building. Classical statistical analysis once insisted on the use of mathematical models for which compact mathematical solutions could be found. This generated a tendency on the part of analysts to restrict their studies to such tractable models, even at the expense of making them unrealistic. With the computer's capabilities we need not be afraid of formulating more realistic models, thereby freeing scientists from the fetters of analytic tractability. As an example, the rate equations governing the time dependence of chemical reactions are usually assumed to be linear first-order differential equations with constant rate coefficients. These give rise to the well-known analytic solutions of mixtures of exponentials that have often been fitted to data obtained from a study of chemical reaction processes. It is well known to the chemist that a model with constant rates is often an oversimplification. Nonlinear rate equations are often more realistic. However, we usually cannot solve the resulting nonlinear rate equations analytically. By contrast, the computer has no difficulty solving more realistic rate equations by numerical integration and fitting these numerical integrals directly to the data. The parameters in the rate equations then become the unknown parameters in the nonlinear regression fit. The rapidity of numerical integration subroutines is essential for this approach to be feasible.

There are many other instances in which numerical analysis can and will replace analytic solutions. Future research will therefore be able to search more freely for information that is at the disposal of scientists. Indeed, they will use the computer as a powerful tool in trying alternative model theories, all of a complex but realistic form, to advance their theories on empirical phenomena.

REFERENCES

1. Massey, F. J. (1951). *J. Amer. Statist. Ass.*, **46**, 68–78.
2. Meyer, H. A. ed. (1956). *Symposium on Monte Carlo Methods*. Wiley, New York.
3. Pearson, E. S. and Hartley, H. O. (1966). *Biometrika Tables for Statisticians*, Vol. 1. Cambridge University Press, Cambridge. (Table 7, Probability integral of the χ^2-distribution and the cumulative sum of the Poisson distribution).

See also EDITING STATISTICAL DATA; ERROR ANALYSIS; GRAPHICAL REPRESENTATION, COMPUTER AIDED; and GENERATION OF RANDOM VARIABLES, COMPUTER.

H. O. HARTLEY

COMPUTER VISION, STATISTICS IN

WHAT IS COMPUTER VISION?

The general goal of computer vision (also known as image understanding) is to derive information about a scene by computer analysis of images of that scene. Images can be obtained by many types of sensors, such as still or video cameras, infrared, laser radar, synthetic aperture radar, millimeter wave radar, etc. Obtaining a description of a scene from one or more images of it can be useful in applications like automatic navigation, virtual reality scene modeling, object tracking, detection and recognition, etc.

Animals and humans have impressive abilities to interact with their environments using vision. This performance constitutes a challenge to vision researchers; at the same time, it serves as an existence proof that the goals of computer vision are achievable. Conversely, the algorithms used by vision systems to derive information about a scene from images can be regarded as possible computational models for the processes employed by biological visual systems. However, constructing such models is not the primary goal of CV; it is concerned only with the correctness of its scene description algorithms, and not whether they resemble biological visual processes.

Computer vision techniques have numerous practical applications, some of them being character recognition, industrial inspection, medical image analysis, remote sensing, target recognition, robot navigation, scene modeling, surveillance, human identification, activity analysis, etc. There have been many successful applications, but many other tasks

are beyond current capabilities, thus providing major incentives for continued research in this area. Since the goal of computer vision is to derive descriptions of a scene from images or videos of that scene, it can be regarded as the inverse of computer graphics, in which the goal is to generate realistic images of a scene, given a description of the scene. The goal of CV is more difficult because it involves the solution of inverse problems that are highly under-constrained ("ill-posed"), not amenable to precise mathematical descriptions, and often computationally intractable. Solutions to these problems have been obtained using a combination of techniques drawn from statistics, physics, applied mathematics, signal and image processing, neural networks, psychophysics, biology, and artificial intelligence.

STATISTICS AND COMPUTER VISION

Computer vision presents numerous challenging problems at the sensor, data and algorithm levels. Traditionally, problems in CV have been grouped into three areas that have vaguely defined boundaries. At the so-called low level, the goal is to extract features such as edges, corners, lines, and segmented regions, track features over a sequence of frames or to compute optical flow. At the intermediate level, using the output of the low-level modules, one is interested in grouping of features, in estimation of depth using stereopsis, and in motion and structure estimation. At the high level, the intermediate-level outputs are combined with available knowledge about the scene, objects and tasks so that descriptions of objects can be derived. Thus, CV can be described as a geometric inference problem since it aims to obtain an understanding of the 3D world that we live in from 2D images of it.

