DIGITAL ANALYSIS OF HIGH RESOLUTION FUNDUS IMAGES

JOACHIM H. NAGEL AND ARTUR V. CIDECIYAN
Departments of Biomedical Engineering and Ophthalmology
University of Miami. Coral Gables and Miami. Florida. USA

ABSTRACT

Fundus photography is a common procedure in ophthalmology providing high resolution images of the inside back portion of the eye to diagnose diseases of the retina and the optic nerve, and to record their progress over time. In many instances, objective, quantitative, reproducible and reliable interpretation of fundus images requires their computerized analysis. A comprehensive system for digital analysis of high resolution fundus images has to address virtually all engineering aspects of medical image processing: restoration, segmentation, pattern recognition, and registration. Based on the specific application of investigating the tapetal-like reflex, a retinal reflection uniquely present in carriers of X-linked retinitis pigmentosa (XLRP), novel approaches to the various stages of image processing are presented, and applications in other areas of medical diagnostics are outlined.

The first step in the analysis of fundus images is the search for a comprehensive mathematical model for the relationship between the biological structure and its recorded 'image'. The multifarious transfer function derived from this model enables the development of image restoration methods to compensate for the imperfections of the imaging system which includes the four components eye, camera, film, and scanner. A Zeiss fundus camera and color transparency film are used for the fundus photography, and a slide scanner is used to digitize the color slides (6μm/pixel). A regularized linear restoration method based on a simplified imaging system model serves to reduce the image noise and to recover the high frequency information. A multi-scale segmentation method is used to separate the tapetallike reflex patches from the retinal background. In order to analyze the reflex progression over time, a high precision registration method compensates for differences between images in terms of translation, rotation, and scaling. The technique is based on a global correlation analysis of log-polar transformed images in a discrete parameter space followed by a refinement and optimization in the 4-dimensional continuous parameter space using the original images.

Analysis of fundus photographs of XLRP carriers showed that the patches making up the tapetal-like reflex have, on the average, 30% higher reflectance compared to non-reflex retina. Mathematical morphology methods applied to segmented images suggest that the reflex is made of elongated structures with a small dimension of less than 14μm (on the retina) and a preferred orientation towards the fovea.

Biomed Eng Appl Basis Comm, 1992 (Dec);4:645-682
Keywords: Fundus image processing, Tapetal-like reflex, Regularization, Restoration, Segmentation, Pattern recognition, Registration.

INTRODUCTION

In the field of ophthalmology, the term 'fundus' refers to the inside back portion of the eye. Photography of the fundus 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 (slides) are usually analyzed by expert ophthalmologists, but computers are being used more and more to aid the experts, especially in quantitation. As computers can only deal with digital information, color fundus photographs must be sampled and digitized. Though digital cameras have become available, still existing shortcomings and the desire to use available data banks will make it necessary to work with the traditional photographs for many more years to come.

The use of digital images in ophthalmology is not new. Many researchers have applied digital image processing methods to quantify abnormalities in fundus photographs. Some examples 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]. Interestingly, most of the literature has concentrated on choices of digital image processing methods, but neglected the fundamental problem of the characterization of the imaging system that produces the digital image data. The relationship between the light intensity originating at the fundus and values of a corresponding pixel in a digital image is almost always assumed to be linear. Furthermore, the noise inevitably contained in all images is usually considered to be signal-independent, additive and white Gaussian. These kinds of simplifying assumptions might very well be reasonable for a specific problem at hand but they are not usually verified as such. Additionally, most of the literature has digitized fundus photographs at resolutions that are much lower than the resolution of the information available.

Therefore, there is a need for an adequate mathematical model that relates the values of pixels in a digital fundus image to the actual light intensity originating in the fundus, and accomplishes this at the limiting resolution of the imaging system. Such fundamental information is very useful in determining quantitatively the real information content of a digital fundus image; specifically, a model allows the prediction of the resulting image of a biological structure. A model is not only useful but absolutely necessary if the distorting or blurring effects of the imaging system need to be removed from a given image, the socalled image restoration problem. A model of the imaging system can also be a very good starting point for the consideration of improvements to the imaging system; e.g. whether the use of a lower noise film will significantly improve the resulting

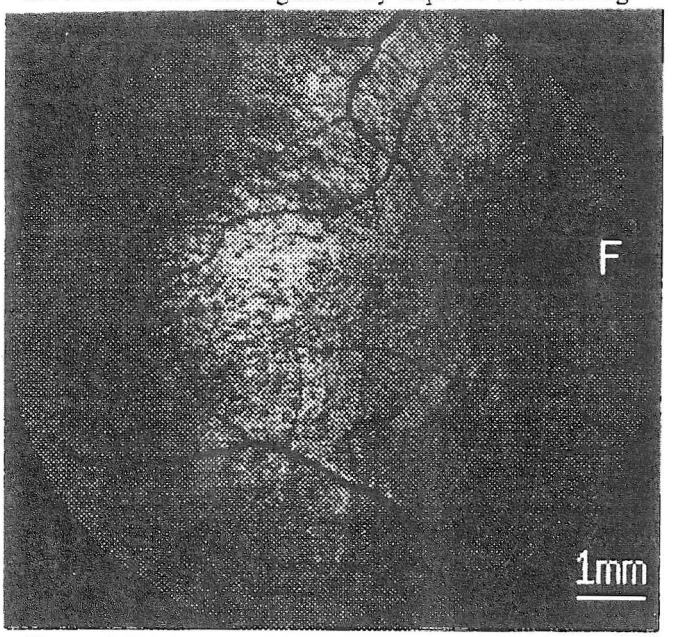


Figure 1. Digitized fundus photograph of the right eye of an X-inked RP carrier with tapetal-like reflex; darker region at the right is the fovea (F), dark lines are retinal blood vessels, and the tapetal-like reflex shows as bright tiny reflections distributed throughout the fundus, especially left of (temporal to) the fovea. The calibration bar is approximately 1 mm, measured on the retinal surface.

Arrived: June 15, 1992; accepted: Oct 14, 1992
Correspondence to: Prof. J.H. Nagel
Department of Biomedical Engineering
University of Miami, P.O. Box 248294
Coral Gables, Florida 33124, U.S.A.

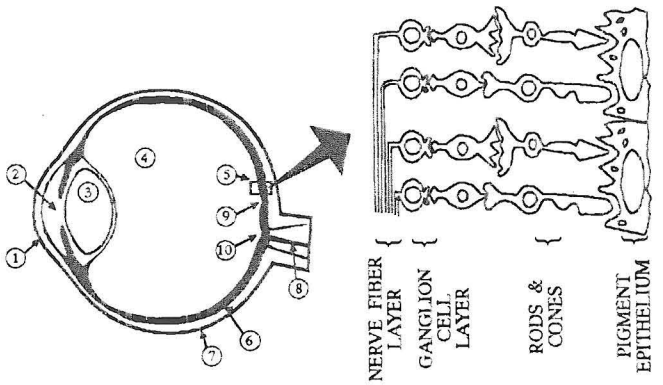


Figure 2: Left, schematic diagram of the major structures of the human eye: (1) cornea, (2) aqueous humor, (3) lens, (4) vitreous humor, (S) retina, (6) choroid, (7) sclera, (8) optic nerve, (9) fovea, (10) optic disc. Right, schematic diagram of major retinal layers.

image quality.

Based on the model and the specific application of investigating X-linked retinitis pigmentosa, novel approaches to the various stages of image processing - restoration, segmentation, pattern recognition, and registration - are presented, and applications in other areas of medical diagnostics, especially diagnosis and treatment of tumors and brain disorders, are outlined.

Background

Retinitis pigmentosa (RP) is the name given to a group of incurable, hereditary diseases that cause progressive degeneration of the retina, the sensory portion of the eye. The retinal degeneration primarily affects the photoreceptor and retinal pigment epithelial cells although the precise molecular and cellular location of the defect is not known in all but a few forms of the disease. More than 100,000 people in the United States and more than 1.5 million people worldwide are estimated to be affected by RP. Patients with RP typically have night blindness, loss of peripheral and eventually central vision, pigmentation within the retina, and abnormal light-evoked electrical responses in electroretinography [12,13].

One of the most severe forms of RP is transmitted on the X-chromosome. Most men with X-linked RP (XLRP) are almost totally blind by age 40 [14]. Women show an attenuated form of the disease, the extent of which is variable from one person to the next. This variability is due to the fact that females are heterozygous with respect to the XLRP gene, i.e. they have one X-chromosome with the RP gene and one normal X-chromosome. Furthermore, one of the two X-chromosomes in each cell of a female is randomly inactivated during early embryonic life (the Lyon hypothesis [15]), and the descendants of each cell contain the same inactivated X-chromosome [16]. Thus each female XLRP patient is a mosaic of two cell types; one type with the diseased X-chromosome being active and the other type with the normal X-chromosome being active. The variability of the expression of the disease seen in females is considered to be consistent with the randomness of the inactivation process [16].

More than 50 years ago, it was noted that there was a golden particulate reflection in the retinas of some women [17, p.164]. Later it was observed that the women with this unusual reflection were also carriers of XLRP [18]. This reflection has been named 'tapetal-like reflex' because it appears similar to the 'tapetal reflex' which is a metallic reflection seen in the eyes of many animals [18-20]. It is believed that the tapetal-like reflex is uniquely seen (and therefore diagnostic) in women who are carriers of XLRP [21]. Fig. 1 shows an example of a tapetal-like reflex as seen in ophthalmoscopy.

Ophthalmoscopists have described the tapetal-like reflex as lying deep to the retinal blood vessels, at or near the retinal pigment epithelium cell layer [18] (see Fig. 2). Neither the origin nor the exact location of this reflection is known. Investigations of post-mortem donor eye tissue have not been helpful since the only report of the histopathology of a carrier of XLRP was in a woman without a tapetal-like reflex [22].

In summary, XLRP is a blinding disease with no known cause and cure. A unique feature of this disease, the tapetal-like reflex seen in carriers, has not been systematically investigated. Such a systematic investigation of the reflex might lead to a better understanding of the mechanism of XLRP.

Specific Goals of this Study

In this study, a non-invasive approach is taken to analyze the tapetal-like reflex from digitized photographic transparencies with the goal of determining its radiometric and geometric properties, and whether it changes with time. This analysis is expected to help narrow down the origin of the reflex, as well as to offer ophthalmologists the techniques which enable them to quantitate the progression of the tapetal-like reflex or other tiny lesions over time.

Some advantages of the non-invasive photographic approach are as follows: as much data as necessary can be collected; the appearance of the reflex can be assessed with serial studies; many members of the same family can be examined; and, families possibly representing different genotypes can be compared. A major disadvantage of the photographic analysis of the tapetal-like reflex is the limited image quality and resolution available in non-invasive imaging techniques. Therefore, in the first part of this paper, specific digital image processing methods will be developed in order to enhance the quality and information content of the digital fundus images. In the second part, these methods will be used to analyze quantitatively the images of the tapetal-like reflex.

Although many researchers have investigated the properties of the components of the imaging system used to take fundus photographs, no information exists in the literature on the overall relationship between the light leaving the retina and its recorded 'image.' Therefore, the first aim is to develop a comprehensive mathematical model of this relationship.

Once the effects of the imaging system are characterized, digital methods may be very useful for the restoration of the images. Thus, the second aim is the development of a restoration method that uses the model of the imaging system to compensate for the imperfections of the imaging system.

Analysis of the tapetal-like reflex,characteristics could be accomplished by using an expert to identify the reflexes. But this approach is usually considered to be subjective, time consuming and not very reproducible. Furthermore, the human visual system is not used to deal with 'smooth' images produced by most restoration methods. Therefore, a third aim is to find an optimal method for the automated separation of the tapetal-like reflex from the retinal background.

To be able to analyze the progression of the tapetal-like reflex over time, one has to be able to exactly overlay two images taken at different times; this is the so-called registration problem. In addition, registration of images is absolutely necessary for validation studies, where multiple images of a scene are used to verify the repeatability of a processing method. Thus, another goal of this study is the development of a method that can accurately register images at very high resolutions.

Once all the necessary methods are developed, they can be applied to the analysis of the radiometric and geometric properties of the patches that make up the tapetal-like reflex. Radiometric properties refer to the intensity and contrast of the reflex, and geometric properties refer to the shape and orientation statistics of the reflex. The new methods will also permit to determine changes of the reflex properties with time.

IMAGING MODEL

A model will be developed in this section for a high resolution digital fundus imaging system which consists of four major components: eye, fundus camera, film and scanner (Fig. 3). The models for each of these components are first derived separately and then the individual component models are combined to arrive at a single mathematical expression that relates the light intensity originating from the fundus to the value of a corresponding pixel in a digital image. The important aspects of the complete model are validated, and finally simulated (synthetic) retinal intensity functions are used to analyze and visualize the deteriorating effects of the imaging system. Unless otherwise specified, all distances and spatial frequencies will refer to the film plane.

1. Eye Model

A casual look at the schematic diagram of the imaging system in Fig. 3 might suggest that the eye is

simply the 'scene' or the 'source of image data', though actually, it is an important component of the imaging system. The illuminating light has to travel through the various layers of the eye before reaching the fundus. Similarly, the light that is reflected from the fundus (i.e. the 'data') has to travel through the same layers of the

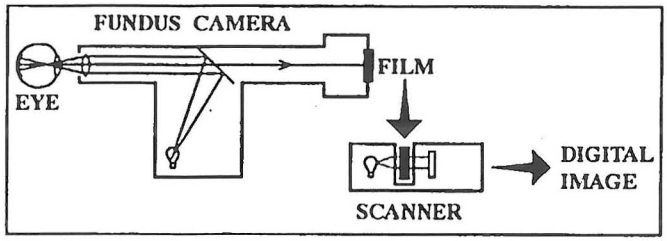


Figure 3. Schematic diagram of the high resolution digital fundus imaging system. The image of the fundus is photographed on film, and the film is digitized in a scanner to produce a digital image.

eye before reaching the camera. Due to the optical characteristics of the eye and the resulting influences on the image formation, the modeling of the eye as an integral part of the imaging system is very important, and it will be discussed first.

Within the relatively small regions of interest, the illumination intensity is uniform. Therefore, the effect of the eye on illumination can be ignored and the light intensity reflecting from the fundus can be considered to originate from a selfluminous fundus.

The effect of transmission of light through the various semi-transparent structures of the eye is approximated by the blurring of a linear system with a spatially-invariant isotropic low-pass transfer function (abbreviated as LSII system). It is assumed that the modulation transfer function (MTF) that completely defines this LSII system may be based on an 'average' eye that will reasonably approximate most 'real' eyes.

The model of the eye in symbolic form is:

$$f_{eye}(x,y) = f(x,y) * h_{eye}(x,y) \qquad (1)$$

where f(x,y) is the luminance of the fundus, $h_{eye}(x,y)$ is the point spread function (PSF) of the eye, $f_{eye}(x,y)$ is the light intensity leaving the eye, and * represents a two-dimensional convolution operation.

MTFs have been published by researchers interested in the optical properties of the human eye [23, 24, 25]. Aerial images (virtual image of a lighted target reflecting back from the fundus) of slits [23] or points [24] have been used commonly to estimate the MTF. Among the many MTFs published for the human eye, the one determined for a 1.5 mm pupil diameter [25] will be used in this study. The choice of the pupil diameter is based on the special construction of the fundus camera which collects only that portion of the light falling within a 1.5 mm diameter area at the center of

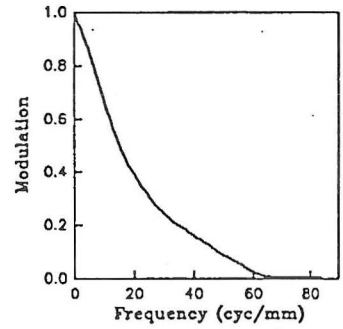


Figure 4. Radial section of the MTF for a human eye with a pupil diameter of 1.5mm (adapted from [25]).

the pupil. The radial section of the eye MTF (shown in Fig. 4) suggests a biological limiting resolution of approximately 60 cyc/mm. This translates loosely to the ability to resolve 3 μm sized objects in the fundus with an ideal camera and recorder.

2. Fundus Camera

The human eye is a spheroidal organ that has relatively opaque layers (sclera/choroid and iris) everywhere except for a small opening, the pupil, which permits the visual observation and the photographic documentation of the fundus with an appropriate optical instrument. Generally, a specialized telescopic system, termed a fundus camera, is preferred over the combination of a contact lens and a microscopic system [26]. The fundus camera is a rather complicated optical instrument because of three important design considerations: there is only one small opening for both illumination and obser-

vation; the spherical fundus must be imaged onto planar photographic film with no appreciable distortion; and,

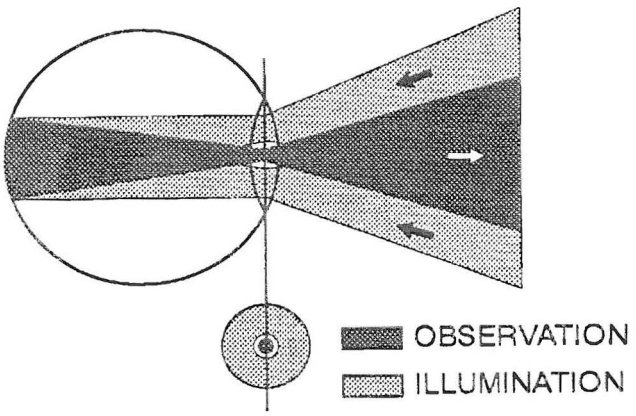


Figure 5. Separation of the illumination and observation light paths in the fundus camera (adapted from [26]).

the depth of focus and the field of view should be as large as possible. Because of the complicated optics and wide field of view, some researchers have claimed that the fundus camera is the main resolution limiting component in a fundus photography system [27].

