

# OPTICAL FLOW FOR GLACIER MOTION ESTIMATION

Christoph Vogel<sup>1</sup>, Andreas Bauder<sup>2</sup> and Konrad Schindler<sup>1</sup>

Photogrammetry and Remote Sensing, ETH Zürich<sup>1</sup> - (vogel, schindler)@geod.baug.ethz.ch  
Laboratory of Hydraulics, Hydrology and Glaciology (VAW), ETH Zürich<sup>2</sup> - bauder@vaw.baug.ethz.ch

Commission III, Working Group III/5

**KEY WORDS:** Glaciology, Estimation, Orthoimage, Automation, Accuracy, Temporal, Experiment

## ABSTRACT:

Quantitative measurements of glacier flow over time are an important ingredient for glaciological research, for example to determine the mass balances and the evolution of glaciers. Measuring glacier flow in multi-temporal images involves the estimation of a dense set of corresponding points, which in turn define the flow vectors. Furthermore glaciers exhibit rather difficult radiometry, since their surface usually contains homogeneous areas as well as weak texture and contrast. To date glacier flow is usually observed by manually measuring a sparse set of correspondences, which is labor-intensive and often yields rather irregular point distributions, with the associated problems of interpolating over large areas. In the present work we propose to densely compute motion vectors at every pixel, by using recent robust methods for optic flow computation. Determining the optic flow, i.e. the dense deformation field between two images of a dynamic scene, has been a classic, long-standing research problem in computer vision and image processing. Sophisticated methods exist to optimally balance data fidelity with smoothness of the motion field. Depending on the strength of the local image gradients these methods yield a smooth trade-off between matching and interpolation, thereby avoiding the somewhat arbitrary decision which discrete anchor points to measure, while at the same time mitigating the problem of gross matching errors. We evaluate our method by comparing with manually measured point wise ground truth.

## 1 INTRODUCTION

It has become common practice to use photogrammetric tools to acquire and analyze the motion of glaciers. The corresponding points in the images, which define the glacier motion, are still often measured manually. In this work we show how to automate the process by applying optical flow techniques to the problem. Optical flow algorithms have advanced to a state at which they can densely compute motion vectors at every pixel despite the difficult radiometry of glaciers. In this work we compare several methods on a data set with available manual measurements and perform a thorough evaluation using different error metrics. The evaluation shows that in areas where the glacier surface is properly visible modern optical flow methods are competitive with human observers.

### 1.1 Glacier Motion

The motion observed at the surface of a glacier is due to gravitational deformation of the ice and sliding at the base. The resulting ice flow transports mass from areas of snow accumulation to lower areas with mass loss by melt. The magnitude of the internal ice deformation depends on thickness and surface slope, whereas basal sliding depends on the basal conditions, which are mainly influenced by the seasonally varying water pressure (Paterson, 1994). This results in an overall smooth spatial pattern of the surface flow field. Observed temporal variations in glacier motion range from hours to seasons and decades (Iken, 1977). Thus, knowledge of surface flow fields of glaciers contributes to understanding glacier dynamics and is important data for many glaciological applications.

Direct measurements by geodetic methods (GPS, total station) can only be performed on a limited number of accessible points. Remote sensing has the potential to provide dense spatial coverage over large areas. A wide range of methods (optical/radar, feature tracking/interferometric, airborne/spaceborne platforms)

have been applied especially on the polar ice sheets of Greenland and Antarctica (Lucchitta and Ferguson, 1986, Bindshadler and Scambos, 1991, Joughin et al., 1998, e.g.). Investigations of spaceborne systems are restricted to the time of satellite passes and appropriate orbits, which limits application for mountain glaciers with more complex and small scale flow fields.

Ortho-images are a standard product of digital photogrammetric analysis and are suitable for periodically repeated glacier mapping. Areas free of seasonal snow show a variety of features and texture suitable for motion analysis. Early glaciological applications have been implemented for semi-automatic and manual use with analytical stereo plotters (Flotron, 1979). More recently photogrammetric methods became popular to determine fully digital deformation measurements (Kaab, 2002). Analysis of pairs or sequences of ortho-images acquired at different points in time allow one to evaluate the horizontal motion field. The vertical component of the ice motion and the mass balance at the surface can then be separated by subsequent numerical flow modeling (Gudmundsson and Bauder, 1999).

### 1.2 Optical Flow

The problem of image alignment and 2D motion estimation has been studied in computer vision over the past 30 years and has reached an impressive level of reliability and accuracy. Numerous variants are already used in the industry performing tasks such as driver assistance, medical image registration and human motion analysis. In this section we give a short overview of the most common approaches for optical flow estimation, and introduce the basic terminology.

In general there are dense as well as sparse techniques. Sparse optical flow is often preferred in time critical applications and usually performs some kind of feature tracking (Tomasi and Kanade, 1991). In a way sparse techniques have already been applied to glacier flow (Debella-Gilo and Kaab, 2011). Here correspondences are established from correlation coefficients, later mismatches are eliminated by thresholding and manual inspection.

