A Way to Account for Models in Image Analysis Illustrated by Motion Extraction

Luc Florack
Image Sciences Institute
Utrecht University, Dept. of Computer Science
Padualaan 14, 3584 CH Utrecht
The Netherlands
Luc.Florack@cs.ruu.nl

Wiro Niessen
Image Sciences Institute
University Hospital Utrecht
Heidelberglaan 100, 3584 CX Utrecht
The Netherlands
Wiro.Niessen@cv.ruu.nl

Abstract

Image analysis requires adequate models, i.e. efficacious symbolic representations of a priori knowledge or hypotheses. These are expressed in terms of basic structural entities (grey-values, derivatives, etc.), defined by a conventional preprocessing of image data. That such entities are in turn subject to models is, however, much less in focus. This article argues in favour of a manifest segregation of task-based models versus image-based models, and of maintaining a transparent relation between the latter and the raw data.

A possible approach is suggested based on topological duality (to model the data) and gauge invariance (to model the task). The procedure is applied to motion extraction. By virtue of model transparency it has indeed proven possible to obtain results for a (2+1)D benchmark sequence outperforming all existing algorithms that have been reported in a comparative study. The extraction of cardiac wall motion from an MR sequence of a canine heart is illustrated using the same method.

1. Introduction

There are roughly two distinct ways in which models enter in image analysis. For specific tasks one resorts to specific models, reflecting all relevant factors involved, such as image formation details, *a priori* knowledge or hypotheses about objects being imaged, etc. Such models could be called *task-based*. They are usually *falsifiable* (much like laws of nature, even though the criterion in image analysis is task performance rather than "ground truth"). *E.g.* in the case of motion extraction in medical imaging one may opt for a physical model of blood as an incompressible medium satisfying a conservation principle, which can indeed be shown to be a valid approximation.

Over the last decade or so the appreciation has grown that the data themselves are to be subjected to models. Important paradigms clearly formalising what one could call *image-based* models have been developed, which are intended to be more or less independent of the details of a task. Examples are various low-level representations (scale-space theory, mathematical morphology, wavelet theory, etc.). More than any, image-based models must adhere to a principle of parsimony, as their primary role is to subserve interpretations. They are neither true nor false, merely *conventional*; different image-based models may equally well support a given task. In view of their uncommittedness they do have to satisfy a few plausible constraints, such as *completeness* (no information should be discarded *a priori*).

The dichotomy is sometimes a bit whimsical, especially in schemes where image data and task objective are confounded from the outset (various nonlinear diffusion schemes and variational principles for image enhancement—arguably a part of image reconstruction—etc.). In any case it is clear that models provide our exclusive control mechanism (one does not "tamper with the data", but one can interpret them at liberty); for this reason it is important to put them into focus. What one would like to have is a *transparent* image-based model, one that allows us to appreciate the consequences of its very choice—after all a merely conventional one—in all stages of the analysis. In addition, perhaps even more importantly, one would like to draw a clear distinction between the "syntactical", image-based model and the "semantical", task-based model. The former defines the description language in terms of which the latter has to be stated and subsequently refined until task performance is satisfactory. This calls for a *manifest* segregation of the two types of models.

It often serves clarity to represent a model in two equivalent ways, a conceptual one that can be easily understood, and a computational one that directly reflects its implementation. It is prudent to anticipate computational aspects by concentrating on operational concepts only; the computational stage is not meant to solve *fundamental* problems.

The approach outlined above encourages modularity, and can easily be generalised. For instance, extraction of motion information from a cine-MR cardiac image could proceed, as will be outlined below, along the lines of an image-based *versus* task-based stratagem. Once a satisfactory algorithm has been obtained its inner workings may be encapsulated in a black-box system, the output of which—a dense vector field—induces new syntagmata (elements of a subsequent image-based model) for a more sophisticated task, such as finding pathology. This requires a new task-based model incorporating expert knowledge of how pathology relates to cardiac motion (and so forth).

The proof of the pudding is in the eating. We concentrate on motion extraction, and show that adequate modelling can indeed produce results that outperform known methods on a benchmark sequence as described by Barron *et al.* [3]. Application to medical image sequences is straightforward as will also be illustrated.

2. Motion analysis

We adhere to the principle explained in the introduction, and distinguish an image-based model (Section 2.1) and a task-based model (Section 2.2). A computational scheme has been taken apart for reasons of clarity (Section 2.3).

2.1. Image-based modelling: topological duality

Instead of manipulating pixels or voxels directly, most state-of-the-art techniques employ some form of preprocessing. One could say that pixels (voxels) conform to a machine-oriented format, which may be distracting in the analysis (especially if continuous models are used) and, moreover, tends to overemphasise irrelevant details (arbitrariness of discretisation and quantisation, noisiness, etc.). To have better control over the data a common procedure is to apply suitable preprocessing filters. This suppresses noise, makes grid artifacts less pronounced, and provides control parameters that may turn out useful in the analysis.

Possibilities are overwhelming given the infinite dimensionality of typical function spaces. Plausible arguments will usually narrow down the filter class to a manageable size, but whatever the arguments, selection always boils down to commitment (because a different choice will typically produce different results in the end). The fact that the filtering procedure *defines* the result—much like an "edge" is the outcome of an edge detector—argues for an approach in which this fact is recognised as a basic truism. A nice mathematical principle that makes the precise interaction of data and filter explicit (rather than focusing on mere output) is *duality*, and a particularly simple instance of this—by virtue of linearity—is *topological duality*, systematically developed by Schwartz in the early fifties [23].

