Registration of Stereo and Temporal Images of the Retina

Nicola Ritter,* *Member, IEEE*, Robyn Owens, *Member, IEEE*, James Cooper, *Member, IEEE*, Robert H. Eikelboom, *Member, IEEE*, and Paul P. van Saarloos, *Member, IEEE*

***Abstract*— The registration of retinal images is required to facilitate the study of the optic nerve head and the retina. The method we propose combines the use of mutual information as the similarity measure and simulated annealing as the search technique. It is robust toward large transformations between the images and significant changes in light intensity. By using a pyramid sampling approach combined with simulated reannealing we find that registration can be achieved to predetermined precision, subject to choice of interpolation and the constraint of time. The algorithm was tested on 49 pairs of stereo images and 48 pairs of temporal images with success.**

***Index Terms*—Hierarchical systems, image registration, information theory, simulated annealing, visual system.**

I. Introduction

IMAGE registration is a task required in many applications of image processing where one must take measurements that involve multiple images. It consists of the alignment of these images so as to either eliminate the differences between them or highlight the salient differences for the purpose of study.

Image registration can be categorized by considering the type of sensor used, the position of the sensors, the time difference between the images, and differences between the image subjects. Using this classification, Brown [1] divides registration into four categories.

1) Multimodal registration, where images of the same subject have been taken by different types of sensors, for example, computed tomography (CT), positron emission tomography (PET), and magnetic resonance (MR) images of the brain.
2) Viewpoint registration, where the same subject is imaged with two similar sensors at the same moment, but from different positions, for example, stereo imaging.
3) Temporal registration, where the same subject is photographed from the same viewpoint, but at different times, for example, the analysis of myocardial function under conditions of rest and stress.
4) Template registration, where a reference image is to be found inside a larger image taken with the same type of sensor, for example, the location of a specific feature on a map.

To this we would add a fifth category.

5) Intersubject registration, where two similar images taken with the same type of sensor are to be aligned, for example, the faces of two people.

Some of the more important registration tasks are found in the field of biomedical computing. For example, when patients are to undergo brain surgery, both PET and MR images of the brain are used to help guide the surgeon's hand. To enable surgeons to use effectively both sets of information, the images must be aligned [2]–[4].

Another major biomedical use of image registration is for retinal images. Retinal or fundus photographs are a standard diagnostic tool in ophthalmology. Serial photographs of the flow of fluorescein dye are used to determine areas of ischemia, hemorrhaging, neovascularization, and occlusions [5] in diseases such as diabetic retinopathy. To determine the progression of glaucoma, the optic-nerve-head topography is assessed from color stereo photographs and the nerve fiber layer is assessed from red-free photographs [6]. Changes in color indicate cupping and notching of the optic disk, and nerve fiber layer loss is shown by changes in the density of the film. Drusen (deposits in the retinal layers), macular and retinal holes, and retinal pigments are some of the other pathologies that can be seen on retinal photographs.

While most of the observations from retinal photographs are subjective in nature, much effort has also been put into making objective measurements: Morgan *et al.* have studied drusen [7], Eikelboom *et al.* [8] have studied the nerve fiber layer. Optic disk cupping has been the study of Chihara and Chihara [9], while Mizuno *et al.* [68] have studied retinitis pigmentosa. For these objective studies it is necessary that the images be registered. This registration generally involves relatively large x translations, due to changes between sittings for temporal images and smaller y translations from changes in position of the chin cup. Rotation occurs due to tilting of the head and through ocular torsion, and scaling is caused by changes in the distance between the camera and the head, due to equipment changes or differing head positions.

Registration, in any field, involves the use of pixel intensity

Manuscript received October 6, 1997; revised March 31, 1998. This work was supported in part by the Lion's Eye Institute. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. Viergever. *Asterisk indicates corresponding author*.

*N. Ritter, R. H. Eikelboom, and P. P. van Saarloos are with the Centre for Ophthalmology and Visual Science, The University of Western Australia, Perth, W.A., Australia, and are also with The Lions Eye Institute, Perth, W.A., Australia (e-mail: nritter@cs.uwa.edu.au).

R. Owens is with the Department of Computer Science, The University of Western Australia, Perth, W.A., Australia.

J. Cooper is with The School of Computer and Information Science, Edith Cowan University, Perth, W.A., Australia.

Publisher Item Identifier S 0278-0062(99)05468-3.

(or other derived measure) within the two images to find a geometric transformation that maps one image onto the other as accurately as possible. Some techniques are suitable only for finding translations alone [10]–[12], some can handle only rigid body transformations [13], while others allow affine transformations that include aspect ratio changes and shearing [14]. A further class of transformations are those called polynomial transformations, where a polynomial function is applied across the whole image [15].

In this paper we use a method for computing global rigid and affine transformations only, and do not tackle techniques or strategies for dealing with polynomial transformations [15], local distortions [16], [17], or the range of methods that deal with elastic deformations [18]–[24]. We will show that mutual information combined with simulated annealing forms a robust method for registration of retinal image pairs and can be found to a predetermined precision.

II. Current Methods for Affine Registration

A general review of registration methods is given by Maintz and Viergever [25] and a comparison of methods for registration of images of the brain by West *et al.* [26].

The registration of two images I and J using an affine transformation can be defined as follows. Let $\mathcal{A}_\alpha$ be one of the possible transformations, described by a matrix, where α is some generalized index. The transformation $\mathcal{A}^*$ is then the $\mathcal{A}_\alpha$ that optimizes a given comparison measure $C(I, \mathcal{A}_\alpha(J))$.

In the case where C is actually a measure of difference between the images, and hence should be minimized, we say that $\mathcal{A}^*$ is the argument that minimizes, over all indexes α, the measure C

$$\mathcal{A}^* = \arg\left\{\min_\alpha \left[C(I, \mathcal{A}_\alpha(J))\right]\right\}. \tag{1}$$

Therefore, the process of image registration can be seen as having two distinct parts: the calculation of a suitable comparison measure and the application of a suitable search technique to find the optimum value of that measure.

A. Preprocessing

Preprocessing of the image involves detection of such features as edges or contours [27], ridges [28], regions [29], feature distance information [3], interest or characteristic points [30], [31], or depth information [32]. The images resulting from this detection are then used in the registration process instead of the original images, allowing the use of comparison measures that would not be robust when applied to the original images.

Christensen *et al.* [20] first produce three-dimensional (3-D) edge-contour maps from CT, MR, and PET image sequences. The new images are binary, showing contours rather than multimodal color image brain scans. Jiang *et al.* also use contour diagrams, but they translate the sensor images into distance images where each pixel is assigned a value associated with the distance of that pixel from the contour.

Chen and Medioni [32] work with registration of range images. A rough solution is found using some (unspecified) simple registration technique, then surfaces are found and, finally, point matching occurs. This somewhat complicated technique is used to produce comprehensive 3-D data from a series of images taken as an object rotates on a platform in front of a camera.

Although preprocessing speeds up the actual registration, it adds a time overhead at the start and introduces an entirely new field of calculation into the image registration problem. Furthermore, these techniques are not generic, as detection of features is not easy in all images. Preprocessing also reduces the amount of information being used to do the actual search and this may limit the accuracy.

An alternative technique involves control-point matching where marked points on the object are used to aid in the image-registration problem. Chen and Medioni [32] use points on 3-D range data images that occur in locally smooth regions, and Backman *et al.* [33] use anatomic landmarks or fiducials marked with ultra-violet sensitive ink. However, clearly, these methods are not possible for images such as those of the retina.

B. Comparison Measures

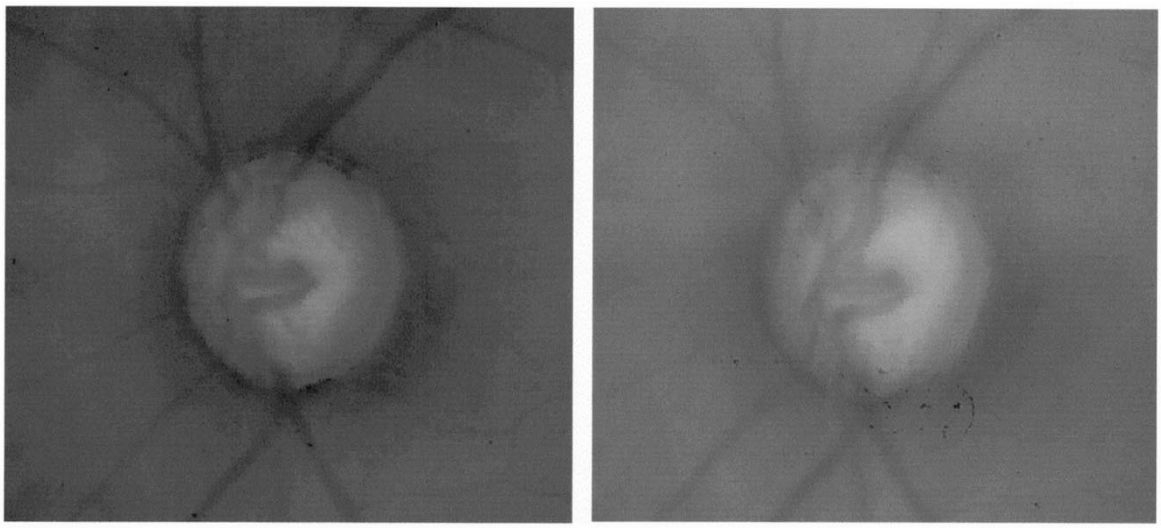
The comparison measure is the calculation used to judge the closeness of fit (or otherwise) between the two images under the current transformation.

The most common measure used in the past has been cross-correlation (for example, [29], [34], and [35]). Crosscorrelation involves taking a template from the first image and then calculating the crosscorrelation between it and various parts (or windows) of the second image. The crosscorrelation measure is given by

$$C_{\mathrm{CC}} = \frac{\sum_k \sum_l T(k,l) J(x+k, y+l)}{\sqrt{\sum_k \sum_l J^2(x+k, y+l)}} \tag{2}$$

where (k,l) refers to the position in the template T and (x,y) is the position in the second image J given by the current transformation. Clearly, the cost of this grows with the square of the size (width) of the template and the square of the size of the search region. Thus, depending on the search strategy (see Section III-B), it can be very slow and is useful only when searching a relatively small sample space of possible transformations. Since it is purely a measure of the closeness of fit between the pixel intensities in the template and the current window into the second image, it is useful only where the lighting remains reasonably constant, there is little white noise, and the sensors are of similar type.

