Image Motion Estimation From Motion Smear—A New Computational Model

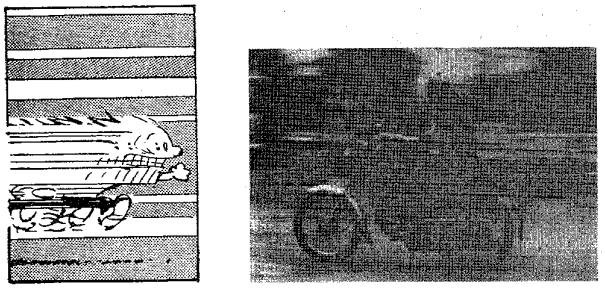
Wei-Ge Chen, N. Nandhakumar, *Senior Member, IEEE*,
and Worthy N. Martin, *Senior Member, IEEE*

Abstract—Motion smear is an important visual cue for motion perception by the human vision system (HVS). However, in image analysis research, exploiting motion smear has been largely ignored. Rather, motion smear is usually considered as a degradation of images that needs to be removed. In this paper, we establish a computational model that estimates image motion from motion smear information—"motion from smear." In many real situations, the shutter of the sensing camera must be kept open long enough to produce images of adequate signal-to-noise ratio (SNR), resulting in significant motion smear in images. We present a new motion blur model and an algorithm that enables unique estimation of image motion. A prototype sensor system that exploits the new motion blur model has been built to acquire data for "motion-from-smear." Experimental results on images with both simulated smear and real smear, using our "motion-from-smear" algorithm as well as a conventional motion estimation technique, are provided. We also show that temporal aliasing does not affect "motion-from-smear" to the same degree as it does algorithms that use displacement as a cue. "Motion-from-smear" provides an additional tool for motion estimation and effectively complements the existing techniques when apparent motion smear is present.

Index Terms—Motion estimation, motion smear (blur), motion-from-smear.

1 Introduction

In this paper, we report the investigation of an important visual cue—motion smear or blur—for image motion estimation. The HVS can perceive motion when presented with an image that is smeared due to motion. This fact is commonly known and has always been exploited by artists, photographers, etc., to indicate speed in their work (Fig. 1). The relative motion of the background in Fig. 1b can be easily perceived by the viewer merely from the motion smear. The greater the smear, the greater is our sensation of motion. Psychophysical and biological investigations have indicated that the HVS and some other biological vision system may have mechanisms to perceive motion from image smear (see, e.g., [1], [2]). In the context of image analysis, however, motion smear has always been considered as a degradation on images. Little has been done to exploit the positive side of this phenomenon, i.e., to use motion smear as a visual cue for image motion estimation, which is the subject of this research. We establish a computational model that functionally emulates the behavior of the HVS, i.e., estimates image motion from motion smear information or, in short, "motion-from-smear."



(a) (b)

Fig. 1. Examples of motion smear in images that provide a sensation of motion to the observer. (a) A cartoon that uses a high degree of smear to indicate high speed. (b) Image of a motorcycle on the road, with a motion smeared background. Motion smear is due to camera panning.

Previous research in motion estimation has assumed that each image (in a motion sequence) is acquired by keeping the shutter open for an infinitesimally short duration of time, i.e., the shutter speed is infinitely fast. Although this assumption is satisfied in many cases where image plane velocities are small relative to the image integration duration, it fails to hold in a number of situations—especially when the available illumination levels are low. In such situations the shutter must be kept open long enough to ensure the integration of sufficient number of photons per pixel to produce images of adequate SNR. There are many applications that do not permit increasing the level of illumination, e.g., astronomy, defense (underwater as well as terrain-based) applications, imaging photoluminescent biological objects or high-speed subatomic particles, etc. In such situations, sensing of moving objects results in significant blur. For example, a scene at a distance of 10m viewed by 6mm × 8mm, 640 ×

- *W.-G. Chen is with Microsoft Corporation, One Microsoft Way, Redmond, WA 98052. E-mail: wchen@microsoft.com.*
- *N. Nandhakumar is with the Department of Electrical Engineering and W.N. Martin is with the Department of Computer Science, University of Virginia, Charlottesville, VA 22903-2442*
  E-mail: {nandhu, martin}@virginia.edu.

Manuscript received Nov. 10, 1994; revised Nov. 30, 1995. Recommended for acceptance by A. Singh.
For information on obtaining reprints of this article, please send e-mail to: transactions@computer.org, and reference IEEECS Log Number P95179.

480 pixel^2 sensor with an 8mm lens mounted on a vehicle moving at 50MPH causes an image plane blur of 10 pixels when a shutter speed of 5ms is used. Indeed, in many low light applications this shutter speed of 5ms may result in an unacceptably noisy image. Thus, in most practical applications, available sensor technology places constraints between sensor conversion efficiency/sensitivity, maximum illumination level, and shutter speed. This practical limitation causes significant image blur that affects the performance of motion estimation. Past techniques in motion estimation have not considered this issue, resulting in techniques that perform well in a controlled environment—but not in practical low light and/or high speed applications. We devise an approach that models and exploits image blur information to overcome this limitation.

1.1 Related Work

1.1.1 Psychological and Biological Study

Psychological investigation has shown that the HVS integrates retinal images for 120 milliseconds to enhance SNR [1], [3], [4]. Due to the integration, motion smear is inevitable. It has been reported that when the HVS is presented with an image blurred due to motion, the amount of smear perceived by human observers increases with observing duration at short durations (up to 20-30ms), but at longer durations the perceived smear actually decreases with observing duration, i.e., the perceived image becomes sharper [1]. It is our conjecture that the HVS is performing a deblurring, or sharpening, function on the image. Deblurring, however, requires knowledge of the motion parameters. Hence, we are of the opinion that motion smear information is probably used for motion sensation and motion deblurring in the HVS.

In addition, biologists have discovered that the European common toad (*Bufo Bufo*) integrates retinal images for a remarkable long duration (1.0-2.5 seconds) in order to enhance SNR of retinal images under illumination levels so low that human observers cannot see anything [2], [5]. The long integration poses two problems for toads in prey-catching. First, the image of a moving prey will appear smeared on the toad's retina. Secondly, the long integration will create a significant delay (three seconds). How can the toad catch a moving worm if the motion information of the pray takes such a long time to reach the brain? Experimental results indicate that the strategy by the toad is to predict the motion of the prey and hit the prey in the anticipated position [2], [5]. Some early processing mechanism enables the toad to estimate the moving trajectory of the prey from motion smear. In his study of cortical connections, Ballard also speculates on the role of smear in motion perception [6]. A neuromechanism for such smear-unsmear behavior of the HVS is discussed in [7].

1.1.2 Image Blur

A great deal of research in image restoration has been conducted on the closely related problem of reconstructing the "original" image from images corrupted by noise, blur, or both. The blur considered may be caused by defocus, optical distortions, object motion, or atmospheric turbulence. A preliminary step in image restoration, namely blur identification, is to identify the blurring process that causes the image degradation (see, e.g., [8], [9], [10]). Usually, the blurring process is modeled as a linear space-invariant system. Blur identification is then equivalent to the identification of the point spread function (PSF) of the linear space-invariant system. A classical technique for blur PSF identification is to identify zeros on the unit circle in the Z domain for some presumed functional forms of the PSF (see, e.g., [8]). This method is limited by the fact that the zeros disappear in the presence of noise and/or even slight deviation of the PSF from the assumed functional form.

When the "original" image can be modeled as an autoregressive (AR) process, another blur identification technique is applicable. The blurring process is modeled as a moving average (MA) process and the blurred image can then be modeled as an ARMA process. Blind system identification techniques can be applied to estimate the ARMA parameters from a single degraded image (see, e.g., [8], [11], [10]). The drawback of estimating blur PSF from a single blurred image is the inherent ambiguity in the blur PSF estimation. For instance, different combinations of scene texture and motion can generate the same "blurred" image. Hence, the inverse problem of estimating motion PSF is ill-posed.

Our "motion-from-smear" algorithm uses two successive frames in a sequence of smeared images. Furthermore, the images are acquired using a special sensor that is designed to produce a stable inverse system of the motion PSF. These two factors eliminate the above problem of ambiguity in motion estimation.

Although both—our formulation of "motion-from-smear" and the blur identification problem in image restoration—estimate motion PSF, our goal is image motion estimation while the latter is to facilitate image restoration. Consequently, our methodology is different from blur identification in several aspects:

- image data acquisition is controlled to ensure that the blur PSF is of a specific parametric form.
- a special sensor design allows derivation of stable well-behaved algorithms.
- unique motion estimates are computed.
- we do not adopt any model, such as AR, for the original image that may restrict the applicability of the formulation.

1.1.3 Image Motion Estimation

Existing techniques are based on two approaches: gradient-based and feature-matching [12]. Different variants exist for each approach, e.g., spatio-temporal approach for gradient-based techniques and recursive region matching for feature-matching techniques. Most gradient-based techniques depend on the relationship between the spatial (image plane) derivatives and temporal derivatives in time-varying imagery. In order to apply gradient-based techniques to practical situations involving noninfinitesimal sampling duration in both space and time, the image motions must be small and apparent texture must be fine grained. (Note that "noninfinitesimal," "small," and "fine" are not absolute terms, but are used to indicate that these attributes are

interrelated in gradient-based techniques.) Unfortunately, smaller motions and finer grained textures are difficult to detect reliably in the presence of image noise. Simple noise reduction operations, such as averaging (spatial and temporal), can severely limit the applicability of gradient-based techniques.

In this regard, feature-based techniques have the advantage of depending on complex image structures that can be detected in conjunction with noise suppression operations. Further, they can be applied to long-range motion situations, i.e., when the image plane velocities are large. With large image velocities, however, reliable detection of the complex image structures indicative of salient features requires the imaging system to be capable of near-infinitesimal temporal sampling.

"Motion-from-smear" is fundamentally different from the families of existing motion estimation methods. It uses an alternative visual cue—motion smear, while all other methods rely, directly or indirectly, on the conventional cue—displacement. This fundamental difference causes "motion-from-smear" to behave differently from the families of existing techniques and thus complements the latter. "Motion-from-smear" provides techniques applicable to motion of longer range than gradient-based techniques and, at the same time, applicable to imagery obtained via practical shutter speeds.

In the rest of the paper, we first introduce a data acquisition scheme and a model for motion smear. Based on this, a "motion-from-smear" algorithm is developed. We then discuss some implementation related issues for the algorithm, followed by preliminary experimental results and descriptions of our future effort.

