

# Revisiting Horn and Schunck: Interpretation as Gauß-Newton Optimisation

Darko Zikic<sup>1</sup>

zikic@in.tum.de

Ali Kamen<sup>2</sup>

ali.kamen@siemens.com

Nassir Navab<sup>1</sup>

navab@in.tum.de

<sup>1</sup> Computer Aided Medical Procedures

Technische Universität München

Munich, Germany

<sup>2</sup> Siemens Corporate Research (SCR)

Princeton, NJ, USA

---

## Abstract

We revisit the classic Horn and Schunck optical flow method, with focus on its interpretation as a Gauss-Newton (GN) minimisation of a non-linear energy. This is in contrast to the traditional derivation by linearisation of the brightness constancy assumption. The proposed interpretation provides a much simpler derivation and a better theoretical understanding of the method, and allows for its variations, by casting it in the least-squares optimisation framework. An important resulting implication is that – in contrast to popular belief – the incremental version of Horn and Schunck actually minimises a non-linear energy. We emphasise this finding by demonstrating the equivalence of incremental Horn and Schunck to several methods committed to minimising the non-linearised energy. Furthermore, we analyse the effect of GN for motion estimation in comparison to methods based on gradient descent. Several examples demonstrate the practical applications of the proposed interpretation. We specify the class of difference measures with sparse Jacobians of the error term as the one which can be efficiently treated in the Horn and Schunck framework. For extension to arbitrary difference measures, we propose a modification based on the analysis of GN for motion estimation, and the concept of preconditioning. We further discuss a modification resulting in decoupled linear systems and the use of compositional updates.

## 1 Introduction

Just as Western philosophy can be seen to “consist of a series of footnotes to Plato” [65], it is tempting to state that the development of optical flow methods in the last three decades is a series of footnotes to the seminal work of Horn and Schunck [16, 17]. While this of course is an exaggeration, it is significant of the importance of Horn and Schunck that optical flow methods are still mostly presented against its background: the vast majority of papers is based on a derivation of Horn and Schunck, followed by the proposed modifications.

Since its introduction, the method of Horn and Schunck has been extended to incremental versions and to hierarchical coarse-to-fine approaches [15, 25, 26]. Also, a lot of work has been dedicated to modification of regularisation to allow for discontinuities, resulting in piece-wise smooth displacements. The two major strategies regarding this point include

anisotropic regularisation, and the use of robust norms, compare for example [10, 25] and [9, 34]. For registration of medical images, the modelling of discontinuities has not received as much attention, and the regularisation model of Horn and Schunck is still widely used [24]. It is interesting to note that in spite of its age, incremental implementations of Horn and Schunck, which utilise a hierarchical coarse-to-fine strategy, still perform comparably well<sup>1</sup>, while featuring very simple and fast implementations, and optimised approaches based on Horn and Schunck can achieve high accuracy [33].

Since the Horn and Schunck method plays such a crucial role in the world of optical flow, we consider it important to have well-founded and diverse interpretations of this method. In this work, we show that the incremental Horn and Schunck (HS) method is equivalent to Gauss-Newton optimisation of the non-linear energy consisting of the SSD criterion and diffusion regularisation (Sec. 2), and focus on this interpretation and its implications.

Surprisingly, although there are several references to the relation between HS and GN scattered in the literature, this view does not seem to be widespread in the community. The fact is neither mentioned in the original publications [16, 17], nor in a retrospective article [18] by Horn and Schunck. There is a long list of renown works (e.g. [2, 9, 11, 12, 25, 26, 28]) treating large displacements and non-linearised methods, which do not employ this interpretation. It is also not treated in several popular reviews [6, 8, 33].

As mentioned, there are several works which exhibit an awareness of the relation between HS and GN. Most of these works focus on proposing specific new methods, so that the comments on the relation between HS and GN are mostly made in passing and not explicit. The early overview of motion estimation by Bergen in [7] is based on hierarchical coarse-to-fine approaches and Gauss-Newton optimisation, however it concentrates on parametric models, and does not treat HS. Bergeron and Dubois [8] propose a block-based approach and in Sec. 4 state that GN “shares the philosophy” of HS. Mémin and Pérez [21, 23] notice the Gauss-Newton character of a formulation which is de facto equivalent to the incremental version of Horn and Schunck - however, without explicitly pointing out this equivalence, or discussing the implications. The closest work is by Konrad [19], which extends the HS model to colour images. In this context it discusses the difference between original and incremental HS, and explicitly claims the equivalence between incremental HS and GN. However, as mentioned, the relation between HS and GN is not the focus of any of the above works.