Another problem with crosscorrelation is the choice of template, which must contain significant data. The choice of template therefore entails some kind of preprocessing involving feature detection. However, some retinal images, such as those in Fig. 1, cause problems with automatic feature detection, due to poorly defined features in the image. Studholme *et al.* [36] use the correlation coefficient across the whole image rather than crosscorrelation shown in (3) at the bottom of the next page where $\bar{I}$ and $\bar{J}$ are the average pixel intensities of the two $M \times N$ images. However, this method will still fail when the two images have very different colors or lighting, particularly when the lighting differs nonlinearly across the image.



(a) (b)

Fig. 1. Poor photographic images of the retina, such as these, are caused by lens, vitreous, or corneal opacities or when the pupil dilation is too small to allow sufficient reflected light from the retina to reach the fundus camera sensor. (a) Left stereo image. (b) Right stereo image.

Yu *et al.* [37] use the sum of the absolute values of the differences (SAD) as their comparison measure

$$C_{\mathrm{SAD}} = \sum_k \sum_l |J(x+k, y+l) - T(k,l)| \tag{4}$$

where (k,l) indexes into the template image and (x,y) is the current position in the second image. Other related possibilities are the normalized sum of differences (NSD)

$$C_{\mathrm{NSD}} = \sum_k \sum_l |[J(x+k, y+l) - \bar{J}] - [T(k,l) - \bar{T}]| \tag{5}$$

the sum of differences squared (SDS) [1]

$$C_{\mathrm{SDS}} = \sum_k \sum_l [J(x+k, y+l) - T(k,l)]^2 \tag{6}$$

and the root mean square difference (RMSD) [3]

$$C_{\mathrm{RMSD}} = \sqrt{\frac{1}{mn} \sum_k \sum_l [J(x+k, y+l) - T(k,l)]^2}. \tag{7}$$

These methods still do not work well when there are large intensity differences between images, when there are large amounts of white noise, and when the two images are from different types of equipment.

Zabih and Woodfill [38] describe two nonparametric measures or ordinal measures which are applied as transforms to the original images: the rank transform and the census transform. With these transforms, for each pixel, the pixels in the neighborhood are assigned a one if they are of intensity less than the central pixel or zero otherwise, producing binary data. The rank transform is then simply the sum of these values, in effect, the number of pixels of lower intensity than the central pixel. Zabih and Woodfill then use the sum of absolute differences as the comparison measure. The census transform arranges the binary data into a bit string and the comparison is then performed on the Hamming distance between two strings: that is, the number of bits that differ. Their results show some success with artificial images.

Bhat and Nayar [39] also describe an ordinal measure for use with stereo registration. As with crosscorrelation, this method involves the comparison of a template from the first image with portions of the second image. Each pixel in the template is given a ranking based on its pixel intensity. These rankings are then compared to the rankings in the current window of the second image. This technique avoids problems of lighting changes between the images, but does not deal with multimodal registration.

Another group of methods can be classed as phase-based comparison methods [4], [41]–[44]. These involve using the Fourier transform and, thus, the optimization search is performed in the frequency domain.

The Fourier transform $\mathcal{F}$ of an image $I(x,y)$ is a complex function at each frequency (ω_x, ω_y)

$$\mathcal{F}(\omega_x, \omega_y) = \mathcal{R}(\omega_x, \omega_y) + \mathbf{i}\mathcal{I}(\omega_x, \omega_y). \tag{8}$$

This can also be expressed in exponential form

$$\mathcal{F}(\omega_x, \omega_y) = |\mathcal{F}(\omega_x, \omega_y)| e^{\mathbf{i}\phi(\omega_x, \omega_y)} \tag{9}$$

where

$$|\mathcal{F}(\omega_x, \omega_y)| = \sqrt{\mathcal{R}^2 + \mathcal{I}^2} \tag{10}$$

$$C_C = \frac{\frac{1}{MN}\sum_x \sum_y I(x,y)J(x,y) - \bar{I}\bar{J}}{\sqrt{\left(\frac{1}{MN}\sum_x \sum_y I^2(x,y) - \bar{I}^2\right)\left(\frac{1}{MN}\sum_x \sum_y J^2(x,y) - \bar{J}^2\right)}} \tag{3}$$

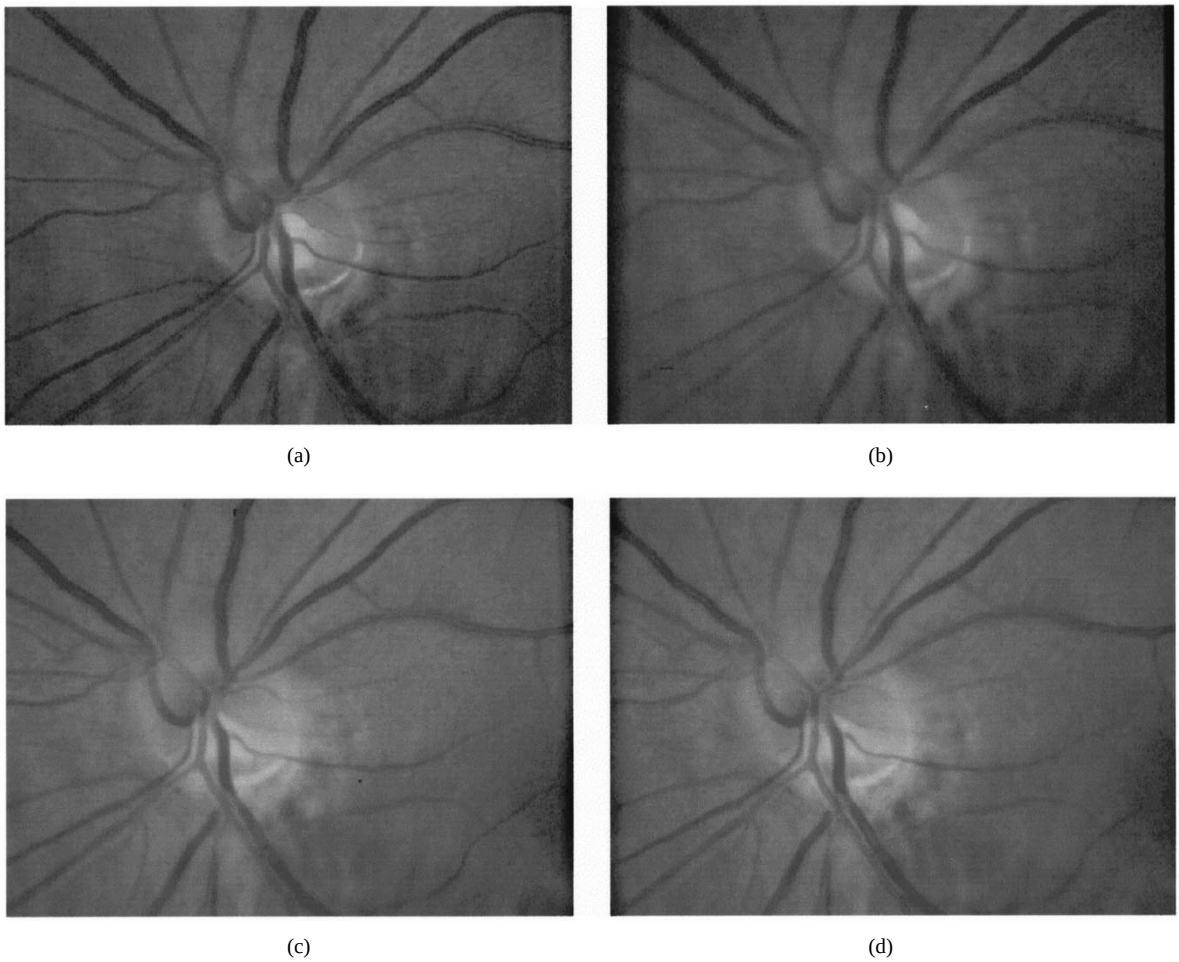


Fig. 2. Temporal stereo pairs of the retina of a patient's left eye, taken in 1995 and 1997, using a Nidek stereo fundus camera. The original images were taken with color-slide film and then digitized with a Polaroid slide digitizer. The images used for registration were 256 grey scale. (a) 1995 left stereo image. (b) 1995 right stereo image. (c) 1997 left stereo image. (d) 1997 right stereo image.

is the amplitude of the transform and

$$\phi(\omega_x, \omega_y) = \tan^{-1}\left[\frac{\mathcal{I}(\omega_x, \omega_y)}{\mathcal{R}(\omega_x, \omega_y)}\right] \tag{11}$$

is the phase.

Two images I and J that are displaced by (d_x, d_y) have the same Fourier amplitude, but a phase difference that is directly related to the displacement

$$\mathcal{F}_J(\omega_x, \omega_y) = e^{-\mathbf{i}(\omega_x d_x + \omega_y d_y)}\mathcal{F}_I(\omega_x, \omega_y). \tag{12}$$

Phase-based methods are more robust to lighting variations [1] than SD-based methods and can be very fast since much of the computation can be performed in special purpose hardware. However, they do not work well with images containing significant amounts of white noise and are restricted under (12) to small translations. Rotation can be included in phase-based methods, but only at the cost of either time or robustness [1].

A new comparison measure has recently emerged from two separate sources ([45] and [14]), with further discussion by Wells *et al.* [46] and Maes *et al.* [47]. This technique is based on the statistical measure called mutual information, described in detail in Section III-A. In essence, it works on the basis of how well one image explains the other.

Mutual information has several major advantages over other comparison measures. It is robust toward changes in light intensity, pixel color, and noise, it can find large transformations, and it works well where there is occlusion [48]. It also works for multimodal registration where the images have been taken from different sensors, such as for MR and CT registration [26]. It can be fully automated and requires no preprocessing of the images and has been found to be both precise and robust [49].

C. Search Techniques

Many common minimization strategies have been used in image registration problems. For example, Mendonça *et al.* [13] use an exhaustive search, Studholme *et al.* [36] use gradient descent, Cideciyan *et al.* [35] use the simplex method, Starink and Backer [31] use simulated annealing, Mandava *et al.* [5] compare simulated annealing and genetic algorithms, whereas Pelizzari *et al.* [2] and Collignon *et al.* [14] use Powell's minimization [51].