Dense techniques on the other hand compute a motion vector at *every* pixel. It is usually assumed that the observed motion between adjacent frames is small, although there are techniques which explicitly try to compute large displacement optical flow. A classic model for optical flow estimation was introduced by (Horn and Schunck, 1981). Many contemporary techniques are variants of this method, see the popular benchmark (Baker et al., 2007), thus we use it as basis for this chapter.

Instead of treating each motion independently, a regularization-based framework is established, which enforces data fidelity and spatial smoothness at the same time. In the original paper the data term is expressed on a per pixel basis, using the well known *brightness constancy assumption* (BCA). The BCA states that the gray value of a moving pixel remains constant throughout the explored frames (Eq. 1):

$$I(x + u, y + v, t + 1) = I(x, y, t), \quad (1)$$

where  $I$  denotes the image brightness at a given pixel and time-step, and  $(u, v)^t$  denotes the optical flow vector. Taylor series expansion at  $(u_0, v_0)$  yields the *optical flow constraint*

$$I_x(u - u_0) + I_y(v - v_0) + I_t = 0. \quad (2)$$

Here  $(I_x, I_y)^t = \nabla I$  and  $I_t = I(x + u_0, y + v_0, t + 1) - I(x, y, t)$  is the *temporal derivative*. Using a variational formulation the data term is finalized as:

$$E_D = \int_{\Omega} \rho_D(I_x(u - u_0) + I_y(v - v_0) + I_t) dx. \quad (3)$$

Integration is performed over the image domain  $\Omega$ . In the original formulation the penalty function  $\rho_D$  is the quadratic penalty  $\rho_D(x) = |x|^2$ . From Eq. 2 we can directly observe that the problem is ill-posed, since only a single constraint per pixel can be extracted. In the literature this is known as the *aperture problem*. The problem is not limited to this specific formulation but also exists when image regions are to be matched, for instance a patch containing a single edge or a textureless area. Furthermore noise can degrade the quality of a matching process. Therefore a smoothness constraint on the motion field is introduced,

$$E_S = \int_{\Omega} \rho_S(\nabla u) + \rho_S(\nabla v) dx, \quad (4)$$

resulting in an overall energy

$$E = E_D + \lambda E_S. \quad (5)$$

The parameter  $\lambda$  controls the amount of smoothing. The Euler-Lagrange equations of the energy term, derived with variational calculus, yield necessary conditions for a minimum of the energy functional. Since the functional is given in terms of continuous variables, variational methods naturally lead to sub-pixel accuracy. The resulting energy minimization problem is highly non-convex, because the *optical flow constraint*, Eq. 2, is only valid in a small local neighborhood. Therefore minimization is performed in a hierarchical coarse-to-fine approach. Using image pyramids, solutions from a lower pyramid level initialize the next-finer level.

The original formulation of (Horn and Schunck, 1981) suffers from the fact that quadratic error functions  $\rho_D$  and  $\rho_S$  lead to over-smoothing at motion boundaries. Different error functions have been proposed. The most popular one nowadays is  $\rho(x) = \sqrt{x^2 + \epsilon^2}$ , a differentiable variant of the  $L_1$ -norm, which is both robust and convex. Larger displacements can be estimated thanks to image warping and postponing any linearisation to the numer-

ical scheme (Brox et al., 2004). Certain versions of variational optic flow can be computed in real-time (Wedel et al., 2008).

In contrast to variational methods, which minimize the energy functional by continuous optimization, combinatorial methods which are dominant in stereo matching are less popular in the literature. Mainly because of the enlarged two-dimensional search space. Notable exceptions are (Lei and Yang, 2009, Lempitsky et al., 2008). Finally, local methods exist, which do not minimize a global energy function at all *e.g.* (Rhemann et al., 2011).

## 2 ALGORITHMS

In this section we describe in more detail the methods selected for our evaluation. We choose three different algorithms as representatives of the large number of optical flow methods developed over the years. All three algorithms are parallelizable and efficient in terms of memory and speed, which is important due to the sheer size of aerial images used for glaciology.

### 2.1 Total variation optical flow

The first algorithm in our set is a representative of the popular variational approaches. The energy functional is based on the classical formulation already presented in Eqs. 3, 4 and 5. However instead of approximating the  $L_1$ -norm in order to make the functional differentiable, the integral equations are addressed directly, using a primal-dual scheme. In (Zach et al., 2007) it was proposed to add additional auxiliary variables  $\mathbf{v}$  to the energy function, to simplify the optimization:

$$\int_{\Omega} \lambda |\nabla \mathbf{u}| + \frac{1}{2\theta} |\mathbf{u} - \mathbf{v}|^2 + |\rho(\mathbf{v})| dx, \quad (6)$$

