Motion Extraction*

An Approach Based on Duality and Gauge Theory

Luc Florack
Department of Computer Science, University of Utrecht
PO Box 80089, NL-3508 TB Utrecht, The Netherlands
Email: Luc.Florack@cs.ruu.nl

April 24, 1998

1 Introduction

Vicious circularities pervade the field of image analysis. For instance, features like "edges" only exist by virtue of a fiducial "edge detector". In turn, such a detector is typically constructed with the aim to extract those features one is inclined to classify as "edges".

The paradox arises from abuse of terminology. The polysemous term "edge" can be used in two distinct meanings: as an operationally defined concept (output of an edge detector), or as a heuristic feature pertaining to our intuition. In the former case the design of edge detection filters is—*strictu sensu*—merely a convention for imposing *structure* on raw data. In the latter case it is our expectation of what an "edge" should be like that begs the question of appropriate detector design. The keyword then becomes *interpretation*.

Clearly all low-level image concepts pertain to structure as well as interpretation. Once defined, structure becomes evidence. Interpretation amounts to a selection among all possible hypotheses consistent with this evidence. Clarity may be served by a *manifest* segregation of the two. A convenient way to achieve this is to embed "structure" into a framework of *duality* and to model "interpretation" by a hermeneutic circle driven by external insight constraining the class of *a priori* feasible interpretations (*gauge conditions*, respectively *gauge invariance*).

Here the proposed framework is applied to motion analysis. Duality accounts for the role of preprocessing filters. The notorious "aperture problem" arises from an intrinsic local invariance (or gauge invariance), which cannot be resolved on the exclusive basis of image evidence. Gauge conditions reflect external knowledge for disambiguation.

In a similar fashion, two stages can be distinguished in an error analysis of the outcome. There are errors of the obvious kind, caused by inadequate modelling ("semantical errors", or "mistakes"), which one would like to remove *altogether*, and subtle but inevitable errors propagated by any structural representation of data of intrinsically finite tolerance. Indeed, the flexibility to alter the gauge (re-interpret the data) and the possibility to carry out a rigorous error propagation study for the data formatting stage is a major rationale behind the current framework, *cf.* Fig. 1 and 2.

*This report is based on a presentation held at the Dagstuhl Seminar, Saarbrücken, Germany, March 16–20 1998.

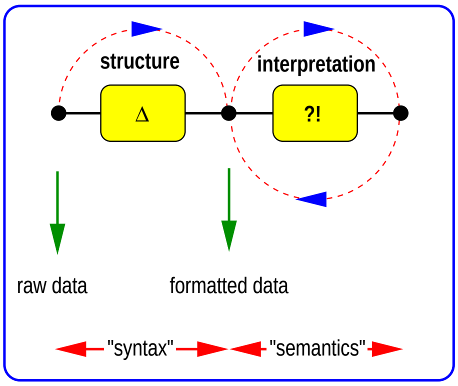

Figure 1: Manifest segregation of structural and semantic representations.

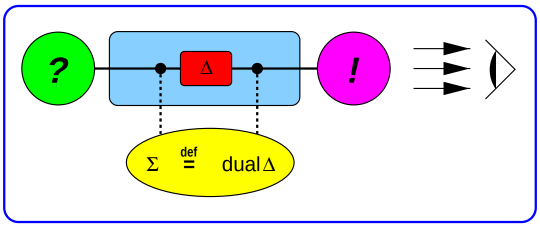

Figure 2: In a duality formalism "structure" means "operationally defined structure". Degrees of freedom captured by raw data $f \in \Sigma$ are identified with probes of a fiducial filter class Δ, *i.e.* with mappings "dual Δ" : $\Delta \longrightarrow \mathbb{R}$ induced by the raw data through exposure to all members $\phi \in \Delta$. In particular, if two source configurations induce identical mappings they are considered equivalent ("metamerism"). This can be exploited so as to hide irrelevant grid details and alleviate the impact of noise (robustness).

The emphasis is on global methodology. Technical details can be found in a recent publication [5].

2 Theory

Duality paradigms account for the inner workings of the filtering stage used to define data format. As such they should be contrasted with conventional "preprocessing" or "regularisation" techniques, in which the emphasis is on suitable preparation of data for subsequent processing. Rigidity of a fixed preprocessing or regularisation stage conflicts with the plasticity required for solving specific tasks of which all details cannot possibly be known in advance. Modelling image structure by duality principles on the other hand manifestly captures the public facts that one always needs a filtering stage to define the basic structural degrees of freedom driving image algorithms (*necessity*), and that outcome always depends crucially on the details of this stage (*criticality*). The latter observation is reflected in the way filters are handled in a dualistic approach, *viz.* as free (albeit mandatory) arguments to algorithms. This view makes the role of filters transparent.