Many methods also incorporate a multiresolution strategy which involves iterative calculations of the registration either from a coarse derivation of the image down to the image itself [30], or using subsampling of the image [48]. Samoilenko and

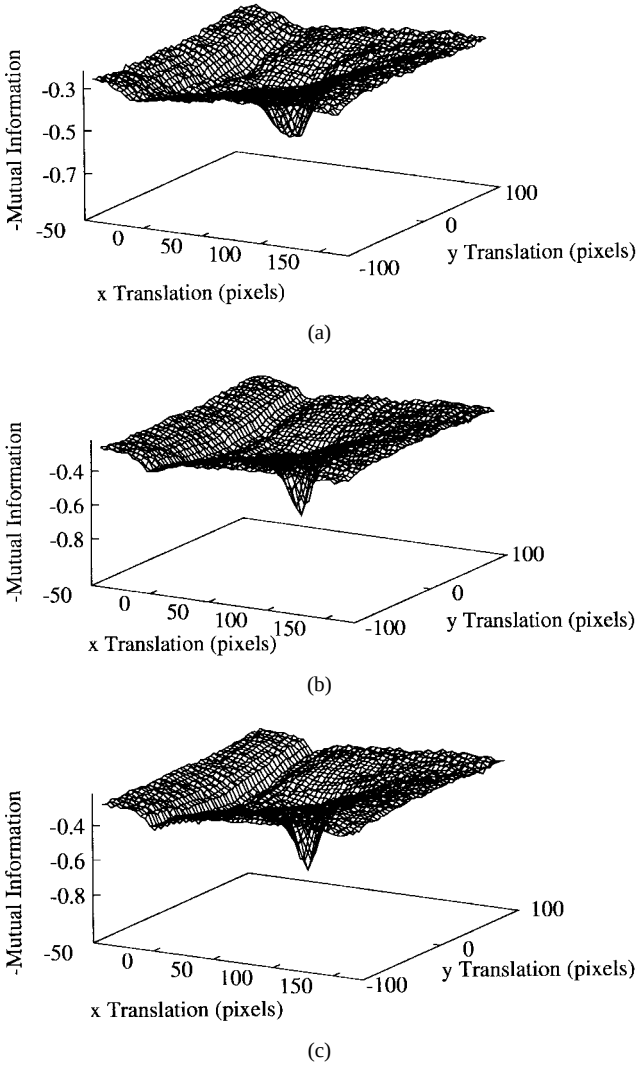


Fig. 3. Crosssections of the mutual-information manifold formed by the parameters of rotation, scaling, and x & y translation, for the retinal images in Fig. 2(b) and (d). A sampling of one in eight pixels was used. (a) shows a crosssection where rotation and scaling were held constant at a quite incorrect value, (b) shows a slightly incorrect value for rotation and scaling, and (c) shows a correct pair of values. While the correct values are obvious, it can be seen that there are many local minima and the function does not approximate the quadratic particularly well. (a) Rotation and scale incorrect by 2 units. (b) Rotation and scale incorrect by 0.5 units. (c) Rotation and scale correct.

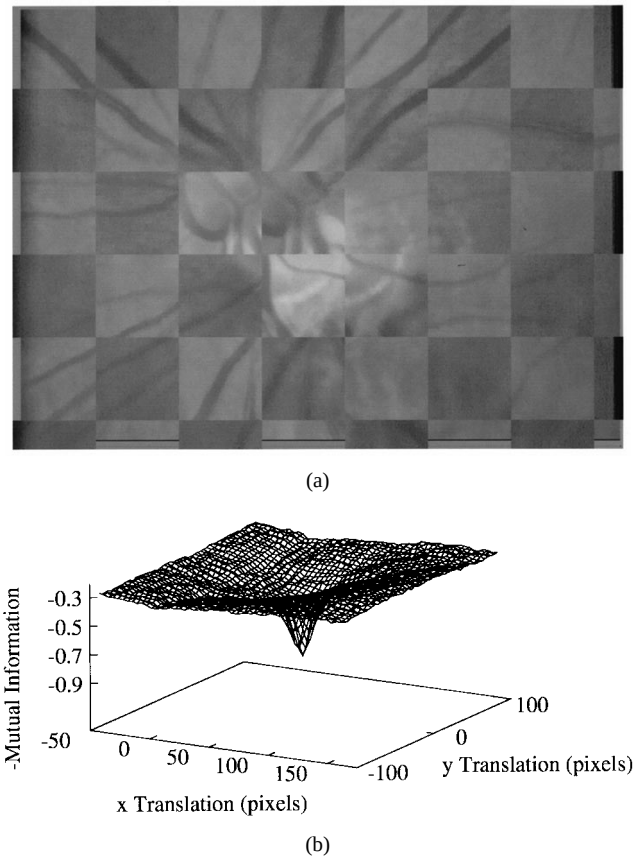


Fig. 4. The graphs in Fig. 3 show a local minimum trench parallel to the y axis. The source of this trench can be seen at the left-hand side of the image in Fig. 4(a) where there is a vertical band of darker pixels in both images. The validity of this hypothesis is demonstrated in Fig. 4(b) which is an identical graph as for that in Fig. 3, but for images from which this pixel band has been clipped. (a) A checkerboard image of the false registration that coincides with the trench that appears in the graphs shown in Fig. 3(b). Removal of the pixel band that appears to cause the false registration shown by the trench in Fig. 3 gives the graph above: the trench shows a significant reduction in depth.

Yakovlev [52] use a method similar to multiresolution where they smear the original image and work on that to find a first approximation to the required transformation.

D. Registration of Retinal Images

A major area of biomedical image registration is that of retinal image photography where the ocular fundus is imaged with either a fundus camera or a scanning laser ophthalmoscope. Examples of fundus camera retinal images can be seen in Figs. 1 and 2. Ophthalmologists analyze images of the retina, optic nerve head and surrounding areas, to find signs of disease [53], and the use of digitized images allows easier study of the areas involved as well as the ability to enhance those images. Barry *et al.* [54] and Yogesan *et al.* [69] use stereo images to study the optic disk for signs of progression of glaucoma. These images need to be registered to enable the clinician to gain a 3-D view of the retina and optic nerve. Photographs are retaken after 12 to 18 months and compared with the original images, again requiring registration. Fundus images also show signs of arterial hypertension and arteriosclerosis [55] and can be used in the diagnosis of such diseases. Peli *et al.* [56] preprocess the first image using an adaptive threshold procedure. This allows selection of vessel points in a template, since the vessels in retinal images are usually much darker than the surrounding tissue. These vessel points are then used with the normalized sum of differences [see (5)] to calculate a comparison measure between the template and the second image. This method produces pixel-level registration (for x and y translation only) in approximately 3 s, using an exhaustive search. However, it is not robust toward large changes in image intensity and white noise, such as those shown in Fig. 1.

Yu *et al.* [37] suggest a method for registering retinal images using the absolute value of the differences of pixel intensities as the comparison measure. They process the images twice, using the optic disc as a feature for coarse alignment of the images and the blood vessels (thinned to stems) as features

for fine alignment. They give no detail as to what search strategy is used and test their method only for pairs of images produced artificially by randomly translating a single retinal image. Thus, their method is not tested for scaling or rotation and there is no difference in intensities between the two images or normal white noise. While this method might be useful for some stereo pairs of images, it is unlikely that it would succeed for temporal images, where the image intensities are generally quite different and both scaling and rotation may have occurred.

We have found consistently that no matter what method is chosen for registration, excellent results are found when comparing an image with a transformed copy of the same image and that these results cannot be used to assume that a registration technique is good for either stereo or temporal pairs of images where differences in light intensity, changes in the imaged objects, and white noise become significant factors.

Cideciyan *et al.* [35] use a multistage method for registration of retinal images. In the first stage they compute the Fourier spectrums of the two images. If the first of the original images is a scaled, rotated, and translated version of the other then, in the Fourier spectrum, images the first image will be a rotated and scaled version of the second with the rotation and scaling parameters being the same as with the original images. In the second stage the log polar transformation of the Fourier spectrum images is computed, which converts the rotation and scaling differences to a translation difference. The rotation and scaling parameters are then computed from the log-polar image pair using crosscorrelation. These parameters are then used to rotate and scale one of the original images. The translation parameters can then be found via crosscorrelation on the resulting pair. Whilst Cideciyan *et al.* report some good results for images taken at the same sitting, images taken at different times, when the x translation may be large, would cause problems with the use of the Fourier spectrum. Changes in light intensity between sittings could also cause problems with the use of crosscorrelation in the second and third stages of their method.

Jasiobedzki [57] suggests the use of *active contours* which conform to edges found by previously processing the images. These contours are active in the sense that they are controlled by an energy level that they try to minimize. This energy level is affected by the shape of the contour and the surrounding image. The active contours for the first image are then mapped into an adaptive adjacency graph, which is simply a network of active contours where nodes are replaced by springs that allow free movement of the contours meeting at that node. If the graph of one image is placed over another image that has the same topology of contours, then the graph will steadily move to match the second image, allowing registration.

The main disadvantage to this method is that the active contours are attracted only to features that they are initially close to. This means that the method would fail whenever the translation between images is large. In temporal images it is possible for the x translation to be as much as 70 pixels. This would make it impossible for the contours to map to the correct features in the second image. Jasiobedzki tests only artificially created pairs of images which, as mentioned previously, do not give a good idea as to whether the method is robust for real pairs of images.

Hart and Goldbaum [58] use control point pairs to register retinal images. The control points used are those defined by the ends of blood vessel segments which have been detected using a blood vessel filter [59]. These are then paired with control points from the second image by an undisclosed method. The next step involves removing those pairs that were more than 100 pixels apart on a coordinate system based on the estimated center of the optic nerve. A further reduction in pairs is then made by correlating the pixel intensities of a 15×15-pixel window centered on the control points. Erroneous pairs are now eliminated by comparing the x and y scaling factors of the pairs. Those where the two factors are quite different are discarded on the assumption that the aspect ratio between the images is unlikely to vary greatly. Pairs that are too close to previously accepted pairs are also eliminated.

A selection of the remaining pairs is then chosen and a least squares method is used to fit the current transformation to these pairs. The estimate of the transformation is then iteratively refined by removing the control point pair with the greatest error and the pair that, with its exclusion, gives the best overall error reduction. The process is then terminated when either there are four control points only left or the mean square error drops below 5.0 pixels.

Hart and Goldbaum's results show that this method does not work with every retinal pair and cannot provide subpixel accuracy. As it uses correlation as a measure of comparison it will also have problems with temporal images that are very different in color or intensity. This is also true for images taken from patients with glaucoma. One of the signs of this disease is the loss of nerve fibers on the retina and topography of the optic disc, which can cause a significant change between the pixel intensities of parts of the two images. A further problem comes from the use of the estimated location of the optic nerve, as errors in this can affect the final registration. Finally, it is documented by Miszalok [55] that retinal blood vessels do not maintain the same position over time.

A more recent comparison method suggested specifically for temporal retinal image pairs is that of edge-point-coordinate detection and is described by Mendonça *et al.* [13]. This method involves using classical edge-detection techniques on the two images and then quantifying these data as two matrices for each image: one recording the horizontal component of the location of edge features and the other storing the vertical component. Registration then becomes the task of aligning the two matrices to get the best fit possible, the measure of fit being calculated from the addition of the matches between the first and second pair of image matrices.

