Computer Vision and Image Understanding **79,** 162–183 (2000)
doi:10.1006/cviu.2000.0845, available online at http://www.idealibrary.com on IDEAL®

Motion-Based Compression of Underwater Video Imagery for the Operations of Unmanned Submersible Vehicles

S. Negahdaripour

Electrical and Computer Engineering Department, University of Miami, Coral Gables, Florida 33124-0640
E-mail: shahriar@miami.edu

and

A. Khamene

Underwater Vision and Imaging Laboratory, Electrical and Computer Engineering Department, University of Miami, Coral Gables, Florida 33124

Received March 23, 1999; accepted February 4, 2000

Unmanned vehicles are employed more and more frequently for a range of scientific and commercial undersea applications. However, the critical dependency on a tether link, mainly for the transmission of live images to the surface for command and control, is a significant technological obstacle limiting vehicle maneuverability. The elimination of the tether requires the capability to compress a massive amount of live video data to meet the bandwidth limitations of acoustic telemetry. This paper addresses motion-compensated compression of underwater video imagery and the extraction of the sought-after transformations by the application of the brightness constancy assumption and a generalized dynamic image model for the analysis of time-varying imagery. Two approaches, suitable for automatic vision-based navigation or operator-assisted missions of unmanned submersible vehicles, are considered. In the former case, the 3D motion and position information extracted from the raw data by the vision-based navigation system is used directly to perform the compression. In the latter case, the compression is achieved using the motion and radiometric information extracted from the live and reconstructed images at the surface station. To evaluate the performance of proposed methods, results from experiments with synthetic and real data are presented and compared to compression methods that are based on the brightness constancy model.

Key Words: motion; underwater imagery; motion-based video compression; unmanned submersible vehicles; ROVs; AUVs.

1077-3142/00 $35.00

1. INTRODUCTION

Vision-based guidance, either automatic or operator-assisted, is a critical capability in the deployment of unmanned submersible vehicles for a variety of operations: exploration of thousands of square miles of unknown deep ocean environments and utilization of their vast resources, geological and archaeological surveys, maintenance and repair of underwater structures, and performance of surveillance and reconnaissance missions for national security.

In human-assisted (or supervised) missions, operators currently depend primarily on visual information from the site images. These are transmitted via a fiber-optic cable in a tether that also provides power and the command and control signals. Elimination of the tether dependency is a technological breakthrough that would tremendously increase the maneuverability and flexibility of these vehicles. The primary obstacle is the limitations of acoustic telemetry, namely the inability to transmit massive amounts of real-time video data via low-bandwidth acoustic channels. For TV-quality monochrome video, a compression ratio of nearly $1250:1$ is the goal [10]. To achieve this, both the spatial and the temporal redundancies of video imagery need to be reduced by employing transform and motion-compensated compression coding techniques.

Time-varying imagery encodes information about various scene events, including the motion of the viewer or objects in a scene, spatial arrangements of objects, and temporal illumination variations. Extraction of information about these factors enables the construction of the image at some time instant $t + \delta t$ from the image at time t. In motion-based compression, knowledge about such events, known a priori and (or) extracted from the image sequence, is exploited in order to arrive at a less redundant representation of the large volume of spatiotemporal data. In particular, information about the image displacements of moving objects can be used to predict their positions in a new frame, starting with positions in the previous frame. In principle, the motion-based prediction can be applied to individual image points, regions, or subblocks. In this context, optical flow or block matching techniques from visual motion studies have important applications in the estimation of the sought-after displacements in an image sequence.

In most terrestrial environments, the scene illumination is relatively steady, both spatially and temporally, and is furnished by a distant source. Variations in scene radiance (and thus image irradiance) due to changes in camera/viewer position are predominant primarily over surface regions with specular reflectance characteristics. In most other regions, temporal shading variations in video sequences are often negligible, gradual, or overshadowed by the surface textures of scene surfaces. As a result, brightness variations in an image sequence are primarily due to motion effects, as is commonly assumed in most optical flow or motion estimation methods based on the so-called brightness constancy assumption (BCM) (e.g., [11]). This somewhat ideal condition in various applications of motion estimation from time-varying imagery is seldom encountered in underwater environments.

In shallow waters with natural sky lighting, shading artifacts can be induced by surface wave actions. Other complexities arise from a higher level of turbidity and scattering due to various suspended particles and marine organisms. In deep ocean operations, artificial lighting is necessary and is provided by a source(s) installed on the vehicle. Consequently, video imagery of deep-sea environments is typically enriched with strong shading variations induced by the light-source motion. Backscatter effects can also become more severe when a short camera–source baseline is inevitable due to operational requirements, including

optimizing power usage. Often, illumination-induced variations are as large as those induced solely by the displacement of the camera. As a result, accurate estimation of image motion requires decoupling and distinguishing between the induced motion-based and illumination-based variations in a sequence.

Feature-based methods for motion analysis are relatively robust with respect to photometric variations, including illumination nonuniformity [32]. These, however, have limited application in deep-sea environments where most natural objects are rounded, lacking the desired geometric characteristics. The unfailing environmental feature for motion-based visual tasks, including tracking, locking, and stabilization, is the (fine) texture of the surfaces of the seafloor, reefs, and various natural objects. To exploit the corresponding visual cues, we are led to the application of intensity-, texture-, and gradient-based methods as more suitable approaches, though extra care is necessary to permit the extraction of accurate information where geometric and photometric variations can be equally dominant.

A significant advantage in the processing of the underwater video imagery over most terrestrial applications, where one often has to deal with multiple object motions, both rigid and nonrigid (e.g., HDTV and video conferencing), is that we are primarily dealing with the 3D rigid body motion of the vehicle relative to a stationary background, namely, the seafloor environment. Though moving objects, e.g., swimming fish, are encountered occasionally, these often affect an image stream over isolated regions, both spatial and temporal. Thus, solving a complex segmentation problem can be avoided. Furthermore, the parameters describing the 3D vehicle motion from one image to the next may be used directly in the compression of the video frames.

This paper explores the application of motion-compensated prediction coding of underwater imagery, to permit efficient transmission of live video from a submersible vehicle to a surface station through low-bandwidth acoustic channels. The critical application is in the human-assisted missions, where the visual information enables the operator to guide and control the vehicle. However, it is also useful in some autonomous operations, where it may still be desirable to track the submersible vehicle (based on the transmitted motion information) or to reconstruct the scene images for high-level guidance, online mission planning, etc.

1.1. Relevant Literature

Motion-based compression has been investigated extensively for terrestrial applications, including teleconferencing, HDTV, multimedia, and entertainment videos (e.g., [4, 6, 8, 14, 30]). It involves exploiting the temporal redundancy in a video sequence, as consecutive images often differ very little when acquired at a high frame rate. According to [30], (1) the critical problem in motion-compensated compression is the image motion estimation algorithm, and (2) most of the algorithms for interframe coding assume three conditions: motion is predominantly a translation parallel to the scene, with no (negligible) camera zoom and object/camera rotation; illumination is spatially and temporally uniform; object occlusion and background uncovering are negligible. Under these conditions, motion estimation is carried out using a number of approaches, including block matching (correlation-based or search methods), gradient-based optical flow estimation, and various recursive techniques. Many such techniques have been either borrowed from the motion vision literature (e.g., [3]) or developed independently in the video coding literature. The main drawback of these methods for underwater applications is the underlying assumption of steady and uniform

illumination conditions, though this has been assumed in one earlier investigation of low-bit coding of underwater imagery [10]. This restrictive constraint has been relaxed in a number of other investigations [17, 18, 20, 21].

The application of a generalized dynamic image model (GDIM) to the analysis of time-varying imagery and a gradient-based formulation to the computation of optical flow has been discussed in [20, 25]. In particular, it has been shown that this model can successfully decouple the geometric and photometric variations in an image sequence, under most imaging conditions and environmental properties commonly encountered in the deep sea [21, 25]. Using the GDIM, a direct method for the computation of 3D motion and its application to automatic optical station-keeping of underwater vehicles have been presented in [22, 23, 26]. A recursive method for motion-compensated prediction based on the simplified form of the GDIM, one that is investigated in this study, has been described in [30, pp. 356–360].

In this paper, the application of the GDIM-based optical flow and 3D motion estimating methods to the compression of underwater video imagery is investigated, in two distinct scenarios involving the operations of untethered submersible vehicles. One of these is tied directly to the main scope of our research in recent years, which is to realize the capabilities of vision-based navigation and positioning for remotely operated vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs). Here, the image compression is achieved based on the direct use of the 3D motion information that is extracted from the raw video data, as a free by-product of the vision-based navigation system. The second case involves operator-assisted missions with or without information from other sensors (e.g., INS, acoustic, and rate gyros), where the primary objective is a high compression ratio with enhanced reconstruction quality. This is achieved by the classical DPCM-type approach where the motion information is extracted by processing the live image and the previous frame from the reconstructed sequence (e.g., at the destination or operator station).