The proposed interpretation is advantageous as it places the Horn and Schunck method in the well understood framework of non-linear least-squares (NLSQ) optimisation, thus providing insight and intuition as to the behaviour of Horn and Schunck. We consider the focus on the interpretation of Horn and Schunck as GN and the resulting theoretical implications to be the major contribution of this paper, and see the discussed variations more as simple examples for possible practical applications of the theoretical considerations. The following implications and modifications are discussed in this paper.

- The incremental HS actually optimises a non-linearised energy formulation, and not - as often stated - a linearised form. This fact is easily overseen in the traditional derivation, and becomes obvious with the proposed interpretation. In fact, the GN optimisation in the HS method can be seen as a specific numerical scheme for time discretisation, comparable to those of other variational methods. We argue that for the task of motion estimation, the linearisation by Gauss-Newton has favourable properties, and emphasise this point by demonstrating that several renown methods exhibit an incremental process identical to the one of Horn and Schunck.

---

<sup>1</sup>Compare the Middlebury Optical Flow evaluation database [4] at <http://vision.middlebury.edu/flow/>.

- We analyse the effect of GN on motion estimation with SSD, compared to methods based on steepest gradient descent. The characteristic effect of GN seems to be the approximate equalisation of the local update magnitudes for locations with differently strong image gradients. This effect results in a high convergence rate of HS in low gradient areas.
- The interpretation as Gauss-Newton allows us to specify the class of difference measures which can be efficiently treated in the HS framework as the ones which feature a sparse Jacobian of the error term. This is the case for the so-called constancy assumptions, but not for statistical difference measures (such as Correlation Ratio, or Mutual Information), which are commonly used for multi-modal registration of medical images.
- As examples for practical applications of the proposed analysis, we perform several variations on the theme of Horn and Schunck. The variations are theoretically justified by identifying GN as a specific preconditioning technique, and subsequent modification of the preconditioner. We present a strategy with a similar effect as GN, but which is applicable to arbitrary difference measures, and a further scheme, which results in decoupled linear systems, which might be preferable for large systems. Within the NLSQ framework, we discuss composition as alternative to the additional update scheme.

## 2 Interpretation of Horn and Schunck as Gauss-Newton

In this section we demonstrate that the incremental method of Horn and Schunck performs a Gauss-Newton optimisation of the energy consisting of the SSD criterion and diffusion regularisation. We start by presenting the traditional derivation of Horn and Schunck in Sec. 2.1, continue with the basic layout of the Gauss-Newton optimisation method in Sec. 2.2, and conclude by showing the proposed equivalence in Sec. 2.3. In Sec. 2.4, we treat some aspects of the arising linear system, which are of importance in this paper.

### 2.1 Horn and Schunck

Traditionally, the Horn and Schunck method is derived in the following manner. Given the image function  $I(x, t)$  over space and time, and the displacement  $u = (u_x, u_y)$ , then the *brightness constancy assumption* states that

$$I(x + u(x), t) = I(x, t + \tau). \quad (1)$$

Please note that we define the spatial coordinate system with respect to  $t + \tau$ , resulting in backward warping, compare e.g. [3]. The linearisation of (1) by Taylor expansion yields

$$I(x, t) + \frac{\partial I(x, t)}{\partial u}^\top u(x) = I(x, t) + \frac{\partial I(x, t)}{\partial \tau} \tau. \quad (2)$$

Cleaning up (2) results in the *optical flow constraint* (OFC)

$$0 = -\frac{\partial I(x, t)}{\partial u}^\top u(x) + \frac{\partial I(x, t)}{\partial \tau} \tau. \quad (3)$$

By using  $\tau = 1$  to indicate the time difference between neighbouring frames, identifying  $I(x, t) = I_S(x)$  and  $I(x, t + 1) = I_T(x)$  with the *source* and the *target* image, and employing that for variational settings we have  $\partial I / \partial u = \nabla I$ , we get

$$0 = -\nabla I_S(x)^\top u(x) + (I_T(x) - I_S(x)). \quad (4)$$

The OFC from (4) is used as the difference term to be minimised in the Euclidean (L2) norm

$$E_D(u) = \frac{1}{2} \int_{\Omega} \left[ I_T(x) - I_S(x) - \nabla I_S(x)^\top u(x) \right]^2 dx. \quad (5)$$

