

A New Framework for Automatic Registration of 2D/3D Texture Images

Ayman El-Baz¹ and Georgy Gimel'farb²

¹Bioengineering Department, University of Louisville, Louisville, KY, USA.

²Department of Computer Science, University of Auckland, Auckland, New Zealand.

Abstract. A new approach to align an image of a textured object with a given prototype is proposed. Visual appearance of the images, after equalizing their signals, is modeled with a Markov-Gibbs random field with pairwise interaction. Similarity to the prototype is measured by a Gibbs energy of signal co-occurrences in a characteristic subset of pixel pairs derived automatically from the prototype. An object is aligned by an affine transformation maximizing the similarity by using an automatic initialization followed by gradient search. Experiments confirm that our approach aligns complex 2D/3D objects better than popular conventional algorithms.

1 Introduction

Image registration aligns two or more images of similar objects taken at different times, from different viewpoints, and/or by different sensors. The images are geometrically transformed to ensure their close similarity. Registration is a crucial step in many applied image analysis tasks, e.g. to fuse various data sources (such as computer tomography (CT) and MRI data in medical imaging) for image fusion, change detection, or multichannel image restoration; form and classify multi-band images in remote sensing; update maps in cartography, perform automatic quality control in industrial vision, and so forth. Co-registered medical images provide more complete information about the patient, help to monitor tumor growth and verify treatment, and allow for comparing the patient's data to anatomical atlases.

Most of the known registration methods fall into two main categories: feature-based and area-based techniques [1]. Feature based techniques rely on salient local structures extracted from images, e.g. specific areas such as water reservoirs and lakes [2, 3], buildings [4], forests [5], or urban areas [6], specific lines like straight segments [7–9], object contours [10–12], coast lines [13, 14], rivers, or roads [15, 16], and specific points, e.g. road crossings [17], centroids of water areas, or oil and gas pads [18]. Scale invariant feature transform (SIFT) proposed by Lowe [19] is most popular at present because it reliably determines a number of point-wise correspondences between two images differing by affine transformation and local contrast / offset signal deviations. But these methods can be used only if objects have distinctive and non-repetitive local features.

Area-based methods such as the classical least square correlation match directly image signals to avoid feature extraction [20]. However, because the objects are assumed identical to within spatially uniform signal deviations, the correlation is too sensitive to non-uniform and spatially interdependent deviations of the corresponding signals due to sensor noise, illumination variations, and/or different sensor types. Alternative phase correlation and spectral-domain (e.g. Fourier-Mellin transform based) methods [21–25] are more robust with respect to the correlated and frequency dependent noise and non-uniform time varying illumination. However, these methods typically allow for only very limited geometric transformations.

More powerful mutual information (MI) based image registration [26, 27] exploits a probabilistic similarity measure that allows for more general types of signal deviations than correlation. The statistical dependency between two data sets is measured by comparing a joint empirical distribution of the corresponding signals in the two images to the joint distribution of the independent signals. Because the MI-based registration performs the best with multi-modal images [27], it is used in many of medical imaging applications. The joint distribution is estimated using Parzen windows [28, 29] or discrete histograms [30]. The main advantage of the MI is insensitivity to monotone variations of correspondence between the object and prototype signals, but the objects should be of almost identical shape apart from their affine geometrical and monotone signal transformations. The MI allows also for some non-monotone signal correspondence variations although they may change the visual appearance too much and hinder registration accuracy.

We consider a more general case of registering a textured object to a prototype with similar but not necessarily identical visual appearance under their relative 2D/3D affine transformations and monotone variations of signal correspondences. The variations are suppressed by equalizing signals in the images. The co-registered equalized images are described with a characteristic subset of signal co-occurrence statistics. The description implicitly “homogenizes” the images, i.e. considers them as spatially homogeneous patterns with the same statistics. In contrast to the feature-based registration, the statistics characterize the whole object. In contrast to the conventional area-based techniques, similarities between the statistics rather than pixel-to-pixel correspondences are measured. Section 2 represents the equalized object and prototype images as samples of a generic Markov–Gibbs random field (MGRF) with pairwise pixel interaction. Gibbs potentials are analytically estimated from co-occurrence statistics for the prototype. Similarity between an affinely transformed object and the prototype is measured with a total Gibbs energy for a characteristic pixel neighborhood. A new algorithm for selecting the neighborhood for the MGRF model is introduced. After an automatic initialization, the affine transformation aligning the object with the prototype is found by the gradient search for the maximum Gibbs energy of the transformed object. Experiments in Section 3 confirm that our method is more efficient for complex textured objects than more conventional SIFT and MI based registration techniques.

