INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

Image Registration Techniques: A Comprehensive Survey

Paresh M. Patel
P.G. Student, Structural Engineering Department
Sardar Vallabhbhai Patel Institute of Technology, Vasad, Gujarat, India
Vishal M. Shah
Assistant Professor, Sardar Vallabhbhai Patel Institute of Technology, Vasad, Gujarat, India

Abstract:
This paper aims to present a overview of recent as well as classic image registration methods. Image registration is the process of overlaying images (two or more) of the same scene taken at different times, from different viewpoints, and/or by different sensors. The registration geometrically aligns two images (the reference and sensed images). The reviewed approaches are classified according to their nature (area-based and feature-based) and according to four basic steps of image registration procedure: feature detection, feature matching, mapping function design, and image transformation and resampling. Registration algorithms compute transformations to set correspondence between the two images the purpose of this paper is to provide a comprehensive review of the existing literature available on Image registration methods. Main contributions, advantages, and drawbacks of the methods are mentioned in the paper. We believe that it will be a useful document for researchers longing to implement alternative Image registration methods for specific applications.

Key words: *Image registration; Feature detection; Feature matching; Mapping function; Resampling*

1. Introduction

Image registration is the process of overlaying two or more images of the same scene taken at different times,from different viewpoints, and/or by different sensors. It geometrically aligns two images—the reference and sensed images. The present differences between images are introduced due to different imaging conditions. Image registration is a crucial step in all image analysis tasks in which the final information is gained from the combination of various data sources like in image fusion, change detection, and multichannel image restoration. Typically, registration is required in remote sensing (multispectral classification, environmental monitoring, change detection, image mosaicing, weather forecasting, creating super-resolution images, integrating information into geographic information systems (GIS)), in medicine (combining computer tomography (CT) and NMR data to obtain more complete information about the patient, monitoring tumor growth, treatment verification, comparison of the patient's data with anatomical atlases), in cartography (map updating), and in computer vision (target localization, automatic quality control), to name a few. During the last decades, image acquisition devices have undergone rapid development and growing amount and diversity of obtained images invoked the research on automatic image registration. A comprehensive survey of image registration methods was published in 1992 by Brown [26]. The intention of our article is to cover relevant approaches introduced later and in this way map the current development of registration techniques. According to the database of the Institute of Scientific Information (ISI), in the last 10 years more than 1000 papers were published on the topic of image registration. Methods published before 1992 that became classic or introduced key ideas, which are still in use, are included as well to retain the continuity and to give complete view of image registration research.We do not contemplate to go into details of particular algorithms or describe results of comparative experiments, rather we want to summarize main approaches and point out interesting parts of the registration methods. In Section 2 various aspects and problems of image registration will be discussed. Both area-based and featurebased approaches to feature selection are described in Section 3. Section 4 reviews the existing algorithms for feature matching. Methods for mapping function design are given in Section 5. Finally, Section 6 surveys main techniques for image transformation and resampling.

2. Image Registration Methodology

Due to the large variety of images that can be registered and the various degradation types, it is impossible to outline an universal method applicable to every registration task, because every case should take into account not only the geometric distortion types but also the peculiar radiometric deformations and noise corruption of the images. Nevertheless, we can outline the main steps needed to register a set of generic digital images:

- **Preprocessing**
  (Smoothing, deblurring, edge detection, segmentation, etc. . .)
- **Feature detection**
  (Extracting points, lines, regions, templates, etc. . .)
- **Feature matching**
  (Find the correct feature pairings avoiding outliers)
- **Transformation estimation**
  (Restore the original image deformation)
- **Resampling**
  (Use the transformation to warp sensed image to the reference)

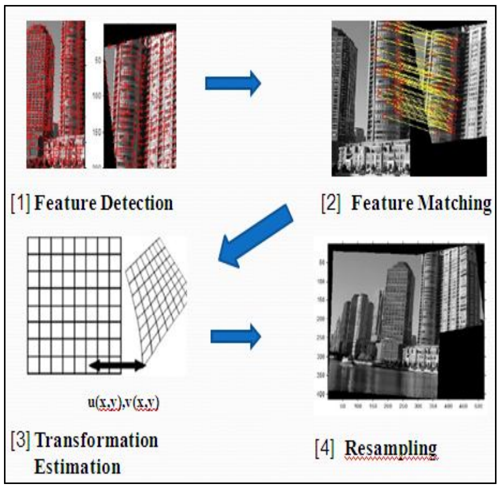

Figure 1: The four main steps needed to register two images:
(1) Feature detection (corners were used in this example)
(2) Feature matching using invariant descriptors (a correspondence is represented by the same pair number),
(3) Transformation estimation, (4) Image Resampling. [1]