Due to the so-called *aperture problem* which states that (5) is under-constrained, a regularisation term is added. For HS, this is a diffusion (first order) regularisation term

$$E_R(u) = \frac{1}{2} \int_{\Omega} \left( \|\nabla u_x(x)\|^2 + \|\nabla u_y(x)\|^2 \right) dx = \frac{1}{2} \int_{\Omega} \text{tr} \left( J_u(x)^\top J_u(x) \right) dx. \quad (6)$$

In summary, the Horn and Schunck method minimises the energy  $E = E_D + \lambda^2 E_R$ , that is

$$E(u) = \frac{1}{2} \int_{\Omega} \left[ I_T(x) - I_S(x) - \nabla I_S(x)^\top u(x) \right]^2 dx + \lambda^2 \frac{1}{2} \int_{\Omega} \text{tr} \left( J_u(x)^\top J_u(x) \right) dx, \quad (7)$$

by solving for the update which fulfils the necessary condition (Euler-Lagrange equation)

$$0 = \frac{\partial E}{\partial u}. \quad (8)$$

We will specify the linear system resulting from (8) in more detail in Sec. 2.4.

### 2.1.1 Incremental Horn and Schunck

In the original formulation of Horn and Schunck in [10], the solution of the linear system resulting from Eq. (8) is considered as the estimated displacement. In modern approaches, an incremental process is applied, in which the solution to (8) in one iteration is used as the initial estimate for the following one. This can be described by expressing the displacement  $u$  as  $u = U + h$ , consisting of the initial estimate  $U$  and an update  $h$ . Then, the displacement is computed by setting  $U = U + h$  after every iteration, until  $h$  becomes sufficiently small. Furthermore, the source image is warped by the current transformation estimate  $\Phi(x) = x + U(x)$  in every iteration. Overall, the following term is optimised in every iteration

$$\frac{1}{2} \int_{\Omega} \left[ I_T(x) - I_S(\Phi(x)) - \nabla I_S(\Phi(x))^\top h(x) \right]^2 dx + \lambda^2 \frac{1}{2} \int_{\Omega} \text{tr} \left( J_u(x)^\top J_u(x) \right) dx. \quad (9)$$

The updates are computed by solving (8) in every iteration. Please note that the original HS method is equivalent to the first iteration step of this process with  $U = 0$ .

Such an incremental approach - in combination with a coarse-to-fine strategy - is employed by Glazer in [11], compare also Anandan [12] for an overview. A closely related, incremental method is discussed by Nagel and Enkelmann [13], which is a generalisation of HS to anisotropic regularisation. Enkelmann [14] presents an extension of [13, 16] to a hierarchical coarse-to-fine setting. Please compare also Konrad [15].

It is noteworthy that most papers following this derivation do not explicitly state the *overall energy* which is minimised, but only the *energy which is minimised in every iteration*.

## 2.2 Gauss-Newton Optimisation

This section follows the presentation in [16]. Consider a non-linear least squares energy, based on the non-linear *error term*  $f$

$$E(x) = \frac{1}{2} f(x)^\top f(x). \quad (10)$$

The Gauss-Newton method iteratively computes the update  $h$  to the initial estimate  $x'$ , based on a linearisation  $l(h) \approx f(x' + h)$

$$l(h) \equiv f(x') + J_f(x')h. \quad (11)$$

Here,  $J_f$  is the *error Jacobian*. Inserting (11) in (10), and with the notation  $f = f(x')$  and  $J = J_f(x')$ , we obtain the approximated energy model  $L(h) \approx E(x' + h)$  as

$$L(h) = \frac{1}{2}f^\top f + h^\top J^\top f + \frac{1}{2}h^\top J^\top J h. \quad (12)$$

The Gauss-Newton method minimises (10) by computing an update  $h$  in each iteration, such that  $L(h)$  from (12) is minimised. This is done by setting

$$0 = \frac{\partial L(h)}{\partial h} = J^\top f + J^\top J h, \quad (13)$$

which results in the linear system

$$J^\top J h = -J^\top f = -\nabla E(x'). \quad (14)$$

After every iteration, the initial estimate is updated by  $x' = x' + \tau h$ , with  $\tau > 0$ .

To summarise: Gauss-Newton computes the updates as critical points of an approximated energy  $L$  in (12), which is based on linearised error term  $l$  from (11).