The input to vision algorithms at the low level is the data obtained from one or more sensors, which is usually corrupted by noise from the sensor or the environment. For example, poor lighting conditions can lead to erroneous results in feature extraction or optical flow computation. Similarly, tracking features or objects in dense visual clutter is a challenging problem. In many of these problems, statistical methods can play very important roles in understanding and modeling the noise processes in order to obtain "optimal" signal estimates and symbolic inferences. Some of the problems which can provide challenges to statisticians are: a) Analysis of non-Gaussian models, b) Object tracking and recognition in cluttered environments, c) Non-stationary image processing, d) Evaluation and performance characterization of algorithms, e) Multi-sensor fusion, f) Robust inference of structure, g) Content analysis in video sequences.

Numerous statistical tools have been applied to computer vision problems with varying degrees of success. One of the most influential models applied to problems in image processing, analysis and understanding is Markov Random Fields (MRFs) [1]. It has led to more meaningful representations that include discontinuities such as edges, lines, etc. An MRF consists of a probability distribution over a set of variables $\{f_i\}$ such that the probability of a specific variable f_i depends only on the states of its neighbors. More precisely, we can define a neighborhood $\mathcal{N}_i$ such that $P(f_i|f_j, j \in \mathcal{N}_i) = P(f_i|f_j, \forall j)$. The relation between MRFs and statistical physics through the Gibbs distribution has led to several interesting optimization algorithms such as simulated annealing [12]. Geman and Geman formulated the image segmentation problem in terms of MRFs in order to smooth images except at places where the image values change rapidly [7].

Tracking an object over a sequence of video frames is another area where ideas from statistics have been applied. Many motion estimation problems have been formulated as posterior state estimation problems, i.e. estimating the position of an object given a set of observation images. They have been typically solved using a Kalman filter or extended Kalman filter [14]. However, the Kalman filter is an optimal estimator in the mean square sense only among the class of linear estimators for a general statistical distribution. For a Gaussian distribution, it is the minimum mean square error estimator. In situations where the state and observation equations are non-linear, the extended Kalman filter has been used. It

uses a linearization of the state equations and the observation equations about the current best estimate of the state to produce "approximate" minimum mean-square estimates of the state. In many tracking applications (which we will discuss later in detail) the observation process is highly non-linear, or even non-analytical. A generalization of the Kalman filter to the non linear case exists based on the Zakai equation [24]. It has been applied to object detection in [8] and to object tracking in [16]. The problem of tracking in visual clutter was addressed in [10] by estimating and propagating the posterior state density from image data using sampling techniques, and was extended in [13] to simultaneous tracking and verification using sequential importance sampling (SIS) [14].

One of the most challenging problems to vision researchers is estimating the 3D structure of a scene from a sequence of images of the scene obtained by a moving camera. This is known as the structure from motion (SfM) problem and has been at the forefront of vision research for over two decades [5,9]. SfM is solved by estimating the scene structure from a set of tracked feature points or from optical flow, both of which can be computed from the sequence of video frames. One of the challenges to solving this problem is inability to understand the errors in estimating the motion between pairs of images and the effect of these errors on structure estimation. Robust solutions to this problem require an understanding of not only the geometrical relationships of the 3D scene to its 2D projections on the image plane, but also the statistical characteristics of the image data [11].

Recently, various robust statistical methods have been applied to computer vision problems. Notable among them are bootstrapping techniques [4] for performance evaluation and the mean shift procedure for analyzing feature spaces [15]. It is not possible to discuss here all the statistical techniques that have been applied to vision problems. We will concentrate on two problems, namely tracking and structure from motion, in order to highlight the importance of statistics to computer vision.

STOCHASTIC FILTERING FOR TRACKING

Conditional Density Propagation

Tracking outlines and features of objects as they move in densely cluttered environments is a challenging problem. This is because elements of the background clutter may mimic features of foreground objects. One of the best-known approaches to this problem is to resolve the ambiguity by applying probabilistic models of object shape and motion to analyze the video stream. *Prior* probability densities can be defined over the curves represented by appropriate parameter vectors $\mathbf{x}$, and also over their motions. Given these priors, and an *observation density* characterizing the statistical variability of the image data $\mathbf{z}$ given a contour state $\mathbf{x}$, a posterior distribution can be estimated for $\mathbf{x}_t$, given $\mathbf{z}_t$ at successive times t. This problem has been studied with thoroughly using Kalman filtering in a relatively clutter-free case [9]. In the presence of clutter, however, there are usually competing observations which tend to encourage a multi-modal, and hence non-Gaussian, density for $\mathbf{x}_t$. If the Kalman filter is applied to this case, it will give an estimator which is optimal only within the class of linear estimators. Besides, the state and observation equations are rarely linear in practice. A well-known probabilistic algorithm for solving this problem is CONDENSATION, which is an acronym for Conditional Density Propagation [10].

