In Le Moigne, J. (ed.), *Proceedings of the Image Registration Workshop*, NASA Goddard Space Flight Center, Greenbelt MD, November 20-21, 1997. NASA Publication #CP-1998-206853. This document was re-typeset in 2013 using LaTeX. A scanned version of the original is available at `http://scanpath.arc.nasa.gov/publications/proc/irw_scanned.pdf`.

Recovery of motion parameters from distortions in scanned images

Jeffrey B. Mulligan
NASA Ames Research Center

Abstract: Scanned images, such as those produced by the scanning-laser ophthalmoscope (SLO), show distortions when there is target motion. This is because pixels corresponding to different image regions are acquired sequentially, and so, in essence, are slices of different snapshots. While these distortions create problems for image registration algorithms, they are potentially useful for recovering target motion parameters at temporal frequencies above the frame rate. Stetter, Sendtner and Timberlake [1] measured large distortions in SLO images to recover the time course of rapid horizontal saccadic eye movements. Here, this work is extended with the goal of automatically recovering small eye movements in two dimensions. Eye position during the frame interval is modeled using a low dimensional parametric description, which in turn is used to generate predicted distortions of a reference template. The input image is then registered to the distorted template using normalized cross correlation. The motion parameters are then varied, and the correlation recomputed, to find the motion which maximizes the peak value of the correlation. The location and value of the correlation maximum are determined with sub-pixel precision using biquadratic interpolation, yielding eye position resolution better than 1 arc minute [2]. This method of motion parameter estimation is tested using actual SLO images as well as simulated images. Motion parameter estimation might also be applied to individual video lines in order to reduce pipeline delays for a near real-time system.

1. Introduction

Video image sequences are often used to track object motion. Unless a special high frame-rate camera is used, the recovered motion is usually sampled in time at the video frame rate (50-60 Hz). While low resolution sampling is adequate for many applications, documentation of high-speed events often requires higher temporal resolution. For images obtained with a scanned system, in which individual pixel values are acquired at different times, it is possible to obtain higher temporal resolution for the motion of extended targets. The sequential nature of the scanning process introduces geometric distortions in the image of a moving target. By measuring these distortions, high temporal resolution information about the target motion can be recovered. This technique is especially useful when *a priori* knowledge about the possible target motions permits a concise description using low-dimensional parametric models, because this reduces the space of possible distortions which must be searched. In the following sections, expressions for the precise form of the motion-induced distortions will be derived.

1.1. Raster scanning

x, y	position in image plane
$\mathrm{s}_x(t), \mathrm{s}_y(t)$	position of scan at time t
f_L	line frequency ($\approx$ 15 kHz)
f_F	frame rate ($\approx$ 60 Hz)
i_L	index of current line
t_S	start time of current line
t_L	time in current line, $t - t_\mathrm{S}$
$\mathrm{v}_{\mathrm{S},x}, \mathrm{v}_{\mathrm{S},y}$	scan velocities

Some imaging systems, using an electronic or mechanical shutter, can simultaneously capture all of the pixels in an image. In a scanned system, however, only a single point is sensed at a given time, and the location of this point is swept over the image area by electronic or mechanical means. Here we present some definitions and conventions that will allow us to precisely describe the scanning process.

The imaging area is defined to be a rectangle indexed by normal Cartesian coordinates x and y. The raster is defined by two scan functions, $\mathrm{s}_x(t)$ and $\mathrm{s}_y(t)$, which represent the instantaneous beam position. These functions are approximated by sawtooth waveforms (see figure 1). By convention, the horizontal dimension is scanned at a relatively high frequency, called the *line frequency,* f_L, while the slower vertical frequency determines the *frame rate,* f_F.

Address for author correspondence: MS 262-2, NASA Ames Research Center, Moffett Field, CA, 94035-1000. Email: jbm@vision.arc.nasa.gov. An electronic version of this paper with additional multimedia figures is available at:
http://vision.arc.nasa.gov/ jbm/papers/irw97.html.

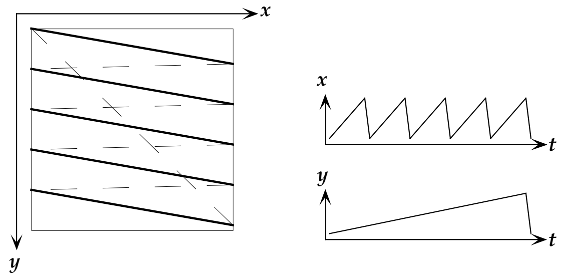

Figure 1: Diagram of raster pattern on the left, with the active portion of each line shown as a heavy solid line, and retrace as a dashed line (see appendix). On the right, the scan functions are shown over time.

Time t=0 in our temporal coordinate system is the beginning of the current frame. By convention, numbering of raster lines begins with 1; The index of the current line, i_L, is

$$\mathrm{i}_\mathrm{L} = \lfloor t f_\mathrm{L} \rfloor \tag{1}$$

We define t_S to be the time of the start of the current line, and t_L to be the time relative to the start of the current line:

$$t_\mathrm{S} = \frac{\mathrm{i}_\mathrm{L}}{f_\mathrm{L}}, \qquad \text{and} \qquad t_\mathrm{L} = t - t_\mathrm{S}. \tag{2a,b}$$

These quantities are illustrated graphically in figure 2.

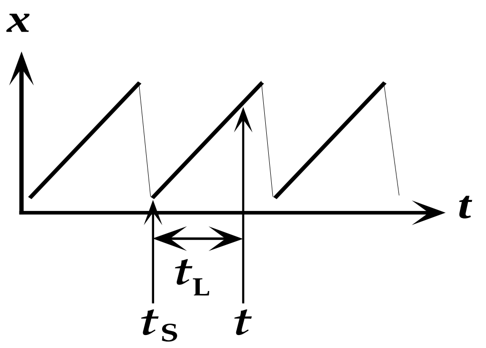

Figure 2: Raster waveform diagram, indicating the current time, t, the start time of the current line, t_S, and the time within the current line, t_L.