We shall content ourselves here with a brief explanation and refer to the literature for details [9, 10, 12, 23]. Let Δ be a linear function space defining our filters. Then the set of all continuous mappings $\Delta \to \mathbb{R}$ is known as its topological dual, and is denoted Δ'. This space is identified with that of the data, Σ say, because it corresponds by construction to

those degrees of freedom that can be segregated by monitoring a sufficient number of filter outputs (the familiar "wysiwyg" philosophy). Thus it implies that members of $\Sigma = \Delta'$ (*i.e.* images) are to be regarded as *equivalence classes*; all fiducial data configurations that produce identical responses after filtering with the entire filter bank are indistinguishable, or *metamerical* (*cf.* Koenderink's example of the "sextuplet image" and his exposition of "metameric black pictures" [16, 17]). Metamerism is perhaps best known from colour perception: many different spectra will map to indistinguishable colours after filtering with the tristimulus curves. Ideally all irrelevant details end up in metamerical classes; in an uncommitted image-based model these coincide with all non-measurable quantities.

The class proposed by Schwartz is used for its many nice properties and genericity. It is denoted $\mathcal{S}(\mathbb{R}^n)$, and consists of all smooth functions of rapid decay on $\mathbb{R}^n$. Members of its topological dual, $\mathcal{S}'(\mathbb{R}^n)$, are also called *tempered distributions*, and correspond to functions of polynomial growth and derivatives of the "Dirac δ-function" (elements that differ by at most a non-measurable amount are considered identical). Thus we henceforth set $\Delta = \mathcal{S}(\mathbb{R}^n)$ and therefore $\Sigma = \mathcal{S}'(\mathbb{R}^n)$. Recall that Σ models the image data, for which Schwartz's choice is certainly generic enough to encompass all functions that are likely to be of interest in practice. There is apparently no need to call for other filters than the extremely nice Schwartz functions!

The linear filtering of a raw image $F \in \mathcal{S}'(\mathbb{R}^n)$ by a filter $\phi \in \mathcal{S}(\mathbb{R}^n)$ is usually written as $F[\phi]$, and can be represented by the *Riesz representation formula* as

$$F[\phi] \stackrel{\text{def}}{=} \int dx\, f(x)\, \phi(x)\ . \tag{1}$$

(The integral is over all of $\mathbb{R}^n$; the "function-under-the-integral" f models the pixel or voxel data.)

One of the main reasons for adopting this image-based model is the fact that it enables us to define *differential structure* to any order and in a *well-posed* sense. This is most convenient given the many task-based models exploiting differential methods (recall that classical differentiation is ill-posed, thus useless). The trick relies on transposition:

$$\nabla F[\phi] \stackrel{\text{def}}{=} -F[\nabla\phi]\ . \tag{2}$$

Notice the minus sign (the idea is that a partial integration reproduces the l.h.s. given the more general r.h.s. in case f is differentiable; ∇F then corresponds to ∇f). Distributional differentiation is akin to a quite general recipe, extremely useful in image models, known in differential geometry as the *carry-along principle* (*e.g.* "pull back" and "push forward" of forms and vectors, respectively [9, 10, 12, 24, 25]).

In the context of a conservation principle, the natural setting for motion analysis [15], the proper differential tool is that of a *Lie derivative*. It can be defined within the framework of topological duality by the same token as any other derivative. Classically, if v is a vector field in spacetime, then the Lie derivative of a scalar function f is defined as its rate of change along the vector field, *i.e.* as the (ill-posed!) directional derivative $\mathcal{L}_v f = \nabla f \cdot v$. For a density field we have $\mathcal{L}_v f = \nabla(f\, v)$, the rate of change of the induced *flux*. In any case we may write

$$\mathcal{L}_v F[\phi] \stackrel{\text{def}}{=} -F[\mathcal{L}_v\,\phi]\ , \tag{3}$$

as the well-posed and operationally meaningful interpretation in the context of our image-based model.

A vanishing Lie derivative expresses conservation, in which case one may scale one component of $v = (w; \vec{v}) \in \mathbb{R}\times\mathbb{R}^{n-1}$ to unity by virtue of linearity (component w corresponds to temporal direction). In the literature one invariably chooses the temporal component—a constraint, albeit locally a very weak one [10, 14]—referring to the spatial component $\vec{v}$ as the motion field. In this case we obtain the following linear systems for scalar, respectively density images:

$$F[\nabla\cdot(\vec{v}\phi)] + F[\frac{\partial}{\partial t}(w\phi)] = 0\ , \tag{4}$$

respectively

$$F[\vec{v}\cdot\nabla\phi] + F[w\frac{\partial}{\partial t}\phi] = 0\ , \tag{5}$$

both subject to the constraint

$$w = 1\ . \tag{6}$$

An important observation is that a scalar model for the raw image implies that the filter behaves as a density, *vice versa*, a plain consequence of duality. This underlines the importance of having an adequate image-based model; after all, any set of raw image data can be explained either as a scalar or as a density field, because the basis (unit scalar, 1, respectively unit 1-form, $dx^1 \wedge \ldots \wedge dx^n$) must be provided by a physical model (examples: greyvalues reflecting a scalar depth map in range imaging, proton density in MRI, etc.).

2.2. Task-based modelling: gauge theory

With the image model of the previous section we are equipped with a powerful description language—essentially all techniques from standard analysis can be applied in a *robust* manner—for solving particular tasks, *in casu* the extraction of motion from a time sequence of images. Note that v defined by the conservation Eqs. (4–5) is merely a representation of image structure; a solution to the equation of vanishing Lie derivative always exists (provided w is not forced to be unity!), and is fully determined by the image's differential structure, *i.e.* measurable quantities of the form $F[\nabla^k \phi]$, $k = 0, 1, 2, \ldots$, which can be computed by simple linear filterings of the raw image. However, defined in this data-intrinsic sense v is *not* a unique vector field, but at best a member of an equivalence class. This pertains to the notorious "aperture problem"; the conservation equation merely determines a *trivial homotopy*—a linking of isosurfaces—but no point-wise connections, *unless* one imposes further constraints in the form of a task-based model.