2 Problem Formulation

2.1 Image Sequence Acquisition

Suppose we have an image $f(x, y; t)$ which denotes an "ideal," unblurred image at time t. Further, suppose we have an imaging system with a noninstantaneous shutter mechanism, i.e., a nonzero exposure time Δt, and denote the observed image by $g(x, y; t, \Delta t)$. These image system assumptions imply that observations can be made at only discrete times. Thus, $g(x, y; t_i, \Delta t_{i,i+1})$ is obtained by the shutter opening at time t_i and the sensor integrating the signal until time $t_{i+1} = t_i + \Delta t_{i,i+1}$ for each (x, y). The subscripts of Δt denote the time when the shutter opens and the time when the shutter closes. We assume that the next observation begins immediately, so that at time $t_{i+2} = t_{i+1} + \Delta t_{i+1,i+2}$, $g(x, y; t_{i+1}, \Delta t_{i+1,i+2})$ is available for analysis. Fig. 2 illustrates the situation.

2.2 A Linear System Model For Motion Smear

Motion smear can be modeled by integration on the image (sensor) plane, i.e.,

$$g(x, y; t_i, \Delta t_{i,i+1}) = \int_{t_i}^{t_{i+1}} f(x, y; t) l(t; t_i, \Delta t_{i,i+1}) dt, \quad (1)$$

where $l(t; t_i, \Delta t_{i,i+1})$ is the normalized sensor gain due to time-varying camera/film sensitivity and/or transmission

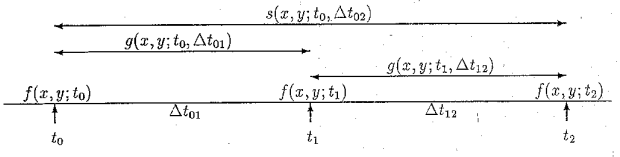


Fig. 2. Graphical description of data acquisition. The original unblurred image at time t_i is denoted by $f(x, y, t_i)$. The shutter remains open for time interval $\Delta t_{i,i+1}$. The blurred image is denoted by $g(x, y; t_i, \Delta t_{i,i+1})$. $s(x, y; t_i, \Delta t_{i,i+2})$ would be the image acquired if the camera shutter were kept open from t_i to t_{i+2}.

coefficient of the lens/filter system. If we assume that the image motion is translational, i.e.,

$$f(x, y; t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y'; t_i)\delta\big(x - v_x(t - t_i) - x', y - v_y(t - t_i) - y'\big)dx'dy', \quad (2)$$

where $t \in (t_i, t_{i+1}]$ and $\mathbf{v} := (v_x, v_y)$ is the image motion vector at (x, y), then substituting (2) into (1) yields

$$\begin{aligned} &g(x, y; t_i, \Delta t_{i,i+1}) = \\ &\int_{t_i}^{t_{i+1}} \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y'; t_i)\delta\big(x - v_x(t - t_i) - x', y - v_y(t - t_i) - y'\big)dx'dy' \right) \\ &l(t; t_i, \Delta t_{i,i+1})dt \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y'; t_i) \left(\int_{t_i}^{t_{i+1}} \delta\big(x - v_x(t - t_i) - x', y - v_y(t - t_i) - y'\big) l(t; t_i, \Delta t_{i,i+1})dt \right) \\ &dx'dy'. \end{aligned}$$

Let

$$h(x, y, x', y'; t_i, \Delta t_{i,i+1}; \mathbf{v}) := \int_{t_i}^{t_{i+1}} \delta\big(x - v_x(t - t_i) - x', y - v_y(t - t_i) - y'\big) l(t; t_i, \Delta t_{i,i+1})dt,$$

then (3) simplifies to

$$g(x, y; t_i, \Delta t_{i,i+1}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y'; t_i) h(x, y, x', y'; t_i, \Delta t_{i,i+1}; \mathbf{v})dx'dy'. \quad (3)$$

Another way of writing $h(x, y, x', y'; t_i, \Delta t_{i,i+1}; \mathbf{v})$ is

$$h(x, y, x', y'; t_i, \Delta t_{i,i+1}; \mathbf{v}) = \begin{cases} l(t; t_i, \Delta t_{i,i+1}) & \text{if } (x', y') = (x, y) - (v_x, v_y)(t - t_i), t_i < t \le t_{i+1}, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

This is a generalization of the linear system model for motion smear that has been widely used in image restoration (see, e.g., [13]). Note $l(t; t_i, \Delta t_{i,i+1})$ is sensor (camera) dependent and, thus, known a priori, while $\mathbf{v}$ is the parameter to be estimated. The motion vector $\mathbf{v}$ is known if $h(x, y, x', y'; t_i, \Delta t_{i,i+1}; \mathbf{v})$ is available because the support of $h(x, y, x', y'; t_i, \Delta t_{i,i+1}; \mathbf{v})$ is uniquely determined by $\mathbf{v}$. We can rewrite the support of $h(x, y, x', y'; t_i, \Delta t_{i,i+1}; \mathbf{v})$ as $y - y' = (x - x')\tan\phi$ where $x' \in (x - L, x]$, $L > 0$ or $x' \in (x, x - L]$, $L \le 0$, $L = v_x \Delta t_{i,i+1}$ and $\phi = \arctan\frac{v_y}{v_x}$. The motion estimation problem, then, is equivalent to recovering $\{L, \phi\}$ at each (x', y'), given blurred images.

Note that, in general, $h(x, y, x', y'; t_i, \Delta t_{i,i+1}; \mathbf{v})$ is space-varying because the motion field is, in general, space-varying. In the following derivation of the "motion-from-smear" algorithm, however, we make the assumption that $h(x, y, x', y'; t_i, \Delta t_{i,i+1}; \mathbf{v})$ is space-invariant (only depends on $(x - x', y - y')$). The space-invariant formulation may be applied locally to images that may have space-varying motion under the condition that the motion is slowly space-varying. Also, in the above model (3) and (4), we have implicitly assumed that $\mathbf{v}$ does not change during $(t_i, t_{i+1}]$.

Consider two successive frames when $i = 0, 1$. We further assume that $\mathbf{v}$ is constant over time intervals on the order of $\Delta t_{01} + \Delta t_{12}$.

Using the space-invariant linear system model, we have

$$g(x, y; t_0, \Delta t_{01}) = f(x, y; t_0) * h(x, y; t_0, \Delta t_{01}; \mathbf{v}) := f(x, y; t_0) * h_{01}(x, y; \mathbf{v}). \quad (5)$$

and

$$g(x, y; t_1, \Delta t_{12}) = f(x, y; t_1) * h(x, y; t_1, \Delta t_{12}; \mathbf{v}) := f(x, y; t_1) * h_{12}(x, y; \mathbf{v}). \quad (6)$$

In the space-invariant case, (2) simplifies to

$$f(x, y; t_1) = f(x, y; t_0) * \delta(x - v_x \Delta t_{01}, y - v_y \Delta t_{01}). \quad (7)$$

In view of (5), (6), and (7), we have

$$s(x, y; t_0, \Delta t_{02}) := g(x, y; t_0, \Delta t_{01}) + g(x, y; t_1, \Delta t_{12}) \quad (8)$$
$$= f(x, y; t_0) * \left[h_{01}(x, y; \mathbf{v}) + h_{12}(x, y; \mathbf{v}) * \delta(x - v_x \Delta t_{01}, y - v_y \Delta t_{01})\right]$$
$$:= f(x, y; t_0) * h_{02}(x, y; \mathbf{v}).$$

Physically, $s(x, y; t_0, \Delta t_{02})$ corresponds to the blurred image acquired if the camera shutter were kept open from t_0 to t_2. We can consider $s(x, y; t_0, \Delta t_{02})$ as an output from a linear system with input $f(x, y; t_0)$ and the PSF $h_{02}(x, y; \mathbf{v})$ (Fig. 3).

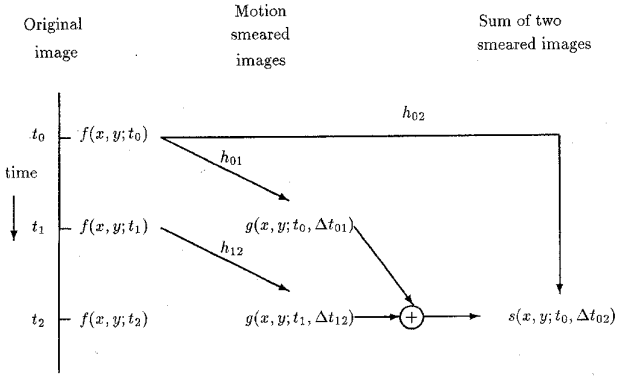


Fig. 3. The meaning of $s(x, y; t_0, \Delta t_{02})$ and $h_{02}(x, y; \mathbf{v})$.

2.3 A Simple Algorithm for Motion-from-Smear

We use the above model to develop a simple algorithm for estimating motion from smear. In later sections, we provide refinements to overcome several of its deficiencies. Given the observed, smeared images $g(x, y; t_i, \Delta t_{i,i+1})$, $i = 0, 1$, we can restore the original image $f(x, y; t_0)$ either:

1) from $g(x, y; t_0, \Delta t_{01})$ with the transformation which is the inverse of that corresponding to $h_{01}(x, y; \mathbf{v})$, or
2) from $s(x, y; t_0, \Delta t_{02})$ with the transformation which is the inverse of that corresponding to $h_{02}(x, y; \mathbf{v})$.

$h_{01}(x, y; \mathbf{v})$ and $h_{02}(x, y; \mathbf{v})$ are commonly constrained by motion and so are their inverse systems. If we hypothesize a motion, then the PSF $h_{01}(x, y; \mathbf{v})$, the restoration of the image $g(x, y; t_0, \Delta t_{01})$, the PSF $h_{02}(x, y; \mathbf{v})$ and the restoration of the image $s(x, y; t_0, \Delta t_{02})$ are all constrained by the hypothesized motion. When the hypothesis is correct, the two restored images will be identical, otherwise the error between the two restored images will be large. The true motion can be found by searching for the minimum of the error, i.e., we can estimate $\mathbf{v}$ by:

$$\hat{\mathbf{v}} = \arg \min_{\mathbf{v}} \mathcal{E}(\mathbf{v}) \quad (9)$$

$$= \arg \min_{\mathbf{v}} \left(\left\| s(x, y; t_0, \Delta t_{02}) * \tilde{h}_{02}(x, y; \mathbf{v}) - g(x, y; t_0, \Delta t_{01}) * \tilde{h}_{01}(x, y; \mathbf{v}) \right\|^2 \right) \quad (10)$$

where $\tilde{h}_{02}(x, y; \mathbf{v})$ and $\tilde{h}_{01}(x, y; \mathbf{v})$ can be implemented by inverse filters in the frequency domain. In order to find the specific forms of the inverse filters, it suffices to find the forms of $h_{02}(x, y; \mathbf{v})$, $h_{01}(x, y; \mathbf{v})$, and their Fourier transforms. For a justification of the one-dimensional version of this approach, see Appendix A.