The implementation of each registration step has its typical problems and we have to decide what kind of features is appropriate for the given task. The features should be distinctive objects which are possibly uniformly spread over the images and easily detectable (sometimes a physical interpretation of the features is demanded). The detected feature sets in the reference and sensed images must have enough common elements, even when images do not exactly cover the same scene or when there are object occlusions or other unexpected changes. The detection methods should have good localization accuracy and should not be sensitive to the assumed image degradation. In an ideal case, the algorithm should be able to detect the same features in all “projections" of the scene regardless of the particular image deformation. Problems caused by an incorrect feature detection or by image degradation can arise in the feature matching step. Physically corresponding features can be dissimilar due to the different imaging conditions and/or sensor spectral sensitivity. The choice of the feature description has to take into account these considerations to be invariant to the assumed degradation and discriminable enough to be able to distinguish among different features. The matching algorithm in the invariant space should be sufficiently robust as not to be influenced by slight unexpected feature variations and single features without corresponding counterparts in the other image should not affect its performance. The type of mapping functions should be chosen accordingly to the a priori known information about the acquisition process and expected image degradations. If no a priori information is available, the model should be flexible and general enough to handle all possible degradations which might appear. The accuracy of the feature detection method, the reliability of feature correspondence estimation, and the acceptable approximation error need to be considered too. Moreover, the decision about which differences between images have to be removed by registration has to be evaluated. It is also desirable not to remove the differences we are searching for if the aim is a global scene change detection; this issue is very important and extremely difficult at the same time. Finally, the choice of the appropriate type of resampling technique depends on the trade-off between the demanded accuracy of the interpolation and the computational complexity. The nearest-neighbour or bilinear interpolation are sufficient in most cases, but some applications require more precise methods which we will investigate later in this survey. Skipping the preprocessing step which typically involves some basic image processing techniques which will not be covered in this report, the following sections will explain the subsequent registration steps.

3. Feature Detection
Choosing points of interest within the image pixel space is not a easy task, because we need to easily identify significant and distinct high-level features such as objects (buildings, people, animals, etc.), lines (coastlines, roads, rivers, etc.) or points (corners, line intersections, high-curvature curve points, etc.). The more invariant these feature are, the more robust and accurate the matching procedure will be, and the number of features should be sufficiently high to increase the discrimination and consequently avoid pairing ambiguities. These kind of features should not be identified directly on image intensity or color values because illumination or sensor noise can vary between sensed and reference images.

3.1. Region Features
These features are often projections of general high-contrast closed-boundary regions of appropriate size represented by their center of gravity [2] and their recognition is invariant with respect to rotation, scaling, skewing, random noise and gray level variation. Regions are usually identified by segmentation procedure, and Goshtasby et al. [3] proposed a new iterative algorithm which is carried out together with registration to fine tune segmentation parameters and obtain subpixel accuracy. Many researchers worked on the identification of water reservoirs, lakes, buildings, forests, urban areas, object shadows, etc. . . Recent research was spent on designing robust features particularly invariant to scale changes: Alhichri [4] proposed the idea of “virtual circles" as features, i.e. local maxima in the distance transform of the edge map [5]; the circles parameters (center and radius) are grouped into collinear sets to which a line is fitted to determine the direction and the Hausdorff fraction is also used as a similarity measure to determine the optimal transformation. This algorithm can find large translation, rotation and scale differences and has linear complexity in terms of the number of virtual circles.

3.2. Line Features
These kinds of features are mainly used in medical and satellite image registration, because they are well suited to identify object contours like anatomical structures or geological elements. Standard edge detection filters like Canny, Harris or Laplacian are often used for these purposes and other applications include general line segments, object contours, coastal lines, roads, etc. . . Region and point features are generally more robust than lines and less diffused, however a survey on existing edge detection methods together with performance evaluation can be found in [6].

3.3. Point Features
Features based on point localization are the most commonly used for image registration procedures as they can provide a highly parametric description of the correspondence based only on point coordinates. Known applications include identification of line intersections, road crossings, centroids of water regions, oil and gas pads, high variance points, local curvature discontinuities, curve inflection points, most distinctive points of similarity measure, etc. . . Most algorithms used for point detection rely on the idea of “corner", a salient point well perceived by human observers and invariant to geometrical transformations, often a line intersection, a centroid of closed-boundary regions or local modulus maxima of the wavelet transform. Methods based on first derivative analysis, such the Harris detector are more robust than the ones using the second derivative or gaussian curvature, because the letters are very sensitive to noise. As today, the best corner detectors are Forstner [7] and SUSAN filters [8], but some recent efforts were also spent on designing a new parametric corner detector which does not employ any derivatives and is able to handle blurred and noisy data [9]. An exhaustive survey on corner detection methods updated to 2001 can be found on [10], but recently researchers have discovered that embedding a “descriptor" into point features greatly enhances the following matching procedure suppressing outliers and incorrect pairings. An extensive survey on local descriptor performance can be found in [11].

3.4. Summary
Feature-based methods are recommended if the images contain enough distinctive and easily detectable objects. (Applications: remote sensing and computer vision). On the other hand area-based methods are used in medical images (low details images). Recently, registration methods using simultaneously both area-based and feature-based approaches have started to appear [12].

4. Feature Matching
The detected features in the reference and sensed images can be matched by means of the image intensity values in their close neighborhoods, the feature spatial distribution, or the feature symbolic description. Some methods, while looking for the feature correspondence, simultaneously estimate the parameters of mapping functions and thus merge the second and third registration steps. In the following sections, the two major categories (area-based and feature-based methods, respectively), are retained and further classified into subcategories according to the basic ideas of the matching methods.

4.1. Area-Based Methods
Sometimes called correlation-like or template matching, these methods somewhat merge the feature detection step with the matching part. These methods deal with the images without attempting to detect salient points, but consider areas of the image as features and windows of predefined size are often used for the correspondence estimation during the second registration step.
The limitations of the area-based methods originate from their basic idea: The rectangular window, which is most often used, suits the registration of images which locally differ only by a translation. If images are deformed by more complex transformations, this type of

the window is not able to cover the same parts of the scene in the reference and sensed images (the rectangle can be transformed to some other shape). Several authors proposed to use circular shape of the window for mutually rotated images. However, the comparability of such simple-shaped windows is violated too if more complicated geometric deformations (similarity, perspective transforms, etc.) are present between images. Another disadvantage of the area-based methods refers to the “remarkableness" of the window content. There is high probability that a window containing a smooth area without any prominent details will be matched incorrectly with other smooth areas in the reference image due to its non-saliency. The features for registration should be preferably detected in distinctive parts of the image. Classical area-based methods like cross-correlation (CC) exploit for matching directly image intensities, without any structural analysis. Consequently, they are sensitive to the intensity changes, introduced for instance by noise, varying illumination, and/or by using different sensor types.