A basic requirement has been mentioned, the *temporal gauge* of Eq. (6). Although (to the best of the authors' knowledge) all of the existing literature adheres to it implicitly, it is actually a *hypothesis*, which could be phrased as "conservation of topological detail": no flowline of v is allowed to reverse its temporal sense (such a reversal, interpreted in causal sense, would appear as an annihilation or creation event of isosurfaces). Indeed, it will be shown below that situations may occur in which the temporal gauge cannot be upheld! For a discussion the reader is referred to the literature [10, 14].

In general, a task-based model should state conditions for selecting relevant degrees of freedom from the ones provided by the image-based model. By virtue of completeness of the latter this is in principle possible by the enforcement of constraints (provided of course no essential information is missing in the measurements). This in turn implies that the structure given by the image-based model is apparently *redundant*, but the nature of this redundancy is quite specific as it is only *relative to the task*. A similar situation is encountered in physics: redundant descriptions are frequently used to model specific field phenomena in so-called *gauge field theories* for essentially the same reason: the most parsimonious description is often a redundant one. In such a description actual physical phenomena are characterised by *gauge invariance*; all redundant degrees of freedom can be removed by adding *gauge conditions*. Such constraints have no physical effect and can be chosen according to mathematical convenience or personal whim.

In the case at hand we may regard Eq. (4) or Eq. (5) as a gauge invariant system, v being the *gauge field*. The manifestation of data evidence does not compel us to choose among the many possible vector fields v as long as they are compatible with the gauge invariant system. In this sense fixing the gauge (removing the invariance) is indeed a mere matter of taste, as in physics. But in order to define a *meaningful* solution one must take into account *external* considerations, in other words, fix the gauge in terms of a unambiguous task-based model. For an in-depth discussion the reader is again referred to the literature [10, 14].

Following common practice we impose the trivial temporal gauge of Eq. (6). In addition we need to fix the spatial gauge by $n - 2$ additional constraints. Several motion constraints induced by physical models have been reported in the literature, such as rigidity or non-elasticity constraints for solid objects [6, 7], incompressibility and continuity conditions for fluid flow [1, 5], or induced by mathematical models, such as smoothness constraints (often in the form of a Tikhonov regularisation), or polynomiality assumptions [2]. Another example of a mathematical model that is often useful is the *normal flow* constraint, *e.g.* if one is interested in motion of the cardiac wall as a surface rather than a point set. This gauge condition is enforced by setting tangential flow to zero: same form as Eq. (4), respectively Eq. (5), but with $(w; \vec{v})$ replaced by $(0; \vec{v}_\perp)$, in which $\vec{v}_\perp$ generically denotes all $n - 2$ independent vectors perpendicular to $\vec{v}$. An unfortunate way to fix the gauge—but not an uncommon one—is to resolve the ambiguity *by a computational scheme*, typically in terms of an intricate combination of differentiation and approximation of the basic conservation equation. It is deceptive as it suggests that the nature of the problem is merely computational; it does not provide much insight, and is thus difficult to verify or falsify. (Example: consider Eqs. (4) and (6) in $(2 + 1)$D, differentiate once, and ignore derivatives of v; a unique solution emerges, but one remains essentially in the dark concerning its physical significance—if any. Moreover it fails to provide cues for its potential failure.)

If the gauge can be established in terms of purely local constraints, one ends up with a system that admits a unique solution that depends only on local derivatives. Details are discussed in the next section.

2.3. Computational Details

Consider normal flow in temporal gauge for simplicity. In order to solve for the complete system of motion equations (n equations for the n components of v: 1 formalising the image-based model, and $n - 1$ for the temporal and spatial gauges, together expressing the task-based model) we need to specify a computational scheme that establishes an explicit mapping

between our objective, the field v, and our data descriptors, $F[\nabla^k \phi]$, $k = 0, 1, 2, \ldots$ The unknown field v appears "under the integral" due to our "wysiwyg" approach (this is clear by rewriting Eq. (3) for the Lie derivative in integral form). Therefore one approach suggests itself, *viz.* to expand the vector field v relative to a function basis, so that the problem is recast into finding unknown coefficients, which can be taken out of the integral. This procedure can be set up so as to maintain as many equations as unknowns. If one discards the gauge conditions it preserves the degree of ambiguity, *i.e.* it does not interfere with the conceptual image/task-distinction, as a computational procedure befits. The remainder of the integrand is of the form f (raw image) times some effective filter. (The use of a complete filter family—*v.i.*—will enable one to rewrite the effective filters in terms of the original ones [10, 14].)

A convenient filter class that is complete and admits a closed-form analysis is the Gaussian family [18], which consists of all normalised Gaussians and their derivatives, though in principle any other complete family should work as well. A convenient expansion for v is a polynomial one[1]:

$$v^\mu_M(x) \stackrel{\text{def}}{=} \sum_{k=0}^{M} \frac{1}{k!} v^\mu_{M;\rho_1 \ldots \rho_k} x^{\rho_1} \ldots x^{\rho_k} + \mathcal{O}(\|x\|^{M+1}) \,, \tag{7}$$

with $\mu = 1, \ldots, n$ and $M = 0, 1, 2, \ldots$ (and summation convention for pairs of identical labels). Such an expansion allows us to take M-th order derivatives if desired (higher orders will expose errors of $\mathcal{O}(1)$), while powers of x times Gaussian derivatives produce simple linear combinations of Gaussian derivatives. Apart from its mathematical attractiveness, the Gaussian family is fully characterised by only two parameters, one spatial and one temporal *inner scale*. Although the high-resolution field v does not depend on scale (not even on the filter family!), scale selection can be exploited in a fairly simple way so as to minimise computational errors, *cf.* Niessen *et al.* [21, 22].

