A Scale Selection Principle for Estimating Image Deformations

*Tony Lindeberg**

Computational Vision and Active Perception Laboratory (CVAP),
Dept. of Numerical Analysis and Computing Science,
KTH (Royal Institute of Technology),
S-100 44 Stockholm, Sweden.

http://www.nada.kth.se/~tony
Email: tony@nada.kth.se

Tech. Rep. ISRN KTH/NA/P–96/16–SE, May 1996, Revised August 1998.

Shortened version in Proceedings 5th International Conference on Computer Vision, (Cambridge, MA), June 1995, pages 134–141. Image and Vision Computing (to appear)

Abstract

A basic functionality of a vision system concerns the ability to compute deformation fields between different images of the same physical structure. This article advocates the *need for incorporating an explicit mechanism for scale selection* in this context, in algorithms for computing descriptors such as *optic flow* and for performing *stereo matching*. A basic reason why such a mechanism is essential is the fact that in a coarse-to-fine propagation of disparity or flow information, it is not necessarily the case that the most accurate estimates are obtained at the finest scales. The existence of interfering structures at fine scales may make it impossible to accurately match the image data at fine scales.

A systematic methodology for approaching this problem is proposed, by estimating the uncertainty in the computed flow estimate at each scale, and then *selecting deformation estimates from the scales that minimize the (suitably normalized) uncertainty over scales*. A specific implementation of this idea is presented for a region based differential flow estimation scheme. It is shown that the integrated scale selection and flow estimation algorithm has the qualitative properties of leading to the *selection of coarser scales for larger size image structures and increasing noise level*, whereas it leads to the *selection of finer scales in the neighbourhood of flow field discontinuities*. The latter property may serve as an indicator when detecting flow field discontinuities and occlusions.

I would like to thank D. Betsis and G. Orban for valuable discussions as well as J. Koenderink and A. van Doorn for providing the torso image in figure 8. This work was partially performed under the Esprit-BRA project InSight and the Esprit-NSF collaboration Diffusion. The support from the Swedish Research Council for Engineering Sciences, TFR, is gratefully acknowledged.

1. Introduction

In several computational visual models, the deformations of brightness patterns constitute an important modelling step. When a camera fixates a surface pattern in the world, the pattern is deformed when mapped onto the camera by the perspective transformation. Depending on the external conditions under which the image data are acquired, the following characteristic situations can be distinguished:

- For a monocular camera that observes the world over time, the definition of a point-to-point correspondence between physical points in the world gives rise to an *optic flow field*.
- For a binocular vision system observing the world from two directions, the definition of a point-to-point correspondence between physical points in the two images gives rise to a *disparity field*.
- In a monocular image of a textured surface, the perspective mapping affects the statistical properties of the texture, which gives rise to *texture gradients*.

The structure of the deformation fields formed in these ways depend on the shape of the object as well as the position and the orientation of the object relative to the observer. Hence, these deformations contain essential geometric information, and a large number of visual modules can be expressed in terms of this framework. Some examples are motion estimation, structure from motion, stereo matching, vergence control, shape estimation from binocular data, and shape from texture.

In general, the geometry of these deformation fields can be modelled by projective transformations. Approximating the projective model by a local first-order approximation (the derivative) gives rise to an affine transformation. With respect to a shape-from-X problem, the interpretation of the zero order component (the translation) of this transformation will typically be that it reveals local depth, whereas the first-order component can typically be used for inferring local surface orientation. To capture higher order shape properties, such as local surface curvature, higher order approximations of the deformation field will usually be required.

The subject of this article is to consider the problem of measuring such local deformations between two-dimensional brightness patterns. Specifically, we will be concerned with the need for a mechanism for automatic scale selection that arises in this context. Traditional methods for flow estimation and stereo matching are usually formulated in a coarse-to-fine hierarchical manner, in which approximate deformation estimates are first computed at coarse scales, and then refined iteratively by computations at finer scales. This approach is usually motivated by the fact that the matching problems are usually simpler at coarse scales, whereas more accurate estimates can be obtained at fine scales. There are also strong psychophysical evidence supporting coarse-to-fine processing in human vision.

In such a coarse-to-fine propagation, however, it is not necessarily the case that the most accurate estimates will be obtained at the finest scales. (Who determines what should be the finest scale?) For example, if we apply a differential flow algorithm at a fine scale, the computation of derivatives can be highly sensitive to spurious noise. Similar problems may arise when performing correlation matching using too small window sizes. For this reason, we argue that it is essential to complement any coarse-to-fine algorithm for estimating image deformations (such as optic flow and stereo matching algorithms) by an explicit mechanism for automatic scale selection. A main requirement on this scale selection mechanism should be the ability to suppress deformation estimates that have been computed at too fine scales and cannot be regarded as reliable.

To approach this problem in a systematic manner, the following methodology is proposed: In the coarse-to-fine propagation, an appropriately normalized measure of uncertainty should be computed and accompany the deformation estimate at each scale. Then, once the coarse-to-fine propagation has been performed, the algorithm should select the deformation estimate from the scale at which the measure of uncertainty assumes its minimum over scales and deliver this information as output. (If motivated by efficiency considerations, this information could also be used for turning off the coarse-to-fine propagation when the uncertainty starts to increase.)

A specific implementation of this idea will be presented based on the the class of region based differential flow estimation schemes proposed by (Lukas and Kanade 1981; Bergen *et al.* 1992). To simplify the presentation, and to avoid making distinctions between different types of deformation fields, we shall throughout this article develop the ideas in the binocular case with two image domains, and therefore often use the term "disparity" to refer to the translation component of the image deformation. With appropriate modifications, however, the general idea presented here, of selecting scale levels for computing image deformations from the scales that minimize the estimated uncertainty over scales, applies to a much wider class of algorithms for performing stereo matching and computing optic flow.

Before starting, let us also point out that because of the generality of this problem domain, these problems have been extensively studied in the computer vision literature, and it is impossible to give a complete review here. Besides the references that will be explicitly cited, the reader is referred to the recent overview by (Barron *et al.* 1994) and the references therein. An earlier version of this manuscript has been presented in (Lindeberg 1994a).

2. Measuring image deformations

A common approach to stereo matching and the computation of three-dimensional shape cues has been to compute image features, such as points and lines, in an initial processing step, and then to use these descriptors as primitives for the subsequent processing steps. Whereas a substantial simplification of the subsequent

processing stages may be the result if reliable image features can be extracted, the selection of what image features to use crucially determines what results can be obtained and is often non-trivial. Therefore, it is of interest to consider methods that operate on the image intensities directly, using only filter-based operations and architecturally simple combinations of their outputs.

A fundamental problem in this context concerns what image operations to use. Is any operation feasible? A systematic approach that has been developed to restrict the class of possibilities is to assume that the first stages of visual processing should be as *uncommitted* as possible and have no particular bias. The essence of the results from *scale-space theory* (Witkin 1983; Koenderink 1984; Yuille and Poggio 1986; Koenderink and van Doorn 1990; Florack *et al.* 1992; Lindeberg 1994d) is that within the class of linear operations, convolution with Gaussian kernels and their derivatives is singled out as a canonical choice.

In this section, we shall use this framework for expressing a hierarchical differential flow field estimation algorithm closely related to (Bergen *et al.* 1992); see also (Werkhoven and Koenderink 1990; Jones and Malik 1992; Proesmans *et al.* 1994; Manmatha 1994; Sato and Cipolla 1994). We start by outlining the components in a multi-scale disparity estimation framework, which in addition to iterative corrections comprises bidirectional matching and explicit usage of confidence measures. Then, we turn to the problem of incorporating a mechanism for scale selection.

2.1. Deformation measurements in scale-space

The linear scale-space representation of a signal $f\colon \mathbb{R}^2 \mapsto \mathbb{R}$

$$L(\cdot;\ t) = g(\cdot;\ t) * f \tag{1}$$

is obtained by convolving f with Gaussian kernels $g\colon \mathbb{R}^2 \times \mathbb{R}_+ \mapsto \mathbb{R}$

$$g(x;\ t) = \frac{1}{2\pi t} e^{-x^T x/2t}, \tag{2}$$

of different standard deviations $\sqrt{t}$. From this representation, Gaussian derivatives are then defined by $L_{x^\alpha}(\cdot;\ t) = \partial_{x^\alpha} L(\cdot;\ t)$ where $\partial_{x^\alpha} = \partial_{x_1^{\alpha_1}} \partial_{x_2^{\alpha_2}}$.

Transformations in the similarity group. This representation is closed under transformations in the similarity group, *i.e.*, if two signals are related by

$$f_L(\xi) = f_R(\sigma \mathcal{R}_\varphi \xi + \Delta x), \tag{3}$$

where $\mathcal{R}_\varphi$ is a rotation matrix, σ represents a positive scaling factor, and Δx represents a translation, then the scale-space representations of f_L and f_R are related by

$$L(\xi;\ t) = R(\sigma \mathcal{R}_\varphi \xi + \Delta x;\ \sigma^2 t). \tag{4}$$

Hence, for similarity transformations, the scale-space representations of f_L and f_R can always be perfectly matched, and assuming that the brightness variations in f_L and f_R are sufficiently rich, such that ambiguities do not occur, it is, in principle, possible to measure similarity transformations exactly.

