Registration of high resolution images of the retina

Artur V. Cideciyan[1,2], Samuel G. Jacobson[1], Colin M. Kemp[1], Robert W. Knighton[1], and Joachim H. Nagel[2]

[1]Department of Ophthalmology, Bascom Palmer Eye Institute
University of Miami School of Medicine
PO Box 016880, Miami, FL 33101

[2]Department of Biomedical Engineering, University of Miami
PO Box 248294, Coral Gables, FL 33124

ABSTRACT

A method of image registration is presented for the case when the deformation between two images can be well approximated with a combination of translation, rotation and global scaling. The method achieves very high accuracy by combining a global optimization in the 4-dimensional discrete parameter space with a local optimization in the 4-dimensional continuous parameter space. The 4-dimensional global optimization is accomplished with two 2-dimensional optimizations. The Fourier magnitude is used to decouple translation from rotation and scaling, and a log-polar mapping of the Fourier magnitude is used to convert rotation and scaling into shifts. Optimal rotation and scaling parameters are determined with a cross-correlation in the log-polar domain. After compensation for rotation and scaling differences, cross-correlation in the spatial domain yields the translation parameters. The four registration parameters are further refined with a local optimization using the correlation coefficient as a similarity measure in the 4-dimensional continuous parameter space. Results are shown from simulations and from registration of retinal images. For simulated images with a signal-to-noise ratio of -5 dB, the accuracy of the registration method is estimated to be better than 0.07 degrees, 0.1%, and 0.3 pixels for rotation, scaling, and translation, respectively. In the case of 512x512 pixel images the computation resource requirements are compatible with high end PCs, i.e., approximately 25 minutes on an Intel 80486/33MHz based IBM/PC compatible.

1. INTRODUCTION

Fundus photography has been used for decades as a non-invasive technique to diagnose and document a number of eye diseases and their progression over time. The resulting color fundus photographs are usually analyzed by expert ophthalmologists, but computers are being used more and more to aid the experts, especially in quantitation. Some examples of the use of digital image processing methods include: the enhancement of the retinal nerve fiber layer[1]; deblurring of images taken through cataractous lenses[2]; segmentation of hemorrhages[3], exudates[4,5], drusen[6], and blood vessels[7]; and, quantitative analysis of the optic nerve head cup[8,9,10] and fluorescein angiograms[11].

We are currently using digital image processing methods to characterize a unique retinal reflection that occurs in the incurable hereditary disease of the retina known as retinitis pigmentosa[12,13,14]. This disease has many genetic types; one of the more severe types is transmitted on the X-chromosome, X-linked retinitis pigmentosa (XLRP). Some carriers of XLRP show a golden scintillating particulate retinal reflection called a tapetal-like reflex[15] (Fig. 1). Although the relation between the tapetal-like reflex and XLRP has been known for over forty years[16], a systematic investigation of the reflex has not been published. Histological studies of post-mortem donor eye tissue have not been helpful to date since the only report of the histopathology of an XLRP carrier was in a woman without a tapetal-like reflex[17].

Figure 1: *Fundus photograph of an XLRP carrier showing the tapetal-like reflex which appears as a bright particulate reflection outside of the fovea (center of image).*

One of the questions we are asking about the tapetal-like reflex is: What happens to the reflex over time? It has been suggested that the reflex "fades" over a "seven- to eight-year period"[18], but there have been no quantitative analyses of change in the appearance of the reflex with

0-8194-0804-2/92/$4.00

time. We have serial fundus photographs of some XLRP carriers from visits separated by years (decades, in one case) and quantitative analysis of these photographs is in progress. In this paper, we describe our approach to the image registration problem, the solution of which is a prerequisite to accurate change detection. The method is broadly applicable to the registration of images produced by other imaging modalities like X-ray CT, PET, SPECT, and MRI.

The following four sections describe our methods, results, and conclusions. In section 2, we consider the need for registration, describe briefly the general approach to the problem, and present in detail our registration method. In section 3, we cover some implementation issues. In section 4, we show results using simulations and actual images, and we present our conclusions in section 5. In the Appendix, we give the derivation of an important equivalence that frees us from having to assume a specific order in the original deformations.

2. REGISTRATION

A high resolution color slide scanner (LS3500, Nikon) is used to digitize color photographic transparencies (slides) of the retina recorded with a fundus camera (FF4, Zeiss) [12,13,14]. This indirect approach makes the image data collected in photographic archives available for digital analysis. Each pixel in a digitized image corresponds approximately to a 2.4 μm by 2.4 μm area measured on the retina; i.e., a typical 30° field is contained in a 4000 by 4000 pixel area. When the same area of the fundus is digitized from two different slides of the same patient (and the same eye) chances are very high that they will not exactly correspond to each other. In general, the differences might be due to the following: a) changes in the underlying anatomy; b) changes in the characteristics of the chemicals of the photographic image; c) changes in the distribution and/or the intensity of the illumination; and d) changes in the geometric relationship between the imaging system and the fundus. Our goal is to determine the changes due to underlying anatomy, therefore we must compensate for the other three effects. Using our model of the complete imaging system[12], we can compensate for the possibility of different films. The slowly-varying illumination component can usually be removed with simple filtering. To deal with the effect of the location and orientation of the fundus camera with respect to the eye being photographed, we need a registration method that can calculate, effectively and accurately, the deformation parameters.

Before determining the extent of deformations, we must first know what *type* of deformations to expect. Small torsional rotations of the eye (<±10°) are possible depending on the location of the fixation point; therefore we have to assume that two images might differ in relative rotation. Global scale (magnification) changes can result from the eye being at different distances from the camera and from changes in the refraction of the patient's eye between visits separated by long intervals. Of course, translation (shift) of the image is always present as it is not practical to use exactly the same photographic geometry from visit to visit. We make the assumptions that nonuniform scaling and parallax are negligible. The first assumption is based on the fact that the high-resolution nature of our data combined with limited computational resources force us to consider relatively small regions of interest (512x512 pixels = 1.2x1.2 mm. on the retina). Within a given small region of interest, the magnification of the fundus camera can be considered to be uniform to a good approximation, although it is not perfectly uniform across the whole field of view[19]. The second assumption (no parallax) is based on the fact that regions of interest are specifically selected void of major retinal blood vessels which appear to be at a different retinal depth than the tapetal-like reflex and the retinal background; the latter two appear to be approximately at the same depth. We assume therefore that translation, rotation and global scaling are enough to explain all the geometric mismatches of interest to us. Furthermore, we use the center of each image as the center of rotation and scaling. As we show in the Appendix, any number of translation, rotation and scaling transformations applied to an image in any sequence, can be equivalently represented by any other number of translation, rotation and scaling transformations, in any other sequence, as long as the parameters are correctly modified.

