André Redert, Emile Hendriks, and Jan Biemond

Correspondence Estimation in Image Pairs

Designing Estimators for Dense Geometric Correspondence Fields

The estimation of correspondences in natural image pairs plays an important role in a large number of applications such as video coding, frame rate conversion, multi-viewpoint image generation, camera calibration, 3D from stereo, and structure from motion. The meaning of the correspondences differs in each application. In video coding, .e., MPEG-2, correspondences are motion vectors from a known image to an unknown image. The luminance of the unknown image is predicted along the vector and thus, the vectors have a photometric meaning. In camera calibration [15], [56], structure from motion [2], [32], [41] and 3D from stereo applications [36], [43], each correspondence is used to extract a 3D scene point. Therefore the correspondences have a geometric meaning. Frame-rate conversion [19] and multi-viewpoint image interpolation [34], [58] lie more or less between the photometric and geometric extremes.

In the area of correspondence estimation, we see a tendency toward simultaneous estimation of pixel-dense geometric correspondence fields, including explicit object segmentation [10], [11], [26], [52], [54], [60], [61]. This is due to several reasons.

First, there is a growing interest in high-accuracy, high-resolution 3D scene acquisition. This has been applied successfully in projects such as the European PANORAMA project [36]. High-resolution analysis requires the estimated fields to be pixel-dense. Simultaneous estimation of all vectors in the field allows for complex field models that contain any kind of interaction, which enhances the accuracy. Algorithms that estimate the vectors one by one impose a causality constraint on the field models. This may enable a fast implementation, but it restricts the accuracy.

Furthermore, in multi-viewpoint image generation, extrapolation [45] provides additional viewpoints to those obtained through interpolation [58], enlarging the geometric role of the correspondences. New image-coding techniques, such as MPEG-4, aim at object-based processing to increase interactability with the image content. In the camera calibration area, fixed calibration [15], [56] is an off-line technique that requires a special calibration object and a sparse correspondence field. Self-calibration techniques allow for online processing without special objects or user intervention. A dense field can be used to obtain high parameter accuracy [48].

The simultaneous estimation of a dense geometric correspondence field is a challenge for two reasons [26]:
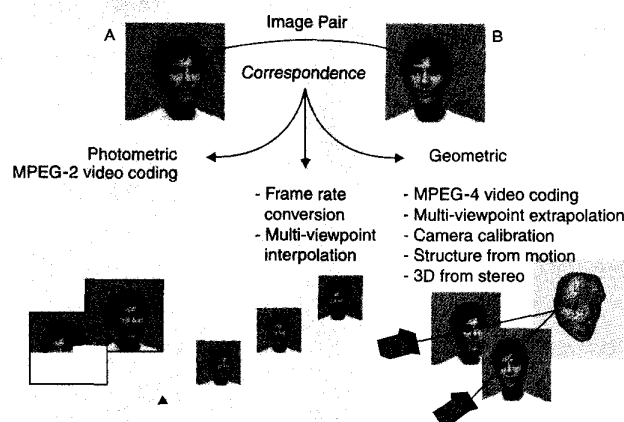
▲ The dimensionality of the solution space is extremely large, in the order of 10^6. This is computationally demanding, but it becomes more feasible regarding the continuing increase in computational power.

▲ The estimation of a geometric correspondence field based on photometric luminance fields of an image pair is not straightforward. The strongest constraint that relates correspondence and luminance is the so-called constant image brightness (CIB) assumption [23]. It states that a corresponding pixel pair has equal luminance. In Fig. 1, two contours of equal luminance are depicted in an image pair. If we take a point P_A on the contour in image A, the question is to which point in image B it corresponds.

For a photometric correspondence, all points on the contour in B would do. But there is only one point that corresponds geometrically, and we are not sure whether it lies on the contour in B, or not. If it does not, this can be due to camera noise, specular reflectivity of scene surfaces, or the use of a stereo camera with unbalanced photometric properties.

Thus, the CIB constraint alone is insufficient for the estimation of dense geometric correspondences. For this reason correspondence estimation is often called an ill-posed problem [4]. Additional geometric constraints are needed, together with an appropriate photometric model that accounts for deviations from the CIB model.

This article provides an overview of current techniques for dense geometric correspondence estimation. We will first formally define geometric correspondence and investigate the different types of image pairs. Then, we briefly look at the classic approaches to correspondence estimation, at their feasibility and flaws for simultaneous dense estimation. We will focus on the Bayesian approach, which is very well suited for this task, and for which several promising algorithms have recently been developed. After having a look at the future of the Bayesian approaches, we conclude with a discussion.

Geometric Correspondence in Image Pairs

Definition of Geometric Correspondence

If the luminance I of a point P_A in image A and a point P_B in image B have been defined by the same scene point, we say that P_A and P_B correspond (Fig. 2). From this point on, we specifically mean geometric correspondence whenever we mention correspondence, unless stated otherwise. Due to object transparency and camera defocus, the luminance of one point in an image may be defined by several scene points at the same time. This holds for both images, giving rise to multiple (many-to-many) correspondences (Fig. 3).

At this point, we are not aware of any attempt to take multiple correspondence into account in dense correspondence estimation. Therefore, from this point on, we will assume that all scene objects are opaque.

Opaque objects that move in front of each other cause occlusion in images. It is possible that a scene point P is visible in image A as P_A, while in image B it is occluded by another scene point, Q, visible in B as Q_B. We define that there is a pseudo-correspondence from P_A to Q_B (Fig. 4). The point P_A is called an occlusion point. Pseudo-correspondences enhance the quality of images generated in multi-viewpoint and frame-rate conversion applications. They provide information about the position of point P in all intermediate images in which P is visible. In 3D-from-stereo and structure-from-motion applications, the models obtained are more complete. It is expected that pseudo-correspondences can be obtained less accurately than real correspondences, since no photometric constraints are available for their estimation. Geometric constraints are the only clue.

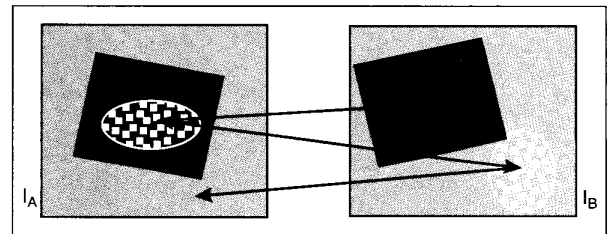## Spatial and Temporal Image Pairs

The image pairs in all applications up to this point can be categorized into two types: spatial and temporal image pairs. A special subset of spatial pairs comprises the paral-
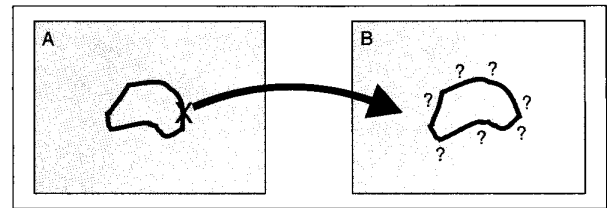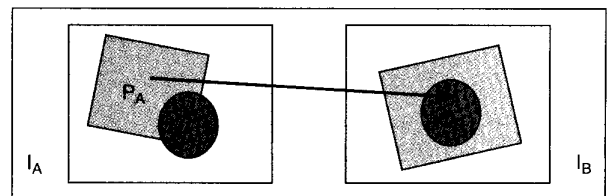

▲ 1. *Photometric similarity is insufficient in geometric correspondence estimation.*
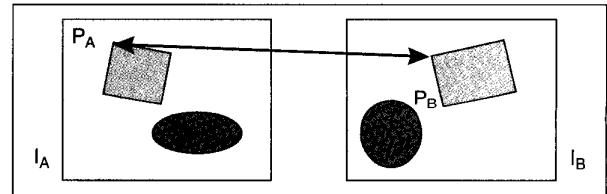
▲ 2. *Correspondence between P_A and P_B.*

▲ 3. *Multiple correspondences.*

▲ 4. *Pseudo-correspondence from occlusion point P_A to some point Q_B.*

lel image pairs. We will discuss these image pairs and illustrate their use in the 3-D-from-stereo and structure-from-motion applications.

Spatial Image Pairs

Spatial image pairs are obtained by simultaneously recording a scene with two cameras (A and B) at different positions (Fig. 5). In spatial image pairs, correspondence estimation is called disparity estimation.

In 3D-from-stereo applications, we can reconstruct a 3D scene point by triangulating a pair of corresponding points, as depicted in Fig. 6. Triangulation requires calibrating the cameras. The relative position, orientation, and some additional parameters of the cameras must be known. An overview of camera-calibration methods appears ("Multi-Camera Systems: Calibration and Applications," pp. 55-65, this issue).

The distance between the cameras is called the baseline. The larger the baseline, the more accurate the triangulation, given the finite accuracy of the estimated correspondences. Large baselines also yield large differences in the image pair—a challenge for the estimation algorithms (Fig. 7).

A special situation arises if the cameras are in a so-called parallel setup. This setup requires that two identical pinhole cameras (no lens distortion or CCD misplacement) are placed with equal orientations, while their positions differ only in the direction of the scan lines. In this way, a parallel image pair is obtained, in which corresponding pixel pairs lie on equal scan lines (Fig. 8).

This restriction on the correspondences reduces the complexity of estimation by an order of magnitude. Many disparity estimation algorithms are developed for parallel image pairs [11], [16], [24], [44], [58], [60].

When cameras providing a spatial image pair are calibrated, the A and B images can be warped in a so-called rectification procedure [37]. The result is a parallel image pair A′, B′, in which disparity can be estimated by means of algorithms for parallel pairs. For uncalibrated cameras, which provide uncalibrated spatial image pairs, this can not be done directly. The one-dimensional nature of disparity estimation can still be applied on the basis of epipolar geometry [15], [56], which will be discussed later.

Temporal Image Pairs

Temporal image pairs are obtained by recording a scene by a single camera that takes a shot at two different time instants, t_A and t_B (Fig. 9). The correspondences are related to the motion of scene objects. Correspondence estimation in temporal image pairs is, therefore, called motion estimation [10], [26], [52].

For scenes with rigid objects, temporal and spatial image pairs can be converted into each other. Those parts in the temporal pair that arise from one scene object can be reconstructed by placing two virtual cameras A′, B′ at appropriate positions around the selected object (Fig. 10). The difference in the positions and orientations of the virtual A′ and B′ cameras relate to the translation and rotation of the selected scene object. If the scene consists of only one object, the spatial and temporal image pairs are the same. This assertion is used in [41] for estimation of structure from motion.