For image and video compression, the motion picture expert group (MPEG) algorithms rely on the nearly optimal energy compaction properties of the discrete cosine transform (DCT), permitting high-quality image reconstruction at relatively low compression rates [12]. For high-compression scenarios, however, the reconstructed data suffers from the visually undesirable "blocking" effect. Alternatively, a separable two-dimensional discrete wavelet transform (DWT) permits the decomposition of an image into detail subbands for horizontal, vertical, and diagonal spatial orientations. The DWT-based compression methodology provides a framework for coding the wavelet coefficients effectively by exploiting the minimum sensitivity of the human visual system (HVS) to the diagonal contrasts. A comparison between the DCT and DWT for underwater imagery has been reported, utilizing the peak signal-to-noise ratio (PSNR) curve versus the compression rate as the performance criterion [9]. It has been shown that the DCT yields a higher PSNR for lower compression rates, but the DWT has superior performance for very high compression rates. Thus, the application of the DWT to video coding, as investigated in this work, becomes attractive as a way to meet the very high compression requirement for the interframes. In [13], the hardware implementation of a transform-based compression algorithm for AUV telemetry has been discussed.

To the best of our knowledge, this paper reports the first attempt to investigate a motion vision-based system for simultaneous passive navigation and high-compression video transmission. In particular, ROV and AUV operations include ideal applications where this technology is useful and tremendously desired. Furthermore, our results provide a quantitative

measure for comparing the PSNR performance of BCM-based and GDIM-based motion estimation algorithms for the coding of underwater imagery, in addition to determining the achievable compression factors using the proposed methods.

The remainder of this paper is organized as follows. In Section 2, we give some background on motion information in time-varying imagery, provide a brief review of motion-compensated prediction as relevant to this study, and review the GDIM and the application to the direct computation of optical flow or 3D motion from the information in the image brightness and spatiotemporal gradients. In Section 3, we present the DWT-based video compression techniques that are investigated, with implementation details given in Section 4. In Section 5, experiments with synthetic and real data are presented to compare results from various GDIM-based and BCM-based methods. Finally, the contributions of this paper are summarized in Section 6.

2. PRELIMINARIES

2.1. Motion Cues in Time-Varying Imagery

In analyzing time-varying imagery and the estimation of motion from image sequences, a viewer-centered Cartesian coordinate system XYZ is commonly assumed, with origin at the focal point F (projection center) and the Z axis along the optical axis. The xy image plane is chosen to be parallel to the XY plane at a distance f from the origin (f is the effective focal length).

Let $I(x, y, t)$ denote the image irradiance at time t for an image point $\boldsymbol{r}(t) = [x, y, f]^T$, which is the projection of a scene point $\boldsymbol{R} = [X, Y, Z]^T$. Assuming perspective projection, we have $\boldsymbol{r} = \frac{1}{Z}\boldsymbol{R}$. We would interchangeably use $I(t)$, $I(x, y)$, and $I(x, y, t)$ where appropriate. As the camera moves relative to the scene, $\boldsymbol{P}$ projects onto a new point $r(t + \delta t) = [x + \delta x, y + \delta y, f]^T$ at the next time instant $t + \delta t$. In addition, the image irradiance of $\boldsymbol{P}$ may change by δI due to a number of factors, including variations in scene illumination. Based on this, we can write

$$I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t) + \delta I(x, y, t). \tag{1}$$

We can decompose the instantaneous camera motion relative to a stationary scene into 3D translational and rotational velocities $\boldsymbol{t} = [t_x, t_y, t_z]^T$ and $\boldsymbol{\omega} = [\omega_x, \omega_y, \omega_z]^T$, respectively, where the rotation is about some axis through the camera focal point F. The image velocity of a point, the so-called optical flow $[u, v, 0]^T$, is given by [16]

$$\begin{bmatrix} u \\ v \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{\delta x}{\delta t} \\ \frac{\delta y}{\delta t} \\ 0 \end{bmatrix} = \boldsymbol{A}\boldsymbol{\omega} + \frac{1}{Z}\boldsymbol{B}\boldsymbol{t}, \tag{2a}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \frac{xy}{f} & -\left(\frac{x^2}{f} + f\right) & y \\ \frac{y^2}{f} + f & -\frac{xy}{f} & -x \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} -f & 0 & x \\ 0 & -f & y \\ 0 & 0 & 0 \end{bmatrix}.$$

Equation (2a) shows the well-known fact that the image motion encodes information about both the instantaneous motion of the camera $\{\boldsymbol{t}, \boldsymbol{\omega}\}$ and the depth map $Z(x, y)$ of the scene surfaces. The use of the differential formulation for the modeling of image displacements is appropriate for small motions of about 1 to 2 pixels per frame. To ensure this, one often requires a high frame rate. Larger motions are typically addressed by the application of a multiresolution scheme, where the motion estimation is carried out incrementally over various image scales (e.g., [5]). We also note the scale-factor ambiguity due to perspective projection geometry, as the translation $\boldsymbol{t}$ and depth Z can be multiplied by the same nonzero constant without changing the above equation. The immediate implication is that the estimation of translational motion $\boldsymbol{t}$ and depth Z is feasible up to a scale factor only. Thus, we may recover the direction of translation $\hat{t} = t/|t|$ and scaled depth $\hat{Z} = Z/|t|$, without loss of generality.

It is often useful to examine the variation in image motion $[\delta u, \delta v, 0]^T$ due to the depth perturbation δZ. One can readily compute this from

$$\begin{bmatrix} \delta u \\ \delta v \\ 0 \end{bmatrix} = -\left(\frac{\delta Z}{Z^2}\right) \boldsymbol{B}\boldsymbol{t} = -\left(\frac{\delta Z}{Z}\right) \begin{bmatrix} u_t \\ v_t \\ 0 \end{bmatrix}. \tag{2b}$$

Thus, the image motion variations as a percentage of the translational component $[u_t, v_t, 0]^T$ are directly proportional to the percentage depth variations. An immediate implication of Eq. (2b), also noting the restriction on the image motion size, is as follows: The error in the image motion, if we assume a scene with constant depth Z_o, is negligible if the discrepancy between the true depth Z and Z_o is small (i.e., for $\delta Z/Z_o \ll 1$, where $\delta Z = Z - Z_o$).

2.2. Motion-Based Compression

Using Eq. (1) and knowledge of the geometric and radiometric transformation fields, $\{\delta x, \delta y\}$ and δI, respectively, one can reconstruct the image $I(t + \delta t)$ from $I(t)$. Effective motion-based video compression involves determining a small number of parameters that describe these transformations with sufficient accuracy.

In the absence of motion and depth discontinuity boundaries, the image motion $\{\delta x, \delta y\}$ varies smoothly over the entire image; see Eq. (2a). In the presence of such discontinuities, the variations would be smooth only within regions defined by these boundaries. For most underwater applications, δI is predominantly due to intensity changes induced by medium attenuation and scene illumination variations. These can be modeled as $\delta I = I\delta m$, where δm is a slowly varying multiplier field. However, discontinuities in δm can exist due to cast-/self-shadows and occlusion. In such cases, δm would vary smoothly within regions separated by the sparse discontinuity boundaries.

A possible representation of $\{\delta x, \delta y\}$ is in terms of the motion vector $\boldsymbol{m} = \{\boldsymbol{t}, \boldsymbol{\omega}\}$ and the depth Z of the scene surfaces according to Eq. (2a). In this case, we may employ a small number of parameters to describe the depth map $Z(x, y)$ either over the entire scene or within selected small subregions. As an example, both the depth Z and multiplier δm fields may be assumed to be constant over image subblocks or described in terms of the coefficients of low-order polynomials over the entire image. Consequently, a few parameters can be used to approximate the geometric and radiometric variations from image $I(t)$ to the next image $I(t + \delta t)$, allowing the construction of $I(t + \delta t)$ from $I(t)$ and these parameters

with reasonable precision. The discrepancy between the estimated and true images, the prediction error $e(t+\delta t) = \hat{I}(t+\delta t) - I(t+\delta t)$, is expected to be highly uncorrelated and can be encoded more efficiently than $I(t+\delta t)$.

Where the application allows, a lossy technique exploiting some form of error quantization can achieve a higher compression ratio. Unfortunately, this leads to image-quality deterioration as the reconstruction errors accumulate. This is overcome by coding the original image, the so-called reset frame, at regular intervals using standard methods for still-picture compression. The duration of the time slot between consecutive reset frames is chosen based on various relevant factors or criteria, e.g., channel capacity and bandwidth, maintaining a minimum PSNR for acceptable reconstruction, etc. The reset frames are coded at a relatively low compression ratio to achieve a better reconstruction quality for all the pursuing frames in the same time slot. Thus, the compression technique involves a hybrid coding scheme, with two modules for motion-compensated prediction error coding and intraframe coding of the reset image.

Motion picture expert group algorithms exploit the DCT for its optimal energy compaction properties [12]. Though high-quality image reconstruction can be achieved for relatively low compression ratios, the visually undesirable "blocking" effect is observed for high-compression applications. Alternatively, a separable two-dimensional DWT may be utilized to decompose an image into detailed subbands with horizontal, vertical, and diagonal spatial orientations. The low sensitivity of the HVS to the diagonal contrast can be exploited to devise efficient DWT-based compression algorithms. A comparison between the DCT-based and DWT-based compression techniques for underwater imagery has been reported, using the PSNR curve versus the compression rate as a criterion [9]. While the DCT performs better for lower compression rates, the DWT achieves higher PSNR for very high compression rates. This motivates the investigation of DWT-based algorithms in applications constrained by very high compression rates, including the underwater video transmission through low-bandwidth acoustic channels.