In summary, a simple algorithm for computing motion from smear is:

A-i Form the sum signal according to (8).
A-ii Hypothesize a motion and compute the inverse filtered images of $s(x, y; t_0, \Delta t_{02})$ and $g(x, y; t_0, \Delta t_{01})$. Compute the mean squared error between the inverse filtered images.
A-iii Repeat A-ii for every possible hypothesis of motion. Choose the hypothesis producing the minimum mean squared error as the correct motion.

2.4 Choosing the Right PSF

Motion smear PSF for a shutter that opens instantaneously at time t_i, stays open for duration $(t_i, t_{i+1}]$ and closes instantaneously at time t_{i+1} takes the form

$$l(t; t_i, \Delta t_{i,i+1}) = \begin{cases} 1 & \text{if } t_i < t \le t_{i+1}, \\ 0 & \text{elsewhere,} \end{cases} \quad (11)$$

which is the boxcar model commonly used in image restoration. Unfortunately, we cannot adopt model (11) because it leads to PSFs that have zeros in the frequency domain and, thus, do not have stable inverse systems. Instead, we use a shutter/filter whose transmission coefficient changes linearly with time during the acquisition of the first frame, i.e.,

$$l(t; t_0, \Delta t_{01}) = \begin{cases} \frac{1}{\Delta t_{12}} \frac{t - t_0}{\Delta t_{01}} & \text{if } t_0 < t \le t_1, \\ 0 & \text{elsewhere} \end{cases} \quad (12)$$

where a is a constant. One cannot choose the same $l(t, t_i; \Delta t_{i,i+1})$ for $(t_0, t_1]$ and $(t_1, t_2]$ because that will cause an identifiability problem for H_{02}/H_{01}. Hence, during the acquisition of the second frame, $t \in (t_1, t_2]$, the transmission of the shutter/filter remains constant, i.e.,

$$l(t; t_1, \Delta t_{12}) = \begin{cases} \frac{1}{\Delta t_{12}} & \text{if } t_1 < t \le t_2, \\ 0 & \text{elsewhere.} \end{cases} \quad (13)$$

This way, neither $h_{01}(x, y; \mathbf{v})$ nor $h_{02}(x, y; \mathbf{v})$ has zeros in the frequency domain. When we let $\Delta t_{01} = \Delta t_{12}$, the Fourier transforms of $h_{01}(x, y; \mathbf{v})$ and $h_{02}(x, y; \mathbf{v})$, which are needed for implementing (10), are given by

$$H_{01}(\Omega_x, \Omega_y; L, \phi) = \frac{e^{-jL\Omega}(2 + 2jL\Omega) - 2}{L^2\Omega^2}, \quad (14)$$

$$H_{02}(\Omega_x, \Omega_y; L, \phi) = \frac{2e^{-jL\Omega} - 2 + 2jL\Omega e^{-2jL\Omega}}{3L^2\Omega^2} \quad (15)$$

where $\Omega = \Omega_x + \Omega_y \tan(\phi)$ and L, ϕ are defined in Section 2.2.

2.5 Dealing with Noise

In the above derivation, we have assumed the observed images to be noise free. Since this is an impractical assumption, we now study the performance of the algorithm when noise is present in the images.

Using the noise-free images $g(x, y; t_i, \Delta_{i,i+1})$, the frequency domain implementation of (9) is

$$\mathcal{E}_o(L, \phi) := \left\| \frac{G_{01}}{H_{01}} - \frac{S_{02}}{H_{02}} \right\|^2 \quad (16)$$

where G_{01}, S_{02}, H_{01}, and H_{02} are the shorthand notation for $G(\Omega_x, \Omega_y; t_0, \Delta t_{01})$, $S(\Omega_x, \Omega_y; t_0, \Delta t_{02})$, $H_{01}(\Omega_x, \Omega_y; \mathbf{v})$, and $H_{02}(\Omega_x, \Omega_y; \mathbf{v})$, the Fourier transforms of $g(x, y; t_0, \Delta t_{01})$, $s(x, y; t_0, \Delta t_{02})$, $h_{01}(x, y; \mathbf{v})$ and $h_{02}(x, y; \mathbf{v})$, respectively. Now consider the noise corrupted versions of $g(x, y; t_i, \Delta t_{i,i+1})$,

$$g_n(x, y; t_0, \Delta t_{01}) = g(x, y; t_0, \Delta t_{01}) + n(x, y; t_0), \quad (17)$$

and

$$g_n(x, y; t_1, \Delta t_{12}) = g(x, y; t_1, \Delta t_{12}) + n(x, y; t_1) \quad (18)$$

where $n(x, y; t_0)$ and $n(x, y; t_1)$ are zero-mean, white noise. Suppose the true values of the motion parameters are: $L = L_0$ and $\phi = \phi_0$. If we were to use (9), substitution of (17) and (18) into (16) yields

$$E(L, \phi) = \left\| \frac{G_n(\Omega_x, \Omega_y; t_0, \Delta t_{01})}{H_{01}(\Omega_x, \Omega_y; L, \phi)} - \frac{S_n(\Omega_x, \Omega_y; t_0, \Delta t_{02})}{H_{02}(\Omega_x, \Omega_y; L, \phi)} \right\|^2 \quad (19)$$

where $G_n(\Omega_x, \Omega_y; t_0, \Delta t_{01})$ is the Fourier transform of $g_n(x, y; t_0, \Delta t_{01})$ and $S_n(\Omega_x, \Omega_y; t_0, \Delta t_{02})$ is the Fourier transform of $s_n(x, y; t_0, \Delta t_{02})$, the noisy version of $s(x, y; t_0, \Delta t_{02})$ as defined in (8), i.e.,

$$s_n(x, y; t_0, \Delta t_{02}) := g_n(x, y; t_0, \Delta t_{01}) + g_n(x, y; t_1, \Delta t_{12}).$$

As discussed in Appendix B, we define $K(L)$ as a parametric function of L (see (38) and (39)). A normalized error

$$\mathcal{E}'(L, \phi) = E[\mathcal{E}]/K(L) \quad (20)$$

is established and a new estimator is defined to be

$$\{\hat{L}, \hat{\phi}\} = \arg \min_{L,\phi} \mathcal{E}'(L, \phi). \quad (21)$$

In summary, in the presence of noise, the simple algorithm defined in Section 2.3 will have to be modified to account for the effect of the noise on (9). We show in Appendix B that the new estimator (21) is consistent, namely it converges to the true value when the sample size is large.

2.6 Improved Search Strategy

The estimator given by (21) involves the minimization of a complicated, nonlinear function. Numerical evaluation reveals that the error surface defined by (21) has many local minima and, hence finding the global minimum is difficult. If we can smooth the error surface, gradient searching algorithms can be effectively used. However, a posteriori smoothing does not help because it first requires an exhaustive search which is precisely what we are trying to avoid. An a priori smoothing mechanism has to be found.

Assume again that the observed images are noise free. We replace the inverse filters used in (16) by regularized inverse filters with a simplified regularizing polynomial [14]

$$H_{01r} = \frac{H_{01}^*}{H_{01}H_{01}^* + \gamma} \quad (22)$$

and

$$H_{02r} = \frac{H_{02}^*}{H_{02}H_{02}^* + \gamma}. \quad (23)$$

Analogous to (9), we define

$$\{\hat{L}, \hat{\phi}\} = \arg \min_{\{L,\phi\}} \mathcal{E}_r = \arg \min_{\{L,\phi\}} \|G_{01}H_{01r} - S_{02}H_{02r}\|^2 \quad (24)$$

When $\gamma \to 0$, $H_{01r} \to \frac{1}{H_{01}}$ and $H_{02r} \to \frac{1}{H_{02}}$ pointwise, respectively. Thus, if we choose γ to be a small value, we should have the shape of the error surface $\mathcal{E}_r$ resemble the shape of $\mathcal{E}$. Our observation is that, in general, the error surface of $\mathcal{E}_r$ in (24) is smoother than that of $\mathcal{E}$ in (9) (see Fig. 4). As verified by our experimental work, $\mathcal{E}_r$ can be minimized by a gradient search scheme. In our work, we adopted the quasi-Gaussian method supplied with analytical gradients. The formulae for the gradients are listed in Appendix C.

The regularized inverse filters in (22) and (23) are special cases of the Wiener filters and, thus, insensitive to noise [14]. Indeed, from experimental work, we found that the normalization procedure (20) that is required for (9) is not necessary for (24). However, noise does have an effect of changing the position of the minimum of $\mathcal{E}_r$ away from the correct value of L, ϕ. Thus, the estimate obtained from (24) may be approximate.

2.7 Complete Algorithm

The overall algorithm for estimating motion from smear is:

B-i Select a location (x_i, y_i) in the images and define an $N \times N$ neighborhood around that location. Heuristically, we choose N to be three times the maximum motion. Denote the image motion at location (x_i, y_i) by L_i, ϕ_i.

B-ii Within the $N \times N$ neighborhood, apply (24) in conjunction with any effective gradient search method to arrive at the approximate estimates of L_i and ϕ_i. Usually, $0.05 \le \gamma \le 0.25$.

B-iii Apply (21) around the approximate estimates to obtain the final estimates of L_i and ϕ_i.

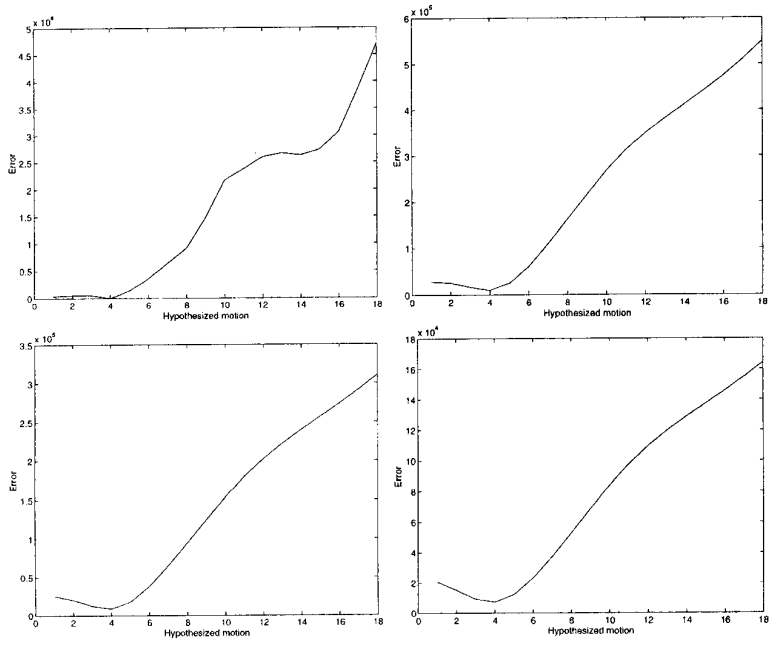


Fig. 4. (a) (Top left) Variation of the error, $\mathcal{E}$, given by (9) where the inverse filter is not regularized. (b) (Top right) Variation of the error, $\mathcal{E}_r$ given by (24) where the inverse filter is regularized with $\gamma = 0.05$. (c) (Bottom left) Variation of $\mathcal{E}_r$ when $\gamma = 0.1$. (d) (Bottom right) Variation of $\mathcal{E}_r$ when $\gamma = 0.2$. For ease of visualization, we demonstrate the one-dimensional case although with two-dimensional motion the error surface exhibits similar behavior. Note that the variation in error is smoother with regularization and the global minimum is preserved. The above behavior is obtained from a particular motion smeared image pair used in our experimental tests. The behavior is typical of that observed in the many experiments described later.