To perform the actual registration a template area is chosen from the first pair of matrices and then matched within the pair of matrices from the second image. This is done using an exhaustive search. While this comparison method clearly solves the problem of different intensities, it leaves two major problems unsolved: it deals only with translation, not rotation or scale changes, and it does not give any information on how to choose the template to be matched. While the largest type of misalignment of retinal images is translational (as can be

seen in Fig. 2), there will be some cases where the eye is not situated at exactly the same distance from the camera, or the subject's head is tilted differently. These effects can lead to scale and rotation changes in the image data.

Noack and Sutton [12] develop a method for use with temporal images taken using a scanning laser ophthalmoscope. The images are first preprocessed to produce gradient images, using the Prewitt kernel with a threshold value that has been identified through prior empirical experiment. The resulting image is a binary image which allows for some optimization of the application of (4).

Since they are dealing with a sequence of frames from a video camera, they use the previous registration as the starting point for the next, moving outwards from this in a spiral: this also improves the speed of matching. The results indicate that the process is robust toward small rotations in the images, although it does not find those rotations. Unfortunately, the results reported are only those from images translated and rotated to produce artificial image pairs. As stated earlier, we have found that this situation produces quite spurious levels of success for any algorithm. Even so, Noack and Sutton's algorithm is accurate only to ± 2 pixels.

III. Registration of Retinal Images

The method of image registration presented in this paper uses mutual information as the comparison measure and simulated annealing with pyramid sampling as the search technique.

A. The Similarity Measure: Mutual Information

Detailed discussions of mutual information can be found in both Cover and Thomas [60] as well as Golomb *et al.* [61]. Consider a set $\mathbf{U}$ of disjoint events $u_1, u_2, \ldots, u_m$. The uncertainty function or entropy is then defined as

$$h(p_1, p_2, \ldots, p_m) = -\sum_{k=1}^{m} p_k \log p_k \tag{13}$$

where p_k is the probability of the kth event, u_k. Note that one can use logarithms of any base since this simply changes the units in which the entropy is measured. The use of base two gives units of bits, and the use of natural logarithms gives units of nats.

The mutual information Ψ between two sets $\mathbf{U}$ and $\mathbf{V}$ is then described as the amount of information that one of the sets gives us about the other. It is a measure of how well $\mathbf{U}$ is explained by $\mathbf{V}$ and is calculated with

$$\begin{aligned}\Psi(\mathbf{U};\mathbf{V}) &= h(\mathbf{U}) - h(\mathbf{U} \mid \mathbf{V}) \\ &= \sum_{u,v} p(u,v) \ln \frac{p(u,v)}{p(u)p(v)}\end{aligned} \tag{14}$$

where $p(u,v)$ is the probability of both u and v occurring.

To apply this mathematics to images we define $\mathbf{G}$ to be the set of available grey levels. Then the entropy of $\mathbf{G}$ for any image is given by

$$h(\mathbf{G}) = -\sum_{k=1}^{m} p_k \log p_k \tag{15}$$

where $p_k = p(g_k)$ is the probability of a particular grey-level event occurring in the image and m is the number of elements in the set $\mathbf{G}$, that is, the number of possible grey levels.

If we then have two images I and J with grey-level sets $\mathbf{G}_I$ and $\mathbf{G}_J$, respectively, then the mutual information between the two images is defined as

$$\begin{aligned}\Psi(I;J) &= \Psi(\mathbf{G}_I;\mathbf{G}_J) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} p(g_i,g_j) \ln \frac{p(g_i,g_j)}{p(g_i)p(g_j)}\end{aligned} \tag{16}$$

where m is the number of grey levels in image I, n is the number of grey levels in image J, $p(g_i)$ is the probability of a particular grey level occurring in image I, $p(g_j)$ is the probability of a grey level occurring in image J, and $p(g_i,g_j)$ is the probability of a particular pair of grey levels occurring at matching locations in the two images.

To calculate this mutual information the probabilities $p(g_i)$, $p(g_j,)$ and $p(g_i,g_j)$ must be estimated. Wells *et al.* [46] use Parzen windowing to estimate the three probabilities. However this method is computationally expensive. A simpler method is that described by Collignon *et al.* [14], [47]. They estimate the probabilities using the distribution of grey-level pixel pairs and pixel values across the common part of the two images. Thus

$$p(g_i,g_j) = \frac{H(g_i,g_j)}{N} \tag{17}$$

where H is the value of the joint histogram formed from the common part of the two images, and N is the number of points in the common part of the sample space. Similarly

$$p(g_i) = \frac{H(g_i)}{N} \tag{18}$$

and

$$p(g_j) = \frac{H(g_j)}{N}. \tag{19}$$

There are 256 possible grey scales in the images of the retina shown in Figs. 1 and 2, leading to histograms of a similar size and to a joint histogram of size 256 $\times$ 256. However, when images are digitized and again when they are translated from color to grey scale there is some error, which means that the actual grey scale of any particular pixel can only be guaranteed to within ± 1 grey scales. The images also do not contain any very bright or very dark pixels. These two facts lead to sparsely filled histogram matrices that contain data that are not absolutely accurate. Therefore, to save time and, without loss of accuracy, we rescale both the images before processing, using the following formula:

$$\forall (x,y) \in O, \qquad I(x,y) = \frac{O(x,y) - O_{\min}}{f} \tag{20}$$

where I is the new image, O is the original image, $O(x,y)$ refers to the grey-level intensity of the image O at position (x,y), $O_{\min}$ is the minimum grey level in that image, and f is a factor that allows binning of the grey scales.

To form a useful measure, the mutual information is negated to allow the use of minimization techniques. The final measure used is

$$C_{\mathrm{MI}} = -\sum_{i=1}^{m}\sum_{j=1}^{n}\begin{cases}0, & \text{if } H(i,j)=0\\ \frac{H(g_i,g_j)}{N}\ln\left(\frac{NH(g_i,g_j)}{H(g_i)H(g_j)}\right), & \text{otherwise}\end{cases} \tag{21}$$

where m is the number of grey levels in image I, which is the binned first image, n is the number of grey-levels in image J, which is the binned and transformed second image, $H(g_i, g_j)$, $H(g_i)$ and $H(g_j)$ are the histograms of the overlapping portions of the images, and N is the number of points sampled in this overlapping portion.

B. Search Strategy: Simulated Annealing with Pyramid Sampling

Assuming only rigid transformations, there are still four parameters (vertical and horizontal translation, rotation, and isotropic scaling) in the search space [1]. For retinal images of size 588 $\times$ 428, such as those shown in Fig. 2, the maximum horizontal translation is approximately 100 pixels, the maximum vertical translation approximately ten pixels and the maximum rotation and scaling around 5° and 5%, respectively. If subunit accuracy (where a unit is one pixel, 1°, or 1%) of 0.1 units is required, then the sample space contains in the order of 10^9 sample points. Therefore, to accelerate the search a pyramid of sampling is used, allowing the process to find increasingly accurate approximations to the minimum.

However, at less than full sampling the mutual information manifold has many local minima. Since the sample space is in four dimensions it is impossible to graph, however, Fig. 3 shows three slices or crosssections through this space for the images shown in Fig. 2(b) and (d). The three graphs show the value of the mutual information versus x and y translation for three different rotation-scale pairs with a sampling of one in eight pixels. The first graph shows fairly incorrect rotation and scale values, the second has slightly incorrect values, and the third correct values found from an exhaustive search.

A noticeable feature of the graphs is the trench formed to the left of the x axis. These trenches are caused by lighting effects and/or features that appear to give good matches across the image under a particular transformation and they are common in mutual-information manifolds. The cause of the trench in Fig. 3 can be seen by looking at the left hand side of the checkerboard image shown in Fig. 4(a). There is a vertical band of darker pixels in both images, causing a false match between the two images under this particular transformation. When the two images are clipped to remove these bands of pixels the trench reduces, as shown in Fig. 4(b).

The nature of the mutual-information function, multiple local minima including trenches, causes problems with many minimization techniques. Wells *et al.* [46] use gradient descent, whereas Collignon *et al.* [47] use Powell's minimization. However, both of these methods assume that the minimum approximates the quadratic form [51], whereas it can be seen that mutual information does not when sparse sampling is used. Collignon *et al.* [14] found that some images were registered at local minima instead of at the correct global minimum and Maes *et al.* [47], who also use Powell's minimization, found that for low resolutions the robustness of Powell's minimization was particularly dependent on the initial order of the parameters.

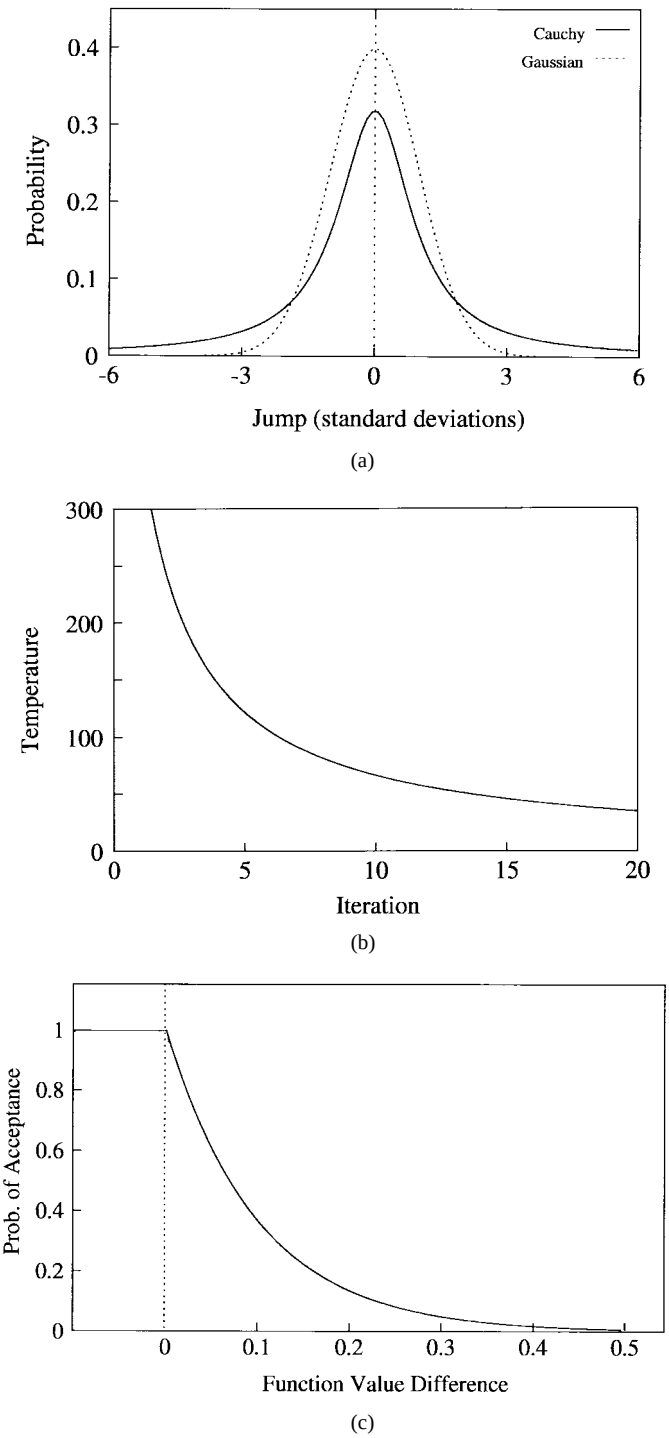


Fig. 5. The three functions required to set up simulated annealing minimization. (a) shows the probability function used to generate possible jumps from the current value to new test values. (b) shows the speed at which the temperature (range) of the generation function is dropped as the minimization continues and (c) shows the function used to decide if a new worse test value of the function should be accepted as the current value. (a) Cauchy generation function (see 22). (b) Annealing schedule (see 23). (c) Acceptance function (see 26).