The approximation order M is also a convenient control parameter in practice; since "ground truth" is generally not available, one can monitor the (pointwise) behaviour of the available terms of the Cauchy sequence $\|v_{M+1}(x) - v_M(x)\|$. Even the highest admissible order M will typically leave points x at which the Cauchy differences do not appear to converge, signifying a "flow discontinuity". Further details of the computational recipe can be found elsewhere [10, 11, 13, 14, 19, 20, 21, 22].

2.4. Simulation and verification

We demonstrate the efficacy of models, both image-based (Section 2.4.1) and task-based (Section 2.4.2).

2.4.1 Image-based models

We again consider normal flow in temporal gauge, and concentrate on the efficacy of image-based models. Several choices are to be made explicit: duality principle, filter class, parameter selection, geometric nature of source field (scalar, density, or other field), spacetime manifold, etc.

Among the possible duality principles we take the linear one, *i.e.* topological duality. However, the space $\mathcal{S}(\mathbb{R}^n)$ is too large for practical use, and so we take, without loss of generality, the Gaussian family $\Delta = \mathcal{G}(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n)$, say. In other words we consider linear scale-space representations of the image data: $\Sigma = \mathcal{G}'(\mathbb{R}^n)$. The assumption of homogeneity and (spatial) isotropy implies that we use linear correlations (or convolutions) only, and leaves us with two parameters to play with, spatial scale, σ, and temporal scale, τ.

To scrutinise all possible choices is beyond the scope of this article. We concentrate on the "scalar *versus* density" option, and illustrate its significance by a simulation study.

We generate periodic sequences of an oscillating, radially symmetric, 2D Gaussian blob in two different ways, one simulating a scalar, and one a density field (think of a video sequence of a suitable object under Lambertian conditions advancing and receding in time, and of proton density cine-MR, respectively). In both cases the initial frame looks like Fig. 1, but evolution differs quantitatively. In the density case we insure that the overall integral is conserved (thus inverse-amplitude will oscillate in phase with blob width), in the scalar case the amplitude is a constant.

Eq. (4) can be solved analytically for normal flow in temporal gauge:

$$\vec{v}_{\text{s,d}}(r,t) = \alpha_{\text{s,d}}(r,t)\, \frac{\vec{r}}{r^2} \,, \tag{8}$$

[1] This is *not* a Taylor expansion; the coefficients depend on M.

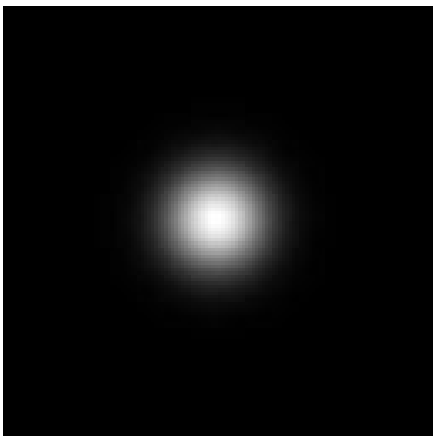

Figure 1. First frame of synthetic sequence.

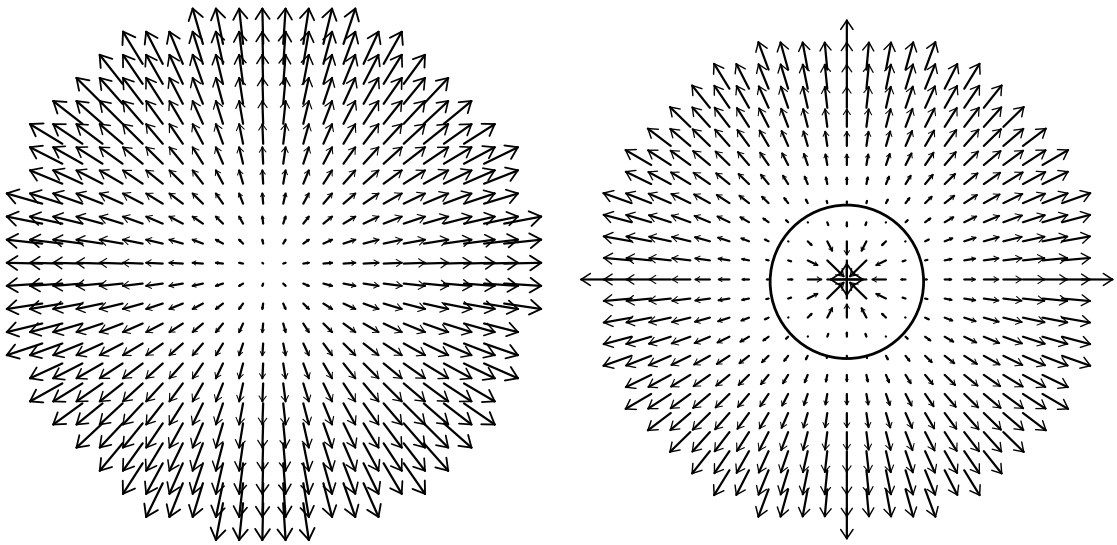

Figure 2. Scalar flow field computed for scalar (left), respectively density stimulus (right).

in which $\alpha_{\mathrm{s,d}}(r,t)$ is a function that differs for scalar and density case ($r = 0$ coincides with the symmetry centre). A study of the scalar case [10, 14] shows that near the origen $\alpha_s(r,t) = \mathcal{O}(r^2)$, thus no problem should arise at the centre, where $\vec{v} \propto \vec{r}$ vanishes smoothly. However, for the density stimulus we find $\alpha_d(r,t) = \mathcal{O}(1)$, revealing a singularity in $\vec{v}$ of order $\mathcal{O}(r^{-1})$. The observations agree with numerical results: Figs. 2–3. Note that the density image shows a flow inversion across a circle, which has been drawn in Fig. 2. According to theoretical prediction [10, 14]—and again confirmed in the experiments—its radius oscillates harmonically in proportion to blob width. This fact, by the way, runs counter to visual impression; perceptual flow alternates uniformly. The numerical results turn out to be extremely robust; the experiments have been repeated with substantial, multiplicative, pixel-uncorrelated Gaussian noise, producing quite similar results. Even with 50% noise the majority of errors for the $M = 1$ scheme does not exceed a few per cent, *cf.* Fig. 3.