Affine image transformations and affine scale-space. To capture higher order deformation fields without bias introduced by the image operations, some extensions are required. Concerning affine transformations, a natural generalization is to use the affine Gaussian scale-space concept based on convolution by non-symmetric Gaussian kernels (Lindeberg 1994d; Lindeberg and Gårding 1994). Alternatively, one could conceive using non-linear affine invariant evolution schemes, such as those proposed by (Sapiro and Tannenbaum 1993; Alvarez *et al.* 1993).

3. Establishing correspondence

A fundamental problem that arises when estimating image deformations concerns how to establish correspondence between different images of the same scene. Whereas the commonly used constant brightness assumption suffers from inherent limitations, we shall nevertheless use it for establishing an initial correspondence. (Then, it can be applied to other differential descriptors, such as the Laplacian.) Hence, assume

$$f_R(\xi) = f_L(\xi + \Delta\xi) = f_L(\xi) + (\nabla f_L)(\xi)\,\Delta\xi + \mathcal{O}(|\Delta\xi|^2)$$

and let us consider only the first-order terms in the local Taylor expansion. This gives rise to (the discrete form of) the well-known motion constraint equation (Horn and Schunck 1981)

$$(\nabla f_L)(\xi)^T(\Delta\xi) + (f_L(\xi) - f_R(\xi)) = \mathcal{O}(|\Delta\xi|^2).$$

Since this analysis is compatible with brightness measurements in scale-space, at any scale t we also have

$$(\nabla L)(\xi;\ t)^T(\Delta\xi) + (L(\xi;\ t) - R(\xi;\ t)) = \mathcal{O}(|\Delta\xi|^2).$$

Least-squares estimation. Assume next that the motion field can be approximated by a constant flow field v over the support region of a window function w. Then, following (Lukas and Kanade 1981; Bergen *et al.* 1992; Barron *et al.* 1994) and several others, integrate the square of this relation using w as window function. This transforms the problem of determining $\Delta\xi$ into the least squares problem

$$\min_{v\in\mathbb{R}^2} \int_{\xi\in\mathbb{R}^2} ((\nabla L)(\xi;\ t)^T v - (L(\xi;\ t) - R(\xi;\ t))^2\, w(\xi)\, d\xi. \tag{5}$$

After expanding the square, exploiting the fact that v is regarded as constant, and dropping the arguments, this problem can be written

$$\min_{v \in \mathbb{R}^2} v^T A v + 2 b^T v + c, \tag{6}$$

where A, b, and c are defined by

$$A = \int_{\xi \in \mathbb{R}^2} (\nabla L)(\nabla L)^T \, w \, d\xi, \tag{7}$$

$$b = \int_{\xi \in \mathbb{R}^2} (R - L) \, (\nabla L) \, w \, d\xi, \tag{8}$$

$$c = \int_{\xi \in \mathbb{R}^2} (R - L)^2 \, w \, d\xi. \tag{9}$$

Ambiguity. When treated pointwise, the motion constraint equation only determines the normal flow parallel to ∇L. On the other hand, if the support region of w contains a sufficiently rich distribution of gradient directions, and if this region moves in a coherent way, which can be approximated by a constant flow field, the solution to (6) may give an estimate close to the true flow field. A natural measure of how scattered the gradient directions are is given by the normalized anisotropy (derived from components a_{ij} from the matrix A)

$$\tilde{Q} = \frac{\sqrt{(a_{11} - a_{22})^2 + 4 a_{12}^2}}{a_{11} + a_{22}}. \tag{10}$$

When all gradient directions are parallel, we have $\tilde{Q} = 1$, whereas $\tilde{Q} = 0$ for maximally scattered distributions. Hence, the indeterminacy in the tangential component of v can be expected to increase with $\tilde{Q}$.

Closed-form solution. Assuming that A according to (7) is non-degenerate, the explicit solution of (6) is

$$v = -A^{-1} b \tag{11}$$

and the residual

$$r = c - b^T A^{-1} b. \tag{12}$$

If A is singular, or close to singular, it is preferable to use the pseudo inverse. For a symmetric two-dimensional matrix of rank one, it is given by

$$A^{\dagger} = \frac{1}{(\mathrm{trace}\, A)^2} A. \tag{13}$$

The pseudo inverse is preferred when the singular values are sufficiently different, or equivalently the normalized anisotropy is sufficiently close to one.

In practice, the window function is chosen as a Gaussian kernel, since then and only then the components of A satisfy scale-space properties under variations

of the scale parameter s of w. (These propagate to the distribution of gradient directions described by A as a composed object.) Concerning the relation between integration scale s for averaging and the local scale t for computing derivatives, one should, in principle, consider a two-parameter variation. In the experiments to be presented, we have throughout used $s = s(t) = \gamma^2 t$ with $\gamma = 2$.

4. Hierarchical and iterative flow field computations

By using scale-space operators at a certain scale t, it is, in general, only possible to capture disparities of the same order of magnitude as $\sqrt{t}$.

- If the disparity is much larger than the local scale, then there may be interfering fine-scale structures between a certain position ξ in one image and its corresponding point $\xi + \Delta\xi$ in the other image, which bias the local linearization.
- If the local scale on the other hand is too large relative to the disparity update, then the shape distortions may have substantial negative influence and lead to unnecessarily inaccurate disparity estimates.

This motivates a coarse-to-fine approach, where initial disparity estimates are computed at coarse scales, propagated to finer scales, and disparity updates then are computed iteratively at finer scales.

When computing the iterative updates, the current disparity estimate $v^{(k)}$ should of course be taken into account when computing the brightness difference $R(\xi_R;\ t) - L(\xi_L;\ t))$ (where ξ_R and ξ_L are related by $\xi_R = \xi_L + v_L(\xi_L;\ t)$ and $\xi_L = \xi_R + v_R(\xi_R;\ t)$) so as to reduce the approximation error in the local linearization. Hence, at any point x, disparity estimates are updated according to

$$v_L^{(k+1)}(x_L;\ t) = v_L^{(k)}(x_L;\ t) + \Delta v_L^{(k+1)}(x_L;\ t) \tag{14}$$

where

$$\Delta v_R^{(k+1)}(x_R;\ t) = \Big(\int_{\xi_L \in \mathbb{R}^2} ((\nabla L)(\xi_L;\ t))((\nabla L)(\xi_L;\ t))^T w_{x_L}(\xi_L;\ s(t))\, d\xi_L \Big)^{\sim 1} \int_{\xi_L \in \mathbb{R}^2} (R(\xi_R;\ t) - L(\xi_L;\ t))\, ((\nabla L)(\xi_L;\ t))\, w_{x_L}(\xi_L;\ s(t))\, d\xi_L, \tag{15}$$

and this updating may, in principle, proceed until a fixed-point has been reached or the scale-space representations of f_L and f_R are in sufficient alignment.

If the transformation is not locally a pure translation, a higher order (*e.g.*, affine) model is required to reduce the approximation error, and corresponding compensations are needed when computing the brightness differences. These iterations can be driven either by the affine Gaussian scale-space representation and shape adaptation or by performing local warping and solving an extension of (6) with the locally constant flow model replaced by a local affine.

4.1. Bidirectional matching and inconsistency measures

The previous matching scheme can be applied in both directions, which gives rise to independent flow field estimates. A natural inconsistency measure is then

$$E_L(x_L;\, t) = v_L(x_L;\, t) + v_R(x_L + v_L(x_L;\, t);\, t), \tag{16}$$

$$E_R(x_R;\, t) = v_R(x_R;\, t) + v_L(x_R + v_R(x_R;\, t);\, t), \tag{17}$$

and a natural measure of the strength of the response

$$R_L(x_L;\, t) = P_L(x_L;\, t)\, P_R(x_L + v_L(x_L;\, t);\, t), \tag{18}$$

$$R_R(x_R;\, t) = P_R(x_R;\, t)\, P_L(x_R + v_R(x_R;\, t);\, t), \tag{19}$$

where $P_L = \operatorname{trace} A_L$ and $P_R = \operatorname{trace} A_R$ denote the average square gradient magnitudes computed in the left and right images, respectively.

5. Formulation of a confidence measure

When to formulate a confidence measure for a flow field computed at a certain scale, it is natural to state the following qualitative requirements:

- Operator strength: The confidence measure should be higher in regions with large brightness variations than in regions with slowly varying grey-levels.
- Mutual consistency: The confidence measure should decrease with the inconsistency between flow fields computed independently in the two images.
- Matching error: The confidence measure should decrease with the error in the least-squares alignment.

In addition, to treat all scales in a uniform manner, we may also require the following:

- Scale invariance: The confidence measure should be invariant under uniform rescalings of the input images.

Based on these general arguments, we can decide to:

- measure the strength of the operator response by R,
- measure the inconsistency by an exponentially decreasing function of E, and
- measure the error in the least-squares alignment by the normalized residual $\tilde{r}$.

Finally, scale invariance can be accomplished by

- measuring $P = \text{trace}\, A$ in R in terms of normalized derivatives $\partial_{\xi_i} = \sqrt{t}\, \partial_{x_i}$, where t represents the local scale parameter for differentiation, and
- measuring the inconsistency E and the normalized residual relative to the current level of scale. Since the latter entities both have dimension [length]^2, scale normalization can be easily achieved by dividing these entities by t.