B-iv If the estimates are obviously erroneous, an exhaustive search based on (21) is conducted at this image location (x_i, y_i).

B-v Move to the next image location (x_{i+1}, y_{i+1}) and repeat the above using the estimates of L_i and ϕ_i obtained at (x_i, y_i) to initialize the gradient search for the new estimates of L_{i+1} and ϕ_{i+1}.

2.8 The Aperture Problem

The aperture problem is well known in computer vision—one cannot determine the component of the optical flow along an isobrightness contour [15]. One would surely speculate that "motion-from-smear" will suffer from the same problem. For example, if the scene has a constant gray level, motion smear would not be perceivable. In the following, we investigate the manifestation of the aperture problem in "motion-from-smear" using some results from system identification theory, from whose point of view the aperture problem is equivalent to the concept of system identifiability.

Assume that the direction of motion is known and consider one dimensional images. According to Proposition 1, in order to uniquely determine motion from $g(x; t_0, \Delta t_{01})$ and $g(x; t_1, \Delta t_{12})$, it is sufficient to have $\{g(x; t_0, \Delta t_{01}), g(x; t_1, \Delta t_{12})\}$, when they are considered as an input-output pair through a linear system H_{02}/H_{01}, to be informative. [17]. The condition of their being informative will depend on the structure of the linear system H_{02}/H_{01}. Due to the difficulty of directly analyzing (14) and (15) which involves transcendental functions in a nonalgebraic way, we consider an approximation in the one-dimensional case. The linear system H_{01} can be approximated by the discrete system

$$\sum_{i=0}^{\lfloor v_x \rfloor} z^{-i} i$$

(v_x is in pixel units) which is the Z-transform of the sampled impulse response function h_{01} (in the one-dimensional case, h_{01} is a simple ramp function) [16]. Similarly the linear system H_{02} can be approximated by

$$\sum_{i=0}^{\lfloor v_x \rfloor} z^{-i} i + \sum_{i=\lfloor v_x \rfloor + 1}^{\lfloor 2v_x \rfloor} z^{-i}$$

the Z-transform of the sampled h_{02}. Thus, the transfer function H_{02}/H_{01} can be approximated by

$$\frac{H_{02}}{H_{01}} = \frac{\sum_{i=0}^{\lfloor v_x \rfloor} z^{-i} i + \sum_{i=\lfloor v_x \rfloor + 1}^{\lfloor 2v_x \rfloor} z^{-i}}{\sum_{i=0}^{\lfloor v_x \rfloor} z^{-i} i}. \tag{25}$$

It follows from [17] that for a linear system with structure like (25) $\{g(x; t_0, \Delta t_{01}), g(x; t_1, \Delta t_{12})\}$ is informative when the input data $g(x; t_0, \Delta t_{01})$ are persistently exciting of order $\lfloor 3v_x \rfloor$. Because H_{01} is nonzero pointwise and $f(x; t_0)$ and $g(x; t_0, \Delta t_{01})$ are related as in (5), it is equivalent to say that if the original image $f(x; t_0)$ has $\lfloor 3v_x \rfloor$ distinctive frequency components, the normal flow can be uniquely identified. This conclusion is consistent with the traditional notion of the aperture problem: if there is no gray level edge, there is only one frequency component that has nonzero energy and thus normal flow is not identifiable; if there is a step edge, there are infinitely many frequency components and consequently the normal flow is identifiable.

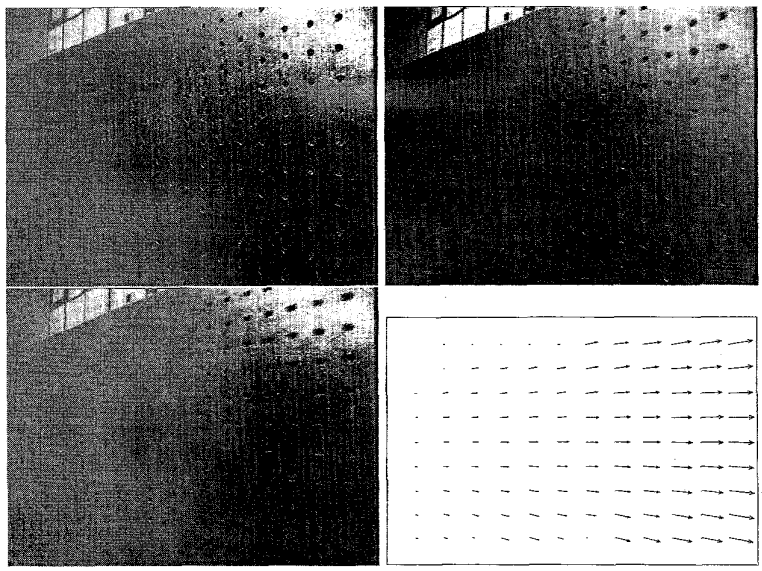


Fig. 5. (a) (Top left) The unblurred image acquired at the initial camera position. (b) (Top right) A simulated motion blurred image frame is created by incrementally translating the camera, acquiring an image at each position, and adding the first 10 images. This corresponds to the first blurred image frame acquired with imaging "duration" $(t_0, t_1]$. (c) (Bottom left) The next 10 images are added to create the second blurred image frame corresponding to imaging "duration" $(t_0, t_2]$. (d) (Bottom right) The estimated motion by "motion-from-smear."

3 Experimental Results

As mentioned before, the energy conversion efficiency of any available sensor technology places constraints between shutter speed and illumination level for a required image SNR. Hence, many practical image motion estimation applications will encounter motion smear. We have designed experiments that attempt to duplicate such real situations. Our experimental investigation of the "motion-from-smear" approach involves three different stages of validation:

1) The algorithm is applied to a two-frame sequence of motion blurred images that are "simulated" as follows. A camera is mounted on an optical rail and is made to view a static scene. The camera is translated by small increments and at each position an image is acquired. The first T images are added to form the first frame $g(x, y; t_0, \Delta t_{01})$ and the next T images are added to form the second frame $g(x, y; t_1, \Delta t_{12})$. This procedure approximates continuous temporal integration of the sensor by a sum of temporally sampled data. The data used and the results obtained are described in Section 3.1.
2) The algorithm is applied to "real data" acquired from a sensor equipped with a filter, the transmission coefficient of which is temporally varying. This sensor conforms to (12) and (13). The sensor is made to view a scene that is rotated at a known constant velocity. The temporally varying transmission coefficient is implemented by rotating a gradient index filter in front of a camera. Implementation of this scheme, and the results obtained are described in Section 3.2.
3) The performance of the "motion-from-smear" approach is experimentally compared with another well-known motion estimation scheme, namely the differential optical flow scheme [15]. The method of comparison and results are presented in Section 3.3.

3.1 Results on Simulated Motion Smeared Image Frames

From (1),

$$g(x, y; t_i, \Delta t_{i,i+1}) = \int_{t_i}^{t_{i+1}} f(x, y; t) l(t; t_i, \Delta t_{i,i+1}) dt \simeq \sum_{k=0}^{T-1} f(x, y; t^{(k)}) l(t^{(k)}; t_i, \Delta t_{i,i+1})$$

where the interval $(t_i, t_{i+1}]$ is divided into T time steps. This is to say that we can simulate $g(x, y; t_i, \Delta t_{i,i+1})$ by summing up unblurred images obtained at incrementally different camera translations corresponding to camera positions at instants $t^{(k)}$. In this experiment, a flat optical table top is viewed at an angle by the camera. Twenty images were taken by translating the camera incrementally along an optical rail. The first ten images are used to create the first frame, $g(x, y; t_0, \Delta t_{01})$. The next ten frames are used to create the second frame, $g(x, y; t_1, \Delta t_{12})$. Fig. 5a is one of the twenty unblurred images of the scene. Figs. 5b and 5c are two frames of motion smeared images. Since the depth field caused by the slanting table top is linear, we expect the image motion field to be continuous, monotonically decreasing with depth except for the upper left corner of the image, where a piece of the background wall appears. The range between the wall and the camera is so large that the apparent motion of the distant wall is virtually zero. Visually, the estimated motion field presented in Fig. 5d agrees with the above analysis. Furthermore, we provide some quantitative evaluation of the estimates. We found the error between the manually measured motion (groundtruth) and the estimated motion vectors has standard deviation of (0.9543, 0.8816),

whereas the magnitudes of the estimated motion vectors range from 0 to 10.87. All measurements are in pixel units.

3.2 Collection of Real Data and Results

We designed a prototype system to implement the sensor specified by (12) and (13). The sensor can be implemented in a number of ways, such as adjusting the gain of the photon amplifier behind the photodiodes of CCD sensors, applying a spatial light modulator in front of the photodiodes, using custom designed CCD arrays, or rotating a gradient index filter in front of the lens of the camera [18]. Suppose $\Delta t = \Delta t_{i,\, i+1}$, $i = 0, 1, 2, \cdots$. To satisfy (12) and (13), the gradient index filter must have a transmission coefficient that varies as shown in Fig. 6a. Let $I(\theta)$ denote the grey level at angle θ. Suppose $0 \le I(\theta) \le I_{max}$ where I_{max} denotes the maximum gray level. The pattern in Fig. 6a satisfies

$$I(\theta) = \begin{cases} \frac{I_{max}}{\pi}\theta & 0 < \theta \le \pi, \\ I_{max} & \pi < \theta \le 2\pi. \end{cases} \tag{26}$$

The gradient index filter is placed closely against the lens of the camera (see Fig. 6b) and is rotated at a constant angular velocity such that

$$\theta(t) = \frac{\pi}{\Delta t}(t - t_0) \bmod 2\pi \tag{27}$$

where t_0 is the time when the data acquisition starts. Note that $\theta(t)\big|_{t=t_1} = \pi$, $\theta(t)\big|_{t=t_2} = 0$ or 2π, and $I(t) \propto I(\theta(t))$. Then

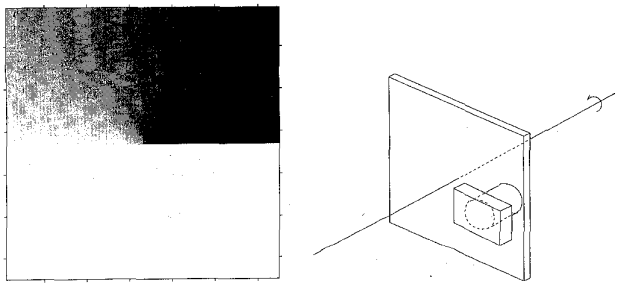
$$I(t) = c\frac{I_{max}}{\Delta t}(t - t_{2i}),\ t_{2i} < t \le t_{2i+1} \tag{28}$$

and

$$I(t) = cI_{max},\ t_{2i+1} < t \le t_{2i+2} \tag{29}$$

where c is a constant and $i = 0, 1, \cdots$. This is the sensor gain required by (12) and (13).