2 MGRF Based Image Registration

Basic notation. We denote $\mathcal{Q} = \{0, \dots, Q-1\}$; $\mathbf{R} = [(x, y, z) : x = 0, \dots, X-1; y = 0, \dots, Y-1; z = 0, \dots, Z-1]$, and $\mathbf{R}_p \subset \mathbf{R}$ a finite set of scalar image signals (e.g. gray levels), a 3D arithmetic lattice supporting digital LDCT data $g : \mathbf{R} \rightarrow \mathcal{Q}$, and its arbitrary-shaped part occupied by the prototype, respectively. A finite set $\mathcal{N} = \{(\xi_1, \eta_1, \zeta_1), \dots, (\xi_n, \eta_n, \zeta_n)\}$ of (x, y, z) -coordinate offsets defines neighbors $\{((x+\xi, y+\eta, z+\zeta), (x-\xi, y-\eta, z-\zeta)) : (\xi, \eta, \zeta) \in \mathcal{N}\} \wedge \mathbf{R}_p$ interacting with each pixel $(x, y, z) \in \mathbf{R}_p$. The set \mathcal{N} yields a neighborhood graph on \mathbf{R}_p to specify translation invariant pairwise interactions with n families $\mathcal{C}_{\xi, \eta, \zeta}$ of cliques $\mathcal{C}_{\xi, \eta, \zeta}(x, y, z) = ((x, y, z), (x+\xi, y+\eta, z+\zeta))$ (see Fig. 1). Interaction strengths are given by a vector $\mathbf{V}^T = [\mathbf{V}_{\xi, \eta, \zeta}^T : (\xi, \eta, \zeta) \in \mathcal{N}]$ of potentials $\mathbf{V}_{\xi, \eta, \zeta}^T = [V_{\xi, \eta, \zeta}(q, q') : (q, q') \in \mathcal{Q}^2]$ depending on signal co-occurrences; here T indicates transposition.

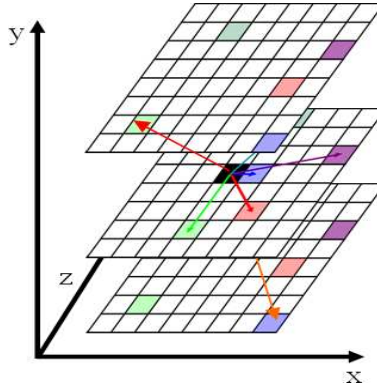


Fig. 1. Pairwise pixel interaction MGRF model.

Image normalization. To account for monotone (order-preserving) changes of signals (e.g. due to different illumination or sensor characteristics), the images are equalized using the cumulative empirical probability distributions of their signals.

MGRF based appearance model. Generic MGRF with multiple pairwise interaction (Fig. 1) [31], the Gibbs probability $P(g) \propto \exp(E(g))$ of an object g aligned with the prototype g° on \mathbf{R}_p is specified with the Gibbs energy

$$E(g) = |\mathbf{R}_p| \mathbf{V}^T \mathbf{F}(g) \quad (1)$$

where $\mathbf{F}^T(g)$ is the vector of scaled empirical probability distributions of signal co-occurrences over each clique family: $\mathbf{F}^T(g) = [\rho_{\xi, \eta, \zeta} \mathbf{F}_{\xi, \eta, \zeta}^T(g) : (\xi, \eta, \zeta) \in \mathcal{N}]$