4.1.1. Correlation-Like Methods

The most commonly used formula the normalized CC and its variants is the following one:

$$\gamma(u,v)=\frac{\sum_{x,y}\left(f(x,y)-\bar{f}_{u,v}\right)\cdot\left(t(x-u,y-v)-\bar{t}\right)}{\sqrt{\sum_{x,y}\left(f(x,y)-\bar{f}_{u,v}\right)^{2}\cdot\sum_{x,y}\left(t(x-u,y-v)-\bar{t}\right)^{2}}}$$

Which represents the correlation coefficient γ between every pixel (u;,v) of the source image and the template $\bar{t}$ is the mean of the template and $\bar{f}_{u,v}$ is the mean of f(x, y) in the region “under" the template. This measure of similarity is computed for window pairs from the sensed and reference images and its maximum is searched. The window pairs for which the maximum is achieved are set as the corresponding ones. If subpixel accuracy is required, interpolation of the CC measure values needs to be used. Although the CC-based registration can exactly align mutually translated images only, it can also be successfully applied when slight rotation and scaling are present. Generalized versions of CC for geometrically more deformed images can also be used: they compute the CC for each assumed geometric transformation of the sensed image window and are able to handle even more complicated geometric deformations than the translation usually similarity transforms. The computational load however, grows very fast with the increase of the transformation complexity. In case the images/objects to be registered are partially occluded the extended CC method based on increment sign correlation can be applied [8]. Similar to the CC methods is the Sequential Similarity Detection Algorithm (SSDA): it uses the sequential search approach and a computationally simpler distance measure than the CC. It accumulates the sum of absolute differences of the image intensity values (matrix l1norm) and applies the threshold criterion (if the accumulated sum exceeds the given threshold, the candidate pair of windows from the reference and sensed images is rejected and the next pair is tested). The method is likely to be less accurate than the CC but it is faster. Recently great interest in the area of multimodal registration has been paid to the correlation ratio based methods. As opposed to classical CC, this similarity measure can handle intensity differences between images due to the usage of different sensors multimodal images. It supposes that intensity dependence can be represented by some function. In case of noisy images with certain characteristic (fixed-pattern noise), projection-based registration working with accumulated image rows and columns outperforms classical CC.

- **Drawbacks**: Flatness of the similarity measure maxima (due to the self-similarity of the images) and high computational complexity. Despite the limitations mentioned above, correlation-like registration methods are still often in use, particularly thanks to their easy hardware implementation, which makes them useful for real-time applications.

4.1.2. Fourier Methods

If an acceleration of the computational speed is needed or if the images were acquired under varying conditions or they are corrupted by frequency-dependent noise, then Fourier methods are preferred against the correlation-like methods. They exploit the Fourier representation of the images in the frequency domain. The phase correlation method is based on the Fourier Shift Theorem and was originally proposed for the registration of translated images. It computes the cross-power spectrum of the sensed and reference images and looks for the location of the peak in its inverse:

$$\frac{\mathcal{F}(f)\,\mathcal{F}(g)^{*}}{|\mathcal{F}(f)\,\mathcal{F}(g)^{*}|}=e^{2\pi i(u\cdot x_0+v\cdot y_0)}$$

The method shows strong robustness against the correlated and frequency dependent noise and non-uniform, time varying illumination disturbances; the computational time savings are more significant if the images to be registered are large. If a change of image scale is present too, the images can be registered using the combination of polar-log mapping of the spectral magnitude (which corresponds to the Fourier {Mellin transform) and the phase correlation or cepstrum filter. Known applications include remote sensing (SPOT images) and medical imaging (MR images). Some authors propose to compute the correlation in frequency domain to handle multimodal images when applied to the edge representations instead of the original gray level images [13]. The authors proposed to compute the correlation in frequency domain. This method can handle multimodal images when applied to the edge representations instead of the original gray level images. Extension of phase correlation to sub-pixel registration by means of the analytic expression of phase correlation on down sampled images was introduced by Foroosh et al. [14].

4.1.3. Mutual Information Methods

The mutual information (MI) methods are the last group of area-based methods to be reviewed here. They have appeared recently and represent the leading technique in multimodal registration. Registration of multimodal images is a dificult task, but often necessary to solve, especially in medical imaging. The comparison of anatomical and functional images of the patient body can lead to a diagnosis, which would be impossible to gain otherwise. Remote sensing often makes also use of more sensor types. Originating from information theory, MI is a measure of statistical dependency between two data sets and is particularly suited to register images acquired with different modalities. MI between two random variables X and Y is given by

$$\boxed{MI(X,Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)}$$

Where,

$$\boxed{H(X) = -E_X\left(\log\left(P(X)\right)\right)}$$

Represents the entropy of a random variable and P(X) is the probability distribution of X. The method is based on the maximization of MI (Figure 3). Often a speed up of the registration is implemented, exploiting the coarse-to-fine resolution strategy (the pyramidal approach) and one of the first researchers proposing this technique were Viola and Wells in 1997 [15]. The authors described the application of MI for the registration of magnetic resonance images as well as for the 3D object model matching to the real scene. MI was maximized using the gradient descent optimization method. The most remarkable results were obtained by Thevenaz and Unser [16], who tried to combine various approaches, solving individual steps of MI registration. They employed the Parzen window for joint probability computation and the Jeeves method or the Marquardt {Levenberg method to maximize the MI; they also used Spline pyramids interpolation to both increase accuracy and reduce complexity. For more in-depth details about these techniques, please refer to above mentioned paper. Another approach involves the use of hierarchical search strategy together with simulated annealing to find the maximum of the MI [17] and, according to the literature, other alternative maximization methods for the MI are multi-resolution hill climbing algorithm, the Brent's method and the Powell's multidimensional direction set.