The concept of duality leaves ample leeway for implementation. One of the simplest options is "topological duality". It basically boils down to linear filtering with smooth, essentially compact filters, though one should always keep in mind that it is not output in itself ("black box") but in connection to its production ("glass box") that is of interest. Originally proposed by Schwartz [10] as a mathematical formalism it may serve as a generic framework for many linear image processing filters used today. Topological duality subjected to a few plausible constraints produces the familiar Gaussian scale-space paradigm (*viz.* postulate a unique, positive filter consistent with Schwartz' theory and require algebraic closure [4]). In §2.1 it is extended to cope with motion.

Once a generic data format has been established, solutions to particular tasks typically depend on fewer degrees of freedom than actually available (provided one has sufficient data and knowledge). Indeed, genericity encourages redundancy, but at the same time has the potential of facilitating the selection of degrees of freedom that are relevant to a specific problem.

Gauge invariant representations—common in physics—are characterised by pointwise redundancies induced by the deliberate use of nonphysical variables (auxiliaries). The idea is that models may become most parsimonious in terms of redundant systems with constraints that cancel the effect of nonphysical degrees of freedom. Variables could be isolated such that the additional constraints (gauge conditions) become obsolete, but only at the price of an increase of model complexity. Gauge theoretical principles and their use in the context of motion are further discussed in §2.2.

2.1 Duality in the Context of Motion

An *a priori* condition for the definition of a dense motion field is local conservation. Some well-defined local characteristic must retain its identity in order to enable us to monitor point trajectories over time. This condition is necessary but does not suffice to define unambiguous motion (except in the case of time varying one-dimensional signals).

Since conservation is a generic principle data format can be defined so as to incorporate it *a priori* ("kinematic structure"). That is, the basic elements in the analysis are of a kinematic nature, encapsulating the "proto-semantics" that enjoys public consensus[1]. Further disam-

[1]The premiss is that one agrees on the quantity that is actually conserved. A direct link with image data

biguation of motion requires specific models, depending on task, image formation details, *et cetera* (the "aperture problem"). In view of specificity this is best left as an interpretation task. In this section only structural aspects are discussed.

Local conservation principles are often stated in terms of a vanishing Lie derivative, which can in turn be expressed in terms of an ordinary derivative and a vector field. For a scalar function f the Lie derivative is given by $\mathrm{L}_{\mathrm{v}} f = \nabla f \cdot \mathrm{v}$, in which ∇ denotes the spatiotemporal gradient operator. For a density field ϱ one has $\mathrm{L}_{\mathrm{v}} \varrho = \nabla(\varrho \cdot \mathrm{v})$ [3]. The original "Horn & Schunck equation" [6, 7, 9] is obtained by identifying f or ϱ with the image function and setting its Lie derivative equal to zero under the additional assumption that the temporal component of the vector field equals one.

Classical derivatives are ill-posed. Their counterparts in the setting of topological duality are not only well-posed but also operationally well-defined. If $f[\phi]$ denotes a linear sample obtained from "raw image" f by linear filtering with filter ϕ, then a derivative sample is defined as

$$\nabla f[\phi] \stackrel{\text{def}}{=} f[\nabla^{\mathrm{T}} \phi] \,,$$

in which A^{T} denotes the transposed of a linear operator A. Generalisation to higher orders is straightforward. The base point associated with such a sample is the filter's centre of gravity, while resolution is the inverse of the filter's width. In any case, derivatives are defined by virtue of a filter paradigm. This allows us to define a Lie derivative as follows:

$$\mathrm{L}_{\mathrm{v}} f[\phi] \stackrel{\text{def}}{=} f[\mathrm{L}_{\mathrm{v}}^{\mathrm{T}} \phi] \,.$$

Again, the definition cannot be unconfounded from a fiducial filter class. It follows that motion, if defined along the lines of Horn & Schunck, has no existence on its own, but only relative to the filter paradigm in use.

