Chapter 55
Combining Steerable Pyramid and Gaussian Mixture Models for Multi-Modal Remote Sensing Image Registration

Peng Ye and Fang Liu

Abstract Multi-modal remote sensing image registration is to align images acquired by different sensors and modalities. It is the fundamental step for following image analysis. Previous multi-resolution methods use spatial pyramids to achieve hierarchical registration with little consideration of the characteristics of pyramid transforms or robust point set registration methods after feature detection. Targeting at both problems, this paper proposes a novel image registration method by combining steerable pyramid and Gaussian mixture models. Steerable pyramid has been proved to be shift-invariant and outperforms traditional pyramid transform. Point set registration methods using Gaussian mixture model has been lately proposed and proved to be more robust and accurate than traditional point set registration methods. Experiments on real multi-modal remote sensing image pair demonstrate the feasibility of proposed method.

Keywords Steerable pyramid · Gaussian mixture models · Multi-modal image registration

55.1 Introduction

Multi-modal remote sensing image registration is the process to align images taken by different sensors and modalities. It provides insight into the analysis of the target otherwise cannot. For example, optical images usually have better resolution and understandable image features, but are greatly influenced by the imaging condition like illumination and clouds. Unlike passive imaging scheme of optical images, the active image mechanism makes SAR images rarely affected by the clouds or illuminations. SAR image are mainly determined by the reflection

P. Ye (✉) · F. Liu
ATR Lab, National Universtiy of Defense Technology, Changsha, China
e-mail: ye.peng.email@gmail.com

W. E. Wong and T. Ma (eds.), *Emerging Technologies for Information Systems, Computing, and Management*, Lecture Notes in Electrical Engineering 236, DOI: 10.1007/978-1-4614-7010-6_55,

characteristics of the target. However, SAR images usually suffer from speckle noise and have worse resolutions than optical image, making them more difficult to process. For better knowing of target, it is desirable to combine images of both modalities for further analysis. And multi-modal image registration is the crucial and fundamental step.

Remote sensing images usually adopt multi-resolution strategy to save the computation time and have a better convergence optimization. Previous multi-resolution methods used pyramid transforms like Gaussian pyramids or Laplace pyramids. Unlike steerable pyramid, these pyramid transforms have been proven to be shift-variant and have limited ability of extension [1]. Furthermore, as the basics of steerable pyramid, steerable filters provide a well theory-defined way for feature extraction, making steerable pyramid capable of being extended to detect feature [2]. The benefit of steerable pyramid is that the shift-invariant multi-resolution transformation and feature detection could be done at the same time.

For multi-modal image pair, it is the structural layout features that remain relatively unchanged and are the foundation for feature-based registration methods. It is therefore needed a point set registration method which preserves the structure layout of features and at the same time tolerates high amount of outliers. Traditional multi-resolution image registration methods used features extracted from different resolution as a scattered point set. Myronenko's research found out that previous point set registration methods lack analysis of the structure layout of point set or have little consideration of robustness [3]. The robustness of the point registration method in remote sensing was usually rather heuristically [4]. Point set registration methods developed on Gaussian mixture models have been lately proposed and proved to be robust and efficient than most traditional methods [3, 5]. The robustness against outliers is greatly valued in the multi-modal image registration case, where images usually present very different image features due to different imaging mechanism.

Remote sensing images are usually of large size with presence of a lot of details and complicated features. As a result, features in remote sensing image should be spatially diverse, easy to detect, high repeatable and less computation-load. Steerable pyramid transform not only could generate multi-resolution images but also could detect meaningful spatial diverse distribution features [4]. However this feature detection is rather coarse making a robust feature matching algorithm necessary. Also, multi-modal image registration makes outliers harder to deal with than mono-modal case. Point set registration methods based on Gaussian mixture models provide good ways to satisfy previous requirements. By combining steerable pyramid transform with Gaussian mixture models based point set registration methods, our method excels traditional registration methods in following aspects:

- the multi-resolution pyramid transform is shift-invariant;
- feature detection could be done within the multi-resolution transformation thus saving computation load;
- robust point set registration with focus on the structure layout of the image which is more suitable for the multi-modal image registration case.

55.2 Combining Steerable Pyramid and Gaussian Mixture Models

In this section, we first lay out the theory foundation of our method—namely steerable filters, steerable pyramid and Gaussian mixture model based point set registration; then we propose our new method.

55.2.1 Steerable Filters and Steerable Pyramid

Steerable pyramid transform, proposed by Simoncelli, is a linear multi-resolution, multi-orientation image decomposition transform [1]. It is developed in order to overcome the limitation of orthogonal separable wavelet decompositions. Because steerable filters are more robust to translation, rotation and noise than the standard Daubechies wavelet filters, they enable steerable pyramid to be shift-invariant and extendable for desired feature detection [1]. Table 55.1 shows the difference between popular pyramid transforms [1].