4.2. Feature-Based Methods

We now proceed with the discussion about registration methods based on more precise features, often points or small blocks of reference and target image. We assume that two sets of features in the reference and sensed images represented by the CPs (points themselves, end points or centers of line features, centers of gravity of regions, etc...) have been detected. The aim is to find the pairwise correspondence between them using their spatial relations or various descriptors of features.

4.2.1. Methods Using Spatial Relations

Methods based primarily on the spatial relations among the features are usually applied if detected features are ambiguous or if their neighborhoods are locally distorted. The information about the distance between the CPs and about their spatial distribution is exploited. The registration is often based on the graph

Matching algorithm, which evaluates the number of features in the sensed image that fall within a given range next to the features in the reference image after a particular transformation. The transformation parameters with the highest score are then set as a valid estimate. Another approach is the clustering technique, which tries to match points connected by edges or line segments (the assumed geometrical model is the similarity transform). For every pair of CPs from both the reference and sensed images, the parameters of the transformation which maps the points on each other are computed and represented as a point in the space of transform parameters. The parameters of transformations that closely map the highest number of features tend to form a cluster, while mismatches fill the parameter space randomly. The cluster is detected and its centroid is assumed to represent the most probable vector of matching parameters. Mapping function parameters are thus found simultaneously with the feature correspondence. Local errors do not influence globally the registration process. Barrow [18] introduced the chamfer matching for image registration: line features detected in images are matched by means of the minimization of the generalized distance between them and Borgefors [19] proposed an improved version, where a better measure of correspondence (the sequential distance transforms together with the root mean square average) was applied. The algorithm also employs a multi-resolution pyramidal decomposition for better efficiency.

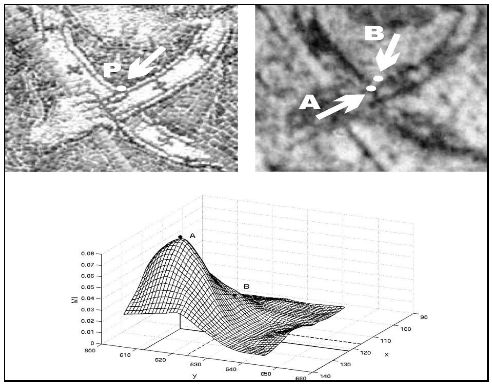

Figure 2: Mutual Information: MI criterion (bottom row) computed in the neighborhood of point P between new and old photographs of the mosaic (top row). Maximum of MI shows the correct matching position (point A). Point B indicates the false matching position selected previously by the human operator. The mistake was caused by poor image quality and image degradations

4.2.2. Relaxation Methods

A large group of registration methods is based on the relaxation approach, as one of the solutions to the consistent labeling problem (CLP): to label each feature from the sensed image with the label of a feature from the reference image, so it is consistent with the labeling given to the other feature pairs. The process of recalculating the pair figures of merit, considering the match quality of the feature pairs and of matching their neighbors, is iteratively repeated until a stable situation is reached. The reference work was done by Ranade and Rosenfeld [20]. Here, the displacement of the feature sets transformed by a certain geometric transformation defines the figures of merit of the feature pairs. This method can handle shifted images and it tolerates local image distortions. Other authors extended the classical relaxation by including the description of the corner features. They used corner sharpness, contrast, and slope, allowing handling translation and rotation distortions in the images, but it is computationally demanding. Other features used for relaxation include line features and their descriptors (coordinates, orientation, and average contrast). Another solution to the CLP problem and consequently to the image registration is backtracking, where consistent labeling is generated in a recursive manner.

4.2.3. Pyramids and Wavelets

We conclude the discussion about feature matching by mentioning some works that try to reduce the computational cost due to the large image size by means of pyramidal approach. First attempts were done back in 1977 when a sub-window was used to find probable candidates of the corresponding window in the reference image and then the full-size window was applied. The appropriate choice of the bind size was crucial to minimize the expected computational cost. Further works proposed to decrease the necessary computational load by taking just a sparse regular grid of windows for which the cross correlation matching is performed. These techniques are simple examples of early pyramidal methods. In general, this coarse-to-fine hierarchical strategy applies the usual registration methods, but it starts with the reference and sensed images on a coarse resolution (generated using Gaussian pyramids, simple averaging or wavelet transform coefficients, among others). Then they gradually improve the estimates of the correspondence of the mapping function parameters while going up to the finer resolutions.

At every level, they considerably decrease the search space and thus save the necessary computational time. Another important advantage resides in the fact that the registration with respect to large-scale features is achieved first and then small corrections are made for finer details. On the other hand, this strategy fails if a false match is identified on a coarser level. To overcome this, a backtracking or consistency check should be incorporated into the algorithms. Combining CC with the pyramidal approach that exploits a summing pyramid (the pixel value at a coarser level corresponds to the summation of the pixel values on the previous level), a median pyramid, and an averaging pyramid was proposed. Some authors combined the SSDA method with pyramidal speed-up extracting features (centroids of closed boundary regions) at every resolution level and founding the parameters of geometric deformation from the histogram of angle differences and line-length ratios. Thevenaz et al. [21] applied a cubic spline based pyramid along with the minimization of the mean square intensity difference between the images and the MI maximization. Recently, wavelet decomposition of the images was recommended for the pyramidal approach due to its inherent multi-resolution character. Methods can differ in the type of the applied wavelet and the set of wavelet coefficients used for finding the correspondence. Most frequently used methods decompose the image recursively into four sets of coefficients (LL, HL, LH, HH) by filtering the image successively with two filters, a low-pass filter L and a high-pass filter H, both working along the image rows and columns. Many comparison tests were carried out to establish which wavelet family could have the best performance and spline biorthogonal and Haar wavelet outperformed others [22]. Other applications include Daubechies wavelets to register Landsat images and AVHRR data [23], while Gabor wavelet transform and Gaussian model were also applied to medical image registration problems [24].