In structure-from-motion applications, the reconstruction of the selected object can be handled in the same way as in the 3D-from-stereo application, with one exception. The two virtual spatial cameras cannot be calibrated offline. As a result, self-calibration techniques have to be used on the basis of the estimated correspondence field [41], [48].
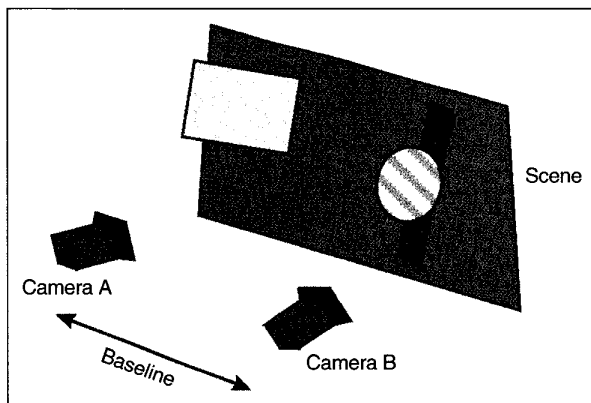
▲ 5. A stereo camera provides a spatial image pair.
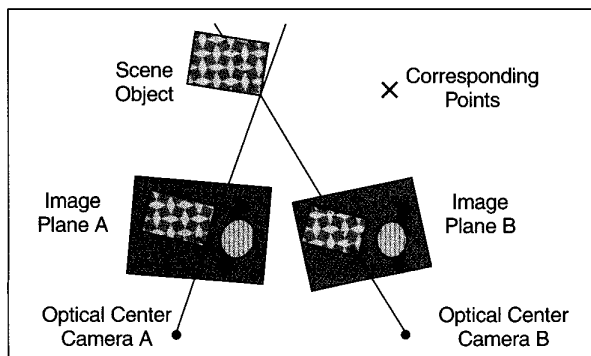
▲ 6. Triangulation of correspondences.

Classic Approaches to Correspondence Estimation

We will briefly discuss the classic approaches to correspondence estimation including: feature detection and matching, block matching, pel-recursive, and optical-flow techniques. For more details we refer the reader to the excellent overview in [53].

Feature-Based Algorithms

Feature-based algorithms [3], [28] first extract predefined features, and then match these (Fig. 11). The separation of detection and matching is a restriction on the

quality that can be obtained. The definition of features is not easy. The most well-known and general feature is the edge. Edge definition and estimation has been investigated over long periods [9]. This approach yields a sparse correspondence field.

Block-Matching Algorithms

In block matching, rectangular blocks of pixels are matched [1], [22] (Fig. 12). For each block in image B, a block is sought in image A most resembling the block in B according to some criterion. A dense field can be obtained by means of interpolation or the use of overlapping blocks. During estimation, a single correspondence vector is used for all pixels within one block. Since the vector models only translation, this approach does not work well for rotated and skewed objects in an image pair.

For large textured areas undergoing relatively uniform motion, large blocks enable high-accuracy correspondence estimation. The uniform-motion restriction, however, limits the resolution obtained. To some extent, this can be overcome by adapting the block size to the image content [25].

Pel-Recursive Algorithms

These algorithms [5], [6] have been developed for image-sequence coding. They obtain a dense field by scanning, i.e., they start the estimation at the upper-left pixel and end at the bottom-right pixel (Fig. 13). First, the luminance of pixel x in image B is predicted from image A by means of the correspondence vector found at the previous pixel in B (pixel 6 in Fig. 13). Then a group of N pixels (here $N=7$) is matched to image A. The group has a "causal" shape in the sense that it contains only pixels with known luminance in B. In the pel-recursive approach, we use an analytical expression to obtain the new vector on the basis of the previous one. It is assumed that the previous vector is a good estimate of the new vector and thus, only small changes are allowed between two vectors.

The regular structure and causality of block matching and pel-recursive techniques allow them to be implemented efficiently in hardware [22], [36]. However, the causality restricts the quality of the correspondences obtained.

Optical-Flow Algorithms

This method is the first approach to the simultaneous estimation of a dense correspondence field [23]. The method relies on the relation between photometric correspondence vectors with components (C_x, C_y) and spatiotemporal derivatives of luminance in an image sequence:

$$\left(C_x \cdot \frac{\partial}{\partial x} + C_y \cdot \frac{\partial}{\partial y} + \frac{\partial}{\partial t} \right) I(x, y, t) = 0 \tag{1}$$

An additional regularization term biases the solution toward a globally smooth correspondence field [23], [57].
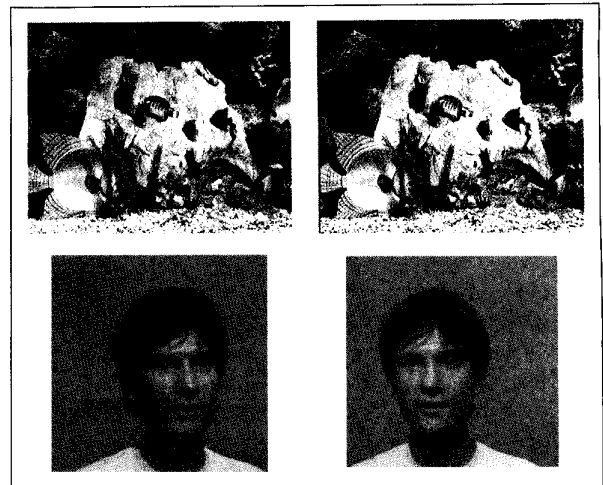
▲ 7. Small (top) and large (bottom) baseline.
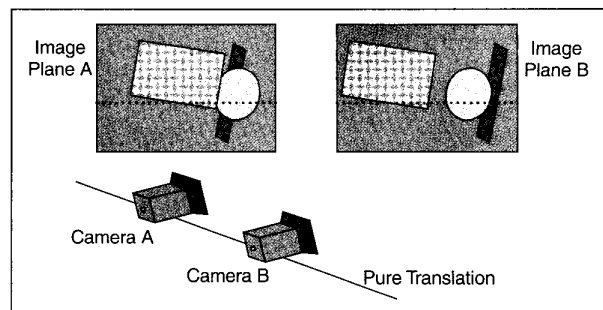
▲ 8. A parallel camera setup provides a parallel image pair.

Discontinuity fields have been incorporated to avoid oversmoothing at object boundaries [21].

The drawback in this approach is that the luminance derivatives are numerically approximated. This requires local spatiotemporal linearity of luminance. In image sequences with large motion, the local linearity is violated. In stereo applications, the temporal axis is replaced by a camera position axis. For a camera baseline of any reasonable size, the position linearity is violated.

Bayesian Approach to Correspondence Estimation

More recent approaches for dense correspondence estimation are the Bayesian methods, applied to temporal image pairs [7], [10], [26], [51], [54], [61] and to spatial pairs [11], [48], [60]. In this approach, the simultaneous estimation of dense correspondence fields is easily combined with object segmentation. The luminance derivatives in the optical flow method are avoided.

In the Bayesian approach or framework, we distinguish four steps, depicted in Fig. 14. The separation of the problem statement in the first three steps and the derivation of a search algorithm in step 4 [13] increases the portability and adaptability of algorithms among different applications and different designers.
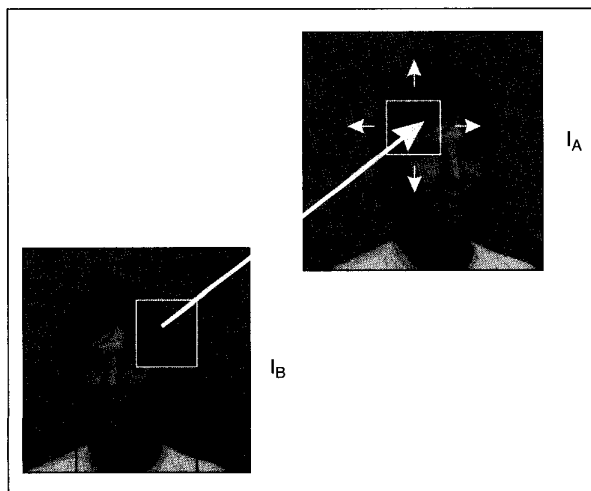
IEEE SIGNAL PROCESSING MAGAZINE
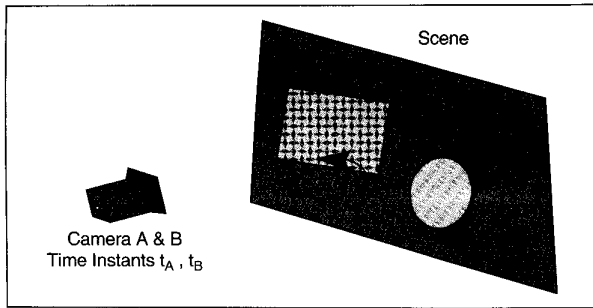
▲ 9. A single camera provides a temporal image pair.
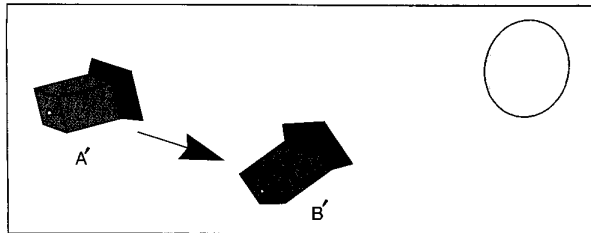
▲ 10. Spatial construction of a rigid object from a temporal image pair.
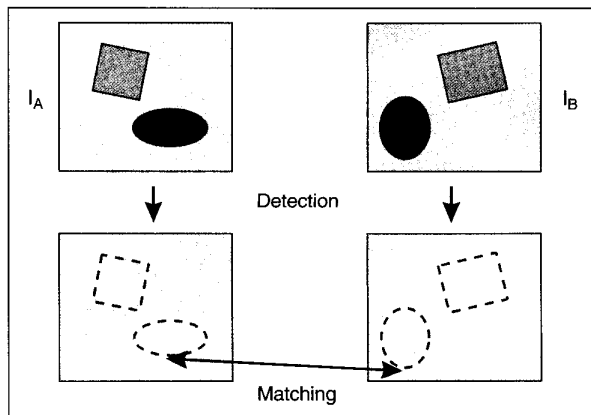
▲ 11. Feature detection and matching.

▲ 12. Block matching.

In the first step, the input images I_A and I_B, and all output fields $\{F_1, F_2, ...\}$ to be estimated are defined. The output fields represent correspondence, occlusion, and possibly discontinuity and segmentation fields.

In step 2, the relations between all these fields are modeled with a joint probability function in $F = \{F_1, F_2, ...\}$, conditioned by the observed image pair i_A, i_B:

$$p_{F|I_A,I_B}(f, i_A, i_B) \qquad (2)$$

It is a density in the continuous fields in F and a mass function in the discrete fields. In the remainder of this article, we will not refer to this explicitly. The design of the joint model is usually decomposed by means of the Bayes rule, hence the name of these approaches:

$$p_{F_1,F_2,F_3|I_A I_B} = p_{F_3|F_1,F_2,I_A,I_B} p_{F_2|F_1,I_A,I_B} \cdot p_{F_1|I_A,I_B} \qquad (3)$$

In the third step, the best solution F_{SOL} is defined by a criterion on the probability function, such as the maximum a posteriori (MAP) criterion.

In the fourth and final step, a search algorithm is formulated that computes the defined solution or a relevant approximation. We will now focus on each of the four steps in the Bayesian framework.

Dense-Field Representations

For the correspondences and occlusions defined earlier, dense-field representations C and O have been developed. For segmentation purposes, additional edge-based segmentation fields S and region-based fields R have been proposed.