### 2.3 Horn and Schunck Is Gauss-Newton

Going back to the optical flow problem, we now set the energy to a linear combination of the sum of squared differences (SSD) and a diffusion regularisation term from (6)

$$E = \frac{1}{2} \int_{\Omega} (I_T(x) - I_S(x + u(x)))^2 dx + \lambda^2 \frac{1}{2} \int_{\Omega} \text{tr} \left( J_u(x)^\top J_u(x) \right) dx. \quad (15)$$

This energy can also be written in the form of (10), with reformulating it as

$$E = \frac{1}{2}f^\top f = \frac{1}{2}f_D^\top f_D + \lambda^2 \frac{1}{2}f_R^\top f_R, \quad (16)$$

with the error term  $f = [f_D, \lambda f_R]^\top$  composed of

$$f_D = [I_T - I_S \circ (\text{Id} + u)], \quad (17)$$

$$f_R = [\nabla u_x, \nabla u_y]. \quad (18)$$

By linearising the above  $f$ , we can arrive at the Gauss-Newton method. For the linearisation of  $f$ , it is important to note that only  $f_D$  is non-linear in  $u$ , because the computation of gradients of  $u_x$  and  $u_y$  is a linear operation in  $u$ . This results in  $l_R = f_R$  for the Gauss-Newton formulation. The linearisation is performed for  $u = U + h$ , with the initial estimate  $U$ , as discussed in Sec. 2.1.1, and results for  $f_D(x + U(x) + h(x)) \approx l_D(h(x))$  in

$$l_D(h(x)) \equiv I_T(x) - I_S(x + U(x)) - \nabla I_S(x + U(x))^\top h(x). \quad (19)$$

With the initial transformation  $\Phi(x) = x + U(x)$ , we get the linearised approximation  $L$  of the energy from (15), which is used by the Gauss-Newton method in every iteration

$$L(h) = \frac{1}{2} l(h)^\top l(h) = \frac{1}{2} [l_D(h), l_R(h)]^\top [l_D(h), l_R(h)] \quad (20)$$

$$= \frac{1}{2} \int_{\Omega} [I_T(x) - I_S(\Phi(x)) - \nabla I_S(\Phi(x))^\top h(x)]^2 dx + \lambda^2 \frac{1}{2} \int_{\Omega} \text{tr} \left( J_u(x)^\top J_u(x) \right) dx. \quad (21)$$

This is identical to the energy from (9) which is minimised in every iteration of the the incremental version of the Horn and Schunck method.

## 2.4 Euler-Lagrange Equations (Linear System)

As we arrive at the same energy to be optimised in every iteration for the incremental Horn and Schunck (9), and for the Gauss-Newton derivation in (21), the corresponding Euler-Lagrange equations (linear systems) are also identical, and are given by

$$\left( \nabla I_S(\Phi(x)) \nabla I_S(\Phi(x))^\top - \lambda^2 \Delta \right) h(x) = \underbrace{(I_T(x) - I_S(\Phi(x))) \nabla I_S(\Phi(x))}_{-\nabla E_D(\Phi)} + \lambda^2 \underbrace{\Delta U(x)}_{-\nabla E_R(\Phi)}. \quad (22)$$

Please note that this corresponds to the GN linear system from (14), with

$$J^\top J = \left( \nabla I_S(\Phi(x)) \nabla I_S(\Phi(x))^\top - \lambda^2 \Delta \right), \quad (23)$$

$$-J^\top f = (I_T(x) - I_S(\Phi(x))) \nabla I_S(\Phi(x)) + \lambda^2 \Delta U(x). \quad (24)$$

Please note that in the above system, the dimensions are coupled through the structure tensor term  $\nabla I_S(\Phi(x)) \nabla I_S(\Phi(x))^\top$ . The solution of (22) can be performed by standard linear solvers, such as Gauss-Seidel [10, 11], or Multigrid methods [12].

## 3 Implications

In this section we discuss three implications resulting from the interpretation of Horn and Schunck as Gauss-Newton optimisation.

### 3.1 Incremental Horn and Schunck Optimises Non-linearised Energy

The equivalence of incremental HS to optimisation of the energy from (15) by the Gauss-Newton method directly reveals that *incremental Horn and Schunck optimises the original non-linearised version of the energy*. The single steps of Gauss-Newton are performed by solving a linear system obtained by linearisation. However, the same holds also for other numerical schemes for variational methods, since the standard strategy for solving the static non-linear problem from (9) is to transform it into a dynamic linear problem, and compute the steady state solution.