\mathcal{N}] where $\rho_{\xi,\eta,\zeta} = \frac{|\mathcal{C}_{\xi,\eta,\zeta}|}{|\mathbf{R}_p|}$ is the relative size of the family and $\mathbf{F}_{\xi,\eta,\zeta}(g) = [f_{\xi,\eta,\zeta}(q, q' | g) : (q, q') \in \mathcal{Q}^2]^T$; here, $f_{\xi,\eta,\zeta}(q, q' | g) = \frac{|\mathcal{C}_{\xi,\eta,\zeta;q,q'}(g)|}{|\mathcal{C}_{\xi,\eta,\zeta}|}$ are empirical probabilities of signal co-occurrences, and $\mathcal{C}_{\xi,\eta,\zeta;q,q'}(g) \subseteq \mathcal{C}_{\xi,\eta,\zeta}$ is a sub-family of the cliques $c_{\xi,\eta,\zeta}(x, y, z)$ supporting the co-occurrence ($g_{x,y,z} = q, g_{x+\xi,y+\eta,z+\zeta} = q'$) in g . The co-occurrence distributions and the Gibbs energy for the object are determined over \mathbf{R}_p , i.e. within the prototype boundary after an object is affinely aligned with the prototype. To account for the affine transformation, the initial image is resampled to the back-projected \mathbf{R}_p by interpolation.

The appearance model consists of the neighborhood \mathcal{N} and the potential \mathbf{V} to be learned from the prototype.

Learning the potentials. The MLE of \mathbf{V} is proportional in the first approximation to the scaled centered empirical co-occurrence distributions for the prototype [31]:

$$\mathbf{V}_{\xi,\eta,\zeta} = \lambda \rho_{\xi,\eta,\zeta} \left(\mathbf{F}_{\xi,\eta,\zeta}(g^\circ) - \frac{1}{Q^2} \mathbf{U} \right); (\xi, \eta, \zeta) \in \mathcal{N}$$

where \mathbf{U} is the vector with unit components. The common scaling factor λ is also computed analytically; it is approximately equal to Q^2 if $Q \gg 1$ and $\rho_{\xi,\eta,\zeta} \approx 1$ for all $(\xi, \eta, \zeta) \in \mathcal{N}$. In our case it can be set to $\lambda = 1$ because the registration uses only relative potential values and energies.

Learning the characteristic neighbors. To find the characteristic neighborhood set \mathcal{N} , the relative energies $E_{\xi,\eta,\zeta}(g^\circ) = \rho_{\xi,\eta,\zeta} \mathbf{V}_{\xi,\eta,\zeta}^T \mathbf{F}_{\xi,\eta,\zeta}(g^\circ)$ for the clique families, i.e. the scaled variances of the corresponding empirical co-occurrence distributions, are compared for a large number of possible candidates. For example in 2D case, Fig. 2 shows a zebra prototype and its Gibbs energies $E_{\xi,\eta}(g^\circ)$ for 5000 clique families with the inter-pixel offsets $|\xi| \leq 50; 0 \leq \eta \leq 50$.

To automatically select the characteristic neighbors, we consider an empirical probability distribution of the energies as a mixture of a large “non-characteristic” low-energy component and a considerably smaller characteristic high-energy component: $P(E) = \pi P_{\text{lo}}(E) + (1 - \pi) P_{\text{hi}}(E)$. Both the components $P_{\text{lo}}(E), P_{\text{hi}}(E)$ are of arbitrary shape and thus are approximated with linear combinations of positive and negative discrete Gaussians (efficient EM-based algorithms introduced in [32] are used for both the approximation and estimation of π).

The intersection of the approximate mixture components gives an energy threshold θ for selecting the characteristic neighbors: $\mathcal{N} = \{(\xi, \eta, \zeta) : E_{\xi,\eta,\zeta}(g^\circ) \geq \theta\}$ where $P_{\text{hi}}(\theta) \geq P_{\text{lo}}(\theta)\pi/(1 - \pi)$. The above 2D example results in the threshold $\theta = 28$ producing 168 characteristic neighbors shown in Fig. 3 together with the corresponding relative pixel-wise energies $e_{x,y}(g^\circ)$ over the prototype:

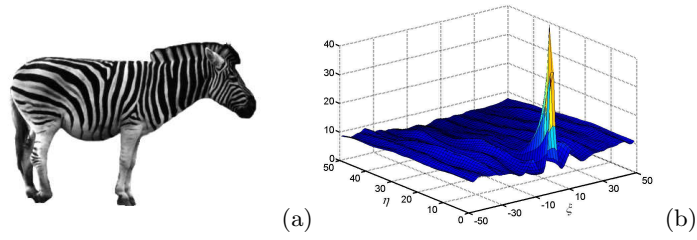


Fig. 2. Zebra prototype (a) and relative interaction energies (b) for the clique families in function of the offsets (ξ, η) .