In summary, these entities can be combined in the following (heuristically chosen) confidence measures:

$$W_L(x_L;\, t) = \frac{R_{L,norm}(x_L;\, t)\, e^{-\omega E_L^2(x_L;\, t)/t}}{\tilde{r}_0 + \tilde{r}(x_L;\, t)}, \tag{20}$$

$$W_R(x_L;\, t) = \frac{R_{R,norm}(x_L;\, t)\, e^{-\omega E_R^2(x_L;\, t)/t}}{\tilde{r}_0 + \tilde{r}(x_R;\, t)}, \tag{21}$$

where the following parameters have been introduced:

- ω (here, ≈ 0.1) determines how large disparity inconsistencies are tolerated relative to the current level of scale, and
- $\tilde{r}_0$ (here, ≈ 0.01) is a non-essential threshold to avoid divisions by zero.

Obviously, this confidence measure can be expected to be small for occluded points. For points whose disparity vectors point outside the available image data, the confidence measure is set to zero in the current implementation of the algorithm.

5.1. Flow field correction and flow field smoothing

At spurious points, it may happen that the disparity estimates according to (15) contain large errors due to noise or the fact the support region of w_x does not contain a sufficiently rich distribution of image structures. To suppress such errors, only disparity updates satisfying

$$|v^{(k+1)}(x;\, t) - v^{(k)}(x;\, t)| < \nu\sqrt{t} \tag{22}$$

are allowed to propagate unaffected (here ν is of the order of 2). Larger updates are truncated. Moreover, in each iteration, the flow field is smoothed using the confidence values W as weights

$$v'(x;\, t) = \frac{\int_{\xi \in \mathbb{R}^2} v(\xi;\, t)\, W(\xi;\, t)\, w_x(\xi;\, s(t))\, d\xi}{\int_{\xi \in \mathbb{R}^2} W(\xi;\, t)\, w_x(\xi;\, s(t))\, d\xi}. \tag{23}$$

This leads to a rapid propagation of disparities from regions with strong intensity variations (typically edges and textured regions where the disparity information can be expected to be maximally reliable) to the interior of smooth regions. Moreover, locally inconsistent disparity estimates as well as spurious deviations are likely to be suppressed, since these can be expected to be assigned low confidence values.

6. Proposed framework for automatic scale selection

Within this framework, disparity estimates can be computed at any scale, using conceptually simple front-end operations. A fundamental problem, however, concerns how to combine the information from different scales. Selecting disparity estimates from the finest scale after a coarse-to-fine hierarchical propagation is not guaranteed to be the best solution, since the flow field estimates at the finest scales can be highly sensitive to noise and other interfering fine-scale structures. In practice, there may be substantial variations within as well as between the disparity fields at the finer scales. Hence, for a flexible vision system required to handle a large number of different situations without any *a priori* information about what are the proper scales to use in a given situation, it is of crucial importance to complement this coarse-to-fine flow estimation framework by an explicit mechanism for scale selection.

Intuitively, such a scale selection mechanism should select coarse-scale disparity fields from noisy data, for which fine-scale correspondences may be impossible to establish. Correspondingly, it should select fine-scale representatives from the disparity fields from sharp data that contain detailed information, so as to produce a maximally accurate disparity field.

Scale selection method. To measure how well two image regions have been aligned by the local least-squares fit (5), it is natural to consider the residual (12). This entity does, however, not contain sufficient information for making such judgements, since it depends upon the local contrast. A straightforward but nevertheless powerful approach is to *select the scale that minimizes the normalized residual*

$$\tilde{r} = \frac{r}{\mathrm{trace}\, A} = \frac{c - b^T A^{-1} b}{\mathrm{trace}\, A} \tag{24}$$

over scales. A basic motivation for dividing the regular residual (12) by the trace of the covariance matrix is that this operation cancels the effect of the amplitude of local brightness variations. Since the dimensions involved are as follows:

Entity	Dimension
A	$[\mathrm{luminance}]^2/[\mathrm{length}]^2$
b	$[\mathrm{luminance}]^2/[\mathrm{length}]$
c	$[\mathrm{luminance}]^2$

it follows that the normalized residual has dimension $[\mathrm{length}]^2$ and can be seen as reflecting a spatial error in the disparity estimate. (A statistical analysis and interpretation of this measure is given in appendix A.)

Concerning the specific selection of normalization factor, $\mathrm{trace}\, A$ constitutes a natural choice, since it is a well-defined function of a differential invariant (the weighted average of the square gradient magnitude in a local neighbourhood).

Qualitative effects. Relating to the abovementioned intuitive requirements, the qualitative effects of this scale selection method are as follows:

- At too coarse scales, a uniform deformation model cannot be expected to hold over the entire region. Also, the shape distortions can be expected to be stronger, thereby increasing the normalized residual.

- At too fine scales, where noise and other fine-scale structures are present, the likelihood that these structures obey the same motion model will be low. Hence, the normalized residual can be expected to increase.

Selecting the minimum leads to a natural trade-off between these effects.

7. Qualitative properties of the scale selection method

Let us now consider the effect of applying this scheme to image data. Figures 1–2 illustrate general properties of the scale selection method when applied to smoothly deformed patterns for which interfering structures are present at finer scales.

Selection of coarser scales for larger size image structures. In figure 1, two synthetic image patterns have been subject to a uniform expansion. The underlying patterns are identical except for the size of the texture elements which differs by a factor of four, and 10 % white Gaussian noise[a] added to each image independently after the deformation. For each pattern, the normalized residual has been computed at different scales for all points in an 8×8 window at the center of the image (of size 64×64 pixels). Such a graph is referred to as the *scale-space signature* of $\tilde{r}$ (Lindeberg 1994c).

Observe that besides small fluctuations due to noise at the finer scales, these signatures become more stable at coarse scales. More importantly, with increasing size of the image structures, the minimum over scales is assumed at coarser scales. This behaviour agrees with the intuitive notion that coarser scales should be selected for patterns containing larger size image structures.

Selection of coarser scales with increasing noise level. In figure 2 the image pattern is the same, whereas the noise level is varied. Observe that with an increasing amount of interfering fine scale structures, the minimum in $\tilde{r}$ over scales is assumed at coarser scales. This behaviour agrees with the intuitive notion that a larger amount of smoothing is required for noisy data than otherwise similar data with less noise.

[a] Throughout this article, we measure the noise level by the ratio between the standard deviation of the noise and the amplitude of the signal.

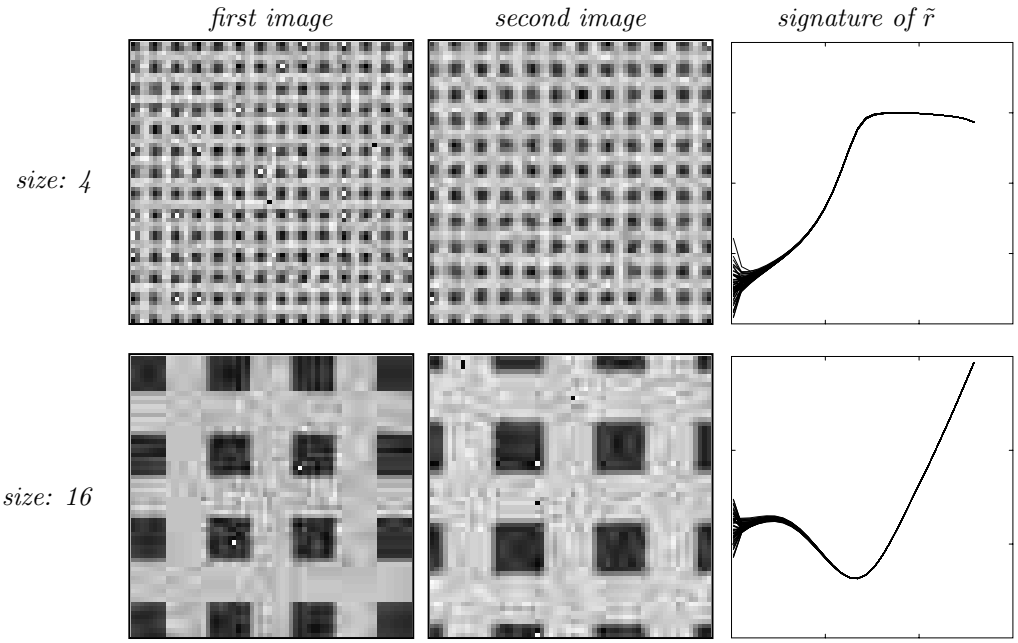

Figure 1: Scale-space signatures of the normalized residual $\tilde{r}$ computed from synthetic expanding patterns with structures at different scales. Notice that with increasing size of the texture elements, the minimum over scales in the normalized residual is assumed at coarser scales.