It is possible to use an exhaustive search. However, this would be very costly if carried out globally and is subject to problems with local minima if performed separately and iteratively on the four separate dimensions.

A better solution is that of simulated annealing which is designed to deal with local minima and does not presuppose any particular shape for the function being searched.

1) Simulated Annealing: Descriptions of simulated annealing can be found in most books on combinatorial problems (see, for example, Dowsland [63]). The Otten and van Ginneken [64] is dedicated entirely to a discussion of simulated annealing and is an excellent source.

The simulated annealing method of finding the minimum of a function involves the use of three other functions: a generation function, an acceptance function, and an annealing schedule. Consider a function in x, $f(x)$ which we are trying to minimize. A particular value x_c is the current value in the minimization process. The next step for minimization using simulated annealing is to find a possible new (test) value x_n. This test value is calculated using the generation function, which provides a randomly generated jump from the current value to the test value. Generation functions are generally bell-shaped probability functions, chosen so that the probability is greater for a small jump than for a large jump. As the process continues the width of the generation function is slowly reduced using the annealing schedule. This is called lowering the temperature of the simulated annealing.[1]

If $f(x_n)$ is lower than $f(x_c)$, then the new value automatically becomes the current value for x, that is, $x_c = x_n$. If $f(x_n)$ is higher than $f(x_c)$ then, to allow escape from local minima, the test value will sometimes be accepted as the new current value. The acceptance or rejection of a test value that is worse is made using the acceptance function.

There are various possible choices of generation function, annealing schedule, and acceptance function, with the choice of the first two being linked so as to ensure convergence to a minimum [65]. The generation function that we have chosen is the Cauchy function

$$G_k(d) = \frac{t_k}{\pi\left(d^2 + t_k^2\right)} \tag{22}$$

where t_k is the current temperature and d the parameter jump being randomly generated. A graph of this function can be found in Fig. 5(a).

The annealing schedule that matches the Cauchy function to ensure convergence is

$$t_k = \frac{t_0}{k} \tag{23}$$

where t_0 is the initial temperature and t_k is the temperature for iteration k where $1 \leq k \leq N$ and N is the number of iterations [65]. A graph of this function can be found in Fig. 5(b). The combination of these two functions is called fast simulated annealing and may be improved upon by newer functions [66], [65], however, since a Cauchy random number is simply the dividend of a pair of Gaussian random numbers [67], it is relatively easy to calculate compared to other random number distributions.

The initial temperature controls the initial range of likely jump sizes. The Cauchy function is fairly close to a Gaussian [see Fig. 5(a)] with equation

$$\mathcal{G}_k(d) = \frac{1}{\sqrt{2\pi t_k}} e^{\frac{-d^2}{2t_k}}. \tag{24}$$

This is a Gaussian with $\mu = 0$ and $\sigma = \sqrt{t_k}$. To ensure that the initial required range is covered by the possible jump size we need the range to be approximately three standard deviations from the mean, so that

$$\begin{aligned} R_0 &\approx 3\sigma = 3\sqrt{t_0} \\ \therefore \quad t_0 &\approx \left(\frac{R_0}{3}\right)^2 \end{aligned} \tag{25}$$

where R_0 is the initial range and t_0 is the initial temperature. The use of this formula will ensure that most of the jumps generated using the temperature t_0 lie within the chosen range, as can be seen in the graph of Fig. 5(a).

The third function required by simulated annealing is the acceptance function. To allow escape from local minima it is necessary to occasionally make a jump from the current parameter value x_c to a test value x_n that is worse, that is, $f(x_n) > f(x_c)$. The probability of accepting such a test value needs to be dependent on the difference between the two values $f(x_n)$ and $f(x_c)$. A commonly used acceptance function [65] is

$$A_k = \begin{cases} 1, & A_k \geq 1 \\ \frac{e^{f(x_n) - f(x_c)}}{Ct_k}, & \text{otherwise} \end{cases} \tag{26}$$

where t_k is the temperature at iteration k and C is a system-dependent constant. A graph of such a function can be seen in Fig. 5(c). The system constant C is used to control this equation: the higher this value the less likely it is that the simulated annealing search engine will jump out of a current minimum. If a function is particularly noisy, with many local minima, then a low value of C works best. The actual decision on whether to accept the test value as the new parameter value is made by comparing a flat-distribution random number with the acceptance value.

To ensure that the simulated annealing finds a good minimum, multiple parameter values are tested for each temperature, giving an inner jump loop (metropolis loop) as well as an outer annealing loop. Thus the simulated annealing algorithm looks like this:

[1] The term temperature comes from the metallurgical process from which simulated annealing gets its name.

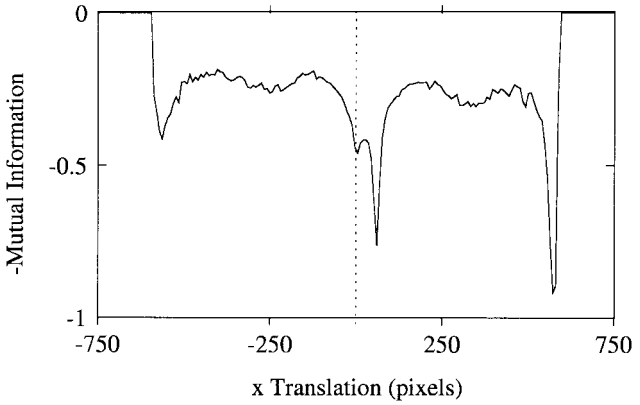


Fig. 6. A graph showing the mutual-information function for x translation when y translation, rotation, and scaling are fixed for the images shown in Fig. 2(b) and (d). Toward the edge of the function, when the images barely overlap, there is a minimum caused by the high level of mutual information between parts of images that have very little detail. These minima cause problems with the minimization of mutual information and it becomes necessary to ensure that the search engine does not become caught in them.

Algorithm 1:

```
Initialize the temperature
FOR annealing iterations
    FOR jump iterations (Metropolis loop)
        Generate a new test value
        IF the test value is better than the current value
            the test value replaces the current value
        ELSE
            Calculate the acceptance probability
            IF this probability is greater than a random num-
                ber the test value replaces the current value
            ENDIF
        ENDIF
    ENDFOR
    Reduce the temperature using the annealing schedule
ENDFOR
```

2) Algorithm for Multiple Parameters: In theory, fast simulated annealing will eventually find the minimum for the four rigid transformation parameters x and y translation, rotation, and isotropic scaling. However, the sample space in which it is searching is still very large and the time taken excessive. From the graphs of Fig. 3 it is clear that the minima for x and y when the other two parameters are fixed are close to the absolute minimum for those parameters. Although graphs are not shown for rotation and scaling, this fact also holds true for these parameters. This leads to an improvement in the algorithm by allowing searching over the four parameters separately.

Three further problems exist. First, since simulated annealing is a random search and since increases in mutual information are allowed to enable the escape from local minima, it is possible that the current position when the iterations end may not be the best position found throughout the search, thus this best position must be kept track of and is the final output of the algorithm. Second, it is possible for the simulated annealing to slowly creep a long way away from the absolute minimum by accepting many small increases in mutual information. This means that searching after this point is being carried out in the wrong place and is wasting processing time. To prevent this happening, a check is made at the end of each inner iteration on whether the current value has crept too far from the best value found so far, and the best value substituted if necessary.

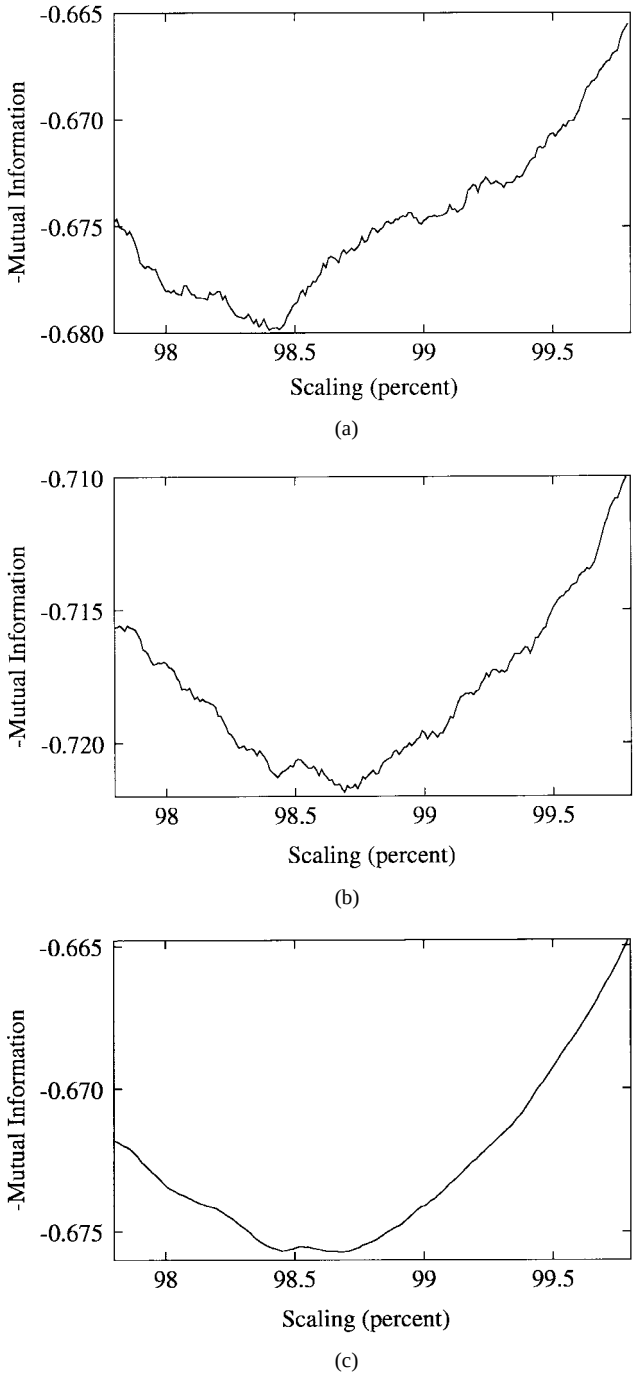


Fig. 7. Graphs of the mutual-information function for scaling when x translation, y translation, and rotation are fixed for the images shown in Fig. 2(c) and (d). While the partial distribution method of producing the grey scale histograms does smooth the function, it does not remove all local minima. This means that we require a minimization technique (such as simulated annealing) that is robust to local minima. (a) Nearest neighbor interpolation. (b) Bilinear interpolation. (c) Partial distribution interpolation: actually, an alternative method of producing the grey-level histogram used to estimate the probabilities required to calculate mutual information.