The *scan velocities,* $\mathrm{v}_{\mathrm{S},x}$ and $\mathrm{v}_{\mathrm{S},y}$, describe the rate at which the scanning beam traverses the image plane. When expressed in units of image widths per second, these are approximately equal to the scan frequencies, f_L and f_F (see appendix for details). We can write simple expressions for the instantaneous scan position in terms of the scan velocities. The horizontal scan position $\mathrm{s}_x(t)$ is:

$$\mathrm{s}_x(t) = t_\mathrm{L}\mathrm{v}_{\mathrm{S},x}. \tag{3a}$$

In most scanning systems, the vertical scan is continuous (partly due to ”mechanical” constraints), and

$$\mathrm{s}_y(t) = t\mathrm{v}_{\mathrm{S},y}. \tag{3b}$$

1.2 Effects of object motion

P	a target point
$\mathrm{p}_x(t), \mathrm{p}_y(t)$	instantaneous position of P
$\dot{\mathrm{p}}_x(t), \dot{\mathrm{p}}_y(t)$	instantaneous velocity of P
$\ddot{\mathrm{p}}_x(t), \ddot{\mathrm{p}}_y(t)$	instantaneous acceleration P
x_0, y_0	position of P at time t=0
$x_\mathrm{P}, y_\mathrm{P}$	position of P in scanned image
t_P	time P is scanned

We consider a fiducial point on the target, located at coordinates (x_0,y_0) at time 0. Let the position at time t be expressed by the functions *pxt* and *pyt*. These positions can be expressed using Taylor series, where $\dot{\mathrm{p}}_x(0)$ is the x velocity at time 0, $\ddot{\mathrm{p}}_x(0)$ is the acceleration, and so on:

$$\mathrm{p}_x(t) = x_0 + \dot{\mathrm{p}}_x(0)t + \frac{1}{2}\ddot{\mathrm{p}}_x(0)t^2 + \ldots \tag{4a}$$

$$\mathrm{p}_y(t) = y_0 + \dot{\mathrm{p}}_y(0)t + \frac{1}{2}\ddot{\mathrm{p}}_y(0)t^2 + \ldots \tag{4b}$$

We wish to know the position of the given point, (x_P,y_P), in the acquired image. When the point's trajectory intersects the raster, the time at which the point is scanned, t_P, will be:

$$t_\mathrm{P} = \frac{y_\mathrm{P}}{\mathrm{v}_{\mathrm{S},y}} + \frac{x_\mathrm{P}}{\mathrm{v}_{\mathrm{S},x}}, \tag{5a}$$

$$\approx \frac{y_\mathrm{P}}{\mathrm{v}_{\mathrm{S},y}}. \tag{5b}$$

By making the approximation, we ignore the dependence on horizontal position. This is justified on the grounds that $\mathrm{v}_{\mathrm{S},x}$ is large, and so this term will be small. By definition, $y_\mathrm{P} = \mathrm{p}_y(t_\mathrm{P})$, and so the value of t_P obtained in equation 5b may be substituted into equation 4b, which can then be solved for y_P. The result can then be used to evaluate equation 4a to obtain x_P.

In general, the raster will not pass directly over the point, and features of finite size will often be represented in more than one scan line. We assume that little target motion occurs during a single line time, so the position of a feature located between two scan lines can be accurately determined by interpolation, and results obtained for points lying directly on the raster will hold for all points.

1.3 Example: constant object velocity

$v_{T,x}, v_{T,y}$	target velocity, $\dot{p}_x(t) = v_{T,x}$

We can use the results of the preceding section to generate simulated distorted images for various motions. We consider first the simple case where the target moves with constant velocity, ($v_{T,x}$,$v_{T,y}$):

$$p_x(t) = x_0 + v_{T,x}t, \tag{6a}$$

$$p_y(t) = y_0 + v_{T,y}t. \tag{6b}$$

Following the strategy outlined above, we construct the following equation for y_P:

$$y_P = y_0 + \frac{v_{T,y} y_P}{v_{S,y}}. \tag{7a}$$

$$= \frac{y_0 v_{S,y}}{v_{S,y} - v_{T,y}}, \tag{7b}$$

$$= y_0 + \frac{v_{T,y} y_0}{v_{S,y} - v_{T,y}}. \tag{7c}$$

We can use this result to derive a corresponding expression for x_P:

$$x_P = x_0 + \frac{v_{T,x} y_0}{v_{S,y} - v_{T,y}}. \tag{8}$$

Several important points may be noted from these equations: first, the deviation in feature position for each component is proportional to y_0, the vertical position of the feature in the image, and to the corresponding component of object velocity. We also notice that when $v_{T,y} \geq v_{S,y}$ (object moving faster than the raster), the solution corresponds to a negative value of t, and does not correspond to a point in the current frame.

Distortions arising from a target speed of $\frac{v_{S,y}}{4}$ are illustrated in figure 3. The left-hand patch shows the image obtained when a square grid target is moved at the right, while the right-hand patch shows the image resulting from upward motion.

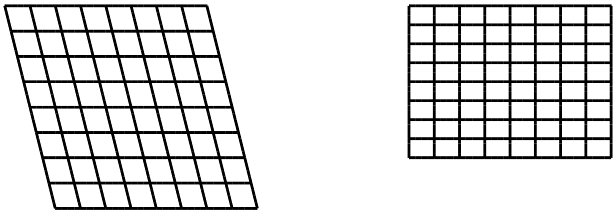

Figure 3: Image distortions of a regular grid for constant velocity motion to the right (left) and upwards (right).

1.4 Example: constant object acceleration

a_x, a_y	target acceleration, $\ddot{p}_x(t) = a_x$

We assume the object accelerates from rest at time 0 with accelerations a_x and a_y:

$$p_x(t) = x_0 + \frac{1}{2} a_x t^2, \tag{9a}$$

$$p_y(t) = y_0 + \frac{1}{2} a_y t^2. \tag{9b}$$

We first consider the case where $a_y = 0$, i.e. a purely horizontal motion. In this case, the vertical position of the fiducial point will not be changed, and the raster will scan the point at time $t_P = \frac{y_0}{v_{S,y}}$. Substituting this value into equation 9a, we obtain:

$$p_x(t_P) = x_0 + \frac{a_x p_y^2(0)}{2v_{S,y}^2}. \tag{10}$$

Equation 10 is quite similar to equation 8, except that here the deviation is proportional to the *square* of the vertical position. This case is illustrated on the left side of figure 4.