Figure 55.1 shows the decomposition (both analysis and synthesis) of steerable pyramid [1]. Initially, an image is separated into low- and high-pass subbands, using filters L_0 and H_0. The lowpass subband is then divided into a set of oriented bandpass subbands and a lower-pass subband. This lower-pass subband is subsampled by a factor of 2 in the X and Y directions. The recursive construction of a pyramid is achieved by inserting a copy of the shaded portion of the diagram at the location of the solid circle.

55.2.2 Gaussian Mixture Models for Feature Point Set Registration

Point set registration methods using GMM are various [3, 5]. We choose Coherent Point Drift (CPD) in this paper. CPD consider the alignment of two point sets as a

Table 55.1 Differences of popular pyramid transforms

	Steerable pyramid	Separable orthogonal wavelet	Laplacian pyramid	Gabor(octave)
Jointly-localized (space/ frequency)	Yes	Yes (can be)	Yes	Not inverse
Translation-invariant (no aliasing)	Yes (approx)	No	Yes (approx)	No
Oriented kernels	Yes	No (not diagonals)	N/A	Yes
Rotation-invariant (steerable)	Yes (approx)	No	N/A	No
Tight frame (self-inverting)	Yes (approx)	Yes	No	No

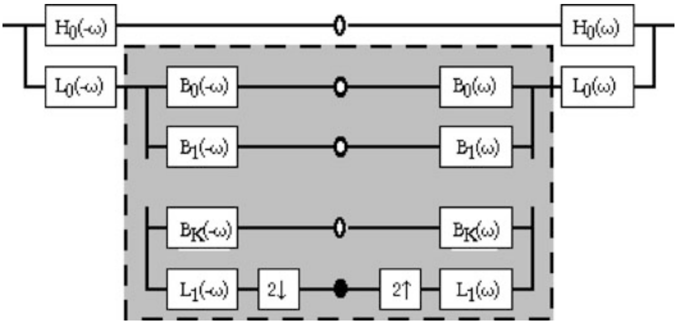

Fig. 55.1 Steerable pyramid

probability density estimation problem, where one point set represents the GMM centroids, and the other represents the data points [3]. Two point sets are aligned when the maximum GMM posterior probability is achieved. Denoting $X = (x_1, \ldots, x_n)^T$ the data point, $Y = (y_1, \ldots, y_n)^T$ the GMM centroids, the GMM probability density function is defined as:

$$p(x) = \omega p(x \mid M + 1) + (1 - \omega) \sum_{m=1}^{M} \mathrm{P(m)} p(x/m) \tag{55.1}$$

where $p(x/m) = \frac{1}{(2\pi\sigma^2)^{D/2}} \exp\left(-\frac{\|x - y_m\|^2}{2\sigma^2}\right)$. $p(x \mid M + 1) = 1/N$ is an additional uniform distribution with weight $\omega, 0 \le \omega \le 1$ added to the mixture models to account for noise and outliers. Equal isotropic covariance σ^2 and equal membership probabilities $P(m) = 1/M$ for all GMM components (m = 1,…,M) are used. And i.i.d (independent identical distribution) data assumption is made. Re-parameterize the GMM centroid locations by a set of parameters θ and estimate them by maximizing the likelihood or, equivalently by minimizing the negative log-likelihood function. The GMM density estimation problem use EM algorithm as optimization method to solve the parameters. By deduction, for the case of affine transformation $T(y_m; B, t) = By_m + t$, where $B_{D \times D}$ is an affine transformation matrix, $t_{D \times 1}$ is the translation vector, the objective function takes the form

$$Q(B, t, \sigma^2) = \frac{1}{2\sigma^2} \sum_{n,m=1}^{N,M} p^{old}(\mathrm{m} \mid x_n) \|x_n - (By_m + t)\|^2 + \frac{N_p D}{2} \log \sigma^2 \tag{55.2}$$

where $N_p = \sum_{n=1}^{N} \sum_{m=1}^{M} p^{old}(m|x_n) \le N$ (with $\mathrm{N} = N_p$ only if $\omega = 0$), p^{old} denotes the posterior probabilities of GMM components calculated using the previous parameter values,

$$p^{old}(m \mid x_n) = \frac{\exp\left(-\frac{1}{2}\left\|\frac{x_n - T(y_m,\theta^{old})}{\sigma^{old}}\right\|^2\right)}{\sum_{k=1}^{M}\exp\left(-\frac{1}{2}\left\|\frac{x_n - T(y_k,\theta^{old})}{\sigma^{old}}\right\|^2\right) + c}, \textit{with } c$$
$$= (2\pi\sigma^2)^{D/2}\frac{\omega}{1-\omega}\frac{M}{N} \tag{55.3}$$

We can directly take the partial derivatives of Q, equate them to zero, and solve the resulting linear system of equations.