Registration can be formally defined as the transformation of one image with respect to another so that the properties of any resolution element of the object being imaged is addressable by the same coordinate pair in either one of the images[20]. Accurate registration is a prerequisite to many image analysis applications, and therefore it has received much attention. Most registration approaches fall into one of two categories, local or global. Local methods determine deformations locally at many points within the images, whereas global methods try to find a single deformation imposed upon the whole image.

2.1 Local methods

Local methods are based on the correspondence between a small subset of image pixels, that are usually called landmark or control points. The power of the local methods is based on the fact that the deformation of each landmark point is determined

independently from other landmark points. Of course, neither the determination of landmark points, nor the determination of individual deformations are easy. One of the simplest solutions to both problems is the interactive approach, in which a human expert chooses corresponding landmark points in both images[21,22]. In most registration applications, the subjectivity of a human expert (as well as time and economic considerations) is objectionable and a fully automated method is preferable.

Various methods have been published that automate both the landmark-choosing and correspondence determination procedures. For example, Anuta[20] describes a method for the registration of remote sensing images. In this method a number of points are chosen on a coarse grid on one image to be used as landmark points (without attention to their local properties) and for each chosen point the corresponding point on the other image is found by using the cross correlation function evaluated in a small neighborhood. In later work, potential landmark points have been defined more carefully as having distinctive features[23,24], and these points are found automatically in the images to be registered. The correspondences of the landmark points can be determined by maximizing a similarity criterion based on deformation invariant properties, e.g. correlation of the invariant moments[25,26].

Landmark based registration methods are commonly used for so-called "rubber-sheet" transformations[27], although they have also been used for the determination of a single global registration for the whole image[21,24]. In our application, the objects of interest (patches of golden reflections) are relatively large (5 to 12 pixels wide) with no distinct boundaries, and the images are noisy, therefore making accurate automated (or manual) landmark-choosing very difficult. Therefore, we have decided to use a global registration method.

2.2 Global methods

In general, the aim of global methods is to find a single m-parameter transformation that optimally registers all (or most) of the pixels of the two images. A similarity criterion is defined that depends on the two images of interest and the m-parameter transformation. The solution to the registration problem is the point in the m-dimensional parameter space that achieves the highest value of the similarity criterion; in other words, registration is an optimization problem. Cross-correlation[28] has been the most commonly used similarity measure because of its proven discriminatory abilities in the presence of white noise, but other measures like phase-correlation[29], sum of absolute differences and sequential similarity detection[30] have also been used.

The process of finding the highest similarity is not always easy. If the parameter space is discretized and can be constrained in both resolution and number of dimensions, then exhaustive search of this space is possible. A good example is the simple translation-only problem, with its 2-dimensional parameter space. The classic approach is to calculate the correlation coefficient for each point in the parameter space (i.e. the cross-correlation function) and determine the two translation parameters for which the correlation coefficient is maximum[28,31]. Higher dimensional parameter spaces can also be approached with an exhaustive search strategy, with obvious penalties in speed. For example, DeCastro and Morandi[29] report implementing an image stabilization system for the observation of the human retina. They assume that the images differ only in translation and rotation (3-dimensional parameter space), and they use phase-correlation as their similarity criterion. The authors determine the phase-correlation function for each discrete rotation value and choose the parameter set that results in the highest phase-correlation.

In the case of even higher-dimensional parameter spaces, an exhaustive search is usually not practical with current-day computational resources. Therefore, so-called "coarse-to-fine" (hierarchical) approaches have been used extensively. An example is given in Herbin et al.[32] where the authors register photographs of dermatological lesions. Their parameter space is 6-dimensional (translation, rotation, scaling, and intensity with 2 parameters); therefore they use either an adaptive random search or a hierarchical search. Adaptive random search can be shown to reach a global optimum, but, as the authors also mention, it is computationally very intensive. In the case of the hierarchical search, the authors first sample the search space coarsely, find a coarse set of parameters, and then they do a full-resolution local search in the neighborhood of the coarse parameter set.

In summary, we can say that the exhaustive search of the parameter space (i.e. trial of each and every possible parameter combination) might be the best method of global registration for a given similarity criterion, but it is usually computationally not feasible. A random search can also reach a globally optimal solution but it usually is also computationally prohibitive, although less so as compared to an exhaustive search. A hierarchical search is usually favorable from a computational point of view, but unfavorable in terms of the assurance of a global optimum.

Next, we present a two stage method for the case of a translation, rotation and global scaling based registration; at the first stage, two 2-dimensional exhaustive searches (instead of a single 4-dimensional exhaustive search) are used to find a globally

optimal registration in a discrete parameter space. In the second stage, a local search in the immediate neighborhood of the globally optimal solution is used to determine the optimal solution in a continuous parameter space, therefore achieving very high resolution registration with a reasonable computational effort.

2.3 Log-polar mapping of the Fourier magnitude

Recently, Apicella et al.[33] described a method called "fast correlation matching" and applied it to the registration of multi-modality images of PET and MRI. The underlying idea is the transformation of the images into a domain in which translation effects disappear, and rotation and global scaling become simple shifts in two independent coordinate directions. Once the rotation and scaling are determined (with a simple cross-correlation), translation can be calculated readily from the appropriately rotated and scaled original images. The theory was originally developed[34] to compensate for the limited pattern recognition abilities of optical correlators[35,36].

The transformation necessary to map the rotation and scaling into independent shifts is a log-polar mapping of the Fourier magnitude[33,37,38,39]. Figure 2 illustrates this method in the simple case of a sinusoid function and its translated, rotated, and scaled version (Panel A). Note that all vertical axes in Figure 2 are inverted so that they correspond to the common use of a digital image with an upper-left origin in computer programs. In panel B, the Fourier magnitudes of the input images are shown (a pair of impulses). As expected, the Fourier magnitude removes all phase factors, including the translation difference between the images; it effectively "centers" the two images to be registered. It can be shown that the rotation of an image by an angle θ corresponds to the rotation of its Fourier magnitude by an angle θ modulo 180°, and the scaling of the image coordinate system by a factor α corresponds to the scaling of the Fourier magnitude by a factor $1/\alpha$. The 180° ambiguity in the rotation angle is based on the fact that the Fourier magnitude of a real image is an even function[40]. For now, we ignore this ambiguity. Later, we will explain how to resolve it.