\The fundus camera used in this study (FF4, Zeiss Corp.) is one of the more traditional and commonly available cameras. It separates the paths of illumination and observation in order to avoid reflection from the cornea overpowering the dim reflection from the fundus. The central 1.5 mm diameter circular area of the pupil is used exclusively for observation whereas the periphery is used for illumination (Fig. 5). The field of view of the fundus camera is 30° of visual angle (approximately 22.5 mm diameter circle on film, and 9 mm diameter circle on the retina).

The camera has a resolution of 8 μm on the optical axis falling off to 15 μm at the periphery as measured on the retinal surface [personal communication from Zeiss Corp.]. The radius dependence of the resolution suggests that the overall transfer function of the camera is spatially variant. However, the very high resolution required for the investigation of the tapetal-like reflex and the limitations in computer resources constrain the analysis to very small regions of a fundus image at any given time. Therefore, it is assumed that the fundus camera can be approximated as an LSII system with a typical low-pass transfer function. Since in this investigation all regions of interest are close to the center, i.e. the optical axis, only a single transfer function is needed for the camera.

The shape of the fundus camera's MTF is approximated by an analytical function that is both reasonable and mathematically tractable. One such function is the isotropic first order exponential decay function (sometimes called the 'first order low-pass Butterworth filter'). This function is defined in the space and spatial-frequency domains as:

$$h_{cam}(r) = \frac{1}{2k} e^{\frac{-|r|}{k}} \qquad (2)$$

$$H_{cam}(w) = \frac{1}{1 + 4\pi^2 k^2 w^2}$$

where $r=(x^2+y^2)^{1/2}$, $w=(u^2+v^2)^{1/2}$, h_{cam} (r) is the PSF, $H_{cam}(w)$ is the MTF, k is a parameter to be determined. It is important to note that the area under the PSF function (which is equal to $H_{cam}(0)$) is unity according to the implied assumption that there is no loss of light in the camera.

The parameter k can be determined using the definition of resolution. 'Resolution' or 'maximum resolving power' is commonly assumed to be the spatial frequency value where the MTF drops to 0.1 [28]. In this study, most regions of interest for the analysis of the tapetal-like reflex will be taken close to the optical axis of the fundus camera. Therefore, the 'worstcase' resolution is assumed to be 10 μm measured on the retina, which corresponds to approximately 25 μm measured on the film plane, and is equivalent to 40 cyc/mm. The value

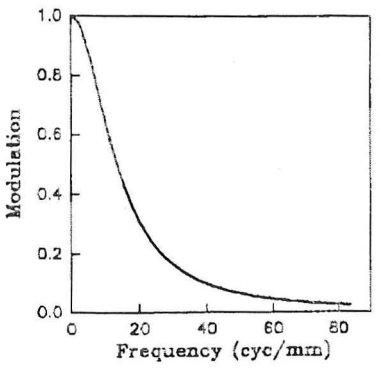


Figure 6. MTF of the fundus camera.

of k can be calculated by replacing 0.1 for $H_{cam}(40)$ in Eq. 2. The value of k is 0.01 mm/cyc and a graphical plot of $H_{cam}(w)$ versus w is shown in Fig. 6.

The model of the fundus camera in symbolic form is:

$$f_{cam}(x,y) = f(x,y) * h_{cam}(x,y) \qquad (3)$$

where f(x,y) is the light intensity entering the fundus camera, $h_{cam}(x,y)$ is the PSF of the fundus camera, and $f_{cam}(x,y)$ is the light intensity reaching the fundus camera detector (i.e. film).

3. Film

The decision to use a specific film to take color fundus photographs is usually a compromise between several competing considerations; sensitivity, contrast, graininess and cost. All fundus images to be analyzed in this study will be Fujichrome 100 RD (Fuji Photo Film Co.).

Photographic film is a detector of exposure which is defined as the integration of light intensity over the exposure time. In physical terms, exposure refers to the total energy of light falling on a unit area of the photographic film. Photographic film 'detects' exposure by changing its opacity (after processing) in relation to the amount of exposure received during photography. The measure of the opacity of film is 'optical density' which is defined as:

$$d = \log_{10}\left(\frac{P_i}{P_t}\right) \qquad (4)$$

where d is the optical density, P_i is the incident (measuring) light intensity on one side of the film, and P_t is the transmitted light intensity on the other side of film.

The relation between the exposure and the optical density of the film is nonlinear. It is called a D-logE curve (or a Hurter-Driffield curve) because this relation is usually specified as density vs. logarithm of exposure. The D-logE curve traced from the manufacturer's data sheet for Fujichrome film [29] is shown in Fig. 7 In this paper, a direct relationship between density and exposure (DE curve) will be used; the DE curve includes the logarithmic transformation as well as the saturation at both ends of the scale. The noise-free response of film to uniform illumination intensity (no blurring) is represented in symbolic form as:

$$d_u = DE(f \cdot t) \qquad (5)$$

where d_u is the optical density of the film, f is the magnitude of the illumination intensity, t is the exposure time, and DE(•) is the DE curve of the film.

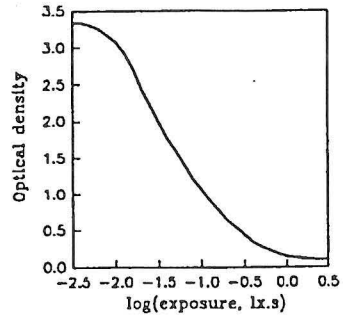


Figure 7. D-logE curve for the Fujichrome 100 RD film (taken from [29]).

When the intensity falling onto the film is not spatially uniform, the response of the film is modified by the input spatial-frequency distribution. Specifically, the photographic film causes a blurring at the higher frequencies in the input intensity function. This is due to the spread of the light intensity from the point of incidence to the surrounding area by reflection, refraction, diffraction and scattering During the development of the film, additional blurring occurs due to the diffusion of reaction products. The two separate blurs before and after the non-linear exposure-to-density conversion are difficult to handle mathematically. Therefore, the so-called 'effective PSF' approximation has been used commonly [30]. The effective PSF approximation is valid for low-contrast photographic images for which the non-linear DE curve can be approximated by a linear function within a small range of an operating point. For a given film, only a single effective PSF is defined that acts in the light intensity domain; i.e. the effective PSF blurs the light intensity before it interacts with the photochemical process of the film. The noise-free response of the film to any input is represented in symbolic form as:

$$d_0(x,y) = DE\left(\left(f(x,y) * h_{film}(x,y) \right) \cdot t \right) \qquad (6)$$

where $d_0(x,y)$ is the gross (noise-free) optical density of the film, f(x,y) is the incoming light intensity distribution, t is the exposure time, and $h_{flm}(x,y)$ is the effective PSF of the film.

A radial section of the MTF for Fujichrome film [29] is shown in Fig. 8. It is interesting to note that the MTF curve has modulation values above 1.0 for some frequencies. That is to say that at some spatial frequencies, the resulting film density will correspond to higher contrast input than the contrast of the actual input. This effect is mainly due to development and it is commonly referred to as the 'adjacency effect' [31].

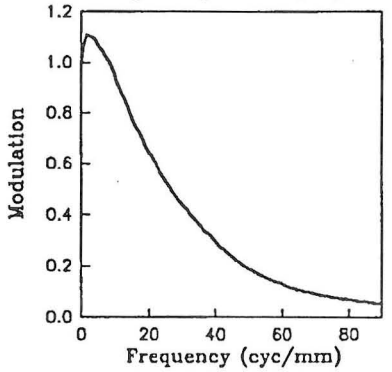


Figure 8. Radial section of the modulation transfer function (MTF) for the Fujichrome film (adapted from [29]).

Microscopically, developed film consists of dye particles on a transparent medium. Exact location, number and sensitivity of these particles are random, and thus cause the so-called 'filmgrain noise.' Film-grain noise is additive, Gaussian and dependent on the signal (i.e. exposure received by the film). A commonly used model [32] for the film-grain noise magnitude is:

$$\sigma_{fgn} = k1_{flm} \, (d_0)^{k2_{flm}} \qquad (7)$$

where σ_{fgn} is the standard deviation of the film-grain noise, $k1_{flm}$ and $k2_{flm}$ are film-specific constants, and d_0 is the gross (noisefree) density of film.

The constant $k2_{flm}$ in Eq. 7 is commonly assumed to be 0.5 [33] and the constant $k1_{flm}$ can be estimated from manufacturer supplied data on film 'granularity.' Granularity is defined as 1000 times the standard deviation of the film density of a uniformly exposed area of the film [28, p.67]. The granularity value for Fujichrome film measured at a gross film density of 1.0 using a 48μm diameter aperture is 11 [29]. At a gross film density of 1.0, $k1_{flm}$ is equal to the standard deviation of the film density (in Eq. 7: if d_0=1.0, then $\sigma_{fgn} = k1_{flm}$). Therefore, $k1_{flm}$ and the manufacturer supplied granularity values are directly related. This relationship, of course, has to take into account the different scanning apertures that are used. The constant $k1_{flm}$ can be estimated as:

$$k1_{flm} = \frac{0.65 \cdot GR_{flm}}{1000} \sqrt{\frac{A_m}{A}} \qquad (8)$$

where GR_{flm} is the specified granularity value of the film, A_m is the area of the scanning aperture used to determine GR_{flm}, and A is the area of the scanning aperture to be used.

As shown later, the value of A can be approximated with 36μm² and the manufacturer's data sheet specifies A_m as 1810μm² [29]. Therefore the value of $k1_{flm}$ for Fujichrome film is 0.051. The factor 0.65 in Eq. 8 is an empirical constant that compensates for the deviation from linearity both due to the low number of grains contained within the very small scanning area and also due to the Schwarzschild-Villiger effect [34]. The Schwarzschild-Villiger effect reduces the expected density fluctuations when the area of illumination is larger than the area of scanning.

A further complication arises from the use of a relatively small scanning aperture which prevents that the filmgrain noise can be approximated with a 'white' Gaussian random field. The noise spectrum of the photographic film was determined experimentally, and a transfer function was calculated that produces realistic film-grain noise by filtering computer-generated white Gaussian noise. The noise properties of the filtered Gaussion noise reproduce the actual scanned filmgrain noise very well. A radial section of the isotropic MTF of this filter is shown in Fig. 9.

The equation for the model of the photographic film in its entirety is:

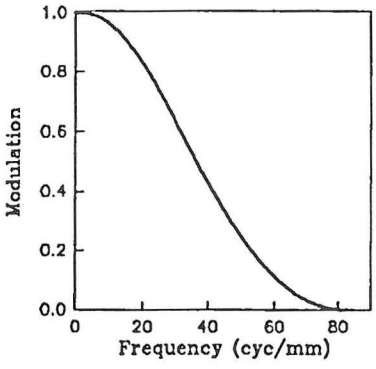


Figure 9. Radial section of the filter that is applied to computer-generated white noise in order to generate realistic film-grain noise.

$$d(x,y) = d_0(x,y) + \left(k1_{film} \cdot d_0(x,y)^{k2_{film}} \right) n_{film}(x,y)$$
$$d_0(x,y) = DE\left(\left(f(x,y) * h_{film}(x,y) \right) \cdot t \right) \quad (9)$$

where d(x,y) is the optical density of the film, $d_0(x,y)$ is an intermediate variable that is equal to the gross (noise-free) optical density of the film, $k1_{film}$, $k2_{film}$ are constants specific to the film, $n_{film}(x,y)$ is zero-mean, unit-variance non-white Gaussian noise source, DE(•) is the film-specific DE curve, f(x,y) is the input intensity distribution, t is the exposure time, and $h_{film}(x,y)$ is the effective PSF of the film.

4. Digitizer

Color fundus photographs were digitized with a very high resolution scanner (LS3500, Nikon Corp.). This scanner is capable of producing red, green and blue digital images by using respective filters in front of the white illumination light during digitization. The tapetal-like reflex is a gold/yellow/light-brown colored reflection on a red/brown background. Considering the three additive colors red, green and blue, it is the amount of added green that changes a color from reddish to yellowish. Therefore, only the green components of digital images were used as they display the highest contrast between the tapetal-like reflex and the background.

The scanner uses a stationary, fixed-focus, fixed-aperture light source and a stationary linear array of 4096 CCD detectors [35]. The photographic slide to be scanned is moved between the light source and the detector array with a stepper motor (see Fig. 10). The size of the unit step of movement is exactly equal to the distance between centers of consecutive CCDs on the linear array. The resulting digital image has square pixels corresponding to an area of 6 μm x 6 μm measured on the film plane.

CCDs respond linearly to the light exposure they receive. Unfortunately, each CCD may have a slightly different sensitivity as compared to other CCDs in the array. Especially at high signal levels, the variation in CCD sensitivities causes much more uncertainty than that caused by other sources of noise [36]. Fortunately, the CCD-to-CCD variations are stable and therefore they can be reduced to negligible levels by compensating (normalizing) the value each CCD produces according to a predetermined sensitivity map. Such a sensitivity compensation is not included in commercially supplied software, therefore custom software was written for this project that determines a sensitivity map (a gain and an offset for each pixel) and scans images using this sensitivity map.

The scanner digitizes a slide by measuring the transmitted proportion of the input light intensity over a fixed time (exposure time). The units of the numbers

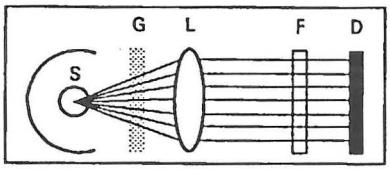


Figure 10. Schematic diagram of the scanner, (S) light source, (G) green filter, (L) lens, (F) film, (D) detector. The film moves into and out of the plane of the paper during scanning.

produced by the scanner (i.e. the value of each pixel in a transmittance domain digital image) will be called scanner-transmittance-units (stu).

There are three aspects of the relation between the spatial distribution of optical density on a film and the resulting digital image. These can be termed the 'uniform-density response,' the 'spatial-frequency response' and the 'noise.' To determine the uniform-density re-

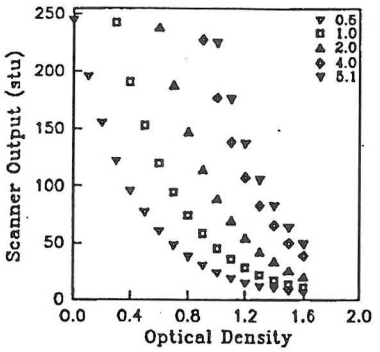


Figure 11.The output of the scanner versus uniform input densities for different values of rit.

sponse, the output of the scanner was measured when various spatially uniform neutral-density filters were used as input. The scanner output was averaged over a 100x100 pixel area to cancel the effects of noise. In addition, the neutral-density filters were located out of the plane of focus to assure uniformity throughout the scanned field. The exact optical density of the filters used in the experiment was determined with a calibrated radiometer (United Detector Technologies) and the light source located within the scanner. The only user adjustable parameter of the scanner is the relative integration time (rit). The rit parameter, which is a rational number between 25/50 and 255/50 (i.e. 0.5 to 5.1), scales the factory-set integration time of each CCD [35]. During the experiments five rit values (0.5,1, 2, 4 and 5.1) were used with the different combinations of uniform-density filters. Fig. 11 shows the results of these experiments. Clearly, there is an exponential relation between the uniform input density and the output digital image in stu. Such an exponential behavior is characteristic of CCD devices because their output is linearly related to the transmitted light intensity for a given exposure time. Of course, transmitted light intensity, in turn, is related exponentially to the optical density of the transparent media (see definition of optical density, Eq. 4). The mathematical relation for the uniform-density response of the scanner can be written as:

$$S_u = rit\,(\,k1_{scn} \cdot 10^{-d} + k2_{scn}\,) \qquad (10)$$

where S_u is the noise-free output of the scanner in the case of uniform density, rit is the relative integration time set by the user during scanning, $k1_{scn}$ and $k2_{scn}$ are scanner specific constants, and d is the uniform optical density of the object being scanned.

To determine $k1_{scn}$ and $k2_{scn}$, Eq. 10 can be put into a linear form:

$$s' = k1_{scn} \cdot T + k2_{scn} \qquad (11)$$

where $s'=s_u/rit$ is the normalized scanner output, and $T=10^{-d}$ is the transminance of the transparent media with optical density d.

The data shown in Fig. 11 was used to fit a least-squares line to a plot of s' versus T. The slope of the line ($k1_{scn}$) was 492.92 and the y-axis intercept of the line ($k2_{scn}$) was -3.38. Fig. 12 shows the same data as in Fig. 11 but with the estimates of the model (Eq. 10) superimposed as solid lines on the experimental data. Clearly, the model approximates the scanner extremely well in case of uniform noiseless input.