where the data term is given by the optical flow constraint Eq. 2,  $\rho(\mathbf{v}) := \langle \nabla I, \mathbf{v} - \mathbf{v}_0 \rangle + I(\mathbf{x} + \mathbf{v}_0, t + 1) - I(\mathbf{x}, t)$  and  $\theta$  is a small constant restricting  $\mathbf{v}$  to be a close approximation of  $\mathbf{u}$ . The convex energy can be minimized by alternating steps updating either  $\mathbf{u}$  or  $\mathbf{v}$ . Fixing  $\mathbf{v}$ , optimization with respect to  $\mathbf{u}$  is achieved using the primal-dual algorithm proposed in (Chambolle, 2004). For a fixed  $\mathbf{u}$  the optimization yields a sum of decoupled pixel-wise energies, which can be minimized individually. More recently an improved primal-dual scheme was introduced by (Chambolle and Pock, 2011), superseding the need for auxiliary variables. Moreover a slightly improved optical flow constraint is used which additionally models varying illumination:

$$\rho(\mathbf{u}, w) = \langle \nabla I, \mathbf{u} - \mathbf{u}_0 \rangle + I_t + \beta w. \quad (7)$$

The parameter  $\beta$  controls the influence of the illumination term. The function  $w$  is assumed to be smooth and therefore requires additional regularization. The final energy was already proposed in (Shulman and Hervé, 1989),

$$\int_{\Omega} \lambda |\nabla \mathbf{u}| + |\nabla w| + \rho(\mathbf{u}, w) dx. \quad (8)$$

Details of the optimization scheme go beyond the scope of this paper, please refer to (Chambolle and Pock, 2011).

### 2.2 Cost-Volume Filtering

The second algorithm we have selected considers optical flow estimation as a labeling problem. The space of possible solutions for the flow vectors is given as a discrete set, and each pixel is assigned to one element of the set. Because the solution space is no longer continuous, the task reduces to a search problem. A naive solution would be to select the label with the lowest data error for

each pixel. Filter based methods, *e.g.* (Yoon et al., 2006), instead apply a local filter on the energy values for different flow vectors (the *cost volume*) before assigning the labels. These techniques can also be seen as an approximation to discrete energy based approaches. In contrast to the spatially global smoothing of a conditional random field, smoothness of the flow field is only defined locally. The intuition why similar solutions are achieved is that pixels far apart usually have only little influence on each other. The method we choose for evaluation was proposed recently by (Rhemann et al., 2011). The algorithm mainly differs from previous filter-based techniques by the choice of the filter weights. The *guided filter* introduced by (He et al., 2010) preserves edges in the input image and has a runtime independent of the filter size. Incorporated into the filter framework the technique achieves high-quality solutions, and is competitive with energy-based methods on standard benchmarks. The method can potentially handle both small scale motion structure and large displacements.

More formally, let the set of labels be denoted by  $\mathcal{L} = \{1, \dots, L\}$  and  $l \in \mathcal{L}$  be a label, in our case a displacement  $(u, v)^t$ . A cost  $C(i, l)$  is assigned to each pixel  $i$  and label  $l$  by evaluating the data term. For a label  $l$  the filtered cost at pixel  $i$ ,  $\bar{C}(i, l)$ , is the weighted sum over the neighborhood  $N(i)$  of  $i$ :

$$\bar{C}(i, l) = \sum_{j \in N(i)} w_{i,j}(I) C(j, l). \quad (9)$$

The final labeling is now given by taking the label  $l_i$  with minimal filtered cost at each pixel  $i$ ,  $l_i = \operatorname{argmin}[\bar{C}(i, l)]$ . The key to high quality results is to use an edge-preserving filter. The weights of the *guided filter* depend on a guidance image  $I$ , in our case the reference image:

$$w_{i,j}(I) = \frac{1}{|N(k)|} \sum_{k|j \in N(k)} 1 + \frac{(I(i) - \mu_k)(I(j) - \mu_k)}{\sigma_k^2 + \epsilon}. \quad (10)$$

Here  $\mu_k$  and  $\sigma_k^2$  are the mean and variance of the region  $N(k)$  centered at pixel  $k$  in  $I$ . The sum is composed of all image regions containing both pixel  $i$  and  $j$ . The edge preservation can be seen by considering a region with a single edge. If pixel  $i$  and  $j$  are on the same side of the edge, the weight  $w_{i,j}(I)$  becomes 2, and close to 0 otherwise. For flat regions with  $\sigma_k^2 < \epsilon$  we have  $w_{i,j}(I) = 1$  for all pixel in the region, which results in a simple averaging filter. An extension to color images is straightforward. For more details please refer to (He et al., 2010).