Suppose that the state of the modeled object at time t is denoted by $\mathbf{x}_t$ and its history by $\mathcal{X}_t = \{x_1, \ldots, x_t\}$. Similarly, let the set of image features at time t be $\mathbf{z}_t$ with history $\mathcal{Z}_t = \{z_1, \ldots, z_t\}$. No functional assumptions are made about the densities or about the relation between the observation and state vectors. It is assumed that the object dynamics follows a temporal Markov chain such that

$$p(\mathbf{x}_t|\mathcal{X}_{t-1}) = p(\mathbf{x}_t|\mathbf{x}_{t-1}), \tag{1}$$

i.e. the new state depends only the immediately preceding state, independent of the earlier history. Observations $\mathbf{z}_t$ are assumed to be independent, both mutually and with

respect to the dynamic process. This is expressed mathematically as

$$p(\mathcal{Z}_{t-1}, \mathbf{x}_t|\mathcal{X}_{t-1}) = p(\mathbf{x}_t|\mathcal{X}_{t-1})\Pi_{i=1}^{t-1} p(\mathbf{z}_i|\mathbf{x}_i), \tag{2}$$

which leads to

$$p(\mathcal{Z}_t|\mathcal{X}_t) = \Pi_{i=1}^{t} p(\mathbf{z}_i|\mathbf{x}_i). \tag{3}$$

The observation process is therefore defined by specifying the conditional density at each time t.

The problem of analyzing the dynamic system (in this case, solving the tracking problem) can be formulated as evaluation of the conditional density $p(\mathbf{x_t}|\mathcal{Z}_t)$. In [10], the following rule for propagating the conditional density was proved:

$$p(\mathbf{x_t}|\mathcal{Z}_t) = k_t p(\mathbf{z}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathcal{Z}_{t-1}), \tag{4}$$

where

$$p(\mathbf{x_t}|\mathcal{Z}_{t-1}) = \int_{\mathbf{x}_t} p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|\mathcal{Z}_{t-1}), \tag{5}$$

and k_t is a normalization constant that does not depend on $\mathbf{x}_t$.

In [13], the authors proposed a similar method using Sequential Importance Sampling (SIS) [14] for estimating the conditional density $p(\mathbf{x_t}|\mathcal{Z}_t)$. The SIS method is a recently proposed technique for approximating the posterior distribution of the state parameters of a dynamic system which is described by observation and state equations. The authors showed that the tracking and verification problems could be solved simultaneously. The visual tracking problem was solved through probability density propagation, and verification was realized through hypothesis testing using the estimated posterior density.

The method of propagating the conditional density using SIS works as follows. If the measurement is denoted by $\mathbf{z}_t$ and the state parameter by $\mathbf{x}_t$, the observation equation essentially provides the conditional distribution of the observation given the state, $f_t(\mathbf{z}_t|\mathbf{x}_t)$. Similarly, the state equation gives the Markov transition distribution from time t to time $t+1$, $q_t(\mathbf{x}_{t+1}|\mathbf{x}_t)$. The goal is to find the posterior distribution of the states $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_t)$ given all the available observations up to t, $\pi_t(\mathcal{X}_t) = P(\mathcal{X}_t|\mathcal{Z}_t)$, where $\mathcal{X}_t = \{\mathbf{x}_i\}_{i=1}^t$ and $\mathcal{Z}_t = \{\mathbf{z}_i\}_{i=1}^t$. One way to represent the approximation of the posterior distribution is by a set of samples and their corresponding weights.

Definition. [14] *A random variable X drawn from a distribution g is said to be* ***properly weighted*** *by a weighting function $w(X)$ with respect to the distribution π if for any integrable function h,*

$$E_g h(X)w(X) = E_\pi h(X).$$

A set of random draws and weights $(x^{(j)}, w^{(j)}), j = 1, 2, \cdots$, is said to be properly weighted with respect to π if

$$\lim_{m\to\infty} \frac{\sum_{j=1}^m h(x^{(j)})w^{(j)}}{\sum_{j=1}^m w^{(j)}} = E_\pi h(X)$$

for any integrable function h.

Suppose $\{\mathcal{X}_t^{(j)}\}_{j=1}^m$ is a set of random samples properly weighted by the set of weights $\{w_t^{(j)}\}_{j=1}^m$ with respect to π_t and let g_{t+1} be a trial distribution. Then the recursive SIS procedure for obtaining the random samples and weights properly weighting π_{t+1} is as follows:

SIS steps: for $j = 1, \cdots, m$,

(A) Draw $X_{t+1} = \mathbf{x}_{t+1}^{(j)}$ from $g_{t+1}(\mathbf{x}_{t+1}|\mathcal{X}_t^{(j)})$. Attach $\mathbf{x}_{t+1}^{(j)}$ to form $\mathcal{X}_{t+1}^{(j)} = (\mathcal{X}_t^{(j)}, \mathbf{x}_{t+1}^{(j)})$.