A scanner cannot respond to higher spatial-frequency inputs the same way as the uniform inputs mentioned above. First of all, the inherent integration of the light intensity (signal) over the sensing area of each CCD limits the ideal frequency response to sinc(πfw),

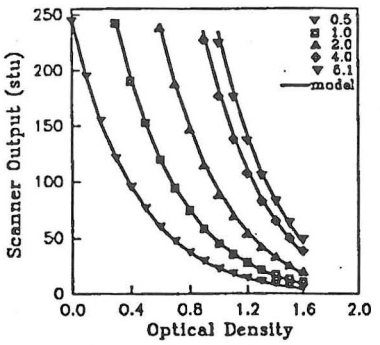


Figure 12.Comparison of model of the scanner with experimental data.

where f is the input spatial frequency in cyc/mm and w is the width of a CCD in mm [36]. In addition, scattered light in collection optics, scattered light from neighboring areas of the transparency, mechanical jitter, and charge diffusion between CCDs cause further blurring of the input. The scanner is approximated as an LSII low-

pass system. The spatial-frequency response of the scanner can be estimated using a sharp edge as input and obtaining the 'edge-response,' the digitized image of the sharp edge. The derivative of the edge-response is the 'line response' (also called 'line spread function' (LSF)). This function is the output of an LSI system to an input that has infinitesimal width and infinite length. The Fourier transform of a section of the LSF perpendicular to the edge is equal to the section of the system MTF perpendicular to the edge [37]. A large number of LSFs in all possible directions are necessary to reconstruct the two-dimensional MTF, unless the system of interest can be assumed to have isotropic response.

A razor blade glued onto a uniform neutral density filter was used to determine the LSF of the scanner in two orthogonal directions. Fig 13 shows sections of the scanner MTF in directions parallel and perpendicular to the CCD array. The fact that two sections are very

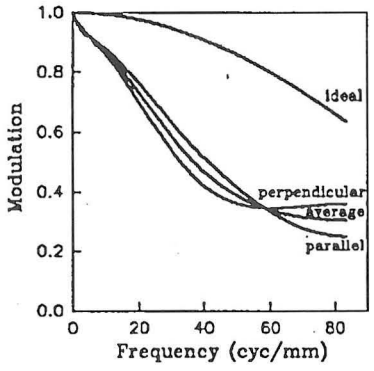


Figure 13.Sections of the scanner MTF in directions parallel and perpendicular to the detector array, their average, and the sinc function corresponding to a scanner made of ideal detectors.

similar reaffirms the validity of the assumption that the scanner has an approximately isotropic response. The average of the two sections (also shown in Fig. 13) was used as the isotropic MTF of the scanner. The noise-free scanner response to any input can be written symbolically as:

$$s_0(x,y) = (rit\ (k1_{scn} \cdot 10^{-d(x,y)} + k2_{scn})) * h_{scn}(x,y) \quad (12)$$

where $s_0(x,y)$ is the noise-free output of the scanner, $k1_{scn}$ and $k2_{scn}$ are scanner-specific constants, rit is the relative integration time, $h_{scn}(x,y)$ is the PSF of the scanner, and d(x,y) is the input film density.

Once the CCD-to-CCD sensitivity changes are normalized, the scanner output has two remaining noise sources [36]. First, there are all the signal independent noise sources which include trapping-state noise, reset noise, background noise, charge transfer noise, output amplifier noise and analog to digital conversion noise

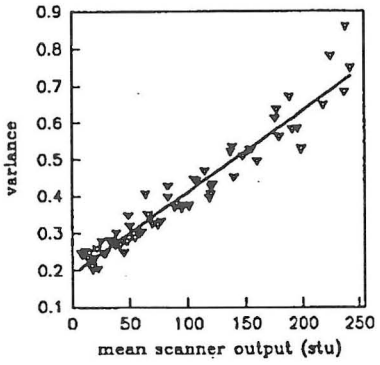


Figure 14.Least-squares fit of a line to a plot of noise variance versus mean scanner output.

(truncation noise). All these noise sources are lumped together and it is assumed that they are approximately Gaussian and additive. Second, there is shot noise (photon noise) which is dependent on the square root of the signal, additive and approximately Gaussian for intermediate to high exposure levels. The sum of two independent Gaussian random variables with zero mean is another Gaussian random variable with zero mean and a variance that is the sum of the individual variances. Therefore the total scanner noise magnitude can be written as:

$$\sigma_{scn} = \sqrt{k3_{scn} + k4_{scn} \cdot s_0} \quad (13)$$

where σ_{scn} is the standard deviation of the scanner noise, $k3_{scn}$ and $k4_{scn}$ are scanner specific constants, and s_0 is the noise-free output of the scanner.

To calculate the values of $k3_{scn}$ and $k4_{scn}$ it is observed from Eq. 13, that the variance of the scanner noise (i.e. standard deviation squared) is expected to be linearly related to the noisefree scanner output. Therefore, the noise-free scanner output is approximated with

a spatial average, and the variance and the mean of the digitized images of uniform density filters are experimentally measured. Fig. 14 shows both the variance versus mean data as well as a least-squares linear fit to the data. The values of $k3_{scn}$ (y-axis intercept of the line) and $k4_{scn}$ (slope of the line) were determined to be 0.1903 and 0.0022, respectively (correlation coefficient was 0.966).

According to the analysis in this section, the scanner model in its entirety is:

$$\begin{aligned} s(x,y) &= s_0(x,y) + \sqrt{k3_{scn} + k4_{scn} \cdot s_0(x,y)}\; n_{scn}(x,y) \\ s_0(x,y) &= \left(rit \left(k1_{scn} \cdot 10^{-d(x,y)} + k2_{scn} \right) \right) * h_{scn}(x,y) \end{aligned} \tag{14}$$

where s(x,y) is the digital image produced by the scanner, $s_0(x,y)$ is an intermediate variable equal to the noise-free scanner output, $k1_{scn}$, $k2_{scn}$, $k3_{scn}$ and $k4_{scn}$ are constants specific to the scanner, $n_{scn}(x,y)$ is zero-mean, unit-variance white Gaussian noise source, rit is the relative integration time set by the user during scanning, d(x,y) is the input optical density of the film, and $h_{scn}(x,y)$ is the PSF of the scanner.

5. Complete Imaging Model

In order to obtain and validate a single comprehensive model of the imaging system that relates the values of the digital fundus image to the actual luminance of the fundus, the models of the individual components need to be combined.

The combination of the individual component models (i.e., Eqs. 1, 3, 9, and 14) in a block diagram is shown in Fig. 15. Some properties of the imaging system can be seen easier in the block diagram than in the equations. Especially, two nonlinear transforms are seen to separate the three domains. Also, the complicated nature of the two noise sources, being non-linearly signal-dependent, becomes obvious. A helpful observation is the existence of three consecutive convolutions with the transfer functions of the eye, the camera and the film in the intensity domain. Because of the associative property of the convolution operation, a cascade of convolutions with different PSFs can be replaced by a single convolution with an effective PSF:

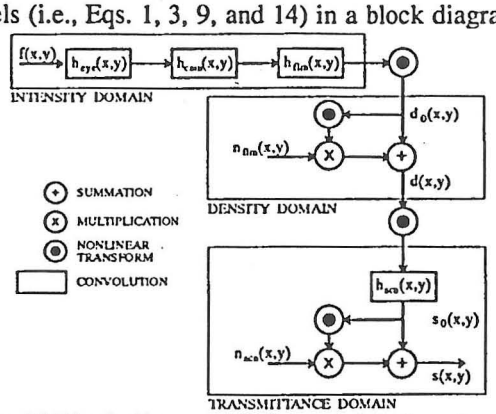


Figure 15. Block diagram of the complete imaging system

$$\begin{aligned} &((f(x,y) * h_{eye}(x,y)) * h_{cam}(x,y)) * h_{film}(x,y) \equiv \\ &\quad \equiv f(x,y) * (h_{eye}(x,y) * h_{cam}(x,y) * h_{film}(x,y)) \\ &\quad \equiv f(x,y) * h_{int}(x,y) \end{aligned} \tag{15}$$

where $h_{int}(x,y)$ is the effective intensity domain PSF which is the convolution of the three individual PSFs. The equivalent MTF is shown in Fig. 16.

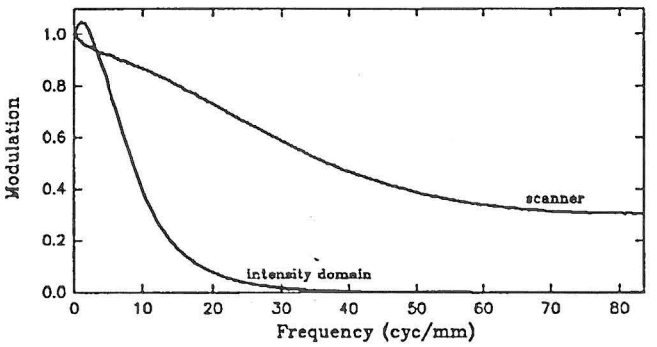
The equations describing the input-output relationship of the complete imaging model are:

$$\begin{aligned} s(x,y) &= s_0(x,y) + \sqrt{k3_{scn} + k4_{scn} \cdot s_0(x,y)}\; n_{scn}(x,y) \\ s_0(x,y) &= (rit (k1_{scn} \cdot 10^{-d(x,y)} + k2_{scn})) * h_{scn}(x,y) \\ d(x,y) &= d_0(x,y) + (k1_{film} \cdot d_0(x,y)^{k2_{film}})\, n_{film}(x,y) \\ d_0(x,y) &= DE ((f(x,y) * h_{int}(x,y)) \cdot t) \end{aligned} \tag{16}$$

where s(x,y) is the digital image, d(x,y) is the optical density of the film, f(x,y) is the luminance of the fundus, t is the exposure time used during photography, $s_0(x,y)$ and $d_0(x,y)$ are intermediate variables representing the noiseless scanner output and the noiseless film density, respectively, $n_{scn}(x,y)$ and $n_{film}(x,y)$ are zero-mean, unit-variance Gaussian noise sources, $h_{scn}(x,y)$ and $h_{int}(x,y)$ are the scanner PSF and the effective intensity domain PSF, DE(•) is the DE curve of the film, rit is the relative integration time set by the user during scanning, $k1_{scn}$, $k2_{scn}$, $k3_{scn}$ and $k4_{scn}$ are scanner specific constants equal to 492.92, -3.38, 0.1903, and 0.0022, respectively, and $k1_{film}$ and $k2_{film}$ are film specific constants equal to 0.051 and 0.5, respectively.

Verification

The equations (Eq. 16) modeling the input-output relationship of the digital fundus imaging system were developed using a series of reasonable assumptions (e.g. LSII systems), a series of experiments (e.g. scanner noise, scanner blur), some data taken from published reports (e.g. MTF of the eye), and some data extrapolated from published reports (e.g. film-grain noise). Naturally, the question arises, how closely the model approximates reality. If the human eye were not an integral part of the imaging system, such a question could be answered by imaging a perfect object (e.g. an edge) and comparing the actual image with the one that would be predicted by the model. Unfortunately, a perfect object does not exist in a biological structure



Figurr 16.The effective intensity domain MTF of eye, camera and film as compared to the scanner MTF.

like the eye. A good physical model of the eye that could be used for validation does not exist either. Therefore, only those parts of the imaging system model which do not include the eye will be verified. At the end of this section, the assumed validity of the eye model will be commented on.

As a first step to validate the model, the magnitude of the total noise in the system for a uniform input luminance was compared with the magnitude of the total noise predicted by the model. A uniform input to the system makes the output (and thus the noise measurements) independent of blurring due to eye, camera, and film (see Fig. 15). Therefore, a total of 256 64x64 pixel regions on 16 slides, each with a different uniform optical density, were digitized. The mean and the standard deviation of the normalized scanner output in stu are plotted (small triangles) in Fig. 17. Each distinct cluster of points represents many areas digitized from the same slide.

The response of the model to many uniform input luminance levels was determined using Eq. 16. Specifically, a uniform retinal luminance in the range 0.025 lx. to 0.375 lx., and an exposure time of 1 s was used. All other constants and nonlinear functions were according to the complete model. Pseudo-random number generators were used to generate the two white Gaussian noise sources using the Box-Mueller method [38, p.216]. One of the white Gaussian noise fields was filtered using the transfer function shown in Fig. 9, rescaled to have zero-mean and unitvariance, and then used as simulated film-grain noise. The relationship between the means and the standard deviations of the resulting images is plotted as a solid line in Fig. 17. It can be seen that the total noise estimate of the model corresponds closely to the measured total noise of the imaging system.

Probably the weakest assumption in the imaging model is the one about the blurring introduced by the fundus camera. A single -parameter analytical function was assumed to approximate the isotropic PSF of the fundus camera, and the parameter of this function was adjusted to force a certain definition of resolution. To verify the validery of the camera PSF, an edge image (a white paper glued onto a black cardboard) was photographed directly with the fundus camera. The edge was located about 3.5 m in front of the camera to assure the image of a sharp reflective step. External flash lighting was used and the internal camera flash was disabled. Three 64x64 pixel regions were scanned from one of the resulting slides. These three 'step responses' include degradations due to all imaging components except the eye. To compare the actual degradations to degradations that would be predicted by the imaging model, three digital luminance edges were simulated. Next, the complete imaging model (except the blurring due to the eye) was applied to the simulated edges. The slopes and shapes of the transition regions were compared. Just like the two representative profile plots shown Fig. 18, all

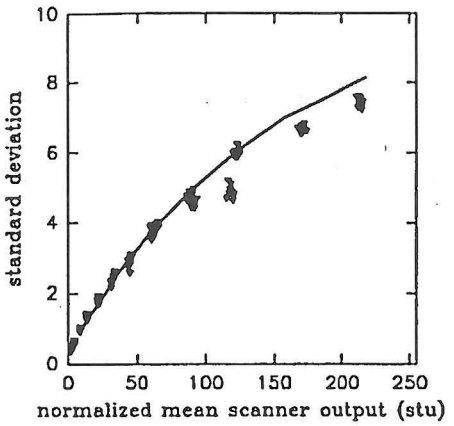


Figure 17.Measured total system noise (triangles) and the predicted total noise (solid line) versus the normalized mean scanner output.

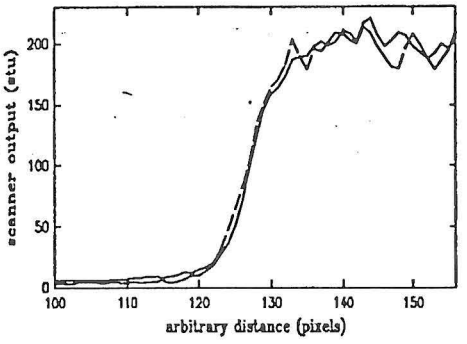


Figure 18.Comparison of profiles through the edge response of the actual imaging system (solid line) to the edge response predicted by the model (dashed line).

the comparisons showed almost identical responses between the actual imaging system and its model.

In summary, it may be stated that within the limited experimental conditions the model has accurately predicted the output of the actual imaging system. Assuming the unvalidated approximations about the blurring introduced by the eye are reasonable, the model is a very complete and realistic model of a digital fundus imaging system.

6. Simulation

The practical relevance of the model for the fundus imaging system - in addition to the possibility to perform an image restoration - is, that it provides answers to questions like: how sharp a change in retinal luminance will appear in a fundus photograph, how small a change in retinal luminance can be detected, and how small a structure of a given luminance can be detected. In order to demonstrate the extent of the degradations caused by the digital fundus imaging system, retinal luminance functions have been simulated. The overall effects of signal dependent noise, nonlinear response and blurring of sharp transitions are shown with the use of a simulated multi-step luminance function. The steps are vertically oriented and the flat regions of the steps were 0.05, 0.15, 0.25, 0.35, and 0.45 lx. The simulated image was degraded according to Eq. 16. Degradation was also performed with a modified imaging model where both noise sources (film and scanner) were set to zero. The simulated luminance functions (profile plots across the center of each image) are shown in Fig. 19.

The nonlinearities introduced by the system are demonstrated by decreasing step sizes in the degraded images. The signal dependence of the noise is also clearly visible as the higher luminance steps have higher noise fluctuations as compared to low luminance areas. The blurring is apparent by the smoothed transitions between steps.

RESTORATION

In order to analyze fundus photographs quantitatively at very high resolution, an image restoration method has to be developed that will invert the most

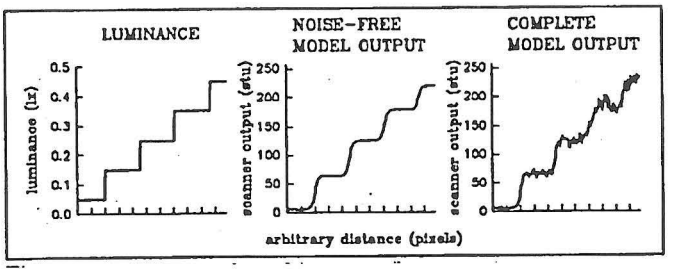


Figure 19.Simulated multi-step reflectance function, and its noiseless and noisy degraded versions (top), graphs of pixel intensities across the center of each image (bottom).

objectionable information degradations caused by the imaging system. Over the past 20 years, many different image restoration methods for various applications have been published based on different assumptions for the degradations caused by the imaging system and the sources and magnitudes of noise. For example, Hunt [39] assumed a non-linear, space-invariant imaging system with signal-independent noise to restore a nonstationary mean, stationary variance scene model with maximum-a-posteriori (MAP) error criterion. Kuan et al. [40] assumed a blurless imaging system with signal-dependent noise, a nonstationary mean, nonstationary variance scene model and attempted restoration using a local linear minimum mean square error (LLMMSE) criterion. On the other hand, most authors [e.g. 41-43] assumed a linear imaging system with additive signal-independent Gaussian noise and a wide-sense stationary scene model with various error measures for their restoration attempts. The common use of the simpler imaging models is due to the immense increase in complexity with nonlinear imaging and/or signal-dependent noise models. Interestingly, the increase in complexity does not usually result in a significantly better restoration. Therefore, the imaging model developed in the previous section will be analyzed as to whether it can be simplified to a generalized linear model with signal-independent noise.