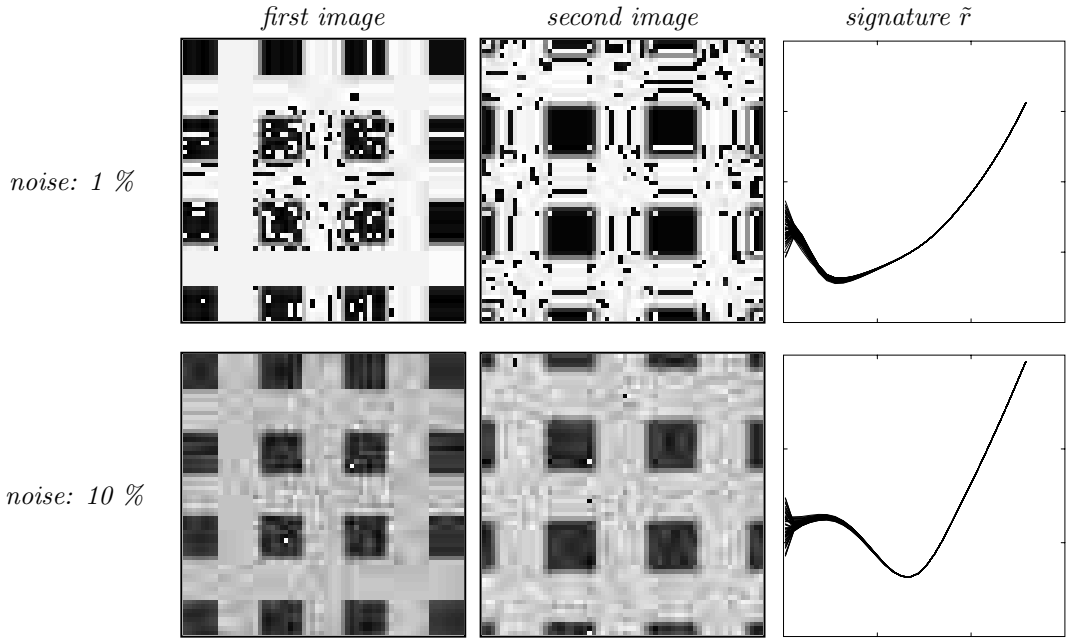

Figure 2: Scale-space signatures of the normalized residual $\tilde{r}$ computed for a synthetic expanding pattern with different amounts of added white Gaussian noise. Observe that with increasing noise level, the minimum over scales in the normalized residual is assumed at coarser scales.

Selection of finer scales near discontinuities in the deformation field. Figure 3 shows the qualitative behaviour of the scale selection method in the neighbourhood of a discontinuity in the flow field. For a "wedding-cake type" random dot stereo pair (consisting of 64×64 dots in a 256×256 image with zero disparity in a central square and uniform non-zero disparity in the periphery) to which 1% white Gaussian noise has been added, the results are shown of accumulating the scale-space signature of the normalized residual in three windows with different distance to the discontinuity. These windows have been uniformly spaced from the image center to one of the discontinuities in the disparity field as shown in figure 3(c).

Observe that with decreasing distance to the discontinuity, the minimum over scales is assumed at finer scales. This qualitative behaviour agrees with the intuitive notion that smaller windows for matching should be selected for image structures near a discontinuity in the disparity field than when matching otherwise similar image structures in a region where the disparity varies smoothly.

Notably, this rapid decrease of the selected scale levels could also provide a clue for detecting flow field discontinuities and signalling possible occlusions.

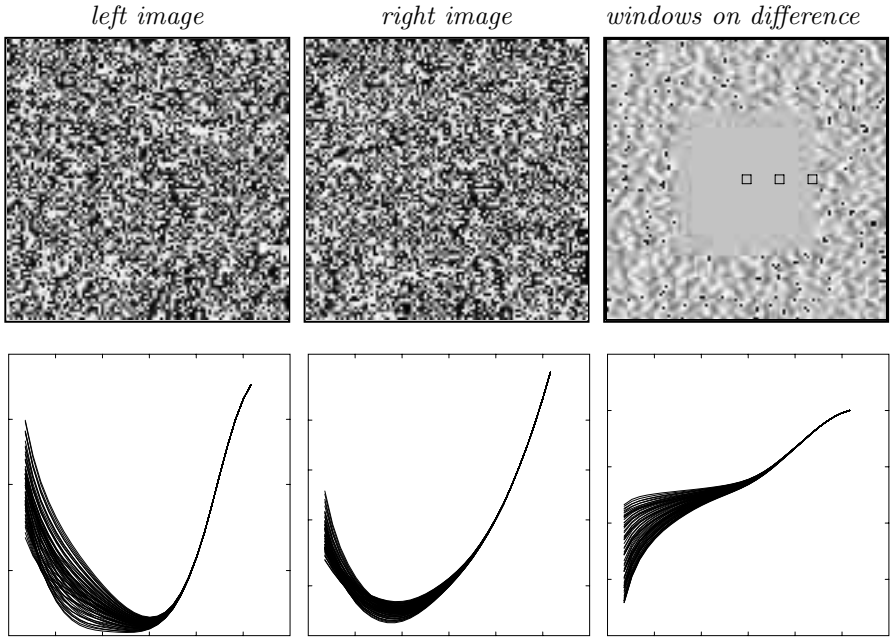

Figure 3: The qualitative behaviour of the scale selection method at a discontinuity in the deformation field. The bottom row shows scale-space signatures of the normalized residual computed in three windows with different distance to the discontinuity (with positions indicated in the upper right image showing the pointwise difference between the two images). Observe that with decreasing distance to the discontinuity, the minimum over scales is assumed at finer scales.

8. Experiments: Estimation of image deformations

Figures 4–5 show the result of applying the composed flow estimation scheme to a synthetic pattern which has been transformed by a pure rotation and a pure expansion, respectively. (In analogy, with previous experiments, 10% white Gaussian noise has been added to each image after the deformation.) To illustrate the consistency between the estimates computed in the two directions, both estimates are shown. Notice first how the qualitative shape of the flow field is captured and that the compensated differences between the images (the result of computing $R(\xi + v_L^{(k)}(x_L;\ t);\ t) - L(\xi;\ t)$ and $L(\xi + v_R^{(k)}(x_R;\ t);\ t) - R(\xi;\ t)$ according to (15)) correspond to errors comparable to the noise level. Moreover, the significance measure is high (represented as dark) in the regions where the computed flow field estimates are valid.

Figure 6 shows corresponding results for a "wedding-cake" random dot stereo pair similar to the one in figure 3. Observe that except for a few points near the discontinuity, the flow compensated differences between the left and right images have been driven down to essentially zero. Moreover, the significance measure is very low near the discontinuities in the disparity field.

Figure 7 shows corresponding results for a detail of a head subject to a rather large (unknown) rotation. Note that except for the upper right corner, where most points either correspond to occluded points or points outside the image, a correct matching has been established without any use of epipolar geometry.

9. Relations to previous work

The least squares formulation (5) for computing optic flow was proposed by (Lukas and Kanade 1981) and has been applied by several researchers; see (Bergen *et al.* 1992; Barron *et al.* 1994) for overviews. The idea of using local image deformations as a primary cue for shape estimation and motion analysis goes back to (Koenderink and van Doorn 1975, 1976) and has been extended in several directions. Besides the references given previously, further examples include (Werkhoven and Koenderink 1990; Koenderink and van Doorn 1991; Arnspang 1991; Jones and Malik 1992; Malik and Rosenholtz 1993; Proesmans *et al.* 1994; Devernay and Faugeras 1994; Sato and Cipolla 1994; Gårding and Lindeberg 1996; van Ee 1995). Whereas a large class of approaches have been developed to compute these deformation fields, most previous algorithms operate either at a fixed scale, or in terms of coarse-to-fine propagation between two given scale levels. A notable exception is the interesting work by (Kanade and Okutomi 1994), who present an algorithm for adapting the window size of a correlation based stereo algorithm based on a statistical model of the disparity field.

The present work aims at addressing the general problem of scale selection that arises in this context and to unify the abovementioned ideas into a coherent framework for flow estimation with integrated scale selection. Specifically, the least-squares formulation in (Lukas and Kanade 1981) has been extended to