2.3. BCM and GDIM for Time-Varying Image Analysis

Most motion-based compression methods exploit merely the geometric transformation field $\{\delta x, \delta y\}$, by implementing some form of the restrictive BCM for the computation of optical flow [30]. These methods treat the radiometric field $\delta I = I(x+\delta x, y+\delta y, t+\delta t) - I(x, y, t)$ as part of the prediction error field, which can be a reasonable strategy for a band-limited error field. Unfortunately, the BCM-based optical flow methods (e.g., [3, 11] explicitly rule out the presence of radiometric variations, and thus they often fail drastically when δI is of the same order as the component of the temporal variation, $I_t = I(x, y, t+\delta t) - I(x, y, t)$, that is solely induced by the camera motion; that is, these methods would yield an erroneous and noisy displacement field $\{\widehat{\delta x}, \widehat{\delta y}\}$ (e.g., see [21]). Consequently, geometric warping of images in a sequence based on such estimates leads to registration errors, and the error field is burdened with accounting for the true radiometric transformation δI, as well as the irradiance adjustment $I(x+\widehat{\delta x}, y+\widehat{\delta y}, t+\delta t) - I(x+\delta x, y+\delta y, t+\delta t)$ that is necessary to correct for the erroneous geometric registration.

Mathematically, this can be expressed as

$$\widehat{\delta I} = \delta I + (I(x+\widehat{\delta x}, y+\widehat{\delta y}, t+\delta t) - I(x+\delta x, y+\delta y, t+\delta t)). \tag{3}$$

Embedded in the error field $\widehat{\delta I}$ are the true radiometric field δI and the correction term

inside the parentheses. The latter is clearly tied to the correlation of the intensity values of neighboring pixels; the finer the surface texture, the more random this error field. It is primarily the contribution from this component that leads to the inefficiency of classical compression techniques based on the BCM. Instead of performing the warping based on an erroneous optical flow, one may code the radiometric correction directly according to

$$\widehat{\delta I} = \delta I + (I(x, y, t + \delta t) - I(x + \delta x, y + \delta y, t + \delta t)). \tag{4}$$

This leads to the application of any standard transform coding method that does not exploit the motion cues (e.g., [6, 30]).

The Taylor series expansion, up to first-order terms, of the left-hand side of Eq. (1) yields

$$I_t \delta t + I_x \delta x + I_y \delta y - \delta I = 0, \tag{5}$$

where $[I_x, I_y, I_t]^T$ is the spatiotemporal gradient at each pixel. This is in the same form as the equation proposed in [7] for the computation of optical flow. The decomposition of δI in terms of the multiplier and offset fields δm and δc,

$$\delta I = I \delta m + \delta c, \tag{6}$$

yields the GDIM for the representation and estimation of geometric and radiometric transformations in an image sequence [25]. Applications of GDIM to the computation of optical flow and direct 3D motion estimation have also been investigated, and the performance against methods based on the BCM has been evaluated [20–22, 25]. The direct 3D motion estimation technique has been employed in the development of a real-time PC-based vision system (running Windows NT) for the automatic station-keeping and navigation of submersible vehicles and for video mosaicking of seafloor imagery [23, 26].

The physical interpretation of the offset and multiplier fields in terms of various scene events is addressed in [25] and is beyond the scope of this paper. It suffices to say that for applications of interest to this investigation, the radiometric transformations are induced predominantly by the medium attenuation and illumination effects, which can be modeled appropriately by a multiplier field δm solely. This leads to the constraint equation

$$I_t \delta t + I_x \delta x + I_y \delta y - I \delta m = 0 \tag{7}$$

for describing the geometric and radiometric transformations in an underwater image sequence. To estimate the sought-after unknown fields, we may minimize the squared error in Eq. (7) over a symmetric region centered at each pixel [21, 25]. This leads to the solution $\boldsymbol{x} = \boldsymbol{M}^{-1}\boldsymbol{d}$, where

$$\boldsymbol{M} = \sum_{\text{Reg}} \begin{bmatrix} I_x^2 & I_x I_y & -I_x I \\ I_x I_y & I_y^2 & -I_y I \\ -I_x I & -I_y I & I^2 \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} \delta x \\ \delta y \\ \delta m \end{bmatrix}, \quad \text{and} \quad \boldsymbol{d} = \sum_{\text{Reg}} \begin{bmatrix} -I_t I_x \\ -I_t I_y \\ I_t I \end{bmatrix}. \tag{8}$$

It has been assumed that $\delta t = 1$, without loss of generality.

Substituting Eq. (2a) into Eq. (7) yields a constraint between the vehicle motion, the scene depth, and the spatiotemporal image gradient at each pixel,

$$I_t - I \delta m + \frac{1}{\hat{Z}}(\boldsymbol{s} \cdot \boldsymbol{t}) + (\boldsymbol{v} \cdot \boldsymbol{\omega}) = 0, \tag{9}$$

where $\boldsymbol{s} = [-f I_x, -f I_y, x I_x + y I_y]^T$ and $\boldsymbol{v} = \frac{1}{f}\boldsymbol{r} \times \boldsymbol{s}$. The use of this equation to estimate

the camera translation direction $\hat{t}$ and yaw motion ω_z and the average scaled depth $\hat{Z}_R$ and multiplier field δm_R over local image regions R has been discussed in [26]. The primary advantage in employing Eq. (9) is to enable the computation of 3D motion, for positioning and navigation applications, directly from the readily available brightness values and spatiotemporal image gradients. Where image compression is the main objective, it may suffice to estimate the optical flow directly from Eq. (8) to obtain the necessary information for image warping and radiometric adjustments. We explore the use of these equations for the motion based video compression of underwater imagery for two different applications of submersible vehicles.

3. MOTION-BASED COMPRESSION OF UNDERWATER IMAGERY: TWO OPERATIONAL SCENARIOS

We envision future generations of intelligent ROVs (or AUVs) with some level of autonomy, either without continuous or with no (little) supervision from a human operator, to be equipped with optical-based navigation and positioning capabilities. Still, it may be desirable to maintain a record of the vehicle's position, based on transmitted motion information, and (or) to reconstruct the scene images for high-level human guidance, online mission planning, etc. In this case, the 3D motion information extracted from the raw video data for automatic vehicle navigation and positioning can be used directly, and consequently the image compression for low-bit transmission can be carried out as a free by-product of the vision-based navigation system.

In this scenario, consecutive frames I^{k-1} and I^k of the raw data are processed by the navigational system to estimate the vehicle motion $\boldsymbol{m}^k$, the depth Z_R^k (distance to the seafloor), and the multiplier field δm_R^k of subblock regions $R = 1, 2, \ldots, N$ [26]. We assume that the scale-factor ambiguity can be resolved by employing some other sensor(s), giving an accurate estimate of *one* range value to the bottom surface.[1] For example, if we know the average depth Z_o, we can compute the sought-after scaling from $|t| = Z_o/\hat{Z}_o$. We further assume that $\boldsymbol{m}$ consists of translation component $\boldsymbol{t}$ and yaw motion w solely, as pitch and roll can be estimated with high precision using readily available gyro systems or inclinometers. These rotations can be discounted, once estimated, by appropriate image warping before the application of our direct 3D motion estimation algorithm [27].

Using the radiometric transformation $\delta I^k = I^k \delta m^k$ and image motion $\{\delta x^k, \delta y^k\}$ for the sequence I^{k-1} to I^k, image I^{k-1} is transformed according to $\hat{I}^k(x, y) = I^{k-1}(x + \delta x^k, y + \delta y^k) - \delta I^k(x, y)$. Ideally, this is the exact replica of I^k. In practice, discrepancies exist due to gray-scale and spatial quantization and inaccuracy in the estimated transformation fields. The error $e^k = \hat{I}^K - I^k$ is calculated and coded using a number of standard techniques (e.g., [30]). The discrete wavelet-transform technique used in our implementation is described in the next section.

Elimination of the tether link is also critical for the human-assisted missions. In these applications of the ROVs, vehicle guidance and control are carried out solely by the operator, who depends mainly or entirely on the (live) scene images on the monitor. High compression ratios become necessary to transmit TV-quality video via an acoustic channel. For this scenario, Eq. (8) or Eq. (9) can still be used to extract the necessary motion and radiometric information. However, improved performance is achieved by determining the

[1] Most, if not all, ROVs and AUVs are equipped with some acoustic sensing device for bottom depth measurements.

transformation from the current raw image to the previous frame reconstructed at the receiver end (e.g., the one displayed on the operator screen), instead of the original data acquired by the vehicle. Mathematically, we now seek the transformation fields that permit the registration of the current image I^k with image $\hat{I}^{k-1}$. The clear advantage is that the degradation of image quality becomes less severe by avoiding the accumulation of reconstruction errors; at every time instant, the estimated transformation is independent of previous observations.