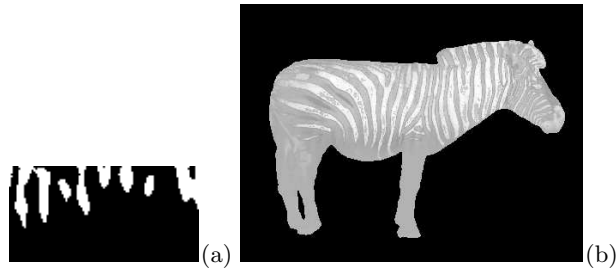


Fig. 3. (a) Characteristic 168 neighbors among the 5000 candidates (a; in white) and the pixel-wise Gibbs energies (b) for the prototype under the estimated neighborhood.

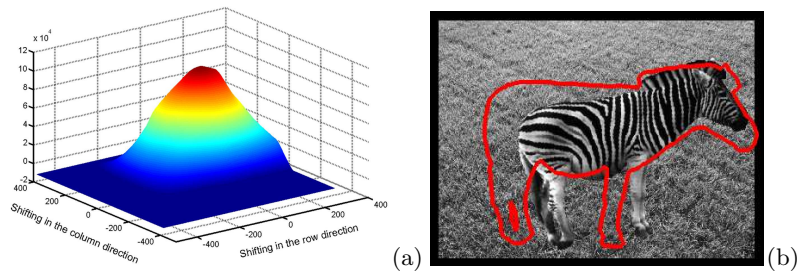


Fig. 4. (a) Gibbs energies for translations of the object with respect to the prototype and (b) initial position of the object with respect to the prototype.

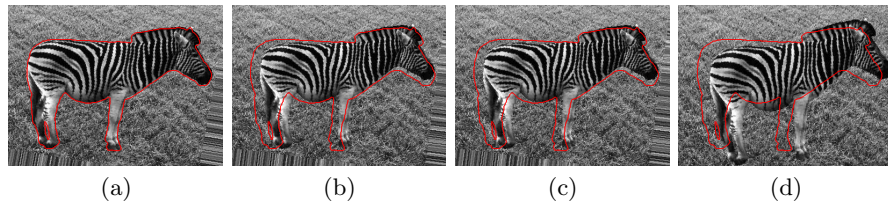


Fig. 5. Our (a), MI-based (b), NMI-based (c), and SIFT-based (d) registration.

$$e_{x,y}(g^\circ) = \sum_{(\xi,\eta) \in \mathcal{N}} V_{\xi,\eta}(g_{x,y}^\circ, g_{x+\xi,y+\eta}^\circ)$$

In similar way, the relative energy in the 3D case is calculating as follows:

$$e_{x,y,z}(g^\circ) = \sum_{(\xi,\eta,\zeta) \in \mathcal{N}} V_{\xi,\eta,\zeta}(g_{x,y,z}^\circ, g_{x+\xi,y+\eta,z+\zeta}^\circ)$$

Appearance-based registration. The object g is affinely transformed to (locally) maximize its relative energy $E(g_{\mathbf{a}}) = \mathbf{V}^T \mathbf{F}(g_{\mathbf{a}})$ under the learned appearance model $[\mathcal{N}, \mathbf{V}]$. Here, $g_{\mathbf{a}}$ is the part of the object image reduced to \mathbf{R}_p by the affine transformation $\mathbf{a} = [a_{11}, \dots, a_{33}]$: $x' = a_{11}x + a_{12}y + a_{13}z + a_{14}$; $y' = a_{21}x + a_{22}y + a_{23}z + a_{24}$; $z' = a_{31}x + a_{32}y + a_{33}z + a_{34}$. The initial transformation is a pure translation with $a_{11} = a_{22} = a_{33} = 1$; $a_{12} = a_{13} = a_{21} = a_{23} = a_{31} = a_{32} = 0$, ensuring the most “energetic” overlap between the object and prototype. For example in 2D case, the energy for different translations (a_{14}, a_{24}) of the object relative to the prototype is shown in Fig. 4(a); the chosen initial position (a_{14}^*, a_{24}^*) in Fig. 4(b) maximizes this energy. Then the gradient search for the local energy maximum closest to the initialization selects the six parameters \mathbf{a} ; Fig. 5 (a) shows the final transformation aligning the prototype contour to the object.

