3-D Time-Varying Scene Capture Technologies—A Survey

Elena Stoykova, A. Aydın Alatan, *Member, IEEE*, Philip Benzie, *Member, IEEE*, Nikos Grammalidis, Sotiris Malassiotis, Joern Ostermann, *Fellow, IEEE*, Sergej Piekh, Ventseslav Sainov, Christian Theobalt, Thangavel Thevar, and Xenophon Zabulis

(Invited Paper)

***Abstract*—Advances in image sensors and evolution of digital computation is a strong stimulus for development and implementation of sophisticated methods for capturing, processing and analysis of 3-D data from dynamic scenes. Research on perspective time-varying 3-D scene capture technologies is important for the upcoming 3DTV displays. Methods such as shape-from-texture, shape-from-shading, shape-from-focus, and shape-from-motion extraction can restore 3-D shape information from a single camera data. The existing techniques for 3-D extraction from single-camera video sequences are especially useful for conversion of the already available vast mono-view content to the 3DTV systems. Scene-oriented single-camera methods such as human face reconstruction and facial motion analysis, body modeling and body motion tracking, and motion recognition solve efficiently a variety of tasks. 3-D multicamera dynamic acquisition and reconstruction, their hardware specifics including calibration and synchronization and software demands form another area of intensive research. Different classes of multiview stereo algorithms such as those based on cost function computing and optimization, fusing of multiple views, and feature-point reconstruction are possible candidates for dynamic 3-D reconstruction. High-resolution digital holography and pattern projection techniques such as coded light or fringe projection for real-time extraction of 3-D object positions and color information could manifest themselves as an alternative to traditional camera-based methods. Apart from all of these approaches, there also are some active imaging devices capable of 3-D extraction such as the 3-D time-of-flight camera, which provides 3-D image data of its environment by means of a modulated infrared light source.**

***Index Terms*—Coded light, digital holography, pattern projection, shape, three-dimensional (3-D) displays, 3-D/stereo scene analysis, 3DTV, time-of-flight, video analysis.**

Manuscript received March 12, 2007; revised June 2, 2007. This work was supported by the European Commission within FP6 under Grant 511568.

E. Stoykova and V. Sainov are with the Central Laboratory of Optical Storage and Processing of Information, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria (e-mail: elena@optics.bas.bg).

A. A. Alatan is with Department of Electrical Engineering, Middle Eastern Technical University, 06531 Ankara, Turkey (e-mail: alatan@eee.metu.edu.tr).

P. Benzie and T. Thevar are with the School of Engineering, University of Aberdeen, Old Aberdeen AB24 3UE, U.K..

N. Grammalidis, S. Malassiotis, and X. Zabulis are with the Informatics and Telematics Institute, Centre for Research and Technology Hellas, GR-57001 Thermi-Thessaloniki, Greece.

C. Theobalt is with the Max-Planck-Institute of Informatics, 66123 Saarbruecken, Germany.

J. Ostermann and S. Piekh are with the Gottfried Wilhelm Leibniz Universität Hannover, D–30060 Hannover, Germany.

Color versions of one or more figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2007.909975

I. Introduction

THREE-DIMENSIONAL television (3DTV) is expected to evoke a revolution in visual technology. Beyond any doubt, precise acquisition of 3-D information of dynamic scenes is crucial for 3DTV implementation. Extensive research has been conducted for capturing, processing, and analysis of 3-D scene data aimed at reliable 3-D reconstructions for more than two decades. A variety of techniques have emerged as a result of the recent advances in image sensor technologies and evolution of digital computation.

The aim of the survey is to shed light on state-of-the-art in the major time-varying 3-D scene capture technologies with an emphasis on their incorporation in functioning 3DTV systems. Having their own pedigree, underlying principles, limitations, and trends for development, the examined 3-D scene-capturing techniques co-exist within the scope of the survey, as separate topics. They are united, however, not only by the same goal they pursue, but also by the use of common mathematical tools to tackle the 3-D reconstruction problem, confronting similar difficulties in many cases. More specifically, the survey proposes concise analysis of approaches for 3-D scene extraction from single and multiple cameras data streams. It also focuses on holographic and pattern projection techniques for real-time collection of point locations and color information of 3-D objects, as an alternative to traditional camera-based methods, and briefly describes potentials of time-of-flight-based systems capable of integrating depth extraction and imaging in real time.

Because it is an ill-posed problem, the challenges for 3-D scene reconstruction from mono-view video sequences are pointed out in Section II. The first part of the section introduces all shape-from-X approaches for 3-D shape extraction but concentrates mainly on shape-from-motion approaches, relying on the relative motion between the camera and the scene, as being quite appropriate for many practical situations. The second part is dedicated to the techniques for capturing of 3-D structures of human face and body, that have many potential applications, still from mono-view sequences. Brief description of methods for face capture and facial motion analysis, as well as for body modeling and body motion tracking and recognition, is also presented. Section III is devoted to hardware issues in multi-camera setups, such as calibration and synchronization, and it addresses the implementation of different classes of multiview reconstruction algorithms for dynamic 3-D reconstruction.

Section IV briefly presents the principle of digital holography, as a means for digital recording of a hologram of an object with further numerical reconstruction of its 3-D image. Pattern projection techniques, in which the information of the object shape and color is encoded in a 2-D pattern that is projected onto and reflected from the object, are discussed in Section V for the cases of structured light and fringe projection. Finally, the principle and some basic issues of time-of-flight range imaging systems are outlined in Section VI.

II. 3-D Single-Camera Techniques

A. 3-D Scene Extraction From a Single Camera

3-D scene extraction from a single-camera video sequence is a well-known problem in computer vision for decades. However, due to its ill-posed nature, other approaches, such as multiview or stereoscopic vision, have been preferred against single-camera techniques. However, it is obvious that conversion of the vastly available mono-view video for the upcoming 3DTV systems is strictly necessary.

1) Shape-From-X for a Single Camera: There are mainly four methods that imitate human 3-D perception to extract the 3-D shape information. These are shape-from-shading [1], shape-from-texture [2], shape-from-defocus/focus [3], and shape-from-motion. In shape-from-shading methods, an energy function is minimized by applying several constraints to overcome the unclear characteristic of the problem for determining the surface gradients from the single-image irradiance equation. For shape-from-texture, the validity of information about the depth in texture mainly depends on two properties, which are homogeneity and isotropy. If a texture is homogenous, both size and density can be used for extracting the shape information, whereas if the texture is isotropic, spatial compression of the texture still provides better shape information. However, texture information is not guaranteed in general scenes. In the shape-from-focus methods, the camera focus varies on an object to determine its depth, which is not available for many cases, as well. The shape-from-motion approach tries to solve for 3-D geometry by using the relative motion between the camera and object, which is an expected situation in practice. It should be emphasized that there is no complete solution to the single-camera 3-D scene extraction problem. All shape-from-X methods have their own advantages and disadvantages. Considering the requirement of applicability of the solution to general scenarios, it can be concluded that shape-from-motion (SfM) methods are more preferable. Hence, the remainder of this section focuses on different solutions to the SfM problem. In all applications that deal with the extraction of the 3-D scene structure, feature matching and tracking plays an important role. Feature correspondences between two (or more) frames of a video sequence are strictly required to geometrically relate the images, for camera calibration or estimating scene structure. A good overview can be found in [4] and [5].

2) Manual and Self-Calibration: In the conversion process of the available content to 3DTV input, such data are expected to lack calibration information. Camera calibration is the process of obtaining camera intrinsic parameters in order to solve for 3-D structural information [6], [7]. Structure and motion problems require a high level of accuracy of the camera matrix (intrinsic parameters) due to the nonlinearity of the problem of the scene reconstruction. Although many algorithms have been proposed for manual calibration, two of them, by Tsai [8] and Zhang [9], received wide acceptance in the computer vision community. These methods are based on utilizing calibrating patterns in order to determine the unknown camera parameters. However, in some cases, it is not possible to utilize such a pattern; hence, the calibration should be performed by only using the available frames, called self-calibration. For self-calibration, it should be noted that the fundamental matrix, which summarizes the geometric relation between views in an algebraic relation, contains both camera-intrinsic parameters and relative orientation between two frames. Therefore, a formulation, which does not vary with the relative motion between these two images, could be defined.

The first approach for self-calibration is proposed by Maybank and Faugeras [10]. In their method, the nonlinear quadratic equations, the so-called Kruppa equations, are constructed by using the fundamental matrices and unknown camera matrices. These equations are solved in different ways [10]–[15]. On the other hand, some methods do not attempt to solve Kruppa equations. They generally determine the camera-intrinsic parameters and the position of the plane of a virtual conic by using the relation between the virtual conic and the camera-intrinsic parameters [16], [17]. These methods later update the projective reconstruction to metric reconstruction. In the method by Pollefeys [18], projective reconstruction is updated to affine reconstruction by using the position of the plane of the virtual conic determined by solving a number of constraints [19]. Then, the affine calibration is updated to a metric one using the estimated camera intrinsic parameters determined by solving the general camera self-calibration equations. This method is called stratified calibration, since one moves between different stratums (i.e., affine to metric) during reconstruction.

3) 3-D Structure Estimation From Two Views:

a) Solving for epipolar geometry for two views: When two views of a camera locate at an arbitrary position, the geometrical relationship between these views is given by the epipolar geometry, where this geometrical relation can be expressed with the fundamental matrix, which is the algebraic representation of the epipolar geometry. Significant research effort has been put into estimating the fundamental matrix from a set of point correspondences so far [6], [20]–[24]. The approaches can be classified into two categories: linear and nonlinear methods. Linear methods formulate the algebraic relation between pixel locations and camera orientation and position in a linear form, with the well-known example of the normalized eight-point algorithm [20], whereas the nonlinear methods strictly impose the constraints of the fundamental matrix. As a criterion for the goodness of the estimation, a cost function can be used to minimize the geometric distance, e.g., minimization of the reprojection error (Gold Standard method) or the Sampson cost function, which minimizes the distance of the points to their corresponding epipolar lines. Minimization of the error is carried out using the Levenberg–Marquardt algorithm, which is a slight variation of the Gauss–Newton iteration method. In the presence of outliers, i.e., badly localized matches or even false matches, the computation of the fundamental matrix is more expensive. Outliers, caused by noise and pixel quantization, decrease the precision of this matrix. Many robust algorithms to reduce the effects of outliers or even to remove the outliers were applied [25], [21]–[24]. Typical examples are the M-estimator, the random sample consensus (RANSAC) [22], MLESAC [23], or NAPSAC [24] algorithms.

b) Solving for rotation and translation: The relationship between the fundamental matrix and the essential matrix can be given as $E = K^T F K$, where K is the camera calibration matrix. It is shown that the decomposition of the essential matrix into rotation R and translation t could be obtained as a cross product $E = t \times R$. A linear algorithm is proposed by using vector algebra properties that uses this relation without imposing the orthonormality of the R matrix to the solution of the rotation [26]. Later, a robust version is also proposed, which considers the orthonormality of the R matrix and utilizes quaternion for its formulation.

c) Estimation of depth at sparse correspondences: One of the most important steps in structure computation is triangulation, in which the position of a point in 3-D is estimated from point correspondences. For the ideal conditions, the back-projected rays should intersect at a sole point in 3-D space in general, and therefore a simple line intersection should suffice. However, due to inevitable noise and errors, it is necessary to employ noise-resistant techniques to estimate the position of the point in 3-D space. In the literature, there are four major methods for triangulation: midpoint method [27], [28], polynomial triangulation method [6], linear methods [29], and iterative linear methods [30]. While a commonly suggested method for triangulation is the selection of the midpoint of the common perpendicular to the back-projected rays of the matched points, this method is used only for Euclidean reconstruction problems, since it is not projective-invariant [27]. In the method of polynomial triangulation, the problem is reduced to finding the roots of a sixth-degree polynomial in one variable by parameterizing the pencil of epipolar lines, and the method is optimal under the assumption of a Gaussian noise model [6]. The linear triangulation method is the most common method used due to the ease of implementation, and it determines the location of a 3-D point from projection matrices and image points by back-projection [29]. In the iterative algorithms, these methods try to find the solution by changing the weights adaptively so that the adapted weight matrix will give a measure of a geometric error function. For a projective reconstruction in which camera matrices are known to a great accuracy, it is recommended to use the polynomial triangulation method which is invariant under projective and affine transformations, computationally cheap, and optimal under Gaussian noise assumption.

d) Determining the dense depth field: Dense-depth-field estimation is one of the most active research areas in computer vision, and there are various depth-estimation methods by different approaches [31]–[36]. It should be noted that disparity estimation is closely related to depth estimation, since dense depth field could be reconstructed by using triangulation once a disparity field is determined. The major dense depth-estimation methods are PDE-based [32], [33], Markov random field-based (MRF) [37], [38], and neighborhood-based [34], [35] methods as well as the maximum flow formulation [36] approach. The aim of PDE-based energy-minimization methods is to estimate the dense disparity map by using a minimization and regularization approach. There are various operators that are used for regularization [39]–[41]. The operator should allow the anisotropic diffusion of the disparity field in order to allow discontinuities across boundaries. In this way, while obtaining smooth depth fields along object boundaries, discontinuous depth fields across object boundaries could also be achieved. The MRF-based methods [37], [38], [42] are probabilistic approaches for dense depth estimation, and they model the unknown depth values, as a random field, whose probability density function (pdf) is assumed to be Gibbsian. After defining the pdf of the depth field by an MRF formulation, a maximum *a posteriori* (MAP) estimator is used for the dense depth-field estimation. According to this formulation, a cost function is derived, and the minimum of this cost function is equal to the MAP estimate of the random depth field [26]. The minimization can be achieved, either in stochastic and deterministic ways, such as simulated annealing and ICM. Neighborhood-based methods [34], [35] try to find the dense disparity map of a stereo image pair by searching extra match points around a starting set of correspondences. Consideration of neighborhood constraints together with a dissimilarity function increases the robustness of the disparity estimation [44]. Finally, a maximum flow formulation of the N-camera stereo correspondence problem can be obtained [36]. Once solved, the minimum cut associated with the maximum flow gives a disparity surface for the whole image.

4) 3-D Structure Estimation From Video: Multiple-frame structure from motion (MFSfM) problem usually involves more than three frames and often a monocular single-camera system which is swept around the scene. The problem has unique properties, which are a mixed bag of blessings and curses. First, the notable characteristic of the problem is the existence of causality and temporal continuity constraints. Hence, the relative position between the camera and scenery changes incrementally, yielding some redundancy. This advantage of the temporal redundancy brings the problem of a narrow baseline, which implies a small disparity between frames, hence little SfM information between the consecutive frames. Temporal redundancy is of limited use, if successful feature tracking and inter-frame correspondence are not achieved. The solution methods that dominate the literature can be categorized on many axes, such as type of features, camera and imaging models, feature geometry and camera motion constraints, and whether the algorithm is batch or online. However, linear versus nonlinear categorization [45] provides the most appropriate division boundary, as it has a scope that is sufficiently broad to capture the overall picture.