55.2.3 Proposed Method

The basic functions of steerable pyramid—steerable filters—are directional derivative operators which come in different sizes and orientations. The number of orientation may be adjusted by changing the derivative order. Steerable filters are highly potential basis function for lots of image processing tasks. Steerable filters are used to detect image features like canny criteria [2]. Mikolajczyk's research found out that for low-dimensional descriptor, steerable filters outperformed other descriptors like differential invariants [6]. In this paper, pixels with large values from the subband image are used as input feature point set. Netanyahu, etc. also used steerable pyramid as multi-resolution transform and initial feature detection, but the feature point registration method they used was heuristic adaption of least median of squares (LMS) estimation, which is vulnerable to a more difficult case like multi-modal registration [4]. Moreover their transformation model was restrained to be rigid and strict to a hand-chosen small value range and the image pairs to be registered were mono-modal, making their method limited to special cases. On the contrary, by introducing a robust point set registration method, our method has much less restriction on transformation model and it is used in multi-modal case. The benefits of CPD are twofold: 1. the point set moves coherently during transformation thus preserving the structure; 2. the robust registration process between estimation-step and maximization step ensures global minimum.

The registration process starts with the coarsest scale, namely the scale with smallest image size. For each scale, top 10 % pixels with the largest value of the subband image are input as initial feature candidates; a following coarse registration is done with this input feature set using CPD algorithm; and the result of this coarse registration serves as initial transformation value for the following scale. This scale-to-scale registration ends when it reaches the final scale with the original image size.

Also the transformation model is set to be increasing in our method. At a coarse scale, the transformation model between two images is chosen to be less complicate than the following scale. We choose rigid for small scales and affine for large scales. This is mainly because an early incorporation of complicate transformation model would actually degrade the image registration or even cause failure [7]. At a coarse scale, the image suffers with less accurate details, thus the feature sets show relatively large structure similarity instead of detail local similarity. Therefore it is indeed needed to use a coarse transformation model at a coarse scale (Liu reviewed transformation models for image registration [8]). This hierarchical transformation model with hierarchical resolution image will ensure the optimization from falling into local minimum trap.

By combining steerable pyramid and Gaussian mixture model, we propose a hierarchical image registration method for multimodal image pair. Multi-modal image registration greatly suffers from outliers brought by feature detection method and different imaging modality. Traditional registration methods done on the original image with a high-order transformation model usually fails because of the outliers. Our method ensures an accurate results by: 1. a coarse registration at a coarse scale, saving time and ensuring a good initial guess for following scale; 2. a coarse transformation model at a coarse scale preventing falling into local optimization trap; 3. a robust point set registration method preserving the structure layout of the image which is more suitable for the multi-modal case.

55.3 Experiments and Analysis

Figure 55.2 shows the original optical-SAR image pair. It could be seen that optical image shows better resolution and better human-understandable features, while SAR image has better discrimination between different materials. The original optical image has a size of 1200*800, SAR image of size 600*400. Both images have been resized for displaying.

Figure 55.3 shows the multi-resolution and feature detection result. Each image has been decomposed into 3 levels. It can be seen that at a coarse scale, only the main structure features are present, thus making a registration with only a coarse transformation like rigid workable. At large scale, though more detail features are present, it is highly contaminated with outliers, which means it can easily fall into

(a) (b)

