Experiments on Robust Image Registration Using a Markov-Gibbs Appearance Model

Ayman El-Baz[1], Aly Farag[1], and Georgy Gimel'farb[2]

[1] Computer Vision and Image Processing Laboratory
University of Louisville, Louisville, KY 40292
{elbaz, farag}@cvip.louisville.edu
http://www.cvip.louisville.edu
[2] Department of Computer Science, Tamaki Campus
University of Auckland, Auckland, New Zealand
g.gimelfarb@auckland.ac.nz

Abstract. A new approach to align an image of a textured object with a given prototype under its monotone photometric and affine geometric transformations is experimentally compared to more conventional registration algorithms. The approach is based on measuring similarity between the image and prototype by Gibbs energy of characteristic pairwise co-occurrences of the equalized image signals. After an initial alignment, the affine transformation maximizing the energy is found by gradient search. Experiments confirm that our approach results in more robust registration than the search for the maximal mutual information or similarity of scale-invariant local features.

1 Introduction

The goal of image registration is to co-align two or more images of the same or similar objects acquired by different cameras, at different times, and from different viewpoints. Thus the images have to be photometrically and geometrically transformed in order to make them closely similar. Co-aligned images provide more complete information about the object and allow for building adequate object models.

Registration is a must in many applications, e.g. medical imaging, automated navigation, change detection in remote sensing, multichannel image restoration, cartography, automatic quality control in industrial vision, and so on [1]. Feature based registration relies on easily detectable local areal, linear, and point structures in the images, e.g. water reservoirs and lakes [2], buildings [3], forests [4], urban areas [5], straight lines [6], specific contours [7], coast lines [8], rivers, or roads [9], road crossings [10], centroids of water areas, or oil and gas pads [11]. In particular, the scale invariant feature transform (SIFT) [12] can reliably determine a collection of point-wise correspondences between two images under the affine geometric transformation and local contrast/offset photometric transformations. But these methods work only with distinctive and non-repetitive local features.

Alternative area-based registration, e.g. the least square correlation obviates the need for feature extraction due to direct matching of all image signals [13].

D.-Y. Yeung et al. (Eds.): SSPR&SPR 2006, LNCS 4109, pp. 65–73, 2006.

However, the correlation assumes spatially uniform contrast/offset transformations and a central-symmetric pixel-wise noise with zero mean. As a result, it frequently fails under non-uniform and spatially interdependent photometric transformations caused by different sensors and varying illumination. Phase correlation and spectral-domain (Fourier-Mellin transform based) methods [14] are less sensitive to the correlated and frequency dependent noise and non-uniform time-variant illumination but allow for only very limited geometric transformations.

Recent image registration by maximizing mutual information (MI) [15] presumes a most general type of photometric transformations, namely, any monotone transformation of the corresponding signals in one of the images. The similarity between two images is measured by the Kullback-Leibler divergence of a joint empirical distribution of the corresponding signals from the joint distribution of the independent signals. This approach performs the best with multi-modal images [15] and thus is widely used in medical imaging. The joint distribution is usually estimated using Parzen windows [16] or discrete histograms [17]. But the MI is invariant also to some non-monotone photometric transformations that change the images too much. The unduly extensive invariance of the MI hinders the registration accuracy.

This paper considers one further area-based registration method assuming that a textured object and its prototype have similar but not necessarily identical visual appearance under affine geometric and monotone photometric transformations of the corresponding signals. The latter transformations are suppressed by equializing both the prototype and the image area matched to it. The equalized prototype is described with a characteristic set of Gibbs potentials estimated from statistics of pairwise signal co-occurrences. The description implicitly considers each image as a spatially homogeneous texture with the same statistics. In contrast to more conventional area-based registration techniques, the similarities between the statistics rather than pixel-to-pixel correspondences are involved.

2 MGRF Based Image Registration

Basic notation. Let $\mathcal{Q} = \{0, \ldots, Q-1\}$; $\mathbf{R} = [(x, y) : x = 0, \ldots, X-1; y = 0, \ldots, Y-1]$ be a finite set of scalar image signals (e.g. gray levels) and a rectangular arithmetic lattice, respectively. The latter supports digital images $g : \mathbf{R} \to \mathcal{Q}$, and its arbitrary-shaped part $\mathbf{R}_{\mathsf{p}} \subset \mathbf{R}$ supports a certain prototype of an object of interest.