4. IMPLEMENTATION DETAILS

The flow charts in Fig. 1 depict the encoding and decoding stages of the compression algorithms. Two different encoders, shown in (a) and (c), have been used for the two

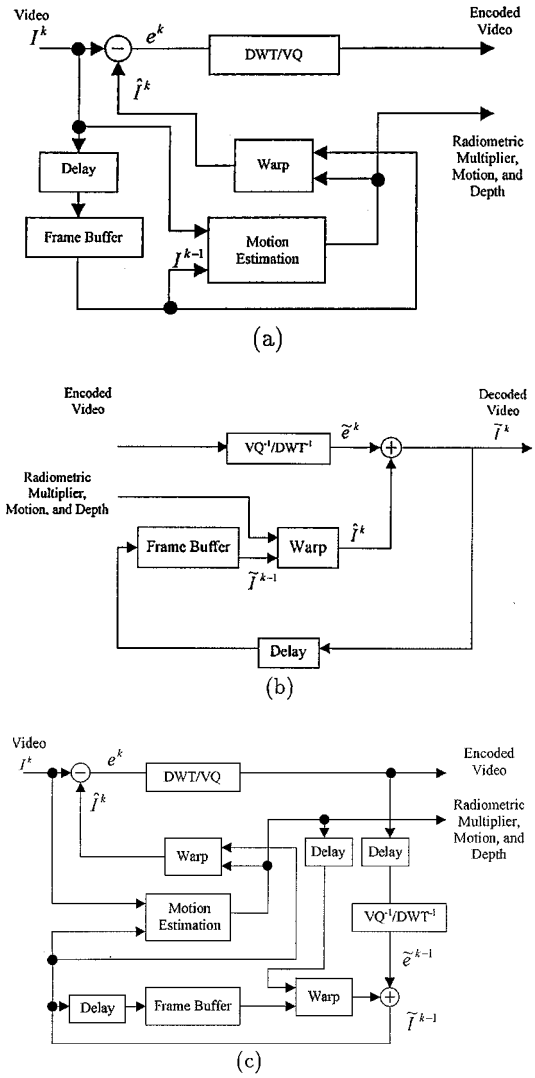

FIG. 1. Flow diagrams of the decoder (a) and the encoders for the two operational scenarios. In (b), we employ the 3D motion information from the vision-based navigation system, coding the registration error of consecutive frames in the raw video. In (c), the DPCM-like encoder utilizes the registration error based on the estimated optical flow for the current frame and the last image decoded at the receiver end (see text for additional details).

operational scenarios described earlier. Each is a hybrid system for motion-compensated prediction error coding to reduce the temporal redundancy and for intraframe coding of the reset image to reduce the spatial redundancy.

The encoder of the first scenario (Fig. 1a) is an open-loop one, where we code the residual error of two consecutive frames of raw data after registration. For the second scenario, a DPCM-like coder has been employed as shown in Fig. 1c. Here, the registration error of the current and the last received/decoded images is coded. As stated, this yields a better reconstruction quality since the motion estimator minimizes the observed error at the receiver end. The same encoders and decoder are used for the reset image, after the frame buffer is cleared.

To compute the prediction error, we require the estimation of multiplier field δm and the image displacements either (1) using the vehicle motion parameters from Eq. (2a) for the first scenario or (2) directly from Eq. (8) for the latter case. For speedup in processing, the current implementation in using Eq. (8) for the estimation of vehicle motion $\boldsymbol{m}$ assumes a constant depth $Z = Z_o$ over the scene, since we are interested in evaluating a system with (near) real-time performance. However, the geometric registration (within the Warp module) can be improved (and the prediction error magnitudes reduced) by determining a 3D depth map Z at the expense of extensive additional computations (e.g., [24, 29]). This improvement is often negligible over most of the image, with the exception of isolated regions where the depth values are grossly different from Z_o; recall from Eq. (2b) that the percentage error in the estimated image displacement depends on the percentage depth variation, and the error is often small when the average image motion is about 1 to 2 pixels. Furthermore, the quantization of the estimated image displacement vectors in the warping stage, as done in our implementation, typically cancels out any such improvements. The prediction error is computed after carrying out the geometric warping and radiometric adjustment using $\delta I = I\delta m$.

Each of the two modules for motion-compensated prediction error coding and intraframe coding has been implemented as a hybrid of a DWT and a vector quantizer (VQ), as described in [2]. In addition to the standard advantages in employing a DWT-based technique (i.e., multiresolution representation, efficient implementation, and high signal energy compaction), this allows us to carry out a direct comparison in performance with the BCM-based compression method in [10] for underwater imagery, using identical coding implementations.

The filter bank of the discrete wavelet transform is constructed from the well-known 9-tap and 7-tap biorthogonal wavelet filter pair, with the LBG algorithm for designing the multiresolution codebook of the VQ. For the training set, we have employed a subset of the images from the three different experiments presented here. The rate range of the codebook is between 0.0625 and 1 bit/pixel. Vectors are quantized using a full-search VQ. For the residual error images in each time slot, four levels of decomposition followed by quantization have been performed. For the reset images, three levels have been sufficient. These yield average compression ratios of approximately 100 : 1 and 16 : 1 for the prediction errors and reset images, respectively. To construct each new frame from the previous frame, the decoder employs the inverse discrete wavelet transform and the image warping and radiometric adjustment based on the transmitted quantized errors and transformation parameters. Alternatively, the inverse discrete wavelet transform is applied directly for reconstructing the reset frame (frame buffer is cleared).

It is useful to comment on the overhead associated with other information that is transmitted. For the direct method, the vehicle motion vector consists of four parameters. In comparison, the flow vectors from the optical-flow-based motion compensation algorithm consist of an estimate over each subblock, assumed to be an 8×8 region. For the 128×128 image

employed in this investigation, we have 16×16 estimates. The radiometric transformation to account for illumination variations in the GDIM is characterized by the multiplier field δm, assumed to be constant over each subblock in the estimation stage (for both the direct and optical flow methods). However, δm typically varies gradually and can be modeled by a smoothly varying two-dimensional function. For a bicubic model, this adds 10 parameters to the amount of data to be transmitted.

5. EXPERIMENTS

Results of experiments with three data sets are reported to compare the performances of the BCM-based and GDIM-based optical flow and 3D vehicle motion estimation methods for motion-based prediction and coding. Image motion calculations were performed on 8×8 subblocks, where GDIM-based methods involved the estimation of the multiplier field δm as well. In the GDIM-based direct technique, the multiplier field for each subblock and the four vehicle motion parameters were estimated using a closed-form solution based on Eq. (9) [26]. A reset rate of 3 Hz, corresponding to a time slot of 10 frames, was assumed.

The results in each experiment include one frame of the sequence to examine the scene characteristics, a sample optical flow from the BCM-based and GDIM-based methods at the end of each time slot, and the corresponding prediction errors. We compare the PSNRs for the prediction errors of these four techniques and that for the coding of the difference image without exploiting the motion cues; see Eq. (4). We also give the PSNR results based on the extraction of motion information from the live image and the previous frame constructed at the receiver end (for the DPCM-like coder in the second scenario). For these last results, we have compared only the two optical-flow-based methods, as the main objective in such applications is to achieve a high-quality video reconstruction with no need for navigation data. For the two direct methods, we additionally present the estimated 3D vehicle interframe motions and overall trajectories.

The first set is a synthetic sequence consisting of 50 images. This was constructed by cropping appropriate sections of a mosaic that we had constructed from a video footage of a shipwreck site. Thus, the data characterizes the average quality of typical underwater imagery. The main purpose for using a synthetic sequence is to evaluate results based on known ground truth. To construct the sequence, an interframe motion of 1 pixel/frame in the X direction was assumed (recall that the estimation of larger motions involves the application of a multiresolution scheme [5], which is beyond the scope of this paper). To simulate radiometric variations due to illumination, each image was scaled using a 2D Gaussian function with σ changing from 30 to 40 pixels in increments of 1 pixel/frame during each time slot. This yields variations that are relatively smooth during each time slot, but sudden at the transition. The second set consists of 50 frames acquired in an indoor water tank that is used for our underwater imaging experiments. The last set is a 300-frame sequence of ocean data acquired by a diver.

The true motion is controlled and known precisely for the synthetic data, somewhat known and controlled for the water tank experiment, and uncontrolled and more or less unknown for the ocean data; though the diver's motion is predominantly, but not exclusively, in the forward and downward direction toward a target on the ocean floor, the camera motion includes all six degrees of freedom. Collectively, these experiments should highlight some important properties of these methods, in addition to providing valuable information about their strengths, weaknesses, and effectiveness for motion-based compression of underwater imagery.