Table 1 shows a list of fields used by several authors in their and our notation. We will now take a close look at each of these fields.

Correspondence and Occlusion Fields

The occlusion points, the real and the pseudo-correspondences, can be represented by several pixel-dense fields. They are all defined on the pixel lattice Λ_P (Fig. 15). The lattices of the images I_A and I_B are denoted by Λ_{PA} and Λ_{PB} respectively.

The correspondence fields C that are most commonly used are defined on one of the image's lattices Λ_{PA}, Λ_{PB} [10], [48], [52], [54], [60], [61]. The C_A field is depicted in Fig. 16. Each entry C_A (P_A) contains a vector with its starting point at the entry P_A on the lattice Λ_{PA}. For pixel accuracy correspondences, the endpoint of the vector lies on the lattice Λ_{PB}. For subpixel accuracy, the vectors end on the continuous domain Λ^*_{PB}.

Most applications benefit from subpixel accuracy, which is reflected in the number of subpixel estimation algorithms that have been developed [7], [10], [52], [54], [60], [61]. For subpixel accuracy, the luminance of the images has to be interpolated to the continuous domain Λ^*_P. In [26], we find (experimentally) that the specific
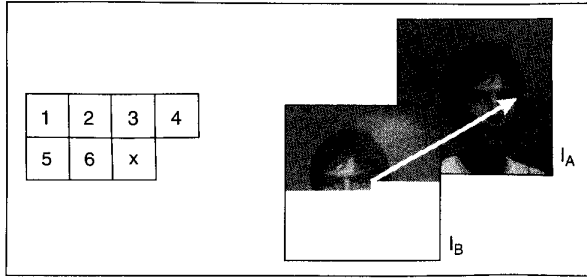
△ 13. Pel-recursive technique.

choice of the interpolation filter does not have much influence on the estimation of correspondences.

If (x_A, y_A) and (x_B, y_B) are the coordinates of a corresponding pixel pair, the value of the correspondence field C_A is:

$$C_A(x_A, y_A) = \begin{bmatrix} x_B - x_A \\ y_B - y_A \end{bmatrix} \tag{4}$$

The value represents the 2-D (vector) displacement of the projection of a scene point between image A and image B. Depending on whether the estimation is performed with pixel or subpixel accuracy, the components of C are integer or real valued.

The C_A field can represent both real correspondences between P_A and P_B and pseudo correspondences from P_A to Q_B. In the latter case, P_A is an occlusion point. The presence of occlusion points can be represented by the occlusion field O_A:

$$O_A(P_A) = \begin{cases} 0 & \begin{array}{l} P_A \text{ is visible in image B} \\ C_A(P_A) \text{ is a real correspondence} \end{array} \\ 1 & \begin{array}{l} P_A \text{ is an occlusion point} \\ C_A(P_A) \text{ is a pseudo - correspondence} \end{array} \end{cases} \tag{5}$$

Fig. 17 shows the binary occlusion fields O_A and O_B. If no occlusions are taken into account [26], the C_A field suffices in the modeling process because it can represent all real correspondences. If occlusions are taken into account, but no pseudo-correspondences are estimated, the C_A field contains all real correspondences and a number of undefined vectors [54]. If pseudo-correspondences are estimated, both C_A and C_B fields are needed to represent all pseudo-correspondences. In [52], the pseudo-correspondences from A to B, contained in C_A, are estimated. At this point, no attempt has been made to estimate all pseudo-correspondences.

For parallel image pairs, all real correspondences and both occlusion fields O_A and O_B can be represented by one field, the so-called chain map [46], provided pixel accuracy is used

and an additional ordering constraint, discussed near the end of this paper, holds true [11], [16], [44], [58].

The C_M field used in [43], [44], [45] is defined on a different domain Λ_{PM}. It is the pixel grid of a virtual image centered between images A and B (Fig. 18). In [26], a more general case is considered where M is placed at an arbitrary position in between A and B. The value of the C_M field is defined similar to (4):

$$C_M(x_M, y_M) = \begin{bmatrix} \dfrac{x_B - x_A}{2} \\ \dfrac{y_B - y_A}{2} \end{bmatrix} \tag{6}$$
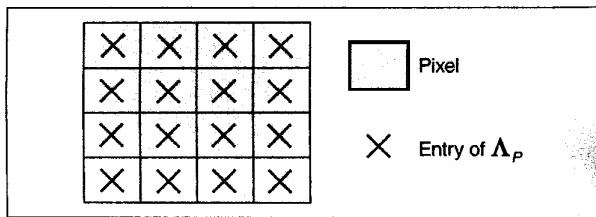
with
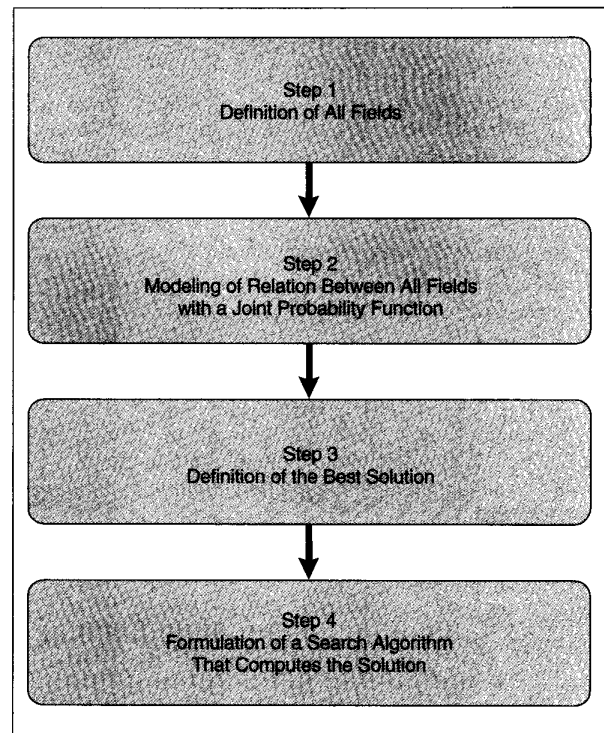
▲ 14. The Bayesian framework.

Table 1. Luminance, Correspondence, Occlusion, Discontinuity, and Segmentation Fields						
	I_A	I_B	C_A	O_A	S_{CA}	R_A
[26] Konrad & Dubois '92	g_t	g_{t+}	D_t		L_t	
[10] Chang et al. '94	g	g'	u,v			x
[54] Tekalp '95	g_k	g_{k-1}	d	O	l	
[60] Woo & Ortega '96	F^j	F^r	D	Φ		
[52] Stiller '97	g_t	g_{t+1}	d_t	see text		l_t

IEEE SIGNAL PROCESSING MAGAZINE

▲ 15. The pixel lattice Λ_P
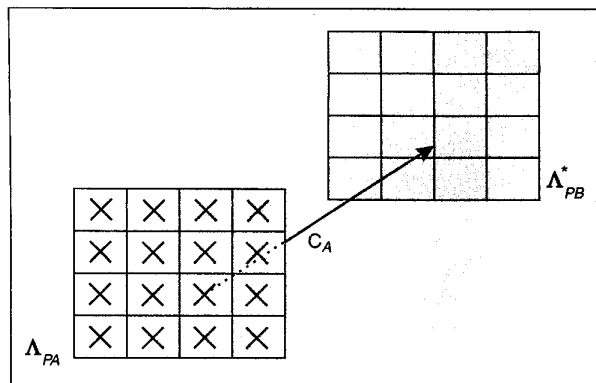
▲ 16. The C_A correspondence field.
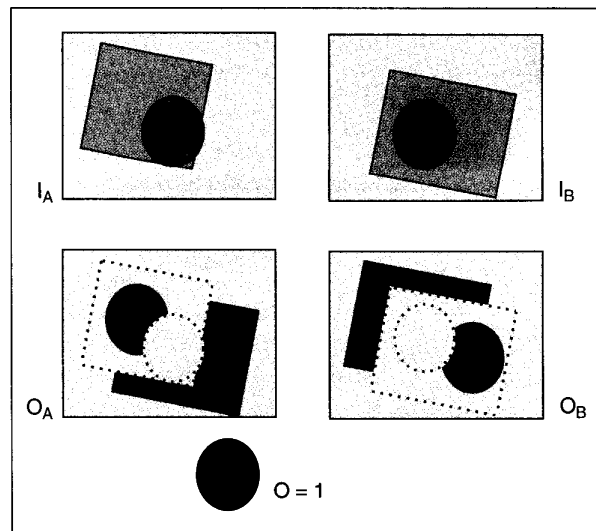
▲ 17. The occlusion fields O_A and O_B.

▲ 18. The C_M correspondence field.

$$\begin{bmatrix} x_M \\ y_M \end{bmatrix} = \frac{1}{2}\begin{bmatrix} x_A - x_B \\ y_A - x_B \end{bmatrix}$$

(7)

The C_M field is attractive because of its symmetry in the A and B images. In 3D-from-stereo applications with parallel image pairs, triangula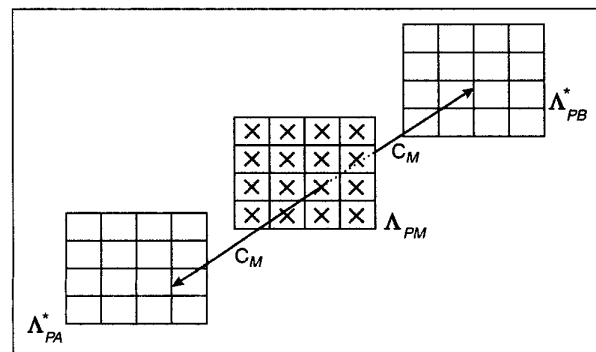tion of the correspondences is very easy [45]. However, it does not allow for an easy incorporation of asymmetric phenomena, such as occlusions and pseudo-correspondences. The field can be used in applications where these phenomena do not play an important role, for example, in face acquisition from stereo images [43].

In some cases, the C_M field cannot represent the real correspondences. The worst-case example is when image B is a 180° rotated version of image A. Then all vectors intersect in the center of C_M.

## Object Segmentation and Correspondence Discontinuity Fields

Some approaches in correspondence estimation do not model discontinuities [23], [26 (MEC algorithm)]. High-quality correspondence estimates are possible if the scene does not contain more than one object of interest, such as face acquisition from stereo [43].

Segmentation and correspondence discontinuities need to be introduced for image pairs with multiple objects. This has led to the introduction of correspondence discontinuity fields S_C, often called line fields [26], [54], [61] and object label fields R [10], [52].

Fig. 19 illustrates the discontinuities S_C in the correspondence fields C in the case of a simple scene with two objects in front of a background. Obviously, the discontinuities coincide with the object boundaries. As Fig. 17 depicts, object boundaries often coincide with boundaries of occlusion areas as well. In [21], experimental results indicate that the incorporation of S or R fields is useful only if occlusion fields O are also taken into account.

The discontinuity fields S are edge-based, for which a domain has to be defined. A widely used domain is defined as all sites between two pixels that are four-connected neighbors, denoted by Λ_{S4}, shown in Fig. 20. Clearly, Λ_{S4} contains two different kinds of sites, corresponding to horizontal edges between upper and lower pixels, and vertical edges between left and right pixels.