We give two examples of methods which are committed to optimising the non-linearised energy ([13] and [14, 15]), and result in iterations processes identical to the incremental Horn and Schunck for an appropriate choice of regularisation. Please note that our focus is on the iteration process and therefore we apply the numerical schemes from [13] and [14, 15] to the

energy from (9), leaving aside the modifications of the regularisation terms in these works. As the first example, in [10] the solution is obtained by deriving a non-linear PDE, and then computing the solution as the steady state of the corresponding linear parabolic system, with an artificial time. The employed numerical scheme for iteratively arriving at the steady state consists of an implicit discretisation of the artificial time, with a subsequent *linearisation* of the PDE term corresponding to the difference measure. The operator of the resulting linear system (for the energy from (9)) is identical to the one of the incremental HS formulation. In [10, 28], the authors also assume a non-linearised energy model, and consider a semi-implicit discretisation of the artificial time. Again, a subsequent linearisation is performed, which results in the incremental Horn and Schunck formulation (for the energy from (9)).

We conclude that the use of GN can be seen as a specific numerical scheme, which is similar to other approaches, for which a linearisation is also performed, but at different stages of the derivation. The comparison of the discussed methods to HS highlights that the improvement of the results in the respective works is likely to stem from the modification of the regularisation terms (and use of other constancy assumptions in [10, 28]), rather than the employed numerical schemes for treatment of the non-linearised energy.

### 3.2 Analysis of the Effect of Gauss-Newton on Motion Estimation

For analysis purposes, we compare the GN scheme to an approach with semi-implicit time discretisation, cf. [24], which is given by

$$(\text{Id} - \lambda^2 \Delta) h(x) = (I_T(x) - I_S(\Phi(x))) \nabla I_S(\Phi(x)) + \lambda^2 \Delta U(x). \quad (25)$$

We chose this scheme rather than the standard steepest gradient descent, since it has better stability and convergence rate properties. The comparison will show that the characteristic effect of GN on motion estimation is the approximate equalisation of local update magnitudes, for locations with differently strong local gradients. We see that the GN scheme in (22) is specified by the structure tensor  $\nabla I_S \nabla I_S^\top$  in the linear operator, in contrast to the identity in (25). Note that the right hand side, which equals  $-\nabla E(\Phi)$ , is identical in (22) and (25) (we omit the argument  $\Phi$  to  $E$  in the following). Therefore, we analyse the effect of GN by analysing the effect of the structure tensor  $\nabla I_S \nabla I_S^\top$  on the right hand side  $\nabla E$ .

To this end, we first note that  $\nabla E = \nabla E_D + \lambda^2 \nabla E_R$  is dominated by  $\nabla E_D$ , which is the driving force until the steady state is reached for  $\nabla E_D = -\lambda^2 \nabla E_R$ . This effect is especially prominent in the early iterations. Furthermore, please note that  $-\nabla E_D = (I_T - I_S) \nabla I_S$  is an eigenvector to  $\nabla I_S \nabla I_S^\top$ . With this, and the approximation  $\|\nabla I_S(\Phi(x))\| \approx |I_T(x) - I_S(\Phi(x))|$  for small  $\Phi$  (think of  $I_S(\Phi(x))$  as locally slightly shifted version of  $I_T(x)$ ), we can express the effect of the structure tensor on the right hand side by

$$\nabla I_S \nabla I_S^\top \nabla E_D = \|\nabla I_S\|^2 \nabla E_D \approx \|\nabla E_D\| \nabla E_D. \quad (26)$$

With (26), the inverse of the application of the structure tensor, which is performed when solving the linear system in (22), results approximately in  $\|\nabla E_D(\Phi(x))\|^{-1} \nabla E_D(\Phi(x))$ .

From the above, we conclude that the characteristic effect of the Gauss-Newton optimisation (22) is that the local gradients  $\nabla E_D(\Phi(x))$ , which form a basis for the updates, experience a rescaling to approximately similar magnitudes - in contrast to the mere smoothing which is achieved by (25). This gives approximately the same weight to contributions from the complete image domain, independent on the local gradient strength, in contrast to (25), where updates are significantly smaller in low contrast regions, resulting in very slow convergence, cf. Fig. 1(a)-1(d), and [36].