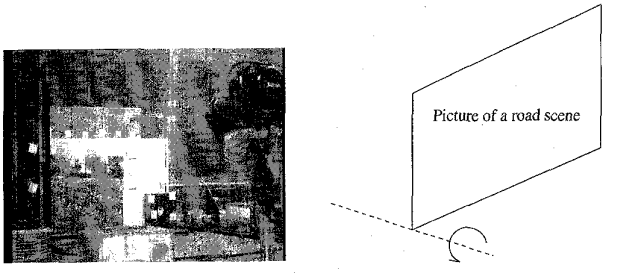
In our low-cost implementation, we printed the gray level pattern (Fig. 6a) onto a transparent film. The transparency is sandwiched between two pieces of Plexiglas to simulate a gradient index filter. The filter is then mounted on the shaft of a DC motor, placed closely against the lens of the camera (Fig. 6b), and rotated at a constant angular velocity according to (27).



(a) (b)

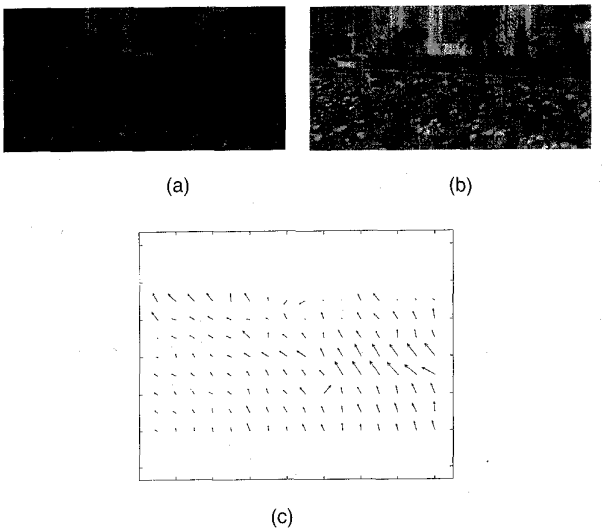
Fig. 6. (a) The gray level pattern of the gradient index filter. A transparency is made from this pattern and rotated in front of the camera. This implements the time-varying transmission coefficient as per (12) and (13). (b) The positioning of the gradient index filter and the camera.

Fig. 7a depicts the experimental setup on the optical table in our laboratory. The object being imaged is a picture of a road scene that is rotated as schematically shown in Fig. 7b. The rotation is driven by another DC motor and the angular velocity of the rotation is kept constant by keeping the voltage applied to the motor constant. The images are recorded on films and later scanned into digital format using a film scanner.



(a) (b)

Fig. 7. (a) The laboratory setup. The camera and the gradient index filter can be seen in the upper right portion of the image. The road scene is rotated by a DC motor (lower right portion of the image) driven by a constant voltage. (b) The object (picture of a road scene) is rotated counterclockwise about the lower left corner.



(a) (b)

(c)

Fig. 8. (a), (b) Real motion smeared image sequence. The scene motion is as indicated in Fig. 6b. (c) the estimated motion from "motion-from-smear."

3.3 Smear vs. Displacement as a Visual Cue

In order to fully justify our motivation to introduce a new motion estimation technique, it is necessary to compare "motion-from-smear" with existing algorithms. We first discuss, in general terms, a fundamental difference between "motion-from-smear" and the existing classes of motion estimation algorithms. Next, we choose a representative method and compare it experimentally with our "motion-from-smear" algorithm.

Two frames of such acquired images are shown in Figs. 8a and 8b. The images are very noisy due to the non-

linearity introduced by the film and the scanner. In calculating the motion vectors, we set the maximum number of iterations to be 60. Due to the severe noise, at some locations of the image, gradient search gives obviously wrong results at the end of the iterations. In this situation, an exhaustive search is performed, where L and ϕ are quantized using an appropriate step size and we search over all possible pairs of the quantized values. Motion field computed by "motion-from-smear" is shown in Fig. 8c, which is as expected for a rotational motion as shown in Fig. 7b. The estimates are quantitatively evaluated as follows. Since the optical axis of the camera is parallel to the axis of scene rotation, the 2D rotational motion results in motion fields that can be modeled by an affine transform. We evaluated the error between the affine model and the estimated motion vectors and found that it has standard deviation of (3.6, 3.6), whereas the magnitudes of the motion vectors range from 1.1 to 22.5. All measurements are in pixel units.

3.3.1 Temporal Aliasing

All of the existing methods, be they gradient based or feature correspondence based, rely on using displacement as a visual cue for image motion [12]. The consequence of this is that they all suffer from temporal aliasing, as pointed out by [19]. For an image sequence acquired assuming an infinitely fast shutter, the minimum temporal sampling (Nyquist) rate is limited by the maximum spatial image frequency as well as the maximum image motion. Temporal aliasing occurs when the Nyquist rate is not achieved. Although temporal aliasing may occur in any image that contains some gray level variation (i.e., nonzero spatial frequency), it is best illustrated with an image containing a repetitive texture. Consider an image satisfying $f(x, y) = f(x + k_x P_x, y + k_y P_y)$, where $k_x, k_y \in \mathbf{Z}$, and $P_x > 0$ and $P_y > 0$ are the periods of the texture pattern in the x and y directions, respectively. Consider a sequence $f(x, y; t)$ that results from the motion of this texture. According to (2), two consecutive frames at time instants t_0 and t_1 are related by

$$f(x, y; t_1) = f(x - v_x \Delta t_{01}, y - v_y \Delta t_{01}; t_0).$$

Hence, if the use of displacement as a motion cue yields the following match between frames in a motion sequence: $f(x, y; t_1) = f(x - D_x, y - D_y; t_0)$, then we must have

$$D_x = v_x \Delta t_{01} + l_x P_x, \quad D_y = v_y \Delta t_{01} + l_y P_y,$$

where $l_x, l_y \in \mathbf{Z}$. Note that we do not know the correct value for l_x, l_y. Hence, the solution for (v_x, v_y) is not unique. If the temporal sampling rate is higher than the Nyquist rate, $l_x = l_y = 0$ is guaranteed, and (v_x, v_y) is uniquely determined.

Assuming that the image contains energy corresponding to the maximum possible spatial image frequency, the Nyquist temporal sampling rate is met only if the image plane displacement is no more than one pixel between successive frames! Since this is usually not satisfied in practical motion estimation applications, temporal aliasing may indeed occur. As a consequence, there are infinitely many legitimate estimates for motion. For example, we generated an image sequence simulating motion using the image with repetitive texture pattern shown in Fig. 9a and motion parameters $v_x = 7$, $v_y = 0$. We applied the classical method that uses displacement as a visual cue—correlation matching, and depict a slice of the cross correlation function in Fig. 9b. Two maxima exist within our search space, each indicating a legitimate estimate for motion.

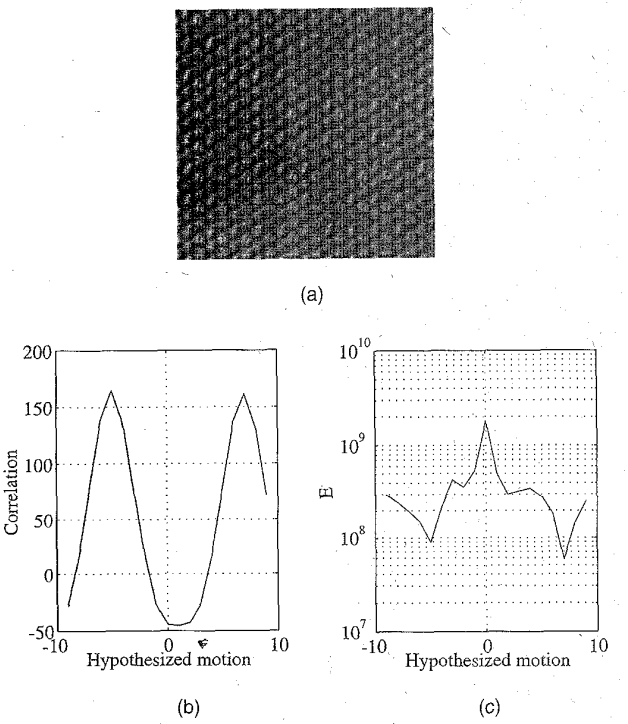


Fig. 9. (a) An image with repetitive texture. This is used to simulate a motion smeared sequence which is then used to compare correlation matching and "motion-from-smear." (b) The cross correlation function reaches maxima at two locations within our search space. (c) The error function in "motion-from-smear" has a unique global minimum within the same search space.

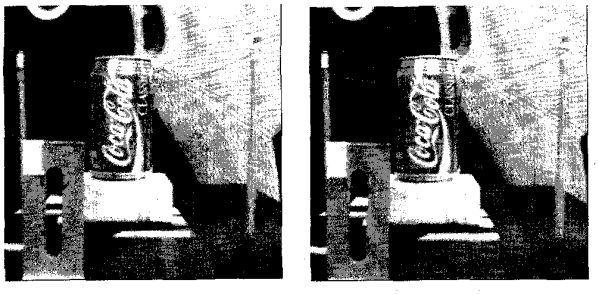
On the other hand, "motion-from-smear" does not suffer from this ambiguity caused by temporal aliasing. The degree of motion smear for each $(D_x, D_y) = (v_x \Delta t_{01} + l_x P_x, v_y \Delta t_{01} + l_y P_y)$ is different for different l_x and l_y, based on which the correct motion estimate can be obtained. Fig. 9c depicts the surface of the error function being minimized in "motion-from-smear" with the same data as those used in Fig. 9b. It has a unique global minimum within the same search space and yields a unique estimate for motion. For twenty simulation experiments, each with different 20dB white Gaussian noise added to the images, the correlation technique gives either $\hat{v}_x = -5$, $\hat{v}_y = 0$ or $\hat{v}_x = 7$, $\hat{v}_y = 0$ whereas "motion-from-smear" always gives only $\hat{v}_x = 7$, $\hat{v}_y = 0$. We note that the perfectly periodic texture used in the above example is the worst, and perhaps rarely encountered case for conventional methods. Also, the correlation technique used above may not perform as well as other optic flow techniques. However, the example does illustrate that motion from smear is less sensitive (if not insensitive) to temporal aliasing when compared to methods based on displacement cues.