The case of vertical accelerations is more complex, due to the interaction between the accelerating motion with the vertical scan. As we did above with equation 7a, we begin by constructing an equation in y_P:

$$y_P = y_0 + \frac{a_y y_P^2}{2v_{S,y}^2}, \tag{11a}$$

which after application of the quadratic formula yields:

$$y_P = \frac{v_{S,y}(v_{S,y} \pm \gamma)}{a_y}, \tag{11b}$$

where

$$\gamma = \sqrt{v_{S,y}^2 - 2a_y y_0}. \tag{12}$$

The smaller of two solutions corresponds to the first coincidence of the raster and the point, while the larger only exists when the acceleration is so large that the point subsequently overtakes the raster. When the acceleration is so large that the raster *never* encounters the point, γ is imaginary.

After more algebra, we can also obtain this result for x_P:

$$x_P = x_0 + \frac{v_{S,y}^2 - a_y y_0 \pm v_{S,y}\gamma}{a_y}. \tag{13}$$

These results are used to compute the images in figure 4. For horizontal acceleration, we see that the vertical grid lines become curved, while for vertical accelerations the grid is compressed nonuniformly.

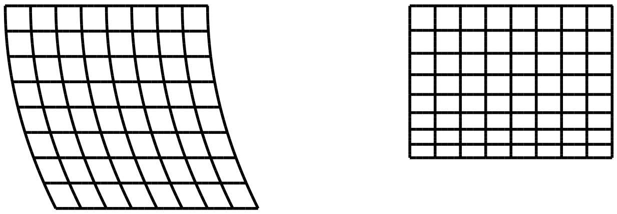

Figure 4: Image distortions resulting from constant acceleration (from rest) of the target to the right (left panel) and upwards (right panel).

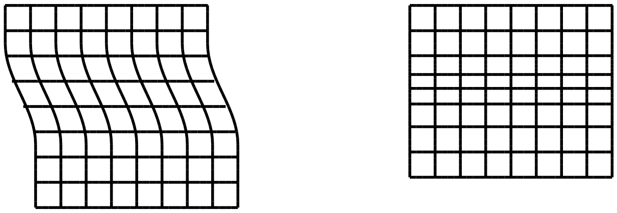

Figure 5: Image distortions resulting from a smoothed step motion to the right (left panel) and upwards (right panel).

In general we observe that horizontal motions generate horizontal shearing of the image proportional to the instantaneous velocity, and that vertical motion similarly generates vertical compression. Figure 5 illustrates the case of a rapid smooth displacement during the middle of the scan. Mathematical details are omitted in the interest of brevity.

2 Methods

2.1 Basic registration methods

The basic registration method, described in detail by Mulligan [2], was developed to process retinal images obtained from a camera-based video ophthalmoscope. It assumes the existence of a *template,* which is a large image of the object. It is assumed that all input images can be registered wholly within the template. Each input image is registered to the template by finding the maximum of the normalized cross correlation. The cross-correlation is computed by taking the product of the Fourier transform of the input and the complex conjugate of the template transform, then taking the inverse Fourier transform. The resulting correlation image is then multiplied by a normalizing image which accounts for the fact that the energy in the template image varies in space. The normalizing image is computed by convolving the pixel-wise squared image of the template with a mask corresponding to the valid input area.

Subpixel interpolation of the correlation maximum is performed by first locating the maximum value in the correlation image, and then performing biquadratic interpolation on the 3 by 3 neighborhood of pixels centered on the maximum. A band-pass filter applied to the input images blurs the the cross-correlation, allowing accurate interpolation. The method has been shown empirically to produce errors less than 0.1 pixel [3, 2]. The peak value of the correlation (which occurs between the sample points) is interpolated using the parameters of the best-fitting quadratic surfaces, computed using the singular value decomposition.

2.2 Motion parameter estimation

Estimation of the target motion profile is done by computing the distortions of the template image resulting from a set of sample guesses. For each guess, the normalized cross correlation between the input image and the distorted template is computed, and the parameter space is searched to find the motion which produces the largest normalized correlation C with the input. We can think of the difference 1-C as representing an ”error” between our guess and the true state of the world, although this may not reach a value of 0 even for the correct motion. The key to a practical solution is minimizing the number of dimensions of the space of possible motions. For example, assuming that the target moved with constant velocity, then there are only two unknown parameters of the warp, $v_{T,x}$ and $v_{T,y}$.

A straightforward but expensive approach is to finely sample the parameter space, and compute the cross-correlation for each candidate motion. The peak value of each correlation is stored in an array, indexed by the motion parameter values. This array can then be examined to find the maximum, corresponding to the best-fitting motion parameters. The space can then be resampled on a finer grid, if desired, to obtain a more precise estimate. The feasibility of such an approach largely depends on the error surface character. If it is smooth, the initial sampling can be quite coarse and yet still provide a good estimate of the maximum through interpolation.

Figure 6 shows velocity space images of the correlation maxima, for uniform translation of a retinal template (shown ahead in figure 8). It can be seen that, at this resolution, the values decrease monotonically with distance from the true parameter values. In such a case, we can obtain reasonable estimates at greatly reduced cost by sampling more coarsely.

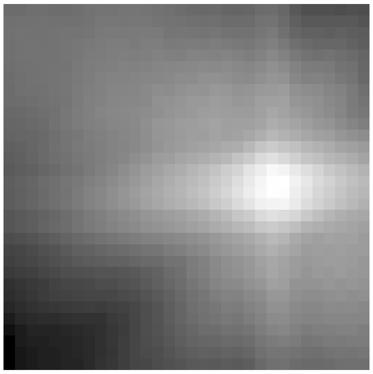

Figure 6: Velocity space image showing the peak value of the normalized cross correlation, evaluated on a 32x32 grid of trial velocities used to warp the template. Input image was warped in accordance with a simulated horizontal motion of 1.25 image widths per frame, sample grid spans the range ±2 image widths per frame. The gray level range has been scaled to span the range of correlation values (0.114 - 0.940).