It should be noted that by transforming from the spatial domain to the Fourier magnitude domain, one effectively moves from a 4-dimensional parameter space to a 2-dimensional parameter space of rotation and scaling. The next step consists of further transformation of the Fourier magnitude so that the effects of rotation and scaling are converted into shifts. The use of polar coordinates is the standard approach; a rotation in rectangular coordinates becomes a cyclic shift in the angular direction of the polar coordinates, and a global scaling in the rectangular coordinates becomes a scaling in the radius direction of the polar coordinates. By taking the logarithm of the radius coordinate, scaling can be converted into a shift. The two operations, performed simultaneously, are called a log-polar transformation which is shown in panel C of Figure 2.

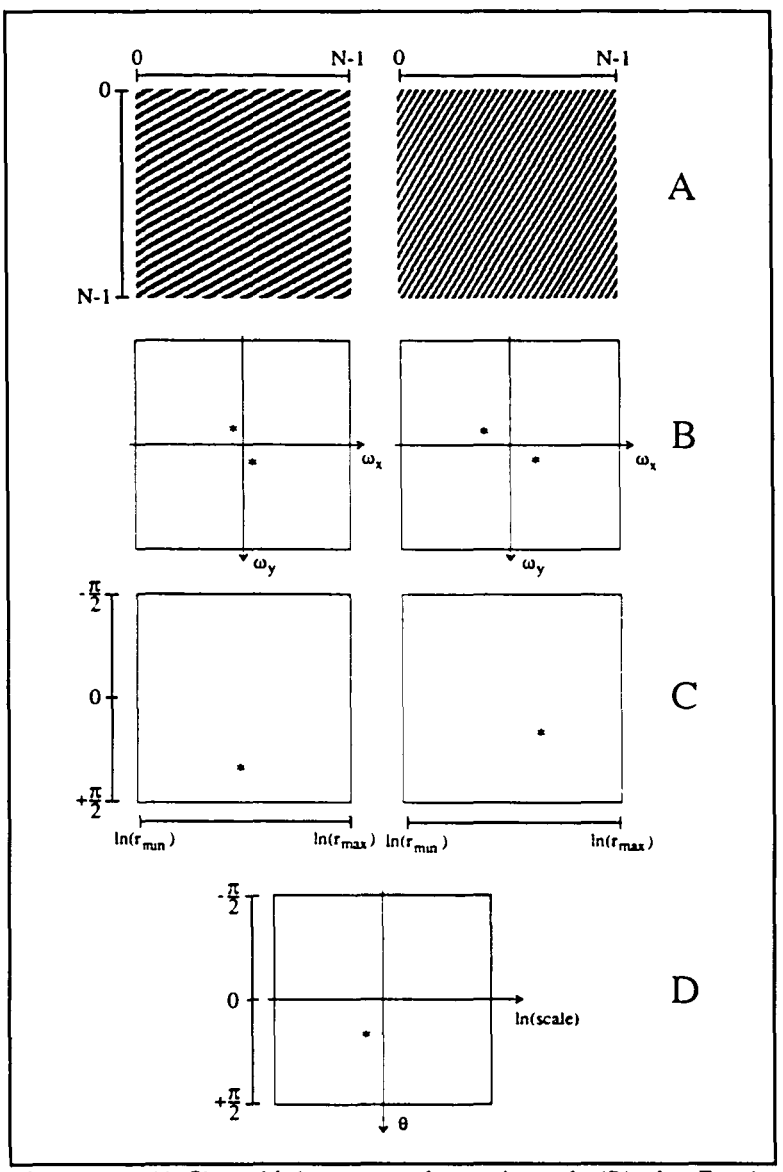

Figure 2: *(A) Sinusoid images to be registered, (B) the Fourier magnitudes, (C) the log-polar mapping of the Fourier magnitude, (D) the cross-correlation of the log-polar Fourier magnitudes.*

The positive peak of the cross-correlation function of the two log-polar Fourier magnitudes corresponds to the optimal rotation and scaling parameters (Figure 2D). The reader's attention is drawn to the fact that the cross-correlation needs to be "cyclic" in the angle coordinate direction, and "linear" in the logarithm-of-radius coordinate direction. This can readily be achieved by zero-padding the logarithm-of-radius direction only, and performing a cyclic cross-correlation with an FFT-based algorithm. Once the rotation and scaling parameters are determined, one of the images is appropriately scaled, rotated and a cross-correlation in the spatial domain is performed to find the optimal translation parameters. Above, we had mentioned that there is a 180° ambiguity in the rotation angle determined. In our case rotations of greater than ±10° are unlikely; therefore, we simply take the rotation angle closest to 0°. If, for an application, the whole range of rotations were possible, the ambiguity could easily be resolved with an additional cross-correlation in the original image domain.

In theory, the cross-correlation of the log-polar Fourier magnitudes can be made as accurate as the spatial domain resolution allows, but, in practice, the already increased space-bandwidth requirements (see section 3.1) of the log-polar transform, coupled with limited computational resources, allow only a relatively coarse (and discrete) parameter space. However, a computationally feasible yet accurate registration can be achieved by a hybrid method that uses the correlation of the log-polar Fourier magnitude to determine the neighborhood of the global optimum in the 4-dimensional discrete parameter space, and then, in this neighborhood, uses an adaptive local search to refine the parameter values to any desired resolution. Of course, it is assumed that in the neighborhood of the global optimum the similarity function is unimodal, and therefore the local search will actually converge to the global maximum and not to a local maximum.

3. IMPLEMENTATION

There are several implementation issues that require attention. Probably the most important one is the choice of the number of samples in the log-polar domain. We explain the compromises between transform domain fidelity and memory requirements in section 3.1. The specifics of the local adaptive search strategy are detailed in section 3.2. In section 3.3 we summarize all the steps necessary to bring two images into registration according to our method.

3.1 Resolution of the log-polar mapping

Let's assume that the Fourier magnitude has NxN samples in the rectangular domain, and N_θxN_ρ samples in the log-polar domain. The choices for N_θ and N_ρ are not trivial for two reasons. Firstly, for a range of $0 \le r < r_{max}$ for the radius, r, the corresponding range is $-\infty < \rho < \ln(r_{max})$, for the logarithm of radius, ρ. This suggests an infinite number of samples. Secondly, the effective resolution in the log-polar domain is space-variant. Therefore, N_θ and N_ρ should be based on a compromise between realistic memory requirements and high resolution in the log-polar domain.