The discontinuity fields S normally have binary values. A "0" indicates continuity, and a "1" represents a discontinuity (Fig. 21). Region-based segmentation fields R contain labels for each pixel in the image lattice Λ_P. In [52], a label field R_A is introduced containing natural numbers. Each region of pixels sharing the same label represents a region that is smooth both in the luminance and correspondence fields. In [10], [52] a correspondence discontinuity field S_{CA} is derived from a label field R_A (Fig. 22). In [10], the R_A field is only used for this purpose, while in [52], additional ordering information al-
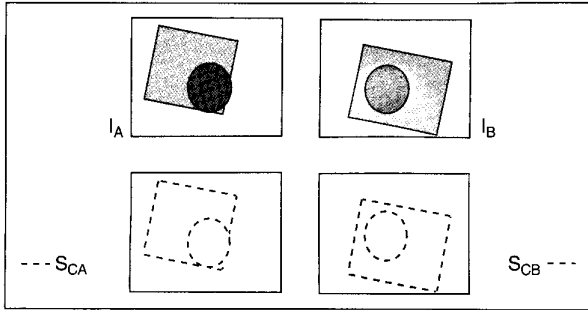
▲ 19. Discontinuities in correspondence fields.
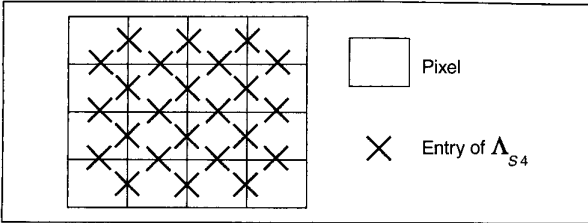
▲ 20. Four-connected edge domain Λ_{S4}

lows for the analytic derivation of the occlusion field O_A, using the C_A field as well. A major difference between the S and R fields is that R fields can not model the open curves as shown within the square in Fig. 21. These open curves may appear in real images, however, as shown in Fig. 23. A single object consisting of a pyramid attached to a plane is recorded by a stereo camera. The fact that the object occludes itself partly in image A leads to open curves of correspondence discontinuities in A.

Modeling the Joint Probability of Several Dense Fields

The design of a joint probability model for several dense fields is by no means an easy task. In general, the modeling process is decomposed at two levels. First, via the Bayes rule, each field can be modeled one at a time. Secondly, we can obtain the global model of each of these fields by combining many equal, simple local models. These assume independence of all entries in a field, or dependence only in a small neighborhood reflecting the Markov property (see "Gibbs and Markov Random Fields").

As an example we take the approach of [54], in which the following joint probability is modeled:

$$p_{C_A, S_{CA}, O_A | I_A, I_B} \tag{13}$$

With the Bayes rule, the joint model is decomposed in several single field models:

$$\frac{p_{I_A | C_A, O_A, I_B} \cdot p_{C_A | S_{CA}} \cdot p_{O_A} \cdot p_{S_{CA} | I_B}}{p_{I_A | I_B}} \tag{14}$$

In this decomposition, several independencies among the fields are assumed. Table 2 shows the joint probability models and Bayes decompositions for the fields in Table 1.

Each of the Bayes factors represents a specific photometric or geometric model, or a combination of both. Examples of photometric models are the CIB assumption and its deviations. Geometric models involve continuity and smoothness of the correspondence field.

We will now present several commonly used models and then combine them into a joint model.

Image Luminance Models

The factor in the denominator of (14) is a constant, given that we have observed the images i_A and i_B. In steps 3 and 4, in the framework, criteria for best solutions and search algorithms are selected that do not need the actual value. Thus this factor is never modeled [10], [26], [52], [54], [60].

The first factor in the numerator is similar to the second factor in all Bayes decompositions in Table 2. It has the form:

$$p_{I_A | I_B, C......} \tag{15}$$

The factor represents the probability of the A image, given that we are able to predict its luminance by the B image, the correspondence, and other fields.

All current models for (15) assume that luminance is a field with independent entries:

$$\text{Independence: } p_{I_A | ......} = \prod_{\substack{all\ pixels \\ in\ image\ A}} p_{pixel\ in\ A | ......} \tag{16}$$

The basic tool for (16) is the CIB assumption:

$$\text{CIB: } I_A(P_A) = I_B(P_B) \tag{17}$$

The CIB assumption is valid if the cameras are noiseless and all objects have diffuse reflection properties. Additionally, in a spatial image pair, the cameras should be photometrically equal. In a temporal pair, the photometry of the camera must not change over time, and light sources cannot move, with respect to other objects.

All current correspondence estimation algorithms assume CIB as a starting point, and model the deviations to some extent. Mostly, the causes for CIB deviations are modeled together by a zero-mean Gaussian [10], [26], [54], [60]:
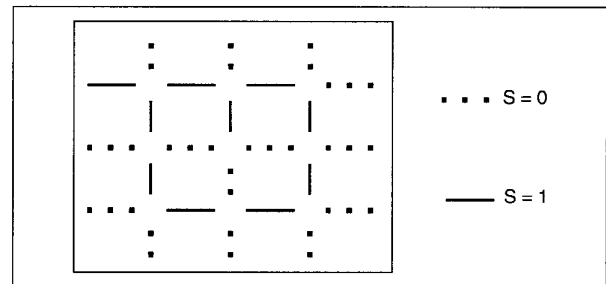
▲ 21. An edge-based correspondence discontinuity field S.

CIB deviations: $p(\Delta I) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\Delta I^2}{2\sigma^2}}$

(18)

In [52], a generalized Gaussian is used, with shape and variance parameters estimated from the images. The shape parameters obtained suggest that a Laplacian outperforms a Gaussian—a result found earlier in [31]. In [39], deviations due to specular reflections of scene surfaces are modeled in a feature-based approach for correspondence estimation. Photometric differences in

cameras can be accounted for in advance by using luminance histogram warping [12].

For occlusion points in image A, no relation, such as (17) or (18), can be established. In [52], the luminance is then modeled with a uniform probability distribution over all gray levels:

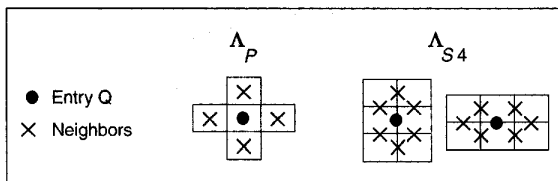Occlusion points: $p(I) = \dfrac{1}{N_{graylevel}}$

(19)

Gibbs and Markov Random Fields

Markov random field (MRF) models can be used to model interactions between entries in a field. For a thorough introduction in MRFs we refer to [18]. Here we will recall the basic characteristics of MRFs, which will suffice for our overview.
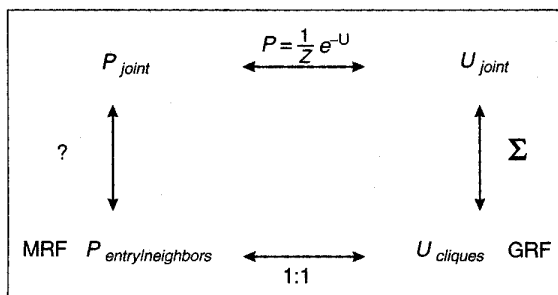
In MRF models, all entries interact with each other only via their direct neighbors. Fig. A shows typical examples for neighborhoods on the Λ_P and Λ_{S4} lattice. Mathematically, if all entries that are neighbors to entry Q are known, the probability distribution for the Q entry does not depend anymore on the rest of the field:

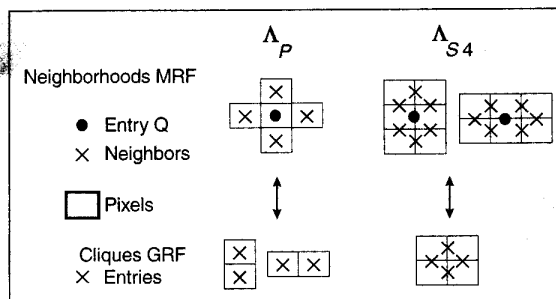$P_{entry\ Q|all\ other\ entries} = P_{entry\ Q|neighbors\ of\ Q}$

(8)

Applying (8) to all entries in the field defines the joint probability for the whole field. However, a practical problem is that the joint probability is not available explicitly. This is solved by the introduction of the Gibbs random field (GRF), see Fig. B. There is a one to one mapping between GRFs and MRFs [18]. A GRF is defined in the energy domain:

▲ C. Neighborhoods and cliques for Λ_P and Λ_{S4}.

$P_{MRF\ joint} = \dfrac{1}{Z} e^{-U_{GRF\ joint}}$

(9)

The joint energy of the GRF is defined as a sum of clique energies:

$U_{GRF\ joint} = \displaystyle\sum_{all\ cliques} U_{clique}$

(10)

A clique is a small group of field entries whose energy is a function of the field values. The neighborhoods in MRFs are related to the cliques in GRFs. Fig. C shows the cliques according to the neighborhoods in Fig. A. The neighbors of an entry Q are all entries that share a clique with Q.

The normalization constant Z in (9) is called the partition function, and is given by (assuming a discrete valued GRF):

$Z = \displaystyle\sum_{all\ different\ fields\ f} e^{-U_{GRF\ joint}(f)}$

(11)

Analytical computation of Z is impossible in general, and so is numerical computation since the space of all different fields is very high dimensional (in the order of 10^6). For the successful use of GRF models, the application should not depend on the actual value of Z.

An example of a GRF model that enforces global smoothness on a correspondence field C_A is (12). Large variations in the C_A field yield a high energy, which leads to a low probability for that field. In (12), the cliques are chosen as depicted in Fig. C for the Λ_P lattice.

▲ A. Typical neighborhoods in a Markov random field.

▲ B. Joint probability of MRF and GRF.

$U_{CA} = \displaystyle\sum_{\substack{all\ P,Q\ that \\ form\ a\ clique}} |C_A(P) - C_A(Q)|^2$

(12)

As an example, in "Modeling Image Luminance: CIB Deviations and Occlusions," the image luminance model (15) is derived on the basis of (16), (18), and (19), resulting in (21).

A similar expression is found in [52], based on a generalized Gaussian. In [26], no occlusions are taken into account, and they obtain only the product series with $O_A = 0$ in (21). The same holds for [54], in which the occlusion point model is discarded.

Correspondence Smoothness Models

Within continuous areas of a correspondence field, it is assumed that the field is also a smooth function of position. Smoothness of correspondence reflects smoothness of scene surfaces. The most basic smoothness constraint penalizes large values of the spatial derivatives of the correspondence field. This means that differences of neighboring entries in the field are penalized. As seen in the sidebar "Gibbs and Markov Random Fields," Gibbs and Markov random field models take into account such interactions between neighboring field entries.