3.3.2 Comparison with Differential Optical Flow

Comparison Using Simulated Motion Smeared Images: In order to compare our approach with other approaches, we must ensure that the same amount of data/information is available to each. Consider the case where the imaging sensor is limited by a maximum shutter speed, as is the case in most practical applications. Let the sensor take at least a duration of Δt to acquire an image. Thus a sequence of n frames requires a period of at least $n\Delta t$. In order to compute as temporally local a motion estimate as possible, i.e., using the least restrictive assumption of temporal constancy of motion, we have to use two, and not greater or fewer than two, successive frames in an image sequence—since a single frame provides ambiguous motion estimates, and the use of more frames requires assumption of motion constancy over a larger duration of time. Hence, we restrict our experimental comparison to consider only other such temporally local motion estimators that are based on analyzing two frames. A representative two-frame algorithm is the differential optical method [15], which we compare with "motion-from-smear."

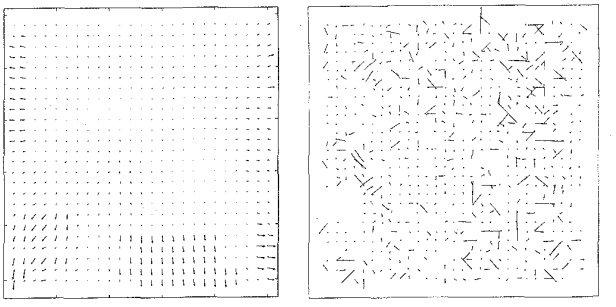
Recall that the amount of motion blur is due to practical limitation of the sensor, i.e., the sensitivity limits maximum shutter speed at any illumination level and for an acceptable SNR. For example, motion smear is more pronounced at lower light levels. Our comparison is designed with this consideration in mind. We note that the differential optical flow method performs best when the lens transmittance does not vary with time while "motion-from-smear" requires lens transmittance to be controlled as per (12) and (13), and generate simulation data accordingly. However, the amount of motion blur in the images used to test the differential optical flow method was the same as that in the images used to test "motion-from-smear."

The simulated images for "Motion-from-smear" and the estimated motion field by "motion-from-smear" are shown in Figs. 10a, 10b, and 11a. The motion field estimated by the differential optical flow method is shown in Fig. 11b. We used the implementation of the differential optical flow method by Barron et al. [19]. The implementation parameters for "motion-from-smear" are $\gamma = 0.25$, window size = 50×50. The implementation parameters for the differential optical flow method are the same as those in [19].



(a) (b)

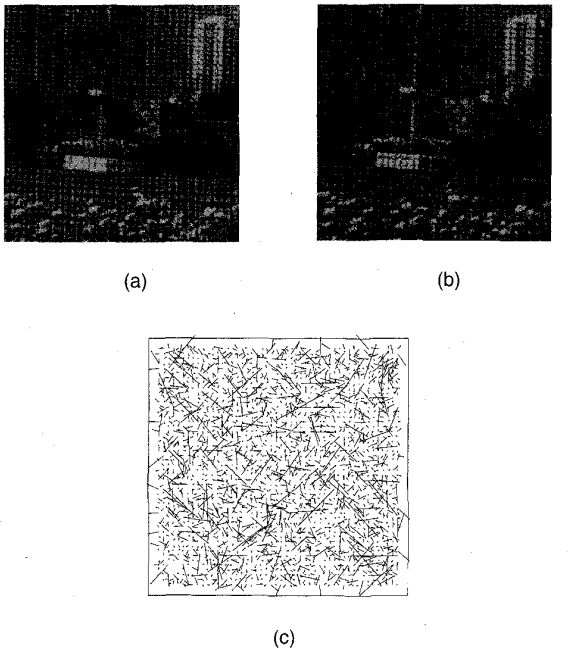
Fig. 10. Two frames of a motion smeared image sequence created from NASA Coke can data.



(a) (b)

Fig. 11. (a) Motion computed by "Motion-from-smear" using images in Fig. 9. (b) Motion computed by differential optical flow using the same data.

Comparison Using Real Blurred Images: We have also performed the comparison on real data acquired using our experimental setup. Two frames of such acquired images of the same scene and the same motion as in Figs. 8a and 8b are shown in Figs. 12a and 12b. Note that although these images contain only part of the scene because the particular implementation of differential optical flow that we use cannot handle images as big as those in Figs. 8a and 8b, this part alone is enough to demonstrate different behaviors of the algorithms under comparison. Fig. 12c shows the results from the differential optical flow algorithm. Note the difference between the results shown in Fig. 8c and Fig. 12c.



(a) (b)

(c)

Fig. 12. (a), (b) Two frames of real motion smeared images. The amount of motion smear is the same as in Figs. 7a and 7b. The experimental setup is the same as that illustrated in Fig. 6b. (c) Motion computed using differential optical flow, for comparison with the results from "Motion-from-smear" shown in Fig. 7c.

4 Conclusion

Motion smear as a visual cue for image motion estimation has been largely ignored. Our work has demonstrated that it is indeed possible to extract motion information from motion smear. Although a rigorous, general proof is not presented, we have shown that temporal aliasing does not affect "motion-from-smear" to the same degree as it does algorithms that use displacement as a cue. Experimental results on images with both simulated smear and real smear, using our "motion-from-smear" algorithm as well as a conventional motion estimation technique, show that "motion-from-smear" provides an additional tool for motion estimation and effectively complements the existing techniques when apparent motion smear is present.

The introduction of a new visual cue for motion estimation open many revenues of research. In our future work, we will explore the fusion of "motion-from-smear" with other motion estimation algorithms. "Motion-from-smear" algorithms that use single frame, multiple frames and innovative data acquisition schemes are being developed. Motion discontinuity, in the context of "motion-from-smear," is also an important issue to consider. A noted constraint for wide applications of "motion-from-smear," in particular using the algorithm developed in this paper, is that the special sensor is not widely available. Although we have demonstrated that a low-cost implementation is possible, in our future effort total elimination of this constraint will be considered.

Appendix A

In the following proposition, we will borrow some concepts from system identification theory. For details, please see [17].

Proposition 1. *Suppose* 1) $s(x)$ *and* $t(x)$ *are two degraded versions of an image* $o(x)$, *i.e.*,

$$s(x) = o(x) * h_s(x) \quad \text{and} \quad t(x) = o(x) * h_t(x),$$

2) $o(x)$ *is persistently exciting,* 3) $h_s(x)$, *and* $h_t(x)$ *and their inverse systems are stable,* 4) *the linear system defined by* $\frac{H_t(\Omega)}{H_s(\Omega)}$ *is strictly globally identifiable with respect to the scaling factor of* Ω. *Let* $\tilde{\cdot}$ *denote the inverse of the corresponding system. Then*

$$\mathcal{E}(i) = \left\| \frac{1}{i} s(x) * \tilde{h}_s(x/i) - \frac{1}{i} t(x) * \tilde{h}_t(x/i) \right\|^2$$

is minimized if and only if $i = 1$.

Proof. Sufficiency: The proof is trivial since when $i = 1$,

$$\begin{aligned}\mathcal{E}(1) &= \left\| s(x) * \tilde{h}_s(x) - t(x) * \tilde{h}_t(x) \right\|^2 \\ &= \left\| o(x) - o(x) \right\|^2 \\ &= 0\end{aligned}$$

Necessity: From Parseval's theorem,

$$\mathcal{E}(i) = \left\| S(\Omega)\tilde{H}_s(i\Omega) - T(\Omega)\tilde{H}_t(i\Omega) \right\|^2$$

Suppose $\mathcal{E}(i)$ is also minimized at some value other than $i = 1$, i.e.,

$$\mathcal{E}(i) = \mathcal{E}(1) = 0, \quad i \neq 1. \tag{30}$$

Since $H_s(\Omega)$ and $H_t(\Omega)$ have stable inverse, we have $H_s(\Omega) \neq 0$ and $H_t(\Omega) \neq 0$, $\forall\Omega$. By the definition of persistent excitation, $O(\Omega) \neq 0$ $\forall\Omega$, and thus $S(\Omega) \neq 0$ and $T(\Omega) \neq 0$, $\forall\Omega$. Therefore, for (30) to be true, $S(\Omega)\tilde{H}_s(i\Omega) = T(\Omega)\tilde{H}_t(i\Omega)$ is required. That is,

$$\frac{H_t(i\Omega)}{H_s(i\Omega)} = \frac{T(\Omega)}{S(\Omega)} = \frac{H_t(\Omega)}{H_s(\Omega)}$$

It follows from the global identifiability of $\frac{H_t(\Omega)}{H_s(\Omega)}$ that $i = 1$—a contradiction. □

When the direction of the motion is known, the two-dimensional "motion-from-smear" problem reduces to a one-dimensional one. In this case, this proposition justifies the algorithm proposed in Section 2.3.