A schematic diagram of the mapping from the rectangular domain to the log-polar domain is shown in Figure 3. As mentioned above, the Fourier magnitude is an even function, so we only need to map half of the plane. Furthermore, because of the difficulties associated with the region around r=0 (which is $\rho=-\infty$), we will only map a semicircular annulus of inner radius r_{min} and outer radius r_{max}. In order to determine N_θ and N_ρ, we will have to determine the size of the resolution elements, $\Delta\theta$ and $\Delta\rho$, in the log-polar domain, and the two radii defining the annulus to be mapped.

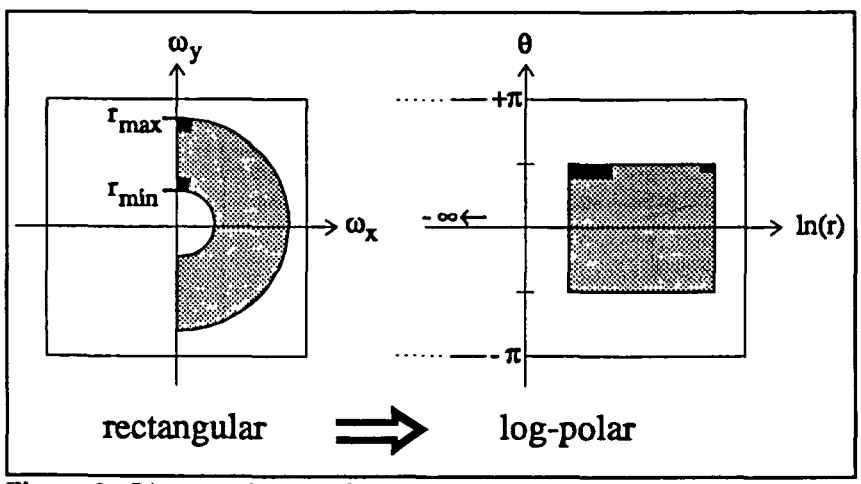

Figure 3: *Diagram showing the semicircular annulus-shaped region of the Fourier magnitude in the rectangular domain mapped to the log-polar domain.*

First, we consider the resolution element size ($\Delta\theta$ and $\Delta\rho$) in the log-polar domain. Without losing generality, we can assume that the size of a resolution element in the rectangular domain is 1. Although the

log-polar domain is uniformly sampled, the effective resolution is space-variant. That is shown schematically with the two pairs of black areas in Figure 3, where the two black areas in the rectangular domain are approximately the same size, but their counterparts in the log-polar domain have drastically unequal sizes. One way to approach this spatially variant resolution is to have the worst-case resolution of the log-polar domain equal the resolution of the rectangular domain. In that case, it can be shown[41,42] that both $\Delta\theta$ and $\Delta\rho$ are equal to $1/r_{max}$, in radians and log-units, respectively.

Second, we consider the range of frequencies to map (r_{max} and r_{min}). In most imaging applications the "signal" is limited to mostly lower frequencies, and the "noise" is distributed over mostly higher frequencies. Therefore, r_{max} should correspond to the application-specific frequency-cutoff above which there is no significant "signal" power. The choice of r_{min} is more flexible because the very low frequency variations usually do not contribute significantly for high resolution registration. Therefore, r_{min} can be chosen in consideration of a realistic N_ρ.

Now, we can calculate the number of samples to be used in the log-polar domain as,

$$\begin{aligned} N_\theta &= \frac{\pi}{\Delta\theta} &&= r_{max} \cdot \pi \\ N_\rho &= \frac{\ln(r_{max}) - \ln(r_{min})}{\Delta\rho} &&= r_{max} \cdot \ln\left(\frac{r_{max}}{r_{min}}\right) \end{aligned} \qquad (1)$$

It is important to reiterate that Eq. (1) is based on unit-resolution in the rectangular domain.

We have derived a comprehensive model of our high resolution fundus imaging system[12,14], which shows the frequency cutoff to be approximately 75 cyc/mm on the retinal surface (note that in earlier work[12,14] the frequencies were measured on the imaging plane). In terms of a unit-resolution rectangular coordinate system, this frequency corresponds to an r_{max} of 93 in the case of 512x512 pixel images with 2.4μm pixels. Because of computational resource limitations, we choose r_{min} as 6, therefore N_θ=292, and N_ρ=254. We use N_θ=256 and N_ρ=256 instead, because most common FFT algorithms require powers of 2 as N. Furthermore, we use bilinear interpolation for the sampling of the log-polar domain, because this type of interpolation is usually considered a good compromise between fidelity and computational requirements.

The log-polar-based registration of two 512x512 images with custom-written software in C (Microsoft C 6.0) on an Intel 80486/33 MHz based IBM/PC compatible computer (433DE, DELL) with 12MB of RAM and running under MS-DOS (ver. 5.0, Microsoft) and a DOS-extender (286|DOS-Extender, Phar-Lap) requires 3.5 minutes (the time includes all operations to calculate the four registration parameters). The software is written in a completely portable fashion that requires simple recompilation to run on different hardware platforms.

3.2 Local adaptive search strategy

In practice, several factors contribute to the loss of accuracy of registration using the log-polar mapping. The major factors are: the loss of phase information due to the Fourier magnitude transformation; frequency domain interpolation errors; and, the ever-existing limited memory constraining the log-polar domain samples. A local search of the parameter space in the neighborhood of the already-determined registration point allows the high-resolution registration that we are aiming for. The local search uses the spatial domain images and therefore has all of the image information available.

The local search amounts to the multidimensional maximization of a similarity measure with respect to the four registration parameters. The correlation coefficient is used as the similarity measure, and it is defined as:

$$corr.coeff. = \frac{\sum_x \sum_y f_1(x,y)\, f_2(x,y)}{\sqrt{\sum_x \sum_y f_1^2(x,y)}\ \sqrt{\sum_x \sum_y f_2^2(x,y)}} \qquad (2)$$

where $f_1(x,y)$ and $f_2(x,y)$ are the two images, and the double summations are performed over all pixels that have valid data in both images after the transformation.