1. Simplified Imaging Model

A linear, spatially-invariant system with additive, signal-independent Gaussian noise can be represented in symbolic form as:

$$g(x,y) = f(x,y) * h(x,y) + \sigma\, n(x,y) \qquad (17)$$

where g(x,y) is the output of the system, f(x,y) is the input, h(x,y) is the PSF, n(x,y) is zero-mean unit-variance Gaussian noise source, and σ is the standard deviation of the noise. It is usually more realistic and more powerful to consider a 'generalized linear system' which includes some point nonlinearities:

$$g(x,y) = A(B(f(x,y)) * h(x,y) + \sigma\, n(x,y)) \qquad (18)$$

where A(•) and B(•) are arbitrary point nonlinearities, that are, at least numerically, invertible. A point nonlinearity is a nonlinear function applied to the value of each pixel independent of other pixels.

The approximation of the complete imaging model (Eq. 16) with a generalized linear imaging model (Eq. 18) is not a simple task because of two major difficulties: first, the complete imaging model has two distinct convolutions (h_{int}, h_{scn}) in two domains separated by two nonlinear transforms; and second, the complete imaging model has two noise sources (n_{flm}, n_{scn}) that are signal-dependent, and are separated by a nonlinear transform.

As a first step in the simplification process, the influences of the two independent noise sources are compared to decide whether one of them might be negligible compared to the other. To compare the effects of the two noise sources, three sets of simulated digital images are generated: the first set contains just the total noise, the second set contains the film-grain noise only, and the third set contains the scanner noise only. In each set, 16 uniform intensity images are synthesized with the intensity being set in the range of 0.025 lx to 0.375 lx. The 'total noise' images include all parameters according to the complete model in Eq. 16. The 'film-grain noise only' images have $k3_{scn}$ and $k4_{scn}$ set equal to zero, and the 'scanner noise only' images have $k1_{flm}$ equal to zero.

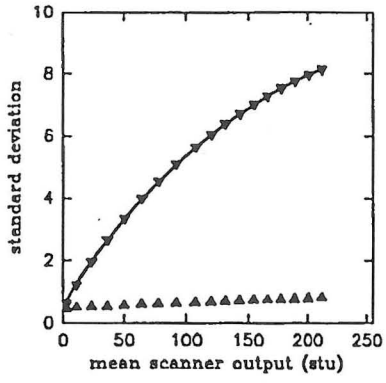


Figure 20. Comparison of film-grain noise (down triangle) and scanner noise (up truangle) magnitudes with total noise (solid line) magnitude in simulated uniform intensity images.

The standard deviation and the mean of the three sets of simulated images are calculated, and shown in Fig. 20. Clearly, the major portion of the total noise is due to the film grain noise. The effect of the scanner noise can be neglected without introducing a major error into the model.

As a second step in the simplification process, it is tested whether one of the two blurring functions may be neglected. According to the complete imaging model there are two independent linear blurs in two separate domains: the effective intensity domain blurring (combined effects of eye, camera, and film), and the scanner blurring acting in the transmittance domain. Fig. 16 shows that the scanner acts almost like an ideal device with negligible blurring because the 'signal' has already been low-pass filtered by the effective intensity domain MTF. Unfortunately, the blurring of the scanner does affect the imaging system output in a significant way that might not be obvious. In stark contrast to the 'signal,' the film-grain noise is broad band with significant high-frequency power. Therefore, the blurring of the scanner effectively reduces the total noise at the output of the imaging system. If the blurring due to the scanner is going to be neglected, then the magnitude of the film-grain noise needs to be appropriately reduced so that the model remains valid. The optimal reduction in the value of $k1_{flm}$ was determined using simulated uniform intensity images (the reduced $k1_{flm}$ is 0.028).

As a third step in the simplification process, the signal dependency of film-grain noise may be neglected. This can be done if the image is assumed to have low contrast, and therefore, the local density dependence of the film-grain noise is approximated by an average density dependence:

$$d_0(x,y)^{k2_{flm}} \approx \bar{d}^{\,k2_{flm}} \qquad (19)$$

where $d_0(x,y)$ is the gross film density, $\bar{d}$ is the average optical density of the low contrast film, and $k2_{flm}$ is a film-grain noise constant defined earlier. This approximation makes the film grain noise dependent on the specific image under consideration, but independent of the spatial variations of the signal level within the image.

The progress of simplifications achieved up to this point can be summarized as: first, the scanner is an ideal device with no noise and no blurring; and second, the image has low contrast and therefore the film-grain noise statistics are stationary acoss a given image, although they might differ from image to image. The imaging model incorporating these assumptions is:

$$\begin{aligned} s(x,y) &= rit\,(\,k1_{scn} \cdot 10^{-d(x,y)} + k2_{scn}\,) \\ d(x,y) &= DE\,(\,f(x,y) * h_{int}(x,y)\,) + \sigma\, n(x,y) \qquad (20) \\ \sigma &= k1_{flm} \cdot \bar{d}^{\,k2_{flm}} \end{aligned}$$

where the exposure time t has been omitted for clarity, $\bar{d}$ is the average optical density of the film, and all other symbols are the same as in Eq. 16. In order to show how the simplified model of Eq. 20 compares to the complete model of Eq. 16 the luminance image of a simulated tapetal-like reflex was used as input and the predicted digital fundus image was calculated using the two models. The difference between the two calculated images had a mean of -0.49 stu and a standard deviation of 3.62 stu. Two corresponding profiles and their difference are shown in Fig. 21. Considerable simplifications had a relatively small effect on the validity of the model.

The imaging model has been simplified considerably, but one more simplification step is necessary. The relation between s(x,y) and d(x,y) (first line in Eq. 20) is a memoryless deterministic one-to-one mapping with a trivial inverse, i.e., if s(x,y) is known then d(x,y) can be calculated, and vice versa. The reason why Eq. 20 is not a generalized linear imaging model is that the blurring, $h_{int}(x,y)$, and the additive noise, n(x,y), are in two different domains (intensity and density, respectively) that are separated by a nonlinear transform, DE(•). There are two realistic simplification paths: either the additive noise in the density domain is replaced by effective additive noise in the intensity domain, or the blurring in the intensity domain is replaced with an effective blurring in the density domain. Both approxi-

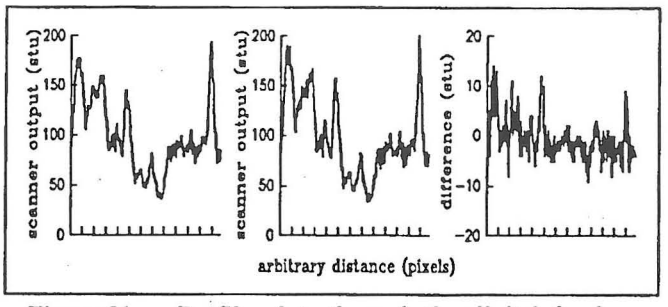


Figure 21. Profile plots through the digital fundus images of the simulated tapetal-like reflex using complete (left) and simplified (middle) models, and a plot of the difference (right).

mations result in generalized linear imaging models.

Approximation #1:

The fundus luminance has already been assumed to have low contrast. Now, it is further assumed that the noise term in Eq. 20 is much smaller than the signal term. Thus, the nonlinear DE curve is linearized about the mean density, and the second line of Eq. 20 is approximated with:

$$d(x,y) \approx DE\,(\, f(x,y) * h_{ins}(x,y) + \sigma_{eff}\, n(x,y)\,) \qquad (21)$$

$$\sigma_{eff} = \left\| \frac{DE^{-1}(\bar{d} + \sigma) - DE^{-1}(\bar{d} - \sigma)}{2} \right\|$$

where $\| \; \|$ is the absolute value operator. In Eq. 21, σ_{eff} is the standard deviation of the effective noise in the intensity domain and $DE^{-1}(\bullet)$ is the (numerical) inverse of the nonlinear DE curve.

Approximation #2:

In this case, the effect of blurring in the intensity domain is approximated with an effective blur in the density domain. Thus the second line of Eq. 20 becomes:

$$d(x,y) \approx DE\,(\, f(x,y)\,) * h_{eff}(x,y) + \sigma\, n(x,y) \qquad (22)$$

where $h_{eff}(x,y)$ is the effective PSF in the density domain, and σ is the same as in Eq. 20. To calculate the $h_{eff}(x,y)$, simulated edge functions of differing mean and contrast (in the intensity domain) were used. For each edge image, first the noiseless density domain image was determined, and then the effective PSF was calculated in the density domain. The resulting PSFs for different edge functions were very similar, and therefore, their average was used as $h_{eff}(x,y)$.

To compare the two possibilities for simplification, a luminance image of the simulated tapetal-like reflex (the same that was used to obtain Fig. 21) was used as input to the complete model and to the two simplified models. The results are interesting. The difference images have means of 0.7 and 8.7 stu, and they have standard deviations of 13.2 and 7.0 stu, respectively for the simplifications #1 and #2. The difference image for simplification #1 looks more like random noise with a large variance, suggesting that the change of the noise source from density domain to intensity domain was not a valid simplifying approximation. The errors caused by simplification #2 are structured with a smaller variance but a larger bias. This can easily be explained by the use of the blurring function in the density domain, which is a compressed domain as compared to the original intensity domain. Fig. 22 shows representative profile plots through the complete and simplified models as well as the difference graphs.

For this study, simplification #2 is chosen as appropriate because it shows a much smaller error variance and because a bias in peak amplitudes is much more acceptable than a random error changing the noise properties.

The simplified imaging model consists of a generalized linear system with space-invariant blurring and additive signal-independent noise. The effective signal-independent noise is a function of the average film density d, which needs to be estimated from the degraded image. With reference to the previous section, the following observations are made: first, both the scanner noise and the film grain noise have zero mean, therefore a spatial average of the image in any domain can be assumed not to be affected by noise sources. Second, both the scanner blurring and the optical blurring are 'passive,' therefore they neither increase nor

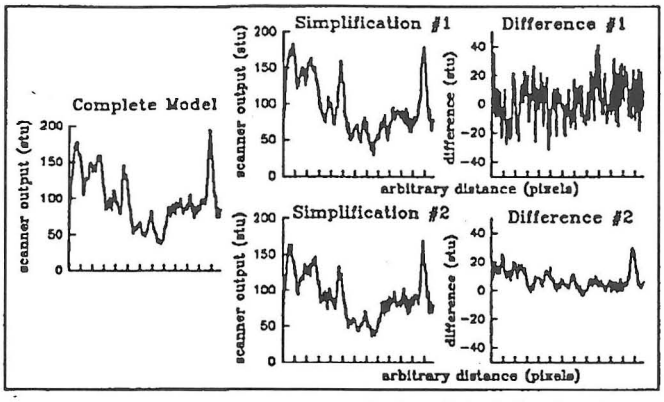


Figure 22.Profile plots through the digital fundus images of a simulated tapetal-like reflex using the complete and the two simplified models, and plots of the differences.

decrease the total (or average) signal power in any domain. Accordingly, with trivial manipulation it can be shown that the average optical density of the image is approximately related to the average value of the degraded digital image according to the following formula:

$$\bar{d} \approx \log_{10}\left(\frac{k1_{scn}}{\frac{\bar{s}}{rit} - k2_{scn}}\right) \qquad (23)$$

where $\bar{s}$ is the average value of the digital image in the transmittance domain, $\bar{d}$ is the average value in the density domain, and $k1_{scn}$, $k2_{scn}$ and rit are the same as in Eq. 16.

The specific procedure to be used for the restoration of fundus images can be summarized as follows. The mean, $\bar{s}$, of the digital image, s(x,y), is calculated first. Next, the average optical density, $\bar{d}$, is estimated using Eq. 23 and the noise standard deviation, σ is calculated using the approximation in Eq. 20. The degraded density domain image, g(x,y), is calculated according to the inverse of the scanner nonlinearity which is:

$$g(x,y) = \log_{10}\left(\frac{k1_{scn}}{\frac{s(x,y)}{rit} - k2_{scn}}\right) \qquad (24)$$

The estimation of the ideal density image, DE(f(x,y)), from the degraded density image, g(x,y), assuming the transfer function $h_{eff}(x,y)$ and the noise σ are known, is the classical linear restoration problem, and it is the subject of the next section. After the estimation of the ideal density image the light intensity radiating from the fundus, f(x,y), may be calculated using the numerical inverse of the DE curve.

2. Regularized Linear Restoration

Image restoration refers to the estimation of the original scene from its degraded image [44]. Usually, an acceptable restoration cannot be accomplished without knowledge of the image formation process with its characteristic blurring, the statistical properties of the noise, a measure of disparity between original scene and its estimate, and some generalized knowledge about the original scene. Given a degraded image, the validity of the (sometimes implicit) assumptions made about these four points dictates the success of restoration algorithms.

In the previous section, the complete imaging model was simplified to a generalized linear system with a linear, space-invariant blurring and additive, signal-independent Gaussian noise. In this section the so-called 'regularized linear restoration' method, that is based on the simplified imaging model will be described. The 'linear' restoration is discussed first, and then, the concept is expanded to 'regularized' linear restoration.

Linear restoration refers to the determination of a linear filter to remove (or reduce) the degradations caused by the imaging system. The result is an estimate of the original (ideal) scene. Linear restoration is defined as:

$$\hat{f}(x,y) = h_L(x,y) * g(x,y) \qquad (25)$$

where $\hat{f}(x,y)$ is the estimate of the original scene, $h_L(x,y)$ is the linear restoration PSF, and g(x,y) is the degraded image.

In the search for a restoration PSF, probably the first one that comes to mind is the convolutional inverse of the imaging system PSF, defined as:

$$h^{-1}(x,y) * h(x,y) = \delta(x,y) \qquad (26)$$

where $h^{-1}(x,y)$ is the convolutional inverse of h(x,y) (and vice versa), and $\delta(x,y)$ is the discrete delta function (an impulse at coordinates 0,0). The so-called 'inverse filter' restoration is:

$$\hat{f}_{inv}(x,y) = h^{-1}(x,y) * g(x,y) \qquad (27)$$

Substituting the linear imaging equation (Eq. 17) for g(x,y):

$$\hat{f}_{inv}(x,y) = h^{-1}(x,y) * \{ f(x,y) * h(x,y) + \sigma\, n(x,y) \} \qquad (28)$$

and taking advantage of the distributive and commutative properties of the convolution operation, it can be shown that the inverse filter estimate is equivalent to:

$$\hat{f}_{inv}(x,y) = f(x,y) + h^{-1}(x,y) * \{ \sigma\, n(x,y) \} \qquad (29)$$

In other words, the inverse filtered image is equal to the original image plus inverse filtered noise. This suggests that if there is no noise, the inverse filter results in perfect restoration. Unfortunately, the inverse of the system PSF does not always exist. Even if it exists, it is rarely possible to calculate it accurately with finite precision computers. In addition, usually noise is present and inverse filtering can rarely produce acceptable restorations. Actually, the shortcomings of inverse filters can be bener appreciated with the frequency domain equivalent of Eq. 28:

$$\hat{F}_{inv}(u,v) = \frac{H(u,v)\,F(u,v) + \sigma\, N(u,v)}{H(u,v)} \qquad (30)$$

$$= F(u,v) + \sigma\, \frac{N(u,v)}{H(u,v)}$$

Clearly, the filter is not defined at the zeros of H(u,v). Even if a pseudo-inverse filter is used, which forces the inverse filter to zero where H(u,v) is zero, the information corresponding to the zeros of the H(u,v) is lost during imaging and cannot be recovered. In addition, at the values of H(u,v) which are much smaller than N(u,v), there will be magnification of noise that swamps the restored image. These are the reasons why restoration is called an "ill-posed inverse problem with no unique solution" [45].

Regularization [46,47,48] refers to a class of solution techniques which convert ill-posed inverse problems to well-posed inverse problems and choose a physically acceptable and meaningful solution among a family of solutions. Regularization concepts are usually based on incorporating a priori knowledge about the true solution into the restoration algorithm. Next, the regularized linear restoration method will be defined, where some of the previous equations are repeated to aid the reader.