Appendix B

In this appendix, we discuss in detail the construction of the estimator defined in (21). Substituting the Fourier transforms of (17), (18), and (8) into (19) yields

$$\begin{aligned}\mathcal{E}(L,\phi) = &\left\| \frac{G(\Omega_x,\Omega_y;t_0,\Delta t_{01}) + N(\Omega_x,\Omega_y;t_0)}{H_{01}(\Omega_x,\Omega_y;L,\phi)} \right. \\ &\left. - \frac{S(\Omega_x,\Omega_y;t_0,\Delta t_{02}) + N(\Omega_x,\Omega_y;t_0) + N(\Omega_x,\Omega_y;t_1)}{H_{02}(\Omega_x,\Omega_y;L,\phi)} \right\|^2 \\ = &\left\| \frac{G(\Omega_x,\Omega_y;t_0,\Delta t_{01})}{H_{01}(\Omega_x,\Omega_y;L,\phi)} - \frac{S(\Omega_x,\Omega_y;t_0,\Delta t_{02})}{H_{02}(\Omega_x,\Omega_y;L,\phi)} \right\|^2 \\ &+ \left\| \frac{N(\Omega_x,\Omega_y;t_0)}{H_{01}(\Omega_x,\Omega_y;L,\phi)} - \frac{N(\Omega_x,\Omega_y;t_0) + N(\Omega_x,\Omega_y;t_1)}{H_{02}(\Omega_x,\Omega_y;L,\phi)} \right\|^2 .\end{aligned} \tag{31}$$

Noise terms $N(\Omega_x, \Omega_y; t_0)$ and $N(\Omega_x, \Omega_y; t_1)$ are independent of the images and zero-mean; therefore, the cross terms are not present in (31). Substituting (16) into (31) and expanding the second integral lead to

$$\begin{aligned}E(L,\phi) = E_o + &\frac{1}{N_x N_y}\int_{\Omega_x}\int_{\Omega_y} \left\{ \left(N_0\tilde{H}_{01}\right)\left(N_0\tilde{H}_{01}\right)^* - \left(N_0\tilde{H}_{01}\right)\left(N_0\tilde{H}_{02}\right)^* - \right. \\ &- \left(N_0\tilde{H}_{01}\right)\left(N_1\tilde{H}_{02}\right)^* - \left(N_0\tilde{H}_{02}\right)\left(N_0\tilde{H}_{01}\right)^* - \left(N_1\tilde{H}_{02}\right)\left(N_0\tilde{H}_{01}\right)^* \\ &+ \left(N_0\tilde{H}_{02}\right)\left(N_0\tilde{H}_{02}\right)^* + \left(N_0\tilde{H}_{02}\right)\left(N_1\tilde{H}_{02}\right)^* + \left(N_1\tilde{H}_{02}\right)\left(N_0\tilde{H}_{02}\right)^* \\ &\left. + \left(N_1\tilde{H}_{02}\right)\left(N_1\tilde{H}_{02}\right)^* \right\} d\Omega_x d\Omega_y \\ := &\mathcal{E}_0 + \mathcal{E}_n\end{aligned} \tag{32}$$

where we have used the shorthand notation N_0, N_1, $\tilde{H}_{01}$, and $\tilde{H}_{02}$ for $N(\Omega_x, \Omega_y; t_0)$, $N(\Omega_x, \Omega_y; t_1)$, $\frac{1}{H_{01}(\Omega_x,\Omega_y;L,\phi)}$, and $\frac{1}{H_{02}(\Omega_x,\Omega_y;L,\phi)}$, respectively. We have also used symbol $\mathcal{E}_n$ to represent the integral on the right hand side of (32) and N_x, N_y

to represent the x and y dimensions of the image, respectively. $\mathcal{E}_o$ will be minimized if and only if $L = L_0$ and $\phi = \phi_0$ (Appendix A). However, $\mathcal{E}$ may not behave so because of the effect of $\mathcal{E}_n$. Consider the first term of $\mathcal{E}_n$. By virtue of Parseval's theorem,

$$\int_{\Omega_x}\int_{\Omega_y}\left(N_0\tilde{H}_{01}\right)\left(N_0\tilde{H}_{01}\right)^* d\Omega_x d\Omega_y = \int_x\int_y \mathcal{F}^{-1}\left[\left(N_0\tilde{H}_{01}\right)\right]\mathcal{F}^{-1}\left[\left(N_0\tilde{H}_{01}\right)^*\right]dxdy. \quad (33)$$

But with

$$\mathcal{F}^{-1}\left[\left(N_0\tilde{H}_{01}\right)\right] = n(x,y;t_0) * \tilde{h}_{01}(x,y;\mathbf{v}) = \int_{x_1}\int_{y_1}\tilde{h}_{01}(x_1,y_1;\mathbf{v})n(x-x_1,y-y_1;t_0)\,dx_1dy_1,$$

(33) changes to

$$\int_{\Omega_x}\int_{\Omega_y}\left(N_0\tilde{H}_{01}\right)\left(N_0\tilde{H}_{01}\right)^* d\Omega_x d\Omega_y = \int_{x_1,x_2,y_1,y_2}\left\{\tilde{h}_{01}(x_1,y_1;\mathbf{v})\tilde{h}_{01}(x_2,y_2;\mathbf{v})\int_x\int_y\left\{n(x-x_1,y-y_1;t_0)n(x-x_2,y-y_2;t_0)\right\}dxdy\right\}dx_1dx_2dy_1dy_2. \quad (34)$$

Taking the expected value of both sides of (34) and dividing by N_xN_y, we have

$$E\left[\frac{1}{N_xN_y}\int_{\Omega_x}\int_{\Omega_y}\left(N_0\tilde{H}_{01}\right)\left(N_0\tilde{H}_{01}\right)^* d\Omega_x d\Omega_y\right] \quad (35)$$

$$= \int_{x_1,x_2,y_1,y_2}\left\{\tilde{h}_{01}(x_1,y_1;\mathbf{v})\tilde{h}_{01}(x_2,y_2;\mathbf{v})\frac{1}{N_xN_y}\int_x\int_y E\left[n(x-x_1,y-y_1;t_0)n(x-x_2,y-y_2;t_0)\right]dxdy\right\}dx_1dx_2dy_1dy_2$$

$$= \sigma_n^2\int_{x_1,y_1}\tilde{h}_{01}(x_1,y_1;\mathbf{v})\tilde{h}_{01}(x_1,y_1;\mathbf{v})\,dx_1dy_1. \quad (36)$$

We have used the property

$$E\left[n(x-x_1,y-y_1;t_0)n(x-x_2,y-y_2;t_0)\right] = \sigma_n^2\delta(x_1-x_2,y_1-y_2)$$

due to the whiteness of the noise. By the same token, other terms of $\mathcal{E}_n$ bear relationships similar to (36), using which, after some simplification, we have

$$E[\mathcal{E}_n] = \sigma_n^2\int_{x_1}\int_{y_1}\left\{\tilde{h}_{01}(x_1,y_1;\mathbf{v})\tilde{h}_{01}(x_1,y_1;\mathbf{v}) - 2\tilde{h}_{01}(x_1,y_1;\mathbf{v})\tilde{h}_{02}(x_1,y_1;\mathbf{v}) + 2\tilde{h}_{02}(x_1,y_1;\mathbf{v})\tilde{h}_{02}(x_1,y_1;\mathbf{v})\right\}dx_1dy_1. \quad (37)$$

Using Parseval's theorem, we can evaluate the integral in (37) in the frequency domain by defining

$$K(L,\phi) = \int_{\Omega_x}\int_{\Omega_y}\left\{\frac{1}{H_{01}}\frac{1}{H_{01}^*} - \frac{2}{H_{01}}\frac{1}{H_{02}^*} + \frac{2}{H_{02}}\frac{1}{H_{02}^*}\right\}d\Omega_x d\Omega_y. \quad (38)$$

Substituting (14) and (15) into (38) yields

$$K(L,\phi) = \int_0^\infty\int_0^\infty\left\{\frac{L^4\Omega^4}{2L^2\Omega^2 + 4L\Omega\sin(L\Omega) + 4(1-\cos(L\Omega))} + \frac{L^4\Omega^4}{\cos(L\Omega)L^2\Omega^2 - L\Omega\sin(2L\Omega) + 2 + 2\cos(L\Omega)} + \frac{L^4\Omega^4}{L^2\Omega^2 + 2L\Omega(\sin(L\Omega) - \sin(2L\Omega)) + 2 - 2\cos(L\Omega)}\right\}d\Omega_x d\Omega_y \quad (39)$$

where $\Omega = \Omega_x + \Omega_y\tan\phi$. It is easy to verify that

$$J(L,\phi)\Big|_{\Omega_x,\Omega_y} = J(L,0)\Big|_{\Omega_x',\Omega_y'}, \quad \forall\phi \quad (40)$$

where $J(L, \phi)$ denotes the integrand in (39),

$$\Omega_x' = \Omega_x\cos\phi - \Omega_y\sin\phi \text{ and } \Omega_y' = \Omega_x\sin\phi - \Omega_x\cos\phi.$$

Due to (40), we always have $K(L, \phi) = K(L, 0)$. Thus, $K(L, \phi)$ is not dependent on ϕ and we instead use $K(L)$ for brevity.

Since the sine and cosine functions are bounded in [–1, 1], long division of (39) yields

$$E[\mathcal{E}_n] = \Theta(L^2), \quad (41)$$

namely $E[\mathcal{E}_n]$ is tightly bounded by L^2. The effect of $\mathcal{E}_n$ follows from (32) and (41): When L is large, $\mathcal{E}_n$, instead of $\mathcal{E}_o$, becomes the dominant term in $\mathcal{E}$, and therefore the minimization of $\mathcal{E}$ as in (9) may give a totally wrong estimate.

If we normalize $E[\mathcal{E}]$ by $K(L)$ and define

$$\mathcal{E}'(L,\phi) = \frac{E[\mathcal{E}]}{K(L)}, \quad (42)$$

we can avoid the effect of $E[\mathcal{E}_n]$. It follows from (32) and (42) that

$$\mathcal{E}'(L,\phi) = \frac{\mathcal{E}_o(L,\phi)}{K(L)} + \sigma_n^2. \quad (43)$$

In view of 1)$K(L) > 0$, 2) σ_n^2 is constant, and 3) $\mathcal{E}_o(L, \phi)$ reaches zero if and only if $L = L_0$, $\phi = \phi_0$, we conclude that $\mathcal{E}'(L, \phi)$ will reach zero if and only if $L = L_0$, $\phi = \phi_0$. Thus, we define

$$\{\hat{L},\hat{\phi}\} = \arg\min_{L,\phi}\mathcal{E}'(L,\phi),$$

which is the same as (21).

In practice, (21) becomes an approximation, because $\mathcal{E}'(L, \phi)$ will be replaced by its sample estimator. However, when the sample size is large enough the approximation is a very close one due to the following fact.

The estimator defined in (21) is convergent, i.e., it converges to a constant value when $N_x \to \infty$, $N_y \to \infty$, because of the convergence of

$$\frac{1}{N_xN_y}\int_{x,y}n(x-x_1,y-y_1;t_i)n(x-x_2,y-y_2;t_i)dxdy \quad i = 0, 1, \cdots.$$

Furthermore, (21) is a consistent estimator, i.e., it not only converges but also converges to the true value. Define

$$\{\bar{L},\bar{\phi}\} = \lim_{N_x\to\infty,N_y\to\infty}\{\hat{L},\hat{\phi}\}.$$

Since

$$\lim_{N_x\to\infty,N_y\to\infty}\frac{1}{N_xN_y}\int_{x,y}n(x-x_1,y-y_1;t_i)n(x-x_2,y-y_2;t_i)dxdy = \sigma_e^2\delta(x_1-x_2,y_1-y_2),$$

we have

$$\{\bar{L}, \bar{\phi}\} = \arg \min_{\{L,\phi\}} \frac{\mathcal{E}_o(L,\phi)}{K(L)} + \sigma_e^2 \tag{44}$$

Since the function to be minimized in (44) is the same as $\mathcal{E}'(L, \phi)$ in (43), we conclude that $\{\bar{L}, \bar{\phi}\} = \{L_0, \phi_0\}$ due to the discussion following (43). In other words, (21) is consistent.