As data term we select two different popular cost functions. Firstly the truncated absolute difference of the gray values and gradient at the matching points in the reference image  $I(\cdot, t)$  and the displaced image  $I(\cdot, t+1)$ , and secondly the negative *normalized cross correlation* (*NCC*). The truncated absolute difference, Eq. 11, has been shown to be robust against illumination changes:

$$C_{\text{SAD}}(i, l) = \alpha \min[|I(i, t) - I(i+l, t+1)|, \tau_1] + (1 - \alpha) \min[|\nabla I(i, t) - \nabla I(i+l, t+1)|, \tau_2]. \quad (11)$$

The truncation values are denoted by  $\tau_1$  and  $\tau_2$  in the equation. In contrast to this pixel-wise data cost the *NCC* is defined as a sum over a region  $N(i)$  centered at pixel  $i$ :

$$C_{\text{NCC}}(i, l) = \sum_{i \in N(i)} \frac{(I(i, t) - \mu_{i,t})(I(i+l, t+1) - \mu_{i+l,t+1})}{\sigma_{i,t}^2 \sigma_{i+l,t+1}^2}, \quad (12)$$

where  $\mu_{i,t}$  and  $\mu_{i+l,t+1}$  are the means of images  $I(\cdot, t)$  and  $I(\cdot, t+1)$  over the region  $N(i)$  and  $\sigma_{i,t}^2$  and  $\sigma_{i+l,t+1}^2$  are the respective variances.

The major challenge in the approach is the huge label space to

be evaluated. In the examples the flow vectors extend over a region of  $\{-75, \dots, 75\}^2$  pixel, which leads to 22500 labels. That problem is tackled by very fast implementations of the filtering: the weighted filtering can be achieved by a sequence of box-filters with a time complexity linear in the number of pixels. Sub-pixel precision is accomplished by upscaling the second image by an integral factor and evaluating the cost function accordingly with respect to the new, increased label space.

### 2.3 Pyramid Lucas-Kanade Optical Flow

Another widely used approach for optical flow estimation was first introduced by (Lucas and Kanade, 1981). Originally developed for image registration the method is still commonly used for feature tracking and template matching. An image patch around a pixel  $i$  undergoes a deformation governed by a parameter vector  $\mathbf{p}$ , in order to minimize the squared difference to a template  $T$ :

$$\sum_{j \in N(i)} (I(f(j, \mathbf{p})) - T(j))^2. \quad (13)$$

Here  $f$  denotes a warping function parameterized by  $\mathbf{p}$ . The key idea, compare also Eq. 2, is to perform gradient descent on the sum of squared distance (*SSD*) energy function, using a Taylor series expansion of the image function at  $\mathbf{p}_0$ :

$$\sum_{j \in N(i)} (I(f(j, \mathbf{p}_0)) + \nabla I \frac{\partial f}{\partial \mathbf{p}} (\mathbf{p} - \mathbf{p}_0) - T(j))^2 \quad (14)$$

Iteratively solving the corresponding normal equations with respect to  $\mathbf{p}$  delivers a solution for the parameter vector. In the 2D case it is common to restrict the parameter set to allow for pure translations or affine deformations only (Baker and Matthews, 2004). In order to achieve dense flow the scheme is applied at each pixel individually and embedded into a hierarchical coarse-to-fine framework.

The scheme estimates the motion at each pixel independently, therefore the implementation can be easily parallelized. Care has to be taken if the normal matrix is close to singular, for instance in textureless regions.

## 3 RESULTS AND DISCUSSION

We evaluate the different optical flow methods on ortho-photos of the Unteraargletscher located in the Bernese Alps, Switzerland. This large valley glacier has two main tributary forming the common tongue with an area of about 23 km<sup>2</sup> and a length of 13 km. The grayscale ortho-images are divided into three parts of 26, 20.5 and 12 mega-pixel, consisting of the individual tributaries and the tongue. One pixel corresponds to one meter ground resolution. Image pairs were acquired in a temporal distance of one whole year, at 1970/71, 1982/83 and 1997/98 (Fig. 1). Details about the acquisition and image processing chain can be found in (Bauder, 2001). The maximal motion observed over the one year period is about 40m at the tongue area and 90m in the area of the tributaries.

We study three algorithms: total variation (*TV-LI*), cost volume filtering (*CF*) and pyramid Lucas-Kanade (*Pyr-LK*). Quantitative evaluation is done by comparing the results with manually measured, sparse correspondences (Bauder, 2001). The dataset consists of 5606 measurements on a 50 pixel grid and is only available for the years 1997/98. The accuracy of the manual measurements is evaluated by measuring several feature points twice and independently, see table 1. For all images the differences are in the range of 1 to 3 pixel (=meters).

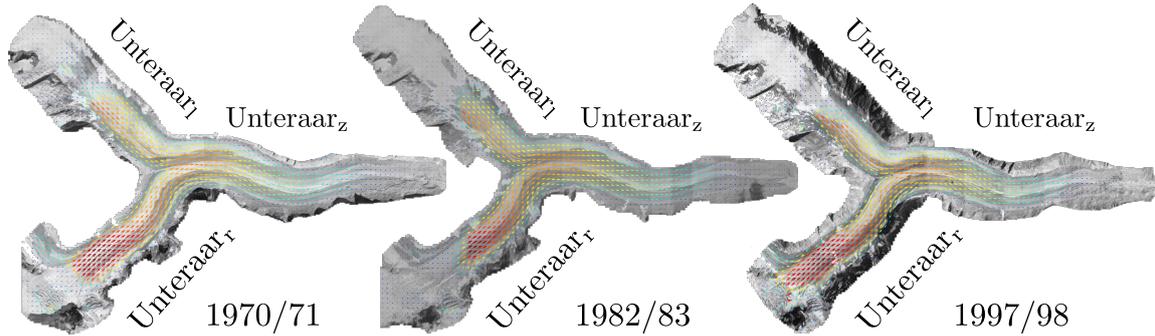


Figure 1: Reconstruction of all image pairs. Flow vectors are color coded w.r.t. their magnitude (*blue*=slow flow, *red*=fast flow).