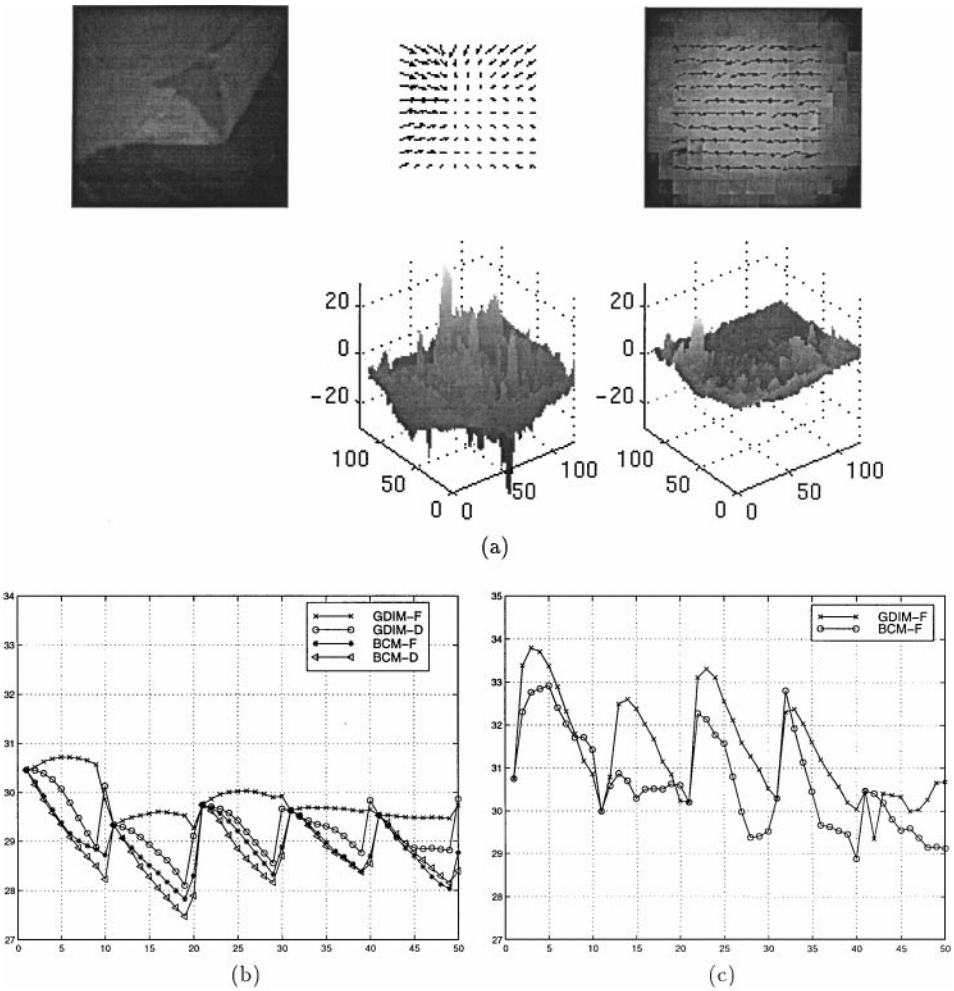

FIG. 2. Synthetic sequence: (a) From left to right and top to bottom, the original image, sample BCM-based and GDIM-based optical flows (superimposed on the estimated δm field), and coded errors from BCM-based and GDIM-based flow methods; (b–c) PSNR results for the first scenario and the DPCM-like encoder, respectively. See text for details.

5.1. Data Set 1

The image in Fig. 2 is one frame of the synthetic sequence. As explained, the motion in this data set is exclusively in the X direction. The BCM-based optical flow estimates depict the impact of the illumination variations. The converging flow vectors can be explained, as each region in the first image is matched with one in the second image with (nearly) equal intensity; recall that the intensity levels drop more quickly in the second image, as the σ of the Gaussian radiometric field is reduced by 1 pixel/frame. The estimated multiplier field δm is shown as a background image of the GDIM-based flow, depicting the smoothness of this field. The prediction errors of the BCM-based and GDIM-based optical flow methods, displayed as surface plots in the middle row, exhibit the size and distribution over the entire image. The means and standard deviations $\{m, \sigma_d\}$ of their distributions are $\{-6.7, 6.5\}$ and $\{0.77, 3.1\}$, respectively.

Figure 2b shows that the GDIM-based flow method gives performance superior to that of the other methods. The PSNR remains steady over each time slot with small variations

at the reset instants. For the other three methods, the monotonic decrease within each slot verifies the significance of the reset frame, to be coded for high-quality reconstruction and transmitted at regular intervals. The PSNR results in Fig. 2c correspond to the GDIM-based and BCM-based compression schemes for the second scenario. As expected, these consistently show performances better than those their counterparts in the previous results.

From Fig. 3, we note that the radiometric variations have adversely impacted the vehicle interframe 3D motion and trajectory that are calculated from the BCM-based method.

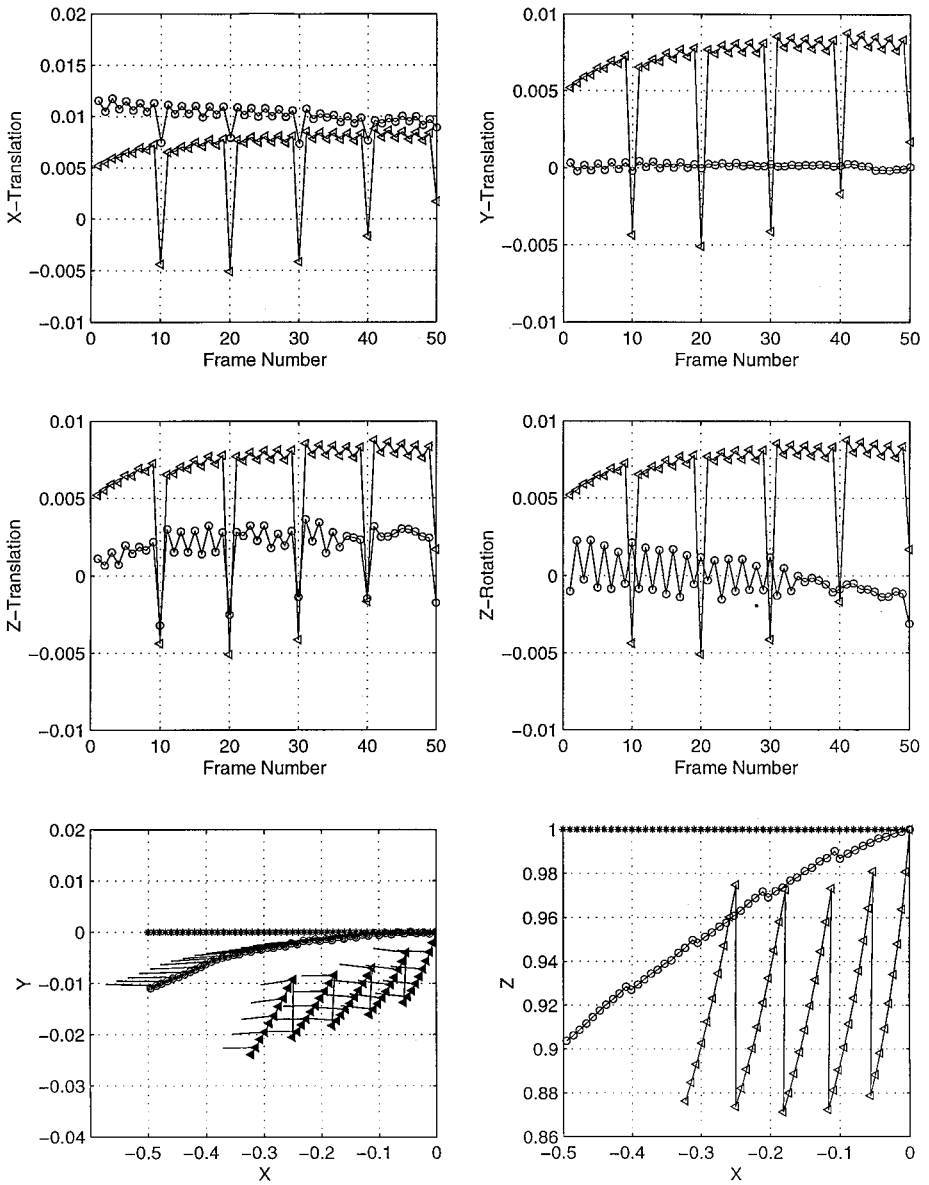

FIG. 3. Synthetic sequence: From left to right and top to bottom, X, Y, Z translation and yaw motion components and the XY and XZ vehicle trajectories from GDIM-based (circles) and BCM-based (triangles) direct methods. Arrows indicate the vehicle heading calculated from the yaw motions. See text for details.

Corresponding estimates from the GDIM-based direct method are fairly accurate. (All motion and trajectory results are expressed in terms of the average scene depth, which is assumed to be one unit of length for the construction of the images in this experiment.)

5.2. Data Set 2

The image in Fig. 4 is a frame of the water tank data. The motion, as recorded visually during the experiment, is predominantly in the X direction, with a relatively small Y component and a very small yaw motion. Equipped with a light source and a downlook camera, the vehicle traveled a distance of about 40 in. at a distance of roughly 50 in. from the bottom surface during the sequence.

