Neurocomputing 172 (2016) 48–52

Contents lists available at ScienceDirect

Neurocomputing

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

Visual summarization of image collections by fast RANSAC

Ye Zhao *, Richang Hong, Jianguo Jiang

School of Computer and Information, Hefei University of Technology, Hefei 230009, China

ARTICLE INFO

Article history:
Received 1 November 2013
Received in revised form
2 April 2014
Accepted 28 September 2014
Available online 8 May 2015

Keywords:
Visual summarization
Random sample consensus
Affinity propagation
Nearest neighbor distance search

ABSTRACT

In this paper we propose a novel approach to select a summary set of images from a large image collection by improved Random Sample Consensus (RANSAC) and Affinity Propagation (AP) clustering. It can automatically select a small set of representatives to highlight all the significant visual properties of a given image collection. The proposed framework mainly composes four stages. First, the scale-invariant feature of each image is extracted by Scale Invariant Feature Transform (SIFT). Second, keypoints of two images are matched and ranked based on nearest neighbor ratio. The representative dataset of RANSAC is established by a minimal number of optimal matches. Third, the target homographic matrix is fitted based on the representative dataset. Mismatches are filtered out via the homographic matrix. Finally, summarization is automatically formulated as an optimization framework by AP clustering. We conduct experiments on a set of Paris which is consisting of 1000 images downloaded from Flickr. The results show that the proposed approach significantly outperforms other methods.

1. Introduction

With the rapid development of Web 2.0 and multimedia technology, huge amount of images on the Internet are created and shared by millions of users every day. Given a query, the image sharing websites may return millions of images. It becomes more important how to quickly retrieve and browse images from these scale-large image collections. Visual summarization is a typical approach to help user efficiently browse the image collections.

Visual summarization can be taken on many forms, such as visual summary of media collections, visual summary of landmarks, and visualization of travel trajectories and routes. In 2006, Torniai et al. [1] present an image set browsing system based on the location and image header metadata. In [2], Jaffe et al. propose a framework for automatically selecting a summary set of photos from a large collection of geo-referenced photographs. This summary algorithm is based on spatial information in photo sets, textual-topical patterns and photographer identity cues. Kennedy et al. [3] use the bag of visual words model to generate various and representative images based on the landmark images. In [4], Chen et al. develop a tourist map that could automatically identify popular locations from community photo collections. Snavenly et al. [5] develop a system for interactively browsing and exploring large unstructured collections of photographs in 3D. Zha et al. [6] proposes a method based on the experience of building many successful applications that are based on mining multimedia content analysis in social multimedia context. Based on a large number user tests, Rudinac et al. [7] develop an approach of automatic selection for images, which jointly uses the analysis of image content, context, popularity, visual aesthetic appeal as well as the sentiment gained from the comments posted on the images. In [8], Yang et al. reformulate the summarization problem into a dictionary learning problem by selecting bases which can be sparsely combined to represent the original image and achieve a minimum global reconstruction error. ImageHive generates a summary image to preserve the relationships between images and avoids occluding their significant parts [9]. To lay out images, Wang [10] presents a treemap-based representation for visualizing and navigating image search and clustering results.

Most existing summarization techniques for large-scale image collections usually use SIFT feature matching [11,12]. SIFT combines the scale-invariant features with the descriptors of gradient direction. It is invariant in rotation, scale changes and affine transformations, while it has many mismatches. Another issue of these summarization approaches is the number of clusters needed to specify before. Fewer clusters result in simpler but lower accuracy of the summary set; more clusters result in higher accuracy but redundancy. Traditional clustering methods, such as *K*-means and *K*-mediods, depend strongly on the initial values, and they stop after obtaining a local optimum, which may lose the representative images. Therefore, a novel visual summarization approach is proposed in this paper. This approach adopts the improved RANSAC geometric calibration for image matching to filter mismatches and reduce time cost. This approach also adopts AP [13] clustering method to replace other traditional clustering methods and to generate the summarization automatically without presetting the number of representative images.

The rest of this paper is organized as follows. In Section 2, we introduce the related technology of visual summarization. Section 3

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2014.09.095
0925-2312/

discusses the proposed visual summary method. Section 4 presents our experimental results, and Section 5 concludes this paper.

2. Feature matching

Existing image summarization methods usually abstract SIFT feature points to match images. However, the effectiveness and the efficiency of this matching are low because of the high dimension of descriptors and a large number of false matches. To improve the matching efficiency and shorten computing time, many researchers propose various improved algorithms based on SIFT. Ke and Sukthankar [14] propose PCA-SIFT to reduce the dimension of SIFT descriptors using PCA (principal component analysis). But the methods of extracting robust features were still very slow [15]. Mikolajca et al. [16] propose the GLOH. It turns the checkerboard block partitions in SIFT into radial partitions of concentric circles. PCA is then used to decrease the dimensionality. The GLOH algorithm has better uniqueness and more complex than SIFT. Bay and Tuytelaars suggest SURF by speeding up robust features and using integral images for image convolutions and Fast-Hessian detector [17]. However, the accuracy of SURF is not good. The matches involve many false matches. Considering the role of the geometrical relationship among visual words in the image recognition process, Tuytelaars et al. [18] use RANSAC [19] for the post-processing of the feature matching, which improves the matching accuracy.