According to a famous theorem a linear continuous sample $f[\phi]$ can be written in integral form as follows:

$$f[\phi] = \int f(x)\, \phi(x)\, dx \,,$$

from which it follows that transposition of a derivative brings in a minus sign: $\nabla^{\mathrm{T}} = -\nabla$ and $\mathrm{L}_{\mathrm{v}}^{\mathrm{T}} = -\mathrm{L}_{\mathrm{v}}$. A subtlety arises in the case of transposing Lie derivatives: if f is a scalar, then ϕ behaves as a density, *vice versa*. Under the assumption of homogeneity the transition from local samples to images is trivial, and leads to similar expressions with spacetime correlations—or, if one prefers, convolutions, in which case the minus signs are absorbed—instead of scalar products. The resulting motion constraint equation is homogeneous and trilinear with respect to input data f (whether scalar or density), correlation filter ϕ, and spacetime vector v.

2.2 Gauge Theory in the Context of Motion

Let v be the desired motion field satisfying the motion constraint equation:

$$\mathrm{L}_{\mathrm{v}} f[\phi] = 0 \,.$$

In principle this fixes 1 component of v per base point, leaving n undefined in $(n+1)$-dimensional spacetime (typically $n = 2$ or 3). In the terminology introduced previously one may say that v is a gauge field with 1 physical and n auxiliary components. It would complicate matters greatly if one would choose to dispense with the auxiliaries beforehand, and in fact could even

requires a careful acquisition protocol and a quantitative reconstruction, *e.g.* proton density cine-MR.

obscure the very motion concept completely, since this is a semantical concept that requires additional knowledge in conjunction with the above constraint equation. Rather, the natural way to proceed is to enforce additional constraints to disambiguate the solution. Such gauge conditions reflect knowledge inspired by the application and other external factors.

However, along with the above equation one additional hypothesis is always tacitly adopted. One could describe it as *conservation of topological detail*. It entails that one takes for granted that the flow induced by v is transversal to spatial frames, so that (by virtue of homogeneity) one can always scale the temporal component to unity: $\mathrm{v} = (1; \vec{\mathrm{v}})$, say. It is the *spatial* part $\vec{\mathrm{v}}$ that is commonly associated with the optic flow or motion vector. One explanation of this "temporal gauge" is that whatever it is that moves cannot reverse its temporal sense and "travel backward in time". But there is an alternative and more natural one, which at the same time shows that the temporal gauge is not self-evident, and even unrealistic if strictly enforced: A time-reversal of a spacetime trajectory can always be given the causal interpretation of a creation or annihilation event. In this way one can *e.g.* account for enhancement of extrema in a scalar image sequence that would otherwise turn into spurious flow singularities.

All validation studies of §3 adhere to the usual temporal gauge. In §3.1 the consequences of this are discussed for the case where it is not appropriate to do so, while §3.2 presents a simulation where it is. Since semantics is de-emphasised the "canonical gauge" will be adopted expressing the normal flow condition. The $n-1$ normal flow equations may be replaced by physical conditions without technical difficulties. (This is done in §3.3.) Recall that this semantical flexibility lies at the core of gauge theoretical formalisms.

3 Validation

In the validation study outlined below one of the aims is to isolate the effect of semantical weaknesses. Errors due to measurement noise and numerical approximations of image derivatives are not quantified in the tables (but of course may contribute significantly to the results listed). Such intrinsic errors set a lower bound on overall errors beyond which no improvement is theoretically possible.

In conformity with scale-space theory all experiments are based on the Gaussian filter family [8], a complete, proper subset of Schwartz' space. For an analysis of error propagation in computing Gaussian derivatives the reader is referred to Blom *et al.* [2]. Below "error" is defined as "deviation from the model". No attempt is made to relate it quantitatively to a theoretical prediction of the aforementioned fundamental limitation. To compensate for this two related experiments are carried out on synthetic data. The first simulates an intentionally deficient model containing a feasible semantical error (§3.1). The second is set up without this deficiency, so that the only potential source of error affecting the solution stems from data noise and numerical manipulations (§3.2). Both simulations are analytically tractable, so that numerical results can be compared with theoretical predictions. In a third study, which is carried out on real image data, there is no direct control over the stimulus, but in this case external knowledge of scene configuration and image formation enables the formulation of analytical gauge conditions (§3.3).