2.3 Successive refinement

ε_i	*i*th estimation error
Δv_i	*i*th parameter sample spacing

After obtaining an initial estimate of the target motion from a coarse sampling of the parameter space (see section 2.2), we may wish to reduce the estimate error by resampling the parameter space more finely in the neighborhood of our current estimate. In the current implementation, all estimates are obtained by using the biquadratic interpolation procedure (developed to localize cross-correlation peaks) to 3x3 sample arrays from the parameter space error surface. The array is sampled at the current estimate (initially 0), and the flanking samples are separated by Δv. After the initial estimate is obtained, a new 3x3 sampling grid is placed at the location of the current estimate. The sample spacing of the new grid is equal to the previous sample spacing, times a fraction β, which can be thought of as the reciprocal of the ”zoom” factor. If beta is large, i.e. close to 1, the error in the new estimate may not be significantly reduced. If it is too small, however, then the true maximum will lie outside of the region spanned by the sample array. A one-dimensional example with a value of β=0.5 is shown in figure 7.

Experimentation has shown that the interpolation procedure can be unstable when the maximum sample value occurs at the array edge; the samples can be fit with a hyperbolic surface, or an ellipsoidal surface which is concave up. Therefore, if the maximum value is not obtained at the center sample, the array is shifted by one sample until the maximum *does* occur at the center. In two parameter estimation, each lateral or vertical shift requires the computation of 3 additional correlations, while diagonal shifts require 5. (This problem does not arise when interpolating cross-correlation images because the entire correlation image is available: we can perform an exhaustive search for the maximum sample value, and insure that it falls at the center of the interpolation array.)

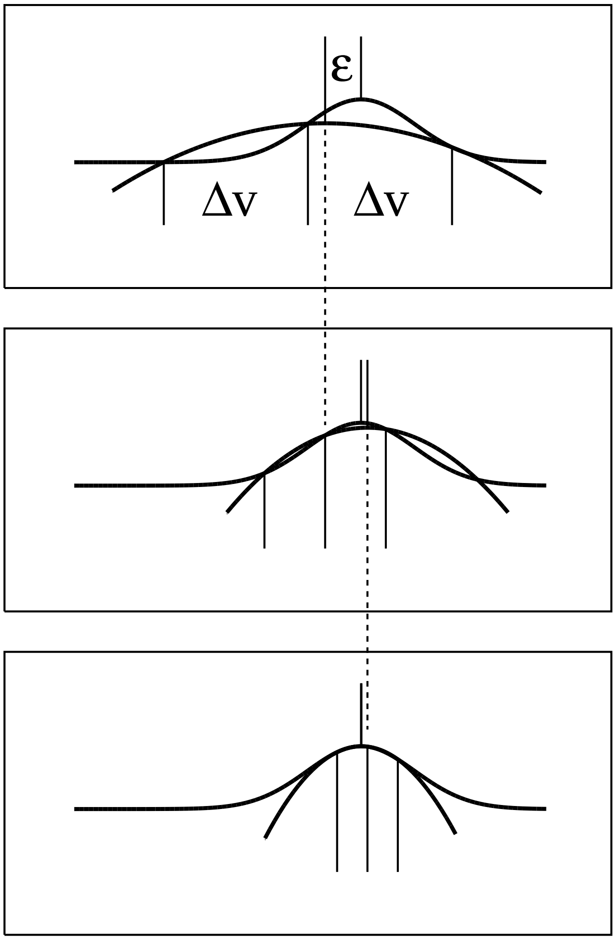

Figure 7: Estimation of the extremal position by quadratic interpolation from 3 samples, illustrated in one dimension. The true underlying error function is a Gaussian. In the initial estimate (top), the quadratic fit is poor, and localization error is significant. This initial estimate is used as the center of a new array of samples, at half the spacing (center). The error is reduced, although the fit is still poor. In the third iteration (bottom), the fit is good and the error is small.

Ideally, we would like for the new sample spacing, Δv_{i+1}, to be approximately equal to the estimation error on the previous iteration, ε_i. While we do not in general have *a priori* information about the nature of ε as a function of Δv, when the error function is smooth the quadratic fit will improve as Δv decreases, and ε will approach zero, or more precisely some small value determined by the numerical precision of the machine and the noise level in the input images. For a particular template image, the dependence of ε on Δv may be precomputed for a set of representative velocities, and the results used to construct a table of β's.

When there are no local extrema on the error surface, this procedure works well. This is an unrealistic assumption, however, because numerical imprecision, input noise, and the structure of the template autocorrelation all add complexity to the error surface. What is needed is the analog of an antialiasing filter in the parameter space domain. Remember that when a signal is sampled at frequency f_S, the signal must first be prefiltered to remove frequencies above the *Nyquist frequency,* $f_S/2$. Failure to do so causes super-Nyquist frequencies to appear as low-frequency ”aliases,” which cannot be distinguished from the actual low frequencies in the input.

When we sample the parameter space at some spacing Δv, we would like to insure that there are no bumps and wiggles occurring between our sample points which will corrupt the interpolation process and cause the procedure to become trapped in a local extremum. The basic idea is similar to the ”coarse-to-fine” approach, proposed for stereo correspondence [4] and motion estimation [5]. In those cases, blurring is performed in the image domain, but here what we would like here is a template transformation which will result in blurring of the error surface. Averaging together the distorted templates corresponding to a neighborhood of the parameter space is not exactly correct because of the (nonlinear) maximum operation that we perform in the construction of the error surface, but something similar may produce a useful result. Rucklidge [6] has described a related approach which is immune to some of the problems of coarse-to-fine strategies, while offering similar computational savings.

2.4 Exploiting features of the cross-correlation

While the successive refinement described in the previous section performs well while computing many fewer correlations than would be required by exhaustive search at the smallest sample spacing, it is still fairly computationally expensive. To reduce the amount of computation required, an intriguing possibility is suggested by the appearance of the correlation images. Figure 8 (top row) shows a portion of a retinal template which we use as our test image, together with a distorted version produced by a simulated eye movement. The third row of figure 8 shows the autocorrelation of the template on the left, together with the cross correlation of the template with the distorted version on the right.