RANSAN is a widely used robust estimation algorithm that chooses a certain number of matching points to estimate a target model and calculates the support set of the target model at random. This process can be repeated until the probability of finding a model with a better support than the current best model decreases below a threshold. We choose the largest support set target model to obtain an optimal result. The advantage of the RANSAC algorithm is its reliability, stability, and accuracy. This algorithm also has strong endurance during the vague extraction of image noise and feature points and is very robust and able to eliminate wrong matching points. However, RANSAC becomes computationally expensive when the amount of data is large, particularly when the error-matching rate is high because of the large number of iterations needed before a correct model is found. Time consumption exponentially increases when the outlier rates are high. Therefore, a considerable amount of research has been devoted to address these shortcomings [20]. These efforts show promise but are ineffective in both accuracy and computing time. MLESAC [21] is approximately 5% more accurate and 15% more computationally burdened than RANSAC in almost all configurations. LO-RANSAC [22] is approximately 10% more accurate than RANSAC because of local optimization but is approximately 5% slower than RANSAC. R-RANSAC with the Td;d test [23] has a similar accuracy to RANSAC and is slightly faster despite the increased iterations. R-RANSAC with SPRT [24] is more accurate than RANSAC but is unable to reduce computing time because of adaptive termination. uMLESAC [25] sustains high accuracy in various datasets but requires 1.5 more iterations than RANSAC because of adaptation. Incorrect matches can be effectively filtered by using the RANSAC geometry check. The application of the RANSAC algorithm is extensive because this algorithm increases image matching performance. However, the computational cost of RANSAC is considerable. A large dataset leads to an increase in false matches and produces a complex estimation model. Time consumption will also exponentially increase [12]. This paper presents a fast RANSAC (FRANSAC) algorithm to address the issue.

3. Our approach

The characteristics of image information in social media are large-scale data, uneven quality, complex background, and noise. In this paper, we first extract features from the images in a given collection. Here we adopt SIFT features. After that, we use RANSAC to filter matches. Finally, summarization is automatically formulated as an optimization summary set by AP clustering.

3.1. Fast random sample consensus (FRANSAC)

The cause of high time consumption of RANSAC is the increase in the estimation complexity of the homography matrix when large amounts of sample data and high outlier ratio exist in the observed and the sample dataset, respectively. We find the correct probability of matching change in the nearest neighbor ratio variation by studying the distinctive feature matching. A smaller nearest neighbor distance leads to a highly accurate matching probability. First, the improved RANSAC algorithm ranks the matching pairs based on the distance of the nearest neighbor and next nearest neighbor. The matching pairs, which require a higher probability to be correct, can be placed in front of the queue. The obtained homography is closer to the real homography because we only selected few optimal points as samples. Therefore, time consumption is greatly reduced under the premise of improving the quality. The object model is fitted by RANSAC in the image matching is the image transformation matrix–homography matrix. The homography matrix between images A and B is as follows:

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \tag{1}$$

where

$$H = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{bmatrix},$$

H has 8 degrees of freedom. We can estimate the H through at least 4 pairs of feature points in theory. The upper type can be written as an 8-parameter equation:

$$\begin{cases} x_2 = \dfrac{h_1 x_1 + h_2 y_1 + h_3}{h_7 x_1 + h_8 y_1 + 1} \\ y_2 = \dfrac{h_4 x_1 + h_5 y_1 + h_6}{h_7 x_1 + h_8 y_1 + 1} \end{cases} \tag{2}$$

$$H_1 = -[C^T C]^{-1} C^T L \tag{3}$$

where

$$H_1 = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 \end{bmatrix},$$

$$L = -\begin{bmatrix} x_2 & y_2 \end{bmatrix}^T,$$

$$C = \begin{bmatrix} -x_1 & y_1 & 1 & 0 & 0 & 0 & -x_2 x_1 & x_2 y_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y_2 x_1 & -y_2 y_1 \end{bmatrix}$$

According to this estimation problem, we need to select 4 pairs of matching points randomly. H can be calculated by above formula and optimized by the purified inliers. The minimal number of sample sets n is defined as follows:

$$n = \min\{N_0, \max\{n_0, n_0 \log_2 \mu N_0\}\} \tag{4}$$

where N_0 is the total number of correspondences, and $N_0 \geq 4$, n_0 is the step of sample number, and μ is the scaling factor.

3.2. Nearest neighbor distance search (NNDS)

The NNDS [26] is defined as the ratio between the nearest neighbor distance and the next nearest neighbor distance. The NNDS can be considered the similarity criterion of two images. The