The third problem again relates to the nature of the mutual-information function. Fig. 6 shows the mutual-information function for the images in Fig. 2(b) and (d) for the x translation when the other three parameters are the best found with an exhaustive search. As can be seen from this graph, anomalies occur toward the edges when the images barely overlap. This is caused by fading at the edge of the images. When there is little detail the small overlap causes a very high level of mutual information between the images. Since the Cauchy generation function shown in Fig. 5(a) can produce very large jumps, it is possible that the simulated annealing search can become trapped if a large x translation is generated. Similarly, a very large scaling, where the scaled image is reduced by, say, 80%, would yield a high level of mutual information. This problem is solved by ensuring that all jumps generated are within reasonable bounds.

The algorithm is now

Algorithm 2:

```
Initialize the temperature
FOR annealing iterations
  FOR jump iterations
    FOR each parameter
      DO
        Generate a new test value
      WHILE the value generated is too large
      IF the test value is better than the current value
        the test value replaces the current value
      ELSE
        Calculate the acceptance probability
        IF this probability is greater than a random
          number the test value replaces the current
          value
        ENDIF
      ENDIF
    ENDFOR
    IF the current value is too poor compared to the best
      value
      Set the current value to the best value
    ELSE
      IF the current value is better than the best
        Set the best value to the current value
      ENDIF
    ENDIF
  ENDFOR
  Reduce the temperature using the annealing schedule
ENDFOR
```

3) Pyramid Sampling: Despite these improvements, simulated annealing cannot be guaranteed to find the absolute minimum of the comparison function based on mutual information. This is because the method is nondeterministic. This problem is exacerbated if infinite decimal places of accuracy are allowed for the four parameter values, as this would entail an infinite run of the algorithm. A further consideration is that not all the parameters have as wide an area to be searched. Generally, the x translation can be very large for temporal retinal image pairs [compare Fig. 2(a) and (c)] y translation in the region of ± 10 pixels and rotation and scaling only around five units each.

For these reasons, we have instituted a system of reannealing [66] incorporating pyramid sampling for speed and allowing the search over fewer parameters, if required. To ensure that the best possible set of parameter values has been found, the annealing is continued to a precision greater than that required and then the results rounded back up to give consistency.

Algorithm 3:

```
FOR decreasing initial temperatures and increasing
sampling rates
    Perform simulated annealing as in Algorithm 2
ENDFOR
Round the solution to the required precision
```

4) Interpolation: Since transformations will not always produce integral values, there is a need for interpolation. When pyramid sampling is used, the low-resolution interpolation can be calculated using nearest neighbor interpolation as this is by far the fastest. However when full or super-sampling is used, this type of interpolation may not be accurate enough. Alternatives are bilinear interpolation and partial-distribution (PD) interpolation [47].

PD interpolation is not a true interpolation, but rather a different method of creating the grey-level histograms required for the probability calculations of (17)–(19). It uses the same weights as those calculated for bilinear interpolation, but these weights are not used to calculate a new grey-level value. Instead, they are added to the histogram entries for the surrounding four pixels [47].

Maes *et al.* [47] found that PD interpolation smoothed the high-sampling mutual-information function sufficiently to allow the use of Powell's minimization as a search engine. Fig. 7 shows that while we also found that it smoothed the mutual-information function, it did not produce a near-quadratic function in all dimensions. The graphs of Fig. 7 show the tip of the mutual-information function for the scaling parameter when x translation, y translation and rotation are fixed. As can be seen, bilinear interpolation is smoother than nearest neighbor and PD interpolation smoother still, however there is still a local minimum, which necessitates the use of simulated annealing rather than other minimization techniques.

5) Summary: The final method involves a pyramid of resolutions combined with simulated reannealing. Until full sampling, the interpolation used is nearest neighbor. Thereafter, the choice of interpolation determines the precision to which the best solution can be found.

This method can be extended to give as accurate a registration as desired: the tradeoff being the time taken. A list of all the required parameters can be found in Table I.

IV. Results and Discussion

The images used in the research were taken with a Nidek stereo fundus camera on slide film, digitized with a Polaroid digitizer, and converted to 256 grey scale. They comprise around 5% of the total surface of the retina centered on

TABLE I
A LIST OF PARAMETERS THAT NEED TO BE PREDETERMINED FOR THE MUTUAL INFORMATION-SIMULATED ANNEALING ALGORITHM USED IN THIS PAPER

1	The number of times that each pair of images will be registered to enable comparison of results.
2	The binning value that reduces the number of grey levels in the images.
3	The maximum fractional creep allowed for the current mutual information compared to the best mutual information found.
4	The maximum bounds on the new values generated.
5	The final precision required.
6	The number of pyramid layers to be used.
7	For each layer:
	(a) The sampling frequency within the images.
	(b) The system constant, C, of (26).
	(c) The type of interpolation to use.
	(d) The parameters (x and y translation, rotation and isotropic scaling) over which the negative mutual information is to be minimised.
	(e) The initial ranges for these parameters.
	(f) The number of outer (annealing) loops.
	(g) The number of inner (jump or Metropolis) loops.

TABLE II
THE FINAL PARAMETERS SHOWN IN THIS TABLE—DESCRIBED IN TABLE I—WERE EXPERIMENTALLY DETERMINED FROM MULTIPLE TRAINING SESSIONS WITH THE IMAGES SHOWN IN FIG. 2. THE SEARCH AT LAYER 1 WAS PERFORMED ON ONLY THE x TRANSLATION. THE SEARCH AT LAYER 5 WAS REPEATED WITH EACH OF THE THREE INTERPOLATION METHODS

Runs done	5
Binning value	4
Fractional creep	0.9
Maximum bounds	$x = 200$, $y = 200$, $r = 20$, $s = 20$
Precision	0.1
Layers	5

Layer	Sampling	Interpolation	Initial Ranges				Loops		System
			x	y	r	s	Inner	Outer	Constant
1	8	nn	80	-	-	-	8	15	0.01
2	8	nn	8	8	8	8	8	15	0.001
3	4	nn	4	4	4	4	8	15	0.001
4	2	nn	2	2	2	2	8	15	0.001
5	1	nn	2	2	2	2	8	15	0.01
5	1	bi	2	2	2	2	8	15	0.01
5	1	pv	2	2	2	2	8	15	0.01

the optic nerve head: an area which is particularly rich in topographic information. The images are those used by Barry *et al.* in their research into glaucomatous changes to the optic-nerve head [54], [69].

The search for the most successful parameters for the simulated annealing was made using the four image pairs shown in Fig. 2: two stereo pairs and two temporal pairs. The actual values used of the parameters described in Table I, are listed in Table II.

The registration was then trialed on 48 pairs of temporal images and 49 pairs of stereo images taken with the same camera. The temporal-image pairs were taken up to three years apart. For each run of the program the final loop was performed for all three types of interpolation, nearest neighbor, bilinear, and partial distribution, to allow comparison of speed. Consistency was tested by doing five runs on each image pair. To enable comparison of these results with a ground truth value, an exhaustive search was performed in the neighborhood of the results obtained through simulated annealing. Since the partial distribution method of creating the histograms produces the smoothest mutual-information function, this was the function that was exhaustively searched.

TABLE III
THE RMSE FOR THE DIFFERENT INTERPOLATION TECHNIQUES, WHERE GROUND TRUTH IS TAKEN TO BE THE RESULT OF THE EXHAUSTIVE SEARCH. THE TIME TAKEN IS THE AVERAGE FOR IMAGE PAIRS SUCH AS THOSE SHOWN IN FIG. 2 WHEN PROCESSED ON A PENTIUM PRO 200

	Stereo image pairs				Temporal image pairs				Time
Layer	RMSE for 5 × 49 runs				RMSE for 5 × 48 runs				Taken
	x	y	r	s	x	y	r	s	(mins)
1	3.44				6.20				0:03
2	2.07	1.81	0.93	2.99	2.67	2.44	1.02	5.71	0:10
3	1.79	1.30	1.15	1.78	1.94	2.54	0.80	3.80	0:38
4	1.22	0.94	0.51	1.02	0.61	0.81	0.28	0.90	2:39
5(nn)	0.76	0.60	0.39	0.67	0.34	0.37	0.17	0.47	9:45
5(bi)	0.59	0.52	0.31	0.58	0.29	0.29	0.15	0.38	17:10
5(pd)	0.35	0.26	0.15	0.32	0.15	0.11	0.08	0.20	18.20

The exhaustive search results were used to calculate the root mean square error (RMSE) values for the 5×97 image runs for the seven layers described in Table II for each type of image pair: temporal and stereo. These results can be seen in Table III and graphs of the distribution of the errors for x translation for temporal image pairs can be found in Fig. 8.