3.1 Simulation Study: Density Stimulus, Scalar Paradigm

Imagine a bell-shaped stimulus with oscillating radius somewhere in the middle of the image. If its amplitude covaries in such a way that its spatial integral ("mass") remains constant over

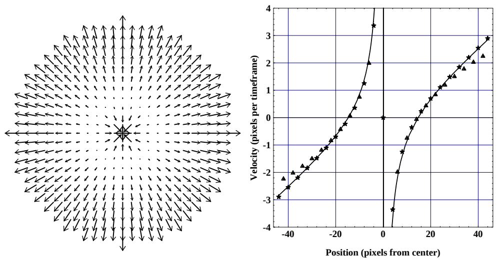

Figure 3: Flow field obtained for the density stimulus (left) and flow magnitudes along a horizontal scanline through the centre (right) showing analytical (solid line), 0-th (triangles) and 1-st order results (stars). Note the singularity and the flow inversion.

time, then the sequence simulates density motion. Think of proton density cine-MR by way of example.

Let us endow the motion constraint equation with the following gauge conditions: (i) the normal flow condition, *in casu* the requirement that the motion field is radial, and (ii) the usual temporal gauge. With the blob centred at the coordinate origin we then have $\mathrm{v} \propto (1; \vec{\mathrm{r}})$, in which the proportionality factor depends on $\|\vec{\mathrm{r}}\|$. If motion is well-defined in the first place this constant should have no singularities, and in view of symmetry the only reasonable motion vector to expect at the origin is the null vector, regardless of the filter paradigm.

Suppose, however, that we make the mistake of modelling the image sequence as a time-varying *scalar* field. Such a misinterpretation is not far-fetched in practice for several reasons. Firstly one may lack adequate knowledge of image formation, so that one does not know whether the image captures a density field at all. Even if it does, one may not understand the exact relation between image values and physical density. Secondly there will be deviations from any definite geometric paradigm, either due to plain noise or to the fact that the paradigm is merely an idealization. In particular some motion sequences are neither densities nor scalars, *e.g.* (typical) shading in optical projection imagery.

For the simulated density the consequence of the scalarity assumption is the appearance of a spurious motion singularity at the centre of the blob. Apart from this there are other qualitative discrepancies between visual percept (one observes alternating contractions and expansions) and prediction (simultaneous inward and outward flow on two sides of a circle oscillating in phase with the blob). Theoretical predictions are in quantitative agreement with numerical computations carried out on a digital rendering of the density sequence, *cf.* Fig. 3. Two numerical schemes have been used, a 0-th and a 1-st order one, the details of which are given elsewhere [5]. Errors turn out to be largest in the immediate neighbourhood of the singularity. It is clear that in realistic sequences with complex topological structure there will be many such problematic neighbourhoods.

If one retains faith in conservation, there are two legitimate explanations for model failure. Either the scalarity assumption fails, or one must allow for point sources and sink-holes, *i.e.* give up the temporal gauge. Note that the numerical schemes yield accurate estimates; failure does *not* have a computational cause. In fact, computations turn out to be quite robust.

3.2 Simulation Study: Scalar Stimulus, Scalar Paradigm

One would expect no problems if the motion constraint equation had been used in the appropriate form applicable to densities. Likewise, no singularities should emerge if we adhere to

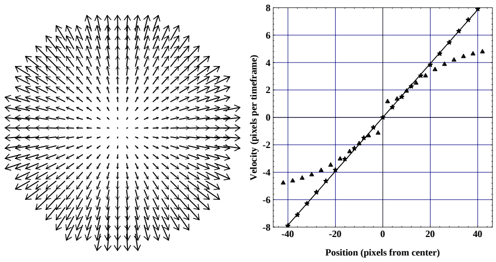

Figure 4: Flow field obtained for the scalar stimulus (left) and flow magnitudes along a horizontal scanline through the centre (right) showing analytical (solid line), 0-th (triangles) and 1-st order results (stars). Note that the field is everywhere well-defined and has no inversions (in agreement with perceptual impression).

the scalar model but slightly adapt the stimulus by taking the blob's amplitude to be constant. Theory then predicts a motion field that (for a harmonically oscillating Gaussian blob) varies linearly with eccentricity. Again, this is confirmed numerically: Fig. 4.

3.3 A Comparative Study

We subject the motion paradigm to a final test to check whether it has any practical advantages over alternative schemes proposed in the literature. To this end we fully exploit the flexibility of semantical modelling enabled by the manifest segregation of stages (Fig. 1), as well as the theoretical properties of the filter paradigm (Fig. 2). In the concrete, we (i) endow the basic structural equations (the motion constraint equation in temporal gauge) with additional spatial constraints reflecting *a priori* knowledge of camera motion and scenery—this should be contrasted with generic schemes that do not incorporate such specific knowledge—and (ii) exploit the scale degree of freedom of the Gaussian family by *scale selection*.