3 Experimental results and conclusions

Due to space limitations, we focus on zebra photos and low dose computed tomography (CT) of human chest commonly perceived as difficult for both the area- and feature-based registration. We compare our approach to three popular conventional techniques, namely, to the area-based registration using MI [27] or normalized MI [30] and to the feature-based registration establishing correspondences between the images with SIFT [19]. Results are shown in Fig. 5.

To clarify why the MI- or NMI-based alignment is less accurate, Fig. 6 compares the MI / NMI and Gibbs energy values for the affine parameters that appear at successive steps of the gradient search for the maximum energy. Both the MI and NMI have many local maxima that potentially hinder the search, whereas the energy is practically unimodal in these experiments. The SIFT-based alignment fails because it cannot establish accurate correspondences between similar zebra stripes (Fig. 7).

In the above example the object aligned with the prototype has mainly different orientation and scale. Figure 8 shows more diverse zebras and their Markov-Gibbs appearance-based and MI-based alignment with the prototype in Fig. 2(a). Visually, the back-projection of the prototype contour onto the objects suggests the better performance of our approach. To quantitatively evaluate the accuracy, masks of the co-aligned objects obtained by manual segmentation are averaged in Fig. 9. The common matching area is notably larger for our approach (91.6%) than for the MI-based registration (70.3%). Similar results obtained for the 3D Lung images are shown in Fig. 10.

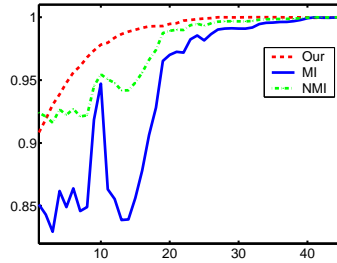


Fig. 6. Gibbs energy, MI, and NMI values at the successive steps of the gradient search.



Fig. 7. Corresponding points by SIFT.

4 Conclusions

In this paper we introduced a new approach to align an image of a textured object with a given prototype whose appearance is modeled with a Markov-Gibbs random field with pairwise interaction. Experimental results confirm that image registration based on our Markov-Gibbs appearance model is more robust and accurate than popular conventional algorithms. Due to the reduced variations between the co-aligned objects, our approach results in more accurate average shape models that are useful, e.g. in image segmentation based on shape priors.

As we mentioned in the experimental result section, the proposed approach is not only limited to zebra photos and lung images but also is suitable for registering starfish photos and brain images. The latter were not included in the paper because of the space limitations, but, the algorithms code, sample data and registration results for the starfishes, brain images will be provided in our web site.

Our future work will focus on integrating our approach in a framework of shape based segmentation; which we believe will enhance the accuracy of segmentation results of the existing approaches.

References

1. B. Zitova and J. Flusser, “Image registration methods: a survey,” *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.

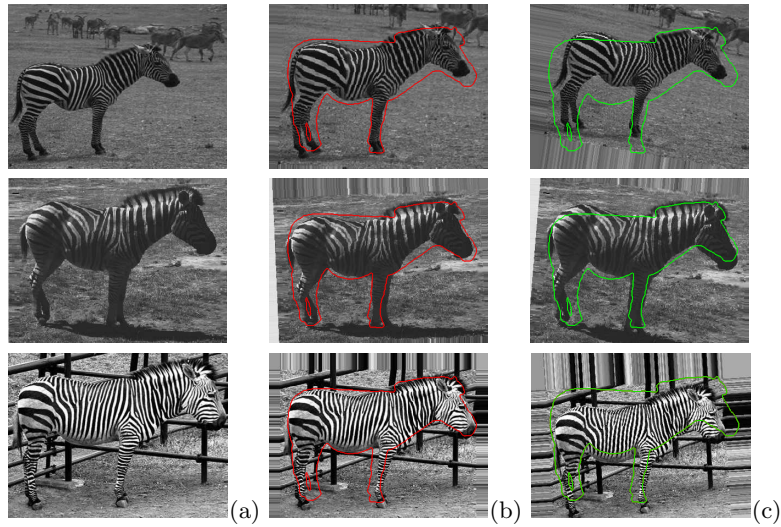


Fig. 8. Original zebras (a) aligned with our (b) and the MI-based (c) approaches.