Fig. 55.2 Original optical-SAR image pair. **a** Optical image. **b** SAR image

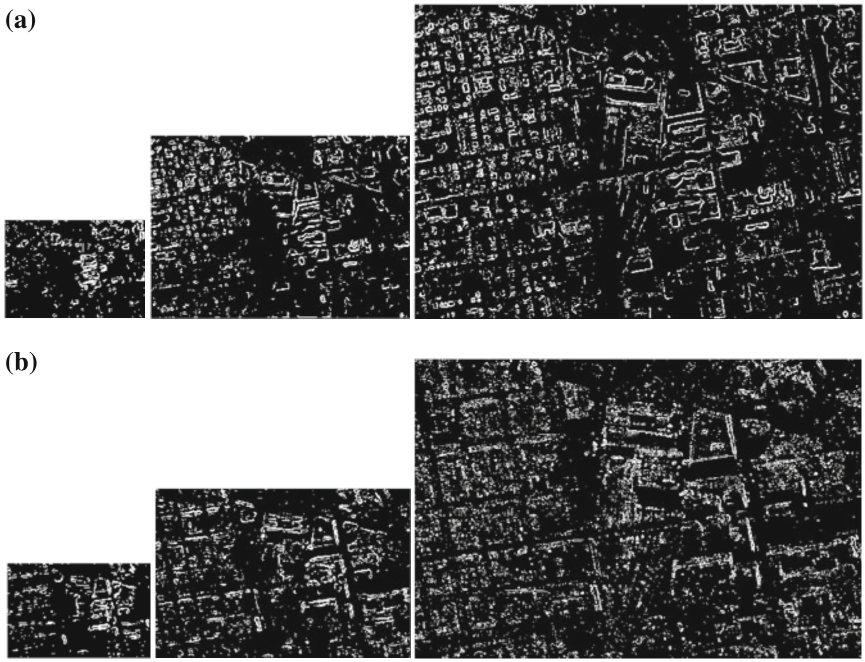

Fig. 55.3 Multi-resolution and feature detection of optical and SAR images. **a** Three level features representation of optical image. **b** Three level features representation of SAR images

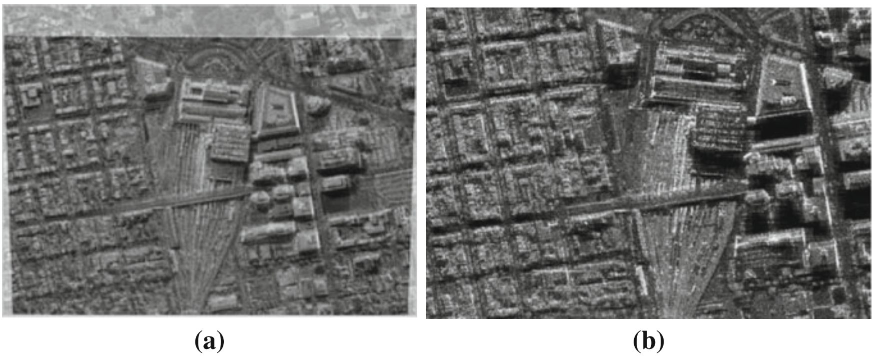

Fig. 55.4 Results of proposed method and CPD. **a** Our method. **b** Directly using CPD

local minimum trap without a good initial value. We set rigid transformation in the smallest size, and affine transformation in the following scales.

Figure 55.4 shows the registration result of using the proposed method and using CPD only. It could be seen that our method aligns the multi-modal image pair correctly, while the result of CPD is wrong (the result is not overlapped because it covers beyond the whole optical image). The RMSE (root mean square error) of our method is 1.34 pixels. Direct implementation of CPD fails mainly because the highly-outliers-contaminated feature point set makes CPD method at the original scale fall into local minimum, while our method assures the correct result with a well computed initial registration result from coarse scale and coarse transformation model.

55.4 Conclusion

In this paper, researchers proposed a novel multi-modal image registration method by combing steerable pyramid and Gaussian mixture model based point set registration method. Steerable pyramid is translation-invariant and rotation-invariant and easy to extend to detect features. Gaussian mixture model based point set registration methods preserve the structure layout of the image and are more robust to outliers than previous point set registration methods. Both methods provide promising characteristics for multi-modal image registration which is bother by the easy trap in local minimum and high amount of outliers. By combining both methods, researchers have achieved better results otherwise could not been done by using any one. Experiments on real optical-SAR image pair demonstrated the method's feasibility. Further research could be done on speed-up computation and extendable feature extraction of steerable pyramid.

References

1. Simoncelli, E.P., Freeman, W.T.: The steerable pyramid: a flexible architecture for multi-scale derivative computation. In: IEEE Second International Conference on Image Processing. Washington (1995)
2. Jacob, M., Unser, M.: Design of steerable filters for feature detection using canny-like criteria. IEEE Trans. Pattern Anal. Mach. Intell. **26**(8), 1007–1019 (2004)
3. Myronenko, A., Song, X.: Point set registration: coherent point drift. IEEE Trans. Pattern Anal. Mach. Intell. **32**(12), 2262–2275 (2009)
4. Netanyahu, N.S., Le Moigne, J., Masek, J.G.: Georegistration of landsat data via robust matching of multiresolution features. IEEE Trans. Geosci. Remote Sens. **42**(7), 1586–1600 (2004)
5. Jian, B., Vemuri, B.C.: Robust point set registration using Gaussian mixture models. IEEE Trans. Pattern Anal. Mach. Intell. **33**(8), 1633–1645 (2011)
6. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **27**(10), 1615–1630 (2005)
7. Yang, G., Stewart, C.V., Sofka, M., Tsai, C.-L.: Registration of challenging image pairs: initialization, estimation, and decision. IEEE Trans. Pattern Anal. Mach. Intell. **29**(11), 1973–1989 (2007)
8. Liu, W., Ribeiro, E.: A survey on image-based continuum-body motion estimation. Image Vis. Comput. **29**(8), 509–523 (2011)