International Journal of Image and Data Fusion
Vol. 1, No. 2, June 2010, 137 158

Taylor & Francis
Taylor & Francis Group

A critical review of image registration methods

Zhen Xiong and Yun Zhang*

Department of Geodesy and Geomatics Engineering, University of New Brunswick, 15 Dineen Drive, PO Box 4400, Fredericton, NB E3B 5A3, Canada

(*Received 2 September 2009; final version received 10 November 2009*)

Image registration is the process of precisely overlaying two (or more) images of the same area through geometrically aligning common features (or control points) identified in the images. It mainly consists of four steps: feature detection, feature matching, transformation function estimation and image resampling. Image registration is usually applied in photogrammetry, remote sensing, computer vision, pattern recognition and medical image registration. This article presents a review of image registration techniques. We emphasise on feature point detection and matching. The goal of this article is to provide the readers an overview of such techniques, a perspective on the technical advances and a reference to relevant research.

Keywords: image registration; feature detection; feature matching; transformation function; image resampling

1. Introduction

Image registration is the process of precisely overlaying two (or more) images of the same area through geometrically aligning common features (or control points) identified in the images (Habib and Ai-Ruzouq 2005, Xiong and Zhang 2009a). Image registration can be more generalised as a mapping between two images both spatially and with respect to intensity (Brown 1992). The images can be taken at different times, from different viewpoints or by different sensors. Therefore, image registration techniques normally can be grouped into four categories: multi-modal registration, template registration, multi-viewpoints registration and multi-temporal registration (Brown 1992, Zitova and Flusser 2003).

The registered images can be used for different purposes, such as (1) integrating or fusing information taken from different sensors, (2) finding changes in the images taken at different times or under different conditions, (3) inferring three-dimensional (3-D) information from images in which either the camera or the objects in the scene have moved and (4) for model-based object recognition (Brown 1992).

Normally, image registration consists of four steps: (1) feature detection and extraction, (2) feature matching, (3) transformation function fitting and (4) image transformation and image resampling (Zitova and Flusser 2003, Xiong and Zhang 2009a).

*Corresponding author. Email: yunzhang@unb.ca

Up to date, in the feature detection, feature extraction and feature matching, we still face huge technical problems. On the other hand, compared with steps 1 and 2, steps 3 and 4 are much easier. So the feature detection, feature extraction and feature matching are hot research topics in the communities of photogrammetry, remote sensing, computer vision, pattern recognition and image processing. Therefore, this article reviews the techniques which are applied in the above four steps, with emphasis on the techniques of feature point extraction and matching.

2. Feature detection

For image registration, sufficient number of control points (common features) is required in order to estimate an optimal geometric transformation between two images. The control points can be selected manually or extracted automatically. They are, normally, any of the following features (Xiong and Zhang 2009a):

- line intersections;
- points of locally maximum curvature (such as building corners);
- gravity centres of closed boundary regions (such as centres of building roofs or traffic islands) and
- centres of windows having locally maximum variance.

Besides point features, linear features and areal features can also be used for image registration, especially for multi-modal image registration (e.g. registration of optical images and laser scanner images) (Brown 1992, Zitova and Flusser 2003, Habib and Ai-Ruzouq 2005).

2.1 *Point feature*

Point features can be extracted in the space domain and frequency domain of an image. A wide variety of point feature detectors in the space domain exist in the literature. They can be categorised into three classes: intensity based, parametric model based and contour based methods (Schmid *et al*. 2000).

2.1.1 *Intensity-based methods*

Feature points can be extracted by using the first or second derivatives of the intensity surface. The derivatives are used for feature detection by Moravec (1977), Beaudet (1978), Dreschler and Nagel (1982), Heitger *et al*. (1992) and Reisfeld *et al*. (1995). Sun and Kweon (1997) developed an algorithm crosses as oriented pair (COP), which uses two oriented cross-operators. Compared with other conventional corner detectors, COP is a very fast, accurate and robust corner detector (based on univalue segment assimilating nucleus, USAN, and gradient). Colour information can make a significant contribution to feature detection. For multi-spectral images, Nicu *et al*. (2006) developed a colour-based corner detection algorithm to detect the most distinctive features.