APPENDIX C

We need to find the gradients of $\mathcal{E}_r$ with respect to L and ϕ. From (24), it follows that

$$\frac{d\mathcal{E}_r}{dL} = 2\,\mathrm{Re}\left\{\left(G_{01}\frac{dH_{01r}}{dL} - S_{02}\frac{sH_{02r}}{dL}\right)\left(G_{01}^* H_{01r}^* - S_{02}^* H_{02r}^*\right)\right\},$$

$$\frac{d\mathcal{E}_r}{d\phi} = 2\,\mathrm{Re}\left\{\left(G_{01}\frac{dH_{01r}}{d\phi} - S_{02}\frac{dH_{02r}}{d\phi}\right)\left(G_{01}^* H_{01r}^* - S_{02}^* H_{02r}^*\right)\right\}$$

where

$$\frac{dH_{01r}}{dL} = \frac{\left(\frac{dH_{01}}{dL}\right) * \gamma - \frac{dH_{01}}{dL} H_{01}^{*2}}{\left(H_{01}H_{01}^* + \gamma\right)^2},$$

$$\frac{dH_{01r}}{d\phi} = \frac{\left(\frac{dH_{01}}{d\phi}\right) * \gamma - \frac{dH_{01}}{d\phi} H_{01}^{*2}}{\left(H_{01}H_{01}^* + \gamma\right)^2},$$

$$\frac{dH_{02r}}{dL} = \frac{\left(\frac{dH_{02}}{dL}\right) * \gamma - \frac{dH_{02}}{dL} H_{01}^{*2}}{\left(H_{02}H_{02}^* + \gamma\right)^2},$$

$$\frac{dH_{02r}}{d\phi} = \frac{\left(\frac{dH_{02}}{d\phi}\right) * \gamma - \frac{dH_{02}}{d\phi} H_{02}^{*2}}{\left(H_{02}H_{02}^* + \gamma\right)^2}.$$

It suffices to find $\frac{dH_{01}}{L}, \frac{dH_{01}}{d\phi}, \frac{dH_{02}}{dL}, \frac{dH_{02}}{d\phi}$. Using (14) and (15), we obtain

$$\frac{dH_{01}}{dL} = \frac{j\Omega e^{-j\Omega L}(2 + 2jL\Omega)}{L^2\Omega^2} - \frac{2e^{-j\Omega L}(2 + j\Omega L) - 2}{L^3\Omega^2},$$

$$\frac{dH_{01}}{d\phi} = \frac{-2\Omega_y\left(1 + \tan(\phi)^2\right)e^{-j\Omega L}}{\Omega^2} - 2\frac{\left(e^{-j\Omega L}(2 + 2j\Omega L) - 2\right)\Omega_y\left(1 + \tan(\phi)^2\right)}{L^2\Omega^3}$$

$$\frac{dH_{02}}{dL} = \frac{-2jL\Omega e^{-jL\Omega} - 2jL\Omega e^{-2jL\Omega} - 4L^2\Omega^2 e^{-2jL\Omega} - 4e^{-jL\Omega} + 4}{3L^3\Omega^2},$$

$$\frac{dH_{02}}{d\phi} = \frac{2\Omega_y\left(1 + \tan\phi^2\right)}{3L\Omega^2}\left(-je^{-jL\Omega} + je^{-2jL\Omega} + 2L\Omega e^{-2jL\Omega}\right) - \frac{4\Omega_y\left(1 + \tan(\phi)^2\right)}{L^2\Omega^3}\left(e^{-jL\Omega} - 1 + jL\Omega e^{-2jL\Omega}\right)\Omega^2 - 2\frac{\left(e^{-j\Omega L}(2 + 2j\Omega L) - 2\right)\Omega_y\left(1 + \tan(\phi)^2\right)}{L^2\Omega^3}.$$

REFERENCES

[1] D. Burr, "Motion Smear," *Nature*, vol. 284, pp. 164-165, Mar. 1980.

[2] A.-C. Aho, K. Donner, S. Helenius, L.O. Larsen, and T. Reuter, "Visual Performance of the Toad (*bufo bufo*) at Low Light Levels: Retinal Ganglion Cell Responses and Prey-Catching Accuracy," *J. Comp. Physiology A*, vol. 172, pp. 671-682, 1993.

[3] C.H. Graham and R. Margaria, "Area and the Intensity-Time Relation in the Peripheral Retina," *Am. J. Physiology*, vol. 113, p. 299, 1935.

[4] J.B. Barlow, "Temporal and Spatial Summation in Human Vision at Different Background Intensities," *J. Physiology*, London, vol. 141, p. 337, 1958.

[5] L.O. Larsen and J.N. Pedersen, "The Snapping Response of the Toad, *bufo bufo*, Towards Prey Dummies at Very Low Light Intensities," *Amphibia-Reptilia*, vol. 2, pp. 321-327, 1982.

[6] D.H. Ballard, "Cortical Connections and Parallel Processing: Structure and Function," *The Behavioral and Brain Sciences*, vol. 9, pp. 67-91, 1986.

[7] K.E. Martin and J.A. Marshall, "Unsmearing Visual Motion: Development of Long-Range Horizontal Intrinsic Connections," *Advances in Neural Information Processing Systems*, S.J. Hanson, J.D. Cowan, and C.L. Giles, eds. San Mateo, Calif.: Morgan Kaufmann, 1993.

[8] H.C. Andrews and B.R. Hunt, *Digital Image Restoration*. Englewood Cliffs, N.J.: Prentice Hall, 1977.

[9] A.K. Katsaggelos and K.-T. Lay, "Maximum Likelihood Identification and Restoration of Images Using the Expectation-Maximization Algorithm," *Digital Image Restoration*, A.K. Katsaggelos, ed., chaps. 6, 7, pp. 143-208. Springer-Verlag, 1991.

[10] G. Pavlovic and A.M. Takalp, "Maximum Likelihood Parametric Blur Identification Based on a Continuous Spatial Domain," *IEEE Trans. Image Processing*, vol. 1, pp. 496-504, Oct. 1992.

[11] *Digital Image Restoration*, A.K. Katsaggelos, ed. Springer-Verlag, 1991.

[12] J.K. Aggarwal and N. Nandhakumar, "On the Computation of Motion from Sequences of Images—A Review," *Proc. IEEE*, vol. 76, pp. 917-935, Aug. 1988.

[13] R.C. Gonzalez and P.A. Wintz, *Digital Image Processing*, second edition. Addison-Wesley, 1987.

[14] N.B. Karayiannis and A.N. Venetsanopoulos, "Regularization Theory in Image Restoration—The Stabilizing Functional Approach, " *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 38, no. 7, pp. 1,155-1,179, July 1990.

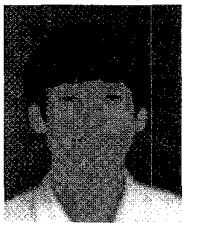
[15] B.K.P. Horn and B.G. Schunck, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.

[16] B.C. Kuo, *Digital Control Systems*. Holt,Rinehart and Winston, 1980.

[17] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, N.J.: Prentice Hall, 1987.

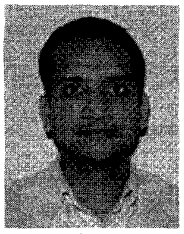
[18] Newport Corporation, *The 1994 Newport Catalog*.

[19] J.L. Barron, D.J. Fleet, and S.S. Beauchemin, "Systems and Experiment: Performance of Optical Flow Techniques," *Int'l J. Computer Vision*, vol. 12, no. 1, pp. 43-47, 1994.

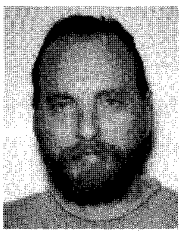


Wei-ge Chen received the BS degree from Beijing University, Beijing, China, in 1989 and the MS degree from the University of Virginia, Charlottesville, in 1992, both in biophysics. He received the PhD degree from the University of Virginia in 1995 in electrical engineering.

Since 1995, Dr. Chen has been with Microsoft Corporation where he works on the development of advanced video compression technology. His research interests include image/video processing, analysis, and compression.



N. Nandhakumar received the BE (Hons) degree in electronics and communication engineering from the P.S.G. College of Technology, University of Madras, India, in 1981, the MSE degree in computer, information, and control engineering from the University of Michigan, Ann Arbor, in 1983, and the PhD degree in electrical engineering from the University of Texas at Austin in 1987. He is currently an assistant professor of electrical engineering at the University of Virginia and also director of the Machine Vision Laboratory. He teaches graduate courses in the areas of computer vision, image processing, and pattern recognition. He was awarded the 1993 Young Faculty Teaching Award by the University of Virginia's Electrical Engineering Department. The MS theses of his students have been recognized by awards from the Masters of Innovation contest by Zenith Data Systems and the B.F. Goodrich Collegiate Inventors Program. The PhD research of his students has resulted in the filing of a patent on a new sensor and algorithms for motion estimation from image blue. Prof. Nandhakumar directs sponsored research in multisensory computer vision, motion computation, stereoscopic image analysis, and applications of machine vision technology in medicine and industry. The results of his research have appeared in more than 70 papers in journals and conference proceedings and book chapters. He is associate editor of the journal *Pattern Recognition* and has been guest editor of several special issues on sensor fusion. He was chairman of the 1994, 1993, and 1992 SPIE Conferences on Sensor Fusion and Aerospace Applications, cochairman of the 1991 SPIE Conference on Applications of Artificial Intelligence, a member of the international advisory committee of the second Asian Conference on Computer Vision, and a member of the program committees of the 1993 IEEE CVPR Conference and of the first and second IEEE International Conferences on Image Processing. He also participated in the organization of other workshops on computer vision, pattern recognition, and image analysis. Prof. Nandhakumar is a senior member of the IEEE and a member of the SPIE.



Worth N. Martin received the BA degree in mathematics, the MA degree in computer science, and the PhD degree in computer science (1981), all from the University of Texas at Austin.

Dr. Martin is an associate professor of computer science at the University of Virginia where he has been a member of the faculty since 1982. His research interests include dynamic scene analysis, digital libraries (primarily access methods for image databases), probabilistic problem solving methods (e.g., genetic algorithms), and applications of computer vision (e.g., computer-aided analysis of imagery from archaeological artifacts). In addition to more than 20 journal articles, his published work included the book *Motion Understanding: Robot and Human Vision*, coedited with J.K. Aggarwal (1988).

Dr. Martin is a senior member of the IEEE and a member of the IEEE Computer Society, the IEEE Technical Committee on Pattern Analysis and Machine Intelligence, the Association for Computing Machinery, the American Association for Artificial Intelligence, and the International Society for Genetic Algorithms. He is also on the editorial board of the journal *Evolutionary Computation*.