Linear algorithms provide direct solution methods. However, this significant advantage is tainted by the fact that the SfM problem is actually nonlinear, so these techniques depend on the linearization of the original problem. In a multiframe context, the dominant algorithms are variants of the popular factorization method [46]. While they come in many different tastes and colors, the basic idea is the same: image measurements can be expressed as a product of two matrices, representing the motion and the structure [47]. On the other hand, nonlinear algorithms rely on the iterative optimization techniques. The minimization of a cost function is performed on both structure and motion parameters. As expected, these algorithms usually have convergence issues and are vulnerable to local minima. However, they provide good numerical accuracy and flexibility. Some of these methods involve the use of recursive techniques, either as an estimate fusing mechanism [48]–[50] or as a state estimator [51], [52]. Others, such as [53], rely on classical optimization techniques, known as bundle adjustment algorithms. The original formulation of factorization is based on the orthographic projection, since it is only possible to linearly decompose structure and motion via this projection [46], [47]. Later, this formulation is also extended to weak perspective [54], paraperspective [55], generalized affine [56] and, finally, a perspective camera [57]–[59]. A sequential version is also proposed for this batch algorithm [60]. Instead of point correspondences, factorization is also formulated for line features as well [61]. In order to cope with occlusions and unreliable features, alternative cost functions are proposed in a technique, known as hallucination, which completes the missing entries [46]. It is also possible to estimate the structure and motion of multiple objects [47].

Nonlinear methods can be divided into two categories: bundle adjustment and recursive methods. While these approaches employ different techniques to obtain a solution, they share common features, such as iterative minimization of a cost function, models, and concerns for trapping into local minima. Bundle adjustment is a catchall name for many techniques to achieve jointly optimal estimates of the 3-D structure, motion, and camera calibration parameters [6], [62]. Optimality implies the minimization of a cost function, which quantifies 2-D reprojection error. As in most iterative optimization techniques, a good initialization is essential.

Fusion methods are indeed a hybrid form between batch and online approaches. The basic idea is constructing a structure estimate by fusing intermediate reconstructions (or subestimates) obtained by processing smaller subsets of the sequence [63]. Moreover, the occlusion problem is handled more easily as a feature should be viewed in only a few frames and not in the entire sequence [49].

Another approach for solving the SfM problem is to consider it in the context of state estimation for dynamical systems [64]. The Kalman filter, or rather the extended Kalman filter (EKF), is a well-known tool for recursive state estimation. In order to design an EKF, the states, a dynamic system model, and an observation model should be determined. In the basic EKF model [65], the states are selected as 3-D coordinates, the system model includes rotational and translational motion between time instants, and the observation model is based on pixel locations of the 3-D points (states), observed with some Gaussian noise. For this approach, since the state vector quickly grows quite large, as every new feature extends this vector by three entries, the state vector should be reduced by one of these two strategies, as explicit and implicit reduction [64].

In the explicit reduction, the measurement equation is solved for some of the states and substituted into the model equations [52]. On the other hand, implicit reduction strategy is employed by fixing some states, or a function of them. Such an approach effectively introduces new constraints and reduces the solution space. While there are many models proposed for system and observations, some seem to have received wider acceptance. In [51], the observation model is a perspective camera with known calibration, the translation model is an nth-order Taylor approximation of the function corresponding to the object centroid motion (equivalently, camera center motion), and the rotation model is mth-order Taylor approximation with quaternion representation. This algorithm is later used as an initialization step in a similar but recursive estimator in [65] and extended with inertial sensor measurements in [66]. Photometric models have also found a niche in [67].

B. Discussion on Single-Camera Techniques

There are many methods for capturing 3-D scenes from a single-camera video sequence. Alhough, each of these methods has its own advantages and disadvantages; currently, the SfM is one step ahead of other techniques. The main reason is that SfM can be used to solve real life problems. On the contrary, the other techniques currently are just used to extract 3-D shapes in controlled 3-D environs. It should also be mentioned that SfM cannot solve all single-camera 3-D scene capture problems, whereas it might solve a fairly good amount of them, compared with other approaches. Even for those unsolved cases, there exist plenty of approximations to yield acceptable 3-D perception in 3DTV systems. Hence, SfM can be assumed to be the most complete solution among all of the single-camera 3-D scene extraction methods, and a good candidate to be utilized in the conversion of the available mono-view content to the upcoming 3DTV systems.

C. Human Face and Body-Specific Techniques

Many 3-D applications are limited to studio environments where the human face or the full body is the prime object. Therefore, capturing the 3-D structure of a human face and body is a very important research area. With *a priori* knowledge about 3-D structure and motion of human faces and bodies, natural limits of human motions can be used in order to make the processing more efficient.

The human-specific techniques can be divided into two areas. One of them is concentrated on the human face and contains the following subtasks:

- face and facial feature detection;
- capturing of 3-D structure of the face;
- analysis of global face motion and mimic.

The other area is associated with the human body:

- 3-D human body modeling;
- human body kinematics and motion analysis;
- human body motion recognition.

Human face and body-specific techniques are not only important because of human presence in the scene, but become more popular in recent days because of progress in 3-D visualization technology. In the area of 3-D visualization and 3-D display systems, robust detection and tracking of the observer's eyes and the observer's view point is necessary to render the correct view according to the observer position.

1) Face and Facial Feature Detection: Face and facial feature detection is the first step in computer vision problems.

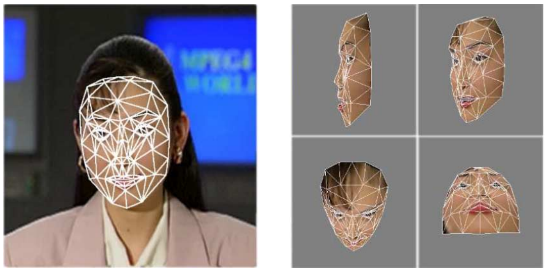

Fig. 1. Capturing of the 3-D face structure from a single camera view.

The face has to be detected from an image or video frame first before performing any further processing. The images on which face analysis has to be performed have a myriad of variations in pose, visible objects (e.g., glasses, moustache, or beard), complex background, different numbers of faces, different facial expressions, occlusions, lighting condition, and camera characteristics.

All of these factors cause some changes in the appearance, and an ideal analysis/capturing system should be invariant to all of them. The following strategies are known and often jointly used [43].

- *Knowledge-based methods* which rely on rules derived from elements of human face structure [68], [69].
- *Feature-invariant methods* which track down invariant features for face detection [70]–[84]
- *Template matching methods* which utilize correlation between the detected face image and face patterns from a set of standard faces and facial features [85]–[90].
- *Appearance or image-based methods* which rely on learned models acquired from collection of training images [91]–[103].

2) Capturing of 3-D Structure of the Face: When modeling a person's face, it is generally represented by a 3-D face model (3-D triangular mesh is commonly used) and an associated composite image (texture image). A common approach to capture the 3-D structure of the face uses 2-D images, which are taken from different directions and contain different views of the face. The standard 3-D face model is used to fit a person's facial structure. Here, the positions of facial features like eyes, mouth, and eyebrows are used to form a 3-D face model (see Fig. 1).

The methods proposed in the literature for generating a 3-D face model can be generally classified as fully manual process, a semi-automatic process, or a fully automatic process. In a fully manual process, the topology of the 3-D mesh has to be manually mapped onto the image of the face. This process is very time consuming. In a semi-automatic process [104]–[110], the certain facial features and feature points are detected or manually selected on the image, and then a standard 3-D face is automatically adapted by a transformation using the location of the detected feature points [111]–[116]. Sometimes a manual fine tuning is also used. In the fully automatic process, a user interaction is not needed, and the time required to model a face is drastically reduced. However, the software-based systems are very sensitive to the image data.

3) Face Motion Analysis: Motion estimation of a human face is important for many applications in computer vision like facial expression analysis, face identification, model-based video coding, and 3-D facial animation systems. Motion estimation or tracking systems can be divided into feature-points-based [117]–[120] or model-based methods [121]–[132]. A feature-points-based approach estimates the displacements of feature points from one video frame to another. The displacements might be estimated using optical flow methods or block-based motion estimation methods. The estimated motion field is used to compute the motion of the object. A model-based tracker, on the other hand, uses a 3-D object model with stored or generated texture and tries to fit the model to the new video frame by minimizing costs. The accuracy of motion estimation is thus dependent on the used object model, on the quality of the initialization process, and on the accuracy of the camera calibration.

4) Body Modeling: There are various methods for acquisition of 3-D human body models. Some commercial systems require special hardware, but they are expensive and cannot be used in certain cases. Recently, using video frames rather than using special hardware is preferred. A number of techniques using video frames has been proposed [133]–[141]. First of all, different views of the subject are obtained from different calibrated cameras or one moving camera. From each view, the 2-D silhouette of the subject is extracted. Then, using volume intersection, different views of the subject are intersected, and a volumetric description of the subject is defined. Finally, a model of the human body is fitted to the volumetric description of the subject. Sticks, ellipses, cylinders, or super-quadrics can be used for the predefined model. In some methods [137], [134], [138], the subject needs to perform some initial movements in order to obtain the model more accurately.

5) Body Motion Tracking: By tracking and recognition of human motion, the simplifying assumptions about the scene and the motion are usually made. The initial position and posture of the person are assumed to be known. Prior to human body motion estimation, the segmentation of the human silhouette from the background should be made. Then, the feature extraction and tracking follows. The prediction of movement is also used in order to solve the occlusion problem. Some tracking techniques try to determine the precise movements of each body part [142], while other methods focus on tracking the human body as a whole [143], [144]. Tracking techniques may also be classified as 2-D and 3D. Using a 2-D approach, the motion in the image plane is analyzed either by exploration of low-level image features or by using a 2-D human body model. 3-D tracking tries to obtain the parameters which describe body motion in three dimensions. 3-D tracking allows 3-D pose recovery, estimation the position of the body parts in 3-D space, and orientation of the body relative to the camera. The 3-D pose parameters are commonly estimated by iteratively matching a set of image features extracted from the current frame with the projection of the model on the image plane. The overview of existing human-motion analysis techniques can be found in [145]–[148].

6) Body Motion Recognition: Human motion recognition may be achieved by analyzing the extracted 3-D pose parameters. Instead of obtaining the exact position of a human body, human motion recognition tries to identify the action performed by a moving person [149]. Most of the known techniques focus on identifying actions belonging to the same category (e.g., specific sport movements, sitting down, standing up, walking, or running) [149]–[151]. Some of the techniques recognize and identify several persons and their interactions [152]–[154]. Some of them are developed to work in special environments and try to use prior knowledge about the layout of the room [155].

III. Multicamera Techniques

A. Acquisition Systems

Some of the most significant multicamera systems that relate to the goals of 3DTV and 3-D reconstruction are related to telepresence and teleconferencing, since these applications intrinsically require live video feeds. Such examples are CMU's 3-D room [156], the view-dependent visual hull system at MIT [157], the multicamera systems at the Keck Kaboratory at the University of Maryland [158], and the Argus System at Duke University [159]. Last but not least, the Stanford University multicamera array [160] is an architecture specialized for facilitating the lightfield rendering approach (see Section III-B).

The teleconferencing system in [161] captures a scene with four cameras mounted around a display. The system in [162] has been pioneering as the first to utilize a large number of video streams to provide a real-time multiview reconstruction. In [163], synthetic views are produced using five streams based on the visual hull method [157]. The system in [164], [165] was initially based on multibaseline dense-depth map computation. Its more recent version [156] is based on visual hull computation using silhouette carving and has been commercialized [166].

Issues that relate to multicamera systems are the calibration and synchronization of the cameras. Typically, multicamera calibration is based on solving the correspondence problem for multiple cameras to estimate their parameters. The calibration toolbox in [167] is designed for single-camera calibration, but can facilitate the calibration of camera clusters if their fields of view overlap. In contrast, the toolboxes in [168] and [169] are dedicated to camera cluster, and utilize an LED as a calibration target to perform their calibration with minimal user interaction.

Multicamera synchronization approaches fall under two major categories: hardware- and software-based synchronization. Hardware-based approaches consist of a control unit dedicated to propagating external signals for triggering the cameras. Point Grey manufactured the Sync Unit [170], which synchronizes multiple cameras on different IEEE-1394 buses. The Objective Imaging OASIS-DC1 [171] is a controller, providing synchronization for a camera that supports trigger signals. In [172], a microcomputer control for the synchronized operation of several high-speed cameras is utilized. Other similar approaches for synchronization at the hardware level are proposed in [173]. Networking the computers that host the cameras facilitates their software-based synchronization [174] by means of computer-clock synchronization protocols [175]. In [176], the proposed system consists of the camera-computers and the triggering-computer. Since there is no proposal to handle the situation in case of failure of the synchronization, the method is mainly dependent on the quality of the Ethernet connection. In [177], the server–client architecture with a simple error-checking technique is utilized. The network latency is estimated and accounted for when sending a trigger signal to the cameras.

B. Multiview Scene Reconstruction

Multicamera systems capture the appearance of a dynamic scene from multiple viewpoints at the same time. A variety of techniques have been proposed in the literature to extract from this footage the so-called free-viewpoint videos, or otherwise, dynamic scene representations that facilitate the rendering of the captured events from arbitrary novel viewpoints. In this survey, the presented taxonomy of multiview stereo reconstruction algorithms is partially based on the review in [178], which classifies them in the following classes: 1) computing a cost function on the 3-D volume; 2) iterative evolvement of a surface based on cost-function optimization; 3) combination of individual views; and 4) feature point reconstruction, usually followed by surface fitting. In addition, the cases of sillhouette-based multiview reconstruction as well as lightfield rendering are considered, although that the last is not actually a reconstruction technique *per se*. The reason that is mentioned is that it can produce results that are useful for the goals of 3DTV.

The usual similarity measure of multiview correspondences is photoconsistency [179], but it requires radiometric/color calibration, which is difficult to achieve and retain. Some multiview methods, which compute individual views and then merge them, utilize correlation-based metrics, which are more robust due to their invariance to global scalings of intensity. More recently, stereo approaches that estimate the surface normal of the imaged surfaces have appeared [180]–[183]. This estimation proves to be enhancing to the accuracy and quantity of established stereo-correspondences, because the pertinent methods compensate for the projective distortion in the images and are thus able to provide better matches.

The first category includes the space-carving [179] and voxel coloring [184] approaches, which provide reconstructions based on the computation of visibility, in terms of determining if a voxel is occupied or not. In these approaches, the optimized cost is local and determined by computation of visibility; otherwise, the notion that the space between a camera and a point of the imaged surface is void.

Voxels are traversed in a fixed visibility ordering, taking full account of occlusions and allowing the input cameras to be far apart and widely distributed. Thus, a special set of invariant voxels are identified, forming a spatial and photometric reconstruction of the scene, which is fully consistent with the input images.

The *orientation-consistency cue*, i.e., the fact that two surface points with the same surface normal and material exhibit the same radiance under orthographic projection and distant lighting, has also been adapted within the voxel coloring framework [185]. The ability of this approach to solve for normals within the voxel framework yields a dramatic improvement in the representation of fine-scale surface detail. In [186], a probability is assigned to each voxel, indicating the likelihood for the voxel to exist. Maximum flow algorithms [187]–[190] and multiway graph cuts [191] have also been applied to extract the optimal surface from a volumetric MRF.

A category of algorithms computes the reconstruction as the result of deforming a surface by the optimization of a cost function. In [192], a depth map is computed individually for each stereo-view and, then, a surface is iteratively deformed. For each instance of this surface, a cost function is computed which is ultimately optimized by the method. The surface is initialized at the visual hull of the object by extracting its silhouette in the acquired images. The method is significant in that it combines the silhouette cue with stereo; however, it is limited in single object rather than scene reconstruction. As in the above method, in [188], the deformed surface is optimized based on the silhouette extraction and, thus, exhibits the same disadvantages. The method differs in that the evaluated cost is correlation-based and that the surface is computed by a volumetric cut. In [191], individual depth maps are merged based on graph cuts so that a global cost is optimized. In [193], the deformation of a surface is based on level-sets. In the approach, a set of partial differential equations are volumetrically defined and evaluated, thereby deforming the optimized surface. The approach differs from other cost-optimization approaches in that the cost is locally, rather than globally, computed. Similar, in terms of locality, are also the methods in [194] and [195], where the optimized surface is deformed by simulated forces until it reaches an equilibrium.