Although highly idealised, the density stimulus is otherwise quite natural (*e.g.* proton density MR would behave similarly under circumstances of ideal acquisition and reconstruction). The emergence of a singularity in this study—which cannot be "blurred out"—should lead one to suspect that the problem will become more severe in real data, indeed even to question the very motion concept: certainly there are no infinite velocities in reality!

In actuality there is of course no problem with the image data, but an incompatibility of image- and task-based models, *viz.* the application of the scalar conservation principle Eq. (4), given the temporal gauge Eq. (6), to an image that apparently

behaves as a density, or, alternatively, the premise underlying Eq. (6) given the assumption of scalarity Eq. (4). The problem can thus be repaired in (at least) two ways: (i) by recognising the most obvious conceptual error, *i.e.* replacing Eq. (4) by Eq. (5) while maintaining the temporal gauge Eq. (6), or (ii) by giving up Eq. (6) while insisting on Eq. (4). The second option is easily overlooked if one takes the temporal gauge for granted. Yet in the end it remains a matter of interpretation: one could legitimately claim that the centre—in general every local extremum—contains a point source (or well) generating (absorbing) isocontours, thus violating "conservation of topological detail" expressed by the temporal gauge. This is by no means far-fetched: think of filming an intermittent light source, or of monitoring growth phenomena in bone tissue in a sequence of CT images acquired (and registered) over a significant time span. This illustrates that the image-based model is not "right" or "wrong" as such, but, in the end, has to be inspired by *general* task considerations.

2.4.2 Task-based models

We study the *Translating Tree Sequence* (TTS) and the *Diverging Tree Sequence* (DTS), respectively [3, 8], which constitute benchmark sequences for which "ground truth" is known: Fig. 4. We exploit the image-based model of Section 2.4.1 for its robustness (by the nature of topological duality) and easy control (scale selection for σ and τ and choice of approximation order M). More specifically we use the scalar model, Eq. (4), and exploit the scale degree of freedom—recall that in theory the field v does *not* depend on these parameters, nor on any other details of the filter family!—by following the suggestion by Niessen *et al.* [21, 22] to select parameter values on the basis of a *stability criterion* (the condition number of the linear system that is to be inverted *does* depend on scale).

The task-based model is derived from knowledge of camera motion. For the TTS sequence the spatial gauge is fixed by the assumption of horizontal camera motion, and by the usual temporal gauge, *i.e.* if $v = (w; \vec{v})$ and $\vec{e}_y$, $\vec{e}_r$ denote spatial basis vectors in vertical, respectively radial direction, then

$$\text{TTS:} \quad w = 1 \quad \text{and} \quad \vec{v} \cdot \vec{e}_y = 0 \,. \tag{9}$$

For the DTS sequence we incorporate knowledge of the focus of expansion (the origin) and postulate a radial field:

$$\text{DTS:} \quad w = 1 \quad \text{and} \quad \vec{v}_\perp \cdot \vec{e}_r = 0 \,. \tag{10}$$

The performance of the proposed motion scheme on these two sequences has been compared with the results reported by Barron *et al.* [3], using the same error measure ε (originally proposed by Fleet and Jepson [8]). As opposed to ours, all techniques discussed in the benchmark study are *identical* for TTS and DTS sequences. The relative uncommittedness of their underlying models makes these techniques suited as *a priori* flow estimators, but at the same time explains their sub-optimality. For details and nomenclature the reader is referred to the literature.

Table 1 concerns methods in which all vectors are maintained, and shows that the proposed method outperforms other approaches despite sub-optimal scale selection (only 2 temporal and 7 spatial scales have been compared). This can be attributed to the adequacy of *a priori* knowledge accounted for in the task-based models, Eqs. (9–10), together with the fact that the image-based model, Eq. (4), guarantees robustness and—in combination with the proposed approximation scheme—provides well-understood parameters that facilitate automatic optimisation.

Finally, in Table 2 we compare our approach with the best performing algorithms that discard uncertain estimates. Our approach performs better than the algorithm of Fleet & Jepson, which is the best performing algorithm reported [3].

The application of the entire procedure to real data is straightforward. Fig. 5 shows results of applying the same algorithm to cine-MR data (the computation has been carried out within a fixed 2D spatial slice for the sake of illustration, but extension to 3D is trivial). Although no "ground truth" is available, the experiment does confirm that stable solutions can indeed be obtained on real data, and that an appropriate handling of image-based model parameters is apparently essential. This illustrates the importance of the *transparency* criterion mentioned in the introduction.

In practice one will often have to search for adequate task-based models following a hermeneutic principle, re-interpreting the data by iteratively adapting the model so as to improve performance (the Bayesian rationale provides a possible framework), rather than rely on a *deus ex machina* gauge condition. *E.g.* in the DTS sequence the focus of expansion provides a cue for the motion field, *vice versa*. In the heart sequence segmentation—manually obtained in Fig. 5—and motion of the endocardial wall are intimately related. The hypothetical and preliminary status of any gauge condition argues strongly in favour of the general line of approach sketched in the introduction, in which the segregation of image- and task-based model is *manifest*.