Assuming a linear spatially-invariant imaging system with signal-independent Gaussian additive noise,

$$g(x,y) = f(x,y) * h(x,y) + \sigma\, n(x,y) \qquad (31)$$

the linear restoration estimate was defined as:

$$\hat{f}(x,y) = h_L(x,y) * g(x,y) \qquad (32)$$

If the restoration were successful, then the estimate, $\hat{f}(x,y)$, should be approximately equal to the original image f(x,y). But, the original image is not known and therefore the merit of the estimate cannot be verified by comparing it to the original image. On the other hand, if the estimate is approximately equal to the original image, then the sum of squared differences between the degraded image, g(x,y), and the blurred estimate, $\hat{f}(x,y)*h(x,y)$, should be approximately equal to the sum of squared noise terms. That is:

$$\sum_x \sum_y \{ g(x,y) - \hat{f}(x,y) * h(x,y) \}^2 \approx \sum_x \sum_y \{ \sigma\, n(x,y) \}^2 \qquad (33)$$

which is equivalent to:

$$\sum_x \sum_y \{ g(x,y) - \hat{f}(x,y) * h(x,y) \}^2 \approx \sigma^2 N^2 \qquad (34)$$

where the images are assumed to be square with NxN pixels. There are many linear filters $h_L(x,y)$ (including the inverse filter) that are solutions to Eq. 34, but most of them produce estimates with extensive magnification of high frequency noise. To enforce a physically acceptable restoration, the concept of regularization is introduced by choosing a solution that minimizes the 'rough-

ness' of the estimate [43]. The roughness is defined as the norm of the high-pass filtered estimate:

$$\Omega(\hat{f}) = \sum_x \sum_y \left(c(x,y) * \hat{f}(x,y) \right)^2 \quad (35)$$

where c(x,y) is the so-called regularizing operator that is a high-pass filter, and it is commonly chosen to be the discrete approximation to the Laplacian [41]. The minimization of Eq. 35 subject to the constraint of Eq. 34 can be accomplished with the method of Lagrange multipliers [49, p.220], and the problem reduces to the minimization of:

$$\Phi(\hat{f}) = \alpha \sum_x \sum_y \left(g(x,y) - \hat{f}(x,y) * h(x,y) \right)^2 + \sum_x \sum_y \left(\hat{f}(x,y) * c(x,y) \right)^2 \quad (36)$$

where α is chosen to satisfy Eq. 34. The minimization is straightforward [49, p.221] and the regularized linear restoration filter in the frequency domain is [44]:

$$H_R(u,v) = \frac{H^*(u,v)}{|H(u,v)|^2 + \gamma |C(u,v)|^2} \quad (37)$$

where $H_R(u,v)$ is the regularized linear restoration filter, H(u,v) is the transfer function of the imaging system, H*(u,v) is the complex conjugate of H(u,v), C(u,v) is the high-pass filter that regularizes the solution, and γ is equal to $1/\alpha$.

In summary, the regularized linear restoration estimate of the original image is:

$$\hat{f}(x,y) = FT^{-1} \left[\frac{H^*(u,v) \cdot G(u,v)}{|H(u,v)|^2 + \gamma |C(u,v)|^2} \right] \quad (38)$$

where $FT^{-1}[\bullet]$ is the inverse Fourier transform. Note that Eq. 38 and Eq. 34 are two simultaneous nonlinear equations where f(x,y) and γ are the unknown variables.

3. Implementation

The solution to the regularized linear restoration (Eq. 38) requires a parameter γ to be chosen to satisfy the constraint of Eq. 34. An iterative method must be used to solve the simultaneous equations. First, the approximate equality of the constraint equation is rewritten in terms of a tolerance:

$$-\varepsilon < \left(\phi(\gamma)^2 - \sigma^2 N^2 \right) < \varepsilon \quad (39)$$

where ε is the tolerance and $\phi(\gamma)^2$ is defined as:

$$\phi(\gamma)^2 = \sum_x \sum_y \left(g(x,y) - \hat{f}(x,y) * h(x,y) \right)^2 \quad (40)$$

The substitution of the solution of the estimate from Eq. 38 into Eq. 40 results in:

$$\phi(\gamma)^2 = \sum_x \sum_y \left(g(x,y) - FT^{-1}\left[\frac{|H(u,v)|^2 \cdot G(u,v)}{|H(u,v)|^2 + \gamma |C(u,v)|^2} \right] \right)^2 \quad (41)$$

which explicitly shows the dependence of $\phi(\gamma)^2$ on γ. Hunt [41] has shown that $\phi(\gamma)^2$ is a monotonic function of γ. Therefore, the nonlinear Eq. 39 can be solved for γ with the following algorithm:

1) Choose a tolerance value ε
2) Choose an initial value for γ
3) Compute $\phi(\gamma)^2$ using Eq. 41
4) If $\phi(\gamma)^2 < (\sigma^2 N^2 - \varepsilon)$ decrement γ, goto (3)
5) If $\phi(\gamma)^2 > (\sigma^2 N^2 + \varepsilon)$ increment γ, goto (3)
6) Else stop.

A Newton-Raphson-like algorithm [38, p.270] was used to speed convergence by adjusting the step size of γ according to the local gradient of $\phi(\gamma)^2$.

4. Results

Restorations of simulated and actual tapetal-like reflex images will be used to test the restoration algorithm. Simulated images, being based on known original images, will be used to show quantitatively the extent of restoration. Since the original (ideal) fundus luminance is not known, the restoration of actual tapetal-like reflex images can only be judged qualitatively.

The 'quality' or the 'information content' of simulated images will be measured using a generalized signal to noise ratio (SNR), which is defined as:

$$SNR = 10 \log_{10} \left(\frac{Var_f}{Var_{f-g}} \right) \qquad (42)$$

Var_f is the variance of the original image f(x,y), and Var_{f-g} is the variance of the difference between the degraded intensity domain image, g(x,y), and the original intensity domain image, f(x,y). It is important to point out that the SNR is not just a function of noise, but also a function of blurring and nonlinearities.

As a direct extension of the SNR definition, the improvement in signal to noise ratio (ISNR) is defined as a measure of restoration performance. ISNR is based on the difference of the SNRs of restored and degraded images, and is defined as:

$$ISNR = 10 \log_{10} \left(\frac{Var_{f-g}}{Var_{f-\hat{f}}} \right) \qquad (43)$$

where $Var_{f-\hat{f}}$ is the variance of the difference between the restored image and the original image.

Image information may be in one of three relevant domains: the intensity domain, the density domain, and the transmittance domain. The intensity domain is the representation of the light intensity originating from the retina, and therefore closely approximates the underlying biology. All of the analysis of the tapetal-like reflex will be based on intensity domain images; therefore, the results of the restoration will also be judged in this domain. The density domain is where the deconvolution part of the restoration will be accomplished, and the transmittance domain is the only data that is usually available in the case of actual fundus images.

The restoration performance on simulated images was determined for four different simulations (A,B,C and D) which used the exact same reflex and noise patterns only differing by reflex to background contrast (A,B and C) and the size of the image (A vs. D). Simulations A and D had reflex to background contrast representative of actual tapetal-like reflex, whereas simulations B and C had less contrast at lower (B) and higher (C) average exposure levels. Table 1 summarizes the restoration results for the four simulations.

In all four cases, restoration resulted in improvements in SNR. A region of the original, degraded and restored versions of simulation A is shown in Fig. 23. The profile plots show how the restoration process reduces the noise and increases the reflex-to-background contrast.

Simulation	Size	avg. exposure	SNR (dB)	ISNR (dB)
A	256x256	0.089	3.28	1.21
B	256x256	0.081	3.81	0.64
C	256x256	0.142	2.81	0.37
D	640x640	0.084	3.20	1.27

Table 1. Summary of restorations performed on simulated tapetal-like reflex images.

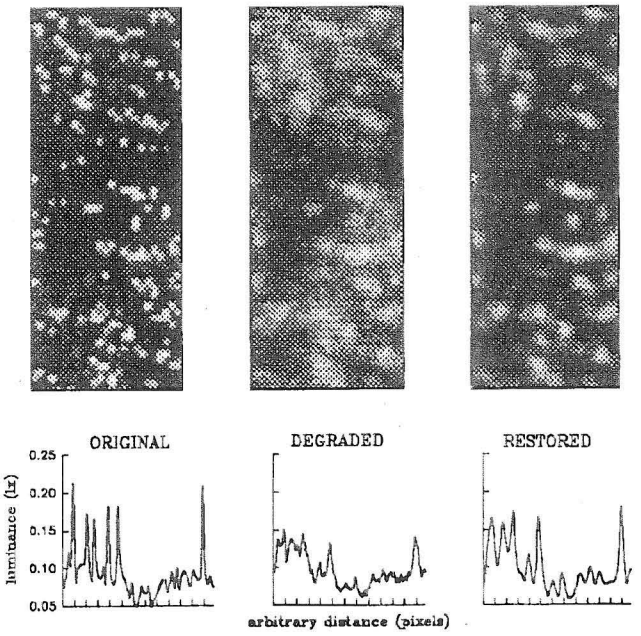


Figure 23. Simulated reflex luminance (top-left), degraded (top-middle) and restored (top-right) versions, and the corresponding profile plots (vertical cut through the image centers).

In the case of simulations, it is possible to judge the results of the restoration process both quantitatively and qualitatively, considering the original image is exactly known. In the case of actual fundus images the original fundus luminance is not known. Therefore, the only method to judge the performance of the restoration

process is to use human intuition to determine whether the results of the restoration are 'reasonable.'

Fig. 24 shows a small region of a real tapetal-like reflex image in the intensity domain, the result of the restoration, and profile plots through the centers of the respective images. This image is a representative sample among the images that were restored. In the restored image there is almost no noise in the traditional sense; i.e. high frequency variations of small amplitude. Furthermore, the high-intensity peaks corresponding to the reflexes of interest have increased contrast and exhibit sharper edges.

SEGMENTATION

Analysis of the tapetal-like reflex involves accumulation of statistics on the sizes and shapes of the individual reflexes, determination of several properties like the contrast against the retinal background, the geometric relationship of neighboring reflexes, and the change of the reflex properties over time. All of these analyses can only be accomplished if pixels belonging to the reflex can be differentiated from the pixels belonging to the background. Therefore, this section will give details of a novel method that is used to separate the pixels, i.e. to perform reliable image segmentation.

Although any one of the many known segmentation strategies (surveyed in [51]) could be adapted for the tapetal-like reflex analysis, a thresholding strategy is

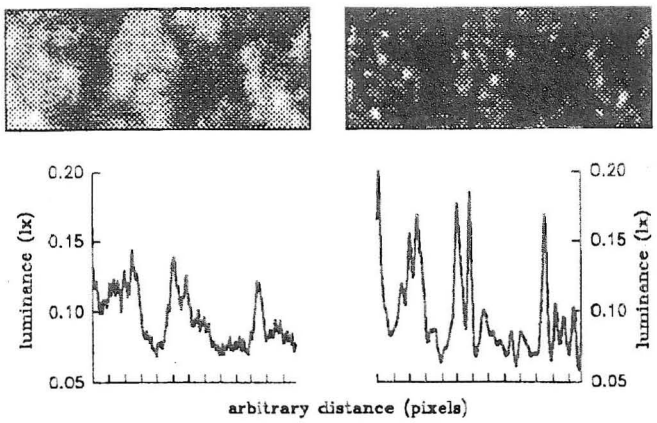


Figure 24. An actual tapetal-like reflex image (top,left), the corresponding restored image (top,right), and horizontal profile plots through image centers.

probably the most obvious, straightforward and simple solution. Classic measures of segmentation performance like percentage of misclassified pixels, distance of misclassified pixels or region uniformity [52] are not very useful in this application. Therefore, several figures of merit are defined that collectively specify the quality or reliability of segmentation results in the case where the ideal result is known. Both object and pixel based figures of merit will be used. The object based figures of merit are:

Measure of true positives:

$$FM_1 = 1 - \frac{n_{fo}}{n_{do}} \qquad (44)$$

where n_{fo} is the number of objects that have been detected falsely, and n_{do} is the total number of detected objects.

Measure of true negatives:

$$FM_2 = 1 - \frac{n_{mo}}{n_{ao}} \qquad (45)$$

where n_{mo} is the number of objects that were missed, and n_{ao} is the total number of actual objects.

Measure of object clustering:

$$FM_3 = \frac{n_{do} - n_{fo}}{n_{ao} - n_{mo}} \qquad (46)$$

The pixel based (classical) figures of merit are:

Measure of true positives:

$$FM_4 = \frac{n_{dp} - n_{fp}}{n_{ap}} \qquad (47)$$

where n_{dp} is the number of detected object pixels, n_{fp} is the number of falsely detected object pixels, and n_{ap} is the number of actual object pixels.

Measure of true negatives:

$$FM_5 = \frac{N - n_{dp} - n_{fp}}{N - n_{ap}} \qquad (48)$$

where n_{dp}, n_{fp} and n_{ap} are as defined above and N is the total number of pixels in the image.

All of the figures of merit are incorporated into a global measure as:

$$FMG = FM_1\, FM_2 \bullet FM_3 \bullet FM_4 \bullet FM_5 \qquad (49)$$

It can be seen by inspection that all figures of merit result in real numbers in the range from 0.0 to 1.0. In each case, 1.0 is the result of perfect segmentation, and a number close to 0.0 is 'bad' segmentation.

Global Thresholding

Global thresholding refers to the generation of a binary image g(x,y) from an input image f(x,y) using a threshold value T and the rule:

$$g(x,y) = \begin{cases} 1 & if \quad f(x,y) > T \\ 0 & if \quad f(x,y) \le T \end{cases} \qquad (50)$$

Assuming f(x,y) is the image of high-intensity objects on a low-intensity background, pixels belonging to objects will be labeled as 1 and pixels belonging to the background will be labeled 0.

Global thresholding used as a segmentation strategy works well for images in which object and background pixels have intensities grouped into two dominant modes. In that case, the threshold (T in Eq. 50) is chosen to separate the two modes. Unfortunately, most real images, including the tapetal-like reflex images, have objects with a range of intensities on a background which has a spatially varying intensity of its own. Imaging system degradations like blurring and noise cause further overlap between intensities of objects and the background. The resulting pixel intensity distributions (histograms) are not bimodal and it is impossible to find a single threshold that separates objects from the background. Therefore, the threshold must be allowed to vary across the image so that it can adjust to local image properties.

Local Thresholding

For local thresholding, the threshold is a spatially variant function:

$$g(x,y) = \begin{cases} 1 & if \quad f(x,y) > t(x,y) \\ 0 & if \quad f(x,y) \le t(x,y) \end{cases} \qquad (51)$$

where f(x,y) is the image to be thresholded, g(x,y) is the binary image which is the result of the thresholding, and t(x,y) is the spatially-variant threshold image, also called the 'threshold surface'.

Assuming there is a method of determining the threshold image, the advantages of local thresholding over global thresholding are clear. Unfortunately, there is no underlying theory of segmentation [51]; therefore, the literature is filled with ad hoc methods of threshold determination that seem to work well in their respective applications.

Most of the methods mentioned in literature [54-56] have in common that in the process of calculating a threshold for each pixel, the authors consider only a relatively small neighborhood around the pixel. The underlying hypothesis clearly is that it is much easier to distinguish between two classes of pixels within a small neighborhood as compared to the whole image. It appears that the result of local thresholding is considerably affected by the 'neighborhood size' that is either explicitly stated or implied. It is much less affected by the specific algorithm used to calculate the threshold. Therefore, a very simple property is chosen for the calculation of the local threshold and efforts are concentrated in analyzing the effect of the neighborhood size.

The local threshold t(x,y) is defined as the average value of the pixel intensities within a moving square window of size (2m+1) pixels, where the parameter 'm' is the 'neighborhood size.' Symbolically, the local threshold is:

$$t_m(x,y) = \frac{1}{(2m+1)^2} \sum_{x-m}^{x+m} \sum_{y-m}^{y+m} f(x,y) \qquad (52)$$

where $t_m(x,y)$ is the spatially-variant threshold that depends on the neighborhood size m. The threshold image is only defined for x and y greater than (m-l) and less

than (N-m) for an NxN pixel image.

In Fig. 25 the local thresholding of a simulated tapetal-like reflex image (restored image of which a part is shown in Fig. 23) is shown with neighborhood sizes of 2, 10, 18, and 24 pixels, and in Table 2, the figures of merit are listed for a wider range of neighborhood sizes.

Two important observations can be made both visually from Fig. 25 and quantitatively from Table 2. First, clustering of many objects occurs as the neighborhood size is increased (FM_3 decreases with larger neighborhood sizes) which suggests that small neighborhoods are better. Second, the number of false objects detected decreases as the neighborhood size is increased (FM_1 and FM_4 increase with larger neighborhood sizes), which suggests that large neighborhoods are better. Interestingly, FMG values show that there is no distinct 'compromise' neighborhood size between the two extremes.

It is concluded that, in order to have a reliable local thresholding method, a single neighborhood size cannot be used, instead the useful features of multiple neighborhood sizes should be combined.

Multi-Scale Thresholding

In the field of image enhancement, some authors [57-61] use an arbitrary function to find the 'optimal' neighborhood size for each pixel in an image, and then they enhance the image contrast locally using the 'optimal' neighborhood of each pixel. In another example [62], the authors determine a window size for each pixel according to the 'signal activity', and they smooth the noise without blurring the edges. In yet another example [63], the authors use a fuzzy logic approach to determine the size of the "contextual region" for each pixel, and they perform histogram equalization within these regions. Although all of these approaches are arbitrary and application-specific, there is a common underlying framework: scale-space theory.

Simply stated, scale-space of an image refers to a family of images derived from the original by progressively removing finer details [64]. Many authors have tried so-called 'multi-resolution segmentation' techniques which attempt pixel groupings by examining members of the scale-space of an image. An interesting and recent example is given by Lifshitz and Pizer [65], where each pixel in an image is associated with a local intensity extremum (maximum or minimum) at a given scale. The paths of the extremum pixels are followed while moving from fine to coarse scales. The amount of blurring necessary for an extremum point to disappear within an enclosing region is a measure of the scale of the extremal region. In simpler terms, the simultaneous consideration of the image at multiple scales allows sensible segmentation at a given scale. The same idea is used to define a novel thresholding method based on consideration of fine and coarse versions of an image.