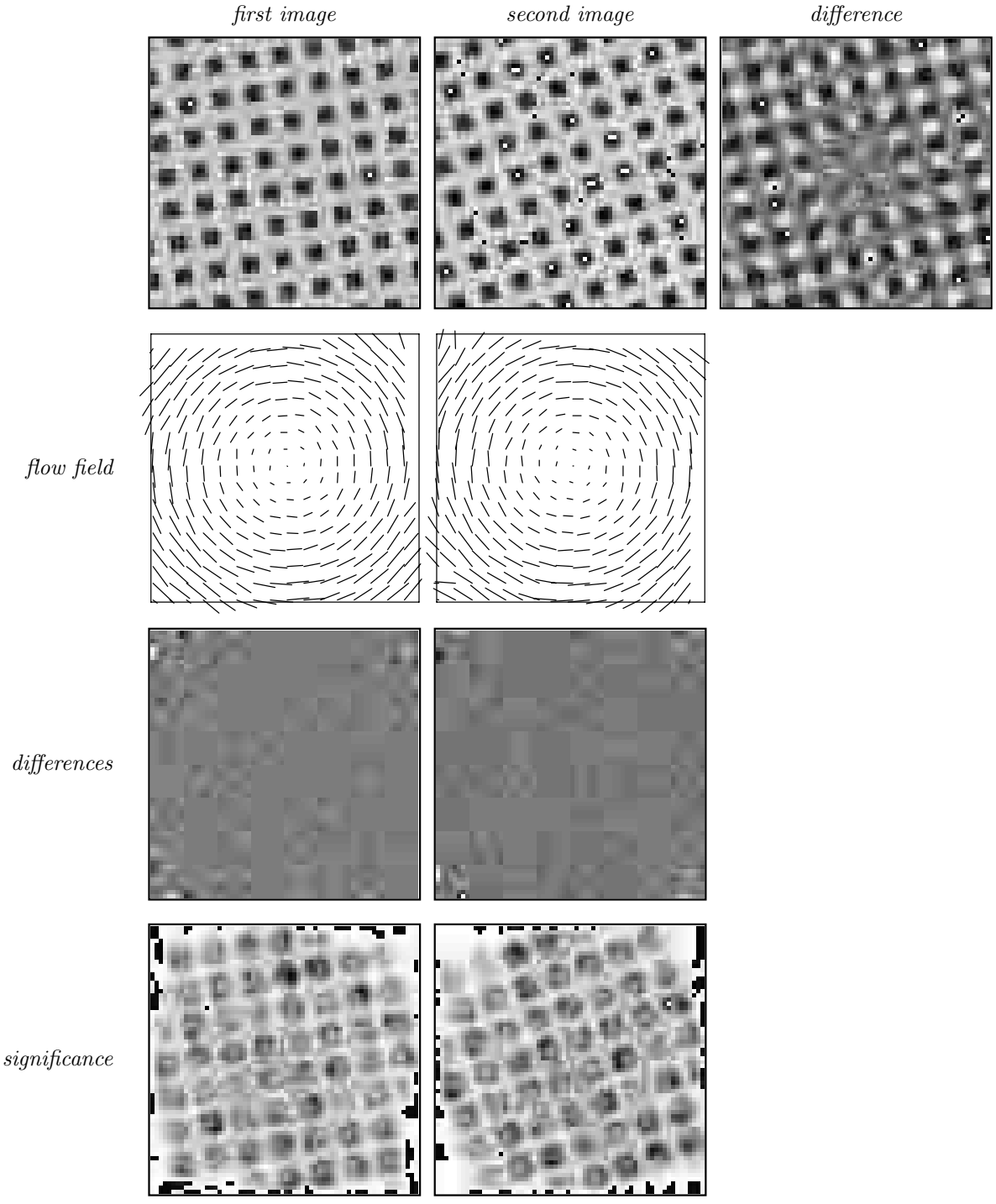

Figure 4: Flow field estimates computed from a synthetic flow field using the proposed scheme with automatic scale selection. From the top row to the bottom row, the images show respectively (a) the original image pair, (b) estimated flow field, (c) compensated differences, (d) significance measure. (Image size: 64*64 pixels. Noise level: 10 %).

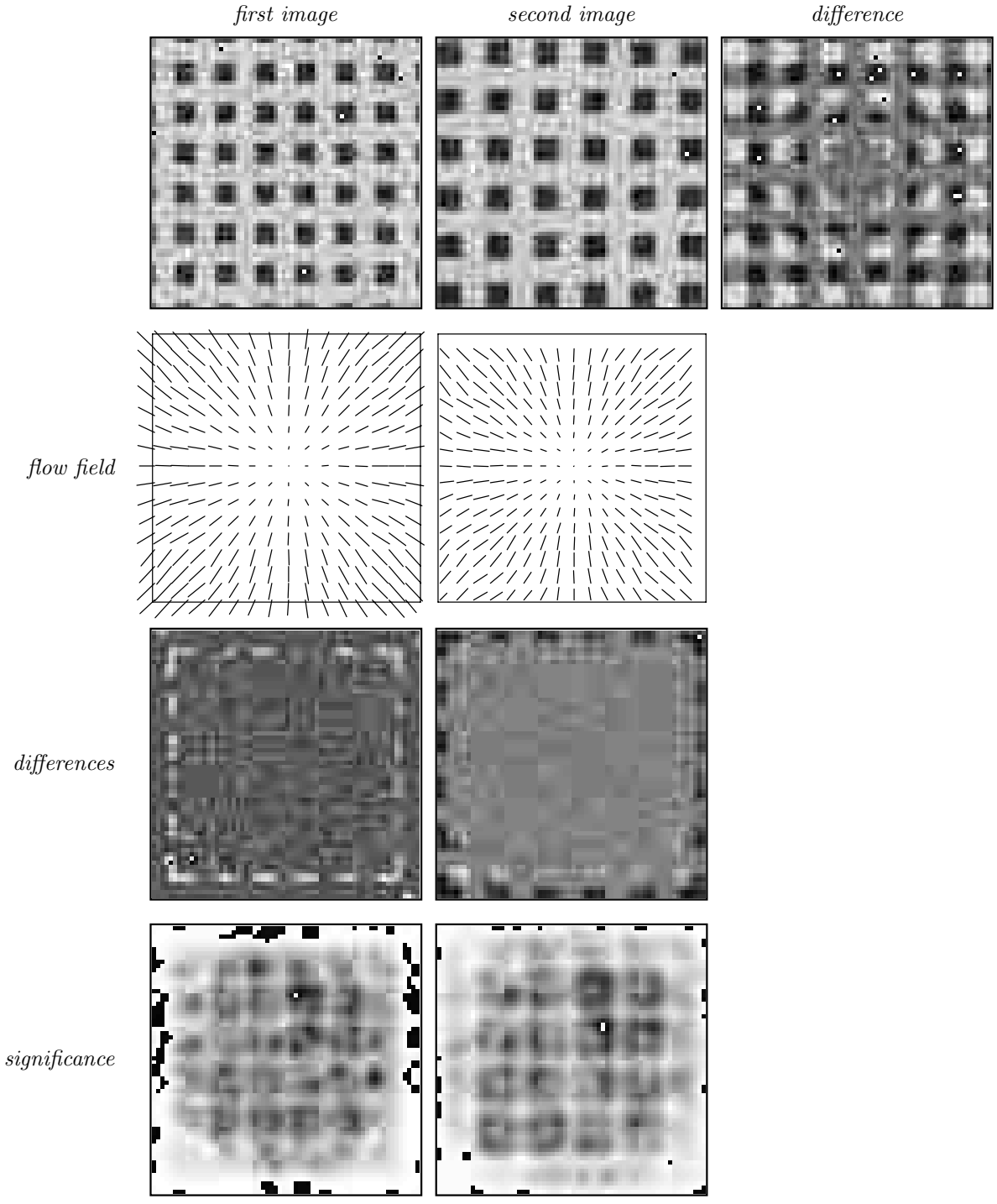

Figure 5: Flow field estimates computed from a synthetic flow field using the proposed scheme with automatic scale selection. From the top row to the bottom row, the images show respectively (a) the original image pair, (b) estimated flow field, (c) compensated differences, (d) significance measure. (Image size: 64*64 pixels. Noise level: 10 %).

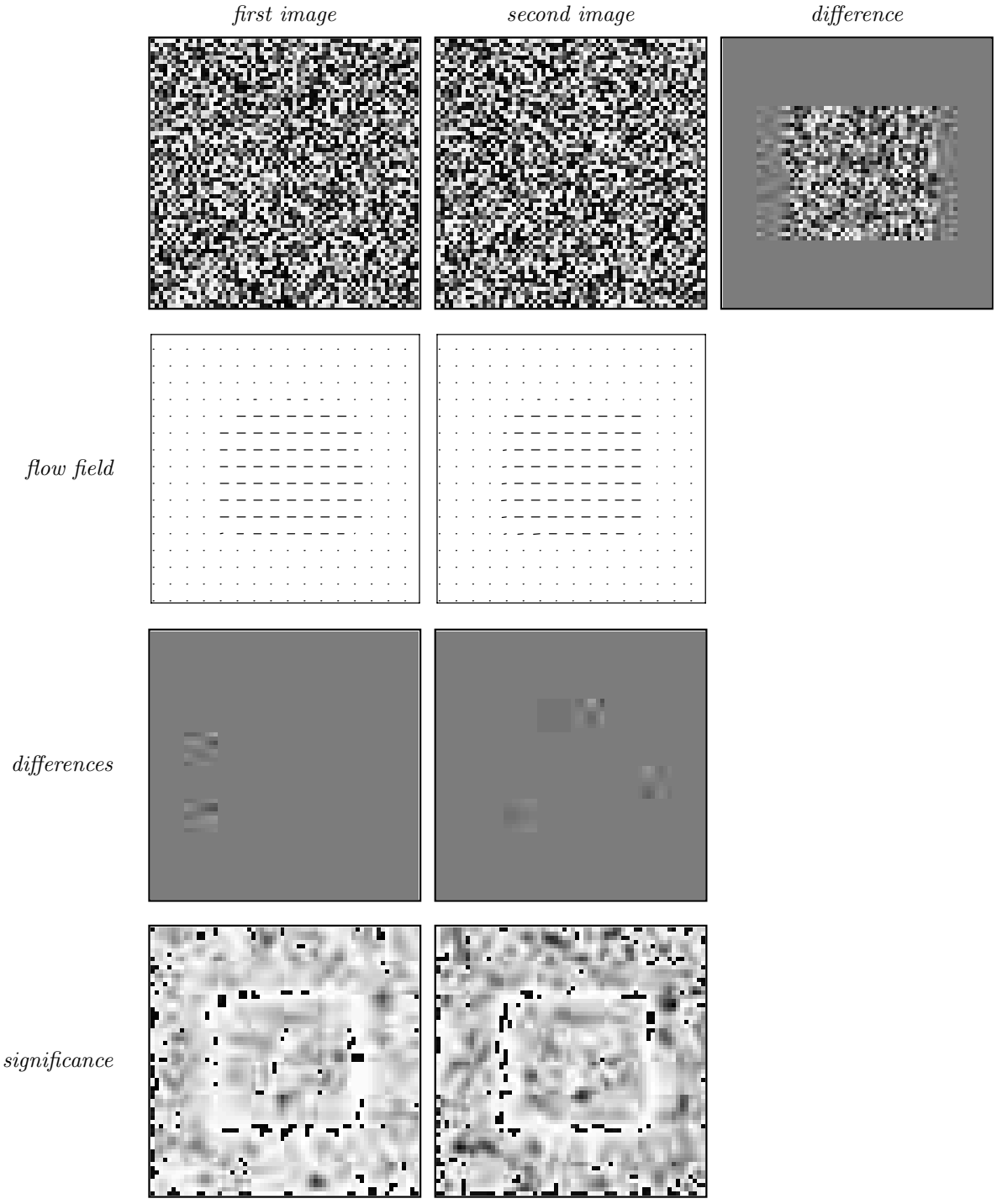

Figure 6: Flow field estimates computed from a synthetic random dot stereogram using the proposed scheme with automatic scale selection. From the top row to the bottom row, the images show respectively (a) the original image pair, (b) estimated flow field, (c) compensated differences, (d) significance measure. (Image size: 64×64 pixels. Noise level: 1 %).

