Neurocomputing 144 (2014) 546–552

Contents lists available at ScienceDirect

Neurocomputing

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

Letters

# A robust aerial image registration method using Gaussian mixture models

Chuan Wu [a], Yuanyuan Wang [b], Hamid Reza Karimi [c,*]

[a] Key Laboratory of Airborne Optical Imaging and Measurement, Image Processing Laboratory, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, PR China
[b] Research Institute of Intelligent Control and Systems, Harbin Institute of Technology, Harbin 150001, PR China
[c] Department of Engineering, Faculty of Engineering and Science, University of Agder, N-4898 Grimstad, Norway

ARTICLE INFO

*Article history:*
Received 31 December 2013
Received in revised form
30 March 2014
Accepted 21 April 2014
Communicated by Hongyi Li
Available online 17 May 2014

*Keywords:*
Image registration
Feature detector
Gaussian mixture models
Aerial images

ABSTRACT

Aerial image registration is one of the bases in many aerospace applications, such as aerial reconnaissance and aerial mapping. In this paper, we propose a novel aerial image registration algorithm which is based on Gaussian mixture models. First of all, considering the characters of the aerial images, the work uses a shape feature detector which computes the boundaries of regions with nearly the same gray-value to extract invariant feature. Then, a Gaussian mixture models (GMM) based image registration model is built and solved to estimate the transformation matrix between two aerial images. Furthermore, the proposed method is applied on real aerial images, and the results demonstrate the improved performance of the proposed algorithm.



## 1. Introduction

Image registration is one of the hottest research topics since it is one of the bases in many computer vision and image processing applications. The main objective of image registration is to geometrically align the pixels in sensed image which are one-to-one corresponding to the pixels in the reference image. The main differences between these two images are introduced due to the different imaging conditions. And image registration is an important step in many image analysis tasks, such as image fusion and multichannel image restoration. These tasks all require that the final information is gained from the combination of various data sources or sensors. Recently, image registration has been widely explored in many areas, including remote sensing [1], image super-resolution [2], industrial monitoring [3–5] and medical image registration [6]. However, seldom work has been focused on the registration problem of aerial images. Aerial images are the photographs of the ground or crafts taken from an elevated position. Due to the bad condition when taking pictures, the aerial images all face various problems. The first one is that the resolution of ariel images might not meet requirements because the distances between the camera on aircraft or the spacecrafts and the objects are too long. Thus, further processing such as image super-resolution should be applied on these ariel images. Another one is ariel images that are always affected by clouds and various lights. In order to avoid these effects and get more information from ariel images, various imaging sensors such as infrared camera are used in image acquisition. Thus multi-sensor image fusion is another important task in the further processing of aerial images. All these tasks all require image registration as an important part in the process.

During the last decades, a lot of approaches have been developed to address the problem. Due to the diversity of images to be registered, we can hardly design a universal method which is applicable to all registration tasks. However, the general image registration methods consist of the four steps, including feature detection, feature matching, transform model estimation and image resampling and transformation. In the feature detection step, the salient and distinctive objects, such as closed-boundary regions, edges, contours, line intersections, corners, are detected. Then, in the second step, the algorithms establish the correspondence between the features detected in the sensed image and those detected in the reference image. After that, the types as well as parameters of the mapping functions are estimated. The mapping functions are used to align the sensed image with the reference image. Furthermore, the sensed image is transformed by means of the mapping functions.

* Corresponding author.
*E-mail address:* hamid.r.karimi@uia.no (H.R. Karimi).

http://dx.doi.org/10.1016/j.neucom.2014.04.012

Feature detectors can be traced back to the Moravecs corner detector [7], which looks for the local maximum of minimum intensity changes by shifting a binary rectangle window over an image. However, it is worth noting that the response of this detector is anisotropic, noisy, and sensitive to edges. To reduce these shortcomings, the Harris corner detector [8] was developed. However, it fails to deal with scale changes, which always occur in images. Smith and Brady [9] considered that pixels in a relatively small region are uniform in terms of brightness if these pixels belong to the same object. Based on this assumption, the SUSAN is implemented by comparing brightness within a circular mask. Therefore, the construction of detectors that can cope with this scaling problem is important. Lowe [10] pioneered a scale invariant local feature, namely the scale invariant feature transform (SIFT). It consists of a detector and a descriptor. The SIFTs detector finds the local maximums of a series of difference of Gaussian (DoG) images. Harris–Laplace [11] region detector locates potentially relevant points, interest points, with the Harris corner detector and then selects the point with a characteristic scale, which is the extremum of the Laplacian over different scales. To deal with the viewpoint changes, Mikolajczyk and Schmid [12] put forward the Harris (Hessian) affine detector, which incorporates the Harris corner detector (the Hessian point detector), scale selection, and second moment matrix based elliptical shape estimation. Tuytelaars and Van Gool, based on the following two motivations, (i) edges are stable under affine transformations; (ii) edge-based region detection is more effective than corner-based region detection, developed an edge-based region detector [13], which considers both curved and straight edges to construct parallelograms associated with the Harris corner points. They also proposed an intensity-based detector [13], which starts from the local extrema of intensity and constructs ellipse-like regions with a number of rays emitted from these extrema. Both the edge- and intensity-based methods preserve the affine invariance. Matas et al. [14] developed the maximally stable extremal region (MSER) detector, which is similar to the watershed-based image segmentation. Kadir and Brady [15] proposed the salient region detector, which is based on the probability density function of intensity values computed over an elliptical region.