4.2.4. Methods Using Invariant Descriptors

As an alternative to the methods exploiting the spatial relations, the correspondence of features can be estimated using their description, preferably invariant to the expected image deformation (see Fig. 3). The description should fulfill several conditions. The most important ones are invariance (the descriptions of the corresponding features from the reference and sensed image have to be the same), uniqueness (two different features should have different descriptions), stability (the description of a feature which is slightly deformed in an unknown manner should be close to the description of the original feature), and independence (if the feature description is a vector, its elements should be functionally independent). However, usually not all these conditions have to (or can) be satisfied simultaneously and it is necessary to find an appropriate trade-off. Features from the sensed and reference images with the most similar invariant descriptions are paired as the corresponding ones. The choice of the type of the invariant description depends on the feature characteristics and the assumed geometric deformation of the images. While searching for the best matching feature pairs in the space of feature descriptors, the minimum distance rule with thresholding is usually applied. If a more robust algorithm is needed, the matching likelihood coefficients [25], which can better handle questionable situations, can be an appropriate solution. Guest et al. proposed to select features according to the reliability of their possible matches. The simplest feature description is the image intensity function itself, limited to the close neighborhood of the feature. To estimate the feature correspondence, authors computed the CC on these neighborhoods. Other types of similarity measures can be used, too. Zheng and Chellapa make use of the correlation coefficients [26]. They assumed the similarity geometric deformation. In their approach, firstly the rotation between images was compensated by the estimation of the illuminant direction and then the coarse-to-fine correlation based registration was performed. In Ref. [27] the MI was used for the improvement of the feature correspondence. Sester et al. [28] proposed to describe forests, used as the region features, by elongation parameter, compactness, number of holes, and several characteristics of the minimum bounding rectangle. Many authors used closed-boundary regions as the features. In principle, any invariant and discriminative enough shape descriptor can be employed in region matching.

Peli proposed simple and fast description by radial shape vector but the usage of this method is limited to star-shape regions only. A generalized shape description in a form of a binary matrix was proposed in Ref. [30]. In Ref. [31], the shape matrices were used for registration of rotated and scaled satellite images. In Ref. [32] a chain code representation of contours was proposed as the invariant description and a chain code correlation-like measure was used for finding the correspondence. A large group of methods uses moment-based invariants for description of closed-boundary region features.

Invariant combination of the basic geometric properties of features can form geometrically oriented descriptors. Govindu et al. represented the extracted contours from possibly rotated images by the slopes of tangents in the contour points. They did not look for contour correspondence, but only for the distributions of the proposed descriptors. By comparison of the corresponding distributions from the reference and sensed images the mutual image rotation can be estimated. They derived a similar type of descriptors for the affine transform, too. Wang and Chen [33] computed the histogram of line-length ratios and the histogram of angle differences of any two line segments in the reference and sensed images. They assumed the similarity transform. Ventura et al. Described image features by various descriptors (ellipticity, angle, thinness, etc.) and represented relations among them by a multivalue logical tree (MVLT). Then they compared the -

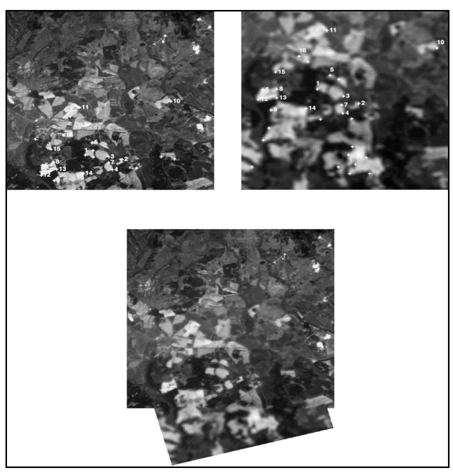

Figure 4: Feature-based method using invariant descriptors: in these two satellite images, control points (corners) were matched using invariants based on complex moments [29]. The numbers identify corresponding CP's. The bottom image shows the registration result

MVLTs of the reference and sensed images to find the feature correspondence.

5. Transform Model Estimation

After the feature correspondence has been established the mapping function is constructed and it should transform the sensed image to overlay it over the reference one. The correspondence of the CPs from the sensed and reference images together with the fact that the corresponding CP pairs should be as close as possible after the sensed image transformation are employed in the mapping function design. The task to be solved consists of choosing the type of the mapping function and its parameter estimation. The type of the mapping function should correspond to the assumed geometric deformation of the sensed image, to the method of image acquisition (e.g. scanner dependent distortions and errors) and to the required accuracy of the registration. In special situations when the geometric deformation is partially known, e.g. when there exists a model for the distortion caused by the acquisition device and/or the scene geometry, the pre-correction based on the inverse of the deformation can be performed. Models of mapping functions can be divided into two broad categories according to the amount of image data they use as their support. Global models use all CPs for estimating one set of the mapping function parameters valid for the entire image (this is also known as rigid deformation). On the other hand, local mapping functions treat the image as a composition of patches and the function parameters depend on the location of their support in the image. It leads to the tessellation of the image, usually a triangulation, and to the definition of mapping parameters for each patch separately (this is also known as elastic deformation). From another point of view, mapping functions can be categorized according to the accuracy of overlaying of the CPs used for computation of the parameters. Interpolating functions map the sensed image CPs on the reference image CPs exactly, whereas approximating functions try to find the best trade-off between the accuracy of the final mapping and other requirements imposed on the character of the mapping function. Since the CP coordinates are usually supposed not to be precise, the approximation model is more common.