Because of the large temporal gap between the acquisitions, changes in image appearance are inevitable, which make optical flow estimation challenging. Difficulties include varying snow coverage, non-uniform melt out or crevasse patterns, moving and tumbling rocks, shadows and generally different illumination conditions. The biggest problem is posed by the varying snow cover extent. The data term measuring visual similarity cannot identify corresponding parts correctly as their appearance is too different. The second major problem are large textureless regions *e.g.* due to snow coverage or shadows. Those parts carry only little, if any, information about the motion. Usually optical flow methods deal with textureless segments by propagating information from neighboring areas into those regions. In our case however, the varying borders of the snow cover and shadow borders are the main source of information here. Since these are completely independent from the motion of the glacier, the quality of the reconstructions is low in these regions (compare Fig. 3).

### 3.1 Error Metrics

Popular error metrics in the optical flow literature (Baker et al., 2007) are the *average end point error* AEP and the *average angular error* AAE. The former measures the distance between two flow vectors in 2D. The latter compares the angle between two flow vectors  $(u, v, 1)$  in homogeneous 3D space. The errors are averaged over all test points. We also report the *normalized root mean squared error* NRMS of the end point error, where the normalization is performed w.r.t. the difference between the maximal and minimal flow magnitude in the image. Further we also report quantiles of the AEP, since the metrics are typically dominated by few gross outliers.

We compare the results of the optical flow algorithms with the sparse manual measurements. The results are shown in table 2. The *TV-LI* algorithm performs best in almost all cases, except for the image *Unteraar<sub>r</sub>*. Here the algorithm cannot handle the varying snow cover at the top of the glacier. A close-up of that part is shown in the first row of Fig. 3, and also on the right of Fig. 4. Cost filtering (*CF*) with *NCC* performs better in that area, although only for parts where the data function delivers reasonable costs. Using the *SAD* data cost already produces unacceptable results, although the cost function works well for smaller images with relaxed environmental conditions. In all cases *Pyramid-LK* performs worst and is not further considered in the evaluation. Without any regularization, the algorithm is not competitive in these complex conditions.

Table 1: Deviations of the control sample set of the manual measurements (*unit: pixel=m*)

|                | Unteraar <sub>z</sub> | Unteraar <sub>r</sub> | Unteraar <sub>l</sub> |
|----------------|-----------------------|-----------------------|-----------------------|
| mean deviation | 0.89                  | 2.96                  | 2.36                  |
| std. deviation | 1.36                  | 3.02                  | 2.23                  |
| sample-size    | 168                   | 166                   | 16                    |

Table 2: Evaluation using different error metrics. (*unit: pixel=m*)

|                   | Unteraar <sub>z</sub> |              |             |             |              |              |
|-------------------|-----------------------|--------------|-------------|-------------|--------------|--------------|
|                   | AEP                   | Q95%         | Q80%        | Q50%        | AAE          | NRMS         |
| <i>TV-LI</i>      | <b>1.34</b>           | <b>4.61</b>  | <b>1.74</b> | <b>0.75</b> | <b>10.3°</b> | <b>5.5%</b>  |
| <i>CF&amp;NCC</i> | 1.69                  | 6.11         | 1.99        | 0.89        | 11.7°        | 8.2%         |
| <i>CF&amp;SAD</i> | 3.80                  | 17.29        | 3.08        | 1.25        | 18.4°        | 24.2%        |
| <i>Pyr.-LK</i>    | 18.44                 | 53.45        | 31.96       | 12.80       | 58.8°        | 62.1%        |
|                   | Unteraar <sub>r</sub> |              |             |             |              |              |
|                   | AEP                   | Q95%         | Q80%        | Q50%        | AAE          | NRMS         |
| <i>TV-LI</i>      | 8.90                  | 44.55        | 14.29       | 2.42        | 14.3°        | 24.9%        |
| <i>CF&amp;NCC</i> | <b>6.51</b>           | <b>33.04</b> | <b>6.55</b> | <b>2.17</b> | <b>14.2°</b> | <b>22.4%</b> |
| <i>CF&amp;SAD</i> | 25.12                 | 87.20        | 63.64       | 4.80        | 45.5°        | 62.6%        |
| <i>Pyr.-LK</i>    | 35.86                 | 76.16        | 53.12       | 33.40       | 76.2°        | 64.1%        |
|                   | Unteraar <sub>l</sub> |              |             |             |              |              |
|                   | AEP                   | Q95%         | Q80%        | Q50%        | AAE          | NRMS         |
| <i>TV-LI</i>      | <b>3.90</b>           | <b>15.65</b> | <b>4.60</b> | <b>1.58</b> | <b>7.5°</b>  | <b>9.3%</b>  |
| <i>CF&amp;NCC</i> | 7.40                  | 45.12        | 5.39        | 1.99        | 14.3°        | 21.2%        |
| <i>CF&amp;SAD</i> | 19.67                 | 98.73        | 34.21       | 3.01        | 32.2°        | 43.3%        |
| <i>Pyr.-LK</i>    | 36.90                 | 103.42       | 58.51       | 28.48       | 66.3°        | 55.9%        |