### 3.3 Class of Difference Measures for Horn and Schunck Framework

For the large systems arising in motion estimation, it is essential that the error Jacobian  $J$ , and thus the term  $J^\top J$  from (23) is sparse. Only then, the solution of the linear system from (22) can be efficiently approximated. Therefore, we conclude that the Horn and Schunck framework can be applied only for difference measures with sparse error Jacobians. The so-called constancy assumptions are in this class, in contrast to statistical difference measures.

For dense motion estimation, the error Jacobian  $J$  expresses the local dependency of the error on the displacements. The sparseness of  $J$  is given for constancy assumptions, which are a generalisation of the brightness constancy, and have the form  $f_D = G(I_T) - G(I_S \circ \Phi)$ , with a differential operator  $G$  [24]. The sparseness of  $J$  for constancy assumptions results from the *locality* of the subtraction in  $f_D$ , with the number of non-zero diagonals of  $J$  depending directly on the order of  $G$ . For statistical difference measures [29] however, such as Correlation Ratio (CR) or Mutual Information (MI), which are used for multi-modal registration in medical settings, the corresponding  $J$  is not sparse. These measures are based on the joint probability distribution of the image intensities, which renders the error *globally* dependent on the displacements. In Sec. 4.1 we propose an extension to arbitrary difference measures, based on the analysis from Sec. 3.2.

## 4 Variations

In this section we discuss several simple variations on the theme of Horn and Schunck. Most are based on the observation that Gauss-Newton belongs to the class of preconditioned gradient descent methods [27]. Simply put, these methods modify the direction of the energy gradient by applying a symmetric positive definite *preconditioning operator*  $P^{-1}$ . The variations in Sec. 4.1, and 4.2 are performed by modifying the standard GN preconditioner  $P = J^\top J$ . Since only the optimisation scheme is modified, the treated energy remains unchanged. For an example of the performance of discussed variations, please compare Fig. 1(e).

### 4.1 Generalising Horn and Schunck to Arbitrary Difference Measures

As discussed in Sec. 3.3, the HS method is restricted to difference measures with a sparse error Jacobian. We can extend the HS method to arbitrary difference measures, by changing the preconditioner  $P = J^\top J$  of Gauss-Newton, based on the analysis from Sec. 3.2, which indicates that the crucial aspect of the GN scheme for motion estimation is the approximate normalisation of the magnitudes of the local updates. The preconditioning term

$$P(x) = \frac{1}{\|\nabla E_D(\Phi(x))\|_\varepsilon} \nabla E_D(\Phi(x)) \nabla E_D(\Phi(x))^\top - \lambda^2 \Delta \quad (27)$$

replaces the structure tensor  $\nabla I_S \nabla I_S^\top$  from (22) by  $\|\nabla E_D\|_\varepsilon^{-1} \nabla E_D \nabla E_D^\top$ , with  $\|x\|_\varepsilon = \sqrt{x^\top x + \varepsilon}$ . The term from (27) has a similar normalisation effect for arbitrary difference measures, as the standard term  $P = J^\top J$  in the case of SSD. Furthermore, it is sparse for any  $E_D$  (in contrast to  $P = J^\top J$ ), which makes it applicable in a wide range of cases. The applicability of the preconditioning from (27) in multi-modal medical settings is demonstrated in [56], and an exemplary plot of the improved convergence of the preconditioning approach in comparison to the semi-implicit scheme for a multi-modal scenario employing MI is given in Fig. 1(g).