Another category of methods computes individual reconstructions from each view and then fuses them in a single result. Typically, the main problem that is encountered is due to calibration errors that cause the "misregistration" of views [162] or the event that a surface point that is visible in more than one views is reconstructed at different coordinates for each view. To cope with the problem, consistency between depth maps is enforced in [196], [191], and [197]. In [198] and [199], the individual results are merged after the individual reconstructions are computed. Finally, in [183], duplicate reconstructions of the same surface point are suppressed along the surface normal.

Algorithms that are based on feature-point reconstruction first extract and match a set of feature points in the available images. In [200], a constrained Delaunay triangulation is used to interpolate 3-D data obtained from stereo, and then tetrahedra that are empty are marked. The boundary of the free space consists of the polyhedral representation of the object. A recursive algorithm for the reconstruction of surface models from sparse 3-D measurements captured from multiple camera views which are consistent with the data visibility is presented in [201]. The algorithm is shown to converge to the real scene structure as the number of views increases and to have a computational cost which is linear in the number of views. In [202], a triangulation of available 3-D points is selected based on its consistency with this set of images of the object. The algorithm starts with an initial rough triangulation and then refines the triangulation until it obtains a surface that best accounts for the images of the object. Surface models of complex scenes are recovered from sparse data in [203] using a principled mechanism for reasoning about the structure of the scene based on quasi-sparse correspondences in multiple images. This approach is able to handle scenes that involve multiple disconnected surfaces which may occlude each other.

In a different category of approaches, dynamic scene geometry is reconstructed by intersecting cones of rays that are formed by reprojecting image silhouettes into 3-D space. This way, a conservative volume estimate of an object in the scene's foreground, the so-called visual hull [204], is obtained which is typically represented as a 3-D triangle mesh [205], a voxel model [206], or which is view-dependently rendered based on epipolarity constraints [207].

The authors of [208] propose a point-based approach to 3-D video that also capitalizes on the visual hull principle. Although simple silhouette intersection can be performed in real time, the resulting geometry is very coarse, and concave surface areas cannot be faithfully reconstructed. By combining silhouette constraints with additional clues, such as photo-consistency [209] or stereo [210] criteria, better results can be achieved. The main advantage of silhouette-based methods is their computational efficiency. However, to obtain visually pleasing models, additional processing is required that prevents real-time reconstruction.

An alternative solution strategy aims at generating synthetic views of scenes by pure image-based recombination of input video frames, for instance, by means of lightfield rendering [211], [212]. Although this allows for intermediate novel viewpoint synthesis of arbitrary scenes, the scene must be sampled with dense camera arrays, and the memory consumption is significant. The required sampling density can be reduced if at least a basic model of the scene's geometry is reconstructed from the footage. Unstructured lumigraph rendering uses a simple geometry proxy in combination with clever camera position-dependent image-blending to create novel views from a sparse and irregular set of input video streams [213]. The system presented in [214] applies a similar lightfield approach to create real-time 3DTV. To this end, the video streams from several closely spaced cameras are transferred to an array of projectors which renders the view-dependent dynamic scene appearance on a specially coated screen. Although the achieved visual quality is very high, the range of allowed virtual view points is very limited.

Recently, a system [215] was proposed that is comprised of a dome of quickly triggered light sources, three high-speed video cameras, and a rotating platform to capture the view- and lighting-dependent appearance of periodic human locomotion.

By having the person perform periodic motion on a treadmill that slowly rotates relative to the cameras, the scene can be captured from a large set of viewpoints. New virtual views of the moving person are rendered by combining nearby captured viewpoints using lightfield interpolation. Although lightfield approaches enable 3-D video renderings of appealing quality, their immense storage requirements and the high number of required input cameras do not make them an ideal choice if large virtual viewpoint changes should be allowed. One way to overcome these limitations is to reconstruct a more detailed geometry model, e.g., time-varying dynamic depth maps, from the input streams. The first approach capitalizing on this idea was the Virtualized Reality system [216], which uses a dome of cameras and a multiview wide-baseline stereo method to reconstruct time-varying scene geometry and texture. The pixel colors at the intermediate rendered views are determined by warping the

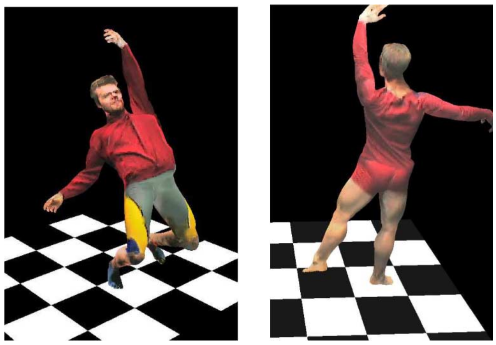

Fig. 2. Model-based free-viewpoint videos rendered in real time.

closest input video frames into the target view based on the reconstructed per-pixel depth information [217]. The main difficulty that stereo-based 3-D video methods are facing is the notoriously hard multiview image correspondence problem that needs to be properly solved in order to get decent scene geometry. The method proposed in [218] therefore suggests a clever boundary matting and depth-map regularization to reduce the artifacts in reconstructed maps. Another possibility to improve 3-D video quality is to combine the multicamera system with a multiprojector system. In [219], the authors enhance space-time stereo reconstruction by additionally projecting multiview dynamic noise patterns into the scene. The authors of [220] propose to use a template human body model, a marker-free optical motion capture approach, and multiview dynamic texture generation to create free-viewpoint videos of human actors that can be rendered in real time (see Fig. 2). Their original approach has been extended such that even time-varying surface reflectance can be reconstructed, thereby enabling the display of virtual humans under arbitrary synthetic lighting conditions [221].

In [178], a quantitative evaluation of six multiview stereo reconstruction algorithms can be found. The algorithms are compared in terms of accuracy and completeness of the representation and information about the computational efficiency is also provided. This review suggests that the choice of an appropriate technique mainly depends on the application, the type of scene, as well as the range of virtual viewpoints that one has in mind. It is also very likely that, for scenes with general foreground and general background, a clever combination of different approaches would probably be most promising.

IV. Holographic Techniques

Holography is a unique technique for recording and reconstructing 3-D information of an object. A hologram is essentially a record of the interference pattern obtained from the superposition of a reference beam and the beam scattered by the object. In classical holography, photographic films are used to record holographic patterns, and the reconstruction is performed optically. However, recent advances in computer and video capture technology have permitted replacing holographic films with charged-coupled devices (CCD) and complementary metal–oxide–semiconductor (CMOS) image sensors to record and numerically reconstruct holograms; this technique is now known as digital holography [222]. In comparison with classical holography, digital holography has the major advantage

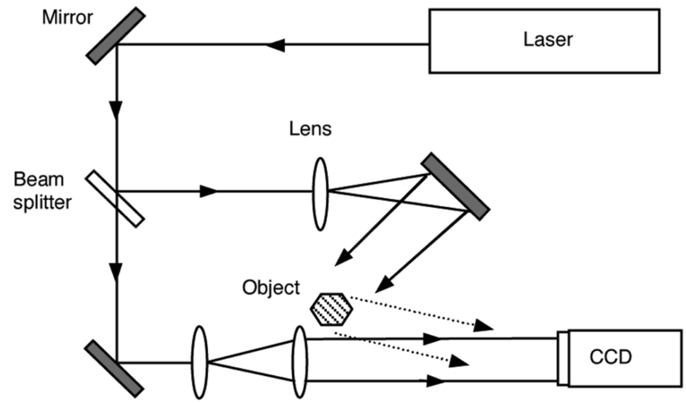

Fig. 3. Apparatus for recording digital holograms.

that it eliminates the need for wet chemical processing and other time-consuming procedures and, thus, recording and numerical reconstruction can be done in almost real time. In addition, numerical data allow for manipulation, replication, and ease of transmission. Therefore, digital holography is seen as the way forward in realizing practical 3-D time-varying scene capture for engineering applications and mass-media 3-D displays. However, it needs to address some technological issues such as image resolution, data storage and retrieval speed, display of real and virtual images, and true color recording and reconstruction before it can gain mainstream acceptance.

The basic process of recording a digital hologram is shown in Fig. 3. The light coming from the object interferes with the reference beam, and this interference pattern is recorded by the CCD camera. To successfully capture a hologram, the angle θ between the reference and the objects waves must not exceed a maximum value given by

$$\theta_{\max} = \sin^{-1}\left(\frac{\lambda}{2x_p}\right) \tag{1}$$

where x_p is the distance between pixel centers and λ is the wavelength of the laser [223]. Equation (1) must be met to satisfy the sampling theorem. The implication of this has limited digital holography to record small objects placed far away from the CCD camera. In order to record large objects, an additional optical lens system is required to reduce the object size [224].

Currently, the pixel size in high-end digital cameras is of the order of a few micrometers, which is still lower than photographic material resolution by at least an order of magnitude. Until high-resolution electronic imaging devices comparable to photo-emulsion becomes available, digital holography will find limited use in 3-D scene capture. However, in the meantime, several numerical methods and algorithms have been investigated to compensate this shortcoming and increase the resolution of image reconstruction [224]–[227]. In one other method, subimages of the holographic interference pattern were captured by using a resolution interface [228]. This interface consists of an array of apertures. The interference of the reference and object beams are projected onto the interface, which in turn projects a distribution of light spots onto the CCD. However, this process has to be repeated by changing the position of the interface until complete information of the interference pattern is acquired. At the end, a high-resolution image is attained. In another method, several shifted under-sampled digital holograms were combined to substantially enhance the resolution of the reconstructed image [229]. This allows the possibility of recording large objects using short object-to-camera distance.

Digital color holography has also been performed, using three lasers in the red, green, and blue (RGB) wavelengths. Amplitude of the interference patterns for each of the wavelengths is captured by a color CCD camera, and reconstruction methods have been adapted to create full-color holograms on the computer [230]. White-light lasers have also been used to create full-color holograms [231]. Color holography requires the object to remain constant as fringe patterns are recorded for each wavelength. Another method of producing color holography is by capturing the object at different distances and wavelengths [232].

It is foreseeable that digital holography of large time-varying objects becomes a reality with advancements in imaging technology. However, parallel efforts in the development of compact, efficient, and cost-effective RGB lasers would be necessary. These high-energy pulsed lasers should have good temporal and spatial coherence and operate at repetition rates in excess of 50 Hz.

V. Pattern Projection Techniques

A. Coded Light Approach

In many applications of 3-D shape capture such as computer graphics, virtual and augmented reality, robot navigation, and manufacturing, it is desirable to obtain the 3-D geometry of moving objects in real time and using low-cost devices.

The requirement for low-cost hardware was addressed with the introduction of the coded light approach [233] employing off-the-shelve components such as LCD projectors and CCD cameras. However, only a few approaches have been proposed aimed at real-time operation, and even fewer additionally permit unlimited movement of the scene. These systems are based on spatial coding with a single static projection pattern (thus commonly called one-shot systems), where the projected light rays or planes are encoded by spatial markings, called subpatterns, within this pattern. The first category of one-shot systems uses a black-and-white projection pattern [234], [235], which has the advantage of relatively accurate acquisition even with strongly colored scenes. However, this comes at the expense of the lateral resolution, which is typically of the area of 64×64 depth values [234] or less [235]. This is due to the low bandwidth of the typically binary projection patterns. To address this limitation, the use of color patterns was also proposed [236], [237], offering an improved lateral resolution of the depth data and the ability to employ error-detecting codes. The main difficulty with this approach is the recognition of the colors of the original pattern from the actual colors reflected by the scene and captured by the camera which may differ substantially from the originally projected ones [235]–[237].

A technique that copes with this problem is the rainbow approach [238], [239], which is based on the projection of a pattern of monochromatic colors and its encoding using the wavelength.

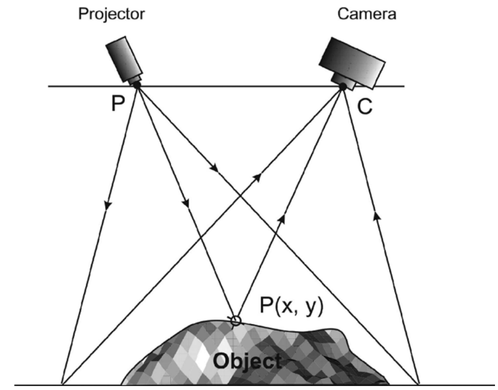

Fig. 4. Schematic of pattern projection profilometry.

The key assumption is that such a pattern yields a reflection which is modified in its intensity, but not in its spectral composition. The main limitation of this approach is the requirement of special cameras which are able to distinguish subtle changes in the wavelength; otherwise, the technique becomes very sensitive to projection and imaging noise.

A new techniques that avoids the above-mentioned difficulties was recently presented in [240] and [241]. This technique exploits the assumption that depth and reflectivity vary smoothly over the surface. Thus, instead of relying on inferring pattern colors from reflected colors, this technique relies on the detection of color transitions in local subpatterns which, given the continuity and smoothness assumptions, are relatively insensitive to the reflectivity of the scene. Furthermore, the proposed system was implemented using an off-the-shelve multimedia projector and color camera. Application of real-time 3-D scene acquisition in the context of 3-D face and gesture recognition was demonstrated in [242].

B. Fringe Projection Approach

The methods for shape capture that rely on a functional relationship between the sought object data and the phase of a periodic fringe pattern (FP), which is projected onto and reflected from the object, occupy a special place as a full-field metrological means. They are characterized with noncomplex setups and processing algorithms that are easy to implement in outdoor and industrial environments. The principle of the measurement is elucidated with the scheme depicted in Fig. 4. The optical axes of both the projector system and the observation system are crossed at a certain plane called the reference plane. In general, the 2-D intensity distribution recorded by an image sensor at moment t is given by

$$I(\vec{r}) = I_B(\vec{r}) + I_V(\vec{r}) f[\varphi(\vec{r}) + \phi(\vec{r}) + N_{\mathrm{ph}}(\vec{r})] + n_a(\vec{r}) \quad (2)$$

where I_B is a slowly varying background intensity at a point $\vec{r}$, I_V is the fringe visibility that is also a low-frequency signal, and $\varphi(\vec{r})$ is the phase term related to the object profile. The phase term $\phi(\vec{r})$ is optional, as it is introduced during the formation of the periodic waveform $f[\cdot\cdot] \in [-1, 1]$ or during the phase evaluation. Both additive noise terms N_{ph} and n_a describe phase and intensity noise contamination. The continuous

FP is imaged over a CCD camera and digitized for further analysis as a 2-D matrix of quantized intensities. The camera spatial resolution is a crucial parameter for techniques based on the principle of optical triangulation. The purpose of computer-aided fringe analysis is to determine the phase across the pattern. Once the phase of the deformed waveform is restored, nonambiguous depth or height values can be computed. Analysis of the deformed image captured with a CCD camera yields the 3-D coordinates of the object provided known positions of the camera, the projector, and the object, i.e., camera calibration is required [243]. The phase retrieval task comprises several obligatory or optional steps such as phase demodulation, including phase unwrapping [244] when necessary, removal of noninformative phase terms, extraction of information about the object [245], and denoising [246], which may be applied during each of these steps.