In Fig. 4 (*left*) we visualize the difference of the solution of the *TV-LI* algorithm and the expert's solution. Areas where the estimates diverge are marked in red, and where they agree in green. The regions of greatest disagreement are found in spots showing limited visual correspondence in the original images. In Fig. 3 we visualize several of those areas. Both algorithms are misled by the ill-posed data term and can not estimate the flow correctly. In the case of the *CF* algorithm these regions have limited impact on neighboring areas, however the estimates within the regions are completely wrong. This can be explained by the edge preserving property of the filter. Vectors from regions separated by an image edge have little influence on each other. On the other hand, the *TV-LI* algorithm generally seems to handle those parts better, and in most cases still produces reasonable motion estimates. Potentially the influence of erroneous areas could be restricted by the use of an anisotropic smoothing kernel.

### 3.2 Illumination Changes

Due to different environmental conditions the assumption of brightness constancy between successive image acquisitions is violated. To alleviate the negative influence of illumination changes, sensor noise, or shadows on the data term, three different methods were tested in case of the *TV-LI* algorithm. *Structure-texture decomposition (STT)* (Aujol et al., 2006), has been shown to be a successful pre-processing step for 2D optical flow estimation (Wedel et al., 2008). Images are decomposed into a structural part, corresponding to larger connected regions of the image, and a texture part containing fine scale details. Further it is assumed that using the textural part for the computation of the optical flow is more

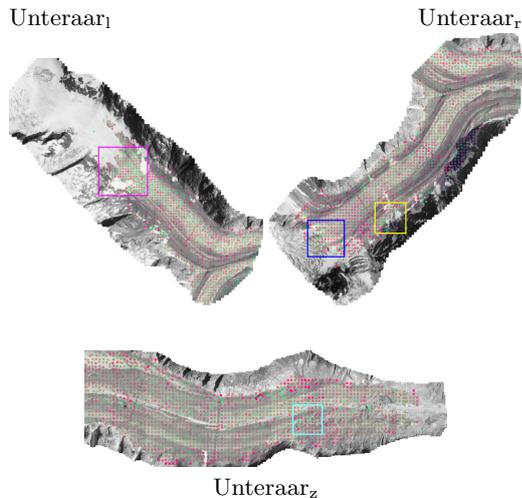


Figure 2: Image similarity at corresponding points according to *TV-LI* and to manual flow estimates. NCC scores are overlain onto the images using color-code. Areas in *red* mark higher similarity with *TV-LI*, *green* areas higher scores of the manual measurements.

robust against shadow and shading artefacts. A study comparing various filters is given in (Vaudrey et al., 2009).

The second method evaluated is an improved optical flow constraint in which the varying illumination is modeled by an additive term (Eq. 7). Additionally the data term is extended to also include gradient images. In table 3 we summarize the experiment. It is obvious that all methods significantly improve our results. In all other experiments the setup “*STT&additive function*” is used, which worked best overall.

### 3.3 Qualitative Evaluation

In a further experiment we attempt to evaluate the quality of the estimated flow fields by investigating the similarity of corresponding image regions. We compute the NCC coefficients in a  $11 \times 11$  and a  $5 \times 5$  window located at matching positions. Note that NCC is not used in the *TV-LI* algorithm, so in that case it constitutes an independent check, whereas the *CF* method uses NCC, which constitutes a bias of the evaluation. We compute correlation scores at all pixel containing manual estimates. Table 4 shows that on average the score produced by the algorithms is noticeably higher than the score of the expert. Results for the *TV-LI* algorithm are consistently between 11% and 28% better than the human expert. The *CF* algorithm achieves at most 16% improvement, and is 3% worse than the human in one case. We conclude that in terms of visual coherence the automatically produced results are at least as good as the annotated ground truth.