Implementation method	TTS		DTS	
	av. error	st. dev.	av. error	st. dev.
Horn & Schunck (original)	38.72	27.67	12.02	11.72
Modified Horn & Schunck	2.02	2.27	2.55	3.67
Uras *et al.* (unthresholded)	0.62	0.52	4.64	3.48
Nagel	2.44	3.06	2.94	3.23
Anandan	4.54	3.10	7.64	4.96
Singh (step 1, $n = 2, w = 2, (N = 4)$)	1.64	2.44	17.66	14.25
Singh (step 2, $n = 2, w = 2, (N = 4)$)	1.25	3.29	8.60	5.60
Florack & Niessen*($M = 1$, multiple σ, τ)	0.49	1.92	1.15	3.32

Table 1. Best performing techniques reported by Barron et al. [3] estimating a dense velocity field. (All angular measures are expressed in degrees.)

Implementation method	TTS			DTS		
	av. error	st. dev.	density	av. error	st. dev.	density
Horn & Schunck (original, $\|\nabla L\| \geq 5.0$)	32.66	24.50	55.9%	8.93	7.79	54.8 %
Modified Horn & Schunck	1.89	2.40	53.2%	1.94	3.89	32.9 %
Lucas and Kanade $(\lambda_2 \geq 1.0)$	0.66	0.67	39.8%	1.94	2.06	48.2 %
Lucas and Kanade $(\lambda_2 \geq 5.0)$	0.56	0.58	13.1%	1.65	1.48	24.3 %
Uras *et al.* $(\det(H) \geq 1.0)$	0.46	0.35	41.8%	3.83	2.19	60.2 %
Nagel $\|\nabla L\|_2 \geq 5.0$	2.24	3.31	53.2%	3.21	3.43	53.5 %
Singh (step 1, $n = 2, w = 2, \lambda_1 \leq 5.0, (N = 4)$)	0.72	0.75	41.4 %	7.09	6.59	3.3 %
Heeger	4.53	2.41	57.8%	4.49	3.10	74.2 %
Fleet & Jepson $(\tau = 2.5)$	0.32	0.38	74.5%	0.99	0.78	61.0 %
Fleet & Jepson $(\tau = 2.0)$	0.23	0.19	49.7%	0.80	0.73	46.5 %
Fleet & Jepson $(\tau = 1.0)$	0.25	0.21	26.8%	0.73	0.46	28.2 %
Florack & Niessen*($M = 1$, multiple σ, τ)	0.16	0.18	60%	0.79	1.13	60 %
Florack & Niessen*($M = 1$, multiple σ, τ)	0.14	0.13	40%	0.43	0.40	40 %

Table 2. Best performing techniques reported by Barron et al. [3] estimating a sparse velocity field.

*TTS: Eqs. (4), (6), (7) & (9). DTS: Eqs. (4), (6), (7) & (10). Scale selection involves 2 temporal and 7 spatial scales following Niessen *et al.* [21, 22].

3. Summary

Simulation results in the context of motion extraction provide support for the general approach outlined in this article, in which it is argued that performance of algorithms—provably the motion algorithm reported here by way of example—may well benefit from a *manifest* segregation of image- and task-based models. The former needs to be suitably defined so as to facilitate analysis and computation; in particular it needs to insure robustness and provide well-understood control parameters that can be exploited for optimisation purposes. The image-based model provides the description language in terms of which specific goals must be formalised. Crucial aspects of an image-based model are therefore (i) simplicity (if one does not understand the language one cannot even formulate a task), (ii) transparency with respect to the image data (because data manipulation proceeds solely via the model), and (iii) robustness and computational realizability. All these are clearly reflected in this article by the central role attributed to *topological duality*.

A task-based model needs to be adequate in that it must account for all relevant *a priori* knowledge or efficacious hypotheses in relation to a specific goal. Its formalisation has to connect to the image-based model, but must not be muddled up with it, because it is merely an interpretation of image evidence that may have to be revised in order to successfully accomplish a task. In this article this has been accomplished by expressing the task in terms of gauge conditions imposed on a gauge invariant system derived from the image-based model.

The successfulness of the outlined approach for accounting for models when applied to motion extraction suggests application to problems in a more general image analysis context.

References

[1] A. A. Amini. A scalar function formulation for optical flow. In J.-O. Eklundh, editor, *Proceedings of the Third European Conference on Computer Vision (Stockholm, Sweden, May 1994)*, volume 800–801 of *Lecture Notes in Computer Science*, pages 125–131, Berlin, 1994. Springer-Verlag.

[2] J. Arnspang. Motion constraint equations based on constant image irradiance. *Image and Vision Computing*, 11(9):577–587, November 1993.

[3] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[4] P. Delogne, editor. *Proceedings of the International Conference on Image Processing 1996 (Lausanne, Switzerland, September 1996)*. IEEE, 1996.

[5] V. Devlaminck and J. Dubus. Estimation of compressible or incompressible deformable motions for density images. In Delogne [4], pages 125–128.

[6] J. D'Haeyer. Determining motion of image curves from local pattern changes. *Computer Vision, Graphics, and Image Processing*, 34:166–188, 1986.

[7] O. D. Faugeras and T. Papadopoulo. A theory of the motion fields of curves. *International Journal of Computer Vision*, 10(2):125–156, 1993.

[8] D. J. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5(1):77–104, 1990.

[9] L. M. J. Florack. Data, models, and images. In Delogne [4], pages 469–472.

[10] L. M. J. Florack. *Image Structure*. Computational Imaging and Vision Series. Kluwer Academic Publishers, 1997. To appear.