Let a finite set $\mathcal{N} = \{(\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)\}$ of (x, y)-coordinate offsets define neighbors $\{((x+\xi, y+\eta), (x-\xi, y-\eta)) : (\xi, \eta) \in \mathcal{N}\} \wedge \mathbf{R}_{\mathsf{p}}$ interacting with each pixel $(x, y) \in \mathbf{R}_{\mathsf{p}}$. The set $\mathcal{N}$ produces a neighborhood graph on $\mathbf{R}_{\mathsf{p}}$ specifying translation invariant pairwise interactions. The latter are restricted to n families $\mathcal{C}_{\xi,\eta}$ of second order cliques $c_{\xi,\eta}(x, y) = ((x, y), (x+\xi, y+\eta))$ of the graph. Interaction strength in each family is specified with the Gibbs potential function $\mathbf{V}^{\mathsf{T}}_{\xi,\eta} = \left[V_{\xi,\eta}(q, q') : (q, q') \in \mathcal{Q}^2\right]$ of the signal co-occurrences in the clique. The total interaction strength is given by the potential vector $\mathbf{V}^{\mathsf{T}} = \left[\mathbf{V}^{\mathsf{T}}_{\xi,\eta} : (\xi, \eta) \in \mathcal{N}\right]$ where T indicates the transposition.

MGRF based appearance model. The monotone (order-preserving) transformations of the image signals may occur due to different illumination or sensor characteristics. To make the registration (almost) insensitive to these transformations, both the prototype and conforming to it part of each image are equalized using cumulative empirical probability distributions of their signals on $\mathbf{R}_\mathsf{p}$. In line with a generic MGRF model with multiple pairwise interaction [18], the probability $P(g) \propto \exp(E(g))$ of an object g aligned with the prototype g° on $\mathbf{R}_\mathsf{p}$ is proportional to the Gibbs energy $E(g) = |\mathbf{R}_\mathsf{p}|\mathbf{V}^\mathsf{T}\mathbf{F}(g)$ where $\mathbf{F}^\mathsf{T}(g) = [\rho_{\xi,\eta}\mathbf{F}^\mathsf{T}_{\xi,\eta}(g) : (\xi,\eta) \in \mathcal{N}]$ is the vector of the scaled empirical probability distributions of signal co-occurrences over each clique family; $\rho_{\xi,\eta} = \frac{|\mathcal{C}_{\xi,\eta}|}{|\mathbf{R}_\mathsf{p}|}$ is the relative size of the family; $\mathbf{F}_{\xi,\eta}(g) = [f_{\xi,\eta}(q,q'|g) : (q,q') \in \mathcal{Q}^2]^\mathsf{T}$ with $f_{\xi,\eta}(q,q'|g) = \frac{|\mathcal{C}_{\xi,\eta;q,q'}(g)|}{|\mathcal{C}_{\xi,\eta}|}$ are the empirical probabilities of signal co-occurrences, and $\mathcal{C}_{\xi,\eta;q,q'}(g) \subseteq \mathcal{C}_{\xi,\eta}$ is a subfamily of the cliques $c_{\xi,\eta}(x,y)$ supporting the same co-occurrences $(g_{x,y} = q,\ g_{x+\xi,y+\eta} = q')$ in g.

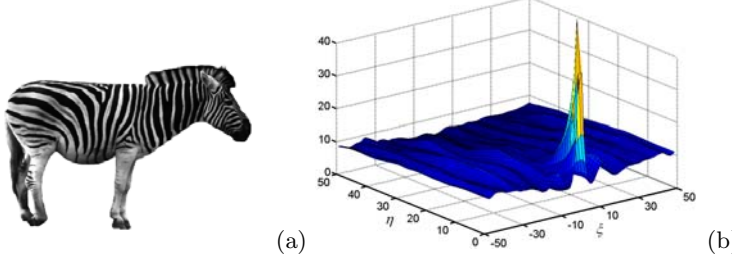
The co-occurrence distributions and the Gibbs energy for the object are determined over $\mathbf{R}_\mathsf{p}$, i.e. within the prototype boundary after an object is geometrically transformed to be aligned with the prototype. To account for the transformation, the initial image is resampled to the back-projected $\mathbf{R}_\mathsf{p}$ by interpolation.