Multi-scale thresholding is based on two inherent properties of the scale-space. The first property is that *at a coarse scale* 'significant' structures of an image clearly stand out as compared to 'insignificant' structures like artifacts and noise. In contrast, the localization of the structure edges (boundaries) and the separation of groups of small structures is not reliable. The second property is that *at a fine scale* successful local-

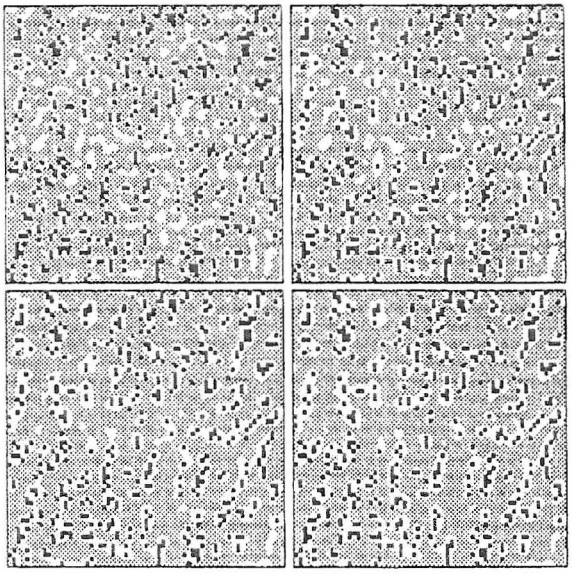


Figure 25. Demonstration of the effects of neighborhood size on local thresholding; m=2 (top,left), m=10 (top,right), m=18 (bottom, left), and m=24 (bottom, right). The original reflexes are superimposed in black on the thresholding results which are white for pixels classified as reflexes and grey for the background.

m (pixels)	FM_1	FM_2	FM_3	FM_4	FM_5	FMG
2 *	0.70	0.94	0.40	0.87	0.71	0.16
4	0.72	0.94	0.39	0.89	0.70	0.17
6	0.74	0.94	0.38	0.90	0.70	0.16
8	0.72	0.95	0.34	0.91	0.69	0.15
10 *	0.73	0.94	0.28	0.92	0.69	0.12
12	0.79	0.93	0.28	0.92	0.69	0.13
14	0.79	0.94	0.26	0.93	0.69	0.12
16	0.84	0.95	0.23	0.94	0.70	0.12
18 *	0.85	0.96	0.22	0.94	0.70	0.12
20	0.85	0.97	0.22	0.95	0.70	0.12
22	0.88	0.97	0.23	0.95	0.70	0.13
24 *	0.86	0.97	0.23	0.95	0.70	0.13

Table 2. The six figures of merit as a function of the neighborhood size 'm' of local thresholding. The entries marked with a '*' are shown in Fig. 25.

ization of the structure edges and separation of groups of small structures is possible, but 'insignificant' structures might be difficult to distinguish from 'significant' structures.

Based on the above two properties, a heuristic generalized segmentation approach is defined as "find the significant structures in a coarse scale and then refine them at a fine scale." The specific implementation of this approach is accomplished by thresholding the image first with a large neighborhood size, and then with a small neighborhood size. The two thresholding results are logically 'AND'-ed to give the final segmentation result. In other words, a pixel is considered to belong to an object, if and only if, it belongs to an object in coarse and fine scales. It is straightforward to show that the logical 'AND' of the two thresholding results is equivalent to using a threshold image that is the maximum of the two threshold images:

$$t_{m1,\,m2}(x,y) = MAX\,[\,t_{m1}(x,y),\ t_{m2}(x,y)] \qquad (53)$$

m1, m2	FM1	FM2	FM3	FM4	FM5	FMG
2, 22	0.91	0.91	0.44	0.84	0.83	0.25

Table 3. The figures of merit of segmentation with the multi-scale thresholding shown in Fig. 26.

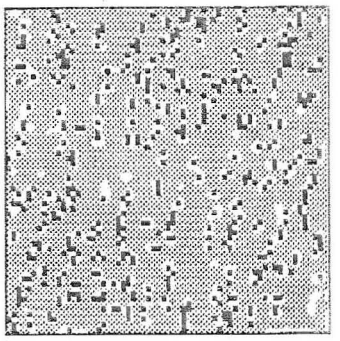


Figure 26. Multi-scale thresholding of the restored image of the simulated tapetal-like reflex (m1=2, m2=22)

where MAX{•} is an operator that is equal to the larger one of its arguments, $t_{m1}(x,y)$ and $t_{m2}(x,y)$ are two threshold images with the neighborhood sizes of m1 and m2, respectively, and $t_{m1,m2}(x,y)$ is the threshold image for the multi-scale thresholding method.

Fig. 26 shows the graphical results and Table 3 specifies the figures of merit for applying the multi-scale thresholding (with m1=2, m2=22) to the restored, simulated image of which a part is shown in Fig. 23.

Both qualitatively and quantitatively, the results of multi-scale thresholding are a considerable improvement over local thresholding with either one of the neighborhood sizes. The number of falsely detected objects and the number of missed objects are reduced to negligible levels. There is still some clustering of tiny objects but that is expected because of the effects of restoration. Some readers might object to the extremely simplistic method of threshold determination (average value) that was used, and wonder if one could do better with another thresholding method. Next, such a choice will be considered. It is important to point out that the multi-scale strategy is not changing, just a change in the method of determining the threshold value within a neighborhood size is being considered.

Most of the popular thresholding methods assume that the pixel values originate from two normally distributed populations, and attempt to fit the sum of two Gaussian distributions to the observed histogram [49, p.360]. Four parameters of the two Gaussians have to be estimated iteratively. In addition to the sometimes un-

stable nonlinear minimization of error, the assumed Gaussian distributions are usually not realistic. A much better approach is a method published by Otsu [66,67]. In this method, the 'between class variance' is calculated for every possible value of the threshold, and the one causing maximum separation between the resulting class of pixels is chosen. The 'between class variance' as a function of the threshold is defined as:

$$\sigma^2_{BC}(t) = n_1 n_2 (\mu_1 - \mu_2)^2 \qquad (54)$$

where σ_{bc}^2 is the between class variance, t is the threshold, n_1 is the number of pixels with intensity less than or equal to t, n_2 is the number of pixels with intensity greater than t, and μ_1 and μ_2 are the means of the two classes generated by the threshold.

The multi-scale thresholding method was implemented using Otsu's method of threshold determination instead of the 'average value' threshold. Fig.27 shows the result of the segmentation and Table 4 shows the corresponding figures of merit. The same image was used as for the 'average value' method in Fig. 26 and Table 3.

Comparing Figures 26 and 27, it is observed that Otsu's method produces segmentation that is more 'tight' (i.e., less clustering, higher value of FM_3). But, it also misses considerably more objects (lower value of FM_2). Overall, Otsu's method does not really result in a better segmentation than the simplistic 'average value' method. Considering the much increased computational requirements of Otsu's method, it is preferable to use the simplistic 'average value' threshold.

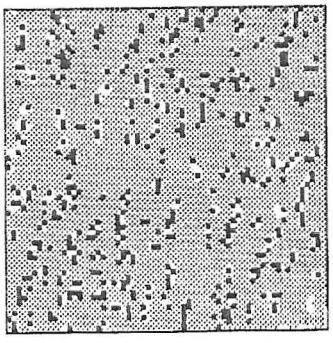


Figure 27. Multi-scale thresholding of the restored tapetal-like simulation using Otsu's method (m1=2, m2=22).

There is currently no means of independently validating the results of the multi-scale segmentation in real reflex images. Therefore, the segmentation results were judged by human intuition, as to their ability to reasonably represent what seems to be the bright objects on a darker background. As the representative example in Fig. 28 shows, the segmentation results do correspond closely to that of human intuition.

REGISTRATION

In order to test the repeatability of the image analysis and to investigate whether the tapetal-like reflex is subject to changes over time, it would be desirable to analyze the exact same region of a fundus using different slides. Therefore, a highly accurate image registration method has been developed that allows a comparison of the same anatomical region from different source images.

When the same area of the fundus is digitized from two different slides of the same patient, it is very likely that they will not exactly correspond to each

m1, m2	FM1	FM2	FM3	FM4	FM5	FMG
2, 22	0.87	0.74	0.60	0.64	0.92	0.23

Table 4. The fgures of merit of segmentation for the multi-scale thresholding shown in Fig. 27.

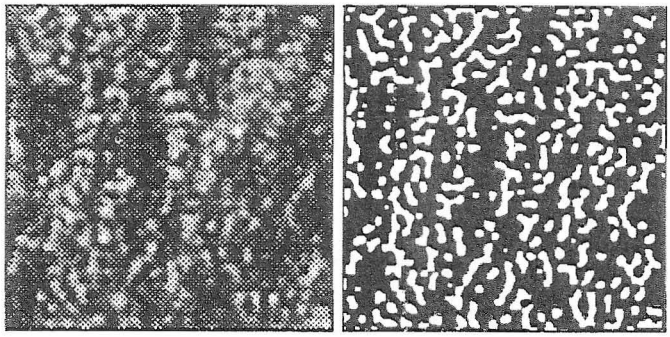


Figure 28. The restored version of an actual tapetal-like reflex image and the corresponding result of segmentation with multi-scale thresholding with m1=2 and m2=22.

other. In general, the differences might be due to: a) changes in the underlying anatomy; b) changes in the characteristics of the photographic film used; 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. Since the goals of registration are to determine the differences due to analysis when there are no underlying anatomical changes, or to determine changes due to the underlying anatomy, one must compensate for the other effects. Using the model of the imaging system, one can compensate for changes in film parameters. The slowly-varying illumination component can usually be removed (if necessary) with simple filtering. Most important of all, a registration method is needed in order to deal with the effect of the location and orientation of the fundus camera with respect to the eye being photographed. This registration method should, effectively and accurately, calculate the deformation parameters.

Before determining the extent of deformations, it must first be known what type of deformations to expect. Fundus images to be registered might differ in relative rotation because small torsional rotations of the eye ($<\pm$ 10°), depending on the location of the fixation point, are possible. 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 because the photographic geometry cannot be held constant for any length of time.

Two simple assumptions are made: (l) nonuniform scaling and (2) parallax effects are negligible. The first assumption is based on the fact that the high-resolution nature of the data combined with limited computational resources forces the consideration of relatively small regions of interest (640x640 pixels≈5°x5°). 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 [68]. 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. It is assumed therefore that translation, rotation and global scaling are sufficient to explain all the geometric mismatches to be encountered when registering fundus images. For methodological convenience, the center of each image is used as the center of rotation and scaling which does not affect the generality of the solution [69,appendix].

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 [69]. Accurate registration is a prerequisite to many image analysis applications, and therefore it has received much attention. In general, 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.

Local methods are based on the correspondence between a small subset of image pixels that are usually called landmark or control points. Of course, neither the determination of which pixels to consider as landmark points in each image, nor the determination of the correspondence between landmark points in two images are easy. A simple solution to both problems is the interactive approach, in which a human expert chooses corresponding landmark points in both images. In general, the human involvement makes the registration process less accurate. Furthermore, 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.

In this application, the objects of interest (patches of golden reflections) are relatively large (5 to 12 pixels wide) with no distinct boundaries, and noisy, thereby making accurate automated or manual landmark-choosing very difficult. Therefore, a global registration method needs to be applied to the registration of fundus images.

In general, the aim of global registration methods is to find a single m-parameter transformation that optimally registers most of the pixels of the two images. For this purpose, 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 [70] has been the most commonly used similarity measure because of its proven discriminatory abilities in the presence of white noise, but other measures like phasec-orrelation [71], sum of absolute differences and sequential similarity detection [72] 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 both in resolution and in 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 [70,73]. Higher dimensional parameter spaces can also be approached with an exhaustive search strategy, but with obvious penalties in speed.

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. In the hierarchical search, the algorithms 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. A hierarchical search is usually favorable from a computational point of view, but unfavorable in terms of the assurance of a global optimum.

A two stage method has been developed in this study for the case of a translation, rotation and global scaling based registration: in 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 [69].

Recently, Apicella et al. [74,75] have described a method called "fast correlation matching" and applied it to the registration of multi-modality (PET and MRI) images of the brain. 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. This is accomplished by a log-polar mapping of the Fourier magnitude of the images. 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.

In theory, the cross-correlation of the log-polar Fourier magnitudes can be made as accurate as the spatial domain resolution allows it, but, in practice, the increased spacebandwidth requirements 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.

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. The maximization is achieved with the "simplex" method [38,p.305], which is straightforward, does not require derivative calculations, and is extremely robust.

The accuracy of registration achieved with this new method for 640x640 pixel images has been shown to be better than 0.07 degrees, 0.1 %, and 0.3 pixels for rota-

tion, scaling, and translation, respectively. A complete and detailed description of the technique, its implementation and validation have been published previously [69].

RESULTS

1. Selection of Fundus Images

Color fundus photographs (slides) used in this study were from one eye each of 4 XLRP carriers, who are referred to as JR, NW, CR, and TH. The ages at the time of photography were 17 (JR), 42 (NW), 44 (CR), and 48 (TH) years.

Two neighboring 5° by 5° (640x640 pixels) regions on the horizontal meridian were considered in each eye. The regions, labeled A and B, were centered on 7.5° and 12.5° temporal to the fovea, respectively (Fig. 29). Before the digitization, all available fundus photographs imaging the region of interest (2 to 16, depending on the patient) were viewed under magnification, and several photographs (2 to 4) which displayed the highest sharpness of the tapetal-like reflex structures were selected. The selected slides were digitized at a coarse resolution, printed, and the location of the fovea as well as the locations of the two regions to be scanned were marked. The location of the center of the fovea was defined as the geometric center of the darkly pigmented foveal region. The two regions on each selected slide were digitized at high resolution and saved to disk. Multiple digital images of each region for each patient were ranked visually. Specifically, the digital images of the region were displayed on a large (20") graphics monitor one after another with a rapid cyclic sequence, and they were visually ranked according to the sharpness of the tapetal-like reflex. For all of the following analysis and results, only the sharpest image of a region was used, except for the repeatability test, where the sharpest and the second-sharpest images were used for each region.

2. Repeatability of Reflex Detection

No absolute validation for the problem of detecting individual tapetal-like reflex patches can be performed since, at this time, there is no possibility of independent confirmation with a different method. Therefore, the repeatability of reflex detection among different images of the same region, which is a measure for the reliability of the image analysis, is considered as a 'relative' validation method.

Two images of region B digitized from different fundus photographs for each of the four patients were used for the repeatability analysis. Each image was restored and segmented. One of the segmented images in each pair was transformed according to predetermined registration parameters, and after the geometric transformation, both images were cropped to 512x512 pixel size to result in full-frame images. The relationship between registered pairs are shown in Fig.s 30 and 31. Each pixel in these figures is coded with one of four colors: white, black, light gray, and dark gray. The two grays represent a 'match' whereas the black and the white indicate a 'mismatch.' Specifically, a pixel is light gray when in both images it was segmented as background, and a pixel is dark gray when in both images it was

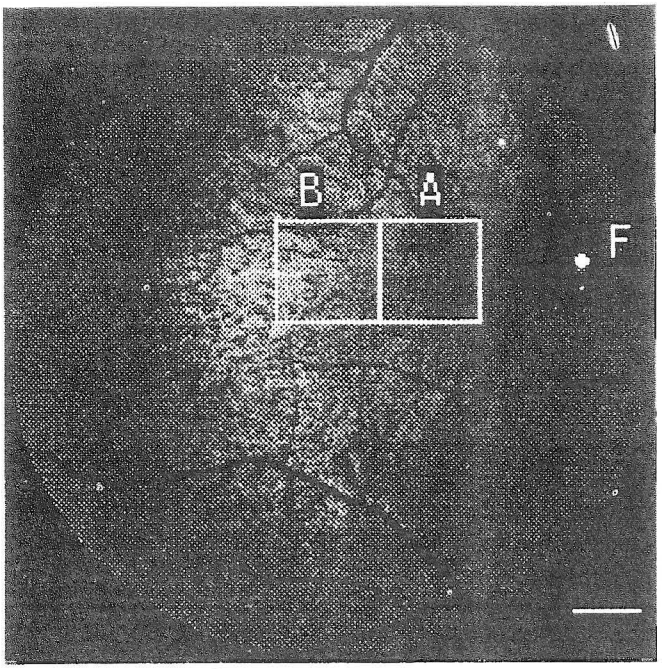


Figure 29.Location of the two regions (A,B) considered. Fovea is marked with a bright dot (F), and the calibration bar is 1 mm measured on the

segmented as reflex. A pixel is white or black when that pixel was segmented as reflex in one image and as background in the other image, or vice versa.

The percentage of gray pixels (light and dark), i.e. same classification in both images, were 85.4%, 86.8%, 85.6% and 82.5% of the total number of pixels for the four image pairs, respectively. Differentiation between 'critical' mismatches and 'noncritical' mismatches shows that the majority of the mismatching pixels can be considered noncritical because they simply change the boundary of an object; i.e., they make an object a little 'fatter' or 'thinner.' In contrast, 'critical' mismatches mean that there is a complete object in one image, and there is no object in the other image. As a side note, figures 30 and 31 also show the high accuracy of the registration method, as tiny errors in registration parameters would show up very distinctly in the 4-color coded images as a bias in the mismatched pixels.