The motion algorithm derived from the theory is the 1-st order scheme detailed elsewhere [5], in which filter scales are selected so as to (pixelwise) minimize the Frobenius norm of the resulting linear system. It is run on benchmark sequences known as the *translating* and the *diverging tree sequence* ("TTS", respectively "DTS"): Fig. 5. Outcome is compared to the comprehensive study of Barron *et al.* [1] using the same error criterion. The tentative gauges reflect the hypothesis that vertical motion is absent, respectively that the focus of expansion is known. (The "hermeneutic principle" relies on the existence of consistent cues conspiring to produce such tentative hypotheses, and on the possibility to test and refine them.) Results are listed in Table 1 (dense flow estimation) and Table 2 (sparse flow estimation discarding uncertain estimates).

4 Conclusion

The strength of duality is that the role of filters is made transparent, thus facilitating the exploitation of filter properties. In the case at hand Gaussian scale-space filters have been used and scale selection has been applied successfully for stable motion extraction.

The gauge field paradigm encourages clarity and parsimony. It has been applied here for a manifest segregation of data evidence (gauge invariant system) and external models (gauge conditions), and has led to a flexible operational scheme for combining motion evidence with

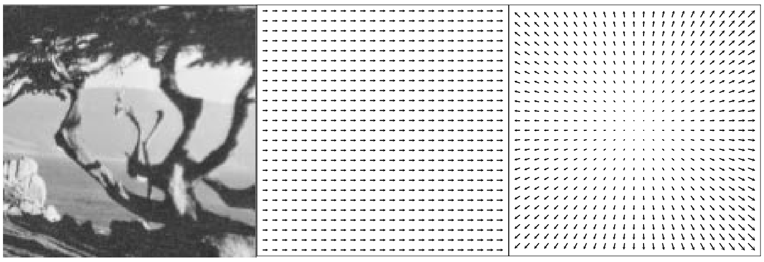

Figure 5: Textured plane and vector field for translation (TTS) and divergence (DTS).

Implementation method	**TTS** μ	**TTS** σ	**DTS** μ	**DTS** σ
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 *et al.* ($M = 1$, scale selection)	0.49	1.92	1.15	3.32

Table 1: Comparison with best performing techniques [1] with dense velocity estimates; μ and σ denote mean and standard deviation of error.

Implementation method	**TTS** μ	**TTS** σ	**TTS** ϱ (%)	**DTS** μ	**DTS** σ	**DTS** ϱ (%)
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 ($\tau = 2.0$)	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 *et al.* ($M = 1$, scale selection)	0.16	0.18	60.0	0.79	1.13	60.0
Florack *et al.* ($M = 1$, scale selection)	0.14	0.13	40.0	0.43	0.40	40.0

Table 2: Comparison with best performing techniques [1] discarding uncertain velocity estimates; μ and σ denote mean and standard deviation of error, ϱ indicates pixel fraction with motion estimates.

prior knowledge. Numerical results vote in favour of the proposed line of approach.

Acknowledgement

Experiments have been carried out by Wiro Niessen at the Image Sciences Institute, Utrecht, The Netherlands.

References

[1] 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.

[2] J. Blom, B. M. ter Haar Romeny, A. Bel, and J. J. Koenderink. Spatial derivatives and the propagation of noise in Gaussian scale-space. *Journal of Visual Communication and Image Representation*, 4(1):1–13, March 1993.

[3] J. M. Fitzpatrick. The existence of geometrical density-image transformations corresponding to object motion. *CVGIP: Image Understanding*, 44:155–174, 1988.

[4] L. M. J. Florack. *Image Structure*, volume 10 of *Computational Imaging and Vision Series*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997.

[5] L. M. J. Florack, W. J. Niessen, and M. Nielsen. The intrinsic structure of optic flow incorporating measurement duality. *International Journal of Computer Vision*, 27(3):263–286, 1998.

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

[7] B. K. P. Horn and B. G. Schunck. "determining optical flow": a retrospective. *Artificial Intelligence*, 59:81–87, 1993.

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

[9] B. G. Schunck. The motion constraint equation for optical flow. In *Proceedings of the 7th International Conference on Pattern Recognition*, pages 20–22, Montreal, Canada, 1984.

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