There have been developed various approaches for phase demodulation which give rise to the corresponding measurement schemes with different complexity, sensitivity, and accuracy. Most of the developed algorithms presume a sinusoidal profile of fringes, being inherently free of errors only at perfect sinusoidal fringe projection. Projection of purely sinusoidal fringes is not an easy task. Using light interference of two enlarged and collimated coherent beams makes large-depth and large-angle measurements possible, however, at the expense of limited lateral field of measurement, inevitable speckle noise, vulnerability to the environmental influences, and overall complexity of the used setup. Use of a conventional imaging system with different types of single-, dual-, and multiple-frequency diffraction gratings, as an amplitude or phase sinusoidal grating or Ronchi grating, enlarges the field of measurement and avoids the speckle noise, however, at the expense of higher harmonics in the projected fringes. The use of a programmable SLM, e.g., LCD [247] or DMD [248], [249], permits to control very precisely the spacing, color, and structure of the projected fringes [250]. Synthetic FPs produced by an SLM, however, suffer from the presence of the higher harmonics. The discrete nature of the projected fringes entails tiny discontinuities in the projected pattern that lead to loss of information.

Over the years, a host of phase evaluation algorithms have been proposed and tested [251]. To fulfill the goal of 3-D capture, an algorithm must ensure real-time precise calculation of the absolute 3-D coordinates of complicated objects characterized with large depth variation and discontinuities such as steps, holes, and protrusions that entail processing of wideband FPs without dominant frequency. Such are closed FPs in which the phase experiences nomonotonous change or FPs, in which the signal is noise-dependent. There exist different classifications of the adopted techniques for phase retrieval such as spatial or temporal methods, single or multiple shot methods, local or global methods, methods with and without a spatial carrier, or methods with or without phase unwrapping. A common feature of temporal analysis methods is that the phase value of a pixel is extracted based on the phase-shifted intensities of this pixel recorded in succession. Spatial analysis methods extract a phase value by evaluating the intensity of a neighborhood of the pixel being studied. A crucial requirement for implementation of any algorithm is the ability for automatic fringe analysis. As the phase retrieval involves nonlinear operations, implementation of many algorithms requires some constraints to be applied.

A popular local or pointwise approach for phase demodulation is the phase-shifting technique [252] based on sinusoidal fringe projection of at least three patterns that have undergone equal (conventional case [253]) or arbitrary (generalized case [254]) phase shifts with respect to each other. In general, the phase is retrieved by the least squares approach [255], which is applied iteratively in the case of unknown phase steps [256], [257]. The main advantage of the phase-shifting technique is its high spatial resolution, accuracy, and dynamic range. It can process patterns with closed fringes. The main drawbacks are the sensitivity to noise and the degrading effect of higher frequency components. Higher harmonics suppression has been achieved by specially designed phase-shifting algorithms [258] or by more flexible techniques such as, for example, a signal subspace method [259]. However, acquisition of several FPs is unacceptable for real-time capture.

A typical spatial analysis method is the Fourier transform method with carrier fringes [260] and without carrier fringes [261]. Introduction of a carrier frequency creates a pattern with open fringes. This is equivalent to separation in the Fourier domain of both counterparts of the fundamental spectrum from each other and from the background intensity contribution around the zero frequency. Due to the global character of the Fourier transform, the phase estimate calculated at an arbitrary pixel depends on the whole recorded FP. The Fourier-transform phase demodulation is a single-shot technique but suffers from filtering problems caused by wideband noisy carriers and a limitation on object height variation [262]. In addition, introduction of a carrier frequency which adapts to the dynamic behavior of the observed object may require expensive equipment. The choice of the filter parameters in the frequency domain is problem-dependent and relies on preliminary information about the noise and bandwidth of the modulating signals. This hampers automatic Fourier transform processing of the FPs. The Fourier transform approach uses the concept of the analytic signal. Contrary to the traditional opinion that it is not possible to find natural isotropic extension of the Hilbert transform beyond one dimension, a novel 2-D quadrature (or Hilbert) transform is developed in [263] and [264] as a combined action of two multiplicative operators: 2-D spiral-phase signum function in the Fourier space and an orientational phase spatial operator. Recently, the space-frequency representations [265], [266] have gained popularity for processing of patterns with great variation of density and orientation of fringes—the case in which the standard Fourier analysis fails. The continuous wavelet transform can be applied both to open and closed fringes and shows promising results as a denoising tool. Both gradient-based and phase-based modifications of the method have been tested [267], [268]. In the wavelet transform analysis, it is not necessary to choose the filter in the frequency domain. The other methods that are capable to ensure localized phase retrieval, as the windowed Fourier transform [269] or the regularized phase-tracking algorithm [270], generally would require *a priori* information about the fringe density and orientation. A frequently met problem is phase demodulation from patterns with partial-field fringes, in which the FP is available in a subregion of the image. The full-field methods as the Fourier transform applied to such a FP leads to artefacts at the borders. This motivates concentration of efforts on the development of phase demodulation techniques from a single FP which in general may consist of closed fringes. From the mathematical point of view phase demodulation of a single FP is an ill-posed problem because of the inherent sign ambiguity [270]. This makes impossible derivation of a unique solution from the observed data without introduction of prior constraints in the demodulation algorithm [271]. Filtering an image or phase unwrapping of noisy images are also ill-posed problems due to unknown information near the borders of the filter and noise-generated inconsistencies. Different local algorithms have been designed as the phase lock loop [272], adaptive quadrature filters [273], [274] or the regularized phase tracking [270]. The latter shows high accuracy both for patterns with open and closed fringes and is capable to process noisy images with irregular shape borders. It fits local plane surfaces to the recovered phase which makes unavoidable averaging over several pixels. Due to the fact that it seeks the phase estimate through minimization of a cost function, this approach involves iterative solving of a set of linear equations and is time-consuming. There have been developed other methods as phase demodulation based on fringe skeletonizing when an extreme map is introduced by locating the fringes minima and maxima [275], phase-stepping recovery of objects by numerical generation of multiple FPs from a single recorded FP [276] or by developing a spatial modification based on assumption of slowly varying phase [277]. The shortcoming of many of the spatial analysis methods is inevitable averaging over several pixels in the neighborhood of the point of interest.

A single-shot measurement without decrease in spatial resolution can be realized by simultaneous acquisition of several FPs which are further processed by known methods. Simultaneous projection of three color patterns (i.e., red, green, and blue) on the object at different angles and Fourier analysis of the deformed image recorded by a single CCD camera is realized in [278]. A phase-shifting method for measuring the 3-D surface of a moving object by projection of a sinusoidal grating pattern and continuous intensity acquisition by three phase-shifted linear array sensors positioned along the projected stripes is proposed in [279]. High-resolution 3-D measurement of absolute coordinates using three phase-shifted FPs coded with three primary colors and recorded at data acquisition speed of 90 fps is presented in [280]. A single-shot fringe projection system based on simultaneous projection of four phase-shifted sinusoidal FPs generated at four different wavelengths in the near infrared is proposed in [281].

In a simple pattern projection system, only one part of the object surface is viewed both by the projector and the image sensor, which yields a solid angle of about 2π for reliable measurement. Measurement of surfaces with almost vertical structures, such as, for example, cylindrical surfaces and of front and back sides of a body requires 360° of observation. In addition, shadowing in objects with a strong surface tilt or distortions caused by nonlinear recording due to specular reflection and diffraction at the object surface makes observation impossible within these parts of the image. To compensate for the loss of information, systems with multiple directions of illumination or observation are required [282]–[284]. Fig. 5 shows reconstruction of a 3-D object from two phase maps corresponding to double symmetrical observation with two CCD cameras [282]. One of the main problems that should be solved for accurate performance of such systems is to make precise transformation of the coordinate systems attached to all sensors into a common global coordinate system [283]. The matching of the point clouds obtained as a result of phase demodulation of the FPs recorded by different sensors can also be done numerically.

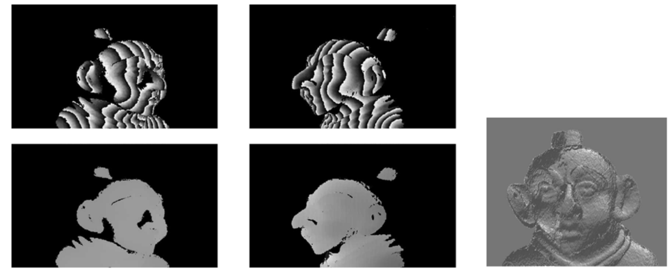

Fig. 5. Top: wrapped phase maps; bottom: unwrapped phase maps and 3-D reconstruction of the object surface. Shadow zones are masked with black.

VI. Time-of-Flight Range Imaging

Range imaging has rapidly gained popularity due to integration of distance measurement with imaging of 3-D objects. Distance measurement predominantly implements time-of-flight (TOF) range or depth acquisition principle going back to radar and lidar remote sensing, where it has been used for more than three decades [285]. In a TOF system, the time, which is required by the probing signal to travel to the target and back to the receiver multiplied by the signal velocity in the propagation medium, gives the distance to the object. Generally speaking, the information about the TOF is encoded in some measurable parameter of the signal which changes after reflection from the object, although the TOF can also be measured directly. Modern TOF systems are low-cost scannerless portable devices that are capable of yielding not only the depth map but also grayscale images of brightness distribution on the 3-D scene in real time. They use RF modulated near-infrared light emitted by an array of LEDs or laser diodes which create uniform illumination of the object [286]; an optical system focuses the incoming light field on an array of pixels, which can register both the intensity and the phase of the reflected light. If modulation of the light field is a periodic waveform $f(2\pi\nu_m t)$, where t is the time and ν_m is a radio frequency, the intensity of the reflected signal as received at a single pixel is $I(t) \propto Vf[2\pi\nu_m(t-2d/c)]$, where V is the modulation amplitude, d is the distance to the object, and c is the speed of light. Hence, the phase shift $\varphi = 4\pi\nu_m d/c$ between the emitted and the reflected modulated light waves carries depth information, whereas the contrast V of the reflected waveform permits to build brightness images as well. The phase shift is determined using a four-step algorithm $\varphi = \arctan[(I(\tau_3) - I(\tau_1))/(I(\tau_0) - I(\tau_2))]$, where $I(\tau_i)$ are four samples of the reflected waveform at $\tau_i, i = 0,1,2,3$, phase shifted at $\pi/2$. In practice, a square modulation waveform is used [287].

The TOF depth sensors suffer from the aliasing effect due to the periodic nature of the modulated signal [287]. A phase shift of 2π sets the so-called ambiguity distance which is equal to 7.5 m at $\nu_m = 20$ MHz. Lowering the modulation frequency increases the ambiguity distance at the expense of decreased spatial resolution. The solution of this problem to ensure both a large range of unambiguous depth determination and reasonable resolution is to use multiple modulation frequencies [287].

The TOF depth sensors outperform other depth sensors as a result of the TOF measurement specifics. Uniform illumination of the scene *a priori* assumes that, as a whole, the system may operate with less light in comparison with coded light and fringe projection techniques whose accuracy depends on the contrast of the projected pattern. In addition, no moving light parts are introduced and the eye-safety problem is avoided. The TOF measurement is texture-independent, and the retrieval of 3-D coordinates requires minimal post-processing procedures and can be done without necessity of a baseline between the light source and the camera. Thus, very fast frame rates can be achieved, and tracking of moving objects can be easily realized. There have been proposed CCD-based TOF systems as described in [288] and [289], but the receiver part of the sensor can be designed as a single low-cost CMOS chip combined with a CPU [290]–[292]. The depth resolution of the working TOF systems which at the first place depends on the amount of light captured by the array of pixels is of the order of a few millimeters. Performance of the TOF depth sensor could be improved by hardware and software noise reduction and precise calibration [293], which ensure removal of the systematic errors and more accurate recovery of the 3-D coordinates.

Several amplitude-modulated continuous-wave TOF range-imaging devices are currently under development, such as the SR2 and the SR-3000 developed by the Centre Suisse d'Electronique et de Microtechnique SA (CSEM) or the CanestaVision DP205 developed by Canesta, Inc. The SR2 [294], [295] implements a CMOS/CCD sensor array with approximately 20 000 pixels and an illumination block of 48 infrared LEDs with $\nu_m =$ 20 MHz. The SR-3000 has about 25 000 pixels [296] and background suppression to avoid saturation effects, which permits its usage in outdoor conditions. The depth resolution varies from 5 to 10 mm. The Canesta system is a CMOS-based device [287] that is able to operate in a dual-frequency mode having four modulation frequencies (13, 26, 52, or 104 MHz).

The achieved depth resolution and their real-time operation make the TOF range-imaging systems suitable for different middle-accuracy applications, such as scene extraction for 3DTV, tracking, recognition, image understanding, autonomous navigation, real-time modeling, and mapping, which all benefit from the direct delivery of geometry information by the depth images.

VII. Conclusion

The type of 3-D scene and its specific requirements for the 3-D dynamic display envisaged for its visualization predetermine the choice of the proper methodology and instrumentation for recording of a 3-D object shape and color and further digital reconstruction. Evaluation of the inherent advantages and drawbacks of the existing time-varying scene-capture techniques shows that all of them might find applications in the coming era of 3DTV. The low-cost single-camera technique solves the problems in mono-view video content and is especially promising as a link between the available conventional video sequences and the 3-D displays. Among the approbated techniques for single-camera capture, the structure-from-motion method can be accepted as the most suitable approach for capturing real life scenes. A perspective branch in this area of research is development of specific scene-oriented approaches with particular applications, such as the human face and body capture techniques. On the other hand, implementation of synchronized multicamera acquisition imposes stringent requirements on the hardware and increases the cost of the overall 3DTV system. A variety of algorithms has been proposed for creation of dynamic scene representations from the data acquired from the multiview points. In this aspect, the virtual free-viewpoint video generation from arbitrarily chosen viewing angles is an important step for 3DTV displays. A low-cost hardware alternative for real-time 3-D shape capture of moving objects is the implementation of different types of pattern projection, such as coded light approach or sinusoidal fringe projection. The extraction of point clouds and color coordinates of the recorded objects out of 2-D images by using a one-shot system relies on development of decoding algorithms from a single image or on the design of special methods for simultaneous acquisition of several images. Finally, digital holography is another candidate for low-cost capture systems, provided that certain issues, such as image resolution, data storage, and true color recording and reconstruction, are successfully solved. Resolution achieved by the modern TOF range-imaging system that are capable of determining the distance map, i.e., the 3-D model of its environment, as well as the local brightness in the scene in real time, make them a useful tool in middle-accuracy applications of computer vision.

The time-varying 3-D scene-capture techniques are still relatively immature to fulfill all of the existing requirements of 3DTV systems and remain a challenging research field. However, their potentials in areas such as computer graphics, virtual reality, robot navigation, metrology, cultural heritage protection, and biomedical investigations are yet to be fully exploited.

Acknowledgment

All the contributors of a previous report, entitled "3D Time-varying Scene Capture Technologies: A Survey" of IST 3DTV NoE are greatly acknowledged. In particular, the authors are indebted to S. Knorr (Technical University of Berlin) and H. E. Sankaran (Tampere University of Technol.), C. E. Erdem, T. Erdem (Momentum), M. Ozkan, C. Turun (Yogurt), A. Weissenfeld (University of Hannover).

References

[1] R. Zhang, P. Tsai, J. Cryer, and M. Shah, "Shape from shading: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 8, pp. 690–706, Aug. 1999.
[2] T. Lindeberg and J. Garding, "Shape from texture from a multi-scale perspective," in *Proc. ICCV*, May 1993, pp. 683–691.
[3] M. Subbarao and G. Surya, "Depth from defocus: A spatial domain approach," *Int. J. Comput. Vis.*, vol. 13, pp. 271–294, 1994.
[4] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *Int. J. Comput. Vis.*, vol. 37, no. 2, pp. 151–172, 2000.
[5] R. Deriche and G. Giraudon, "A computational approach for corner and vertex detection," *Int. J. Comput. Vis.*, vol. 10, no. 2, pp. 101–124, 1993.