The multi-dimensional maximization is achieved with the "simplex" method[43,44], which is not necessarily the fastest, but is straightforward, does not require derivative calculations, and can be extremely robust. The method requires five initial points that define the initial shape of the simplex in 4-dimensional (continuous) parameter space. The registration parameters determined with the log-polar mapping are used as one point of this simplex, and the other four points of the simplex are calculated from the first point as shown below:

Simplex point 2: $(x_0+0.5, y_0, \theta_0, \alpha_0)$
Simplex point 3: $(x_0, y_0+0.5, \theta_0, \alpha_0)$
Simplex point 4: $(x_0, y_0, \theta_0+0.5, \alpha_0)$
Simplex point 5: $(x_0, y_0, \theta_0, \alpha_0+0.05)$

where $(x_0, y_0, \theta_0, \alpha_0)$ are the x-coordinate and y-coordinate translations in pixels, rotation angle in degrees, and global scaling factor, respectively. The arbitrary initial shape of the simplex might affect the number of iterations necessary for the optimization, but it does not affect convergence.

The number of similarity function evaluations required to achieve a prespecified convergence criterion varies from image to image. In our case, the algorithm requires on the average 170 evaluations to reach a change in the value of the similarity measure (so-called fractional tolerance) of 1×10^{-7}, which is equal to the machine resolution. For the above mentioned hardware/software combination, each similarity function evaluation (including the transformation of one of the images using bilinear interpolation) takes 7.5 seconds (approximately 21 minutes total). The above specified tolerance is very strict for most imaging applications, and it can be relaxed to achieve faster convergence.

3.3 Summary of the registration method

1) Calculate the 2D FFT for each image;
2) take the magnitude of the FFT;
3) sample a semi-circular annulus of the Fourier magnitude limited by application-specific frequency cutoffs onto a log-polar grid in each image;
4) zero-extend the log-polar Fourier magnitudes in the logarithm-of-radius coordinate direction only;
5) perform 2D cyclic cross-correlation using an FFT-based algorithm;
6) determine the peak of the cross-correlation function and the corresponding rotation angle and scale factor;
7) appropriately rotate and scale one of the original images;
8) cross-correlate the rotated and scaled image with the other;
9) determine the peak of the cross-correlation and the two corresponding translation parameters; and
10) use a local search to find the four optimal (continuous) registration parameters.

4. RESULTS

In this section, we describe three sets of results. Firstly, the registration results are given in the case of actual retinal images to which exactly known deformations are applied. Secondly, simulated images are used, and the effect of noise on registration accuracy is explored. Lastly, registration is applied to actual retinal images with unknown registration parameters.

4.1 Simulation: Experiment 1

In this set of experiments, we considered five different 640x640 pixel digitized regions from the fundus photographs of two patients. We applied five different combinations of rotation, scaling, translation transformations to each one of the five images, generating 25 images with exactly-known transformations. Next, we cropped the edge-regions of all images, making sure that each image is a 512x512 pixel full-frame image. Our registration method was applied to the twenty-five transformed images and their corresponding untransformed images. The mean value of the absolute error between the calculated and actual transformations are summarized in the following table.

	rotation angle (degrees)	scale factor	x translation (pixels)	y translation (pixels)
Mean absolute error using log-polar method	0.2	0.02	0.4	0.7
Mean absolute error after local search	0.001	0.0000	0.002	0.002

There are several points to be made. First of all, it should be emphasized that these results are based on "perfect" conditions: the images used for registration were identical with the exception of the interpolation errors introduced during the simulated deformation. Second, the noise in these retinal images acts as "information" when a second (deformed) image is produced. Therefore, the registration with the local search was more accurate than would be expected in case of natural images with lower bandwidth. The above results can be considered as validating the registration method under "perfect" conditions.

4.2 Simulation: Experiment 2

Two issues were raised in the previous section: bandwidth of images used and the effect of random noise. To clarify these issues we generated three simulated images (640x640 pixels) consisting of bright, randomly sized and shaped objects on a slowly (and randomly) varying background. These simulated images, when put through the model of our imaging system, produced images that qualitatively looked similar to the retinal images of interest to us. Therefore, our simulations were considered to represent a realistic "noiseless scene". From each simulated image five new images were made using different combinations of rotation, scaling and translation transformations. The edge regions of the original and transformed images were cropped so that 512x512 pixel full-frame image resulted. White zero-mean Gaussian noise of various magnitudes was added to the three original simulations as well as to the fifteen transformed simulations. Registration was performed between each transformed image and the corresponding untransformed image, for each noise level. The noise level is reported as a signal-to-noise ratio[45] (SNR), in dB units:

$$SNR = 10 \log_{10} \left(\frac{\sigma^2_{image}}{\sigma^2_{noise}} \right) \tag{3}$$

where σ^2_{image} is the variance of the original (noiseless) image, and σ^2_{noise} is the variance of the noise added.

The following table lists the registration results obtained as a function of noise added. Upper entries in each cell represent the mean absolute error using the log-polar method only, and the lower entries represent the mean absolute error when the log-polar method was followed with a local search for optimum parameters.

SNR (dB)	rotation angle (degrees)	scale factor	x translation (pixels)	y translation (pixels)
∞	0.1 **0.003**	0.01 **0.0001**	0.2 **0.03**	0.2 **0.03**
5	0.1 **0.01**	0.01 **0.0003**	0.2 **0.04**	0.2 **0.05**
0	0.1 **0.02**	0.01 **0.0003**	0.2 **0.1**	0.2 **0.1**
-5	0.2 **0.07**	0.01 **0.001**	0.4 **0.3**	0.4 **0.3**

The first interesting observation is that the accuracy of the log-polar method is not affected by the noise level up to SNRs of 0 dB, and the accuracy drops only slightly for the SNR of -5 dB. The second interesting observation is that in the case of translation at high noise levels (SNR=0 and -5 dB) the accuracy with or without the local search is very similar (0.4 pixels without vs. 0.3 pixels with local search at -5 dB). On the other hand, in the case of rotation and scaling, there is a factor of 3 and 10 increase, respectively, in the accuracy due to the local search at an SNR of -5 dB. The question arises whether the local search is worth doing at high noise levels. It is important to point out that a 1% error in the scale factor corresponds to up to 5 pixels error in some parts of the image. In comparison, the local search reduces the error to subpixel levels. A similar statement can be made for rotation also. Therefore, it can be concluded that the local search does significantly improve registration accuracy even in very high noise levels with an SNR of -5 dB.