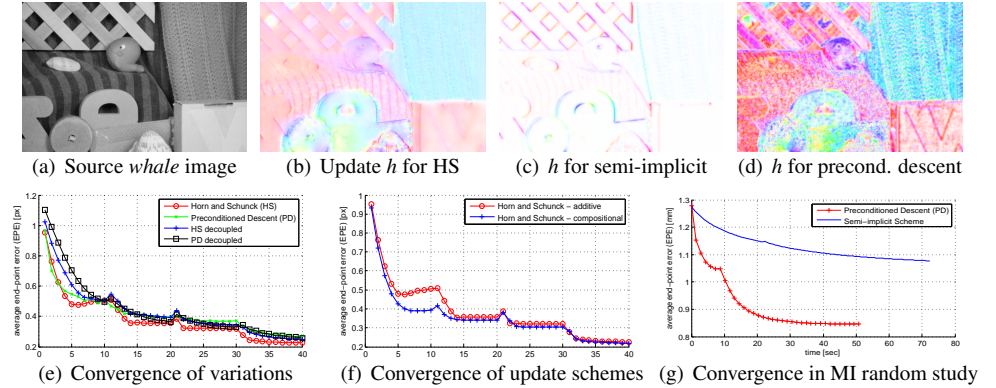


Figure 1: In (a)-(d) we demonstrate on the *whale* data set from [4] with a given ground truth, the effect of the following schemes on the structure of updates  $h$  (1st iteration, original resolution) by visualising the color-coded updates: (b) incremental HS; (c) semi-implicit scheme from (25); (d) the PD variation from (27). Note the dependence of update magnitudes on image gradients in (c). In (e), we demonstrate that the discussed variations perform similar to HS (on image pyramid with 4 levels). In (f), an example for the compositional update scheme is given. In (g), the results of a random study on brain images with ground truth (taken from [56]), show the applicability of PD in multi-modal scenarios using MI.

## 4.2 Decoupling the Linear System

As noted in Sec. 2.4, the structure tensor is responsible for the coupling of the dimensions in the system (22). With the analysis from Sec. 3.2, we can modify the preconditioning term to

$$P(x) = \|\nabla I_S(\Phi(x))\|^2 \text{Id} - \lambda^2 \Delta, \quad (28)$$

such that it has a similar normalising effect as the original HS scheme from 2.4, however, without coupling the dimensions. For the generalised system from (27), we get accordingly

$$P(x) = \|\nabla E_D(\Phi(x))\| \text{Id} - \lambda^2 \Delta. \quad (29)$$

The decoupled version has the advantage that the single dimensions can be treated independently, which facilitates high-level parallelisation. Furthermore, the resulting  $d$  linear systems are of size  $n \times n$ , instead of one  $dn \times dn$  system for the coupled version (with  $d = \# \text{dimensions}$ , and  $n = \# \text{image points}$ ). This reduction of the minimally required memory can present an advantage for treatment of large 3D systems, which result for registration of medical data sets (e.g.  $d=3$ ,  $n=512^3$ ), for systems with limited memory, such as GPUs.

## 4.3 Additive and Compositional Update Schemes

An alternative to the standard addition of updates  $u = U + h$  is the compositional rule  $\text{Id} + u = \Phi \circ (\text{Id} + h)$ , which is motivated by the ratio that the composition is a more appropriate operation for transformations. This can be easily integrated into the GN derivation by replacing  $f_D(I_T, I_S, x + U(x) + h(x))$  by  $f_D(I_T, I_S, \Phi(x + h(x)))$ . With the notation  $I'_S = I_S \circ \Phi$  for the source image warped by the current transformation estimate  $\Phi$ , we get

that  $f_D(I_T, I_S, \Phi(x + h(x))) = f_D(I_T, I'_S, x + h(x))$ , and the corresponding linearisation reads

$$I_D(h(x)) \equiv I_T(x) - I'_S(x) - \nabla (I'_S(x))^\top h(x) = I_T(x) - I_S(\Phi(x)) - \nabla (I_S(\Phi(x)))^\top h(x). \quad (30)$$

The subtle difference to (19) is that with the assumption of the compositional updates we obtain  $\nabla (I_S(\Phi(x)))$ , which is the gradient of the warped source image, while for the additional update in (19), we were dealing with the warped gradient of the source image ( $\nabla I_S)(\Phi(x))$ .

This discussion corresponds to the one in [4] for the Lucas-Kanade algorithm. For dense motion estimation, the compositional update mode was discussed for the so-called demons method in the context of medical image registration in [13, 30, 32], where improved accuracy was observed. Please compare Fig. 1(f) for an example.

## 5 Summary

In a nutshell, the intention of this paper is to promote the interpretation of the incremental Horn and Schunck as Gauss-Newton optimisation of a non-linearised energy. This interpretation is useful and simple, and it is surprising that it is not wider spread in the community. Apart from its simplicity, the advantage of the interpretation is that it places Horn and Schunck in the well-understood framework of least-squares optimisation. This way, not only does the derivation become more elegant (Step 1: state the energy; Step 2: apply GN), but also the interpretation provides a better understanding of the behaviour of HS. For example, it becomes immediately clear that the incremental HS actually minimises the non-linearised energy, the class of treatable difference measures is easily identified, and the relation to other methods becomes clearer. As an example that the theoretical considerations can also result in practical modifications, we present several variations on the theme of Horn and Schunck, which extend the framework to arbitrary difference measures, lead to decoupled systems, or employ different update schemes.