[6] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[7] O. Faugeras, *Three Dimensional Computer Vision: A Geometric Viewpoint*. Cambridge, MA: MIT Press, 1993.
[8] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3-Dmachine vision metrology using off-the-shelf TV cameras and lenses," *IEEE J. Robot. Autom.*, vol. 3, no. 4, pp. 323–344, Aug. 1987.
[9] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
[10] O. Faugeras, L. Quang-Tuan, and S. Maybank, "Camera self-calibration: Theory and experiments," in *Proc. 2nd Eur. Conf. Comput. Vis.*, 1992, pp. 321–334.
[11] L. Quang-Tuan and O. Faugeras, "Self-calibration of a moving camera from point correspondences and fundamental matrices," *Int. J. Comput. Vis.*, vol. 22, no. 3, pp. 261–289, 1997.
[12] C. Zeller and O. Faugeras, Camera Self-Calibration From Video Sequences: The Kruppa Equations Revisited INRIA, Tech. Rep. 00073897, 1996.
[13] R. Hartley, "Kruppa's equations derived from the fundamental matrix," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 133–135, Feb. 1997.
[14] L. Cheng, A New Approach to Solving Kruppa Equations for Camera Self-Calibration ICPR, 2002.
[15] A. Fusiello, "Uncalibrated euclidean reconstruction: A review," *Image Vis. Comput.*, vol. 18, pp. 555–563, 2000.
[16] A. Heyden and K. Astrom, "Euclidean reconstruction from constant intrinsic parmeters," in *Proc. 13th Int. Conf. Pattern Recognit.*, 1996, pp. 339–343.
[17] M. Pollefeys and L. van Gool, "Self-calibration from the absolute conic on the plane at infinity," in *Proc. Comput. Anal. Images Patterns*, 1997, pp. 175–182.
[18] M. Pollefeys and L. van Gool, "A stratified approach to self-calibration," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, 1997, pp. 407–412.
[19] M. Pollefeys, L. van Gool, and A. Oosterlinck, "The modulus constraint: A new constraint for self calibration," in *Proc 13th Int. Conf. Pattern Recognit.*, 1996, pp. 349–357.
[20] Q. T. Luong, R. Deriche, O. Faugeras, and T. Papadopoulo, On Determining the Fundamental Matrix: Analysis of Different Methods and Experimental Results INRIA, Tech. Rep. Raport de Recherche 1894, 1993.
[21] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *Int. J. Comput. Vis.*, vol. 7, no. 2, pp. 161–195, 1998.
[22] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography.," *Commun. ACM*, pp. 381–385, 1981.
[23] P. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comp. Vis. Imag. Underst*, vol. 78, pp. 138–156, 2000.
[24] D. Myatt, P. Torr, S. Nasuto, J. Bishop, and R. Craddock, "NAPSAC: High noise, high dimensional robust estimation—it's in the bag," in *Proc. Brit. Mach. Vis. Conf. BMVC*, 2002, pp. 138–156.
[25] Z. Zhang, R. Deriche, O. D. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artif. Intell.*, vol. 78, no. 1–2, pp. 87–119, 1995.
[26] A. M. Tekalp, *Digital Video Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[27] P. Beardsley, A. Zisserman, and D. Murray, "Navigation using affine structure from motion," in *Computer Vision*, ser. LNCS Series 801. Berlin, Germany: Springer-Verlag, 1994, pp. 85–96.
[28] R. Haralick and L. Shapiro, *Computer and Robot Vision*. Reading, MA: Addison-Wesley, 1992, vol. 1.
[29] R. Hartley, R. Gupta, and T. Chang, "Stereo from uncalibrated cameras," in *Proc. CVPR*, 1992, pp. 761–764.
[30] R. Hartley and P. Sturm, "Triangulation," *Comput. Vis. Image Underst.*, vol. 68, no. 2, pp. 146–157, 1997.
[31] U. Dhond and J. Aggrawal, "Structure from stereo—A review," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 6, pp. 1489–1510, Jun. 1989.
[32] L. Alvarez *et al.*, Dense Disparity Map Estimation Respecting Image Discontinuities: A pde and Scale-Space Based Approach INRIA Sophia-Antipolis, Tech. Rep. 3874, Jan. 2000.
[33] C. Strecha, T. Tuytelaars, and L. van Gool, "Dense matching of multiple wide-baseline views," in *Proc. 9th IEEE ICCV*, 2003, pp. 1194–1201.
[34] L. Falkenhagen, "Hierarchical block-based disparity estimation considering neighborhood constraints," in *Proc. Int. Workshop SNHC 3-DImaging*, Greece, 1997, pp. 115–122.
[35] G. V. Meerbergen, M. Vergauwen, M. Pollefeys, and L. van Gool, "A hierarchical symmetric stereo algorithm using dynamic programming," *Int. J. Comput. Vis.*, vol. 47, no. 1/2/3, pp. 275–285, 2002.
[36] S. Roy and I. Cox, "A maximum-flow formulation of the N-camera stereo correspondence problem," in *Proc. Int. Conf. Comput. Vis.*, Bombay, India, 1998, pp. 492–499.
[37] A. Alatan and L. Onural, "Estimation of depth fields suitable for video compression based on 3-D structure and motion objects," *IEEE Trans. Image Process.*, vol. 7, no. 6, pp. 904–908, Jun. 1998.
[38] A. Alatan and L. Onural, "Object based 3-D motion and structure estimation," in *Proc. Int. Conf. Image Process.*, 1995, vol. 1, pp. 390–393.
[39] Charbonnier *et al.*, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *Proc. IEEE ICIP*, Austin, TX, Nov. 1996, pp. 168–172.
[40] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.
[41] H. Nagel and J. Malik, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 8, pp. 565–593, Aug. 1986.
[42] A. Alatan and L. Onural, "Gibbs random field model based 3-D motion estimation by weakened rigidity," in *Proc. IEEE Int. Conf. Image Process.*, 1994, vol. 2, pp. 494–497.
[43] M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 34–58, Jan. 2002.
[44] Z. Zhang *et al.*, A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry INRIA, Tech. Rep. 2273, May 1994.
[45] T. Jebara, A. Azarbayejani, and A. Pentland, "3-D structure from 2-D motion," *IEEE Signal Process. Mag.*, vol. 16, no. 3, pp. 66–84, May 1999.
[46] C. Tomasi and T. Kanade, "Shape and motion from image streams: A factorization method," *J. Comput. Vis.*, vol. 9, no. 2, pp. 137–154, 1992.
[47] T. Kanade and D. Morris, "Factorization methods for structure from motion," *Phil. Trans. R. Soc. Long A.*, pp. 1153–1173, 1998.
[48] J. L. Barron and R. Eagleson, "Recursive estimation of time varying motion and structure parameters," *Pattern Recognit.*, vol. 29, no. 5, pp. 797–818, 1995.
[49] J. Thomas and J. Oliensis, "Dealing with noise in multiframe structure from motion," *Comput. Vis. Image Underst.*, vol. 76, no. 2, pp. 109–124, Nov. 1999.
[50] A. Chowdury and R. Chellappa, "Statistical error propagation in 3-D modeling from monocular video," in *Proc. CVPR Workshop Stat. Anal.*, 2003, pp. 89–92.
[51] T. Broida and R. Chellappa, "Experiments and uniqueness results on object structure ad kinematics from a sequence of monocular images," in *Proc. Workshop Visual Motion*, Mar. 1989, pp. 21–30.
[52] A. Azarbayejani and A. Pentland, "Recursive estimation of motion, structure and focal length," in *Proc. 2nd CAD Based Vis. Workshop*, 1994, pp. 90–97.
[53] R. Szeliski and S. B. Kang, Recovering 3-D Shape and Motion From Image Streams Using Non-Linear Least Squares Cambridge Res. Lab., Tech. Rep., 1993, Digital Equipment Corporation.
[54] C. J. Poelman and T. Kanade, A Paraperspective Factorization Method for Shape and Motion Recovery Carnegie Mellon Univ., Tech. Rep. CMU-CS-93-219, Dec. 1993.
[55] C. Poelman and T. Kanade, "A paraperspective factorization method for shape and motion recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 3, pp. 206–218, Mar. 1997.
[56] J. Fujiki, T. Kurata, and T. Tanaka, "Iterative factorization method for object recognition," *Proc. SPIE*, vol. 3554-18, pp. 192–201, 1998.
[57] B. Triggs, "Factorization methods for projective structure and motion," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, San Fransisco, CA, 1996, pp. 845–849.
[58] P. Strum and B. Triggs, "A factorization based algorithm for multi-image projective structure and motion," in *Proc. 4th Eur. Conf. Comput. Vis.*, Cambridge, U.K., Apr. 1996, vol. 1065, pp. 709–720.
[59] B. Triggs, Some Notes on Factorization Methods for Projective Structure and Motion 1998, unpublished.
[60] T. Morita and T. Kanade, "A sequential factorization method for recovering shape and motion from image streams," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 8, pp. 858–867, Aug. 1997.
[61] L. Quan and T. Kanade, "A factorization method for affine structure from line correspondences," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, U. San Fransisco, CA, 1996, pp. 803–808.
[62] M. Pollefeys, "Tutorial on 3-Dmodeling from images," presented at the ECCV, 2000, unpublished.

[63] J. Oliensis, "A critique of structure from motion algorithms," *Comput. Vis. Image Underst.*, vol. 80, no. 2, pp. 172–214, Nov. 2000.
[64] S. Soatto and P. Perona, "Reducing structure from motion: A general framework for dynamic vision part 1: Modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 9, pp. 933–942, Sep. 1998.
[65] T. J. Broida, S. Chandrasekhar, and R. Chellappa, Recursive 3-D motion estimation from a monocular image sequence 1996, unpublished.
[66] R. Chellappa, G. Qian, and S. Srinivasan, "Structure from motion: Sparse versus dense correspondence methods," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 1999, pp. 492–499.
[67] H. Jin, P. Favaro, and S. Soatto, "A semi-direct approach to structure from motion," *Visual Comput.*, vol. 19, pp. 1–18, 2003.
[68] G. Yang and T. S. Huang, "Human face detection in complex background," *Pattern Recognit.*, vol. 27, pp. 53–63, 1994.
[69] C. Kotropoulus and I. Pitas, "Rule-based face detection in frontal views," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, 1997, vol. 4, pp. 2537–2540.
[70] I. Craw, H. Ellis, and J. R. Lishman, "Automatic extraction of face-feature," *Pattern Recog. Lett.*, pp. 183–187, 1987.
[71] X. Jiang, M. Binkert, B. Achermann, and H. Bunke, "Towards detection of glasses in facial images," *Pattern Anal. Appl.*, vol. 3, pp. 9–18, 2000.
[72] S. Sirohey, Human Face Segmentation and Identification Univ. of Maryland, Tech. Rep. CS-TR-3176, 1993.
[73] D. Chetverikov and A. Lerch, "Multiresolution face detection," *Theoret. Found. Comput. Vis.*, vol. 69, pp. 131–140, 1993.
[74] H. Graf, T. Chen, E. Petajan, and E. Cosatto, "Locating faces and facial parts," in *Proc. 1st Int. Workshop Autom. Face Gesture Recognit.*, 1995, pp. 41–46.
[75] C. Han, H. Liao, K.-C. Yu, and L.-H. Chen, "Fast face detection via morphology-based pre-processing," in *Proc. 9th Int. Conf. Image Anal. Process.*, 1998, pp. 469–476.
[76] K. Yow and R. Cipolla, "Feature-based human face detection," *Image Vis. Comput.*, vol. 15, no. 9, pp. 713–735, 1997.
[77] T. Leung, M. Burl, and P. Perona, "Finding faces in cluttered scenes using random labeled graph matching," in *Proc. 5th IEEE Int. Conf. Comput. Vis.*, 1995, pp. 637–644.
[78] M. Augusteijn and T. Skujca, "Identification of human faces through texture-based feature recognition and neural network technology," in *Proc. IEEE Conf. Neural Netw.*, 1993, pp. 392–398.
[79] Y. Dai and Y. Nakano, "Face-texture model based on SGLD and its application in face detection in a color scene," *Pattern Recogn.*, vol. 29, no. 6, pp. 1007–1017, 1996.
[80] J. Terrillon, M. Shirazi, H. Fukamachi, and S. Akamatsu, "Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recogn.*, 2000, pp. 54–61.
[81] D. Chai and K. Ngan, "Locating facial region of a head and shoulders color image," in *Proc. Third Int. Conf. Automatic Face and Gesture Recognition*, 1998, pp. 124–129.
[82] D. Saxe and R. Foulds, "Toward robust skin identification in video images," in *Proc. Second Int. Conf. Autom. Face and Gesture Recogn.*, 1996, pp. 379–384.
[83] R. Kjeldsen and J. Kender, "Finding skin in color images," in *Proc. Second Int. Conf. Autom. Face and Gesture Recogn.*, 1996, pp. 312–317.
[84] M. Yang and N. Ahuja, "Detecting human faces in color images," in *Proc. IEEE Int. Conf. Image Process.*, 1998, vol. 1, pp. 127–130.
[85] V. Govindaraju, S. Srihari, and D. Sher, "A computational model for face location," in *Proc. Third IEEE Int. Conf. Comput. Vis.*, 1990, pp. 718–721.
[86] P. Sinha, "Object recognition via image invariants: A case study," *Investigative Ophthalmology and Visual Sci.*, vol. 35, no. 4, pp. 1735–1740, 1994.
[87] B. Scassellati, "Eye finding via face detection for a foevated, active vision system," in *Proc. 15th National Conf. Artificial Intelligence*, 1998, pp. 969–976.
[88] J. Miao, B. Yin, K. Wang, L. Shen, and X. Chen, "A hierarchical multiscale and multiangle system for human face detection in a complex background using gravity-center template," *Pattern Recognit.*, vol. 32, no. 7, pp. 1237–1248, 1999.
[89] Y. Kwon and N. da Vitoria Lobo, "Face detection using templates," in *Proc. Int. Conf. Pattern Recognit.*, 1994, pp. 764–767.
[90] A. L. Yuille, P. W. Hallinan, and D. S. Cohen, "Feature extraction from faces using deformable templates," *Int. J. Comput. Vis.*, vol. 8, no. 2, pp. 99–111, 1992.
[91] A. Lanitis, C. Taylor, and T. Cootes, "An automatic face identification system using flexible appearance models," *Image and Vis. Computing*, vol. 13, no. 5, pp. 393–401, 1995.
[92] H. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, 1996, pp. 203–208.
[93] T. Darrell, G. Gordon, M. Harville, and J. Woodfill, "Integrated person tracking using stereo, color, and pattern detection," in *IEEE Proc. Int. Conf. Comput. Vis. and Pattern Recognit.*, 1998, pp. 185–195.
[94] M. LaCascia, S. Sclaroff, and V. Athitsos, "Fast, reliable head tracking under varying llumination: An approach based on registration of textured-mapped 3-D models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 322–336, 2000.
[95] H. A. Rowley, S. Baluja, and T. Kanade, "Rotation invariant neural network-based face detection," in *Proc. IEEE Int. Conf. Comput. Vis. and Pattern Recognit.*, 1998, pp. 38–44.
[96] S.-H. Lin, S.-Y. Kung, and L.-J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 114–132, 1997.
[97] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 103–108, Jan. 1990.
[98] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cog. Neurosci*, vol. 3, pp. 71–86, 1991.
[99] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: An application to face detection," in *Proc. IEEE Int. Conf. Comput. Vis. and Pattern Recognit.*, 1997, vol. 6, pp. 130–136.
[100] J. Terrillon, M. Shirazi, M. Sadek, H. Fukamachi, and S. Akamatsu, "Invariant face detection with support vector machines," in *Proc. 15th Int. Conf. Pattern Recognit.*, 2000, vol. 4, p. 4B.
[101] D. Roth, M.-H. Yang, and N. Ahuja, "A snow-based face detector," in *In Advances in Neural Information Processing Systems 12 (NIPS 12)*. Cambridge, MA: MIT Press, 2000.
[102] H. Schneiderman and T. Kanade, "Probabilistic modeling of local appearance and spatial relationships for object recognition," in *IEEE Conf. Comput. Vis. and Pattern Recognit.*, 1998, vol. 6, p. 45.
[103] H. Schneiderman and T. Kanade, "A statistical model for 3-D object detection applied to faces and cars," in *IEEE Conf. Comput. Vis. and Pattern Recognit.*, 2000, pp. 746–751.
[104] C. Bregler, "Learning and recognizing human dynamics in video sequences," in *Proc. IEEE Int. Conf. Comput. Vis. and Pattern Recogn.*, San Juan, Puerto Rico, 1997, pp. 568–574.
[105] N. Werghi and Y. Xiao, "Posture recognition and segmentation from 3-D human body scans," in *Proc. Int. Conf. 3-D Data Processing Visualization and Transmission*, Padova, Italy, 2002, pp. 636–639.
[106] P. Eisert and B. Girod, "Analyzing facial expressions for virtual conferencing," *IEEE Comput. Graph. Appl.*, pp. 70–78, Sept. 1998.
[107] N. Sarris, N. Grammalidis, and M. Strintzis, "Building three-dimensional head models," *Graphical Models*, vol. 63, no. 5, pp. 333–368, 2001.
[108] Z. Liu, Z. Zhang, C. Jacobs, and M. Cohen, Rapid modeling of animated faces from video Microsoft Research, Tech. Rep. MSR-TR-2000-11, 2000.
[109] Y. Sheng, A. H. Sadka, and A. M. Kondoz, "Automatic 3-D face synthesis using single 2-D video frame," *Electron. Lett.*, vol. 40, no. 19, pp. 1173–1175, Sept. 2004.
[110] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. H. Salesin, "Synthesizing realistic facial expressions from photographs," in *Int. Conf. Comput. Graphics and Interactive Techniques*, 1998, pp. 75–84.
[111] R. Koch, "Adaptation of a 3-D facial mask to human faces in videophone sequences using model based image analysis," in *Picture Coding Symp. PCS'91*, 1991, pp. 285–288.
[112] M. Kampmann and R. Farhoud, "Precise face model adaptation for semantic coding of videophone sequences," in *Picture Coding Symp.*, Berlin, Germany, Sept. 1997, pp. 85–87.
[113] J. Noh and U. Neumann, "Expression cloning," in *Comput. Graphics, Proceedings of ACM SIGGRAPH 2001*, Los Angeles, CA, Aug. 2001, pp. 277–288.
[114] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quart. Appl. Math.*, vol. 2, no. 2, pp. 164–168, Sept. 1944.
[115] M. J. D. Powell, "Radial basis functions for multivariate interpolation: A review," in *Algorithms for Approximation*. Oxford: Clarendon Press, 1987.
[116] A. del Valle and J. Ostermann, "3-D talking head customization by adapting a generic model to one uncalibrated picture," in *ISCAS*, Sydney, Australia, 2001, pp. 325–328.
[117] D. Lowe, "Robust model-based motion tracking through the integration of search and estimation," *Int. J. Comput. Vis.*, vol. 8, no. 2, pp. 113–122, 1992.
[118] Liu and Z. Zhang, "Robust head motion computation by taking advantage of physical properties," in *Workshop on Human Motion*, 2000, p. 73.