[11] L. M. J. Florack. The intrinsic structure of optic flow in the context of the scale-space paradigm. In T. Moons, E. Pauwels, and L. Van Gool, editors, *From Segmentation to Interpretation and Back: Mathematical Methods in Computer Vision*, Lecture Notes in Computer Science. Springer-Verlag, Berlin, 1997. Selected papers from the Computer Vision and Applied Geometry workshop in Nordfjordeid, Norway (in press).

[12] L. M. J. Florack. Scale-space and measurement duality. In Sporring et al. [26], chapter 5, pages 61–74.

[13] L. M. J. Florack and M. Nielsen. The intrinsic structure of the optic flow field. Technical Report ERCIM-07/94-R033 or INRIA-RR-2350, INRIA Sophia-Antipolis, France, July 1994. URL: http://www-ercim.inria.fr/publication/technical_reports.

[14] L. M. J. Florack, W. J. Niessen, and M. Nielsen. The intrinsic structure of optic flow incorporating measurement duality. To appear in the International Journal of Computer Vision, 1997.

[15] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[16] J. J. Koenderink. Local image structure. In P. Johansen and S. Olsen, editors, *Theory & Applications of Image Analysis*, volume 2 of *Series in Machine Perception and Artificial Intelligence*, pages 15–21. World Scientific, Singapore, 1992. Selected papers from the 7th Scandinavian Conference on Image Analysis.

[17] J. J. Koenderink. What is a "feature"? *Journal of Intelligent Systems*, 3(1):49–82, 1993.

[18] J. J. Koenderink and A. J. v. Doorn. Receptive field families. *Biological Cybernetics*, 63:291–298, 1990.

[19] W. J. Niessen, J. S. Duncan, L. M. J. Florack, B. M. t. Haar Romeny, and M. A. Viergever. Spatiotemporal operators and optic flow. In T. S. Huang and D. N. Metaxas, editors, *Workshop on Physics-Based Modeling in Computer Vision*, pages 78–84, Cambridge, Massachusetts, June 18–19 1995. IEEE Computer Society Press.

[20] W. J. Niessen, J. S. Duncan, B. M. t. Haar Romeny, and M. A. Viergever. Spatiotemporal analysis of left ventricular motion. In *Medical Imaging 95*, pages 192–203, San Diego, February 1995. SPIE.

[21] W. J. Niessen, J. S. Duncan, M. Nielsen, L. M. J. Florack, B. M. t. Haar Romeny, and M. A. Viergever. A multiscale approach to image sequence analysis. *Computer Vision and Image Understanding*, 65(2):259–268, February 1997.

[22] W. J. Niessen and R. Maas. Optic flow and stereo. In Sporring et al. [26], chapter 3, pages 31–42.

[23] L. Schwartz. *Théorie des Distributions*. Hermann, Paris, second edition, 1966.

[24] M. Spivak. *Calculus on Manifolds*. W. A. Benjamin, New York, 1965.

[25] M. Spivak. *Differential Geometry*, volume 1–5. Publish or Perish, Berkeley, 1975.

[26] J. Sporring, M. Nielsen, L. M. J. Florack, and P. Johansen, editors. *Gaussian Scale-Space Theory*, volume 8 of *Computational Imaging and Vision Series*. Kluwer Academic Publishers, Dordrecht, 1997.