From Table III it can be seen that the halving of RMSE comes at the cost of double the processing time. For temporal images, as can be seen in the graphs shown in Fig. 8, nearest

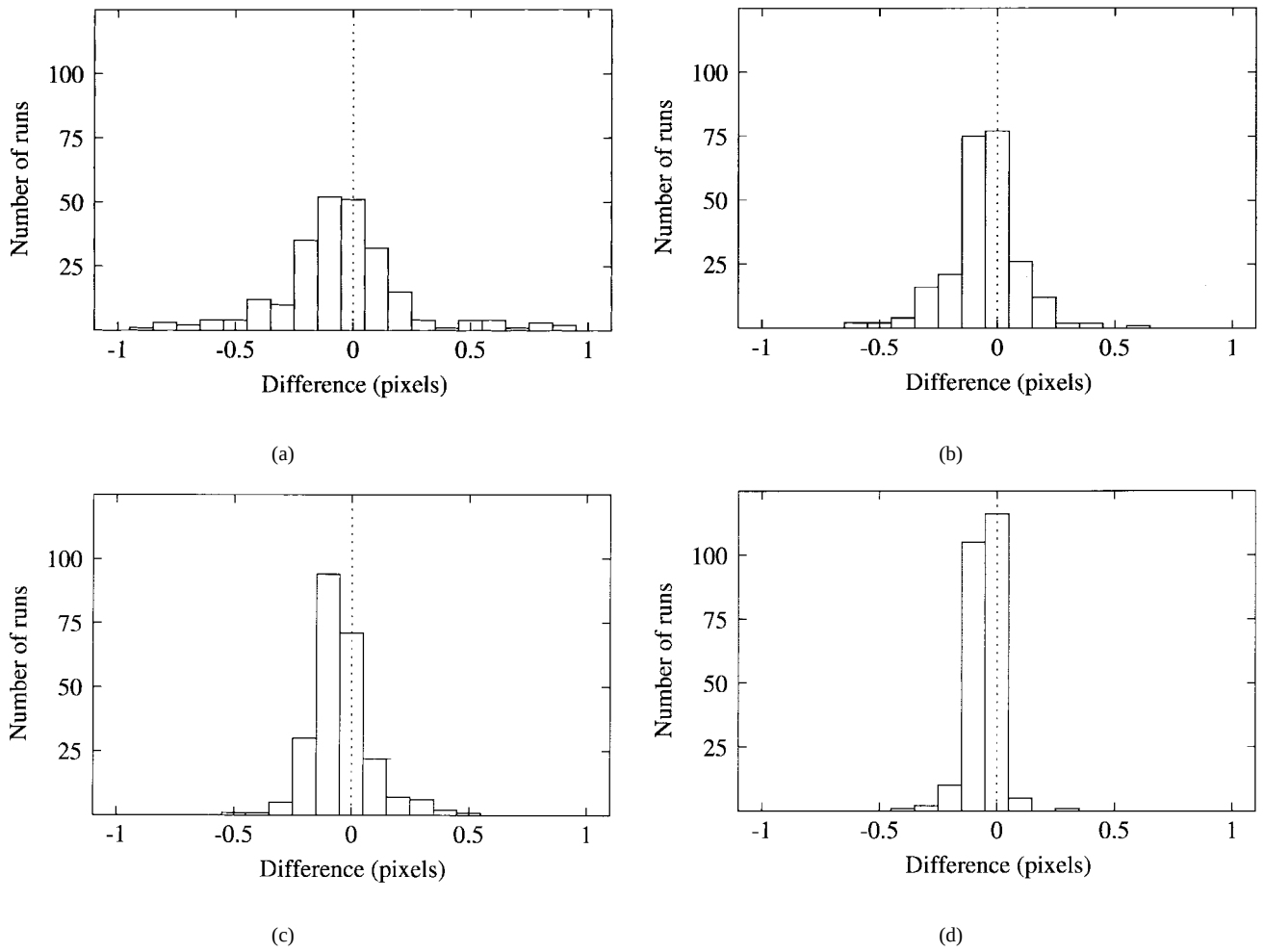


Fig. 8. The distributions of the differences between the four parameter values found using the searches and those found with an exhaustive search for x translation for temporal image pairs. The graphs for stereo image pairs and for the other four parameters showed similar results. (a) The fourth layer in the pyramid used nearest neighbor interpolation and only sampled every second pixel. The cumulative time for this was 3:30 min, which is significantly faster than that taken for partial distribution with full sampling. Although the results for layer four are less consistent than those for partial distribution, the differences from ground truth are all still less than one pixel. (b) The fifth layer of the pyramid performed using nearest neighbor interpolation. (c) The fifth layer of the pyramid performed using bilinear interpolation. (d) The fifth layer of the pyramid performed using partial distribution to generate the grey-level histograms.

neighbor interpolation when sampling only every second pixel produces unit-level accuracy (where a unit is one pixel, 1° of rotation and 1 % of scaling) in an accumulated time of only 3:30 min, as compared to partial distribution which takes 21:50 min. For stereo images the fourth layer is less accurate, with the maximum error being approximately 3.8 pixels.

Checkerboard images of the registration were produced to allow visual assessment of the registration found. The checkerboard comparison of the images from Fig. 2(b) and (d) can be seen in Fig. 9. Note that the dislocations in the center of the image are due to changes in the optic nerve over the two year gap, during which time the patient had an operation to relieve pressure related to glaucoma.

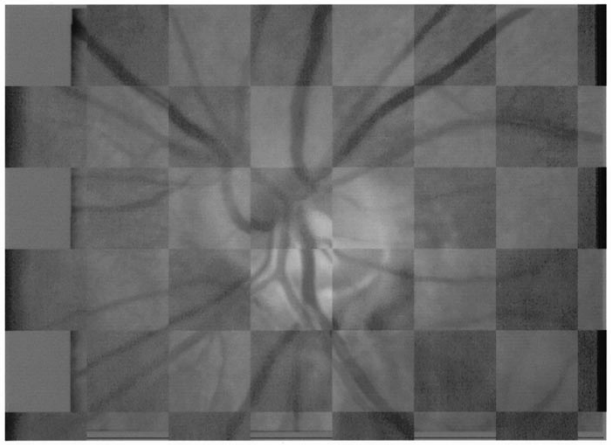


Fig. 9. A checkerboard image of the correct registration for the images shown in Fig. 2(b) and (d). The registration was calculated using the simulated annealing parameters given in Table II and confirmed with an exhaustive search.

V. Conclusion

As previously ascertained by other researchers [14], [26], [45]–[48], mutual information proves to be a successful measure of goodness of fit for registration of images. It is not affected by large changes in light intensity, as can occur between stereo and temporal images of the retina. It requires no preprocessing of the images and no assumptions as to what the images contain. This makes it a robust image registration tool.

We have used a pyramid-sampling based reannealing search technique to improve the speed and precision of the registration. Within the limits imposed by the resolution and noise levels of the input data, this algorithm allows the required precision of the process to be predetermined and is affected only by the considerations of interpolation technique and available time. While partial distribution is slightly slower than bilinear interpolation, it does produce the most consistent and precise results. However, it takes almost twice the time of nearest neighbor interpolation which produced results that were within ± 1 unit of those found with an exhaustive search for 99% of the registration runs.

A comparison of our simulated annealing with that of Mandava *et al.* [50] would be of interest.

ACKNOWLEDGMENT

The authors would like to thank C. Barry and the McCusker Glaucoma Foundation for making available the images of the retina, and the reviewers for their many useful suggestions.