[119] S. Shapiro, A. Zisserman, and M. Brady, "3-D motion recovery via affine epipolar geometry," *Int. J. Comput. Vis.*, vol. 16, no. 2, pp. 147–182, 1995.
[120] M. J. Black, "Robust Incremental Optical Flow," Ph.D. Dissertation, Yale University, New Haven, 1992.
[121] P. Li, P. Roivanen, and R. Forchheimer, "3-D motion estimation in model-based facial image coding," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 6, pp. 545–555, 1993.
[122] G. D. Hager and P. N. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 10, pp. 1025–1039, 1998.
[123] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *ECCV 92: Proc. 2nd Eur. Conf. Comput. Vis.*, 1992, pp. 237–252.
[124] S. Birchfield, "Elliptical head tracking using intensity gradients and color histograms," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. and Pattern Recognit.*, 1998, p. 232.
[125] S. Basu, I. Essa, and A. Pentland, "Motion regularization for model-based head tracking," in *Int. Conf. Pattern Recognit.*, 1996, p. 611.
[126] J. Xiao, T. Kanade, and J. F. Cohn, "Robust full-motion recovery of head by dynamic templates and re-registration techniques," in *FGR 02: Proc. Fifth IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2002, p. 163.
[127] M. L. Cascia, S. Sclaroff, and V. Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on robust registration of texture-mapped 3-D models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 4, pp. 322–336, Apr. 2000.
[128] M. Malciu and F. Preteux, "A robust model-based approach for 3-D head tracking in video sequences," in *Proc. SPIE*, 2000, vol. 4121, pp. 51–62.
[129] A. Bottino, "Real time head tracking from uncalibrated monocular views," in *Proc. 5th Asian Conf. Comput. Vis. (ACCV 2002)*, Melbourne, Australia, Jan. 2002, pp. 23–25.
[130] I. A. Essa and A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 757–763, 1997.
[131] D. DeCarlo and D. N. Metaxas, "Optical flow constraints on deformable models with applications to face tracking," *Int. J. Comput. Vis.*, vol. 38, no. 2, pp. 99–127, 2000.
[132] F. Dornaika and J. Ahlberg, "Fast and reliable active appearance model search for 3-D face tracking," *IEEE Trans. Syst., Man, Cybern. B*, vol. 34, no. 4, pp. 1838–1853, 2004.
[133] R. Rosales, M. Siddiqui, J. Alon, and S. Sclaroff, Estimating 3-D body pose using uncalibrated cameras Comput. Sci. Dep., Boston Univ., Tech. Rep. 2001-008, Jun. 2001.
[134] I. M. M. Trivedi, E. Hunter, and P. Cosman, "Human body model acquisition and tracking using voxel data," *Int. J. Comput. Vis.*, vol. 53, no. 3, pp. 199–223, 2003.
[135] F. Remondino, "Human body construction from image sequences," in *Lecture Notes in Comput. Sci.*, 2002, vol. 2449, pp. 50–57.
[136] A. Laurentini and A. Bottino, "A silhouette based technique for the reconstruction of human movement," *Comput. Vis. and Image Underst.*, vol. 83, pp. 79–95, 2001.
[137] I. Kakadiaris and D. Metaxas, "3-D human body model acquisition from multiple views," in *Proc. 5th Int. Conf. Comput. Vis.*, 1995, pp. 618–623.
[138] G. Cheung, S. Baker, J. Hodgins, and T. Kanade, "Markerless human motion transfer," in *Second Int. Symp. on 3-D Data Processing, Visualization and Transmission (3DPVT'04)*, 2004, pp. 373–378.
[139] C.-E. Liedtke, H. Busch, and R. Koch, "Automatic modelling of 3-D moving objects from a TV image sequence," in *SPIE/SPSE Symp. on Electronic Imaging*, Santa Clara, CA, Feb. 1990, pp. 230–239.
[140] B. Rosenhahn, L. He, and R. Klette, "Automatic human model generation," in *Comput. Analysis of Images and Patterns (CAIP)*, 2005, pp. 41–48.
[141] D. Grest, D. Herzog, and R. Koch, "Human model fitting from monocular posture images," in *Proc. VMV 2005*, Erlangen, Germany, Nov. 2005, p. 107.
[142] H. Sidenbladh, M. Black, and L. Sigal, "Implicit probabilistic models of human motion for synthesis and tracking," in *Proc. European Conf. Comput. Vis.*, Copenhagen, Denmark, 2002, p. 784.
[143] R. Okada, Y. Shirai, and J. Miura, "Tracking a person with 3-D motion by integrating optical flow and depth," in *Proc. 4th IEEE Int. Conf. Automatic Face and Gesture Recognition*, Grenoble, France, 2000, pp. 336–341.
[144] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, 1997.
[145] C. Cédras and M. Shah, "Motion-based recognition: A survey," *Image and Vis. Computing*, vol. 13, no. 2, pp. 129–155, 1995.
[146] J. Aggarwal and Q. Cai, "Human motion analysis: A review," *Comput. Vis. and Image Understanding*, vol. 73, no. 3, pp. 428–440, 1999.
[147] D. Gavrila, "The visual analysis of human movement: A survey," *Comput. Vis. and Image Underst.*, vol. 73, no. 1, pp. 82–98, 1999.
[148] T. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Comput. Vis. and Image Underst.*, vol. 81, no. 3, pp. 231–268, 2001.
[149] A. Ali and J. Aggarwal, "Segmentation and recognition of continuous human activity," in *Proc. IEEE Workshop on Detection and Recognition of Events in Video*, Vancouver, Canada, 2001, p. 28.
[150] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 3, pp. 257–267, 2001.
[151] J. Davis and A. Bobick, , N. M.-T. D. Thalmann, Ed., "A robust human-silhouette extraction technique for interactive virtual environments," in *Lecture Notes in Artificial Intelligence*. Heidelberg, New York: Springer-Verlag, 1998, pp. 12–25.
[152] S. Park and J. Aggarwal, "Recognition of human interaction using multiple features in grayscale images," in *Proc. 15th Int. Conf. Pattern Recognit.*, Barcelona, Spain, 2000, p. 1051.
[153] K. Sato and J. K. Aggarwal, "Tracking and recognizing two-person interactions in outdoor image sequences," in *IEEE Workshop on Multi-Object Tracking*, Vancouver, Canada, 2001, p. 87.
[154] I. Haritaoglu, D. Harwood, and L. Davis, "W4: real-time system for detecting and tracking people," in *Proc. IEEE Comp. Soc. Conf. Comput. Vis. and Pattern Recognit.*, 1998, p. 877.
[155] D. Ayers and M. Shah, "Monitoring human behavior from video taken in an office environment," *Image and Vis. Computing*, vol. 19, no. 12, pp. 833–846, 2001.
[156] G. Cheung, T. Kanade, J. Bouguet, and M. Holler, "A real time system for robust 3-D voxel reconstruction of human motions," in *IEEE CVPR*, 2000, pp. 714–720.
[157] W. Matusik, C. Buheler, R. Raskar, S. Gortler, and L. McMillan, "Image-based visual hulls," in *SIGGRAPH*, 2000, pp. 369–374.
[158] P. Baker and Y. Aloimonos, "Complete calibration of a multi-camera network," in *IEEE Workshop Omnidirectional Vis.*, 2000, pp. 134–141.
[159] D. Brady, R. Stack, S. Feller, L. Cull, D. Kammeyer, and B. R. , "Information flow in streaming 3-D video.," in *Three-Dimensional Video and Display Devices and Systems*. : SPIE PRESS, 2000, vol. CR76.
[160] B. Wilburn, N. Joshi, V. Vaish, M. Levoy, and M. Horowitz, "High speed video using a dense camera array," in *IEEE CVPR*, 2004, pp. II-294–II-301.
[161] P. Kauff and O. Schreer, "An immersive 3-D video-conferencing system using shared virtual team user environments," in *ACM Collaborative Virtual Environments*, 2002, pp. 105–112.
[162] J. Mulligan, X. Zabulis, N. Kelshikar, and K. Daniilidis, "Stereo-based environment scanning for immersive telepresence," *IEEE CSVT*, vol. 14, no. 3, pp. 304–320, 2004.
[163] H. Baker, D. Tanguay, I. Sobel, D. Gelb, M. Gross, W. Culbertson, and T. Malzbender, "The Coliseum immersive teleconferencing system," presented at the Int. Workshop Immersive Telepresence, Juan-les-Pins, France, 2002 [Online]. Available: http://citeseer.ist.psu.edu/baker02coliseum.html
[164] T. Kanade, A. Yoshida, K. Oda, H. Kano, and M. Tanaka, "A stereo engine for video-rate dense depth mapping and its new applications," in *IEEE CVPR*, 1996, pp. 196–202.
[165] P. Narayanan, P. Rander, and T. Kanade, "Constructing virtual worlds using dense stereo," in *ICCV*, 1998, pp. 3–10.
[166] M. Billinghurst, A. Cheok, S. Prince, and H. Kato, "Projects in VR: Real world teleconferencing," *IEEE Comput. Graphics and Applcations*, vol. 22, pp. 11–13, 2002.
[167] J. Y. Bouguet, Camera Calibration Toolbox for Matlab [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc
[168] T. Svoboda, D. Martinec, and T. Pajdla, "A convenient multi-camera self-calibration for virtual environments.," in *PRESENCE: Teleoperators and Virtual Environments*. : MIT Press, 2005, vol. 14.
[169] X. Zabulis, J. P. Barreto, N. Kelshikar, R. Molana, and K. Daniilidis, "Volumetric multi- camera scene acquisition with partially metric calibration for wide-area tele-immersion," in *NSF Workshop on Collaborative Virtual Reality and Visualization*, 2003 [Online]. Available: http://www.cis.upenn.edu/teleimmersion/research/downloads/EasyCal
[170] P. G. Research [Online]. Available: http://www.ptgrey.com/products/sync/index.html
[171] O. Imaging [Online]. Available: http://www.objectiveimaging.com/OASIS_DC1.htm
[172] N. Liu, Y. Wu, X. Tan, and G. Lai, "Control system for several rotating mirror camera synchronization operation," in *Int. Congress on High-Speed Photography and Photonics*, 1997, pp. 695–699.