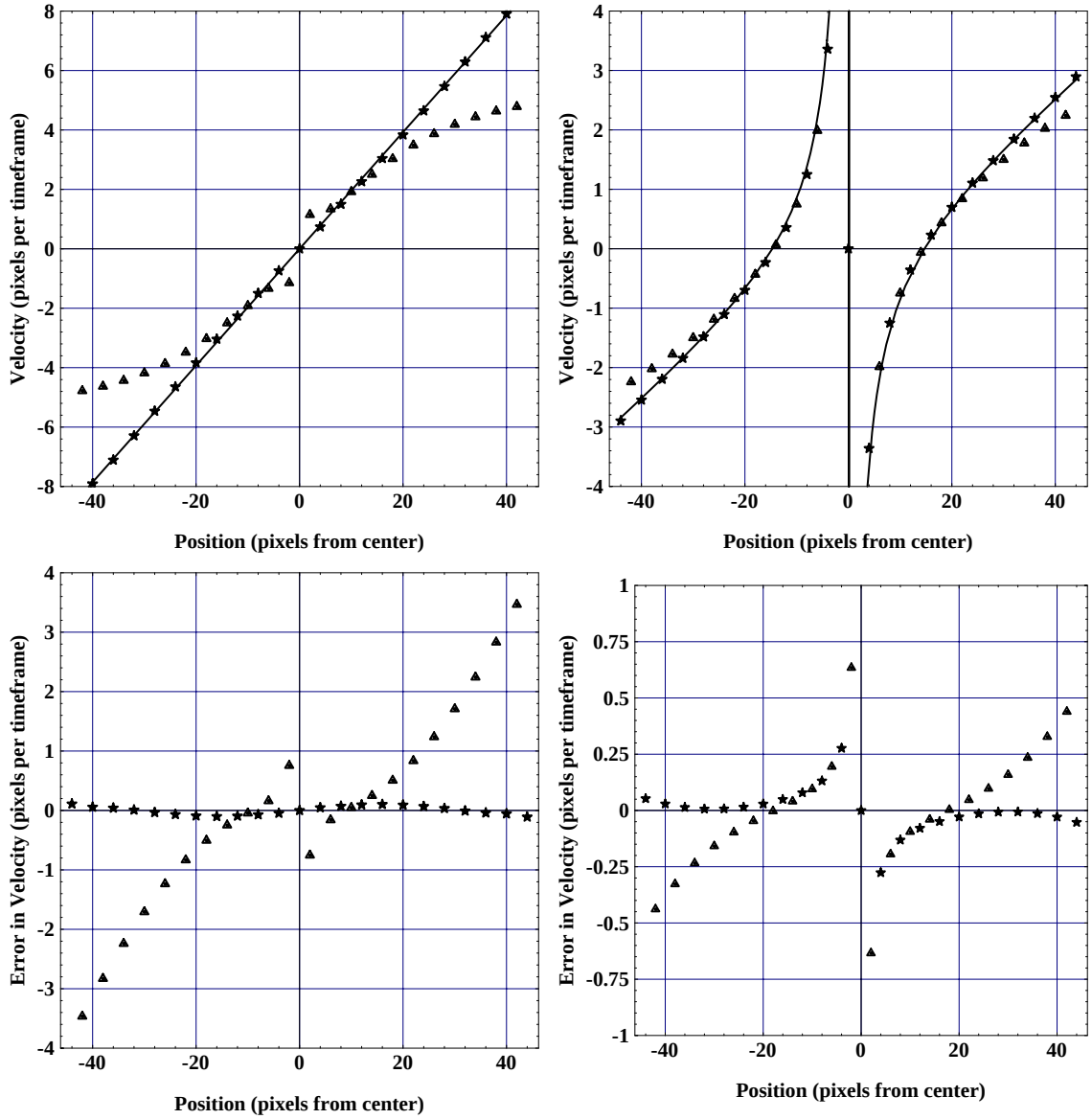

Figure 3. Cross-section of flow field (top) and error magnitude (bottom) along horizontal scanline through the centre for scalar (left) and density stimulus (right). Solid line: analytical curve; triangles: M=0; stars: M=1.

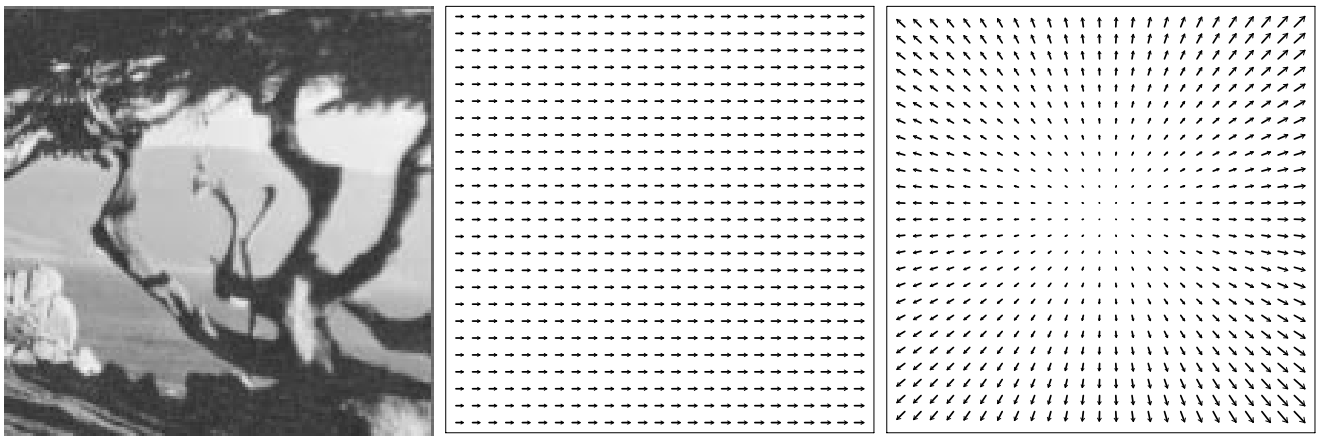

Figure 4. Textured surface and vector field for pure translation and divergence.

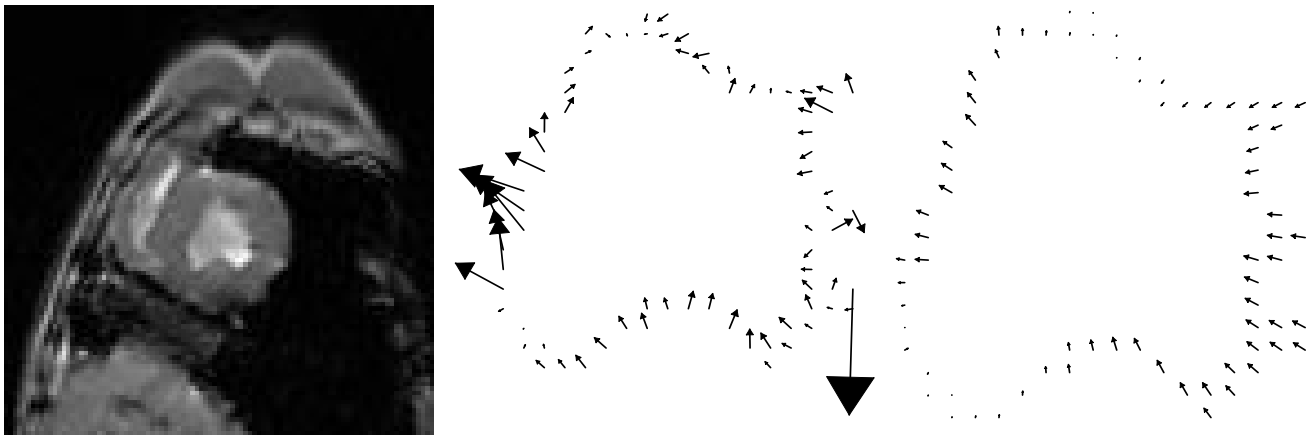

Figure 5. Left: one frame from an MR sequence of a canine heart. Resolution: 100 by 100 pixels, 16 frames per cardiac cycle. Middle: calculation using the M=1 scheme at fixed scales, restricted to the—manually segmented—endocardial wall: $\sigma = 2, \tau = 1$. Right: likewise, using scale selection on the basis of a stability criterion.