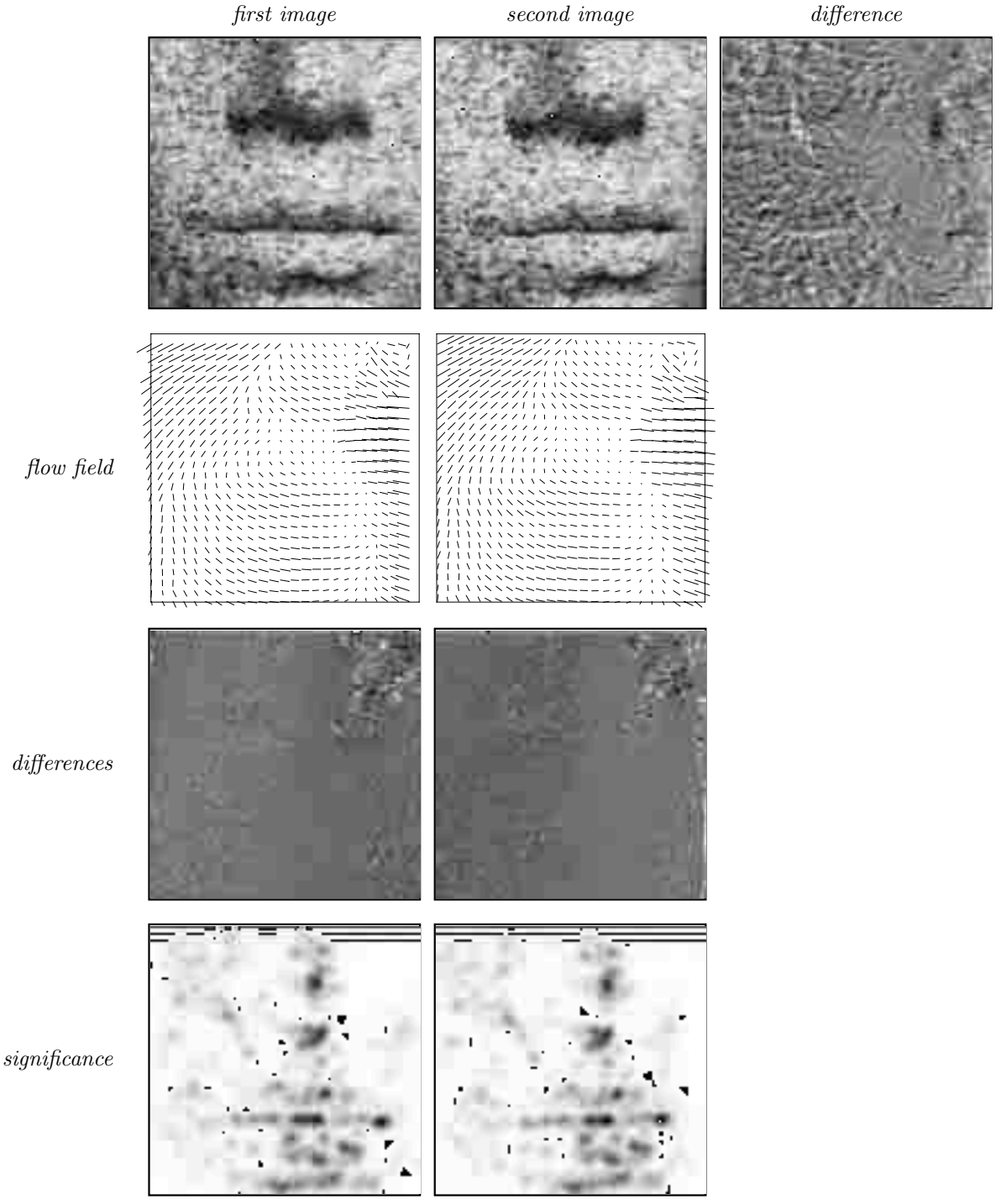

Figure 7: Flow field estimates computed from a detail of a statue using the proposed scheme with automatic scale selection. From the top row to the bottom row, the images show respectively (a) the original image pair, (b) estimated flow field, (c) compensated differences, (d) significance measure. (Image size: 200×200 pixels.)

involve minimization over the scale dimension as well.

10. Summary and discussion

We have argued that the incorporation of an explicit mechanism for automatic scale selection is an essential complement to traditional techniques for stereo matching and flow estimation. The main purpose of such a scale selection mechanism is to allow for local adaptive determination of the scales at which the deformation estimates should be extracted, and to suppress the influence of the erroneous deformation estimates that may be obtained if too fine scales are used. Specifically, such a mechanism makes it unnecessary to provide external information about the scales to which the coarse-to-fine propagation should be performed.

For general flow estimation and stereo matching algorithms, it was proposed that this problem could be addressed in the following general way by:

- Evaluating the evolution properties over scales of the estimated uncertainty of the deformation estimates, and
- selecting the deformation estimates from the scales at which the estimated uncertainty assumes local minima (or boundary minima) over scales.

Then, a specific implementation of this approach was presented based on a region based differential scheme for hierarchical and iterative computation of optic flow, which incorporated explicit mechanisms for flow averaging and confidence measurements. The confidence measure was constructed as a combination of the strength of the operator response, the internal consistency between the flow estimates in the forward and backward directions of the bidirectional matching scheme, as well the error in the least squares alignment measured in terms of the residual.

It was shown that for this specific scheme, automatic scale selection could be performed by minimizing a suitably normalized residual over scales. Specifically, the combined scale selection and flow estimation scheme resulting from this approach has the intuitively appealing qualitative properties of leading to:

- selection of coarser scales for large size image structures,
- selection of coarser scales with increasing noise levels, and
- selection of finer scales near flow discontinuities.

The underlying reason why this behaviour can be expected is that the normalized residual can be expected to be high when a local translation model does not agree with the image data. Thus, by selecting scales that minimize the normalized residual over scales we are much less likely to select deformation estimates computed at inappropriate scales.

Notably, the resulting scheme is expressed in terms of only the following types of image operations: large support scale-space smoothing, small support derivative computations, pointwise combinations of the outputs from abovementioned

steps, and the detection of local minima across scales. Hence, the algorithm lends itself naturally to straightforward implementation in a visual front-end.

Another interesting aspect of the resulting approach is that the problem of computing the deformation flow fields has been reformulated as the problem of finding a suitable transformation of the image operators that brings the output of the image operators into alignment (corresponding to a fixed point of (15). In an active situation, the control signals obtained from this construction can serve as a natural vergence mechanism. If a corresponding translation scheme is applied in the log-polar domain, it provides a lower order approach for measuring the other primitive transformations in the similarity group, *i.e.*, rotations and size changes. When extended to local full affine models, and when complemented by an affine Gaussian scale-space representation (see section 2.1. and section 11.), the scheme allows for estimation of full local affine transformations without bias effects due to the image operators. Philosophically, such a viewpoint to the image matching problem can be seen as finding the transformation that best explains the data.

Concerning the possible relevance of the proposed methodology to biological vision, one may speculate whether mechanisms for scale selection exist in biological vision. In this context, it is worth noting that in their investigations of the receptive fields of motion sensitive cells in Area MT (V5) of the macaque monkey, (Raiguel *et al.* 1995) have observed that receptive field sizes can typically vary by a factor of 10 in area among the cells in a locally confined region of visual space. Hence, it seems plausible that the information may at least be available in confined regions for making the types of local judgements that the proposed scale selection principle is based on.

11. Extensions

The main message of this article has been to advocate the need for a scale selection mechanism in visual modules for computing optic flow and performing stereo matching, and to propose a systematic methodology for addressing this problem. Whereas a specific implementation was presented to support the suggested approach, no claims are made that this implementation constitutes any optimal choice. (For example, we are currently developing a point based scheme based on a similar approach.) Besides possible technical improvements of the uncertainty measure used, this scale selection methodology lends itself to more general extensions in the following directions:

Higher order deformation models. Whereas the specific scheme presented has been concerned with a local translation model, there is nothing in principle that prevents it from being extended to higher-order deformation models, such as local affine deformations. This will, in general, increase the accuracy and allow for unbiased estimation of the higher order deformations. As mentioned previously, these general ideas can also be applied on other flow estimation and stereo matching algorithms.

Incidentally, minimization of normalized error measures over scales applies also to other types of problems, such as junction localization (Lindeberg 1994b).

Relaxing the constant brightness assumption. Throughout this work, we have applied the constant brightness assumption implying that the image brightness is assumed to be the same for all views of a physical point. Since this approach suffers from inherent limitations (see (Pentland 1990) for illustrative examples), one could, of course, more generally apply the same technique to other functions of the local N-jet. For example, if we apply it to the output from the Laplacian operator, we cancel the effects of illumination variations up to first order. More generally, if we apply it to several differential invariants simultaneously, we will be much more likely to obtain useful results from point based optic flow techniques.