[173] B. Holveck and H. Mathieu, Infrastructure of the GrImage Experimental Platform: The Video Acquisition Part INRIA, Tech. Rep. rT-0301, 2004.
[174] G. Litos, X. Zabulis, and G. Triantafyllidis, "Synchronous image acquisition based on network synchronization," in *IEEE Workshop on Three-Dimensional Cinematography*, 2006, p. 167.
[175] D. L. Mills, "Internet time synchronization: The network time protocol," *IEEE Trans. Commun.*, vol. 10, pp. 1482–1492, 1991.
[176] T. Svoboda, H. Hug, and L. van Gool, "ViRoom—Low cost synchronized multicamera system and its self-calibration," in *Pattern Recognit., DAGM Symp.*, 2002, pp. 515–522.
[177] P. Rai, K. Tiwari, P. Guha, and A. Mukerjee, "A cost-effective multiple camera vision system using FireWire cameras and software synchronization," in *Int. Conf. High Performance Computing*, 2004, pp. 17–20.
[178] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proc. Comput. Vis. and Pattern Recognit. (CVPR)*, 2006, vol. 1, pp. 519–526 [Online]. Available: citeseer.ist.psu.edu/seitz06comparison.html, [Online]. Available:
[179] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," *IJCV*, vol. 38, no. 3, pp. 197–216, 2000.
[180] A. Bowen, A. Mullins, R. Wilson, and N. Rajpoot, "Light field reconstruction using a planar patch model," in *Scandinavian Conf. Image Processing*, 2005, pp. 85–94.
[181] R. Carceroni and K. Kutulakos, "Multi-view scene capture by surfel sampling: From video streams to non-rigid 3-D motion, shape & reflectance," *Int. J. Comput. Vis.*, vol. 49, no. 2–3, pp. 175–214, 2002.
[182] O. Faugeras and R. Keriven, "Complete dense stereovision using level set methods," in *Eur. Conf. in Comput. Vis.*, 1998, pp. 379–393.
[183] X. Zabulis and K. Daniilidis, "Multi-camera reconstruction based on surface normal estimation and best viewpoint selection," in *IEEE Int. Symp. on 3-D Data Processing, Visualization and Transmission*, 2004, pp. 733–40.
[184] W. Culbertson, T. Malzbender, and G. Slabaugh, "Generalized voxel coloring," in *Vis. Algorithms: Theory and Practice*, 1999, pp. 100–115.
[185] A. Treuille, A. Hertzmann, and S. Seitz, "Example-based stereo with general BRDFs," in *8th Eur. Conf. Comput. Vis. (ECCV 2004)*, 2004, vol. II, pp. 457–469.
[186] A. Broadhurst, T. Drummond, and R. Cipolla, "A probabilistic framework for space carving," in *ICCV*, 2001, pp. 388–393.
[187] S. Roy and I. J. Cox, "A maximum-flow formulation of the n-camera stereo correspondence problem," in *ICCV*, 1998, pp. 492–502 [Online]. Available: citeseer.ist.psu.edu/roy98maximumflow.html, [Online]. Available
[188] G. Vogiatzis, P. Torr, and R. Cippola, "Multi-view stereo via volumetric graph-cuts," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. (CVPR-2005)*, 2005, pp. 391–399 [Online]. Available: citeseer.ist.psu.edu/vogiatzis05multiview.html, [Online]. Available
[189] S. Sinha and M. Pollefeys, "Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation," in *ICCV*, 2005, pp. 349–356.
[190] Y. Furukawa and J. Ponce, High-fidelity image-based modeling Univ. of Illinois at Urbana-Champaign, Tech. Rep. 2006-02, 2006.
[191] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," in *ECCV (3)*, 2002, pp. 82–96 [Online]. Available: citeseer.ist.psu.edu/kolmogorov02multicamera.html
[192] C. H. Esteban and F. Schmitt, "Silhouette and stereo fusion for 3-D object modeling," in *CVIU, 96(3)*, 2004, pp. 367–392 [Online]. Available: citeseer.ist.psu.edu/706732.html
[193] J. Pons, R. Keriven, and O. Faugeras, "Modelling dynamic scenes by registering multi-view image sequences," in *Proc. IEEE Conf. in Comput. Vis. and Pattern Recognit.*, 2005, pp. 822–827 [Online]. Available: citeseer.ist.psu.edu/pons05modelling.html
[194] P. Fua and Y. G. Leclerc, "Object-centered surface reconstruction: Combining multi-image stereo and shading," *Int. J. Comput. Vis.*, vol. 16, no. 1, pp. 35–55, 1995.
[195] T. Yu, N. Xu, and N. Ahuja, "Shape and view independent reflectance map from multiple views," *Int. J. Comput. Vis.*, vol. 73, no. 2, pp. 123–138, June 2007.
[196] R. Szeliski, "A multi-view approach to motion and stereo," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit.*, 1999, vol. 1, pp. 157–163.
[197] P. Gargallo and P. Sturm, "Bayesian 3-D modeling from images using multiple depth maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,*, Jun. 2005, vol. 2, pp. 885–891 [Online]. Available: http://perception.inrialpes.fr/Publications/2005/GS05, [Online]. Available
[198] P. Narayanan, P. Rander, and T. Kanade, "Constructing virtual worlds using dense stereo," in *Proc. 6th IEEE Int. Conf. Comput.Vis. (ICCV'98)*, Jan. 1998, pp. 3–10.
[199] M. Goesele, B. Curless, and S. Seitz, "Multi-view stereo revisited," in *Proc. 2006 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Washington, DC, 2006, pp. 2402–2409.
[200] O. D. Faugeras, E. L. Bras-Mehlman, and J. D. Boissonnat, "Representing stereo data with the delaunay triangulation," *Artif. Intell.*, vol. 44, no. 1–2, pp. 41–87, 1990.
[201] A. Manessis, A. Hilton, P. Palmer, P. McLauchlan, and X. Shen, "Reconstruction of scene models from sparse 3-D structure," in *Proc. 2000 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2000, vol. 1, pp. 666–673.
[202] D. D. Morris and T. Kanade, "Image-consistent surface triangulation," in *Comput. Vis. and Pattern Recognit. (CVPR 2000)*, June 2000, vol. 1, pp. 332–338, IEEE Comput. Soc..
[203] C. J. Taylor, "Surface reconstruction from feature based stereo," in *ICCV'03: Proc. Ninth IEEE Int. Conf. Comput. Vis.*, Washington, DC, 2003, pp. 184–190.
[204] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 2, pp. 150–162, Feb. 1994.
[205] W. Matusik, C. Buehler, and L. McMillan, "Polyhedral visual hulls for real-time rendering," in *Proc. EGSR*, 2001, pp. 116–126.
[206] S. Moezzi, L.-C. Tai, and P. Gerard, "Virtual view generation for 3-D digital video," *IEEE Multimedia*, vol. 4, no. 1, pp. 18–26, Jan.–Mar. 1997.
[207] W. Matusik, C. Buehler, R. Raskar, S. Gortler, and L. McMillan, "Image-based visual hulls," in *Proc. SIGGRAPH'00*, pp. 369–374, ACM, 2000.
[208] S. Wuermlin, E. Lamboray, O. Staadt, and M. Gross, "3d video recorder," in *Proc. IEEE Pacific Graphics*, 2002, pp. 325–334.
[209] S. Vedula, S. Baker, and T. Kanade, "Spatio-temporal view interpolation," in *Proc. EGSR*, 2002, pp. 65–75.
[210] M. Li, H. Schirmacher, M. Magnor, and H.-P. Seidel, "Combining stereo and visual hull information for on-line reconstruction and rendering of dynamic scenes," in *Proc. IEEE MMSP*, 2002, pp. 9–12.
[211] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. ACM SIGGRAPH'96*, 1996, pp. 31–42.
[212] T. Fujii, T. Kimoto, and M. Tanimoto, "A new flexible acquisition system of ray-space data for arbitrary objects.," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 2, pp. 218–224, Feb. 2000.
[213] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen, "Unstructured lumigraph rendering," in *Proc. SIGGRAPH'01*, 2001, pp. 425–432, ACM.
[214] W. Matusik and H. Pfister, "3-D TV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes," *Proc. SIGGRAPH'04*, vol. 23, no. 3, pp. 814–824, 2004.
[215] P. Einarsson, C.-F. Chabert, A. Jones, W.-C. Ma, B. Lamond, T. Hawkins, M. Bolas, S. Sylwan, and P. Debevec, "Relighting human locomotion with flowed reflectance fields," *Rendering Techniques*, pp. 183–194, 2006.
[216] T. Kanade, P. Rander, and P. J. Narayanan, "Virtualized reality: Constructing virtual worlds from real scenes," *IEEE Multimedia*, vol. 4, no. 1, pp. 34–47, 1997.
[217] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, 2005.
[218] C. L. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *Proc. SIGGRAPH'04*, vol. 23, no. 3, pp. 600–608, 2004.
[219] M. Waschbüsch, S. Würmlin, D. Cotting, F. Sadlo, and M. Gross, "Scalable 3-D video of dynamic scenes," *The Visual Comput. (Special Issues on Pacific Graphics 2005)*, vol. 21, no. 8–10, pp. 629–638, 2005.
[220] J. Carranza, C. Theobalt, M. Magnor, and H.-P. Seidel, "Free-viewpoint video of human actors," *ACM Trans. Graphics (Proc. SIGGRAPH'03)*, vol. 22, no. 3, pp. 569–577, Jul. 2003.
[221] C. Theobalt, N. Ahmed, H. Lensch, M. Magnor, and H.-P. Seidel, "Seeing people in different light—Joint shape, motion and reflectance capture," *IEEE Trans. Visual. Comput. Graph.*, to be published.
[222] U. Schnars and W. Jueptner, "Direct recording of holograms by a CCD target and numerical reconstructions," *Appl. Opt.*, vol. 33, no. 2, pp. 179–181, 1994.
[223] U. Schnars and W. Jueptner, "Digital recording and reconstruction of holograms in hologram interferometry and shearography," *Appl. Opt.*, vol. 33, no. 20, pp. 4373–4377, 1994.
[224] U. Schnars, T. Kreis, and W. Jueptner, "Digital recording and reconstruction of holograms: Reduction of the spatial frequency spectrum," *Opt. Eng.*, vol. 35, no. 4, pp. 977–982, 1996.
[225] U. Schnars, "Direct phase determination in hologram interferometry with use of digitally recorded holograms," *J. Opt. Soc. Amer. A*, vol. 11, pp. 2011–2015, 1994.

[226] I. Yamaguchi and T. Zhang, "Phase-shifting digital holography," *Opt. Lett.*, vol. 22, pp. 1268–1270, 1997.
[227] Y. Takaki and H. Ohzu, "Fast numerical reconstruction technique for high-resolution hybrid holographic microscopy," *Appl. Opt.*, vol. 38, no. 11, pp. 2204–2211, 1999.
[228] M. Jacquot, P. Sandoz, and G. Tribillion, "High resolution digital holography," *Opt. Commun.*, vol. 190, pp. 87–94, 2001.
[229] S. Zhang, "Application of super-resolution image reconstruction to digital holography," *EURASIP J. Appl. Signal Process.*, pp. 1–7, 2006.
[230] I. Yamaguchi, T. Matsumura, and J. Kato, "Phase-shifting color digital holography," *Opt. Lett.*, vol. 27, pp. 1108–1110, 2002.
[231] J. Kato, I. Yamaguchi, and T. Matsumura, "Multicolor digital holography with an achromatic phase shifter," *Opt. Lett.*, vol. 27, pp. 1402–1405, 2002.
[232] D. Alfieri, G. Coppola, S. D. Nicola, P. Ferraro, A. Finizio, G. Pierattini, and B. Javidi, "Method for superposing reconstructed images from digital holograms of the same object recorded at different distance and wavelength," *Opt. Commun.*, vol. 260, pp. 113–116, 2006.
[233] J. Salvi, J. Pages, and J. Batlle, "Pattern codification strategies in structured light systems," *Pattern Recognit.*, vol. 37, no. 4, pp. 827–849, 2004.
[234] P. Vuylsteke and A. Oosterlinck, "Range image acquisition with a single binary-encoded light pattern," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 2, pp. 148–164, Feb. 1990.
[235] P. M. Griffin, L. S. Narasimhan, and S. R. Yee, "Generation of uniquely encoded light patterns for range data acquisition," *Pattern Recogn.*, vol. 25, no. 6, pp. 609–616, 1992.
[236] K. L. Boyer and A. C. Kak, "Color-encoded structured light for rapid active ranging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 1, pp. 14–28, Jan. 1987.
[237] R. A. Morano, C. Ozturk, R. Conn, S. Dubin, S. Zietz, and J. Nissano, "Structured light using pseudorandom codes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 322–327, Mar. 1998.
[238] J. Tajima, "Rainbow range finder principle for range data acquisition," in *IEEE Workshop on Ind. Appl. of Mach. Vis. and Mach. Intell.*, 1987, vol. 1, pp. 381–386.
[239] Z. J. Geng, "Rainbow three-dimensional camera: New concept of high-speed three dimensional vision systems," *Opt. Eng.*, vol. 35, no. 2, pp. 376–383, 1996.
[240] F. Forster, "Real-Time Range Imaging for Human-Machine Interfaces," Ph.D. dissertation, Tech.Univ. München, Munich, Germany, 2005.
[241] F. Forster, M. Lang, and B. Radig, "Real-time range imaging for dynamic scenes using colour-edge based structured light," in *Proc. Int. Conf. Pattern Recognit.*, 2002, vol. 3, pp. 645–628.
[242] C. Boehnen and P. Flynn, "Accuracy of 3-D scanning technologies in a face scanning context," in *Proc. 5th Int. Conf. 3-DDigital Imaging and Modeling*, 2005, pp. 310–317.
[243] J. Li, L. Hassebrook, and C. Guan, "Optimized two-frequency phase-measuring profilometry light-sensor temporal-noise sensitivity," *J. Opt. Soc. Amer. A*, vol. 20, pp. 106–115, 2003.
[244] A. Baldi, F. Bertolino, and F. Ginesu, "On the performance of some unwrapping algorithms," *Opt. Lasers Eng.*, vol. 37, no. 4, pp. 313–330, 2002.
[245] Q. Hu and K. Harding, "Conversion from phase map to coordinate: Comparison among spatial carrier, Fourier transform, and phase shifting methods," *Opt. Lasers Eng.*, vol. 45, pp. 342–348, 2007.
[246] F. Berryman, P. Pynsent, and J. Cubillo, "A theoretical comparison of three fringe analysis methods for determining the three-dimensional shape of an object in the presence of noise," *Opt. Lasers Eng.*, vol. 39, no. 1, pp. 35–50, 2003.
[247] C. Quan *et al.*, "Shape measurement by use of liquid-crystal display fringe projection with two-step phase shifting," *Appl. Opt.*, vol. 42, pp. 2329–2335, 2003.
[248] P. Huang, C. Zhang, and F.-P. Chiang, "High-speed 3-D shape measurement based on digital fringe projection," *Opt. Eng.*, vol. 42, pp. 163–168, 2003.
[249] J. Harizanova and A. Kolev, "Comperative study of fringes generation in two-spacing phase-shifting profilometry," in *Proc. SPIE*, 2005, vol. 6252, pp. 625221–625225.
[250] D. Mehta *et al.*, "Simple multifrequency and phase-shifting fringe-projection system based on two-wavelength lateral shearing interferometry for three-dimensional profilometry," *Appl. Opt.*, vol. 44, pp. 7515–7521, 2005.
[251] B. Dorrio and J. Fernandez, "Phase-evaluation methods in whole-field optical measurement techniques," *Meas. Sci. and Techn.*, vol. 10, pp. R33–R55, 1999.
[252] T. Kreis, *Handbook of Holographic Interferometry*. Weinheim: Wiley-VCH GmbH, 2004.
[253] K. Creath, "Phase-measurement interferometry techniques," *Prog. Opt.*, vol. 26, pp. 349–393, 1988.
[254] A. Patil, R. Langoju, and P. Rastogi, "An integral approach to phase shifting interferometry using a super-resolution, frequency estimation method," *Opt. Expr.*, vol. 12, no. 20, pp. 4681–4697, 2004.
[255] C. Morgan, "Least-squares estimation in phase-measurement interferometry," *Opt. Lett.*, vol. 7, pp. 368–370, 1982.
[256] A. Patil and P. Rastogi, "Approaches in generalized phase shifting interferometry," *Opt. Lasers Eng.*, vol. 43, pp. 475–490, 2005.
[257] L. Z. Cai, Q. Liu, and X. L. Yang, "Generalized phase-shifting interferometry with arbitrary unknown phase steps for diffraction objects," *Opt. Lett.*, vol. 29, pp. 183–185, 2004.
[258] Y. Surrel, "Design of algorithms for phase measurements by the use of phase stepping," *Appl. Opt.*, vol. 35, no. 1, pp. 51–60, 1996.
[259] A. Patil, P. Rastogi, and B. Raphael, "Phase-stepping interferometry by a covariance-based method," *Appl. Opt.*, vol. 44, pp. 5778–5785, 2005.
[260] M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement 3-D object shapes," *Appl. Opt.*, vol. 22, pp. 3977–3982, 1983.
[261] T. Kreis, "Digital holographic interference phase measurement using the fourier-transform method," *J. Opt. Soc. Amer A.*, vol. 3, pp. 847–856, 1986.
[262] X. Su and W. Chen, "Fourier transform profilometry: A review," *Opt. Lasers Eng.*, vol. 35, pp. 263–284, 2001.
[263] K. G. Larkin, D. Bone, and M. Oldfield, "Natural demodulation of two-dimensional fringe patterns. i. general background of the spiral phase quadrature transform," *J. Opt. Soc. Amer. A.*, vol. 18, no. 8, pp. 1862–1870, 2001.
[264] K. Larkin, D. Bone, and M. Oldfield, "Natural demodulation of two-dimensional fringe patterns. ii. stationary phase analysis of the spiral phase quadrature transform," *J. Opt. Soc. Amer A.*, vol. 18, no. 8, pp. 1871–1881, 2001.
[265] C. Sciammarella and T. Kim, "Determination of strains from fringe patterns using space-frequency representations," *Opt. Eng.*, vol. 42, pp. 3182–3193, 2003.
[266] A. Dursun, S. Ozder, and F. Ecevit, "Continuous wavelet transform analysis of projected fringe patterns," *Meas Sci. Techn.*, vol. 15, pp. 1768–1772, 2004.
[267] L. Watkins, "Phase recovery from fringe patterns using the continuous wavelet transform," *Opt. Lasers Eng.*, vol. 45, no. 2, pp. 298–303, 2007.
[268] J. Zhou, "Wavelet-aided spatial carrier fringe pattern analysis for 3-D shape measurement," *Opt. Eng.*, vol. 44, p. 113602, 2005.
[269] K. Qian, "Windowed Fourier transform for fringe pattern analysis," *Appl. Opt.*, vol. 43, pp. 2695–2702, 2004.
[270] J. L. Marroquin *et al.*, "Regularization methods for processing fringe-pattern images," *Appl. Opt.*, vol. 38, no. 5, pp. 788–794, 1999.
[271] M. Servin, J. Marroquin, and F. Cuevas, "Fringe-follower regularized phase tracker for demodulation of closed-fringe interferograms," *Journ. Opt. Soc. Amer A.*, vol. 18, pp. 689–695, 2001.
[272] M. Gdeisat, D. Burton, and M. Lalor, "Fringe pattern demodulation with a two-frame digital phase-locked loop algorithm," *Appl. Opt.*, vol. 41, pp. 5471–5578, 2002.
[273] J. Marroquin, M. Servin, and R. Rodriguez-Vera, "Adaptive quadrature filters and the recovery of phase from fringe pattern images," *J. Opt. Soc. Amer. A.*, vol. 14, pp. 1742–1753, 1997.
[274] M. Servin, J. Quiroga, and J. Marroquin, "General n-dimensional quadrature transform and its application to interferogram demodulation," *J. Opt. Soc. Amer. A.*, vol. 20, pp. 925–934, 2003.
[275] C. Quan *et al.*, "Phase extraction from a single fringe pattern based on guidance of an extreme map," *Appl. Opt.*, vol. 44, pp. 4814–4821, 2005.
[276] M. Liebling, T. Blu, and M. Unser, "Complex-wave retrieval from a single off-axis hologram," *J. Opt. Soc. Amer. A.*, vol. 21, pp. 367–377, 2004.
[277] M. de Angelis *et al.*, "Profile measurement of a one-dimensional phase boundary sample using a single shot phase-step method," *Opt. Lasers Eng.*, vol. 43, pp. 1305–1314, 2005.
[278] O. A. Skydan, M. J. Lalor, and D. Burton, "Using coloured structured light in 3-D surface measurement," *Opt. Lasers Eng.*, vol. 43, pp. 801–814, 2005.
[279] S. Yoneyama, Y. Morimoto, M. Fujigaki, and Y. Ikeda, "Three dimensional surface profile measurement of moving object by a spatial-offset phase stepping method," *Opt. Eng.*, vol. 42, pp. 137–142, 2003.
[280] S. Zhang and S.-T. Yau, "High-resolution, real-time 3-D absolute coordinate measurement based on a phase-stepping method," *Opt. Exp.*, vol. 14, pp. 2644–2654, 2006.
[281] V. Sainov, E. Stoykova, and J. Harizanova, "Real time phase stepping pattern projection profilometry," *Proc. SPIE*, vol. 6341, p. 63410P-1/6, 2006.