(B) Compute the "incremental weight" u_{t+1} by

$$u_{t+1}^{(j)} = \frac{\pi_{t+1}(\mathcal{X}_{t+1}^{(j)})}{\pi_t(\mathcal{X}_t^{(j)}) g_{t+1}(\mathbf{x}_{t+1}|\mathcal{X}_t^{(j)})}$$

and let $w_{t+1}^{(j)} = u_{t+1}^{(j)} w_t^{(j)}$.

It can be shown [14] that $\{\mathcal{X}_{t+1}^{(j)}, w_{t+1}^{(j)}\}_{j=1}^m$ is properly weighted with respect to π_{t+1}. Hence the above SIS steps can be recursively applied to get a properly weighted set for any future time instant when corresponding observations are available. It is not difficult to show that given a set of properly weighted samples $\{\mathcal{X}_t\}$ with respect to the joint posterior distribution $\pi_t(\mathcal{X}_t)$, the "marginal" samples formed

by the components of $\mathbf{x}_i$ in $\{\mathcal{X}_t\}$ are properly weighted by the same set of weights with respect to the marginal posterior distribution $\pi_t(\mathbf{x}_i)$. Once the properly weighted samples of the joint distribution are obtained, the marginal distributions are approximated by the "marginal" samples weighted by the same set of weights.

Verification

Assume now that there are C classes $\{\omega_1, \ldots, \omega_C\}$ to which the tracked object can belong (e.g., C different people). Then given an observation Z, the Bayesian maximum *a posteriori* (MAP) probability rule chooses $\omega = \max_i P(\omega_i|Z)$, where $P(\omega_i|Z)$ is the posterior probability of the class ω_i given Z and can be computed as

$$P(\omega_i|Z) = \int_A p_i(X|Z)dX, \tag{6}$$

where $p_i(X|Z)$ is the posterior density of class ω_i, A being some properly defined region. Further details can be found in [13].

We now illustrate tracking using SIS. Fig. 1 (left column) shows sample frames of a video sequence in which two persons are moving around; the face templates of these persons are to be verified from the video. In the middle and right columns, the templates are overlapped on the video. For easy visualization, a black block is used for the template corresponding to the face of the man in the white shirt (denoted by M1), and a white block for the template corresponding to the face of the second man (denoted by M2). The middle column illustrates the situation where the algorithm is correctly initialized, meaning that the templates are correctly put on their respective persons. The figures show that tracking is maintained for M1 over the entire sequence, and is able to recover from occlusion for M2 (since the two people switched positions). The right column in Fig. 1 shows a case in which we switch the hypotheses by putting the templates on the wrong persons. We observe that M2 eventually gets dropped into the cluttered background, while M1, after first sticking to the wrong person, is attracted to the right person.

Zakai Equation

Another stochastic filtering approach to the tracking problem is the use of the Zakai

Figure 1. Left column: Sample frames of a sequence. The top row is a frame from the beginning of the sequence, while the bottom row is a frame from the end of the sequence. Middle column: Templates overlaid on the video when the hypotheses are true. Right column: Results when the hypotheses are false.

equation [3,17,24], which can be regarded as a generalization of the Kalman filter to the non-linear case. T. Duncan, R. Mortensen and M. Zakai derived equations that must be solved in order to find the optimal filter (in the same least squares sense as the Kalman filter) which, given a set of not necessarily linear observations, produces the best estimates of the required coordinates. This is possible provided a certain second-order partial differential equation can be solved. For a long time, this remarkable result was mostly of theoretical interest. One of its first applications to image processing and computer vision can be found in [8], where the Zakai equation and wavelets were used to address the problem of tracking an object over a sequence of frames. The smoothness of the wavelets was used in the derivation of the equation describing the evolution of the conditional density giving the filter.

We will now provide a brief outline of the theory of the Zakai equation and its application to the tracking problem. Let X_t be a stochastic process in $\Re^n$ satisfying the stochastic equation

$$dX_t = h(X_t)dt + g(X_t)dB_t, \tag{7}$$

where $h : \Re^n \to \Re^n$ and $g : \Re^n \to \{n \times m$ matrices$\}$ are twice-differentiable functions modeling the state noise structure, and B_t is a Brownian motion in $\Re^m$. If the state vector X_t represents geometric parameters of an object, such as its coordinates, then the tracking problem is solved if we can compute the state updates given information from the observations. We are interested in estimating some statistic ϕ of the states of the form

$$\pi_t(\phi) = E[\phi(X_t|\mathcal{Z}_t)] \tag{8}$$

given the observation history $\mathcal{Z}_t$ up to time t.