REFERENCES

[1] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surveys*, vol. 24, no. 4, pp. 325–375, Dec. 1992.
[2] C. A. Pelizzari, G. T. Y. Chen, D. R. Spelbring, R. R. Weichselbaum, and C.-T. Chen, "Accurate three-dimensional registration of CT, PET and/or MR images of the brain," *J. Comput. Assist. Tomogr.*, vol. 13, no. 1, pp. 20–26, Jan. 1989.
[3] H. Jiang, R. A. Robb, and K. S. Holton, "A new approach to 3-D registration of multimodality medical images by surface matching," *SPIE Visualization Biomed. Comput.*, vol. 1808, pp. 196–213, 1992.
[4] A. Collignon, D. Vandermeulen, P. Suetens, and G. Maral, "3D multi-modality medical image registration using feature space clustering," in *Computer Vision, Virtual Reality and Robotics in Medicine*, N. Ayache, Ed. Berlin, Germany: Springer-Verlag, 1995, pp. 195–204.
[5] H. R. Novotny and D. L. Alvis, "A method of photographing fluorescence in circulating blood in the human retina," *Circulation*, vol. 24, pp. 82–86, 1961.
[6] R. N. Wienreb, "Assessment of the optic nerve and nerve fiber layer in glaucoma," *J. Glaucoma*, vol. 2, pp. 135–137, 1993.
[7] W. H. Morgan, R. L. Cooper, I. J. Constable, and R. H. Eikelboom, "Automated extraction and quantification of macular drusen from fundal photographs," *Aust. New Zealand J. Ophthalmol.*, vol. 22, pp. 7–12, 1994.
[8] R. H. Eikelboom, R. L. Cooper, and C. J. Barry, "Performance of two films for densitometry of retinal photographs," *Graefes Archive Clinic. Experiment. Ophthalmol.*, vol. 231, pp. 514–520, 1993.
[9] E. Chihara and K. Chihara, "Covariance of optic disc measurements and ocular parameters in the healthy eye," *Graefes Archive Clinic. Experiment. Ophthalmol.*, vol. 232, pp. 265–271, 1994.
[10] M. S. Mort and M. D. Srinath, "Maximum likelihood image registration with subpixel accuracy," in *Proc. SPIE: Applications Digital Image Processing XI*, 1988, vol. 974, pp. 38–43.
[11] N. D. A. Mascarenhas and G. J. Erthal, "Image registration by sequential tests of hypotheses: The relationship between Gaussian and binomial models," *Comput. Graph.*, vol. 16, no. 3, pp. 259–264, 1992.
[12] J. Noack and D. Sutton, "An algorithm for the fast registration of image sequences obtained with a scanning laser opthalmoscope," *Phys. Med. Biol.*, vol. 39, pp. 907–915, 1994.
[13] A. M. Mendonça, A. Campilho, and J. M. R. Nunes, "A new similarity criterion for retinal image registration," in *Proc. IEEE 1994 Int. Conf. Image Processing ICIP-94*, Los Alamitos, CA, Nov. 1994, pp. 696–700.
[14] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Maral, "Automated multi-modality image registration based on information theory," in *Inform. Processing Med. Imag.*, Y. Bizais *et al.* Eds. Amsterdam, The Netherlands: Kluwer, 1995, pp. 263–274.
[15] R. Szeliski and S. Lavallée, "Matching 3-D anatomical surfaces with nonrigid deformations using octree-splines," *Int. J. Comput. Vision*, vol. 18, no. 2, pp. 171–186, 1996.
[16] J. Flusser, "An adaptive method for image registration," *Pattern Recognit.*, vol. 25, no. 1, pp. 45–54, 1992.
[17] J. Feldmar and N. Ayache, "Rigid, affine and locally affine registration of free-form surfaces," *Int. J. Comput. Vision*, vol. 18, no. 2, pp. 99–119, 1996.
[18] D. J. Burr, "A dynamic model for image registration," *Comput. Graph. Image Processing*, vol. 15, pp. 102–112, 1981.
[19] R. Bajcsy and S. Kovačič, "Multiresolution elastic matching," *Comput. Vision Graph. Image Processing*, 1989, vol. 46, pp. 1–21.
[20] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, "3D brain mapping using a deformable neuroanatomy," *Phys. Med. Biol.*, vol. 39, pp. 609–618, 1994.
[21] R. Sivaramakrishna and R. Gordon, "Image registration using minimization," in *Proc. IEEE Wescanex'95 Communications, Power, Computing Conf.*, 1995, pp. 181–184.
[22] F. L. Bookstein, "Biometrics, biomathematics and the morphometric synthesis," *Bull. Math. Biol.*, vol. 58, no. 2, pp. 313–365, 1996.
[23] Y. Amit and A. Kong, "Graphical templates for model registration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 225–236, Mar. 1996.
[24] C. Davatzikos, J. L. Prince, and R. N. Bryan, "Image registration based on boundary mapping," *IEEE Trans. Med. Imag.*, vol. 15, pp. 112–115, Feb. 1996.
[25] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Med. Image Anal.*, vol. 2, no. 1, pp. 1–36, 1998.
[26] J. West, J. M. Fitzpatrick, M. Y. Yang, B. M. Dawant, C. R. Maurer Jr., R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. A. van den Elsen, S. Napel, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. G. Hill, D. J. Hawkes, C. Studholme, J. B. A. Maintz, M. A. Viergever, G. Malandain, X. Pennec, M. E. Noz, G. Q. Maguire Jr., M. Pollack, C. A. Pelizzari, R. A. Robb, D. Hanson, and R. P. Woods, "Comparison and evaluation of retrospective intermodality brain image registration techniques," *J. Comput. Assist. Tomogr.*, vol. 21, no. 4, pp. 554–566, July 1997.
[27] J. M. Henderson, K. R. Smith, and R. D. Bucholz, "An accurate and ergonomic method of registration for image-guided neurosurgery," *Computer. Med. Imag. Graph.*, vol. 18, no. 4, pp. 273–277, 1994.
[28] J. B. A. Maintz, P. van den Elsen, and M. A. Viergever, "Evaluation of ridge seeking operators for multimodality medical image matching," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 353–365, Apr. 1996.
[29] C.-M. Lau, T. Adah, and Y. Wang, "Coregistration of PET/MR brain images by multi-feature correlation matching," in *Proc. Fifteenth Southern Biomedical Engineering Conf.*, Mar. 1996.
[30] T. Wakahara, "An iterative image registration technique using local affine transformation," *Syst. Comput. Japan*, vol. 21, no. 12, pp. 78–89, 1990.
[31] J. P. P Starink and E. Backer, "Finding point correspondences using simulated annealing," *PR*, vol. 28, no. 2, pp. 231–240, 1995.
[32] Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," *Image Vision Comput.*, vol. 10, no. 3, pp. 145–155, Apr. 1992.
[33] R. B. Backman, C. Pudney, N. Spadaccini, F. Wood, P. Hartmann, and R. Owens, "Registration of anatomic landmarks during respiration using ultraviolet and structured lighting," in *Proc. IEEE Frontiers Biomedical Visualization—'95*, 1995, pp. 42–49.
[34] G. Q. Maguire, M. E. Noz, E. M. Lee, and J. H. Schimpf, "Correlation methods for tomographic images using two and three dimensional techniques," in *Proc. 9th Conf. Information Processing Medical Imaging*, June 1985, pp. 266–279.
[35] A. V. Cideciyan, S. G. Jacobson, C. M. Kemp, R. W. Knighton, and J. H. Nagel, "Registration of high resolution images of the retina," in *Proc. SPIE: Medical Imaging VI: Image Processing*, vol. 1652, pp. 310–322, 1992.
[36] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "Multiresolution voxel similarity measures for MR-PET registration," in *Information Processing in Medical Imaging*, Y. Bizais *et al.*, Eds. Amsterdam, The Netherlands: Kluwer, pp. 287–298, 1995.
[37] J. J.-H. Yu, B.-N. Hung, and C.-L. Liou, "Fast algorithm for digital retinal image alignment," in *Proc. IEEE Ann. Int. Conf. Engineering Medicine Biology Society, Images Twenty-First Century*, Nov. 1989, vol. 2, pp. 374–375.
[38] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Computer Vision—ECCV'94*, (Lecture Notes Computer Science) J.-O. Eklundh, Ed. Berlin, Germany: Springer-Verlag, 1994, vol. 801, pp. 151–158.
[39] D. N. Bhat and S. K. Nayar, "Ordinal measures for visual correspondence," in *Proc. 1996 IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Los Alamitos, IEEE Computer Society Press, CA, June 1996, pp. 351–357.

[40] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *Proc. IEEE 1975 Int. Conf. Cybernetics Society*, Sept. 1975, pp. 163–165.
[41] D. J. Lee, T. F. Krile, and S. Mitra, "Digital registration techniques for sequential fundus images," *Proc. IEEE Applications of Digital Image Processing X*, 1987, vol. 829, pp. 293–300.
[42] S. P. Kim and W.-Y. Su, "Subpixel accuracy image registration by spectrum cancellation," in *Proc. IEEE 1993 Int. Conf. Acoustics, Speech Signal Processing, ICASSP-93*, Apr. 1993, vol. 5, pp. 153–156.
[43] 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, Aug. 1996.
[44] H. Shekarforoush, M. Berthod, and J. Zerubia, "Subpixel image registration by estimating the polyphase decomposition of cross power spectrum," in *Proc. 1996 IEEE Computer Society Conf. Computer Vision Pattern Recognition*, Los Alamitos, CA, June 1996, pp. 532–537.
[45] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," in *Proc. 5th Int. Conf. Computer Vision*, Los Alamitos, CA, 1995, pp. 16–23.
[46] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multimodal volume registration by maximization of mutual information," *Med. Image Anal.*, vol. 1, no. 1, pp. 35–51, Mar. 1996.
[47] F. Maes, A. Collignon, D. Vandermeulen, G. Maral, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, pp. 187–198, Apr. 1997.
[48] N. Ritter, R. Owens, K. Yogesan, and P. P. van Saarloos, "The application of mutual information to the registration of stereo and temporal images of the retina," in *Proc. Advanced Topics Artificial Intelligence: 10th Australian Joint Conf. Artificial Intelligence, AI'97*, Dec. 1997, pp. 67–76, .
[49] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures," *Med. Phys.*, vol. 24, no. 1, pp. 25–36, Jan. 1997.
[50] V. R. Mandava, J. M. Fitzpatrick, and D. R. Pickens III, "Adaptive search space scaling in digital image registration," *IEEE Trans. Med. Imag.*, vol. 8, pp. 251–262, Sept. 1989.
[51] W. H. Press, S. A. Teukolsky, W. T. Vettering, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. New York: Cambridge Univ. Press, 1992.
[52] V. I. Samoilenko and A. V. Yakovlev, "Use of correlation procedures for image registration with local bright pixels," *Radioelectron. Commun. Syst.*, vol. 36, no. 11, pp. 15–18, 1993.
[53] M. H. Goldbaum, S. Chatterjee, S. Chaudhuri, and N. Katz, "Digital image processing for opthalmology," in *Noninvasive Diagnostic Techniques in Opthalmology*, B. R. Masters, Ed. Berlin, Germany: Springer-Verlag, 1990, ch. 29, pp. 548–568.
[54] C. J. Barry, P. House, M. Cuypers, and R. H. Eikelboom, "Flicker stereo chronoscopy in glaucoma: A preliminary report Part 2: Method, results and discussion," *J. Biomed. Imag.*, vol. 3, no. 1, pp. 4–9, 1997.
[55] V. A. Miszalok, "Fundus imaging and diagnostic screening for public health," in *Noninvasive Diagnostic Techniques in Opthalmology*, B. R. Masters, Ed. Berlin, Germany: Springer-Verlag, 1990, ch. 26, pp. 510–515.
[56] E. Peli, R. A. Augliere, and G. T. Timberlake, "Feature-based registration of retinal images," *IEEE Trans. Med. Imag.*, vol. 6, pp. 272–278, Sept. 1987.
[57] P. Jasiobedzki, "Registration of retinal images using adaptive adjacency graphs," in *Proc. Sixth Annual IEEE Symposium Computer-Based Medical Systems*, June 1993, pp. 40–45.
[58] W. E. Hart and M. H. Goldbaum, "Registering retinal images using automatically selected control point pairs," in *Proc. 1994 IEEE Int. Conf. Image Processing ICIP-94*, Nov. 1994, pp. 576–580.
[59] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Trans. Med. Imag.*, vol. 8, pp. 263–269, Sept. 1989.
[60] T. M. Cover and J. A. Thomas, *Elements in Information Theory*. New York: Wiley, 1991.
[61] S. W. Golomb, R. E. Peile, and R. A. Scholtz, *Basic Concepts in Information Theory and Coding: The Adventures of Secret Agent 00111*. New York: Plenum, 1994.
[62] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
[63] K. A. Dowsland, "Simulated annealing," in *Modern Heuristic Techniques for Combinatorial Problems*, C. R. Reeves, Ed. New York: Wiley, 1993, ch. 2, pp. 20–69.
[64] R. H. J. M. Otten and L. P. P. P. van Ginneken, *The Annealing Algorithm*. Boston, MA: Kluwer, 1989.
[65] X. Yao, "A new simulated annealing algorithm," *Int. J. Comput. Math.*, vol. 56, pp. 161–168, 1995.
[66] L. Ingber, "Very fast simulated re-annealing," *Math. Comput. Modeling*, vol. 12, no. 8, pp. 967–973, 1989.
[67] P. McCullagh, "Möbius transformation and Cauchy parameter estimation," *Ann. Stat.* , vol. 24, no. 2, pp. 787–808, Apr. 1996.
[68] K. Mizuno, A. Majima, K. Ozawa, and H. Ito, "Fundus photography in red-free light (Rhodopsin photography)," *Vision Research*, vol. 8, pp. 281–282, 1968.
[69] K. Yogesan, C. J. Barry, L. Jitzkaia, R. H. Eikelboom, P. P. van Saarloos, P. H. House, and W. H. Morgan, "Software for 3-D visualization/analysis of optic-disc images," *IEEE Trans. Eng. Biol. Med.*, vol. 18, pp. 43–49, Jan. 1999.