The appearance model consists of the neighborhood $\mathcal{N}$ and the potential $\mathbf{V}$ to be learned from the prototype. The approximate MLE of $\mathbf{V}$ is proportional to the scaled centered empirical co-occurrence distributions for the prototype [18]:

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

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 $\rho_{\xi,\eta} \approx 1$ for all $(\xi,\eta) \in \mathcal{N}$. In our case it can be set to $\lambda = 1$ because the registration needs only relative potential values and energies.

Learning the characteristic neighbors. To find the characteristic neighborhood set $\mathcal{N}$, the top relative energies $E_{\xi,\eta}(g^\circ) = \rho_{\xi,\eta}\mathbf{V}^\mathsf{T}_{\xi,eta}\mathbf{F}_{\xi,\eta}(g^\circ)$ for the clique families, i.e. the scaled variances of the corresponding empirical co-occurrence distributions, have to be separated for a large number of low-energy candidates.



(a) (b)

Fig. 1. Zebra prototype (a) and the relative interaction energies (b) for the clique families in function of the offsets (ξ,η)

Figure 1 shows a zebra prototype and its Gibbs energies $E_{\xi,\eta}(g^\circ)$ for the 5,100 clique families with the inter-pixel offsets $|\xi| \le 50$; $0 \le \eta \le 50$.

To automatically select the characteristic neighbors, let us 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: $P(E) = \pi P_{\rm lo}(E) + (1-\pi)P_{\rm hi}(E)$. Because both the components $P_{\rm lo}(E)$, $P_{\rm hi}(E)$ can be of arbitrary shapes, we closely approximate them with linear combinations of positive and negative Gaussians. For both the approximation and the estimation of π, we use the efficient EM-based algorithms introduced in [19].

The intersection of the approximated low- and high-energy distributions gives an energy threshold θ for selecting the characteristic neighborhood $\mathcal{N} = \{(\xi,\eta) : E_{\xi,\eta}(g^\circ) \ge \theta\}$, that is, the threshold solves the equation $P_{\rm hi}(\theta) = P_{\rm lo}(\theta)\pi/(1-\pi)$. The above example results in the threshold $\theta = 28$ producing the 168 characteristic neighbors shown in Fig. 2 together with the corresponding relative pixel-wise energies $e_{x,y}(g^\circ)$ over the prototype:

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

Appearance-based registration. Let $g_{\mathbf{a}}$ denote a part of the object image reduced to $\mathbf{R_p}$ by the affine transformation $\mathbf{a} = [a_{11},\ldots,a_{23}]$: $x' = a_{11}x + a_{12}y + a_{13}$; $y' = a_{21}x + a_{22}y + a_{23}$. To align with the prototype, the object g should be

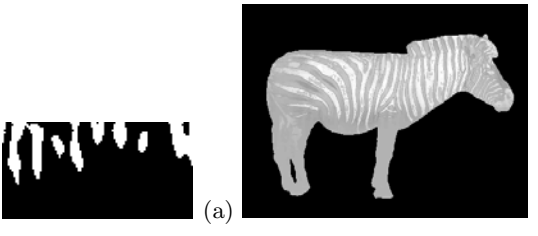

(a) (b)

Fig. 2. Characteristic 168 neighbors among the 5100 candidates (a; in white) and the gray-coded relative pixel-wise Gibbs energies (b) for the prototype under the estimated neighborhood

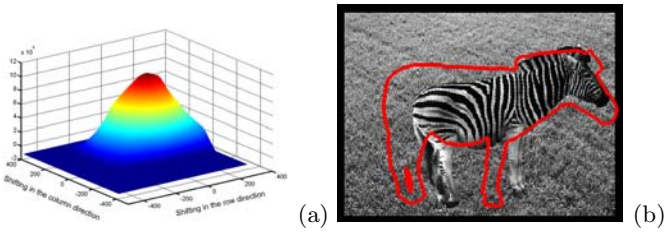

(a) (b)

Fig. 3. Gibbs energies for the object's translations (a) with respect to the prototype and the resulting initial relative position of the object

affinely transformed to (locally) maximize its relative energy $E(g_{\mathbf{a}}) = \mathbf{V}^{\mathsf{T}}\mathbf{F}(g_{\mathbf{a}})$ under the learned appearance model $[\mathcal{N}, \mathbf{V}]$.

The initial transformation is a pure translation with $a_{11} = a_{22} = 1$; $a_{12} = a_{21} = 0$, ensuring the most "energetic" overlap between the object and prototype. The energy for the different translations (a_{13}, a_{23}) of the object relative to the prototype and the chosen initial position (a^*_{13}, a^*_{23}) maximizes this energy are shown in Fig. 3.