In [16], the authors used the Zakai equation for 3D object tracking. They used an approximate shape model of an object for tracking and motion estimation and showed that it is possible to derive a simplified form of the Zakai equation. The branching particle propagation method was used for computing the solution [2]. This demonstrated that it is possible to construct a sequence of branching particle systems U_n which converges to the solution of the Zakai equation p_t, i.e. $\lim_{n\to\infty} U_n(t) = p_t$.

Statistical Methods in Motion Analysis

Error Analysis in Structure from Motion. Reconstructing the 3D structure of a scene from a video sequence has been one of the most prominent areas of research in computer vision and is known as structure from motion. The first step toward solving this problem is to estimate the motion between corresponding points in two frames of the video sequence. If the frames are close enough in time, the motion can be estimated using optical flow [5]. In general, however, determining corresponding points automatically is extremely difficult because of poor image quality, similarities between textures, changes of viewpoint, etc. It is important to understand the effects of the errors which arise and which propagate through the reconstruction process. We will now briefly describe the problem and outline the statistical approaches which have been applied to it.

Consider a coordinate frame attached rigidly to a camera, with origin at the center of perspective projection and z-axis perpendicular to the image plane. Assume that the camera is in motion relative to a stationary scene with translational velocity $V = [v_x, v_y, v_z]$ and rotational velocity $\Omega = [\omega_x, \omega_y, \omega_z]$. We further assume that the camera motion between two consecutive frames of the video sequence is small, and use the small-motion approximation to the perspective projection model for motion field analysis. If $p(x,y)$ and $q(x,y)$ are the horizontal and vertical velocity fields of a point (x,y) in the image plane, they are related to the 3D object motion and scene depth z by [5]

$$\begin{aligned} p(x,y) &= (xv_z - fv_x)/z(x,y) + \frac{1}{f}xy\omega_x \\ &\quad - \left(f + \frac{1}{f}x^2\right)\omega_y + y\omega_z \\ q(x,y) &= (yv_z - fyv_y)/z(x,y) + \left(f + \frac{1}{f}y^2\right)\omega_x \\ &\quad - \frac{1}{f}xy\omega_y - x\omega_z, \end{aligned} \tag{9}$$

where f is the focal length of the camera. Examination of these equations reveals that only the translational component of the image velocity depends on the 3D location of the scene point; the rotational component depends only on the image position (x,y). Also, the image velocity field is invariant under equal scaling of the depth z and the translational velocity vector V; this is known as the scale ambiguity in 3D reconstruction, and shows that we can determine the relative motion and scene structure only up to a scale factor. Since only the direction of the translational motion can be obtained from (9), the equations can be rewritten as

$$\begin{aligned} p(x,y) &= (x - fx_f)h(x,y) + \frac{1}{f}xy\omega_x \\ &\quad - \left(f + \frac{1}{f}x^2\right)\omega_y + y\omega_z \\ q(x,y) &= (y - fy_f)h(x,y) + \left(f + \frac{1}{f}y^2\right)\omega_x \\ &\quad - \frac{1}{f}xy\omega_y - x\omega_z, \end{aligned} \quad (10)$$

where $(x_f, y_f) = (\frac{v_x}{v_z}, \frac{v_y}{v_z})$ is known as the *focus of expansion* (FOE), and $h(x,y) = \frac{v_z}{z(x,y)}$ is the inverse scene depth.

Analysis of these equations shows that errors in estimating the motion $\mathbf{u} = [p_1, q_1, \ldots, p_N, q_N]$ between two corresponding points will affect the results of the 3D reconstruction $\mathbf{z} = [h_1, \ldots, h_N, x_f, y_f, \omega_x, \omega_y, \omega_z]$, where N is the number of points tracked in each image (in the dense case, it is the total number of pixels in the image). It should be noted that the system of equations (10) is non-linear, and the unknown vector $\mathbf{z}$ lies in an extremely high-dimensional space ($(N+5)$-dimensional). Nevertheless, it is possible to derive precise expressions for the error covariance $\mathbf{R_z}$ in $\mathbf{z}$ as a function of the error covariance $\mathbf{R_u}$ in terms of the parameters in (10). Define