4.3 Actual fundus images

As mentioned in the introduction, our goal is the accurate registration of fundus images of XLRP carriers taken years apart so that any changes in the appearance of the tapetal-like reflex (Fig. 4) may be detected. Of course, when there are underlying possible changes in image content, it is very difficult (visually or using a computer) to judge the accuracy of the registration method. Many times misregistration might be mistaken as change, or actual change might be mistaken as misregistration. Therefore, we wanted to further validate the accuracy of the method by registering different fundus photographs taken at the same photography session. These images should have no underlying change, and if all of our assumptions are correct, the difference between registered pairs should consist of image noise.

For this purpose we selected six pairs of fundus photographs. Each pair was taken at the same photography session and imaged approximately the same region of the eye. Corresponding areas were digitized (640x640 pixels) from each of the six pairs of images, and they were registered using the method described in this paper. The calculated registration parameters were used to transform one of the images, and then edge regions were cropped so that full-frame 512x512 registered pairs of images result. The following table lists the registration parameters, and the standard deviation of the difference image calculated from each registered pair.

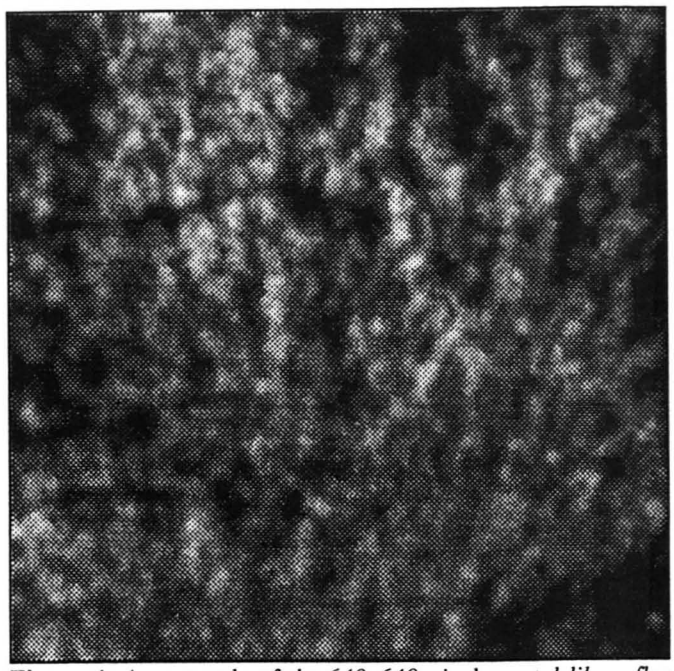

Figure 4: *An example of the 640x640 pixel tapetal-like reflex image used for registration experiments.*

	Registration parameters				
Image #	rotation angle (degrees)	scale factor	x translation (pixels)	y translation (pixels)	Std. dev. of diff. (grey levels)
1	3.08	0.998	-15.3	-27.0	3.54
2	1.76	0.997	29.2	48.4	5.33
3	1.77	1.006	2.6	23.2	4.35
4	-0.61	0.999	17.0	-72.4	3.73
5	0.66	0.994	-8.0	-27.9	5.87
6	0.64	0.990	0.9	-20.2	5.10

As expected, the rotational differences between images were small. The difference in scale was less than 1% in each case, which corresponds to the expectation that there should be no significant scale differences between photographs taken at the same session. The difference images between registered pairs were inspected visually, and in each case they consisted of "noise" with no apparent "structure" suggestive of misregistration. Furthermore, the standard deviations of these difference images (average standard deviation=4.7) were comparable to the value of 4.6 calculated from the difference of simulated images generated with the model of the imaging system.

5. CONCLUSIONS

In this paper, we presented a highly accurate method of registration of digital images. The method assumes that one of the images is a translated, rotated and scaled version of the other. The method combines a global optimization in the 4-dimensional discrete parameter space with a local optimization in the neighborhood of the global optimum. Correlation coefficient is used as a similarity criterion in both local and global optimizations. The accuracy of the registration method for 512x512 pixel images is estimated to be better than 0.07 degrees, 0.1 %, and 0.3 pixels for rotation, scaling, and translation, respectively. The software requires approximately 25 minutes to register a pair of 512x512 pixel images on a 80486/33 MHz based IBM/PC compatible.

6. SOFTWARE

The registration software described in this paper is available free of charge to interested individuals associated with educational institutions. Contact Mr. Cideciyan at the above listed address or at acidec@umbio.med.miami.edu.

7. ACKNOWLEDGMENTS

This research was supported in part by Public Health Service research grant EY05627 (Dr. Jacobson); the National Retinitis Pigmentosa Foundation, Inc. (Bethesda, Maryland); and, The Chatlos Foundation, Inc. (Longwood, Florida). Mr. Cideciyan was supported by a University of Miami Fellowship during part of the above-reported work. Dr. Jacobson is a Research to Prevent Blindness Dolly Green Scholar.

The authors thank Dr. J.S. Kippenhan and Dr. A. Apicella for critical review of this manuscript.

8. APPENDIX

All registration methods which assume that the image deformations are well approximated with translation, rotation, and scaling, use a certain sequence of these transformations in their attempts at parameter determination. Usually, the actual deformation that was imposed upon an image is not known, and a sequence is chosen based on methodological convenience. In this Appendix, we will show that in the case of translation, rotation, and global scaling, any ordering (or number) of these transformations can be represented with any other ordering (or number). For example, a rotation followed by a translation can be represented with a translation followed by a rotation. Interestingly, this property is not valid when the more general form of scaling is used. We will first define the geometric transformation conventions and nomenclature, and then we will prove the claim made above.