Then the gradient search for the local energy maximum closest to the initial point in the affine parameter space selects the six parameters $\mathbf{a}$. Figure 4 (a) illustrates the final alignment by back-projecting the prototype's contour to the object.

3 Experimental Results and Conclusions

Experiments have been conducted with several types of images. Below we discuss results obtained for zebra images available on the Internet (they include both artificial collages and natural photos) and for natural medical images such as dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) of human kidneys and low dose computed tomography (LDCT) images of human lungs. These image types are commonly perceived as difficult for both the area- and feature-based registration. The like results have been obtained for other images of complex textured objects, e.g. starfish images available on the Internet and MRI of human brain. In total, we used in these experiments 24 zebra, 40 starfish, 200 kidney, 200 lungs, and 150 brain images.

We compared our approach to three popular conventional techniques, namely, to the area-based registration using the MI [15] or the normalized MI [17] and to the feature-based registration by establishing inter-image correspondences with the SIFT [12]. Results for the above zebra image are shown in Fig. 4. The SIFT-based alignment fails because the SIFT could not establish accurate correspondences between the similar zebra stripes (see Fig. 5).

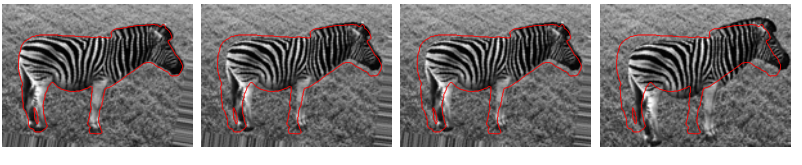


Fig. 4. From left to right: our, MI-, NMI (normalized MI)-, and SIFT-based registration

Fig. 5. Corresponding points by SIFT

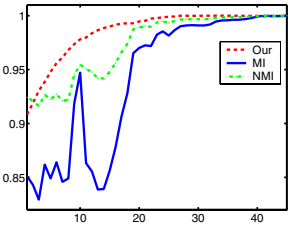


Fig. 6. Gibbs energy, MI, and NMI values at the successive steps of the gradient search

The lower accuracy of the MI- and NMI-based alignment comparing to our approach can stem from a notably different behavior of the MI / NMI and the Gibbs energy values in the space of the affine parameters. Figure 6 presents these 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 close to unimodal in this case.

In the above example the object aligned with the prototype differed mainly by its orientation and scale. Figure 7 shows more diverse zebra objects and results of their Markov-Gibbs appearance-based and MI-based alignment with the prototype in Fig. 1(a). The results are illustrated by the back-projection of the prototype contour onto the objects. Visually, these results suggest that our approach has the better performance. To quantitatively evaluate the registration accuracy, the manually segmented masks of the co-aligned objects are averaged in Fig. 8. The common matching area for our approach (91.6%) is considerably larger than for the MI-based registration (70.3%).

Fig. 7. Original zebras (a) aligned with our (b) and the MI-based (c) approach

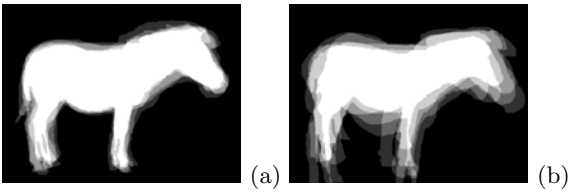


Fig. 8. Overlap between the object masks aligned with our (a; 91.6%) and the MI-based approaches (b; 70.3%)

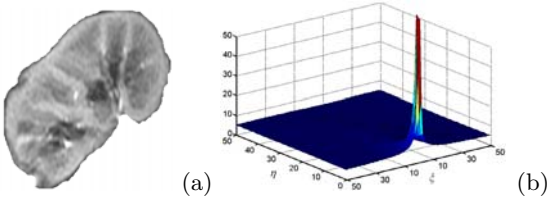


Fig. 9. Kidney image (a) and relative interaction energies (b) for the clique families in function of the offsets (η, ξ)