$$\begin{aligned} A_{\bar{i}p} &= [-(x_{\bar{i}} - x_f)\mathbf{I}_{\bar{i}}(N) \mid h_{\bar{i}} \;\; 0 \;\; -\mathbf{r}_{\bar{i}}], \\ &= [A_{\bar{i}ph} | A_{\bar{i}pm}], \\ A_{\bar{i}q} &= [-(y_{\bar{i}} - y_f)\mathbf{I}_{\bar{i}}(N) \mid 0 \;\; h_{\bar{i}} \;\; -\mathbf{s}_{\bar{i}}], \\ &= [A_{\bar{i}qh} | A_{\bar{i}qm}] \end{aligned} \quad (11)$$

where $\bar{i} = \lceil i/2 \rceil$ is the upper ceiling of i ($\bar{i}$ then represents the number of feature points N, and $i = 1, \ldots, n = 2N$), $\mathbf{r}_i = [x_i y_i, -(1+x_i^2), y_i]^T$, $\mathbf{s}_i = (1+y_i^2, -x_i y_i, -x_i)^T$, and $\mathbf{I}_n(N)$ denotes a 1 in the n^{th} position of an array of length N that has zeros elsewhere. The subscripts p in $A_{\bar{i}p}$ and q in $A_{\bar{i}q}$ denote the fact that the elements of the respective vectors are derived from the p^{th} and q^{th} components of the motion in (10). In [23], the authors proved that if p and q were corrupted by additive IID white Gaussian noise with variance r^2, i.e. $\mathbf{R_u} = r^2\mathbf{I}_{2N\text{X}2N}$, then

$$\mathbf{R_z} = r^2\mathbf{H}^{-1}, \quad (12)$$

where

$$\mathbf{H} = \sum_{\bar{i}=1}^{N} \left(A_{\bar{i}p}{}^T A_{\bar{i}p} + A_{\bar{i}q}{}^T A_{\bar{i}q} \right). \quad (13)$$

An extension this result has been recently proposed; a more general expression was derived using the implicit function theorem, without the strong assumptions of (12) and (13). In [19, 20] the authors proved that

$$\begin{aligned} \mathbf{R_z} = \mathbf{H}^{-1} \Bigg(\sum_{\bar{i}=1}^{N} (& A_{\bar{i}p}{}^T A_{\bar{i}p} R_{u\bar{i}p} \\ & + A_{\bar{i}q}{}^T A_{\bar{i}q} R_{u\bar{i}q}) \Bigg) \mathbf{H}^{-T}. \end{aligned} \quad (14)$$

The importance of the expressions in (12), (13) and (14) lies in the fact that they provide precise mathematical expressions for the errors in reconstruction in terms of the parameters of the basic equations in (10). These expressions can then be used to obtain robust engineering solutions to the 3D reconstruction problem.

Statistical Bias in Motion Estimates

As mentioned earlier, noise in the image intensities causes errors in the estimation of features such as points, lines, edges, etc. It has recently been proposed that the estimation of these features is biased, which causes them to be perceived incorrectly [6]: the appearance of the pattern is altered, and

this provides a possible explanation for many geometrical optical illusions. For example, consider the estimation of a point $\mathbf{x}$ as an intersection of two straight lines. It is possible to obtain a linear system of equations represented in matrix form by $\mathbf{Ix} = \mathbf{C}$, where $\mathbf{I}$ is a $n \times 2$ matrix of n measurements of image gradients, and $\mathbf{C}$ is an n-dimensional vector. The coordinates of $\mathbf{x}$ can then be obtained by a least squares (LS) solution. It is well known that the LS solution to a linear system of the form $Ax = b$ with errors in the measurement matrix A is biased. In our case, the matrix $\mathbf{I}$ of estimated image gradients will almost always have measurement errors; hence the estimate of the position of the point of intersection will be biased. Under IID noise in the parameters of $\mathbf{I}$, an exact expression for the bias was derived in [6], and through experiments, it was shown that this could be used to explain many of the commonly occurring illusions.

This result about the bias in the estimation of image features can be extended to prove that 3D depth estimates are also biased, and through simulations, it can be shown that the effect of this bias is significant [21]. Consider once again (10). In cases where the FOE (x_f, y_f) is known, it is possible to obtain a linear system of equations for N points. Since many SfM algorithms work by first estimating the camera motion and then the depth, this situation often occurs in practice. Once an over-determined system of linear equations has been obtained, its LS solution introduces bias. In [21], the authors derived an expression for the bias and analyzed the effects of different camera motions on it.

The use of total least squares (TLS) does not help us to avoid this bias, because the TLS estimate is unbiased only if the error in estimating A is equal in variance to the error in estimating b [22], and this would be very difficult to maintain in (10). Also, estimating the bias of a TLS estimate is extremely cumbersome, and the covariance of an unbiased TLS estimate is larger than that of the LS estimate, in first order approximation as well as in simulations. Hence there is no fundamental gain in choosing the TLS over the LS solution.