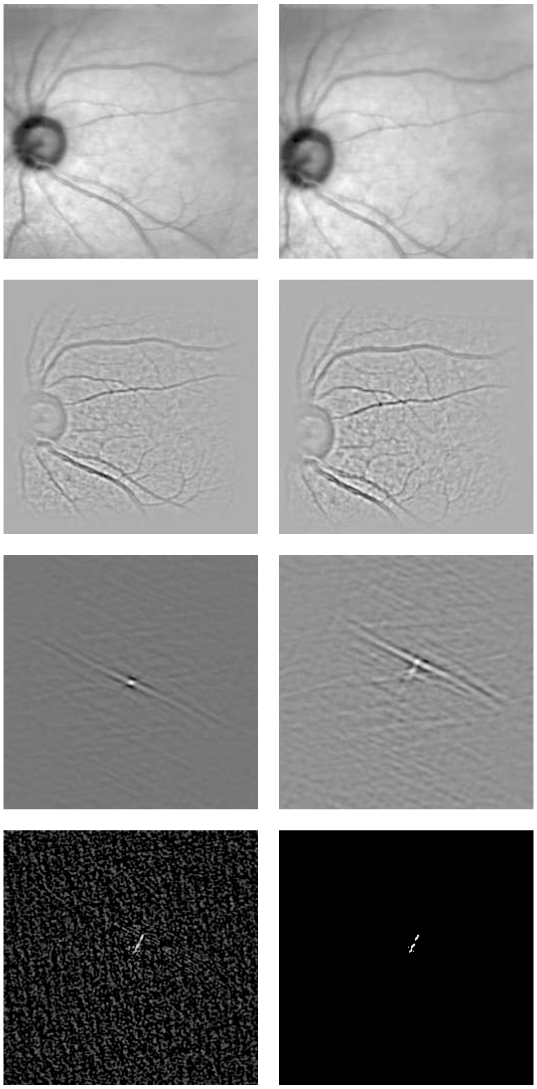

Figure 8: Top row: a portion of a retinal image template (left), together with a version distorted by a simulated eye movement of constant velocity. Second row: the images from the first row are band-pass filtered to accentuate the retinal blood vessels, and windowed to reduce edge artifacts. Third row: the autocorrelation of the template (left), and the cross correlation of the template and the input. Because the gray level range has been matched to the image extreme values, more detail is seen in the cross-correlation image on the right, which has a lower peak value. The long diagonal streak is related to the prominent vessel to the lower right of the optic disk. Bottom row: the cross-correlation is deconvolved with the template autocorrelation (left). The sharp diagonal feature is related to the motion trajectory, and may be identified by thresholding (right).

In the cross-correlation image, the central peak is smeared out in the direction of motion. To see why this should be so, recall that when scanning a moving object, the position in the top of the frame is different from the position in the bottom of the frame. The peak in the cross-correlation image indicates the relative location of the object; as the object moves during the scan, this peak shifts accordingly. When the autocorrelation of a template image patch is roughly independent of the its location (e.g. high-pass noise), the cross correlation image can be thought of as the convolution of the autocorrelation with the motion trajectory. For less uniform images, the trajectory may show gaps, corresponding to portions of the template in which there is little or no energy.

We can attempt to solve for the trajectory by deconvolving the cross-correlation image with autocorrelation of the template; the result of this operation is shown in the bottom row of figure 8. The bottom row of figure 8 also shows the result of clipping the deconvolved image from below at 0, rescaling the positive values to the range 0-1, and thresholding at a value of 0.4. A simple machine vision algorithm might look for a line segment in this image, and obtain a rapid estimate of $v_{T,x}$ and $v_{T,y}$ from the location of the endpoint.

3 Application to SLO images

The registration methods described were developed to track eye movements using images obtained with a camera-based video ophthalmoscope [2], and were subsequently applied to images obtained from a scanning-laser ophthalmoscope (SLO) [7]. The full-field correlation method differs from previous efforts to obtain eye position data from SLO images [8], which relied upon the identification and localization of small discrete features, such as vessel bifurcations. Stetter *et al.* [1] computed high temporal resolution profiles of horizontal saccades from the profiles of a few major blood vessels oriented roughly vertically in the image, exploiting the shear distortion produced by the interaction between the vertical scan and horizontal eye movements. In this section we will begin by considering the problems encountered applying full-field correlation techniques to these images, and discuss the use of image distortions to recover complete two dimensional motion trajectories.

3.1 Template construction

In the preceding section, we assumed the existence of a template image, e.g. a large image of the target object constructed as a mosaic of a large number of input images. Various techniques have been proposed to automate the construction of image mosaics from sets of fundus images [9, 10, 11], but these have generally been more concerned with visualization of gross anatomical features than the preservation of metric structure. Here, templates are constructed from an input sequence using a bootstrap procedure. All images are first band-pass filtered; the high-pass component of the filter serves to accentuate the retinal blood vessels, and remove low-frequency artifacts due to non-uniform illumination of the retina, while the low-pass component attenuates high-frequency camera noise. These images are optionally windowed with a Gaussian-blurred rectangle slightly smaller than the input to reduce edge artifacts. The first preprocessed image is used as the initial template, and the second frame is registered with respect to it. The computed displacement is then applied to the second frame to form a scrolled copy. Sub-pixel translations are performed by appropriate phase shifts in the Fourier domain [12]. The shifted image is then summed into an accumulation buffer, while a similarly shifted mask of 1's is summed into a pixel count buffer. The current template is a weighted average of all of the previously registered frames, computed as the pixel-wise quotient of the accumulation and count images. Frames whose normalized correlation with the template is below a threshold (typically around 0.5) are excluded. A representative template image is shown in figure 9.

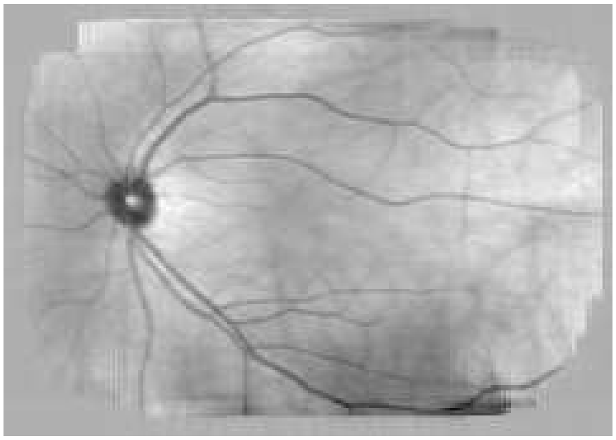

Figure 9: Registration template constructed from 120 SLO field images.