Enforcing consistency. The scheme presented so far estimates local image deformations in a neighbourhood of each image point. The only coupling between the flow field estimates at different points is *via* the local least squares model and the flow field smoothing, which limit the spatial variation of the computed flow fields. Clearly, for such independently computed estimates, it is not guaranteed that shape descriptors derived from them correspond to a coherent surface. For example, for surface orientation estimates computed from such descriptors, it is not guaranteed that the curl is zero, which is a necessary requirement for a vector field to have a potential, and hence for the surface orientations to correspond to a coherent surface.

Figure 8 shows an example of enforcing such consistency on monocular data by fitting a pointwise (and hence parameter free) depth map to surface orientation estimates computed from a slight modification of the shape from texture method in (Lindeberg and Gårding 1993; Gårding and Lindeberg 1996) based on the weak isotropy assumption (Gårding 1993). For each point, the surface orientation has been obtained from a *centered second moment matrix*

$$\nu = \int_{\xi \in \mathbb{R}^2} (\nabla L)(\nabla L)^T \, w \, d\xi - (\overline{\nabla L})(\overline{\nabla L})^T \tag{25}$$

where $\overline{\nabla L} = \int_{\xi \in \mathbb{R}^2} \nabla L \, w \, d\xi$ and w is a Gaussian window function. Then, from (a modification of) the weak isotropy assumption—that ν in the surface should be a constant times the unit matrix—the slant angle has been computed as $\sigma = \arccos \sqrt{\frac{1-\tilde{Q}(\nu)}{1+\tilde{Q}(\nu)}}$, (where $\tilde{Q}(\nu)$ is defined from ν in a similar way as $\tilde{Q}$ is defined from A in (10)) and the tilt direction has been determined as the eigendirection of μ corresponding to the maximum eigenvalue. The resulting surface orientation estimates are shown in figure 8(a) as a set of ellipses, which should be interpreted as projected circles.

Finally, a pointwise depth map has obtained by a least-squares fitting of a pointwise depth function to the data by minimizing the curl of the estimated field of surface orientation estimates, in a way analogous to the experimental

technique used in (Koenderink *et al.* 1992) and the integrability constraint exploited in certain shape-from-shading algorithms (see (Horn and Brooks 1989) for an overview). Observe how the qualitative shape of the torso is captured by these very conceptually simple operations. We are currently studying the further analysis and integration of these components into a unified framework for shape estimation.

Estimated normals *Surface model*

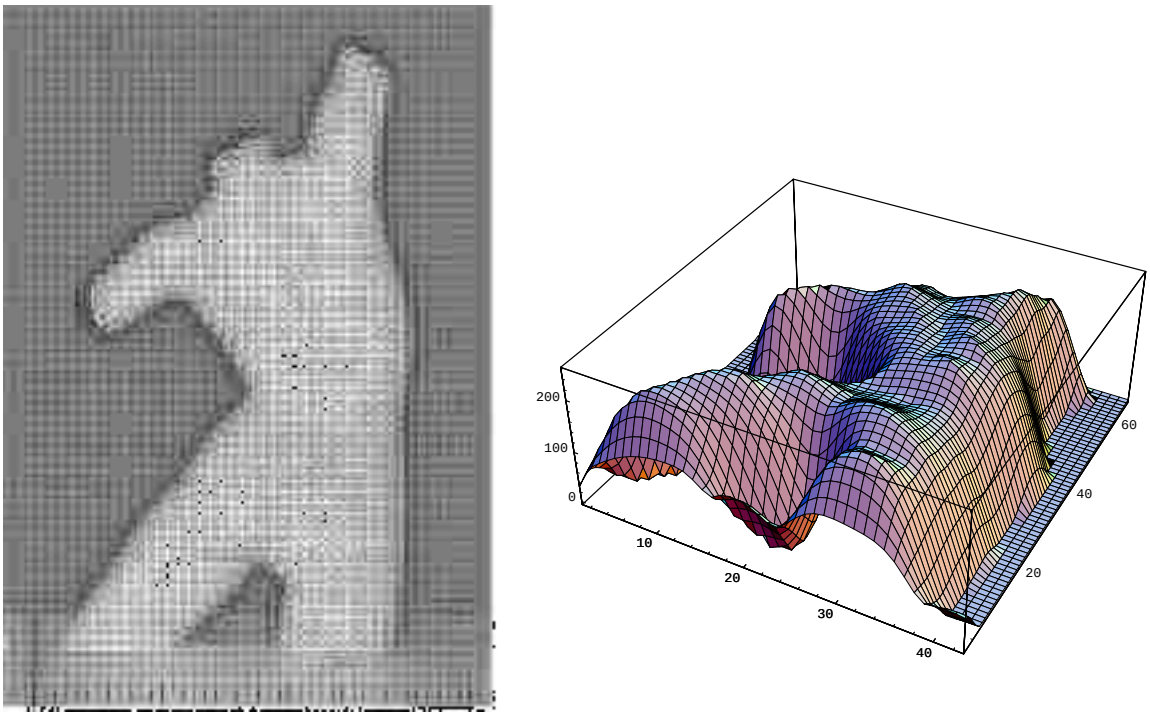

Figure 8: (left) Surface normals estimated from local affine deformations measured by centered second moment matrices. (right) Surface model constructed from the additional requirement that the field of surface normals should correspond to a depth function (*i.e.* have a zero curl).

A Appendix: Statistical analysis

Given a disparity estimate computed according to the methodology in section 3., we would like to estimate how accurate this estimate is.

The estimate is computed from a local Taylor expansion of the brightness constancy relation (5) at any level t in scale-space

$$(\nabla L)(\xi;\, t)^T(\Delta\xi) + (L(\xi;\, t) - R(\xi;\, t)) = \mathcal{O}(|\Delta\xi|^2). \tag{26}$$

This relation is integrated over a region of interest (5)

$$\min_{v\in\mathbb{R}^2} \int_{\xi\in\mathbb{R}^2} (((\nabla L)(\xi;\, t))^T v - (L(\xi;\, t) - R(\xi;\, t))^2\, w(\xi)\, d\xi. \tag{27}$$

which leads to the least squares estimation problem (6), with A, b and c according to (7), (8) and (9):

$$\min_{v\in\mathbb{R}^2} v^T A v + 2b^T v + c. \tag{28}$$

The solution (11) of (28) is given by

$$v = -A^{-1}b. \tag{29}$$

Definition of error measures. Given a field of pointwise disparity estimates $\Delta\xi$ obtained from (26), one may at a first glance consider the following entity as a natural measure of the error in the least squares estimate (29):

$$\mathcal{E} = \int_{\xi\in\mathbb{R}^2} |v - \Delta\xi|^2\, w(\xi)\, d\xi = \int_{\xi\in\mathbb{R}^2} (v - \Delta\xi)^T (v - \Delta\xi)\, w(\xi)\, d\xi. \tag{30}$$

This measure does, however, not take the following facts into account: (i) The brightness constancy relation (26) only implies a one-dimensional constraint on two-dimensional local disparity estimate. (ii) For points where the gradient magnitude is high, the brightness constancy relation (26) provides a stronger constraint on the local disparity than in regions where the gradient magnitude is low, and the local image structure is more strongly influenced by the information in the higher order derivatives of the intensity function L.

To capture these effects, let us instead measure the error in the local disparity estimate only in terms of it component parallel to the local gradient direction (*i.e.*, we are only sensitive to errors in the pointwise normal flow), and let us weight the local errors using the gradient magnitude as weight. This leads to the following weighted error measure, where the integrated square gradient magnitude in the numerator serves as normalization for the weight in the denominator:

$$\begin{aligned}
\mathcal{E}_{\nabla L} &= \frac{\int_{\xi\in\mathbb{R}^2} |(\nabla L)^T(v - \Delta\xi)|^2\, w(\xi)\, d\xi}{\int_{\xi\in\mathbb{R}^2} |\nabla L|^2\, w(\xi)\, d\xi} \\
&= \frac{\int_{\xi\in\mathbb{R}^2} (v - \Delta\xi)^T (\nabla L)(\nabla L)^T (v - \Delta\xi)\, w(\xi)\, d\xi}{\int_{\xi\in\mathbb{R}^2} (\nabla L)^T(\nabla L)\, w(\xi)\, d\xi}
\end{aligned} \tag{31}$$

To derive a more explicit expression for this entity, let us make the following linear approximation, which disregards the higher order terms in (26):

$$(\nabla L)^T (\Delta \xi) \approx R - L. \tag{32}$$

Then, after expansion, the numerator of (31) can be written

$$\begin{aligned} \mathcal{E}_{\nabla L, \text{num}} &= \int_{\xi \in \mathbb{R}^2} |(\nabla L)^T v - (R - L)|^2 \, w(\xi) \, d\xi \\ &= v^T A v - 2 b^T v + c, \end{aligned} \tag{33}$$

with A, b and c according to (7), (8) and (9). Using $v = -A^{-1} b$ and

$$\mathcal{E}_{\nabla L, \text{denom}} = \int_{\xi \in \mathbb{R}^2} (\nabla L)(\nabla L)^T \, w(\xi) \, d\xi = \text{trace}\, A, \tag{34}$$

it follows that $\mathcal{E}_{\nabla L}$ in (31) can be written

$$\mathcal{E}_{\nabla L} \approx \frac{c - b^T A^{-1} b}{\text{trace}\, A} = \tilde{r} \tag{35}$$

corresponding to the definition of the normalized residual $\tilde{r}$ in (24). In other words, the normalized residual used for automatic scale selection can be seen as an approximate estimate of the error in the local least squares estimation problem. Specifically, the scale selection principle proposed in section 6. corresponds to matching the image data at the scale at which the error estimate is minimized.