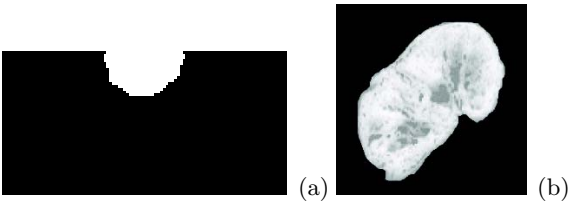


Fig. 10. (a) Most characteristic 76 neighbors among the 5,100 candidates (a; in white) and the pixel-wise Gibbs energies (b) for the prototype under the estimated neighborhood

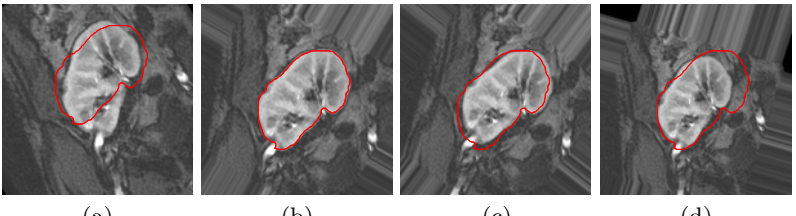
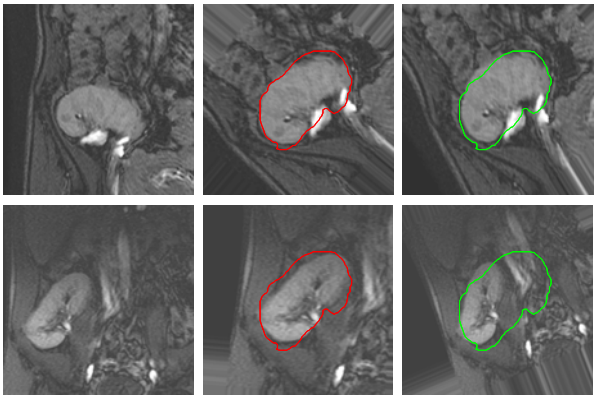


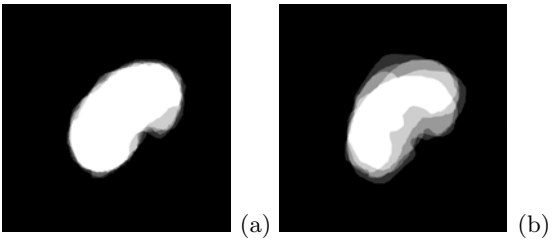
Fig. 11. Initialization (a) and our (b), the MI- (c), and the SIFT-based (d) registration

Similar results obtained for the kidney images are shown in Figs. 9–13: the common matching area 90.2% is for our approach vs. 62.6% for the MI-based one. Therefore, image registration based on our Markov-Gibbs appearance model is more robust and accurate than popular conventional algorithms. Due to reduced variations between the co-aligned objects, it results in more accurate average shape models that are useful, e.g. in image segmentation based on shape priors.



(a) (b) (c)

Fig. 12. Original kidneys (a) aligned with our (b) and the MI-based (c) approach



(a) (b)

Fig. 13. Overlap between the object masks aligned with our (a; 90.2%) and the MI-based (b; 62.6%) approach

References

1. B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.
2. 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, 1991.
3. 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.
4. 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.
5. 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.
6. 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.

7. 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, 1997.
8. 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.
9. 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, 2004.
10. S. Growe 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, 1997.
11. J. Ton and A. K. Jain, "Registering landsat images by point matching," *IEEE Trans. Geoscience and Remote Sensing*, vol. 27, pp. 642–651, 1989.
12. D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. of Computer Vision*, vol. 60, pp. 91–110, 2004.
13. Pope and J. Theiler, "Automated Image Registration (AIR) of MTI Imagery," *Proc. SPIE 5093*, vol. 27, pp. 294–300, 2003.
14. 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.
15. P. Viola, "Alignment by Maximization of Mutual Information," *Ph.D. dissertation, MIT*, Cambridge, MA, 1995.
16. 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, USA, 1998*, pp. 833–837, 1998.
17. 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.
18. G. Gimelfarb and A. A. Farag, "Texture Analysis by accurate identification of simple Markov models," *Cybernetics and Systems Analysis*, vol. 41, no. 1, pp. 37–49, 2005.
19. G. Gimelfarb, A.A. Farag and A. El-Baz, "Expectation-Maximization for a linear combination of Gaussians," *Proc. of 18$^{\text{th}}$ IAPR Int. Conf. on Pattern Recognition (ICPR-2004), Cambridge, UK, August 2004*, pp. 422–425, 2004.