Once a template has been constructed, the registration quality can be assessed by creating a video sequence in which the input images are overlaid on the template, after having been translated to provide the best match to the template. When this is done well, what is seen is a stationary template pattern, with a smaller noisy window moving over it. Registration errors are manifested as motions of the template structure within the moving window, while the position of the window is related to the computed direction of gaze. A static version of this is presented in figure 10, where a single nominally registered input image is superimposed over the template shown in figure 9.

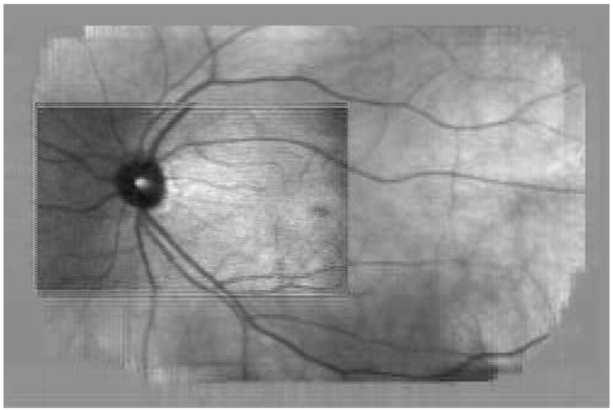

Figure 10: Template image shown in figure 9 overlaid with a single input image. Misregistration can be observed at the upper border, while registration at the right and lower edges appears to be good.

When the quality of registration is visualized as described above, non-rigidity of the template structure is often observed, indicating the occurrence of registration errors. There are two potential sources of these errors. The first is distortion introduced by a moving eye (the topic of this paper). The second results from template construction errors. Stetter *et al.* measured the static SLO distortion by holding a sheet of graph paper in front of the instrument and observing the image transformation resulting from the SLO optics. They used the measured distortion to correct the image feature locations used in their method.

If not corrected for, distortions such as that observed by Stetter *et al.* will corrupt the template construction process, because globally correct matches will not be obtained. The resulting template distortion depends not only on the inherent imaging system distortion, but also on the positions which are sampled and the order in which they used to compute the template. Another potential source of distortion results from the fact that the retina is a spherical surface, which is projected to a plane by the imaging system. This distortion is negligible for excursions of less than 4 degrees [13], but must be incorporated into the registration process for accurate construction of large templates. Error-free templates are critical for obtaining eye movement estimates which are limited only by the noise in the images.

3.2 Parametric models of eye movements

Before attempting to estimate high temporal resolution eye movements from SLO images, it is useful to first consider what is known about eye movements (an excellent survey is provided by Kowler [14]). The highest velocities reached by the eye are during the execution of *saccadic* eye movements, which will therefore be of most relevance to the present work. Saccades are rapid, ballistic changes in fixation, which can be superimposed over other smooth movements. Small saccades which occur during ”steady” fixation are sometimes referred to as *microsaccades.* Lawful relations are observed between saccadic amplitude, duration, and velocity [15, 16, 17], so that saccades in a given direction can be described by a family of functions. A good analytic description of these functions is provided by the density function of the gamma distribution [18], which has three parameters which are related to duration, peak velocity, and skewness or asymmetry. (The situation is more complicated for oblique saccades, where asynchronies between different muscle group activations can produce ”looping” saccades.)

Because the largest saccadic velocities are well below the scan velocities of the SLO, reasonable template registration is obtained even during saccades. Thus, the positions sampled at the frame rate (obtained by correlating the uncorrected images with the template) will allow us to identify the occurrence of large saccades, and for large saccades whose durations spans several frames we can make good predictions about average velocity and acceleration within each frame, speeding the initial portion of the parameter search. Microsaccades are too brief to be detected in this way, but may be revealed by dips in the value of the correlation of the raw input with the template.

3.3 Ocular torsion

In addition to rotating about horizontal and vertical axes to redirect gaze, the eye is also capable of rolling around the line of sight, a movement which is known as *ocular torsion.* There is a systematic variation of torsion with deviation of gaze [19], as well as small fluctuations during fixation [20]. The occurrence (and measurement) of torsional eye movements complicates the registration of fundus images, although a search procedure similar to that described above for finding image warps has been able to recover simulated torsions with an accuracy of around 0.1 degree [2], which is comparable to results obtained from video images of the iris [21, 22, 23, 24]. Ott and Eckmiller [24] used SLO images to measure ocular torsion during smooth pursuit by measuring the change in the slopes of lines connecting image features. They do not mention the existence and correction of the scan induced shear in the moving target image, which is significantly smaller than the measured effects, but significantly larger than the attainable measurement precision.

Fortunately, torsional eye movements are relatively slow, so that there is very little rotation during the scan of a single frame. Furthermore, this slowness means that the torsional state of the eye can be well predicted from the state in the preceding frames. Therefore, the only distortions which have been considered in this paper are those resulting from pure translation of the entire target, although the methods are completely general.

3.4 Superresolution

When the target contains frequencies above the Nyquist limit of half the sampling rate which are passed by the point spread function of the imaging system, *aliasing* will occur. This term refers to the fact that samples from a signal above the Nyquist frequency are indistinguishable from samples from a corresponding signal composed entirely of frequencies below the Nyquist frequency. This is true for single images, but not strictly so when we consider multiple images with slight positional offsets. When the target moves by a small amount, all of the sub-Nyquist frequencies in the image move by the same amount. Aliases of super-Nyquist frequencies, on the other hand, have a different motion, like that of a Moire pattern. Reconstructing the super-Nyquist frequencies from multiple samples of the aliases is known as *superresolution* [25, 26, 27] (the term has also appeared in the optics literature to describe physical situations in which the resolution is increased by a factor of 2 or $\sqrt{2}$ above the nominal diffraction limit [28]).

Can we apply superresolution techniques to SLO images? In other words, are there any aliases of retinal image frequencies above the Nyquist limits imposed by the SLO raster? The following anecdotal observation suggests that the answer is yes: Stevenson [29] has noted that the individual raster lines of the SLO can be resolved by the eye (of the subject). This means that the spot size of the illumination beam is smaller than the line spacing. The retinal image is effectively low pass filtered by a filter whose kernel is the spot profile of the laser beam convolved with the eye's optical point spread function. This low-pass filter functions as an effective anti-aliasing filter only when its width is comparable to the line spacing. When the raster lines are clearly resolved, two images whose vertical position differs by half the line spacing can be viewed as interlaced components of an image with twice the vertical resolution, and by analogy a sequence of slightly-misregistered images can be combined to produce a high-resolution image, provided the input images can be accurately registered.