The BCM-based optical flow depicts the impact of the illumination variations due to light-source motion, while the same estimate from the GDIM-based method is qualitatively correct according to the motion of the vehicle. The means and standard deviations of the prediction errors are {14.9, 6.8} and {4.3, 2.8} for the BCM-based and GDIM-based flow

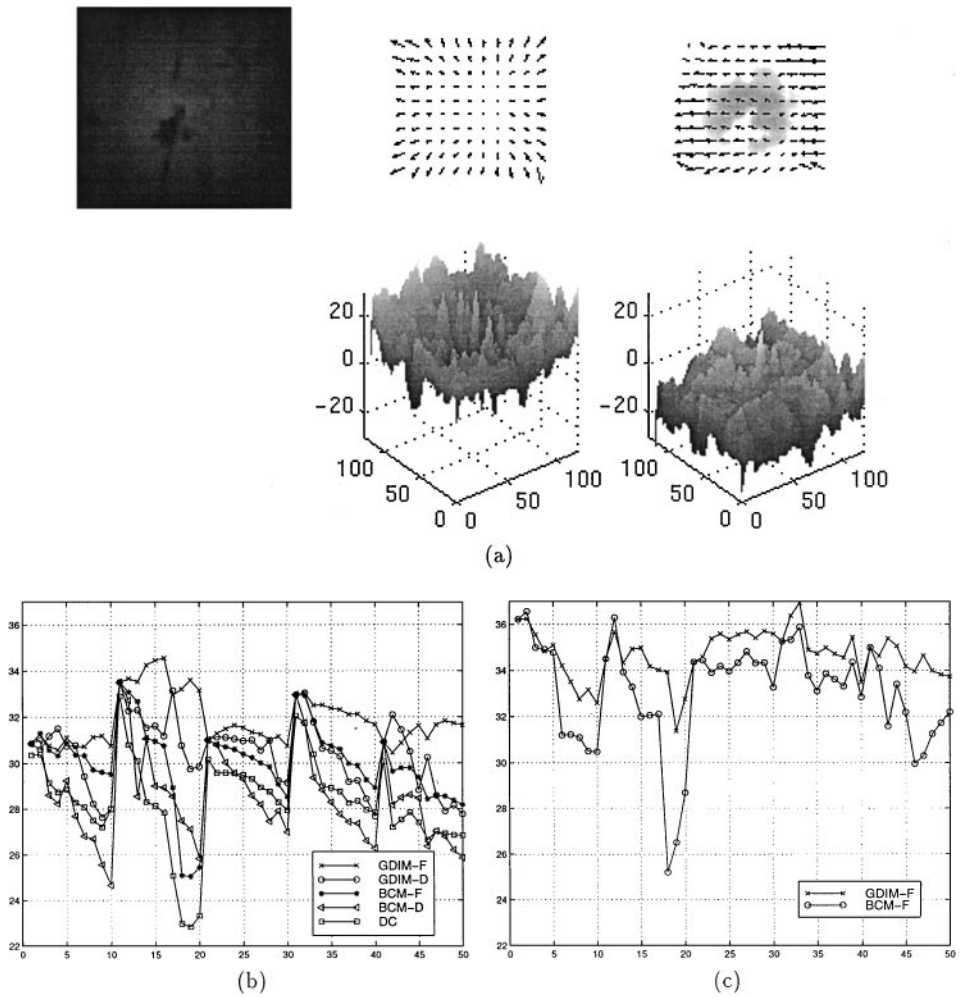

FIG. 4. Water tank sequence: Same results as in the previous experiment, including sample BCM-based and GDIM-based optical flows, coded errors, and PSNR plots.

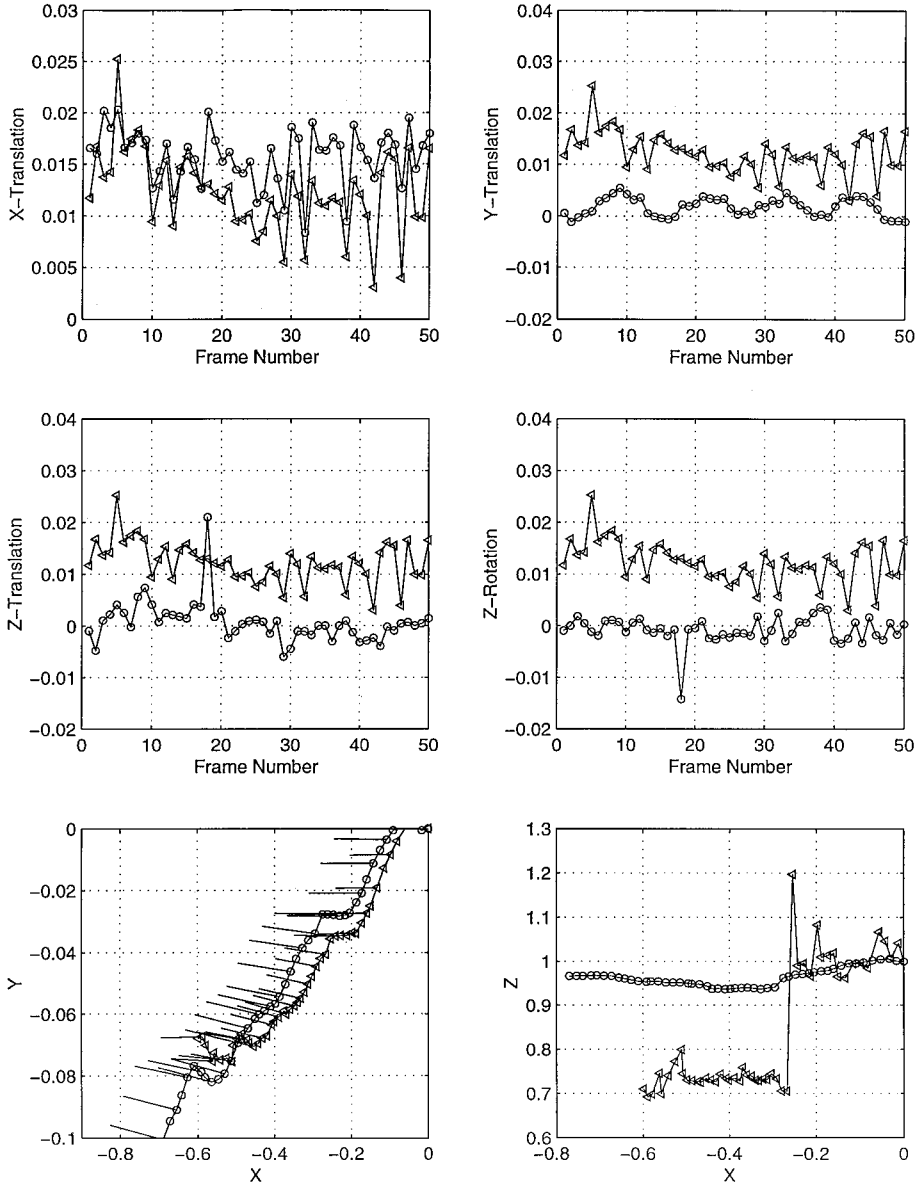

FIG. 5. Water tank sequence: Same motion and trajectory results as in the previous experiment, using the GDIM-based (circles) and BCM-based direct (triangles) methods.

methods, respectively. The PSNR plots in Figs. 4b–4c are relatively consistent with the results in the previous experiment. To evaluate the motion and trajectory results in Fig. 5, we first recall that these are expressed in units of the initial average distance from the bottom, which is about 50 in. (i.e., a motion of 0.02 means 0.02×50 in. $= 1$ in.). Thus, the GDIM-based direct method suggests a total traveled distance of nearly 40 in. in the X direction, 5 in. in the Y direction, and a fairly steady distance from the botton (see Fig. 5). These are in good agreement with the experimental conditions. In contrast, the BCM-based solution underestimates the total traveled distance in the XY plane, in addition to making significant errors in the Z direction (see the XZ trajectory).