SIS for SfM

We previously discussed the use of SIS techniques for propagating the posterior density function for tracking applications. The SIS procedure has also been applied to the problem of structure estimation, by formulating it as a state estimation problem [18]. We briefly describe this formulation of the problem, the approach, and some results.

The problem can be formulated as first estimating the camera motion using geometric rigid body constraints like the epipolar constraint [9], and then recovering scene structure using the motion estimates. Two coordinate systems are required to model the motion. One coordinate system, denoted by C, is attached to the camera and uses the center of projection of the camera as its origin. The Z axis of C is along the optical axis of the camera, with the positive half-axis in the looking direction. The X-Y plane of C is perpendicular to the Z axis, with the X and Y axes parallel to the borders of the image plane, and the X-Y-Z axes of C satisfy the right-hand rule. The other coordinate system is a world inertial frame, denoted by I, which is fixed on the ground. Five parameters are employed to describe the motion of the camera:

$$\mathbf{x}_t = (\psi_x, \psi_y, \psi_z, \alpha, \beta)$$

Here (ψ_x, ψ_y, ψ_z) are the rotation angles of the camera about the coordinate axes of the inertial frame I, and (α, β) are the elevation and azimuth angles of the camera translation direction, measured in the world system I.

Given the above motion parameterization, a state space model can be used to describe the behavior of a moving camera:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{n}_x \tag{15}$$

$$\mathbf{y}_t = \mathrm{Proj}(\mathbf{x}_t, \mathcal{S}_t) + \mathbf{n}_y \tag{16}$$

where $\mathbf{x}_t$ is the state vector and $\mathbf{y}_t$ is the observation at time t. Proj$(\cdot)$ denotes the perspective projection, a function of camera motion $\mathbf{x}_t$ and scene structure $\mathcal{S}_t$. n_x denotes the dynamic noise in the system, describing the time-varying property of the state vector. If no prior knowledge about the motion is available, a random walk is a suitable

alternative for modeling the camera position.

Based on this state space model, the authors designed an SIS method for finding an approximation to the posterior distribution of the motion parameters. The method was based on computing the likelihood function $f(\mathbf{y}_t|\mathbf{x}_t)$ by taking advantage of the epipolar constraint. The results of 3D reconstruction from a video sequence of a face is shown in Figure 2. The first image shows one frame of the video sequence and the remaining images show different views of the reconstructed 3D model.

CONCLUSION

The area of research concerned with extracting useful 2D and/or 3D information from one or more images is known as computer vision. It is an interdisciplinary field which draws ideas from mathematics, physics, biology and computer science, among others. The input data to most vision algorithms consists of images, which are corrupted by noise from the sensors or the environment. Statistical concepts have been applied to understand and model the characteristics of this noise. In this article we have reviewed some of the relevant literature on uses of statistics in computer vision, and have discussed in detail two of the most important vision applications, tracking and 3D reconstruction.

REFERENCES

1. Chellappa, R. and Jain, A. K. (1993). *Markov Random Fields: Theory and Applications*. Academic Press.
2. Crisan, D., Gaines, J., and Lyons, T. (1998). Convergence of the branching particle method to the solution of the Zakai equation. *SIAM Journal of Applied Mathematics*, **58**, 1568–590.
3. Duncan, T. E. (1967). *Probability Densities for Diffusion Processes with Applications to Nonlinear Filtering Theory*. PhD Thesis, Stanford University.
4. Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall.
5. Faugeras, O. D. (1993). *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press.