An example of a Gibbs random field (GRF) model that enforces global smoothness on correspondence field C_A is:

$$U_{C_A} = \alpha \sum_{Q \in \Lambda_{S4A}} \left| C_A(P_{Q1}) - C_A(P_{Q2}) \right|^2 \tag{22}$$

Fig. 24 illustrates (22). For all neighboring entries (P_{Q1}, P_{Q2}) on the P domain, the cliques, the difference in the C_A entries are squared and added. In (22), the cliques are indexed by the entries Q of domain Λ_{S4}. Large fluctuations in the correspondence field yield high energies that result in a low probability for that field. The scale factor α is used in all approaches to regulate the influence of the smoothness constraint with respect to other constraints. Each constraint has its own parameter, and all parameters are usually determined by experiment.

As shown in "Gibbs and Markov Random Fields," the energy U_{CA} in (22) results in a Bayes factor:

$$p_{C_A} = \frac{1}{Z} e^{-U_{C_A}} \tag{23}$$

As discussed in Fig. 24, the value of Z cannot be computed. However, it is a constant and can be discarded from the modeling process in a similar way as the denominator in (14).

In [27], the square in (22) is replaced by more general functions that more or less incorporate discontinuities without modeling them explicitly.

It is often assumed that the smoothness of correspondence C is correlated with the smoothness of luminance I. Overviews of these photometric-geometric models can be found in [14], [33], and [50]. Basically, these models relax smoothness constraints across luminance edges, resulting in so-called oriented smoothness constraints.

Smoothing the correspondence field C_A while taking the discontinuities according to S_{CA} into account can be performed by a compound GRF model involving both fields [26]:

$$U_{C_A|S_{CA}} = \alpha \sum_{Q \in \Lambda_{S4A}} \left| C_A(P_{Q1}) - C_A(P_{Q2}) \right|^2 [1 - S_{CA}(Q)] \tag{24}$$

Variations of (24) can be found in [54] and [61]. In [10] and [60], a segmentation field R_A is used as in Fig. 22. In [60], the occlusion field O_A is used as approximation to R_A.

Modeling Image Luminance: CIB Deviations and Occlusions

As an example, we will derive a model for the I_A image given the I_B image, the C_A correspondence field, and the O_A occlusion field. According to (16), we model each pixel independently:

$$p_{I_A|C_A,O_A,I_B} = \prod_{P_A \bullet \Lambda_{PA}} p_{I_A(P_A)|C_A,O_A,I_B} \tag{20}$$

In the figure, the image pair is depicted, along with the correspondence and occlusion fields C_A and O_A. To incorporate the occlusion pixel model (19) in (20), we need to know which pixels in A are occlusion points, and for the Gaussian CIB deviation model (18) which are not. This information is contained in occlusion field O_A (5). In (18), the I term refers to the luminance difference of a pair of corresponding pixels in A and B. For each non-occlusion pixel in image A, we need a real correspondence vector that originates from that pixel. These vectors are the real correspondence vectors contained in the C_A field. If we apply (18) and (19) in (20) using the C_A and O_A fields (4) and (5), we obtain

$$p_{I_A|C_A,O_A,I_B} =$$

$$\prod_{\substack{P_A \in \Lambda_{PA} \\ \text{with } O_A=1}} \frac{1}{N_{graylevel}} \cdot \prod_{\substack{P_A \in \Lambda_{PA} \\ \text{with } O_A=0}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\{I_A(P_A)-I_B(P_A+C_A(P_A))\}^2}{2\sigma^2}}$$

$$\tag{21}$$

Correspondence Discontinuity Models

For the discontinuity field S_{CA} in (24), several models have been proposed, in which three different types can be distinguished. First, discontinuities can be penalized independently for each entry in the field:

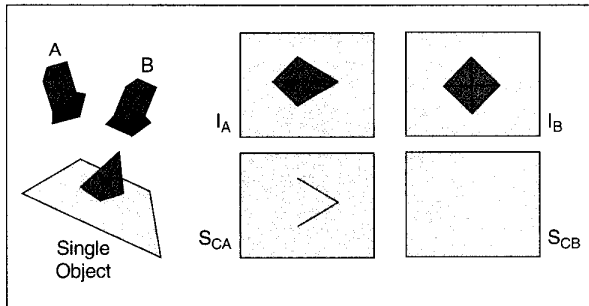$$U_{S_{CA}} = \alpha \sum_{Q \in \Lambda_{S+A}} S_{CA}(Q) \tag{25}$$

In [10], a model similar to (25) is used based on label field R_A and Fig. 22.

A second model for discontinuities is that they often coincide with luminance discontinuities [26], [61], reflecting a combined photometric-geometric constraint:

$$U_{S_{CA}|I_A} = \alpha \sum_{Q \in \Lambda_{S+A}} \frac{S_{CA}(Q)}{\left| I_A(P_{Q1}) - I_A(P_{Q2}) \right|^2} \tag{26}$$
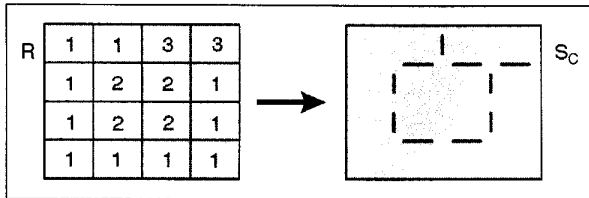
▲ 22. Extraction of a correspondence discontinuity field S_c from a label field R.

▲ 23. Open curves of correspondence discontinuities.

In the case of a zero in the denominator, no discontinuity is allowed.

In [52], a discontinuity field is defined on Λ_{S8}, that also includes entries for diagonally neighboring pixels (eight-connectedness). The field is extracted from a label field R_A in a similar way as that in Fig. 22.

The third model for discontinuities is globally connected curves. The connectivity is modeled in general by means of a GRF. Fig. 25 shows an example of the cliques and their energy in [54]. In [26], similar cliques are defined, including additional cliques that forbid pixels from being surrounded by discontinuities.

Occlusion Models

For the occlusion field, two types of models are generally applied. First, the presence of occlusions can be penalized [53], similar to (25) for S_C:

$$U_{O_A} = \alpha \sum_{P_A \in \Lambda_{PA}} O_A(P_A) \tag{27}$$

Penalizing occlusion points promotes the presence of real correspondences in an image pair.

The second model for occlusions encourages connectivity of occlusion points [60]. Such a model can be obtained by applying (25) on discontinuities S_O extracted from the O field, as depicted in Fig. 22.

Combination into a Joint Model

To combine the models (21), (24), and (26) and the occlusion model (27) into a joint probability model, we will convert the latter three to the probability domain. For the occlusion model, this results in:

$$p_{O_A} = \frac{1}{Z} e^{-U_{O_A}} \tag{28}$$

Similar to (23), the partition function Z is a constant and can be neglected. For the discontinuity adaptive correspondence smoothness model (24) we have:

Table 2. Joint Probability Models and Bayes Decompositions		
	Joint Model	Bayes Decomposition
[26] Konrad & Dubois '92	$p_{C_A, S_{CA}|I_A, I_B}$	$p_{I_B|I_A}^{-1} \quad p_{I_B|C_A, S_{CA}, I_A} \quad p_{C_A|S_{CA}} \quad p_{S_{CA}|I_A}$
[10] Chang et al. '94	$p_{C_A, R_A|I_A, I_B}$	$p_{I_B|I_A}^{-1} \quad p_{I_B|C_A, R_A, I_A} \quad p_{C_A|R_A} \quad p_{R_A}$
[54] Tekalp '95	$p_{C_A, S_{CA}, O_A|I_A, I_B}$	$p_{I_A|I_B}^{-1} \quad p_{I_A|C_A, O_A, I_B} \quad p_{C_A|S_{CA}} \quad p_{O_A} \quad p_{S_{CA}|I_B}$
[60] Woo & Ortega '96	$p_{C_A, O_A|I_A, I_B}$	$p_{I_A|I_B}^{-1} \quad p_{I_A|C_A, O_A, I_B} \quad p_{C_A|O_A} \quad p_{O_A}$
[52] Stiller '97	$p_{C_A, R_A|I_A, I_B}$	$p_{I_B|I_A}^{-1} \quad p_{I_B|C_A, R_A, I_A} \quad p_{C_A, R_A|I_A}$
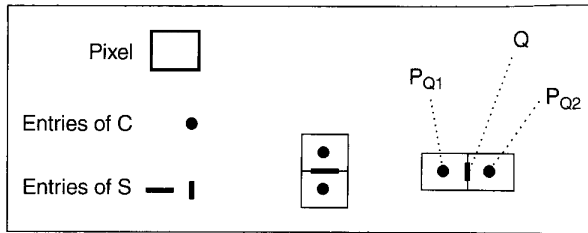
24. Cliques in a GRF for correspondence smoothing.
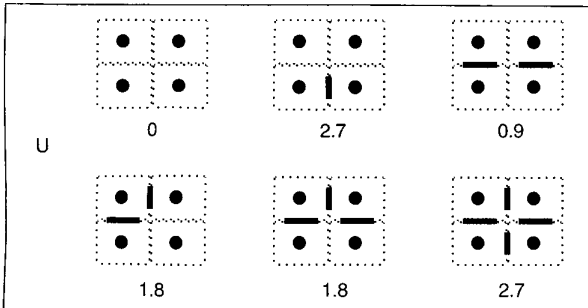
▲ 25. Cliques to enforce connectivity of correspondence discontinuities.

$$p_{C_A|S_{CA}} = \frac{1}{Z} e^{-U_{C_A,S_{CA}}}$$

(29)

In this case, the partition function Z is not a constant, but a function of the conditioning field SCA [18], which itself is not constant during estimation. This can be neglected, as is done in [10], [26], [54], [60].

In general, non-constant partition functions arise when two or more output fields interact with each other and are modeled in the energy domain. A way to circumvent this is to refrain from applying the Bayes rule on those fields. Instead, we combine the energies of (24) and (26) to form a model that is joint in these two fields:

$$p_{C_A,S_{CA}|I_A} = \frac{1}{Z} e^{-U_{C_A,S_{CA}} + U_{S_{CA},I_A}}$$

(30)

In [51], this is applied on the correspondence and segmentation fields C_A and RA (Table 2).

However, (21) and (30) contain circular dependencies and cannot be combined by means of the Bayes rule. If an attempt is made, the wrong image B appears in one of the Bayes factors (Table 2). One solution is to transform (21) to the energy domain via U=-ln p, and then add all energies to form a joint model:

$$p'_{C_A,S_{CA},O_A,I_A,I_B} = \frac{1}{Z} e^{(-U_{I_A,C_A,O_A,I_B} + U_{C_A,S_{CA}} + U_{S_A,I_A} + U_{O_A})}$$

(31)

For the energy terms in (31), no relation with any Bayes factor can be established for the decomposition of p '. In the optical-flow-based approach in [21], a joint model is designed similarly.

In (31), the joint model is constructed by adding energies freely instead of using the Bayes rule. This is at the cost of some explicitness in the modeling process, but it allows for the integration of several useful constraints.

Criteria for Best Solutions

The best solution f_{SOL} can be defined in many ways. In the area of simultaneous estimation of dense correspondence fields, two criteria are commonly used: the maximum a posteriori (MAP) criterion and the mean field (MF) criterion.