Perturbation analysis. In the abovementioned analysis, the input images f_L and f_R are regarded as given, and so are their scale-space representations L and R. The error measure (31) estimates the error in the regional disparity estimate v based on the internal consistency between the pointwise flow estimates $\Delta \xi$.

An alternative approach to error analysis is to investigate how sensitive the estimated disparity is to perturbations of f_L and f_R. Such an analysis, however, requires further *a priori* information about the statistical properties of f_L and f_R, based on specific knowledge about the structure of the world as well as the imaging conditions. We do not exclude that the proposed scale selection methodology could benefit from such an approach, provided that an appropriate statistical model can be formulated. In this treatment, we have deferred from such an approach, partly because of the need to formulate accurate statistical models of the two-dimensional signals f_L and f_R, and partly because the closed form expressions for the error estimates become much more involved.

References

L. Alvarez; F. Guichard; P.-L. Lions, and J.-M. Morel. "Axioms and Fundamental Equations of Image Processing". *Arch. for Rational Mechanics*, 123(3):199–257, Sep. 1993.

J. Arnspang. *Motion Constraint Equations in Vision Calculus.* Doctoral dissertation. Doctoral dissertation, Dept. Med. Phys. Physics, Univ. Utrecht, NL-3508 Utrecht, Netherlands, 1991.

J. J. Barron; D. J. Fleet, and S. S. Beachemin. "Performance of Optical Flow Techniques". *Int. J. of Computer Vision*, 12(1), 1994.

J. R. Bergen; P. Anandan; K. J. Hanna, and R. Hingorani. "Hierarchical Model-Based Motion Estimation". In G. Sandini, editor, *Proc. 2nd European Conf. on Computer Vision*, volume 588 of *Lecture Notes in Computer Science*, pages 237–252, Santa Margherita Ligure, Italy, May. 1992. Springer-Verlag.

F. Devernay and O. D. Faugeras. "Computing Differential Properties of 3-D Shapes from Stereoscopic Images without 3-D Models". In *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, pages 208–213, Seattle, Washington, 1994.

L. Florack and M. Nielsen. "The Intrinsic Structure of the Optic Flow Field". Technical Report ERCIM–07/94–R033, 1994.

L. M. J. Florack; B. M. ter Haar Romeny; J. J. Koenderink, and M. A. Viergever. "Scale and the Differential Structure of Images". *Image and Vision Computing*, 10(6):376–388, Jul. 1992.

J. Gårding and T. Lindeberg. "Direct computation of shape cues using scale-adapted spatial derivative operators". *Int. J. of Computer Vision*, 17(2):163–191, 1996.

J. Gårding. "Shape from Texture and Contour by Weak Isotropy". *J. of Artificial Intelligence*, 64(2):243–297, Dec. 1993.

B. K. P. Horn and M. J. Brooks, editors. *Shape from Shading.* MIT Press, Cambridge, Massachusetts, 1989.

B. K. P. Horn and B. G. Schunck. "Determining Optical Flow". *J. of Artificial Intelligence*, 17:185–204, 1981.

D. G. Jones and J. Malik. "A computational framework for determining stereo correspondences from a set of linear spatial filters". In G. Sandini, editor, *Proc. 2nd European Conf. on Computer Vision*, volume 588 of *Lecture Notes in Computer Science*, pages 395–410, Santa Margherita Ligure, Italy, May. 1992. Springer-Verlag.

T. Kanade and M. Okutomi. "A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment". *IEEE Trans. Pattern Analysis and Machine Intell.*, 16(9):920–932, 1994.

J. J. Koenderink and A. J. van Doorn. "Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer". *Optica Acta*, 22(9):773–791, 1975.

J. J. Koenderink and A. J. van Doorn. "Geometry of binocular vision and a model for stereopsis". *Biological Cybernetics*, 21:29–35, 1976.

J. J. Koenderink and A. J. van Doorn. "Receptive Field Families". *Biological Cybernetics*, 63:291–298, 1990.

J. J. Koenderink and A. J. van Doorn. "Affine structure from motion". *J. of the Optical Society of America*, pages 377–385, 1991.

J. J. Koenderink. "The structure of images". *Biological Cybernetics*, 50:363–370, 1984.

J. J. Koenderink; A. J. van Doorn, and A. Kaeppers. "Surface perception in pictures". *Perception and Psychophysics*, 52:487–496, 1992.

T. Lindeberg and J. Gårding. "Shape from Texture from a Multi-Scale Perspective". In H.-H. Nagel et. al., editor, *Proc. 4th Int. Conf. on Computer Vision*, pages 683–691, Berlin, Germany, May. 1993. IEEE Computer Society Press.

T. Lindeberg and J. Gårding. "Shape-adapted smoothing in estimation of 3-D depth cues from affine distortions of local 2-D structure". In J.-O. Eklundh, editor, *Proc. 3rd European Conference on Computer Vision*, volume 800 of *Lecture Notes in Computer Science*, pages 389–400, Stockholm, Sweden, May. 1994. Springer-Verlag.

T. Lindeberg. "Direct Estimation of Affine Deformations of Brightness Patterns Using Visual Front-End Operators with Automatic Scale Selection". Technical Report ISRN KTH/NA/P--94/33--SE, Dept. of Numerical Analysis and Computing Science, KTH, Stockholm, Sweden, Nov. 1994. Also in Proc. 5th International Conference on Computer Vision.

T. Lindeberg. "Junction detection with automatic selection of detection scales and localization scales". In *Proc. 1st International Conference on Image Processing*, volume I, pages 924–928, Austin, Texas, Nov. 1994. IEEE Computer Society Press.

T. Lindeberg. "Scale Selection for Differential Operators". Technical Report ISRN KTH/NA/P--94/03--SE, Dept. of Numerical Analysis and Computing Science, KTH, Stockholm, Sweden, Jan. 1994. Extended version in *Int. J. of Computer Vision*, vol 30, number 2, 1998. (In press).

T. Lindeberg. *Scale-Space Theory in Computer Vision.* The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, Dordrecht, Netherlands, 1994.

B. D. Lukas and T. Kanade. "An iterative image registration technique with an application to stereo vision". In *Image Understanding Workshop*, 1981.

J. Malik and R. Rosenholtz. "A differential method for computing local shape-from-texture for planar and curved surfaces". In *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, pages 267–273, 1993.

R. Manmatha. "Measuring the Affine Transform using Gaussian filters". In J.-O. Eklundh, editor, *Proc. 3rd European Conference on Computer Vision*, volume 801 of *Lecture Notes in Computer Science*, pages 159–164, Stockholm, Sweden, May. 1994. Springer-Verlag.

M. Otte and H.-H. Nagel. "Optical Flow Estimation: Advances and Comparisons". In J.-O. Eklundh, editor, *Proc. 3rd European Conference on Computer Vision*, volume 800 of *Lecture Notes in Computer Science*, pages 51–60, Stockholm, Sweden, May. 1994. Springer-Verlag.

A. P. Pentland. "Photometric Motion". In *Proc. 3rd Int. Conf. on Computer Vision*, pages 178–187, Osaka, Japan, Dec. 1990. IEEE Computer Society Press.

M. Proesmans; L. van Gool; E. Pauwels, and A. Oosterlinck. "Determination of Optical Flow and its Discontinuities using Non-Linear Diffusion". In J.-O. Eklundh, editor, *Proc. 3rd European Conference on Computer Vision*, volume 801 of *Lecture Notes in Computer Science*, pages 295–304, Stockholm, Sweden, May. 1994. Springer-Verlag.

S. Raiguel; M. van Hulle; D.-K. Xiao; V. L. Marcar, and G. A. Orban. "Shape and spatial distribution of receptieve fields and antagonistic motion surrounds in the middle temporal area (V5) of the Macaque". *Eur. J. Neurosci.*, 7:2064–2082, 1995.

G. Sapiro and A. Tannenbaum. "Affine invariant scale-space". *Int. J. of Computer Vision*, 11(1):25–44, 1993.

J. Sato and R. Cipolla. "Extracting the Affine Transformation from Texture Moments". In J.-O. Eklundh, editor, *Proc. 3rd European Conference on Computer Vision*, volume 801 of *Lecture Notes in Computer Science*, pages 165–172, Stockholm, Sweden, May. 1994. Springer-Verlag.

R. van Ee. *Stability of Binocular Depth Perception.* PhD thesis. , Helmholtz Institute, Utrecht University, Princetonplein 5, Utrecht, the Netherlands, March 1995.

P. Werkhoven and J. J. Koenderink. "Extraction of motion parallax structure in the visual system". *Biological Cybernetics*, 1990.

A. P. Witkin. "Scale-space filtering". In *Proc. 8th Int. Joint Conf. Art. Intell.*, pages 1019–1022, Karlsruhe, West Germany, Aug. 1983.

A. L. Yuille and T. A. Poggio. "Scaling Theorems for Zero-Crossings". *IEEE Trans. Pattern Analysis and Machine Intell.*, 8:15–25, 1986.