As the results show, the segmentation of different images of the same region are highly repeatable. On the average, 85% of the total number of pixels are classified equally in two independent images. The majority of the remaining 15% do not have common classification, probably due to slight differences in image quality.

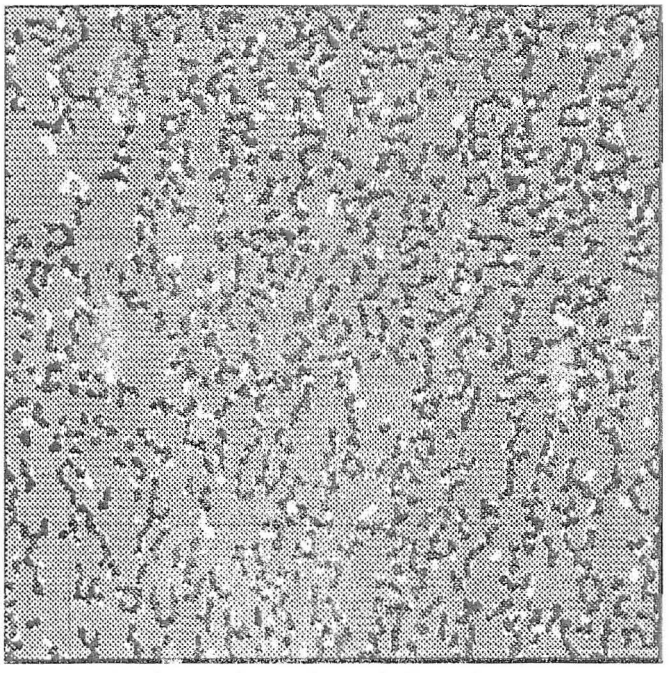


Figure 30.Comparison of two independent segmentations of region B in the right eye of patient TH.

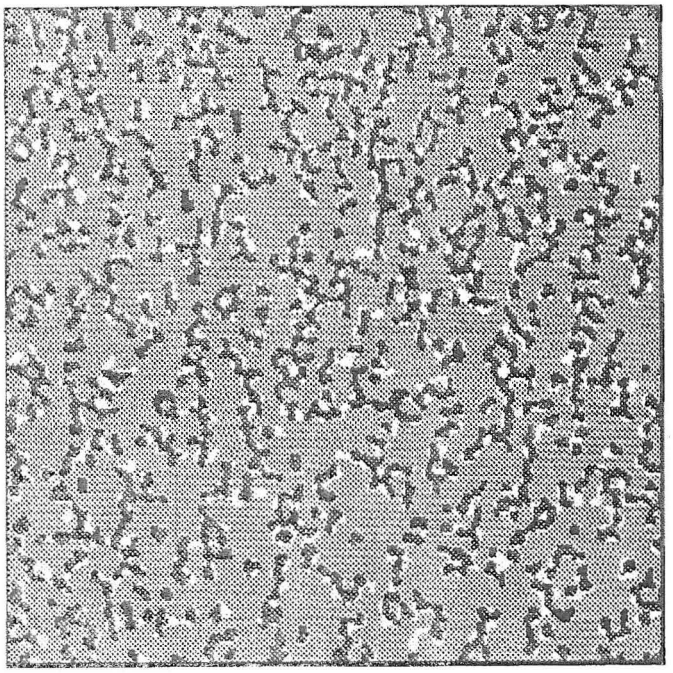


Figure 31.Comparison of two independent segmentations of region B in the right eye of patient NW.

3. Radiometric Properties of the Reflex

The reflectance of the tapetal-like reflex is a basic property that needs to be characterized. Using the methods described in this study, an absolute reflectance value for the tapetal-like reflex cannot be determined, but instead, a relative reflectance can be calculated. The relative reflectance is defined as the local contrast of a patch of reflex in the intensity domain. That is, the average intensity of a reflex divided by the average intensity of the surrounding non-reflex retina. This measure of local contrast would be equal to the relative reflectance assuming the illumination is uniform within the small region surrounding each patch of reflex.

Fig. 32 shows the distributions of local contrast of the tapetal-like reflex in region B of 4 patients. The vertical axis in this figure corresponds to the number of reflexes detected with the segmentation method, and the horizontal axis is the local contrast. The reader is reminded that the segmentation process requires a post-processing operation where the objects below a local

contrast threshold are discarded. This threshold was approximately 1.15, therefore values below that are not shown in the figure.

The mean and standard deviations of the distributions in Fig. 32 are 1.3±0.1 for each one of the four patients. This would suggest that the reflectance of the tapetal-like reflex is on the average 30% higher than the reflectance of the neighboring non-reflex retina. Since the linearization required for the restoration process was shown to reduce the intensity domain amplitudes of some peaks, the actual reflectance of the tapetal-like reflex might actually be even higher.

4. Geometric Properties of the Reflex

Reflex Size

One of the important questions that may be raised about the tapetal-like reflex is: "Does the reflex consist of randomly sized patches, or is it based on multiples of a unit sized patch?" An accurate answer to this question might increase understanding of the origin of the reflex.

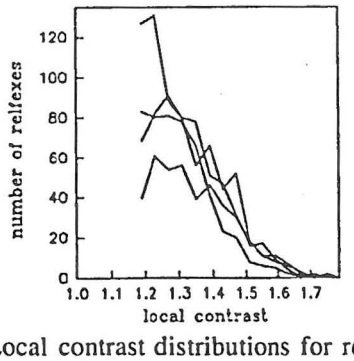


Figure 32.Local contrast distributions for region B of patients CR, TH, NW, and JR.

One possible method of analyzing the distribution of particle size in segmented tapetal-like reflex images would be to simply calculate area statistics of the detected reflex patches. However, such a method would not be able to differentiate between large blobs, long thin straight objects, and long thin convoluted objects. To derive reflex patch size distributions, a method called 'granulometrics' was used, which is derived from mathematical morphology [81]. In this method, segmented images are filtered morphologically with structuring elements (stelts) of ever increasing size. The number of pixels remaining after the morphological filtering is recorded as a function of stelt size. When the pixel counts are normalized by the number of original object pixels, and subtracted from 1, a cumulative probability function results. The derivative of the cumulative probability distribution is the probability density function (pdf) of particle size in a given segmented image.

Fig. 33 shows the plots of pdf for the granulometric analysis of region B in four patients. The horizontal axis is the diameter (in pixels) of the disc-shaped stelt used for morphological filtering. The values plotted for each size give the probability of a pixel to belong to an object of no-less than that size, and less than the next larger size. The mean and the standard deviation values were 6.3±1.4, 5.9±1.4, 6.6±1.5, and 6.0±1.4 pixels for patients CR, TH, NW, and JR, respectively.

Granulometric analysis suggests that in region B of the fundi of 4 XLRP carriers, segmented tapetal-like reflex images result in objects with a basic 'size' of approximately 6 pixels which is equal to 14μm on the retina. It is important to note that the previous statement does not mean that the segmented structures are all 6 pixel diameter discs. Instead, it suggests that the most likely dimension of an arbitrarily shaped object is 6 pixels or an integer multiple of 6 pixels. For example, a 'worm-like' structure made from 6 pixel diameter discs put next to each other on an s-shaped line can be considered an object with a basic 'size' of 6 pixels. Another example would be a 'y-shaped' structure with its linear segments having the width of 6 pixels.

Reflex Directionality

One of the first impressions when looking at a tapetal-like reflex is that it radiates from the fovea to-

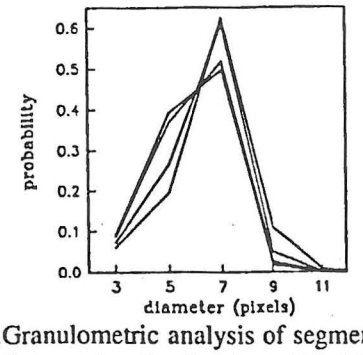


Figure 33.Granulometric analysis of segmented images from region B of patients CR, TH, NW, and JR.

wards the periphery. This impression may be due to two possible reasons: either the reflex consists of small elongated objects that have a preferred orientation towards the fovea, or it consists of randomly oriented and shaped small objects that are located such that spaces in between reflexes (the retinal background) have a preferred orientation towards the fovea. Just as in the previous section, mathematical morphology is used to investigate the possible directionality of the tapetal-like reflex.

For the directional analysis, the 'bounding angle' of a region in the fundus is defined as the angle between two lines that originate at the fovea and intersect the two most extreme pixels of the region being analyzed (Fig. 34). Three lines dividing the bounding angle equally are used to divide the region of interest into four subregions. Each one of the subregions is assigned a 'foveal orientation' as the direction along the line that bisects the angle between the two lines defining the subregion. Each subregion was morphologically filtered with a line-shaped structuring element of 21 pixels length, 1 pixel width and a sequence of orientations. The orientations were chosen with 5 degree increments from -90 degrees to +90 degrees, where 0 degrees corresponded to the 'foveal orientation.'

Fig.s 35 through 38 show the percent of object pixels left after the morphological filtering as a function of structuring element orientation in the case of four patients. In each graph, 4 line plots represent the results obtained from the four subregions. The results show a definite orientation preference of the objects making up the tapetal-like reflex: in region B of the four patients analyzed, elongated objects have a preferred orientation towards the fovea.

6. Biological Implications

A basic assumption in this study is that the tapetal-like reflex results from the partial expression of the same gene defect that causes XLRP in men. A further assumption is that the tapetal-like reflex originates in or near the retinal pigment epithelium (RPE) and photoreceptors, because XLRP is known to cause the degeneration of these cells. Support for the second assumption arises from clinical observations that place the reflex in

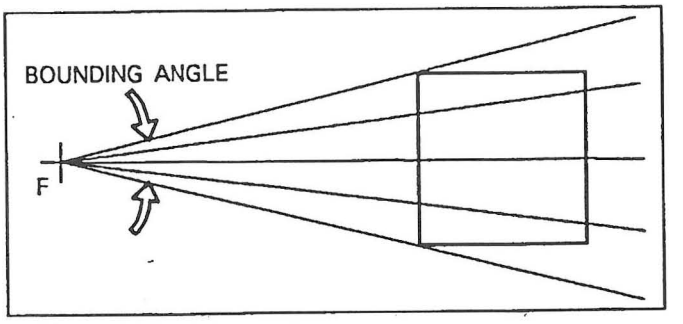


Figure 34.Definition of bounding angle, and how it is used to divide up a region of interest. (F) is the location of the fovea.

the 'deep retina' [18,20]. Although the exact basis of the tapetal-like reflex cannot be determined by the results presented in this paper, it is interesting to analyze whether these results can support any of the existing theories about the origin of this retinal reflection.

The digital image analysis presented has led to the conclusions that the reflex can have a unit size of 6 pixels or smaller (see below); it is made up of elongated clusters that are oriented toward the fovea; and in one of

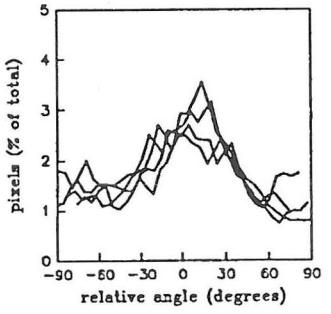


Figure 35.Patient CR, left eye, region B.

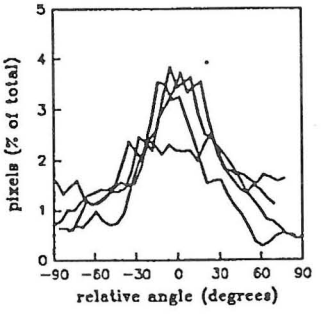


Figure 36.Patient TH, right eye, region B.

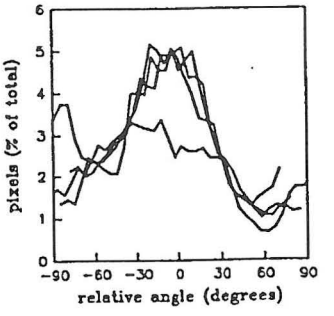


Figure 37.Patient NW, right eye, region B.

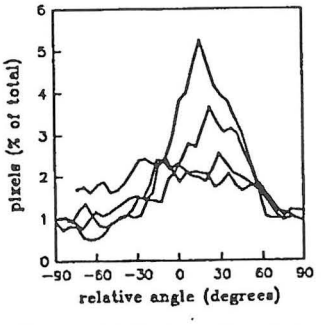


Figure 38.Patient JR, right eye, region B.

the examined subjects, it did not undergo dramatic changes in appearance over at leat two decades [85]. Assuming the reflex is not of extracellular origin, then it must derive from RPE, rods or cones. The foveal orientation of the reflex patches weighs against an RPE cell origin. In carriers of X-linked ocular albinism, the patches of hypopigmented RPE cells are oriented towards the optic disc [86] and there is evidence that this orientation results from a pattern of embryological development [86]. On the other hand, a unit size of 6 pixels (14 μm) is consistent with the size of RPE cells in the retinal region studied [87]. When some of the sharpest tapetal-like reflexes are inspected visually, however, some tiny (2x2 pixel) bright reflections persist in multiple images of the same region. Because of their very small size, the restoration process smooths these structures out and therefore they do not show up in the analysis. Such small sized objects would not be consistent with an RPE cell hypothesis.

A photoreceptor origin to the reflex may be more plausible. In the vicinity of region B (12 degrees temporal to the fovea), rod inner segments have approximate diameters of 2 μm and cone inner segments are approximately 5 μm in diameter. In this region, rods are close to their highest concentration (120,000 rods/mm^2) whereas cone densities are much lower (7000 cones/mm^2) [88]. In terms of the digitized fundus images used in this study, the rods correspond to less than 1 pixel sized structures distributed almost 1 rod per pixel. Cones would correspond to approximately 2x2 pixel structures distributed with a center-to-center distance of about 5 pixels. A cone hypothesis is favored because we recently found that the earliest detectabl functional abnormality in affected family members (men and carriers) of a large XLRP pedigree is in the cone photoreceptor system [89]. Furthermore, most of the carriers in this family had a tapetal-like reflex. It is not known (and would be interesting to evaluate) whether cone function of carriers deteriorates over time in regions where the reflex appears to be stable over decades.

CONCLUSIONS

In this study, a noninvasive imaging approach was taken to systematically investigate the tapetal-like reflex seen in carriers of XLRP. Digital image processing methods were developed and applied to the analysis and understanding of this reflex. Several previously unknown properties of the reflex were quantified, and based on these properties a preliminary hypothesis was suggested about its origin. The methods developed in this study are general in nature and therefore can be applied to a broad range of image analysis problems.

The first aims of this study was the determination of a mathematical model for the relationship between the light intensity leaving the fundus and its recorded image, including all four components of the imaging system: eye, camera film, and scanner. The model, which combines linear blurs, nonlinear point processes, and signal-dependent noise sources, is accurate, as validated with several experiments, and it showed that the eye and the fundus camera are the imaging system components causing most of the blurring, whereas the film is responsible for most of the image noise and the nonlinearities. Although the parameter values and functional relationships of the model were developed for a specific imaging system, the equations for the model remain valid for most other film-based imaging systems. The current trend towards direct digital imaging does not change the fact that film-based imaging is still the most common mode both in ophthalmology and in radiology. Furthermore, in medical and non-medical fields a major fraction of the photographic records collected over the past century might not be reproducible. Therefore good models of film-based imaging systems are very useful in understanding and analyzing the full information content of these records. The concept presented in this study is, however, not limited to photographs. Replacing the photographic film properties with the characteristics of other image detectors and electronic screens, the approach remains valid and clearly shows, how establishing the mathematical relationship between pixels in the image and the originating biological structures can improve upon medical image analysis by permitting adequate image restoration.

A major aim of this study was the development of

a digital image restoration method which would reduce the image noise, recover the diminished high-frequency information, and invert the nonlinearities. The restoration method is based on the simplification of the complete nonlinear imaging system model to a generalized linear model. Restoration results showed the expected suppression of noise, recovery of high-frequency information, and inversion of nonlinearities. The assumptions made in simplifying the imaging system model were application-specific but the methodology of the simplification was very general and should be applicable to most image restoration problems considering a simplified model is usually necessary because of computational limitations. Furthermore, in optical imaging, the properties of biological structures of interest are related to the intensity of light originating from them. Therefore, analysis based on the estimates of the original scene intensity, as done in this work, is a considerable improvement over the usual approach of simply smoothing a digitized image.

In order to automate the detection of the tapetal-like reflex patches, a novel object-background segmentation method, called multi-scale segmentation, was developed. The method is based on local thresholding of images using two differently sized neighborhoods around each pixel; a pixel was considered to be part of an object only if it was above the thresholds calculated within both of the neighborhoods. This segmentation method allows good separation of distinct objects as well as a low rate of falsely detected objects. Its performance was quantified using simulated images and several object-based figures-of-merit which were more informative than classical pixel-based measures. The results of the segmentation were shown to be highly repeatable when applied to different images of the same region of the fundus. The multi-scale segmentation method is simple, fast, reliable, and general. It can be adapted to the segmentation of other images where there are many variable intensity objects distributed on a variable intensity background. For example cell number, shape and size analysis in histological specimens is an application where multi-scale segmentation may be applied.