Maximum a Posteriori

The most widely used criterion is the MAP criterion [7], [10], [26], [52], [54], [60]:

$$f_{MAP} = \arg\max_{\hat{f}} p_{F|I_A,I_B}(\hat{f}, i_A, i_B)$$

(32)

The MAP criterion selects the solution that has highest probability given the observed images i_A and i_B. Since these are constants in the maximization in (32) we have:

$$p_{F|I_A,I_B} \propto p_{F,I_A,I_B} \propto p_{F,I_B|I_A} \propto p_{F,I_A,I_B}$$

(33)

The MAP solution can be obtained by maximizing any of the probability functions in (33).

Mean Field

The mean field (MF) criterion is used less frequently [61] than the MAP criterion and is defined as:

$$f_{MF} = \int_{\hat{f}} \hat{f} \cdot p_{F,I_A,I_B}(\hat{f}, i_A, i_B) d\hat{f}$$

(34)

It yields the average or expected solution, which is equal to the first moment of the probability function on the output variables conditioned by the observed image pair. It can only be used for continuous output variables, since discrete variables such as binary occlusion and discontinuity fields can not be averaged.

The MAP and MF criteria are special cases of a family of criteria, which is parameterized by some cost function [55]. The mean expected cost (MEC) criterion in [26] is equal to the mean field criterion. The discrete segmentation field in their MAP algorithm was removed for the MF algorithm. The MF criterion has been reported to yield results similar to the MAP criterion whenever both criteria can be applied [26].

Search Algorithms
MAP Search Algorithms
Since the dimensionality of (32) is extremely large, the probabilities get extremely small. Even for the actual MAP solution f_{MAP}, it may be on the order of $10^{-1,000,000}$.

Therefore, energy is used rather than probability in the numerical evaluation of (32). As an example, we will consider the joint model in (33):

$$p_{F,I_A,I_B} = \frac{1}{Z} e^{-U_{F,I_A,I_B}} \tag{35}$$

Maximization of probability is equivalent to minimization of energy, which does not require the computation of the partition function Z:

$$f_{MAP} = \arg \min_{f} U_{F,I_A,I_B}(\hat{f}, i_A, i_B) \tag{36}$$

Many search algorithms are available for the minimization in (36). They are either exact or approximate, and either deterministic or stochastic.

The most well-known technique for these kinds of minimizations is the downhill or gradient-descent method [42]. It is a deterministic method that easily gets stuck in local minima. To avoid local minima, stochastic methods are available, such as simulated annealing (SA). In SA, an estimate to the solution is perturbed at random. Better estimates (less energy) are always accepted; worse estimates are accepted now and then, governed by a temperature parameter. If the temperature is decreased from T_0 to zero infinitely slowly [18], the exact solution to (36) is reached. In practice, the temperature is lowered much faster and an approximation is obtained.

To use the SA algorithm, we need to define a temperature cooling schedule and a random perturbation generator. At this point, there are no general rules to help the designer. In [52], the cooling schedule is chosen to be exponentially decreasing. A table is presented with several perturbations, such as small changes in the correspondence fields and flipping of the binary values of the occlusion and discontinuity fields.

Many different versions of SA have been presented, e.g. the Metropolis algorithm [54]; iterated conditional modes (ICM) [10], [21], [54]; and so-called Gibbs sampler methods [18]. The interested reader is referred to the specific articles for details.

The only exact and deterministic algorithm for the MAP solution is the dynamic programming (Viterbi) algorithm [11], [16], [20], [24], [44], [58]. It can be used for disparity estimation in parallel image pairs. It requires that (36) is separable in all scan lines, excluding interactions between scan lines, such as smoothing.

Fig. 26 shows the MAP solutions obtained by an exact DP algorithm without vertical smoothing, and an approximate SA algorithm including vertical smoothing that obtains the dense field simultaneously.

Adaptations to the DP algorithm have been made in [35] and [47] that include vertical smoothing to some extent, without the need for simultaneous estimation.

Genetic algorithms (GA) have been used for correspondence estimation. In [17], the estimation is done separately for each scan line. For dense simultaneous esti-
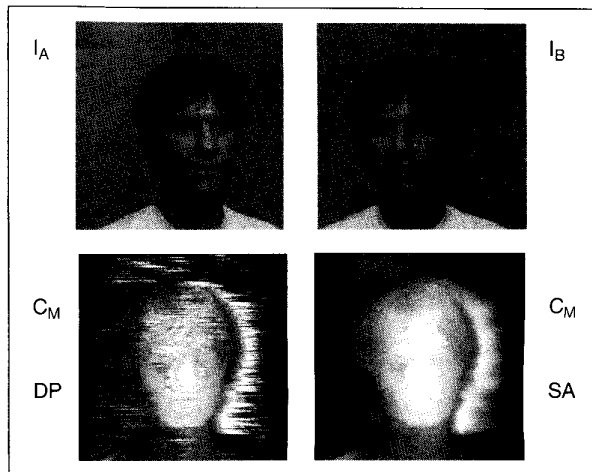
▲ 26. MAP solutions obtained by DP and SA search algorithms.

mation, the GA approach is not feasible since it requires several solution estimates to be maintained at the same time. This demands a tremendous amount of memory and computational power.

An extended Kalman-type recursive estimator is derived in [7] and also used in [8] to obtain the MAP estimate for the displacement vector field.

MF Search Algorithms

The mean field theory (MFT) is used in [61] to obtain the MF solution. It is based on the following approximation to (34):

$$f_{single} \approx \int_{\hat{f}_{single}} \hat{f}_{single} \cdot p_{F_{single}|F_{rest},I_A,I_B}(\hat{f}_{single}, f_{rest}, i_A, i_B) d\hat{f}_{single} \tag{37}$$

MFT means that if the mean solution of all fields f_{rest} is given except for a single entry of one field f_{single}, we can obtain an approximation to this single variable. Evaluation of (37) only requires integration over a single variable of the output space, while (34) requires integration over the entire solution space. The marginal probability model in (37) can easily be obtained from joint models on the basis of Gibbs Markov random fields [18].

Given an approximation to the complete solution, we can obtain a better approximation for each single variable with (37), in order to obtain the next approximation of the complete solution.

In [26], a different technique is used to obtain the MF solution, which is based on the so-called Gibbs sampler [18]. A Gibbs sampler provides a sequence of different realizations $f_{Gibbs,i}$ of the fields to be estimated, according to the probability model in (34). A statistical average of N of these realizations is an approximation to the mean solution:

$$f_{MF} \approx \frac{1}{N} \sum_{i=1}^{N} f_{Gibbs,i} \tag{38}$$

Hierarchical Approach

The MAP and MF search algorithms for dense fields yield a large computational burden. Although stochastic methods are designed to avoid local minima, the restrictions for a feasible implementation (fast cooling schedules and a low number of iterations) still lead to problems with local minima.

A general approach that provides faster convergence and, at the same time avoids local minima, is the hierarchical approach. Due to its good results for natural images, it is used in a wide variety of correspondence estimation algorithms [1] ,[10], [14], [21], [26], [28], [34], [38], [43], [52].

Fig. 27 depicts the hierarchical approach. The observed images are downsampled to lower-resolution versions. The original images are at level 0; the resolution decreases with level number. At the lower resolution level L the estimation starts. After estimation, the fields are upsampled to the resolution of level L-1. These fields are then used as an initial estimate for the estimation at this level. This continues until estimation is performed at full resolution level 0.

Many different upsampling, downsampling, and estimation schemes can be chosen. This involves the selection of new lattices for the lower-resolution fields, suitable filters and possibly level-dependent search algorithms. In most cases, the influence of these choices is small compared to other choices made in the four steps in the Bayesian framework.

Most popular schemes for the lower-resolution lattices are the 2:1 schemes [26], [52], in which both x and y axes are subsampled with a factor two. Schemes with non-integer ratios also exist. In [30], the effect of these schemes on computational efficiency is investigated.

Many different filters are used for downsampling the images, such as Gaussian [28], [38] and low-pass FIR filters [26]. In [52], bilinear filters are used for upsampling of the correspondence fields, and nearest-neighbor interpolation filters for discrete label fields.

Generally, for the estimation at different levels, the same algorithm is applied at each level. However, some authors include level dependencies, such as increased smoothness constraints [38], or removal of discontinuity fields at lower resolution levels [26]. In [29], special types of Markov random field probability models are investigated, for which the efficiency of level-independent estimation schemes is optimal. Using other models, however, may still result in near optimal schemes [52].

Future Model Improvements

It is expected that computational power will continue to increase. With this, the introduction of more fields and their simultaneous estimation becomes feasible. These fields enable more complex models with additional photometric and geometric constraints, and thus, enhanced results. We will discuss several models and constraints

which have previously been used in correspondence estimation, but not yet in approaches for simultaneous estimation of dense fields.

Two Correspondence Fields

The introduction of both $_{CA}$ and CB fields enables the estimation of all pseudo-correspondences. This is useful in several applications as discussed in the definition of pseudo-correspondence. Previously, both fields have been estimated separately to remove outliers in real correspondences [34], [36].

Image Restoration

The models for image luminance shown in this article all assume independence between pixels. A more complex image model is used in [8], where estimation of correspondence is combined with the restoration of an image sequence. These models include noise and blurring due to the camera system, that are part of the CIB deviation model.

In [8], the estimation is performed recursively for each scan line. In simultaneous estimation of dense fields, the improved image model requires the introduction of additional dense luminance fields, for example, in the case of image noise.

Specular Scene Surfaces

A different source of CIB deviations is specular reflectivity of scene surfaces. In [39], this is taken into account
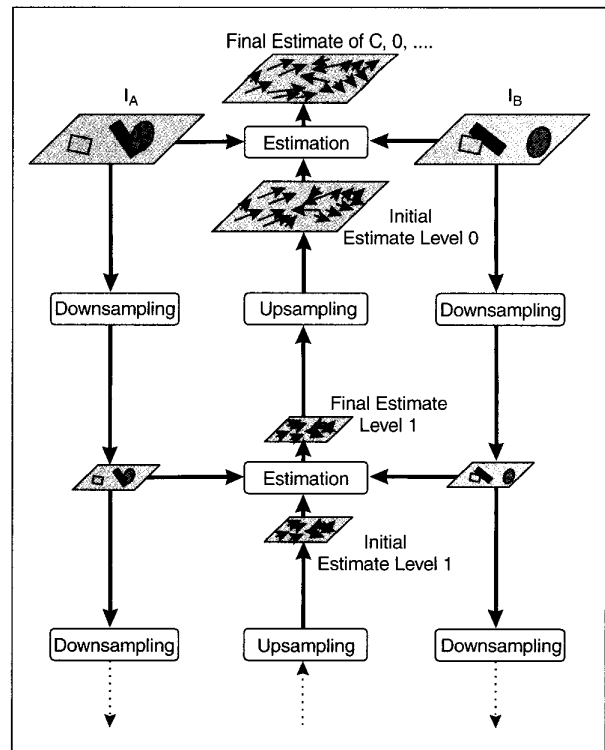
▲ 27. Hierarchical estimation.

in a feature-based approach for correspondence estimation. In dense simultaneous estimation, an additional field for specular reflectivity is needed.