In Fig. 4 we visually compare the matching quality of the *TV-LI* algorithm to the human expert. For this purpose we warp patches

Table 3: AEP of the *TV-LI* algorithm using different illumination models. (*unit: pixel=m*)

|                                  | Unteraar <sub>z</sub> | Unteraar <sub>r</sub> | Unteraar <sub>1</sub> |
|----------------------------------|-----------------------|-----------------------|-----------------------|
| none                             | 5.84                  | 13.84                 | 30.56                 |
| <i>STT</i>                       | 1.35                  | 4.58                  | <b>8.70</b>           |
| additive function                | <b>1.32</b>           | 4.61                  | 10.32                 |
| <i>STT&amp;additive function</i> | 1.34                  | <b>3.90</b>           | 8.90                  |
| <i>STT&amp;gradients</i>         | 1.33                  | 4.40                  | 8.80                  |

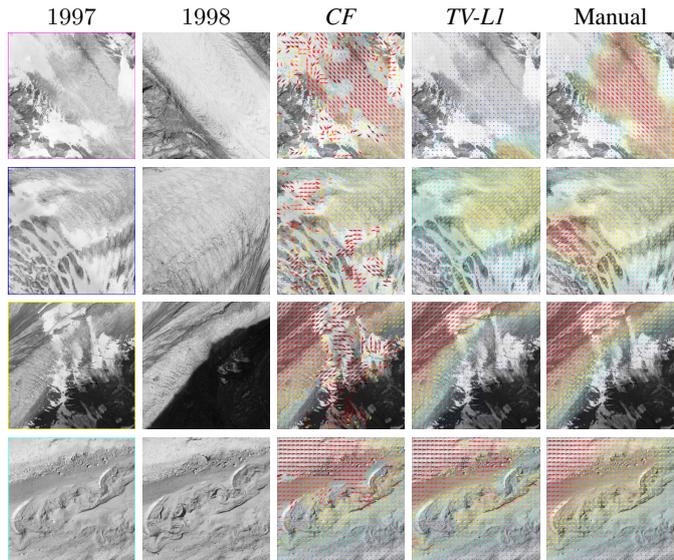


Figure 3: Depicted are patches in which the computed flow-fields differ grossly from the manual estimates. Errors occur if the true motion is not visible on the surface (*col. 1,2*). Here the objective to match corresponding image content can distort the flow fields (*col. 3,4*). Flow vectors are color coded with respect to their magnitude (*blue=slow flow, red=fast flow*).

Table 4: Mean NCC scores of different algorithms

|                   | Unteraar <sub>z</sub> |             | Unteraar <sub>r</sub> |             | Unteraar <sub>1</sub> |             |
|-------------------|-----------------------|-------------|-----------------------|-------------|-----------------------|-------------|
|                   | 11×11                 | 5×5         | 11×11                 | 5×5         | 11×11                 | 5×5         |
| <i>TV-LI</i>      | <b>0.49</b>           | <b>0.49</b> | <b>0.45</b>           | <b>0.46</b> | <b>0.36</b>           | <b>0.36</b> |
| <i>CF&amp;NCC</i> | 0.45                  | 0.40        | 0.37                  | 0.43        | 0.32                  | <b>0.36</b> |
| human expert      | 0.45                  | 0.42        | 0.35                  | 0.39        | 0.30                  | 0.31        |

of size  $200 \times 200$  using the flow vectors. Ideally the warps should be identical to their counterpart in the reference image. For simple comparison we show the difference image between reference and warped patch. In all cases, warps created by the algorithm appear more similar to the reference. The global deviation is smaller and the error patches appear smoother.

Fig. 2 compares NCC-scores of the *TV-LI* flow and the human expert. In regions colored red the score of the algorithm was higher, green regions mark areas with a higher NCC-score for the expert. Clusters of green spots are located in areas affected by changes in environmental conditions, where expert knowledge is needed to overrule the observations. Otherwise red (*TV-LI* better) clearly dominates.

## 4 CONCLUSIONS AND FUTURE WORK

We have evaluated three different optical flow techniques on the task of estimating the motion of glaciers. Especially results obtained from the *TV-LI* based method look promising. Based on an independent image matching metric, and also by visual inspection, the estimates appear to be more accurate than manually measured correspondences. Regions with little visual coherence however can lead to distorted results. A solution could be to simply detect and exclude these areas from the estimation process. Going further, the regularization term could be adapted to the task by adopting a numerical model from glaciology. Finally an extension of the algorithm to the 3D domain could be interesting.

## REFERENCES

Aujol, J.-F., Gilboa, G., Chan, T. and Osher, S., 2006. Structure-texture image decomposition – Modeling, algorithms, and param-