5.1. Global Mapping Models

One of the most frequently used global models uses bivariate polynomials of low degrees. Similarity transform is the simplest model—it consists of rotation, translation and scaling only.

$$\boxed{\begin{aligned} u &= s(x\cos(\varphi) - y\sin(\varphi)) + t_x \\ v &= s(x\sin(\varphi) + y\cos(\varphi)) + t_y \end{aligned}}$$

This model is often called ‘shape-preserving mapping’ because it preserves angles and curvatures and is unambiguously determined by two CPs.

Slightly more general but still linear model is an affine transform,

$$\boxed{\begin{aligned} u &= a_0 + a_1x + a_2y \\ v &= b_0 + b_1x + b_2y, \end{aligned}}$$

Which can map a parallelogram onto a square? This model is defined by three non-collinear CPs, preserves straight lines and straight line parallelism. It can be used for multiview registration assuming the distance of the camera to the scene is large in comparison to the size of the scanned area, the camera is perfect (a pin-hole camera), the scene is flat, and the present geometric distortion has no local factors.

If the condition on the distance of the camera from the scene is not satisfied the perspective projection model, Should be used. This model exactly describes a deformation of a flat scene photographed by a pin-hole camera the optical axis of which is not perpendicular to the scene. It can map a general quadrangle onto a square while preserving straight lines and is determined by four independent CPs. Slight violations of these assumptions may lead to the use of the second or the third-order polynomial models. Higher order polynomials usually are not used in practical applications because they may unnecessarily warp the sensed image in areas away from the CPs when aligning with the reference image. In general, the number of CPs is usually higher than the minimum number required for the determination of the mapping function. The parameters of the mapping functions are then computed by means of the least-square fit, so that the polynomials minimize the sum of squared errors at the CPs. Such mapping functions do not map the CPs onto their counterparts exactly. This approach was proved to be very effective and accurate for satellite images, for instance.

$$\boxed{\begin{aligned} u &= \frac{a_0 + a_1x + a_2y}{1 + c_1x + c_2y} \\ v &= \frac{b_0 + b_1x + b_2y}{1 + c_1x + c_2y} \end{aligned}}$$

5.2. Local Mapping Models

However, a global polynomial mapping cannot properly handle images deformed locally. This happens, for instance, in medical imaging and in airborne imaging. The least square technique averages out the local geometric distortion equally over the entire image which is not desirable. Local areas of the image should be registered with the available information about the local geometric distortion in mind. Several authors have shown the superiority of the local or at least locally sensitive registration methods above the global ones in such situations (Goshtasby , Ehlers and Fogel , Wiemker , and Flusser , among others). The weighted least square and weighted mean methods gain the ability to register images locally by introducing the slight variation to the original least square method. The local methods called piecewise linear mapping and piecewise cubic mapping, together with the Akima’s quintic approach, apply the combination of the CP-based image triangulation and of the collection of local mapping functions each valid within one triangle. These approaches belong to the group of the interpolating methods.

5.3. Mapping by Radial Basis Functions

Radial basis functions are representatives of the group of global mapping methods but they are able to handle even locally varying geometric distortions. The resulting mapping function has a form of a linear combination of translated radially symmetric function plus a low-degree polynomial.

$$\boxed{u = a_0 + a_1 x + a_2 y + \sum_{i=1}^{N} c_i g(\mathbf{x}, \mathbf{x}_i)}$$

And similarly for v:

Originally they were developed for the interpolation of irregular surfaces. Their name 'radial' reflects an important property of the function value at each point-it depends just on the distance of the point from the CPs, not on its particular position. Multiquadrics, reciprocal multiquadrics, Gaussians, Wendland's functions, and thin-plate splines are several examples of the radial basis functions used in image registration.

5.4. Elastic Registration

Another approach to the registration of images with considerable complex and/or local distortions is not to use any parametric mapping functions, where the estimation of the geometric deformation is reduced to the search for the 'best' parameters. This idea was introduced by Bajcsy et al. and is often called elastic registration. The images are viewed as pieces of a rubber sheet, on which external forces stretching the image and internal forces defined by stiffness or smoothness constraints are applied to bring them into alignment with the minimal amount of bending and stretching. The feature matching and mapping function design steps of the registration are done simultaneously. This is one of the advantages of elastic methods, because feature descriptors invariant to complicated deformations are not known and the feature correspondence is difficult to establish in the traditional way. The registration is achieved by locating the minimum energy state in an iterative fashion. A pyramidal approach is often applied. The external forces can be derived from the local optimization of the similarity function which is defined by the intensity values or by the correspondence of boundary structures, among others. In Ref. [34], no external forces were used and the prescribed displacements, derived from the correspondence of boundary structures, were incorporated to the elastic image deformation. Disadvantage of elastic registration is in situations when image deformations are much localized. This can be handled by means of fluid registration. Fluid registration methods make use of the viscous fluid model to control the image transformation. The reference image is here modelled as a thick fluid that flows out to match the sensed image under the control of the derivative of a Gaussian sensor model. This approach is mainly used in medical applications .The weakness of this approach is blurring introduced during the registration process. Another example of non-rigid methods is diffusion-based registration, level sets registration, and optical flow based registration.