Image Sequences

In this article, we deal only with pairs of images. Image sequences, both in temporal and spatial (multiple camera) directions, can be used to apply additional constraints.

In [32], correspondences are estimated in a sequence, where pairs of images $(t,t+t)$ are treated one by one. The results are integrated (enhanced) by a Kalman filter. Recursive approaches apply temporal consistency constraints [52] to enhance the estimation in the current image pair on the basis of the previously estimated fields. In [38] and [57], the recursive approach is applied on combined temporal/spatial image quadruples. In [40], spatial image triples are used to obtain accurate feature-based correspondences from the edges of curved objects. In [20], multi-camera spatial images are used. Due to a specific camera setup (all in one line) a single correspondence field is sufficient in the estimation process.

Simultaneous estimation of correspondences in general image sequences requires that multiple sets of all fields are present simultaneously. The number of fields is thus, (at least) linear with the number of images in the sequence.

Epipolar Geometry

Epipolar geometry provides a very powerful restriction on correspondences that have general validity. In dense correspondence estimation it has been widely applied on parallel image pairs, and on spatial image pairs from calibrated cameras. These pairs are equivalent up to a warping of the image lattices [37]. In [48], a first attempt is made to apply the epipolar constraint on spatial pairs from uncalibrated cameras. On temporal image pairs, the constraint has been applied in a block-based approach [51]. In simultaneous estimation of dense fields, epipolar geometry has not been used yet on temporal pairs.

We will explain the geometry and its applications in correspondence estimation. For more detailed overviews of epipolar geometry, we refer to [15] and [56].

Fig. 28 shows the interior of a stereo pinhole camera, including the optical centers and the image planes. The baseline is the line through the optical centers of the cameras (the term baseline is also used for the distance between the optical centers). Any plane that contains the baseline is called an epipolar plane. All scene points in such a plane are projected on a line in each of the images. These lines are the epipolar lines. A pair of epipolar lines that share the same epipolar plane are called conjugate epipolar lines. If two points from the image pair correspond, they should lie on conjugate epipolar lines. This is called the epipolar constraint. It reduces the set of possible correspondence candidates for a point in image A from all points in image B to only those on the conjugated epipolar line in B (Fig. 29). For pinhole cameras, the
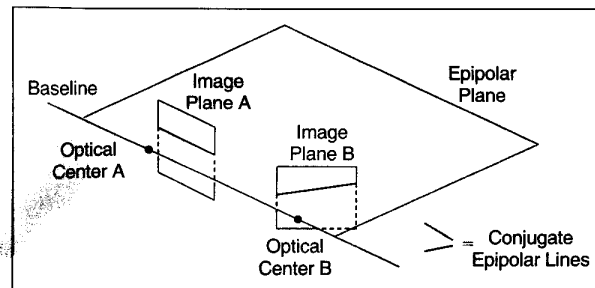
▲ 28. Epipolar geometry in a spatial pair.
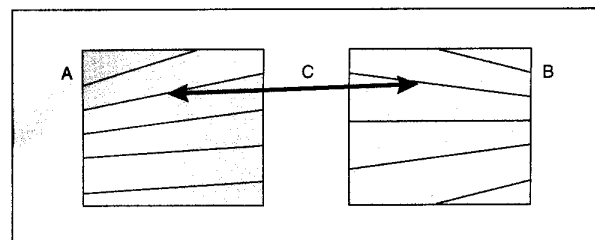
▲ 29. Correspondence is restricted to conjugate epipolar lines.

epipolar lines are straight lines. Due to lens distortion [59], the epipolar lines may become curved.

Parallel Image Pair

In a parallel image pair, the scan lines coincide with the epipolar lines. The epipolar constraint is applied by removing the y components from a correspondence field, which is then called a disparity field. Disparity estimation algorithms for parallel image pairs are widely available [11], [16], [24], [44], [58], [60].

Spatial Image Pair

For a spatial pair from uncalibrated cameras, the epipolar constraint can still be imposed if the geometry is estimated along with the correspondences. In [41], the pinhole camera geometry is estimated in a preprocessing step using sparse feature (corner) detection and matching.

In the simultaneous estimation of correspondences in uncalibrated spatial image pairs, the epipolar constraint has been applied recently [48] for cameras with lens distortion. In this approach, a field models the angle of the local tangent to the epipolar lines, and is estimated simultaneously (Fig. 30). The curvature of the epipolar lines is extracted from both images and, interpreted as lens distortion, penalized. The advantage of this approach is that it does not require feature extraction and estimation of predefined distortion parameters in a preprocessing step.

Temporal Image Pairs

Epipolar geometry can also be used in temporal image pairs. Consider a scene with a number of rigid objects moving differently. Each of these objects has its own epipolar geometry, shown in Fig. 31. This can be seen when the spatial construction in Fig. 10 is applied, which results in Fig. 32. To apply the epipolar constraint in tem-

poral pairs with multiple objects, one needs fields both for epipolar geometry E and its discontinuities S_E, visible in Fig. 31. From the rod-like object, it is clear that epipolar geometry may provide a strong clue for the global clustering of different image regions into one scene object. A similar constraint has been used in object rigidity checking on the basis of a sparse set of correspondences [49].

Ordering Constraint

A strong and useful relation exists between correspondence, its discontinuities, and epipolar geometry. If there is an interval without correspondence discontinuities along a pair of conjugate epipolar lines, then the ordering constraint holds at this interval. The constraint means that scene points appear in the same order along the intervals in A and B (Fig. 33). The ordering constraint does not always hold across discontinuities. Fig. 33 shows an example in which a thin object moves fast in front of a background (temporal pair), or is present in front of it, recorded by a stereo camera with large baseline (spatial pair).

In disparity estimation in parallel image pairs, the ordering constraint is often applied globally, regardless of discontinuities [11], [44], [58]. This simplifies the algorithms because a strong constraint can be applied without the need for discontinuity estimation. Additionally, it enables the use of the deterministic search algorithm dynamic programming (DP). In temporal and uncalibrated spatial image pairs, the ordering constraint has not been applied yet.

Discussion

Applications of Geometric Correspondences

Correspondence fields for image pairs play an important role in a wide variety of applications. These include MPEG-4 object-based coding, multi-viewpoint image generation, camera calibration, and structure-from-motion and 3D-from-stereo applications.

All of the applications require geometric correspondences. Such a correspondence represents a 3D scene point, while a photometric correspondence just represents photometric similarity between image points.

The high-resolution and high-accuracy estimation of geometric correspondences requires complex (non-causal) dense field models. This, on its turn, requires algorithms that estimate all correspondences in the dense fields simultaneously.

### Simultaneous Estimation of Dense Fields

The classic approaches to correspondence estimation, feature detection and matching, block matching, pel-recursive algorithms, and optical-flow methods are not well-suited for simultaneous estimation of dense geometric correspondence fields.

More recently, several promising algorithms for this task have been developed using the Bayesian approach. They use explicit probability models of the images, the correspondence fields, and their segmentation. Explicit models increase the portability and adaptability of algorithms among different applications and different designers.

As discussed in this article, several models for interacting fields cannot be combined using the Bayes rule. Instead, the models can be combined in the energy domain. This is at the cost of the explicitness of the modeling, but an advantage is that the joint model can be synthesized more freely and can include more submodels and constraints.

The submodels can be categorized into photometric and geometric models. Photometric models include image luminance and its discontinuities in relation to those of the correspondence field. Geometric models, needed for geometric correspondence estimation, currently include a priori models for occlusions, and continuity and smoothness of correspondence.
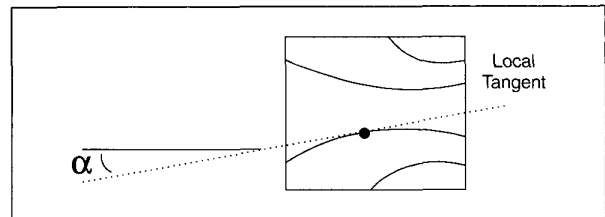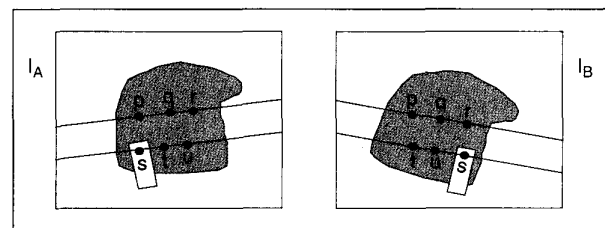

▲ 30. Epipolar geometry field.
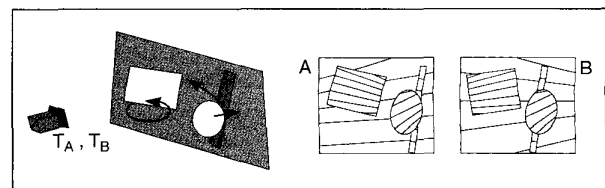
▲ 31. Epipolar geometry in temporal pairs.
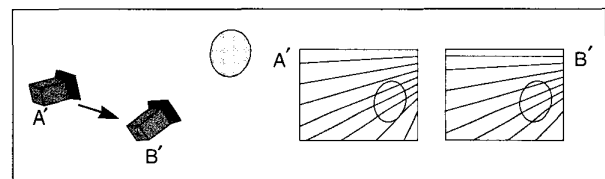
▲ 32. Spatial construction of one rigid object in a temporal image pair.

▲ 33. The ordering constraint.

Future Improvements

The increase in computer power will enable the simultaneous estimation of more fields in the future. Additional constraints can be introduced to enhance the quality of the estimation results.

The improvements discussed in this article include the estimation of all pseudo-correspondences, the incorporation of image restoration models, modeling of specular reflectivity of scene surfaces, the use of image sequences instead of pairs, and the application of epipolar geometry. The last provides one of the most powerful constraints in correspondence estimation. It has been applied widely on parallel image pairs, and recently on uncalibrated spatial pairs. For the simultaneous estimation of dense correspondence fields for temporal image pairs, it has not yet been used.

Acknowledgements

André Redert is a Ph.D. student, *Emile Hendriks* is an Associate Professor, and *Jan Biemond* (IEEE Fellow) is a Professor at the Information and Communication Theory Group in the Department of Electrical Engineering at the Delft University of Technology, Delft, The Netherlands. They can be reached at +31 15 278 6269; fax: +31 15 278 1843; e-mail: P.A.Redert, E.A.Hendriks, J.Biemond@its.tudelft.nl; URL: http://www-ict.its. tudelft.nl.

References

[1] M. Accame, F.G.B. De Natale, and D.D. Giusto, "Hierarchical block matching for disparity estimation in stereo sequences," *Proc. ICIP95*, pp. 374-377, 1995.

[2] A. Azarbayejani and A.P. Pentland, "Recursive estimation of motion, structure, and focal length," *IEEE Trans. PAMI*, vol. 17, no. 6, pp. 562-575, Jun. 1995.