3.5 Application to laser surgery

There is a current need for real-time retinal tracking and stabilization in the medical community. Laser surgery consists of delivering a brief, high energy pulse to a small retinal area. The target area is selected by the physician, who presses a button indicating that an aiming cursor has been placed over the target area in an ophthalmoscopic retinal image. There is a small, but finite, time window (a few hundred milliseconds) between the physician's decision that the cursor is placed, and the actual delivery of the laser pulse. During this interval, the patient could make an eye movement causing the pulse to be delivered to the wrong part of retina, possibly injuring the fovea or optic nerve head! A safer system would would track eye movements and permit the physician to indicate the target location on a stabilized image, while stabilization optics keep the laser accurately targeted in the presence of patient eye movements.

A nearly real-time stabilizer might be constructed by registering individual SLO raster lines to the template as they become available from the instrument, eliminating the 16 msec pipeline delay needed to buffer an entire video field. Because the video line rate of 15 kHz is much higher than the bandwidth of natural eye movements, it should be possible to make accurate predictions by correlating each line with just a few lines of the template in the neighborhood of the expected position. Computing a small number of 1-D correlations between selected lines is considerably less computation than the complete 2-D cross correlation, and is likely to be tractable with modern digital signal processing circuits. The problem is complicated somewhat if one wishes to measure ocular torsion, although this may not be necessary to produce a useful surgical stabilizer.

4 Conclusions

While more work is needed to refine motion parameter estimation to extract high time resolution eye movement data from SLO image sequences, these results suggest that the technique is practical. Increased temporal resolution, combined with the high precision made possible by sub-pixel interpolation, make this approach competitive with other high-resolution techniques, such as the more invasive search coil [30].

5. Acknowledgements

Special thanks to Tina Beard for comments on early drafts of the manuscript and help with the figures. This work was supported by NASA grant 199-06-12-39, and NASA Aeronautics RTOP's 505-64-53 and 548-60-12. SLO imagery was kindly provided by David Grosof, Washington University and Central Institute for the Deaf, whose work received support from the J. Epstein Foundation, Jewish Hospital Research Fund, St. Louis MO, and an unrestricted grant from Research to Prevent Blindness to Washington University Ophthalmology.

5. Appendix: Minutiae of raster scanning

τ_A	active time
τ_R	retrace duration
α_L	line scan duty cycle, $\tau_{AL} f_L$
α_F	frame scan duty cycle, $\tau_{AF} f_F$

In our discussion of raster scanning, we ignored the fact that raster retrace requires a finite duration, and stated that the scan velocities (in image widths per second) were approximately equal to the scan frequencies in Hz. In this appendix we will present a more precise formula for the scan velocities, and discuss some additional issues concerning raster geometry.

Typically, image data are acquired, or displayed, only during a portion of the scan periods, called the *active time,* of duration τ_A; the remainder of each scan period is spent returning the beam to the starting position, known as *retrace,* of duration τ_R. These quantities are shown with respect to a typical scan waveform in figure 11. In the raster diagram in figure 1 (section 1.1), the active portions of the scan lines are shown as solid lines, while the retrace portions are shown as dashed lines. The *duty cycle* α of the active segment for each scan is defined as the product of the active time and corresponding scan frequency:

$$\alpha_F = \tau_{AF} f_F, \qquad \text{and} \qquad \alpha_L = \tau_{AL} f_L. \qquad (14a,b)$$

The *scan velocities,* $v_{S,x}$ and $v_{S,y}$, are equal to the corresponding scan frequency divided by the corresponding duty cycle times the image width (equal to 1 due to our choice of units):

$$v_{S,x} = \frac{f_L}{\alpha_L}, \qquad v_{S,y} = \frac{f_F}{\alpha_F}. \qquad (15a,b)$$

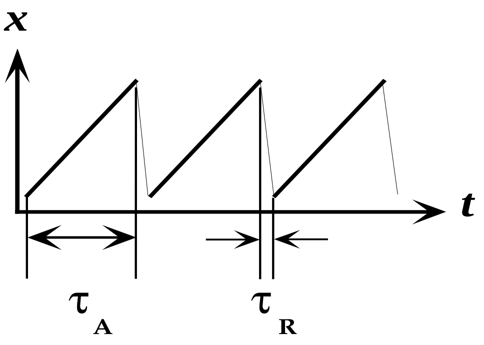

Figure 11: Diagram of raster waveform showing active time τ_A and retrace time τ_R.

In the television standard defined by the National Television Standards Committee (NTSC) [31, 32], the scan frequencies have a value of $f_F \approx 60$ Hz, and $f_L = 262.5\ f_F$. The significance of the horizontal frequency being a half-integral multiple of the frame rate is that two successive ”frames” are vertically offset by half the line spacing, known as *vertical interlace* (see figure 12). These two half-frames are referred to as *odd* and *even fields,* determined by the parity of the line numbers. Treating fields as ”frames” is equivalent to superimposing vertical square wave motion on the target at half the field rate, with amplitude of 0.5 line, and so may easily be corrected *post hoc.*

The raster shown in figure 1 depicts a continuous vertical scan, resulting in the image samples being taken from slightly oblique lines. Images manipulated in a digital frame buffer, however, are often assumed to be samples obtained from a rectangular sampling grid, corresponding to a discontinuous vertical scan (see figure 13). In this case,

$$s_y(t) = t_S v_{S,y}. \qquad (16)$$

The small errors introduced by this approximation vanish as $f_L >> f f$; when necessary they may be corrected by resampling an appropriately sheared version of the image.

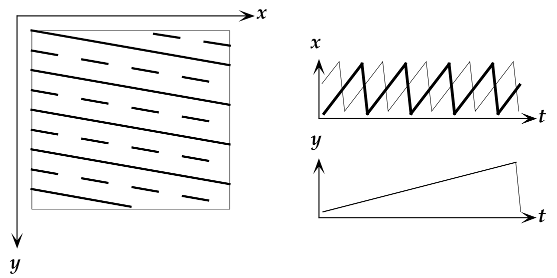

Figure 12: Raster diagram similar to figure 1, showing an interlaced scan.