Translation of a point in an image by Δx in the x-coordinate direction and by Δy in the y-coordinate direction is mathematically defined as

$$\begin{aligned} x^* &= x + \Delta x \\ y^* &= y + \Delta y \end{aligned} \tag{A1}$$

where (x^*,y^*) are the coordinates of the new point, (x,y) are the coordinates of the original point, and Δx, Δy are the amounts of the translation in the two respective coordinate directions. Eq. (A1) can be expressed in square-matrix form (using homogeneous coordinates) for both concise notation and easy concatenation of several transformations. That is,

$$\begin{bmatrix} x^* \\ y^* \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{A2}$$

where the square transformation matrix is called a translation matrix symbolized as

$$T_{\Delta x,\Delta y} \equiv \begin{bmatrix} 1 & 0 & \Delta x \\ 0 & 1 & \Delta y \\ 0 & 0 & 1 \end{bmatrix} \tag{A3}$$

The translation of a complete image can be concisely written as

$$X^* = T_{\Delta x,\Delta y}\, X \tag{A4}$$

where X is an N^2x3 matrix that refers to all pixel coordinates of an NxN image, and X^* is an N^2x3 matrix that refers to the transformed coordinates of each pixel; that is,

$$X = \begin{bmatrix} x_1 & x_2 & \cdot & x_1 & \cdot & \cdot & x_N \\ y_1 & y_1 & \cdot & y_2 & \cdot & \cdot & y_N \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \qquad X^* = \begin{bmatrix} x_1^* & x_2^* & \cdot & x_1^* & \cdot & \cdot & x_N^* \\ y_1^* & y_1^* & \cdot & y_2^* & \cdot & \cdot & y_N^* \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{A5}$$

In the same way, we can define a global scaling matrix S_α that scales an image by a factor of α in both coordinate directions

$$S_\alpha \equiv \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{A6}$$

Similarly, a rotation of the image about the center of the coordinate system by an angle θ is represented by the matrix

$$R_\theta \equiv \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{A7}$$

In many registration applications, a certain sequence of these three transformations is used with the aim of compensating for the actual (unknown) deformation. The following question arises: Do we have to know the sequence (and the number) of the transformations that caused the actual geometric deformations, or can we represent any transformation with a standard set? We will show that any number of translation, rotation and scaling transformations applied to an image in any sequence can be equivalently represented by a translation, a rotation and a scale. The equivalent scale factor is the product of all applied scale factors and the equivalent rotation angle is the sum of all applied rotation angles. The equivalent translation depends on the number, type and sequence of the transformations applied. In symbolic form:

$$T_{\Delta x_1,\Delta y_1} R_{\theta_1} S_{\alpha_1} T_{\Delta x_2,\Delta y_2} R_{\theta_2} S_{\alpha_2} \cdots T_{\Delta x_n,\Delta y_n} R_{\theta_n} S_{\alpha_n} \equiv T_{\Delta w,\Delta z} R_{\theta_1+\theta_2+\cdots+\theta_n} S_{\alpha_1\alpha_2\cdots\alpha_n} \tag{A8}$$

where Δw, Δz are functions of all Δx_i, Δy_i, α_i and θ_i for i=1...n, as well as the specific sequence of the transformations.

The derivation of Eq. (A8) is straightforward through the use of the following identities:

$$R_\theta S_\alpha = S_\alpha R_\theta \tag{A9}$$

$$R_\theta T_{\Delta x,\Delta y} = T_{(\Delta x\cos\theta+\Delta y\sin\theta),(\Delta y\cos\theta-\Delta x\sin\theta)} R_\theta \tag{A10a}$$

$$T_{\Delta x,\Delta y} R_\theta = R_\theta T_{(\Delta x\cos\theta-\Delta y\sin\theta),(\Delta y\cos\theta+\Delta x\sin\theta)} \tag{A10b}$$

$$S_\alpha \, T_{\Delta x,\Delta y} = T_{(\alpha \Delta x),(\alpha \Delta y)} \, S_\alpha \tag{A11a}$$

$$T_{\Delta x,\Delta y} \, S_\alpha = S_\alpha \, T_{(\Delta x/\alpha),(\Delta y/\alpha)} \tag{A11b}$$

$$T_{\Delta x_1,\Delta y_1} \, T_{\Delta x_2,\Delta y_2} = T_{(\Delta x_1+\Delta x_2),(\Delta y_1+\Delta y_2)} \tag{A12}$$

$$R_{\theta_1} \, R_{\theta_2} = R_{(\theta_1+\theta_2)} \tag{A13}$$

$$S_{\alpha_1} \, S_{\alpha_2} = S_{(\alpha_1 \alpha_2)} \tag{A14}$$

Eq.s (A9), (A10a), (A10b), (A11a) and (A11b) allow the grouping of similar transformations; i.e. the left side of Eq. (A8) can be transformed to a number of translations, followed by a number of rotations, followed by a number of scalings. A group of a given type of transformation can be further reduced to a single transformation using Eq.s (A12), (A13), and (A14). There are two important points that need mentioning:

a) using Eq. (A8) it is straightforward to show that any sequence of translation, rotation and global scaling can be represented with any other sequence, as long as the correct parameter values are chosen; and

b) the validity of Eq. (A8) is due to the choice of global (uniform) scaling, as opposed to the more general form of scaling with two different scale factors in the two coordinate directions. The general form of scaling would invalidate Eq. (A9) and thereby invalidate Eq. (A8).

9. REFERENCES

1. E. Peli, T.R. Hedges, B. Schwartz, "Computerized enhancement of retinal nerve fiber layer", Acta Ophthalmologica, 64:113-122, 1986.
2. E. Peli, T. Peli, "Restoration of retinal images obtained through cataracts", IEEE Trans. Med. Imag., 8(4):401-406, 1989.
3. R.P. Phillips, T. Spencer, P.G.B. Ross, P.F. Sharp, J.V. Forrester, "Quantification of diabetic maculopathy by digital imaging of the fundus", Eye, 5:130-137, 1991.
4. N.P. Ward, S. Tomlinson, C.J. Taylor, "Image analysis of fundus photographs", Ophthalmology, 96(1):80-86, 1989.
5. J. Gilchrist, "Analysis of early diabetic retinopathy by computer processing of fundus images - a preliminary study", Ophthal. Physiol. Opt., 7(4):393-399, 1987.
6. M.H. Goldbaum, N.P. Katz, M.R. Nelson, L.R. Haff, "The discrimination of similarly colored objects in computer images of the ocular fundus", Invest. Ophthalmol. & Vis. Sci., 31(4):617-623, 1990.
7. S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters", IEEE Trans. Med. Imag., 8(3):263-269, 1989.
8. M.J. Cox, I.C.J. Wood, "Computer-assisted optic nerve head assessment", Ophthal. Physiol. Opt., 11:27-35, 1991.
9. V.R. Algazi, J.L. Keltner, C.A. Johnson, "Computer analysis of the optic cup in glaucoma", Invest. Ophthalmol. & Vis. Sci., 26:1759-1770, 1985.
10. R. Varma, G.L. Spaeth, "The PAR IS 2000: A new system for retinal digital image analysis", Ophthalmic Surgery, 19(3):183-192, 1988.
11. P. Nagin, B. Schwartz, G. Reynolds, "Measurement of fluorescein angiograms of the optic disc and retina using computerized image analysis", Ophthalmology, 92(4):547-552, 1985.
12. A.V. Cideciyan, J.H. Nagel, S.G. Jacobson, "Modeling of high resolution digital retinal imaging", Proc. 13th Ann. IEEE/EMBS Conf., Orlando, FL, Oct. 31-Nov. 3, pp.264-266, 1991.
13. A.V. Cideciyan, S.G. Jacobson, J.H. Nagel, "Multi-scale segmentation of retinal images", Proc. 13th Ann. IEEE/EMBS Conf., Orlando, FL, Oct. 31-Nov. 3, pp.267-268, 1991.
14. A.V. Cideciyan, An image processing study of the tapetal-like reflex in X-linked retinitis pigmentosa heterozygotes, Ph.D. Dissertation, University of Miami, Coral Gables, FL, 1992.
15. I. Mann, Developmental abnormalities of the eye, Cambridge University Press, London, pp.164-166, 1937.