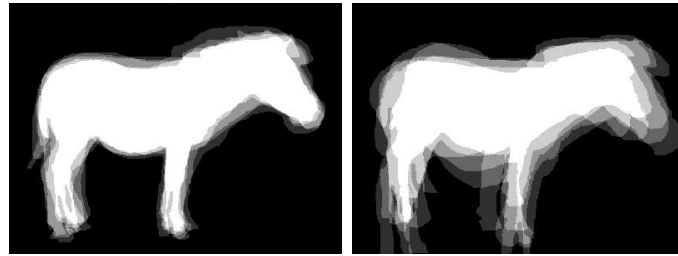


Fig. 9. Overlap between the object masks aligned with our (a; 91.6%) and the MI-based approaches (b; 70.3%).

2. A. Goshtasby and G. C. Stockman, "Point pattern matching using convex hull edges," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 15, pp. 631–637, 1985.
3. M. Holm, "Towards automatic rectification of satellite images using feature based matching," *Proc. Int. Geoscience and Remote Sensing Symp. IGARSS'91*, Espoo, Finland, 1991, pp. 2439–2442.
4. J. W. Hsieh, H. Y. M. Liao, K. C. Fan, M. T. Ko, and Y. P. Hung, "Image registration using a new edge-based approach," *Computer Vision and Image Understanding*, vol. 67, pp. 112–130, 1997.
5. M. Sester, H. Hild, and D. Fritsch, "Definition of ground control features for image registration using GIS data," *Proc. Symp. on Object Recognition and Scene Classification from Multispectral and Multisensor Pixels*, CD-ROM, Columbus, Ohio, 1998.
6. M. Roux, "Automatic registration of SPOT images and digitized maps," *Proc. IEEE Int. Conf. on Image Processing ICIP'96*, Lausanne, Switzerland, 1996, pp. 625–628.

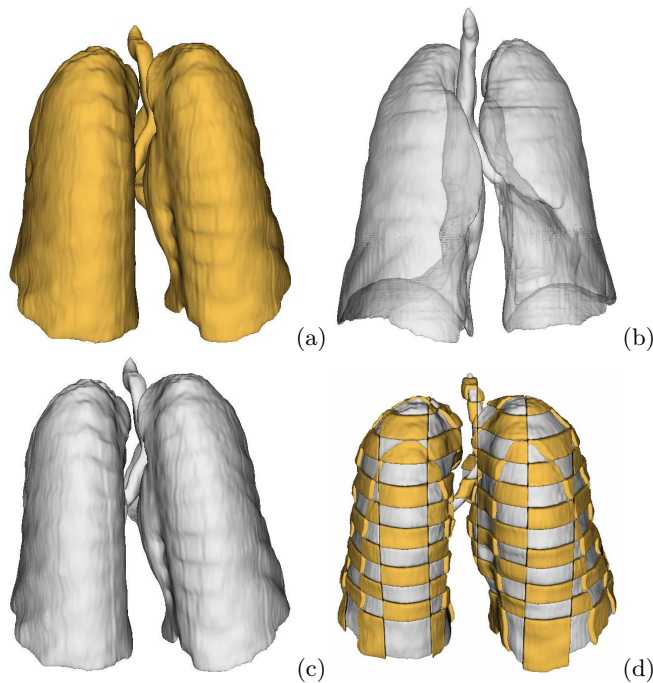


Fig. 10. (a) Reference data, (b) target data, (c) transformed data using 12 degree of freedom affine transformation, and (d) checkerboard visualization to show the quality of rigid alignment. Note that it can be seen that the connectivity at the edges of the lung, between the two volumes is not smooth when using the global model only, this is due to the local deformation which comes from breathing and heart beats.