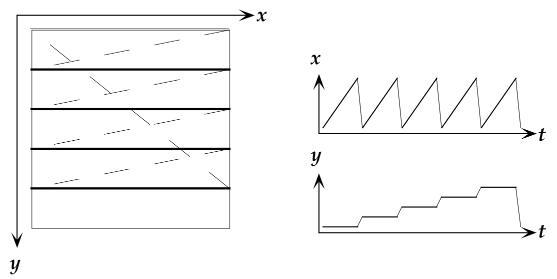

Figure 13: Diagram showing a rectangular raster approximation, and the associated waveforms.

References

[1] Stetter, M., Sendtner, R. A., and Timberlake, G. T., “A novel method for measuring saccade profiles using the scanning laser ophthalmoscope,” *Vision Research* **36**, 1987–1994 (1996).

[2] Mulligan, J. B., “Image processing for improved eye tracking accuracy,” *Behav. Res. Methods Instrum. Comput.* **29**, 54–65 (1997).

[3] Mulligan, J. B. and Beutter, B. R., “Eye-movement tracking using compressed video images,” **1**, 163–166 (1995).

[4] Marr, D. and Poggio, T., “A computational theory of human stereo vision,” *Proc. Roy. Soc. Lond. B* **204**, 301–328 (1979).

[5] Anandan, P., “A computational framework and an algorithm for the measurement of visual motion,” *International Journal of Computer Vision* **2**, 283–310 (1989).

[6] Rucklidge, W., “Efficient guaranteed search for gray-level patterns,” in [*Proc. CVPR*], 717–723, IEEE Computer Society Press (1997).

[7] Webb, R. H., Hughes, G. W., and Delori, F. C., “Confocal scanning laser ophthalmoscope,” *Applied Optics* **26**, 1492–1499 (1987).

[8] Wornson, D. P., Hughes, G. W., and Webb, R. H., “Fundus tracking with the scanning laser ophthalmoscope,” *Applied Optics* **26**, 1500–1504 (1987).

[9] Peli, E., Augliere, R. A., and Timberlake, G. T., “Feature-based registration of retinal images,” *IEEE Transactions on Medical Imaging* **MI-6**, 272–278 (1987).

[10] Cideciyan, A. V., “Registration of ocular fundus images,” *IEEE Engineering in Medicine and Biology* , 52–58 (1995).

[11] Mahurkar, A. A., Vivino, M. A., Trus, B. L., Kuehl, E. M., III, M. B. D., and Kaiser-Kupfer, M. I., “Constructing retinal fundus photomontages,” *Investigative Ophthalmology & Visual Science* **37**, 1675–1683 (1996).

[12] Bracewell, R. R., [*The Fourier transform and its applications*], McGraw-Hill (1965).

[13] Ott, D. and Daunicht, W. J., “Eye movement measurement with the scanning laser ophthalmoscope,” *Clinical Vision Science* **7**, 551–556 (1992).

[14] Kowler, E., “The role of visual and cognitive processes in the control of eye movement,” in [*Eye movements and their role in visual and cognitive processes*], Kowler, E., ed., 1–70, Elsevier Science (1990).

[15] Ditchburn, R. W., [*Eye-movements and visual perception*], Clarendon Press (1973).

[16] Boghen, D., Troost, B. T., Daroff, R. B., Dell’Osso, L. F., and Birkett, J. E., “Velocity characteristics of normal human saccades,” *Investigative Ophthalmology* **13**, 619–622 (1974).

[17] Bahill, A. T. and Stark, L., “The trajectories of saccadic eye movements,” *Scientific American* **240**(1), 108–117 (1979).

[18] Opstal, A. J. V. and Gisbergen, J. A. M., “Skewness of saccadic velocity profiles: A unifying parameter for normal and slow saccades,” *Vision Research* **27**, 731–745 (1987).

[19] Nakayama, K., “A new method of determining the primary position of the eye using listing’s law,” *American Journal of Optometry and Physiological Optics* **55**, 331–336 (1978).

[20] van Rijn, L. J., van der Steen, J., and Collewijn, H., “Instability of ocular torsion during fixation: cyclovergence is more stable than cycloversion,” *Vision Research* **34,**, 1077–1087 (1994).

[21] Curthoys, I. S., Moore, S. T., McCoy, S. G., Halmagyi, G. M., Markham, C. H., Diamond, S. G., Wade, S. W., and Smith, S. T., “Vtm – a new method of measuring ocular torsion using image-processing techniques,” *Annals of the New York Academy of Sciences* **656**, 826–828 (1992).

[22] Bos, J. E. and de Graaf, B., “Ocular torsion quantification with video images,” *IEEE Transactions on Biomedical Engineering* **41**, 351–357 (1994).

[23] Groen, E., Nacken, P. F. M., Bos, J. E., and de Graaf, B., “Determination of ocular torsion by means of automatic pattern recognition,” *IEEE Transactions on Biomedical Engineering* **43**, 471–479 (1996).

[24] Ott, D. and Eckmiller, R., “Ocular torsion measured by tv- and scanning laser ophthalmoscopy during horizontal pursuit in humans and monkeys,” *Investigative Ophthalmology & Visual Science* **30**, 2512–2520 (1989).

[25] Irani, M. and Peleg, S., “Improving resolution by image registration,” *CVGIP: Graphical Models and Image Processing* **53**, 231–239 (1991).

[26] Mann, S. and Picard, R., “Virtual bellows: constructing high quality stills from video,” *Proc. ICIP* (1994).

[27] Cheeseman, P., Kanefsky, B., Kraft, R., Stutz, J., and Hanson, R., [*Super-resolved surface reconstruction from multiple images*], Kluwer (1996).

[28] Cox, I. J. and Sheppard, C. J. R., “Information capacity and resolution in an optical system,” *Journal of the Optical Society of America A* **3**, 1152–1158 (1986).

[29] Stevenson, S. B., [*personal communication*] (1995).

[30] Robinson, D. A., “The mechanics of human saccadic eye movement,” *Journal of Physiology* **174**, 245–264 (1964).

[31] Loughren, A. V., “Recommendations of the national television system committee for a color television signal,” *J. Soc. Motion Picture and Television Eng.* **60**, 321–336,596 (1953).

[32] Poynton, C. A., [*A technical introduction to digital video*], John Wiley and Sons (1996).