Detection is followed by feature description. To represent points and regions, which are detected by the above methods, a large number of different local descriptors have been developed. The simplest descriptor is a vector of pixel values. The earliest local descriptor could be the local derivatives [16]. Florack et al. [17] incorporated a number of local derivatives and constructed the differential invariants, which are rotational invariant, for local feature representation. Schmid and Mikolajczyk [18] extended local derivatives as the local gray-value invariants for image retrieval. Freeman and Adelson [19] proposed steerable filters, which are linear combinations of a number of basis filters, for orientation and scale selection to handle tasks in image processing and computer vision research. Wavelets, which are effective and efficient for multiresolution analysis, can also represent local features. Past research has shown the effectiveness of the SIFT descriptor, which is a 128 dimensional vector created by first computing the gradient magnitude and orientation in the neighborhood of the keypoint. This feature is invariant to changes in partial illumination, background clutter, occlusion, and transformations in terms of rotation and scaling. Shape context, a robust and simple algorithm to find correspondences between shapes, is a 3D histogram of edge point locations and orientations introduced by Belongie et al. [20]. Based on the phase and amplitude of steerable bandpass filters, Carneiro and Jepson [21] proposed phase-based local features, which improve invariance to illumination changes. Ke and Sukthankar [22] simplified the SIFT descriptor by utilizing principal component analysis (PCA) to normalized gradient patches to achieve fast matching and invariance to image deformations. This method is named as PCA-SIFT. Lazebnik et al. [23] put forward the rotation invariant feature transform (RIFT), which divides each circular normalized patch into concentric rings, each of which is associated with a gradient orientation histogram. A recent study reports the significance of the gradient location and orientation histogram (GLOH), proposed by Mikolajczyk and Schmid [18], which is an extension of the SIFT descriptor. Similar to the PCA-SIFT, GLOH also applies PCA to reduce the dimension of the descriptor. Preliminary experiments have demonstrated the effectiveness of these descriptors.

Feature matching is an important step to measure the similarity or the dissimilarity between two images, which are represented by two sets of local features, where a similarity metric is constructed based on the correspondences of the local features. In most applications, the following three matching methods are applied: (i) threshold-based matching, (ii) nearest neighbor matching, and (iii) nearest neighbor distance ratio matching. Threshold-based matching finds all possible candidate points in other image for each point in the reference image, in case that the distance between the descriptors of the candidate point and the reference point is below a specified threshold. Nearest neighbor matching algorithms find the point with the closest descriptor to a reference point. Nearest neighbor distance ratio matching utilizes the ratio between the distance of the nearest and the second-nearest neighbors for a reference point. Using which form of matching method depends on a specific application. If a simple and fast strategy is required, the threshold-based matching is often the best choice; if an accurate and effective algorithm is a prerequisition, the nearest neighbor distance ratio matching has distinct advantages.

The iterative closest point (ICP) algorithm is one of most common approaches to feature-based image registration and shape matching problem because of its simplicity and performance. Nonetheless, it has its own limitations. The non-differentiable cost function associated with ICP introduces the local convergence problem which requires sufficient overlap between the data sets and a close initialization. Also, a naive implementation of ICP is known to be prone to outliers which prompted several more robust variations [24,25]. Another elegant method is the partial Hausdorff distance registration [26] which incorporates an underlying robust mechanism similar to the least median of squares technique in robust regression. However, its dependence on a single critical point makes it sensitive to noise and the max of the min approach in the definition is not suitable for performing numerical optimization. Another interesting class involves methods that align two point sets without establishing the explicit point correspondence, and thus achieve more robustness to the missing correspondences and outliers. The idea is to model each of the two point sets by a kernel density function and then quantify the (dis)similarity between them using an information theoretic measure. This (dis)similarity is optimized over a space of coordinate transformations yielding the desired transformation. For instance, Tsin and Kanade propose a kernel correlation based point set registration approach where the cost function is proportional to the correlation of two kernel density estimates. Chui and Rangarajan [27] show that the mixture model can be used to develop a general point matching framework, which includes a unified probabilistic treatment of noise, outliers and the regularization parameters. They use deterministic annealing to simplify the mixture model and enhance the robustness by directly controlling the fuzziness of the correspondence. However they choose one sparsely distributed point-set as the template density modeled by a Gaussian mixture and treat another relatively dense point-set as sample data. Instead of the asymmetric point matching case, Jian [28] treats the problem using mixtures in a symmetrical manner. In this way, the two point-sets, model and scene, are