[282] J. Harizanova and V. Sainov, "Three-dimensional profilometry by symmetrical fringes projection technique," *Opt. Lasers Eng.*, vol. 44, pp. 1270–1282, 2006.
[283] C. Reich, R. Ritter, and J. Thesing, "3-D shape measurement of complex objects by combining photogrammetry and fringe projection," *Opt. Eng.*, vol. 39, no. 1, pp. 224–231, 2000.
[284] R. Kowarschik *et al.*, "Adaptive optical three-dimensional measurement with structured light," *Opt. Eng.*, vol. 39, no. 1, pp. 150–158, 2000.
[285] M. Hebert and E. Krotkov, "3-D measurements from imaging laser radars: How good are they?," *Int. J. Image and Vis. Computing*, vol. 10, pp. 170–178, 1992.
[286] B. Stann, A. Abou-Auf, S. Frankel, M. Giza, W. Potter, W. C. Ruff, P. Shen, D. Simon, M. Stead, Z. Sztankay, and L. Lester, "Research progress on scannerless ladar systems using a laser diode transmitter and FM/CW radar principles," *SPIE Laser Radar Technol. Appl.*, vol. 4377, pp. 12–22, 2001.
[287] S. Gokturk, H. Yalcin, and C. Bamji, "A time-of-flight depth sensor system description, issues and solutions," in *IEEE Comput. Vis. Pattern Recognit. Workshop*, 2004 [Online]. Available: http://www.canesta.com/technicalpapers.htm
[288] R. Miyagawa and T. Kanade, "CCD-based range-finding sensor," *IEEE Trans. Electron Devices*, vol. 44, no. 10, pp. 1648–1652, 1997.
[289] 3DV Systems web pages. [Online]. Available: http://www.3dvsystems.com
[290] R. Jeremias, W. Brockherde, G. Doemens, B. Hosticka, L. Listl, and P. Mengel, "A CMOS photosensor array for 3-D imaging using pulsed lasers," in *Proc. IEEE Int. Solid-State Circuit Conf.*, 2001, vol. 1, pp. 252–253.
[291] C. Bamji and E. Charbon, "Systems for CMOS-compatible three-dimensional image sensing using quantum efficiency modulation," U.S. Patent 6 580 496, Jun. 17, 2003.
[292] T. Ushinaga, I. Halin, T. Sawada, S. Kawahito, M. Homma, and Y. Maeda, "A QVGA-size CMOS time of-flight range image sensor with background light charge draining structure," *Proc. SPIE*, vol. 6056, Jan. 2006.
[293] T. Kahlmann and H. Ingensand, "Calibration and improvements of the high-resolution range-imaging camera Swiss Ranger TM," *SPIE*, vol. 5665, pp. 144–155, 2005.
[294] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE J. Quantum Electron.*, vol. 37, no. 3, pp. 390–397, Mar. 2001.
[295] T. Oggier, M. Lehmann, R. Kaufmann, M. Schweizer, M. Richter, P. Metzler, G. Lang, F. Lustenberger, and N. Blanc, "An all-solid-state optical range camera for 3-D real-time imaging with sub-centimeter depth resolution (swissrangertm)," *SPIE*, vol. 5249, pp. 534–545, 2004.
[296] T. Oggier, B. Bottgen, F. Lustenberger, G. Becker, B. Riegg, and A. Hodac, "Swissranger SR3000 and first experiences based on miniaturized 3D-TOF cameras," *First Range Imaging Res. Day at ETH Zurich*, pp. 534–545, 2005.

Elena Stoykova received the Ph.D. degrees from Bulgarian Academy of Sciences (BAS), Sofia, Bulgaria, in 1988, in quantum electronics.

She was a post-doctoral research associate at DLR—German Aerospace Center in 1993 and a visiting researcher at the University of North Paris between 1997 and 1999. In 1998 she received a position of a Senior Scientist in the Institute of Electronics—BAS. In 2002, she joined as a Scientific secretary the Central Laboratory of Optical Storage and Processing of information to the BAS. For more than ten years she was a lecturer in major Universities in Sofia and Plovdiv, Bulgaria. Her current research interests include interferometry, digital holography, Monte-Carlo simulation and error analysis.

A. Aydın Alatan (S'91–M'07) was born in Ankara, Turkey, in 1968. He received the B.S. degree from Middle East Technical University, Ankara, Turkey, in 1990, the M.S and DIC degrees from Imperial College of Science, Medicine and Technology, London, U.K., in 1992, and the Ph.D. degree from Bilkent University, Ankara, Turkey, in 1997, all in electrical engineering.

He was a post-doctoral research associate at Center for Image Processing Research at Rensselaer Polytechnic Institute between 1997 and 1998 and at New Jersey Center for Multimedia Research, New Jersey Institute of Technology, Newark, NJ, between 1998 and 2000. Since August 2000, he joined the Faculty of Electrical and Electronics Engineering at the Department at Middle East Technical University, Ankara, Turkey.

Philip Benzie (M'07) received the B.Eng. degree (hons.) in electrical and electronic engineering and the Ph.D. degree from Aberdeen University, Aberdeen, U.K., in 2001 and 2006, respectively. His Ph.D. dissertation was on the application of finite element analysis to holographic interferometry for nondestructive testing.

He is currently working at Aberdeen University as a Research Fellow and Section Leader for holographic displays in the European Network of Excellence wide 3DTV project. His current research interests include holographic imaging, 3DTV, nondestructive testing, and underwater holography.

Nikos Grammalidis received the B.S. and Ph.D. degrees in electrical and computer engineering from the Aristotle University of Thessaloniki, in 1992 and 2000, respectively.

He is a Researcher in the Informatics and Telematics Institute, Thessaloniki, Greece. Prior to his current position, he was a researcher in 3-D Imaging Laboratory at the Aristotle University of Thessaloniki. His main research interests include image compression, 3-D data processing, multimedia image communication, 3-D motion estimation, stereo and multiview image sequence coding. His involvement with those research areas has led to the co-authoring of more than 15 articles in refereed journals and more than 50 papers in international conferences. Since 1992, he has been involved in more than 10 projects, funded by the EC and the Greek Ministry of Research and Technology.

Sotiris Malassiotis was born in Thessaloniki, Greece, in 1971. He received the B.S. and Ph.D. degrees in electrical engineering from the Aristotle University of Thessaloniki, in 1993 and 1998, respectively.

From 1994 to 1997 he was conducting research in the Information Processing Laboratory of Aristotle University of Thessaloniki. He is currently a senior researcher in the Informatics and Telematics Institute, Thessaloniki. He has participated in several European and National research projects. He is the author of more than 20 articles in refereed journals and more than 30 papers in international conferences. His research interests include range image analysis, pattern recognition, and computer graphics.

Jörn Ostermann (SS'86–M'88–SM'00–F'05) studied electrical engineering and communications engineering at the University of Hannover, Hannover, Germany, and Imperial College London, London, U.K. He received the Dipl.-Ing. and Dr.-Ing. degrees from the University of Hannover in 1988 and 1994, respectively.

From 1988 till 1994, he worked as a Research Assistant at the Institut für Theoretische Nachrichtentechnik conducting research in low bit rate and object-based analysis-synthesis video coding. In 1994 and 1995, he worked in the Visual Communications Research Department, AT&T Bell Laboratories on video coding. He was a member of Image Processing and Technology Research within AT&T Labs—Research from 1996 to 2003. Since 2003, he has been a Full Professor and Head of the Institut für Informationsverarbeitung , University of Hannover. From 1993 to 1994, he chaired the European COST 211 sim group coordinating research in low bit rate video coding. Within MPEG-4, he organized the evaluation of video tools to start defining the standard. He chaired the Adhoc Group on Coding of Arbitrarily-shaped Objects in MPEG-4 Video. Jörn was a scholar of the German National Foundation. He published more than 50 research papers and book chapters. He is coauthor of a graduate level text book on video communications. He holds 22 patents. His current research interests are video coding and streaming, 3-D modeling, face animation, and computer-human interfaces.

Dr. Ostermann received the AT&T Standards Recognition Award and the ISO award in 1998. He is a member of the IEEE Technical Committee on Multimedia Signal Processing and past chair of the IEEE CAS Visual Signal Processing and Communications (VSPC) Technical Committee. Joern served as a Distinguished Lecturer of the IEEE Circuits and Sytems Society.

Sergej Piekh received the Dipl.-Ing. degree in microelectronics in 1997 from Belarussian National Technical University.

From 1998 to 2001 he did research on semiconductor thin films using spectroscopic ellipsometry for the optoelectronic applications at Belarussian National Technical University. Since 2003 he has been a research assistant at Institut für Informationsverarbeitung at the University of Hannover, Hannover, Germany, conducting research in low bit-rate and object-based analysis-synthesis video coding. His current research interests include image processing, object based video coding, 3-D modeling, face animation and human-computer interface.

Ventseslav Sainov is the scientific leader of the holographic and laser interferometry group, and the Director of the CLOSPI-BAS. He has expertise in the fields of light sensitive materials, holography, holographic and laser interferometry, including shearography, and experience in non-destructive testing, 3-D micro/macro measurements, optical and digital processing of interference patterns.

Christian Theobalt received the M.Sc. degree in artificial intelligence from the University of Edinburgh, Edinburgh, U.K., and the Diploma (M.S.) degree in computer science from Saarland University, Saarbrücken, Germany, in 2000 and 2001 respectively. From 2001 to 2005 he was a researcher and Ph.D. candidate in Hans-Peter Seidel's Computer Graphics Group at MPI Informatik in Saarbrücken.

He is head of the research group *3-D Video and Vision-based Graphics* within the Max-Planck Center for Visual Computing and Communication (Saarbrücken/Stanford). In 2005, he received his Ph.D. (Dr.-Ing.) from Saarland University. His research interests include free-viewpoint and 3-D video, marker-less optical motion capture, 3-D computer vision, and image- and physics-based rendering.

Thangavel Thevar was born in Singapore on August 15, 1964. He received the B.Sc. (hons.) degree in electrical engineering and Ph.D. in laser engineering from the University of Aberdeen, UK., in 1989 and 1993, respectively.

In 1993 he joined Singapore Polytechnic as a lecturer. During this period, he went on a sabbatical to NASA Langley Research Center, working on the development of 2-mm lasers. From 1997 to 2000 he worked as a research scientist for Holographics Inc., New York. He was involved in the development of a novel laser based non-destructive inspection technique for the aircraft industry. In 2000 he joined Light Age Inc., New Jersey as a senior scientist where he worked on the development of Alexandrite lasers for scientific, medical, and military applications. In 2005 joined the University of Aberdeen as a lecturer in the engineering department. His current research includes laser development, holography, and non-destructive testing."

Xenophon Zabulis received the B.A., M.S. and Ph.D. degrees in computer science from the University of Crete, Crete, in 1996, 1998 and 2001, respectively.

He has worked as a Postdoctoral Fellow at the Computer and Information Science Department, at the interdisciplinary General Robotics, Automation, Sensing and Perception laboratory and at the Institute for Research in Cognitive Science, both at the University of Pennsylvania. He is currently a Research Fellow at the Institute of Informatics and Telematics—Centre of Research and Technology Hellas, Greece.