Many algorithms use the auto-correlation function for feature detections (Förstner and Gülch 1987, Harris and Stephens 1988, Förstner 1994). A matrix related to the auto-correlation function is widely used in feature detection. This matrix averages derivatives of the signal in a window around a point (x, y):

$$\begin{bmatrix} \sum_{(x_k,y_k)\in W} (I_x(x_k, y_k))^2 & \sum_{(x_k,y_k)\in W} I_x(x_k, y_k) I_y(x_k, y_k) \\ \sum_{(x_k,y_k)\in W} I_x(x_k, y_k) I_y(x_k, y_k) & \sum_{(x_k,y_k)\in W} (I_y(x_k, y_k))^2 \end{bmatrix} \tag{1}$$

where $I(x, y)$ is the image function and (x_k, y_k) are the points in the window around (x, y).

Harris detector is a typical representative of algorithms of using the above matrix for feature detection. If a grey scale, 2-dimensional (2-D) image I is used; taking an image patch over the area (u, v) and shifting it by (x, y), the sum of squared differences (SSD) between these two patches, S is given by

$$S = \sum_u \sum_v (I(u, v) - I(u - x, v - y))^2 \tag{2}$$

The Harris matrix (denotes A) is found by taking the second derivative (the Hessian) of S around $(x, y) = (0, 0)$. A is given by

$$A = \begin{bmatrix} \langle I_x^2 \rangle & \langle I_x I_y \rangle \\ \langle I_x I_y \rangle & \langle I_y^2 \rangle \end{bmatrix} \tag{3}$$

where the angle brackets denote averaging (summation over (u, v)), and the typical notation for partial derivatives is used. If a circular window (or circularly weighted window, such as a Gaussian) is used, then the response will be isotropic (Harris and Stephens 1988).

The strength of the corner is determined by 'how much' the second derivative is. This is done by considering the eigenvalues (λ_1 and λ_2) of A. Based on the magnitudes of the eigenvalues, the following inferences can be made (Harris and Stephens 1988):

(1) If $\lambda_1 = 0$ and $\lambda_2 = 0$, then there are no features of interest at this pixel (x, y).
(2) If $\lambda_1 = 0$ and λ_2 is some large positive values, then an edge is found.
(3) If both λ_1 and λ_2 are large, distinct positive values, then a corner is found.

Harris and Stephens (1988) note that exact computation of the eigenvalues is computationally expensive (since it requires a square root) and instead suggest the following function M_c, where κ is a tunable parameter which determines how 'edge-phobic' the algorithm is.

$$M_c = \lambda_1 \lambda_2 - \kappa(\lambda_1 + \lambda_2)^2 \tag{4}$$

$$M_c = \det(A) - \kappa \operatorname{trace}^2(A) \tag{5}$$

Therefore, the algorithm does not actually have to compute the eigenvalue decomposition of the matrix A and instead it is sufficient to evaluate the determinant to find corners, or rather interest points in general. The value of κ has to be determined empirically, and in the literature values in the range 0.04–0.06 have been reported as feasible. If $M_c > 0$, it is a corner, otherwise, it is not a corner (Harris and Stephens 1988).

Lowe (2004) developed a scale invariant feature transform (SIFT) method for extracting distinctive invariant features from images that can be used to perform reliable

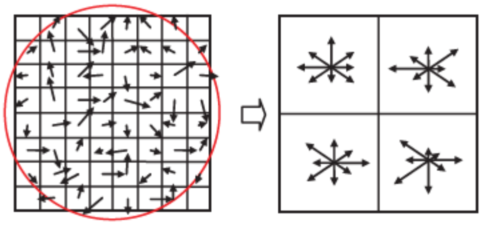

Figure 1. SIFT keypoint descriptor, a 2×2 descriptor array (right) computed from an 8×8 set of samples (left) (Lowe 2004). Notes: The gradient magnitudes and orientations at individual image sample points (pixels), indicated by the overlaid circle, are weighted by a Gaussian window. Four orientation histograms (right) with the length of each arrow corresponding to the sum of the gradient magnitudes near that direction within the region which corresponds to a 4×4 sub regions (sample points) (left) are shown in the right diagram.