7. Y. C. Hsieh, D. M. McKeown, and F. P. Perlant, "Performance evaluation of scene registration and stereo matching for cartographic feature extraction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, pp. 214–237, 1992.
8. S. Moss and E. R. Hancock, "Multiple line-template matching with EM algorithm," *Pattern Recognition Letters*, vol. 18, pp. 1283–1292, 1997.
9. W. H. Wang and Y. C. Chen, "Image registration by control points pairing using the invariant properties of line segments," *Pattern Recognition Letters*, vol. 18, pp. 269–281, 1997.
10. X. Dai and S. Khorram, "Development of a feature-based approach to automated image registration for multitemporal and multisensor remotely sensed imagery," *Proc. Int. Geoscience and Remote Sensing Symp. IGARSS'97*, Singapore, 1997, pp. 243-245.
11. V. Govindu, C. Shekhar, and R. Chellapa, "Using geometric properties for correspondenceless image alignment," *Proc. Int. Conf. on Pattern Recognition (ICPR'98)*, Brisbane, Australia, 1998, pp. 37–41.
12. H. Li, B. S. Manjunath, and S. K. Mitra, "A contour-based approach to multisensor image registration," *IEEE Trans. Image Processing*, vol. 4, pp. 320-334, 1995.
13. H. Maitre and Y. Wu, "Improving dynamic programming to solve image registration," *Pattern Recognition*, vol. 20, pp. 443–462, 1987.

14. D. Shin, J. K. Pollard, and J. P. Muller, "Accurate geometric correction of ATSR images," *IEEE Trans. Geoscience and Remote Sensing*, vol. 35, pp. 997–1006, 1997.
15. S. Z. Li, J. Kittler, and M. Petrou, "Matching and recognition of road networks from aerial images," *Proc. 2nd European Conf. on Computer Vision ECCV'92*, St Margherita, Italy, 1992, pp. 857–861.
16. E. H. Mendoza, J. R. Santos, A. N. C. S. Rosa, and N. C. Silva, "Land Use/land Cover Mapping in Brazilian Amazon Using Neural Network with Aster/terra Data," *Proc. Geo-Imagery Bridging Continents*, Istanbul, Turkey, 2004, pp. 123–126.
17. S. Grove and R. Tonjes, "A knowledge based approach to automatic image registration," *Proc. IEEE Int. Conf. on Image Processing ICIP'97*, Santa Barbara, California, 1997, pp. 228–231.
18. J. Ton and A. K. Jain, "Registering landsat images by point matching," *IEEE Trans. Geoscience and Remote Sensing*, vol. 27, pp. 642–651, 1989.
19. D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. of Computer Vision*, vol. 60, pp. 91–110, 2004.
20. Pope and J. Theiler, "Automated Image Registration (AIR) of MTI Imagery," *Proc. SPIE 5093*, vol. 27, pp. 294–300, 2003.
21. P. E. Anuta, "Spatial registration of multispectral and multitemporal digital imagery using Fast Fourier Transform," *IEEE Trans. Geoscience Electronics*, vol. 8, pp. 353–368, 1970.
22. Q. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transform for image registration and recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, pp. 1156–1168, 1994.
23. B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation and scale-invariant image registration," *IEEE Trans. Image Processing*, vol. 5, pp. 1266–1271, 1996.
24. G. Wolberg and S. Zokai, "Robust image registration using log-polar transform," *Proc. IEEE Int. Conf. on Image Processing*, Canada, September, 2000.
25. H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Trans. Image Processing*, vol. 11, pp. 188–200, 2002.
26. J. Inglada and A. Giros, "On the Possibility of Automatic Multisensor Image Registration," *IEEE Trans. Geoscience and Remote Sensing*, vol. 42, no. 10, October 2004.
27. P. Viola, "Alignment by Maximization of Mutual Information," *Ph.D. dissertation*, MIT, Cambridge, MA, 1995.
28. P. Thevenaz and M. Unser, "Alignment An efficient mutual information optimizer for multiresolution image registration," *Proc. IEEE Int. Conf. on Image Processing ICIP'98*, Chicago, IL, 1998, pp. 833–837.
29. P. Thevenaz and M. Unser, "A pyramid approach to sub-pixel image fusion based on mutual information," *Proc. IEEE Int. Conf. on Image Processing ICIP'96*, Lausanne, Switzerland, 1996, pp. 265–268.
30. C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition*, vol. 32, pp. 71–86, 1999.
31. G. Gimelfarb and A. A. Farag, "Texture Analysis by Accurate Identification of Simple Markov Models," *Ukrainian Journal of Cybernetics and Systems Analysis*, vol. 41, no. 1, pp. 37–49 January 2005.
32. G. Gimelfarb, A.A. Farag and A. El-Baz, "Expectation-Maximization for a linear combination of Gaussians," *Proc. of IAPR International Conference on Pattern Recognition, ICPR-2004, Cambridge, UK, August 2004*, pp. 422425.