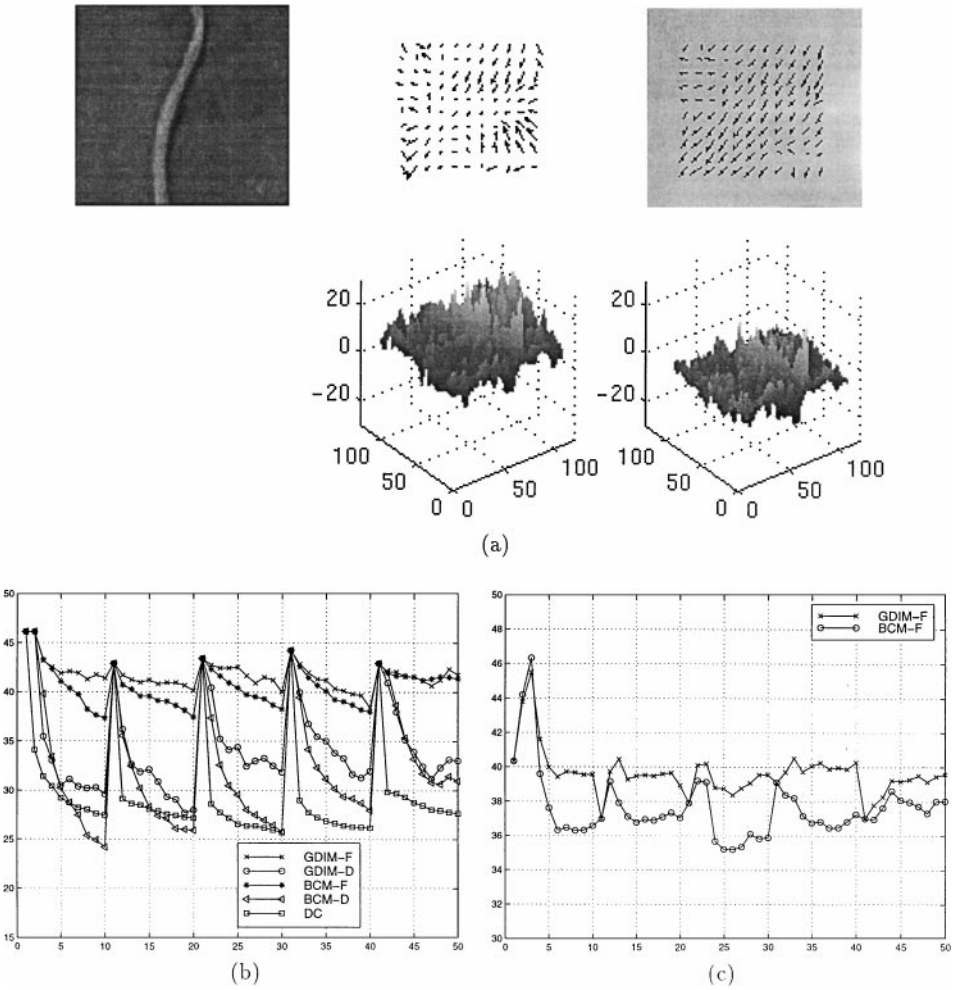

FIG. 6. Ocean sequence: Same results as in the previous experiments, including sample BCM-based and GDIM-based optical flows, coded errors, and PSNR plots.

5.3. Data Set 3

Figure 6 is a sample frame of the 300-frame ocean data, recorded in midday off of the Boca Raton coast. The camera was carried by a diver in relatively shallow waters, about 25 feet deep, at a distance of about 1 to 3 feet from the bottom. During this portion of the sequence, the diver swims down and forward to track a garden hose, laid on the ocean floor.

The prediction errors of the BCM-based and GDIM-based flow methods have means and standard deviations $\{13.2, 7.8\}$ and $\{-0.35, 6.2\}$, respectively. The PSNR plots are given for the first 50 frames only, as the same patterns repeat during the complete sequence. Most of these results are consistent with those from the earlier two experiments, with one exception. To explain, we first comment that the radiometric variations of the sequence are negligible, as the data was recorded in clear shallow water under natural sun light. Thus, the BCM-based optical flow method is expected to provide good results. Furthermore, the uncontrolled motion of the camera, carried by the diver, contains all six degrees of freedom. However, the direct methods have been devised for computing only four motion

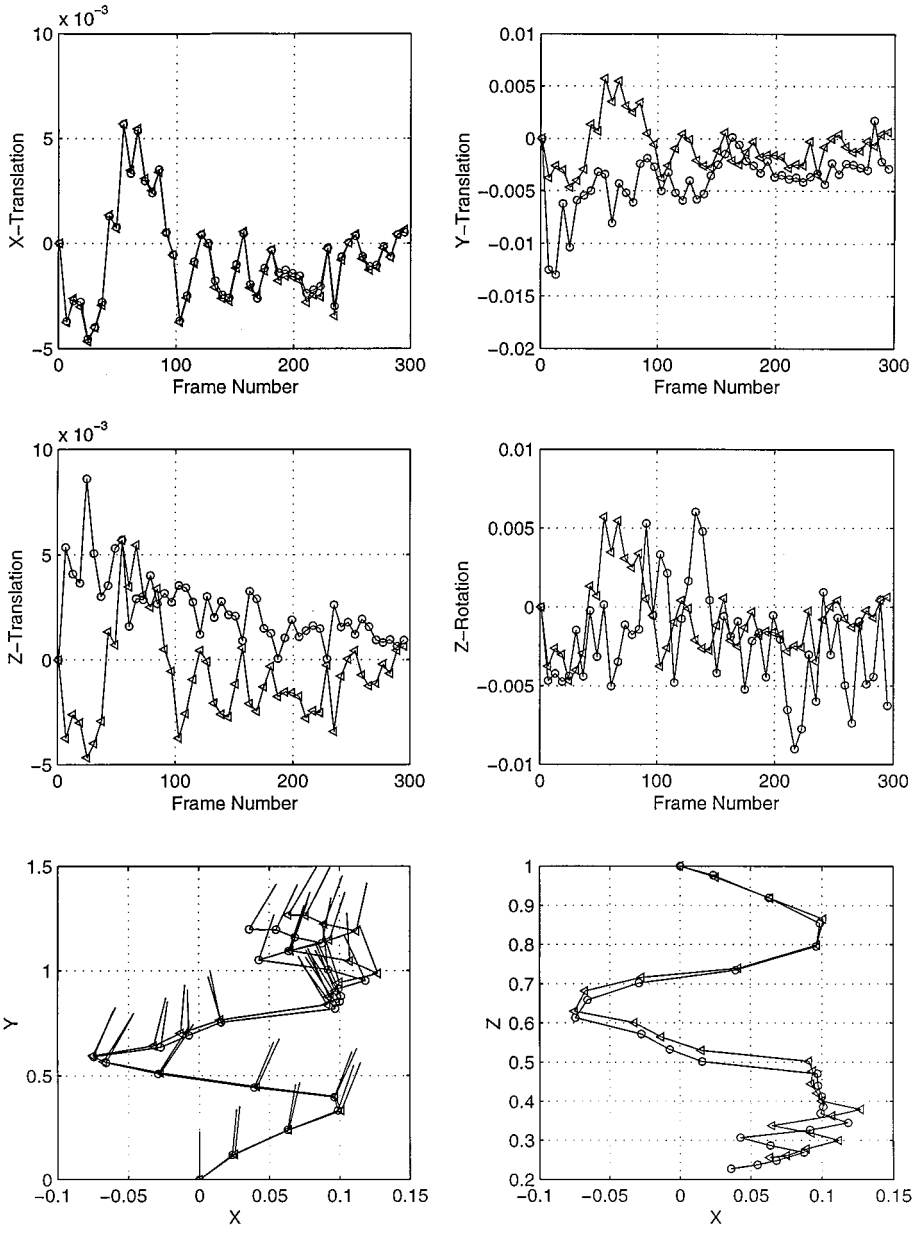

FIG. 7. Ocean sequence: Same motion and trajectory results as in the previous two experiments, using the GDIM-based (circles) and BCM-based direct (triangles) methods.

components; pitch and roll are not calculated. This suggests that the estimated motions are expected to be inaccurate where the instantaneous pitch and roll components of the camera motion are significant. Thus, though the estimated diver trajectories in Fig. 7 appear to be qualitatively correct (i.e., they exhibit forward and downward motion toward the hose on the ocean floor), the quantitative accuracy of the diver's position during the sequence cannot be evaluated. These conclusions may be confirmed by, or are in agreement with, the PSNR results: The GDIM-based direct method does not perform as well as the BCM-based optical flow method, though this does not follow immediately from the one erroneous BCM-based

flow sample. Collectively, it appears that the complex camera motions have been more "destructive" for the GDIM-based direct method than the illumination-based radiometric variations for the BCM-based optical flow method. Not surprisingly, the BCM-based direct method performs worst due to both the complex motion and the radiometric effects.

5.4. Analysis

Experiments with the three data sets have led to interesting and valuable conclusions. First, the GDIM-based optical flow has consistently outperformed the other three methods in every case. In particular, these results show that the reset image is less significant in improving the PSNR of this method. This suggests that a longer time slot between consecutive reset frames may be allowed without compromising the reconstruction quality.

The GDIM-based direct method performed better than the BCM-based methods in the first two experiments, but not in the third. This was not surprising, as the current algorithm was developed for estimating only four motion components. Our own experience and a large number of other investigations have verified that pitch and roll motions are difficult to estimate when coupled with translational motion in the XY direction (e.g., [1]). As a result, we have concluded that, to develop a robust motion estimation algorithm, one ought to restrict the sought-after degrees of freedom to those attainable with good accuracy. Since other sensors including gyros can measure the pitch and roll of the vehicle with high accuracy, these can be integrated with the vision system to develop a more effective optics-based sensor [27].

The BCM-based optical flow method has performed more competitively than expected in terms of the PSNR. We attribute this to at least two factors. First, the low-detailed low-contrast nature of underwater imagery makes the BCM-based motion-compensated prediction error coding more immune to mismatch errors. Referring to Eq. (3), this means that $\widehat{\delta I}$ maintains the smooth nature of the radiometric transformation field δI, though the average magnitude may somewhat increase due to the contribution from the second term (see the error plots for each of the three experiments). Consequently, the prediction error of the BCM-based methods may still be coded efficiently to yield an acceptable PSNR. However, the reconstructed sequence often reveals deterioration in local regions which are hidden in the scalar PSNR measure, as verified in our experiments. Second, our experiments were performed on 128×128 images by smoothing the original data acquired at four times this resolution. The blurring effect has improved the error immunity further.

