer features of a pretrained CNN model for scene classification is also reported [16].

On the other hand, in order to meet the requirement for aerial images with high resolution, binary-based descriptors are developed for a higher descriptor matching speed. An Accelerated Binary Robust Invariant Scalable Keypoints (ABRISK) algorithm and spatial analysis of corresponding control points [30] are proposed for image registration. The matching performance of ABRISK is inferior to that of SIFT. In addition, all the above improved SIFT methods and binary-based descriptors mainly focus on the aerial image matching techniques rather than the construction of the descriptors. As a popular bi-