[3] S.T. Barnard and W.B. Thompson, "Disparity analysis of images," *IEEE Trans. PAMI*, vol. 2, no. 4, pp. 333-340, Jul. 1980.

[4] M. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proc. IEEE*, vol. 76, no. 8, pp. 869-889, Aug. 1988.

[5] J. Biemond, L. Looijenga, D.E. Boekee, and R.H.J.M. Plompen, "A pel-recursive Wiener-based displacement estimation algorithm," *Signal Processing*, vol. 13, no. 4, pp. 399-412, Dec. 1987.

[6] L. Böröczky, "Pel-recursive motion estimation," Ph.D. thesis, Department of Electrical Engineering, Delft Univ. Technol., Delft, The Netherlands, 1991.

[7] J.C. Brailean and A.K. Katsaggelos, "A recursive nonstationary MAP displacement vector field estimation algorithm," *IEEE Trans. Image Processing*, vol. 4, no. 4, pp. 416-429, April 1995.

[8] J.C. Brailean and A.K. Katsaggelos, "Simultaneous recursive displacement estimation and restoration of noisy-blurred image sequences," *IEEE Trans. Image Processing*, vol. 4, no. 9, pp. 1236-1251, Sept. 1995.

[9] J. Canny, "A computational approach to edge detection," *IEEE Trans. PAMI*, vol. 8, no. 6, pp. 679-698, Nov. 1986.

[10] M.M. Chang, M.I. Sezan, and A.M. Tekalp, "An algorithm for simultaneous motion estimation and scene segmentation," *Proc. ICASSP94*, no. 5, pp. 221-224, 1994.

[11] I.J. Cox, S.L. Hingorani, and S.B. Rao, "A maximum likelihood stereo algorithm," *Computer Vision and Image Understanding*, vol. 63, no. 3, pp. 542-567, May 1996.

[12] I.J. Cox, S. Roy, and S. Hingorani, "Dynamic histogram warping of image pairs for constant image brightness," *Proc. ICIP95*, pp. 366-369, 1995.

[13] J.N. Driessen, "Motion estimation for digital video," Ph.D. thesis, Department of Electrical Engineering, Delft Univ. Technol., Delft, The Netherlands, 1992.

[14] W. Enkelmann, "Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences," *Computer Vision, Graphics, and Image Processing*, vol. 43, pp. 150-177, 1988.

[15] O. Faugeras, *Three-Dimensional Computer Vision, a Geometric Viewpoint*, Boston, MA: MIT Press, 1993.

[16] R.E.H. Franich, "Disparity estimation in stereoscopic digital images," Ph.D. thesis, Department of Electrical Engineering, Delft Univ. Technol., Delft, the Netherlands, 1996.

[17] R.E.H. Franich, R.L. Lagendijk, and J. Biemond, "A disparity space image path," *Proc. of the International Workshop on Three-Dimensional Imaging*, pp. 122-127, 1995.

[18] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. PAMI*, vol. 6, no. 6, pp. 721-741, 1984.

[19] G. de Haan, "Motion estimation and compensation," Ph.D. thesis, Department of Electrical Engineering, Delft Univ. Technol., Delft, the Netherlands, 1992.

[20] N. Grammalidis and M.G. Strintzis, "Disparity and occlusion estimation in multiocular systems and their coding for the communication of multiview image sequences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 3, pp. 328-344, Jun. 1998.

[21] F. Heitz and P. Bouthemy, "Multimodal estimation of discontinuous optical flow using Markov random fields," *IEEE Trans. PAMI*, vol. 15, no. 2, pp. 1217-1232, Dec. 1993.

[22] E.A. Hendriks and Gy. Marosi, "Recursive disparity estimation algorithm for real-time stereoscopic video applications," *Proc. ICIP96*, pp. 891-894, 1996.

[23] B.K.P. Horn, "Robot vision," Boston, MA: MIT Press, McGraw-Hill Book Company, 1986.

[24] S.S. Intille and A.F. Bobick, "Disparity-space images and large occlusion stereo," *MIT Media Lab Perceptual Coding Group*, Technical Report no. 220, 1994.

[25] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experiment," *IEEE Trans. PAMI*, vol. 16, no. 9, pp. 1207-1212, Sept. 1994.

[26] J. Konrad and E. Dubois, "Bayesian estimation of motion vector fields," *IEEE Trans. PAMI*, vol. 14, no. 9, pp. 910-927, Sept. 1992.

[27] S.Z. Li, "On discontinuity-adaptive smoothness priors in computer vision," *IEEE Trans. PAMI*, vol. 17, no. 6, pp. 576-586, Jun. 1995.

[28] J. Liu and R. Skerjanc, "Stereo and motion correspondence in a sequence of stereo images," *Signal Processing: Image Communication*, vol. 5, pp. 305-318, 1993.

[29] M.R. Luettgen, W.C. Karl, A.S. Willsky, and R.R. Tenney, "Multiscale representations of Markov random fields," *IEEE Trans. PAMI*, vol. 41, no. 12, pp. 3377-3396, Dec. 1993.

[30] A. Lundmark, N. Wadström, and H. Li, "Hierarchical subsampling giving fractal regions," Technical Report ISSN 1400-3902, LiTH-ISY-R-2035, Linköping University, Sweden, 1998.

[31] P. Maragos, "Morphological correlation and mean absolute error criteria," *Proc. ICASSP'89*, pp. 1568-1571, 1989.

[32] L. Matthies, T. Kanade, and R. Szeliski, "Kalman filter-based algorithms for estimating depth from image sequences," *Int. J. Computer Vision*, vol. 3, pp. 209-238, 1989.

[33] H.H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Trans. PAMI*, vol. 8, no. 5, pp. 565-593, Sept. 1986.

[34] J.-R. Ohm, K. Grueneberg, E. Hendriks, M.E. Izquierdo, D. Kalivas, M. Karl, D. Papadimatos, and A. Redert, "A real-time hardware system for stereoscopic videoconferencing with viewpoint adaptation," in *Signal Processing: Image Communication*, vol. 14, no.1-2, pp. 147-171, 1998.

[35] Y. and T. Kanade, "Stereo by intra- and inter-scanline search using dynamic programming," *IEEE Trans. PAMI*, vol. 7, no. 2, pp. 139-154, 1985.

[36] PANORAMA, European ACTS AC092 project, http://www.tnt.uni-hannover.de /project/eu/panorama/, 1995-1998.

[37] D.V. Papadimitriou and T.J. Dennis, "Epipolar line estimation and rectification for stereo image pairs," *Proc. of the International Workshop on Three-Dimensional Imaging*, pp. 128-133, 1995.

[38] I. Patras, E.A. Hendriks, and G. Tziritas, "Construction of multiple views using jointly estimated motion and disparity fields," *Proc. VCIP97*, SPIE vol. 3024, pp. 380-387, 1997.

[39] F. Pedersini, A. Sarti, and S. Tubaro, "Synthesis of virtual views using non-Lambertian reflectivity models and stereo matching," *Proc. ICIP95*, vol. 2, pp. 358-361, 1995.

[40] F. Pedersini, A. Sarti, and S. Tubaro, "3D surface reconstruction from horizons," *Proc. of the International Workshop on Synthetic-Natural Hybrid Coding and Three-Dimensional Imaging*, pp. 85-88, 1997.

[41] M. Pollefeys, R. Koch, M. Vergauwen, and L. Van Gool, "Flexible acquisition of 3D structure from motion," *Proc. IEEE Image and Multidimensional Digital Signal Processing Workshop '98*, pp. 195-198, 1998.

[42] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C, the Art of Scientific Computing*, Cambridge: Cambridge University Press, 1992.

[43] P.A. Redert, B. Kaptein, M.J.T. Reinders, I. Van den Eelaart, and E.A. Hendriks, "Extraction of semantic 3D models of human faces from stereoscopic image sequences," *Acta Stereologica*, Ljubljana, Slovenia, scheduled for publication.

[44] P.A. Redert, C.J. Tsai, E.A. Hendriks, and A.K. Katsaggelos, "Disparity estimation with modeling of occlusion and object orientation," *Proc. VCIP98*, SPIE vol. 3309, pp. 798-808, 1998.

[45] P.A. Redert, E.A. Hendriks, and J. Biemond, "Synthesis of multi-viewpoint images at non-intermediate positions," *Proc. ICASSP97*, pp. 2749-2752, 1997.

[46] P.A. Redert and E.A. Hendriks, "An efficient disparity map format for real-time interpolation in multi-viewpoint stereoscopic video systems," In *Proc. of the International Workshop on Synthetic-Natural Hybrid Coding and Three-Dimensional Imaging*, pp. 155-158, 1997.

[47] P.A. Redert and E.A. Hendriks, "Disparity estimation by searching parallel and orthogonal to the epipolar lines," *Proc. Fourth Annual Conference of the Advanced School for Computing and Imaging*, pp. 184-188, 1998.

[48] P.A. Redert and E.A. Hendriks, "Self calibration of stereo cameras with lens distortion," *Proc. IEEE Image and Multidimensional Digital Signal Processing Workshop '98*, pp. 163-166, 1998.

[49] D.P. McReynolds and D.G. Lowe, "Rigidity checking of 3D point correspondences under perspective projection," *IEEE Trans. PAMI*, vol. 18, no. 12, pp. 1174-1185, Dec. 1996.

[50] M.A. Snyder, "On the mathematical foundations of smoothness constraints for the determination of optical flow and for surface reconstruction," *IEEE Trans. PAMI*, vol. 13, no. 11, pp. 1105-1114, Nov. 1991.

[51] E. Steinbach, S. Chaudhuri, and B. Girod, "Robust estimation of multi-component motion in image sequences using the epipolar constraint," *Proc. ICASSP97*, pp. 2689-2692, 1997.

[52] C. Stiller, "Object-based estimation of dense motion fields," *IEEE Trans. Image Processing*, vol. 6, no. 2, pp. 234-250, Feb. 1997.

[53] A.M. Tekalp, *Digital Video Processing*, Englewood Cliffs, NJ: Prentice Hall, 1995.

[54] Section 8.3 of [53].

[55] C.W. Therrien, *Discrete Random Signal and Statistical Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, 1992.

[56] E. Trucco and A. Verri, *Introductory Techniques for 3D Computer Vision*, Englewood Cliffs, NJ: Prentice Hall, 1998.

[57] C.J. Tsai and A.K. Katsaggelos, "Optical flow estimation for multi-channel video sequences," *Proc. VCIP97*, SPIE vol. 3024, pp. 360-368 , 1997.

[58] B.L. Tseng and D. Anastassiou, "A theoretical study on an accurate reconstruction of multiview images based on the Viterbi algorithm," *Proc. ICIP95*, pp. 378-381, 1995.

[59] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Trans. PAMI*, vol. 14, no. 10, pp. 965-980, Oct. 1992.

[60] W. Woo and A. Ortega, "Stereo image compression with disparity compensation using the MRF model," *Proc. VCIP96*, SPIE vol. 2727, no. 1, pp. 28-41, 1996.

[61] J. Zhang and J. Hanauer, "The mean field theory for image motion estimation," *Proc. ICASSP93*, vol. 5, pp. 197-200, 1993.