16. H.F. Falls, C.W. Cotterman, "Choroidoretinal degeneration", Arch. Ophthalmol., 40:685-703, 1948.
17. R.B. Szamier, E.L. Berson, "Retinal histopathology of a carrier of X-chromosome-linked retinitis pigmentosa", Ophthalmology, 92(2):271-278, 1985.
18. G.A. Fishman, A.B. Weinberg, T.T. McMahon, "X-linked recessive retinitis pigmentosa: clinical characteristics of carriers", Arch. Ophthalmol. 104:1329-1335, 1986.
19. T. Behrendt, K.E. Doyle, "Reliability of image size measurements in the new Zeiss fundus camera", Am. J. Ophthalmol., 59:896-899, 1965.
20. P.E. Anuta, "Spatial registration of multispectral and multitemporal digital imagery using fast Fourier transform techniques", IEEE Trans. Geo. Elec., 8:353-368, 1970.
21. M. Merickel, "3D reconstruction: the registration problem", Comp. Vis. Graph. Image Proc., 42:206-219, 1988.
22. K.D. Toennies, J.K. Udupa, G.T. Herman, I.L. Wornom, S.R. Buchman, "Registration of 3D objects and surfaces", IEEE Comp. Graph. Appl., 10(3):52-62, 1990.
23. S.T. Barnard, W.B. Thompson, "Disparity analysis of images", IEEE Trans. Pat. Anal. Mach. Intel., PAMI-2:333-340, 1980.
24. G.C. Stockman, S. Kopstein, S. Benett, "Matching images to models for registration and object detection via clustering", IEEE Trans. Pat. Anal. Mach. Intel., PAMI-4(3):229-241, 1982.
25. R.Y. Wong, E.L. Hall, "Scene matching with invariant moments", Comp. Graph. Image Proc., 8(1):16-24, 1978.
26. A. Goshtasby, "Template matching in rotated images", IEEE Trans. Pat. Anal. Mach. Intel., PAMI-7(3):338-344, 1985.
27. A. Goshtasby, "Registration of images with geometric distortions", IEEE Trans. Geo. Rem. Sens., 26(1):60-64, 1988.
28. W.K. Pratt, "Correlation techniques of image registration", IEEE Trans. Aero. Elec. Sys., 10:353-358, 1974.
29. E. de Castro, C. Morandi, "Registration of translated and rotated images using finite Fourier transforms", IEEE Trans. Pat. Anal. Mach. Intel., PAMI-9(5):700-703, 1987.
30. D.I. Barnea, H.F. Silverman, "A class of algorithms for fast image registration", IEEE Trans. Comp., C-21(2):179-186, 1972.
31. E. de Castro, G. Cristini, A. Martelli, C. Morandi, M. Vascotto, "Compensation of random eye motion in television ophthalmoscopy: preliminary results", IEEE Trans. Med. Imag., MI-6(1):74-81, 1987.
32. M. Herbin, A. Venot, J.Y. Devaux, E. Walter, J.F. Lebruchec, L. Dubertret, J.C. Roucayrol, "Automated Registration of Dissimilar Images: Application to Medical Imagery", Comp. Vis. Graph. Image Proc., 47:77-88, 1989.
33. A. Apicella, J.S. Kippenhan, J.H. Nagel, "Fast multi-modality image matching", Proc. SPIE Vol. 1092, Medical Imaging III: Image Processing, Jan. 31-Feb. 3, pp.252-263, 1989.
34. D. Casasent, D. Psaltis, "Position, rotation, and scale invariant optical correlation", Applied Optics, 15(7):1795-1799, 1976.
35. A. VanderLugt, "Signal detection by complex filtering", IEEE Trans. Inform. Theory, IT-10(2):139, 1964.
36. C.S. Weaver, J.W. Goodman, "Technique for optically convolving two functions", Applied Optics, 5:1248-1249, 1966.
37. D. Casasent, D. Psaltis, "Deformation invariant, space-variant optical pattern recognition", in Progress in Optics, E.Wolf (ed.), North-Holland Publishing Co., Amsterdam, pp.290-356, 1978.
38. J. Altmann, H.J.P. Reitbock, "A fast correlation method for scale- and translation-invariant pattern recognition", IEEE Trans. Pat. Anal. Mach. Intel., PAMI-6(1):46-57, 1984.
39. A.S. Jensen, L. Lindvold, E. Rasmussen, "Transformation of image positions, rotations, and sizes into shift parameters", Applied Optics, 26(9):1775-1781, 1987.
40. R. Bracewell, The Fourier transform and its applications, McGraw-Hill Book Co., New York, pp.47, 1965.
41. D. Casasent, D. Psaltis, "Space-bandwidth product and accuracy of the optical Mellin transform", Applied Optics, 16(6):1472, 1977.
42. D. Casasent, D. Psaltis, "Accuracy and space bandwidth in space variant optical correlators", Optics Comm., 23(2):209-212, 1977.
43. W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical recipes in C, Cambridge University Press, Cambridge, pp.305-309, 1988.
44. J.A. Nelder, R. Mead, "A simplex method for function minimization", Computer Journal, 7(4):308-313, 1965.
45. A.K. Jain, Fundamentals of digital image processing, Prentice Hall, Englewood Cliffs, pp.59, 1989.