6. Image Resampling

The mapping functions constructed during the previous step are used to transform the sensed image and thus to register the images. The transformation can be realized in a forward or backward manner. Each pixel from the sensed image can be directly transformed using the estimated mapping functions. This approach, called a forward method, is complicated to implement, as it can produce holes and/or overlaps in the output image (due to the discretization and rounding). Hence, the backward approach is usually chosen. The registered image data from the sensed image are determined using the coordinates of the target pixel (the same coordinate system as of the reference image) and the inverse of the estimated mapping function. The image interpolation takes place in the sensed image on the regular grid. In this way neither holes nor overlaps can occur in the output image because missing values are interpolated from the ones already present (Figure 5).

The interpolation itself is usually realized via convolution of the image with an interpolation kernel. An optimal interpolant like the 2D sinc function is hard to implement in practice because of its infinite extent. Thus, many simpler interpolants of bounded support have been investigated in the literature. In order to reduce the computational cost, preferably separable interpolants have been considered. The separability enables to replace an m×m two-dimensionalconvolution by (m + 1) mono-dimensional convolutions which is much faster. The nearest neighbour function, the bilinear and bicubic functions are the most widely used, but also quadratic splines, cubic B-splines, higher-orderB-splines, Catmull Rom cardinal splines, Gaussians and truncated sinc functions are applied in some cases. Higher order polynomial kernels like quintic or septic showed only marginal improvement in comparison with cubic interpolation at an highly increased computational cost. Several survey papers on resampling techniques have been published in the last years. A detailed investigation and comparison of methods was carried out in. Most recently, Thevenaz et al. Have proposed a different approach to image resampling. Unlike the other methods, their resampling functions do not necessarily interpolate the image gray levels. They rather interpolate values calculated as certain functions of the gray levels. The authors have demonstrated that this approach outperforms traditional interpolation techniques in terms of final image quality. Even though the bilinear interpolation is outperformed by higher-order methods in terms of accuracy and visual appearance of the transformed image, it offers probably the best trade-off between accuracy and computational complexity and thus it is the most commonly used approach.

Cubic interpolation is recommended when the geometric transformation involves a significant enlargement of the sensed image. Nearest neighbor interpolation should be avoided in most cases because of artifacts in the resampled image. The only exclusion is when the image to be transformed contains low number of intensities and we do not want to introduce \synthetic" gray levels” colors by higher order interpolation.

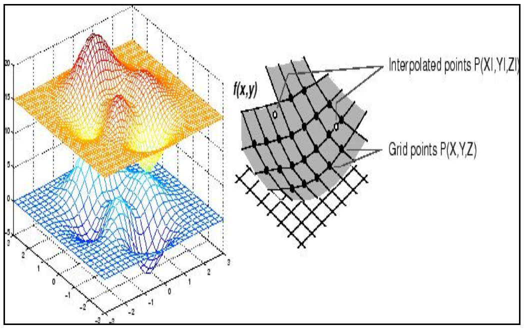

Figure 5: Generic data interpolation: this _gure illustrates the data interpolation based on already present values (grid points)

7. Conclusion and Outlook for Future Work

Image registration is one of the most important tasks when integrating and analyzing information from various sources. It is a key stage in image fusion, change detection, super-resolution imaging, and in building image information systems, among others [1]. This paper gives a survey of the classical and up-to-date registration methods, classifying them according to their nature as well as according to the four major registration steps. Although a lot of work has been done, automatic image registration still remains an open problem. Registration of images with complex nonlinear and local distortions, multimodal registration, and registration of N-D images (where N= 2) belong to the most challenging tasks at this moment. When registering images with non-linear, locally dependent geometric distortions, we are faced with two basic problems—how to match the CPs and what mapping functions to use for registration. While the second one can be solved at least on theoretical level by using appropriate radial basis functions, the first problem is generally unsolvable due to its nature. Since the between-image deformations can be arbitrary, we cannot use any automatic matching method. Another conceptual question here is how can we distinguish between image deformations and real changes of the scene. In multimodal registration, MI technique has become a standard reference, mainly in medical imaging. However, being an area-based technique, the MI has principal limitations. To overcome them, some authors combined the MI with other, preferably feature-based, methods to gain higher robustness and reliability. To speed up the computation, they often employed pyramidal image representation along with fast optimization algorithms. Unfortunately, when the images have significant rotation and/or scaling differences, these methods either fail or become extremely time expensive.

The future development in this field could pay more attention to the feature-based methods, where appropriate invariant and modality-insensitive features can provide good platform for the registration. Besides, we trust that new application-specific methods utilizing particular sensor characteristics appear soon in remote sensing. The major difficulty of N-D image registration resides in its computational complexity. Although the speed of computers has been growing, the need to decrease the computational time of methods persists. The complexity of methods as well as the size of data still grows (the higher resolution, higher dimensionality, larger size of scanned areas). Moreover, the demand for higher robustness and accuracy of the registration usually enforces solutions utilizing the iterations or backtracking, which also produces increase of computational complexity of the method. In the future, the idea of an ultimate registration method, able to recognize the type of given task and to decide by itself about the most appropriate solution, can motivate the development of expert systems. They will be based on the combination of various approaches, looking for consensus of particular results.