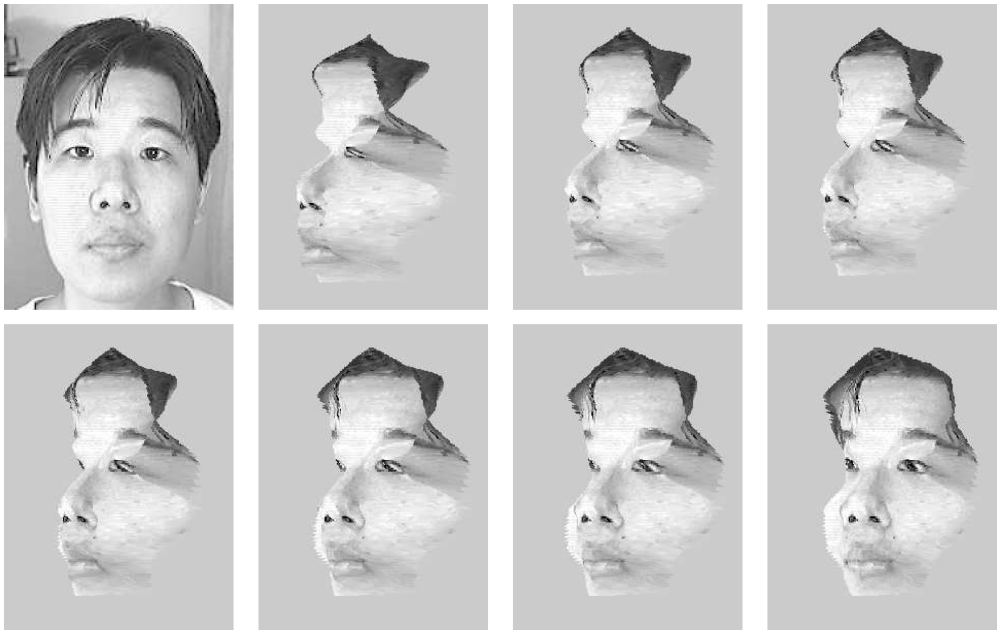

Figure 2. One frame from the original video sequence followed by the reconstructed 3D model viewed from different positions using the SIS procedure.

6. Fermuller, C., Malm, H., and Aloimonos, Y. (May 2001). Statistics explains geometrical optical illusions. Technical report, CS-TR-4251, University of Maryland, College Park.
7. Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **6**, 721–741.
8. Haddad, Z. S. and Simanca, S. R. (1995). Filtering image records using wavelets and the Zakai equation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **17**, 1069–1078.
9. Hartley, R. I. and Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge University Press.
10. Isard, M. and Blake, A. (1998). Conditional density propagation for visual tracking. *International Journal of Computer Vision*, **29**, 5–28.
11. Kanatani, K. (1996). *Statistical Optimization for Geometric Computation: Theory and Practice*. North-Holland.
12. Kirpatrick, S. Gelatt, C. D., Jr., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, **220**, 671–680.
13. Li, B. and Chellappa, R. (2000). Simultaneous tracking and verification via sequential posterior estimation. In *Comp. Vision and Pattern Recognition*, pages II, 110–117.
14. Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.*, **93**, 1032–1044.
15. Meer, P. Stewart, C.V., and Tyler, D. E. (2000). Robust computer vision: An interdisciplinary challenge. *Computer Vision and Image Understanding*, **78**, 1–7.
16. Moon, H., Chellappa, R., and Rosenfeld, A. (2001). 3d object tracking using shape-encoded particle propagation. In *International Conference on Computer Vision*, pages II, 307–314.
17. Mortensen, R. E. (1966). *Optimal Control of Continuous-time Stochastic Systems*. PhD thesis, University of California, Berkely.
18. Qian, G. and Chellappa, R. (2001). Structure from motion using sequential Monte Carlo methods. In *Int. Conf. on Computer Vision*, pages II, 614–621.
19. Chowdhury, A. Roy and Chellappa, R. (October 2003). Stochastic approximation and rate-distortion analysis for robust structure and motion estimation. *International Journal of Computer Vision*, pages 27–53.
20. Chowdhury, A. Roy and Chellappa, R. (July 2004). An information theoretic criterion for evaluating the quality of 3d reconstruction. *IEEE Trans. on Image Processing*, pages 960–973.
21. Chowdhury, A. Roy and Chellappa, R. Statistical bias in 3d reconstruction from a monocular video. *IEEE Trans. on Image Processing*, Accepted.
22. Van Huffel, S. and Vandewalle, J. (1991). *The Total Least Squares Problem*. SIAM Frontiers in Applied Mathematics.
23. Young, G. S. and Chellappa, R. (1992). Statistical analysis of inherent ambiguities in recovering 3-D motion from a noisy flow field. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **14**, 995–1013.
24. Zakai, M. (1982). On the optimal filtering of diffusion processes. *Z.Wahrsch.verw.Geb.*, **11**, 230–243.

FURTHER READING

Doucet, A., de Freitas, N., and Gordon, N. (2000). *Sequential Monte Carlo Methods in Practice*. Springer.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer.

See also COMPUTER-INTENSIVE STATISTICAL METHODS and MONTE CARLO METHODS.

RAMA CHELLAPPA
AMIT K. ROY CHOWDHURY

CONCAVE AND LOG-CONCAVE DISTRIBUTIONS

A real function g defined on the interval (a, b) $(-\infty \leqslant a < b \leqslant \infty)$ is convex if

$$g(\alpha x + (1-\alpha)y) \leqslant \alpha g(x) + (1-\alpha)g(y) \quad (1)$$

whenever $\alpha \in [0,1]$ and $x, y \in (a, b)$ (*see* GEOMETRY IN STATISTICS: CONVEXITY). A function g is concave if $-g$ is convex. A positive valued function g is said to be log-concave if $\log g$ is concave, and log-convex if $\log g$ is