matching between different views of an object or scene. SIFT is one of the best algorithms for the extraction of the feature points. The features are highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from many images. Following are the major steps of computation used to generate the image features (Figure 1; Lowe 2004):

(1) Scale-space extreme detection: the first stage of computation searches over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian (DOG) function to identify potential interest points that are invariant to scale and orientation.
(2) Keypoint localisation: at each candidate location, a detailed model is fit to determine the location and scale. Keypoints are selected based on the measures of their stability. the keypoint, with low contrast (<0.03) or the ratio between the largest magnitude eigenvalue and smallest one is very large, e.g. 10, will be eliminated.
(3) Orientation assignment: one or more orientations are assigned to each keypoint location based on the local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale and location for each feature, thereby providing invariance to these transformations.
(4) Keypoint descriptor: the local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination (Figure 1).

The SIFT keypoints are particularly useful due to their distinctiveness, which enables the correct match for a keypoint to be selected from a large database of other keypoints. This distinctiveness is achieved by assembling a high-dimensional vector representing the image gradients within a local region of the image. The keypoints have been shown to be invariant to image rotation and scale, and robust across a substantial range of affine distortion, addition of noise and change in illumination (Lowe 2004).

The features described by SIFT descriptor use only a monochrome intensity image; therefore, further distinctiveness could be derived from including illumination-invariant colour descriptors (Funt and Finlayson 1995, Brown and Lowe 2002, Lowe 2004).

Local texture measures appear to play an important role in human vision and could be incorporated into feature descriptors in a more general form than the single spatial frequency used by the SIFT.

The SIFT operator has another two drawbacks in the case of stereo matching in photogrammetry. First of all, the DOG detects mainly blob-like interest points, while the significant points, such as the corners of buildings and saddle points near the edges of roads, could not be successfully extracted, and this disadvantage is critical to the 3-D reconstruction. Second, the interest points DOG detected may be not dense enough to fulfil the generation of digital surface model through image matching and the exterior orientation (Zhu *et al*. 2007).

Mikolajczyk and Schmid (2004) developed a scale and affine invariant interest point detector. This scale and affine invariant detector is based on the following results:

(1) Interest points extracted with the Harris detector can be adapted to affine transformations and give repeatable results (geometrically stable).
(2) The characteristic scale of a local structure is indicated by a local extreme over the scale of normalised derivatives (the Laplacian).
(3) The affine shape of a point neighbourhood is estimated based on the second moment matrix.

This detector first computes a multi-scale representation for the Harris interest point detector, then selects the local maximal points over scales. The scale invariant detector is extended to affine invariant by estimating the affine shape of a point neighbourhood. An iterative algorithm modifies location, scale and neighbourhood of each point and converges to affine invariant points. This method can deal with significant affine transformations including large-scale changes. The characteristic scale and the affine shape of neighbourhood determine an affine invariant region for each point (Mikolajczyk and Schmid 2004).

The scale invariant detector can deal with larger scale changes better than the affine invariant detector, but it fails for images with large affine transformations. The affine invariant points provide reliable matching even for the images with significant perspective deformations. However, the stability and convergence of affine regions is the subject of further investigation as well as their robustness to occlusions (Mikolajczyk and Schmid 2004).

2.1.2 *Parametric model-based methods*

Parametric model-based methods can extract features with high accuracy. Rohr (1992) developed a template parametric model, where the parameters of the model are adjusted by a least squares method. For the 'L' corners, the parameters include the angle of L-corner, the angle between the symmetry axis of the L-corner and the x-axis, the grey values, the position of the point and the amount of blur (Schmid *et al*. 2000). Deriche and Blaszka (1993) improved Rohr's method by substituting an exponential for the Gaussian smoothing function. Baker *et al*. (1998) and Parida *et al*. (1998) also developed parametric methods for feature extraction (Schmid *et al*. 2000).

2.1.3 *Contour-based methods*

Up to date, many contour-based algorithms have been developed (Freeman and Davis 1977, Beus and Tiu 1987, Liu and Srinath 1990). A number of frequently cited approaches

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase® Smarter legal research.