8. References

1. L.G. Brown, A survey of image registration techniques, ACM Computing Surveys 24 (1992) 326–376.
2. J. Flusser, T. Suk “A moment-based approach to registration with afine geometric distortion" IEEE Transactions on Remote Sensing, pag. 382-387 (1994).
3. A. Goshtasby, G.C. Stockman, C.V. Page “A region-based approach to registration with subpixel accuracy" Pattern Recognition Letters 24, pag. 1181-1190 (2003)
4. H.S. Alhichri, M. Kamel “Virtual circles: a new set of features for fast image registration" IEEE Transactions on Remote Sensing, pag. 390-399 { (1986)
5. P.F. Felzenszwalb, D.P. Huttenlocher “Distance Transforms of Sampled Functions"(2004).
6. D. Ziou, S. Tabbone “Edge detection techniques - An overview” (1997).
7. W. Forstner, E. Gulch”fast detector of distinct points, corners and circular features" Proceedings of ISPRS Workshop, pag. 281{305 (1986).
8. S.M. Smith, J.M. Brady “SUSAN - A New Approach to Low Level Image Processing" International Journal of Computer Vision (1997).
9. B. Zitov_a, J. Kautsky, G. Peter, J. Flusser “Robust detection of significant points in multiframe images" Pattern Recognition Letters 20, pag. 199-206 (1999)

10. K. Rohr “Landmark-based image analysis using geometric and intensity models" Computational Imaging and Vision series 21 (2001)
11. K. Mikolajczyk, C. Schmid “A performance evaluation of local descriptors" Proceedings of International Conference CVPR'03 (2003)
12. P. Hellier, C. Barillot, Coupling dense and landmark-based approaches for non rigid registration, IRISA research report, PI 1368:30, France, 2000.
13. Q. Chen, M. Defrise, F. Deconinck “Symmetric phase-only matched filtering of Fourier-Mellin transform" IEEE Transactions on Pattern Analysis 16, pag. 1156{1168. - (1994)
14. H. Foroosh, J.B. Zerubia, M. Berthod “Extension of phase correlation to subpixel registration" IEEE Transactions on Image Processing 11, pag. 188-200 -(2002)
15. P. Viola, W.M. Wells “Alignment by maximization of mutual information" International Journal of Computer Vision 24, pag. 137-154 - (1997)
16. P. Thevenaz, M. Unser “A mutual information optimizer for multiresolution image registration" Proceedings of the IEEE ICIP'98, Chicago, pag. 833-837 (1998)
17. N. Ritter, R. Owens, J. Cooper, R.H. Eikelboom, P.P. Saarloos “Registration of stereo and temporal images of the retina" IEEE Transactions on Medical Imaging 18, pag. 404-418 - (1999)
18. H.G. Barrow, J.M. Tenenbaum, R.C. Bolles, H.C. Wolf “Parametric correspondence and chamfer matching for image matching" Proceedings of the 5th IJCAI, pag. 659-663 - (1977)
19. G. Borgefors “Hierarchical chamfer matching: a parametric edge matching algorithm" IEEE Transactions on Pattern Analysis, pag. 849-865 - (1988)
20. S. Ranade, A. Rosenfeld “Point pattern matching by relaxation" Pattern Recognition 12, pag. 269-275 - (1980)
21. C.O.S. Sorzano, P. Thevenaz, M. Unser “Elastic registration of biological images using vector-spline regularization" IEEE Transactions on Biomedical Engineering 52:(4), pag. 652-663 - (2005)
22. R. Turcajova, J. Kautsky “A hierarchical multiresolution technique for image registration" Proceedings of SPIE Mathematical Imaging { (1996)
23. J. le Moigne “Registration of multi-sensor remote imagery using wavelet coefficients" Proceedings SPIE Wavelet Applications 2242, pag. 432-443 - (1994)
24. A. Goshtasby Piecewise linear mapping functions for image registration" Pattern Recognition 19, pag. 459-466 - (1986)
25. J. Flusser, Object matching by means of matching likelihood coefficients, Pattern Recognition Letters 16 (1995) 893–900.
26. Q. Zheng, R. Chellapa, A computational vision approach to image registration, IEEE Transactions on Image Processing 2 (1993) 311–325.
27. B. Zitova´, J. Flusser, F. Sroubek, Application of image processing for the conservation of the medieval mosaic, Proceedings of the IEEE International Conference on Image Processing ICIP’02, Rochester, MI, 2002, pp. 993–996.
28. M. Sester, H. Hild, D. Fritsch, Definition of ground control features for image registration using GIS data, Proceedings of the Symposium on Object Recognition and Scene Classification from Multispectral and Multisensor Pixels, CD-ROM, Columbus, Ohio, 1998, 7 pp.
29. . Flusser, B. Zitova´, Combined invariants to linear filtering and rotation, International Journal of Pattern Recognition and Artificial Intelligence 13 (1999) 1123–1136.
30. A. Goshtasby, Description and discrimination of planar shapes using shape matrices, IEEE Transactions on Pattern Analysis and Machine Intelligence 7 (1985) 738–743.
31. A.Goshtasby, G.C. Stockman, C.V. Page, A region-based approach to digital image registration with subpixel accuracy, IEEE Transactions on Geoscience and Remote Sensing 24 (1986) 390–399.
32. H. Li, B.S. Manjunath, S.K. Mitra, A contour-based approach to multisensor image registration, IEEE Transactions on Image Processing 4 (1995) 320–334.
33. W.H. Wang, Y.C. Chen, Image registration by control points pairing using the invariant properties of line segments, Pattern Recognition Letters 18 (1997) 269–281.
34. W. Peckar, C. Schnorr, K. Rohr, H.S. Stiehl, Two step parameterfree elastic image registration with prescribed point displacements, Journal of Mathematical Imaging and Vision 10 (1999) 143–162