For these methods, the average effective compression ratio has been measured to be approximately 50 : 1, which may be improved by a factor of 2–3 with the implementation of a lossless entropy coding stage (e.g., Hoffman coding). Based on this assumption, we may realize a 100 : 1 to 150 : 1 compression rate with the addition of a lossless entropy coder, but no other modification. This can accommodate transmission at about $1/3$ of the full spatial resolution, which can be sufficient for most underwater applications.

6. SUMMARY

Despite growing demand for the deployment of unmanned underwater vehicles for a variety of scientific, commercial, and military operations, the tether link to supply live video, power, command, and control will continue to limit the utility and maneuverability of these vehicles. The inability to transmit the massive volume of live images by acoustic telemetry

is the primary obstacle to the removal of the tether. For live TV-quality monochrome video, a compression ratio in excess of 1000 : 1 needs to be achieved. In reducing the dependency on live video transmission for continuous operator support, these vehicles need to be made more intelligent and independent by implementing various signal processing algorithms that exploit the information from various smart and accurate sensors for automatic guidance and control. Even without a tether, two distinct operational modes in the applications of submersible vehicles, operator-assisted and automatic computer-controlled, are of special interest. In the former case, the guidance and control of the vehicle would be carried out by a human operator who relies on the live video transmitted through an acoustic channel. Thus, achievement of the highest quality reconstruction is a primary objective. In the latter scenario, we envision an intelligent vehicle that is equipped with a vision system for guidance and control during navigation, positioning, search missions, etc. The estimation of vehicle 3D motion is required to perform the primary tasks, and this information can be exploited simultaneously in video compression and transmission to a surface station for tracking the vehicle and its performance, to reconstruct the scene images for high-level human guidance, for online mission planning, etc.

We have investigated the motion-compensated compression of underwater imagery based on two dynamic image models, the BCM and GDIM. We have described four techniques based on the computation of optical flow or the direct estimation of vehicle 3D motion that are suitable for either of the two operational scenarios. A hybrid system has been simulated, consisting of two modules for motion-compensated prediction error coding and intraframe coding to reduce temporal and spatial redundancies, respectively. In each module, we have employed a hybrid of DWT and VQ stages, as described in [2]. Based on an implementation suitable for a 128×128 resolution, we have achieved average compression ratios of approximately 100 : 1 and 16 : 1 for the residual errors and reset images, respectively (recall that the decomposition levels depend on the image resolution). Overall, an average compression ratio of about 50 : 1 has been achieved, but may be improved 2–3 times with the addition of a lossless entropy coding stage. This would accommodate the transmission of monochrome video at about $1/3$ the standard resolution without any other modification to the proposed technique.

Various results have been presented from experiments with three selected image sequences, to evaluate the four methods in terms of the PSNR of the reconstructed images. The GDIM-based optical flow technique has provided superior performance with less sensitivity to the reset frames to maintain a steady reconstruction image quality (in terms of the PSNR). Although the BCM does not account for the illumination-induced radiometric variations, its performance has been competitive mainly due to the low-contrast characteristics of underwater images. The proposed GDIM-based direct method has proven to be effective in terms of the required computations and good PSNR performance. For accurate computation of vehicle motion based on this technique, integration of information from accurate pitch and roll sensors has been proposed [27]. This would enhance performance for automatic vision-guided navigation and positioning, in addition to achieving better PSNR rating in motion-based prediction coding for video transmission.

ACKNOWLEDGMENTS

This paper describes research work with funding support partly from the ONR under Grant N00014-95-1-0676 and the NSF under Grant BES-9711528.

REFERENCES

1. G. Adiv, Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field, in *Proc. IEEE CVPR, San Francisco, June 1995*.
2. M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, Image coding using wavelet transform, *IEEE Trans. Image Process.* **1**(2), 1992.
3. J. L. Baron, D. J. Fleet, and S. S. Beauchemin, Performance of optical flow techniques, *Int. J. Compet. Vision* **12**(1), 1994.
4. K. B. Benson and D. G. Fink, *HDTV—Advanced Television for the 1990s*, McGraw–Hill, New York, 1991.
5. J. R. Bergen, P. Anandan, Keith J. Hanna, and Rajesh Hingorani, Hierarchical model-based motion estimation, in *2nd ECCV, Santa Margherita Ligure, Italy, May 1992.*
6. V. Bhaskaran and K. Konstantibides, *Image and Video Compression Standards: Algorithms and Architectures*, Kluwer Academic, Boston, 1995.
7. N. Cornelius and T. Kanade, Adapting optical flow to measure object motion in reflectance and X-ray image sequences, in *ACM SIGGRAPH/SIGART Interdisciplinary Workshop on Motion: Representation and Perception, Toronto,* April 1983.
8. F. Dufaux and F. Moscheni, Motion estimation techniques for Digital TV: A review and a new contribution, *Proc. IEEE* **83**(6), 1995.
9. D. F. Haog and V. K. Ingle, Underwater image compression using the wavelet transform, in *Proceedings Oceans'94, Brest, France, September 1994*.
10. D. F. Hoag, V. K. Ingle, and R. J. Gaudette, Low-bit-rate coding of underwater video using wavelet-based compression algorithms, *IEEE J. Ocean. Engrg.* **22**(2), 1997.
11. B. K. P. Horn and B. G. Schunck, Determining optical flow, *Arti. Intell.* **17**, 1981.
12. ISO/IEC JTC 1/SC 29/WG 11, *ISO/IEC CD 13818-2: Information Technology: Generic of coding moving pictures and associated audio information, Part 2: Video*, ISO, Geneva 1993.
13. D. M. Kocak and F. M. Caimi, DSP hardware implementation of transform based compression algorithm for AUV telemetry, in *Proc. Oceans'98, Nice, France, September/October 1998*.
14. J. Koga, K. Iiunuma, A. Hirani, Y. Iijima, and T. Ishiguro, Motion compensated interframe coding for video conferencing, in *Proc. National Telecomm. Conference, 1981*.
15. Y. Linde, A. Buzo, and R. M. Gray, An algorithm for vector quantization, *IEEE Trans. Acoust. Speech Signal Process.* **37**, 1989.
16. H. C. Longuet-Higgins and K. Prazdny, The interpretation of a moving retinal image, *Proc. Roy. Soc. London Sec. B* **208**, 1980.
17. M. Mattavelli and A. Nicoulin, Motion estimation relaxing the constant brightness constraint, in *Proc. ICIP, 1994*.
18. N. Mukawa, Motion field estimation for shaded scenes, in *Proc. ICIP-89, Singapore, September 1989*.
19. S. Negahdaripour and B. K. P. Horn, Passive navigation, *IEEE Trans. Pattern Anal. Mach. Intell.* **9**(3), 1987.
20. S. Negahdaripour, A. Shokrollahi, and M. Gennert, Relaxing the brightness constancy assumption in computing optical flow, in *Proc. ICIP, Singapore, September 1989*.
21. S. Negahdaripour and C. Y. Yuh, A generalized brightness change model for computing optical flow, in *Proc. International Conference on Computer Vision, Berlin, May 1993.*
22. S. Negahdaripour and L. Jin, Direct recovery of motion and range from images of scenes with time-varying illumination, in *Proc. Int. Symp. Comp. Vision, Coral Gables, Florida, November 1995*.
23. S. Negahdaripour, X. Xun, and A. Khamene, Applications of direct 3D motion estimation for underwater machine vision systems, in *Proc. Oceans'98, Nice, September 1998*.
24. S. Negahdaripour, S. Zhang, and A. Khamene, On Shape and motion recovery from underwater imagery for 3D mapping and motion-based video compression, *Proc. Oceans'98, Nice, September 1998*.
25. S. Negahdaripour, Revised interpretation of optical flow: Integration of geometric and radiometric transformations for dynamic scene analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(9), 1998.
26. S. Negahdaripour, X. Xu, and L. Jin, Direct estimation of motion from sea floor images for automatic station-keeping of submersible platforms, *IEEE J. Ocean. Engrg* **24**(3), 1999.

27. S. Negahdaripour and A. Khamene, Robust closed-form solutions for estimating the motion of planar surfaces from images; case of reduced motion set, submitted for publication.
28. A. Netravali, Algorithms for the estimation of three-dimensional motion, *AT&T Tech. J.* **64**(2), 1995.
29. G. P. Stein, Geometric and Photometric Constraints: Motion and Structure from Three Views, Ph.D. thesis, Electrical and Computer Engineering Department, MIT, Cambridge, MA, June 1998.
30. A. Netravali and B. Haskell, *Digital Pictures: Representation, Compensation, and Standards*, Plenum New York, 1995.
31. M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice Hall International, Englewood Cliffs, NJ, 1995.
32. Y. F. Wang, N. Karandikar, and J. K. Aggarwal, Analysis of video image sequences using point and line correspondences, *Pattern Recog.* **24**, 1991.