One of the important aims of this study was the development of a method that achieves high accuracy registration of fundus images. The registration is based on the combination of two optimizations: a global optimization that determines the best registration parameters in a 4-dimensional discrete parameter space followed by a local optimization that refines the registration parameters within a 4-dimensional continuous parameter space. The registration method was shown to achieve an accuracy of better than 0.07 degrees in rotation, 0.1% in scaling, and 0.3 pixels in translation for real fundus images, and it was shown to be efficient enough to be used on IBMcompatible personal computers for the case of 640x640 pixel images. Similar to the segmentation method, the registration method was very general in nature, and is applicable to the registration of any medical and nonmedical images where the deformations can be approximated with the three assumed transformations translation, rotation, and scaling.

The methods developed in this study were applied to fundus photographs in order to perform an analysis of the tapetal-like reflexes seen in XLRP and to investigate their origin. The results showed that the reflectance of the patches that make up the tapetal-like reflex were on the average 30% higher than the reflectance of the neighboring non-reflex retina. Mathematical morphology methods applied to segmented images suggested that the reflexes origin from elongated structures with a dimension of approximately 14μm and a preferred orientation towards the fovea. Based on these results and other known facts about the tapetal-like reflex and XLRP, a cone photoreceptor origin is favored as the source of the reflex.

This is the first study, known to the authors, analyzing features seen in high resolution digital fundus images. The methods developed for automatic, reproducible detection of tiny patches of the tapetal-like reflex from fundus photographs has an immediate clinical application: the quantification of the progression of the tapetal-like reflex as well as many other fundus lesions that result from systemic or ocular diseases.

The presented techniques for image segmentation, registration, and morphological analysis can also readily

be adapted to other areas of medical imaging. Some examples of possible applications in the medical field include images from multiple modalities (e.g. MRI vs. PET) to correlate metabolic activity to anatomic structure and images of cancerous lesions (e.g. dermatological) for quantitation of growth or regression.

In diagnosis and treatment of cancer, automated segmentation and multimodality image registration are critical research topics that have not yet found adequate solutions. For the multimodality detection and delineation of neoplasms - important e.g. for adequate approaches to serial evaluation ol tumor response to treatment - reliable automatic image processing (in contrast to currently used manual delineation of tumor margins) would facilitate the process of volume measurement, minimize errors, and would make available the combined diagnostic power of computerized tomography, positron emission tomography, conventional and metabolic chemical shift magnetic resonance imaging, and immunoimaging.

ACKNOWLEDGEMENT

The authors wish to thank Drs. S.G. Jacobson, C.M. Kemp, and R.W. Knighton for their critical advice and support. This study was supported in part by Public Health Services grant EY05627, the National Retinitis Pigmentosa Foundation, Inc. (Baltimore, MD), and the Chatlos Foundation, Inc. (Longwood, FL).

REFERENCES

1. Peli E, Hedges TR, Schwartz B, "Computerized enhancement of retinal nerve fiber layer", Acta Ophthalmologica, 64:113-122, 1986.
2. Peli E, Peli T, "Restoration of retinal images obtained through cataracts", IEEE Trans. Med. Imag., 8(4):401-406, 1989.
3. Phillips RP, Spencer T, Ross PGB, Sharp PF, Forrester JV, "Quantification of diabetic maculopathy by digital imaging of the fundus", Eye, 5:130-137, 1991.
4. Ward NP, Tomlinson S, Taylor CJ, "Image analysis of fundus photographs", Ophthalmology, 96(1):80-86, 1989.
5. Gilchrist J, "Analysis of early diabetic retinopathy by computer processing of fundus images - a preliminary study", Ophthalmol. Physiol. Opt., 7(4):393-399, 1987.
6. Goldbaum MH, Katz NP, Nelson MR, Haff LR, "The discrimination of similarly colored objects in computer images of the ocular fundus", Invest. Ophthalmol. Vis. Sci., 31(4):617-623, 1990.
7. Chaudhuri S, Chatterjee S, Katz N, Nelson M, Goldbaum M, "Detection of blood vessels in retinal images using two-dimensional matched filters", IEEE Trans. Med. Imag., 8(3):263-269, 1989.
8. Cox MJ, Wood ICJ, "Computer-assisted optic nerve head assessment", Ophthalmol. Physiol. Opt., 11:27-35, 1991.
9. Algazi VR, Keltner JL, Johnson CA, "Computer analysis of the optic cup in glaucoma", Invest. Ophthalmol. Vis. Sci., 26:1759-1770, 1985.
10. Varma R, Spacth GL, "The PAR IS 2000: a new system for retinal digital image analysis", Ophthalmic Surgery, 19(3):183-192, 1988.
11. Nagin P, Schwartz B, Reynolds G, "Measurement of fluorescein angiograms of the optic disc and retina using computerized image analysis", Ophthalmology, 92(4):547552, 1985.
12. Marmor MF, et al., "Retinitis pigmentosa: a symposium on terminology and methods of examination", Ophthalmology, 90(2):126-131, 1983.
13. Pagon RA, "Retinitis pigmentosa", Surv. Ophthalmol., 33:137-177, 1988.
14. Wright AW, "Toward the identification of genes in Xlinked retinitis pigmentosa", in "Progress in Retinal Research", N. Osborne and G. Chader (eds.), Vol.9, Oxford, Pergamon Press, pp.197-227, 1990.
15. Lyon MF, "Gene action in the X-chromosome of the mouse (Mus.musculus. L.)", Nature, 190:372-373, 1961.
16. Jay B, "X-linked retinal disorders and the Lyon hypothesis", Trans. Ophthalmol. Soc. U.K., 104:836-844, 1985.

17. Mann I, "Developmental abnormalities of the eye", Cambridge University Press, London, 1937.
18. Falls HF and Cotterman CW, "Choroidal degeneration: a sex-linked form in which heterozygous women exhibit a tapetal-like retinal reflex", Arch. Ophthalmol., 40:685-703, 1948.
19. Nicol JAC, "Tapeta lucida of vertebrates", in "Vertebrate photoreceptor optics", J.M. Enoch, F.L. Tobey (ed.s), Springer-Verlag, Berlin, pp. 401-431, 1981.
20. Francois J, "Chorioretinal degeneration or retinitis pigmentosa of intermediate sex-linked heredity", Doc. Ophthalmol., 16:111-127, 1962.
21. Fishman GA, Weinberg AB, McMahon TT, "X-linked recessive retinitis pigmentosa", Arch. Ophthalmol., 104:1329-1335, 1986.
22. Szamier RB, Berson EL, "Retinal histopathology of a carrier of X-chromosome-linked retinitis pigmentosa", Ophthalmology, 92(2):271-278, 1985.
23. Campbell FW, Gubish RW, "Optical quality of the human eye", J. Physiol., 186:558-578, 1966.
24. Santamaria J, Artal A, Bescos J, "Determination of the point-spread function of human eyes using a hybrid opticaldigital method", J. Opt. Soc. Am. - A, 4(6):1109-1114, 1987.
25. Gubish RW, "Optical performance of the human eye", J. Opt. Soc. Am., 57(3):407-415, 1967.
26. Leutwein K, Littmann, H "The fundus camera", in "Refraction and Clinical Optics", A. Safir (ed.), New York, Harper & Row, pp.457-466, 1980.
27. Frisen L, "Resolution at low contrast with a fundus camera. Comparison of various photographic films", Invest. Ophthalmol., 12(11):865-869, 1973.
28. Langford M, "Advanced photography", 5th Edition, Focal Press, London and Boston, 1989.
29. Fuji Photo Film Co., Ltd., "Fuji film data sheet. Fujichrome 100".
30. Lee HC, "Review of image-blur models in a photographic system using the principles of optics", Opt. Eng., 29(5):405-421, 1990.
31. Dainty JC and Shaw R, "Image science", Academic Press, San Diego, CA, 1988.
32. Naderi F and Sawchuck AA, "Estimation of images degraded by film-grain noise", Appl. Opt., 17(8):12281237, 1978.
33. Falconer DG, "Image enhancement and film-grain noise", Optica Acta, 17(9):693-705, 1970.
34. Kraus W, "The illumination of the emulsion for photographic granularity measurements", Photo. Sci. Eng., 12(4):217-218, 1968.
35. Nikon Corp., "Nikon 35mm film scanner LS-3500. User's manual".
36. Janesick JR, Elliott T, Collins S, Blouke MM, Freeman J, "Scientific charge-coupled devices", Opt. Eng., 26(8):692-714, 1987.
37. Maguire GQ, Noz MF, Schimpf JH, Horii SC, "Gamma camera MTF measurements using an image chain analysis", Comp. Biomed. Res., 16:149-159, 1983.
38. Press WH, Flannery BP, Teukolsky SA, Vetterling WT, "Numerical recipes in C", Cambridge University Press, Cambridge, 1988.
39. Hunt BR, "Bayesian methods in nonlinear digital image restoration", IEEE Trans. Comp., C-26(3):219-229, 1977.
40. Kuan DT, Sawchuk AA, Strand TC, Chavel P, "Adaptive noise smoothing filter for images with signal dependent noise", IEEE Trans. Pat. Anal. Mach. Intel. PAMI-7(2):165-177, 1985.
41. Hunt BR, "The application of constrained least squares estimation to image restoration by digital computer", IEEE Trans. Comp., C-22(9):805-812, 1973.
42. Trussel HJ, "Convergence criteria for iterative restoration methods", IEEE Trans. Acous. Speech Sig. Proc., ASSP- 31(1):129-136, 1983.
43. Thompson AM, Brown JC, Kay JW, Titterington DM "A study of methods of choosing the smoothing parameter in image restoration by regularization", IEEE Trans. Pat. Anal. Mach. Intel., 13(4):326-339, 1991.
44. Biemond J, Lagendijk RL, Merserau RM, "Iterativ methods for image deblurring", Proc. IEEE, 78(5):856-883 1990.
45. Sezan MI and Tekalp AM, "Survey of recent developments in digital image restoration", Opt. Eng.,

29(5):393-404, 1990.

46. Phillips DL, "A technique for the numerical solution of certain integral equations of the first kind", J. Ass. Comput. Mach. 9:84-97, 1962.
47. Tikhonov AN, Arsenin VY, "Solutions of ill-posed problems", John Wiley & Sons, Inc., New York, 1977.
48. Miller K, "Least squares methods for ill-posed problems with a prescribed bound", SIAM J. Math. Anal., 1:52-74, 1970.
49. Gonzalez RC, Wintz P, "Digital image processing", 2nd Edition, Addison-Wesley Publishing Company, Reading, Massachusetts, 1987.
50. Pratt WK, "Digital image processing", 2nd Ed., John Wiley & Sons, Inc., New York, 1991.
51. Haralick RM, Shapiro LG, "Image segmentation techniques", Comp. Vis. Graph. Image Proc., 29: 100-132, 1985.
52. Derin H, Won CS, "A parallel image segmentation algorithm using relaxation with varying neighborhoods and its mapping to array processors", Comp. Vis. Graph. Image Proc., 40:54-78, 1987.
53. Maragos P, Schafer RW, "Morphological systems for multidimensional signal processing", Proc. IEEE, 78(4):690-710, 1990.
54. Chow CK, Kaneko T, "Automatic boundary detection of the lefl ventricle from cineangiograms", Comp. Biomed. Res., 5:388-410, 1972.
55. Kittler J, Illingworth J, Foglein J, "Threshold selection based on a simple image statistic", Comp. Vis. Graph. Image Proc., 30:125-147, 1985.
56. Parker JR, "Gray level thresholding in badly illuminated images", IEEE Trans. Pat. Anal. Mach. Intel., 13(8):813819, 1991.
57. Gordon R and Rangayyan RM, "Radiographic feature enhancement, information content, and dose reduction in mammography and cardiac angiography", Proc. 5th Ann. Conf. of IEEE/EMBS, 10-12 Sept, pp.161-165, 1983.
58. Gordon R and Rangayyan RM, "Feature enhancement of film mammograms using fixed and adaptive neighborhoods", Appl. Opt., 23(4):560-564, 1984.
59. Rangayyan RM, Nguyen HN, "Pixel-independent image processing techniques for enhancement of features in mammograms", 8th Ann. Con£ of IEEE/EMBS, Fort Worth, Texas, pp.1113-1117, 1986.
60. Dhawan AP, Buelloni G, Gordon R, "Enhancement of mammographic features by optimal adaptive neighborhood image processing", IEEE Trans. Med. Imag., MI-5(1):8-15, 1986.
61. Morrow WH, Rangayyan RM, Desautels JEL, "Feature-adaptive enhancement and analysis of highresolution digitized mammograms", Ann. Intl. Conf. of IEEE/EMBS, pp.165-166, 1990.
62. Song WJ and Pearlman WA, "Edge-preserving noise filtering based on adaptive windowing", IEEE Trans. Circ. Sys., 35(8):1048-1055, 1988.
63. Ogawa K, Saito A, Nakajima M, Ando Y, Hashimoto S, "Regional adaptive histogram equalization using fuzzy sets", SPIE, Medical Imaging IV, Vol.1233, pp.374-383, 1990.
64. Lindeberg T, "Scale-space for discrete signals", IEEE Trans. Pat. Anal. Mach. Intel., 12(3):234-254, 1990.
65. Lifshitz LM, Pizer SM, "A multiresolution hierarchical approach to image segmentation based on intensity extrema", IEEE Trans. Pat. Anal. Mach. Intel., 12(6):529540, 1990.
66. Otsu N, "A threshold selection method form gray-level histograms", IEEE Trans. Sys. Man Cyber., SMC-9(1):6266, 1979.
67. Otsu N, "Discriminant and least squares threshold selection", 4th Intl. Joint Conf. on Pattem Recognition, Nov. 7-10, Kyoto, Japan, pp.592-596, 1978.
68. Behrendt T, Doyle KE, "Reliability of image size measurements in the new Zeiss fundus camera", Am. J. Ophthalmol., 59:896-899, 1965.
69. Cideciyan AV, Jacobson SG, Kemp CM, Knighton RW, Nagel JH, "Registration of high resolution Images of the retina", SPIE Proceedings "Medical Imaging VI: Image Processing", Vol.1652, pp.310-322, 1992
70. Pratt WF, "Correlation techniques of image registration", IEEE Trans. Aero. Elec. Sys., 10:353-358, 1974.
71. de Castro E, Morandi C, "Registration of translated and rotated images using finite Fourier transforms",

IEEE Trans. Pat. Anal. Mach. Intel., PAMI-9(5):700-703, 1987.

72. Barnea DL, Silverman HS, "A class of algorithms for fast image registration", IEEE Trans. Comp., C-21(2):179-186, 1972.
73. de Castro E, Cristini G, Martelli A, Morandi C, Vascotto M, "Compensation of random eye motion in television ophthalmoscopy: preliminary results", IEEE Trans. Med. Imag., MI-6(1):74-81, 1987.
74. Apicella A, "Image matching by correlation analysis: biomedical applications", Ph.D. dissertation, University of Miami, Coral Gables, FL, 1989.
75. Apicella A, Kippenhan JS, Nagel JH, "Fast multimodality image matching", Proc. SPIE Vol. 1092, Medical Imaging III: Image Processing, Jan. 3 1-Feb. 3, pp.252-263, 1989.
76. Casasent D, Psaltis D, "Deformation invariant, spacevariant optical pattern recognition", in Progress in Optics, E.Wolf (ed.), North-Holland Publishing Co., Amsterdam, pp.290-356, 1978.
77. Altmann J, Reitbock HJP, "A fast correlation method for scale- and translation-invariant pattern recognition", IEEE Trans. Pat. Anal. Mach. Intel., PAMI-6(1):46-57, 1984.
78. Jensen AS, Lindvold L, Rasmussen E, "Transformation of image positions, rotations, and sizes into shift parameters", Appl. Opt., 26(9):1775-1781, 1987.
79. Nelder JA, Mead R, "A simplex method for function minimization", Comp. J., 7(4):308-313, 1965.
80. Jain AK, "Fundamentals of digital image processing", Prentice Hall, Englewood Cliffs, 1989.
81. Dougherty ER, Pelz JB, "Morphological granulometric analysis of electrophotographic images - size distribution statistics for process control", Opt. Eng., 30(4):438-445, 1991.
82. Osch LV, Schooneveld MV, Blecker-Wagemakers EM, "Golden tapetal reflex in male patients with X-linked retinitis pigmentosa", Ophthal. Paed. Gen., 11(4):287-291, 1990.
83. Marmor MF, "Long-term follow-up of the physiologic abnormalities and fundus changes in fundus albipunctatus", Ophthalmology, 97:380-384, 1990.
84. Eastman Kodak Co., "Manufacturers data sheet on Kodachrome 64".
85. Cideciyan AV, "Digital analysis of high resolution fundus images", Dissertation, University of Miami, 1992.
86. Lang GE, Rott HD and Pfeiffer RA, "X-linked ocular albinism", Ophthal. Paed., 11(4):265-271, 1990.
87. Kaczurowski MI, "The pigment epithelium of the human eye", Am. J. Ophthalmol., 53:79-92, 1962.
88. Curcio CA, Sloan KR, Kalina RE, Hendrickson AE, "Human photoreceptor topography", J. Comp. Neur., 292:497-523, 1990.
89. Jacobson SG, Roman AJ, Cideciyan AV, Robey MG, Iwata T, Inana G, "X-linked retinitis pigmentosa: functional phenotype of an RP2 genotype", Inves. Ophthalmol. Vis. Sci., in press.