## References

- [1] L. Alvarez, J. Weickert, and J. Sánchez. Reliable estimation of dense optical flow fields with large displacements. *International Journal of Computer Vision (IJCV)*, 2000.
- [2] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision (IJCV)*, 1989.
- [3] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision (IJCV)*, 2004.
- [4] S. Baker, D. Scharstein, J.P. Lewis, S. Roth, M.J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *Proc. International Conference on Computer Vision (ICCV)*, 2007.
- [5] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision (IJCV)*, 1994.
- [6] S.S. Beauchemin and J.L. Barron. The computation of optical flow. *ACM Comput. Surv.*, 1995.

- [7] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. pages 237–252, 1992.
- [8] C. Bergeron and E. Dubois. Gradient-based algorithms for block-oriented map estimation of motion and application to motion-compensated temporal interpolation. *IEEE Transactions on Circuits and Systems for Video Technology*, 1991.
- [9] M.J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding (CVIU)*, 1996.
- [10] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *Proc. European Conference on Computer Vision (ECCV)*, 2004.
- [11] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International Journal of Computer Vision (IJCV)*, 2005.
- [12] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr. A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. *International Journal of Computer Vision*, 2006.
- [13] C. Chéfd’hotel, G. Hermosillo, and O. Faugeras. Flows of diffeomorphisms for multimodal image registration. 2002.
- [14] W. Enkelmann. Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences. *Computer Vision, Graphics, and Image Processing*, 1988.
- [15] F.C. Glazer. *Hierarchical motion detection*. PhD thesis, Amherst, MA, USA, 1987.
- [16] B.K.P. Horn and B.G. Schunck. Determining optical flow. Technical report, Massachusetts Institute of Technology, 1980.
- [17] B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artificial Intelligence*, 1981.
- [18] B.K.P. Horn and B.G. Schunck. Determining optical flow: a retrospective. *Artificial Intelligence*, 1993.
- [19] J. Konrad. Use of colour in gradient-based estimation of dense two-dimensional motion. In *Computer Vision: Systems, Theory and Applications*, 1993.
- [20] K. Madsen, H.B. Nielsen, and O. Tingleff. Methods for non-linear least squares problems. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, 2004.
- [21] E. Mémin and P. Pérez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Transactions on Image Processing (TIP)*, 1998.
- [22] E. Mémin and P. Pérez. A multigrid approach for hierarchical motion estimation. In *Proc. International Conference on Computer Vision (ICCV)*, 1998.
- [23] E. Mémin and P. Pérez. Hierarchical estimation and segmentation of dense motion fields. 2002.

- [24] Jan Modersitzki. *Numerical methods for image registration*. Oxford University Press, 2004. ISBN 0-19-852841-8.
- [25] Enkelmann W. Nagel, H.H. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 1986.
- [26] H.H. Nagel. Constraints for the estimation of displacement vector fields from image sequences. In *IJCAI'83: Proceedings of the Eighth international joint conference on Artificial intelligence*, 1983.
- [27] J. Nocedal and S.J. Wright. *Numerical optimization*. Springer, 2000.
- [28] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision (IJCV)*, 2006.
- [29] A. Roche, G. Malandain, and N. Ayache. Unifying maximum likelihood approaches in medical image registration. *International Journal of Imaging Systems and Technology*, 2000.
- [30] R. Stefanescu, X. Pennec, and N. Ayache. Grid powered nonlinear image registration with locally adaptive regularization. 2004.
- [31] D. Sun, S. Roth, and M.J. Black. Secrets of optical flow estimation and their principles. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [32] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 2009.
- [33] J. Weickert, A. Bruhn, T. Brox, and N. Papenberg. A survey on variational optic flow methods for small displacements. Technical report, Mathematical Image Analysis Group, Saarland University, Germany, 2005.
- [34] M. Werlberger, W. Trobin, T. Pock, H. Bischof, A. Wedel, and D. Cremers. Anisotropic huber-l1 optical flow. In *Proc. British Machine Vision Conference (BMVC)*, 2009.
- [35] Alfred North Whitehead. *Process and Reality. An Essay in Cosmology*. Macmillan Publishing Co., Inc., 1929.
- [36] D. Zikic, M. Baust, A. Kamen, and N. Navab. Generalization of deformable registration in riemannian sobolev spaces. *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2010.