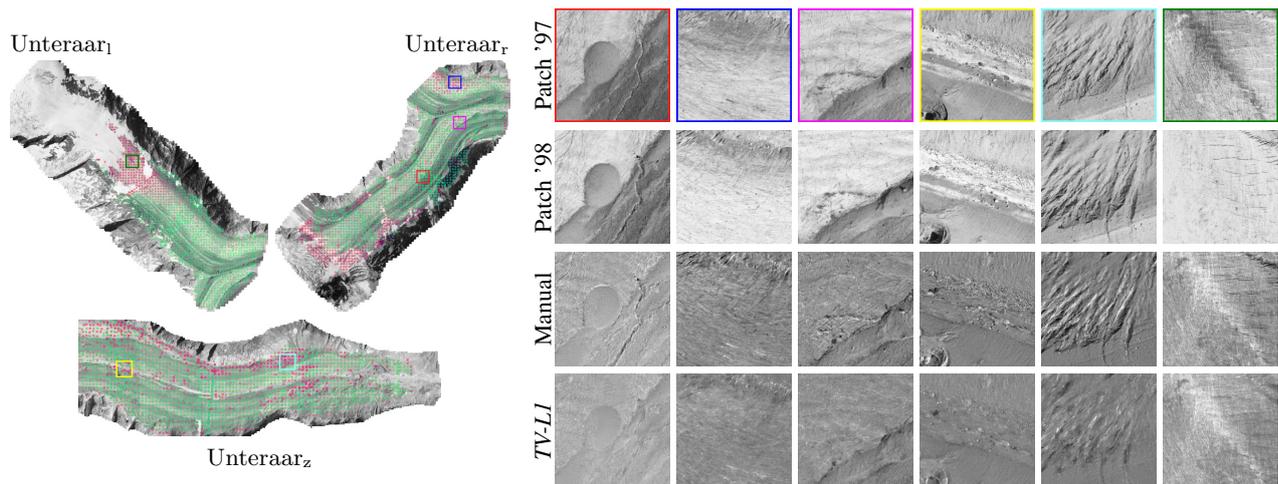


Figure 4: Qualitative comparison of the patch-similarity warped by using the flow estimates of *TV-L1* and according to the manual measurements. *Right*: Magnified windows show several selected areas, together with difference images between target and warped source image. Warps based on the *TV-L1* flow have a greater similarity. *Left*: Each Dot denotes a manual measurement. Areas with high similarity between *TV-L1* estimates and manual measurements are drawn in *green*, deviating regions in *red*.

eter selection. *IJCV* 67(1), pp. 111–136.

Baker, S. and Matthews, I., 2004. Lucas-kanade 20 years on: A unifying framework. *IJCV* 56(1), pp. 221 – 255.

Baker, S., Roth, S., Scharstein, D., Black, M. J., Lewis, J. and Szeliski, R., 2007. A database and evaluation methodology for optical flow. *ICCV*.

Bauder, A., 2001. Bestimmung der Massenbilanz von Gletschern mit Fernerkundungsmethoden und Fließmodellierungen. Dissertation No 14110, ETH Zürich.

Bindschadler, R. A. and Scambos, T. A., 1991. Satellite-image-derived velocity field of an Antarctic Ice Stream. *Science* 252(5003), pp. 242–246.

Brox, T., Bruhn, A., Papenberg, N. and Weickert, J., 2004. High accuracy optical flow estimation based on a theory for warping. In: *ECCV*.

Chambolle, A., 2004. An algorithm for total variation minimization and applications. *JMIV* 20(1-2), pp. 89–97.

Chambolle, A. and Pock, T., 2011. A first-order primal-dual algorithm for convex problems with applications to imaging. *JMIV* 40(1), pp. 120–145.

Debella-Gilo, M. and Kaab, A., 2011. Sub-pixel precision image matching for measuring surface displacements on mass movements using normalized cross-correlation. *Remote Sens. Environ.* 115(1), pp. 130–142.

Flotron, A., 1979. Verschiebungsmessungen aus Luftbildern. In: *Festschrift Peter Kasser*, Vol. 41, Mitteilungen der Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie, ETH Zürich.

Gudmundsson, G. H. and Bauder, A., 1999. Towards an indirect determination of the mass-balance distribution of glaciers using the kinematic boundary condition. *Geografiska Annaler* 81A(4).

He, K., Sun, J. and Tang, X., 2010. Guided image filtering. In: *ECCV*.

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

Iken, A., 1977. Variations of surface velocities of some Alpine glaciers measured at intervals of a few hours. Comparison with Arctic Glaciers. *ZGG* 13(1/2), pp. 23–35.

Joughin, I. R., Kwok, R. and Fahnestock, M. A., 1998. Interferometric estimation of three-dimensional ice-flow using ascending and descending passes. *IEEE TGARS* 36(1), pp. 25–37.

Kaab, A., 2002. Monitoring high-mountain terrain deformation from repeated air- and spaceborne optical data: examples using digital aerial imagery and aster data. *ISPRS Journal of Photogrammetry and Remote Sensing* 57(1-2), pp. 39–52.

Lei, C. and Yang, Y.-H., 2009. Optical flow estimation on coarse-to-fine region-trees using discrete optimization. In: *ICCV*.

Lempitsky, V. S., Roth, S. and Rother, C., 2008. Fusionflow: Discrete-continuous optimization for optical flow estimation. In: *CVPR*.

Lucas, B. D. and Kanade, T., 1981. An iterative image registration technique with an application to stereo vision. In: *IJCAI*.

Lucchitta, B. K. and Ferguson, H. M., 1986. Antarctica: measuring glacier velocity from satellite images. *Science* 234(4780), pp. 1105–1108.

Paterson, W. S. B., 1994. *The Physics of Glaciers*. third edn, Pergamon, New York. pp. 480.

Rhemann, C., Hosni, A., Bleyer, M., Rother, C. and Gelautz, M., 2011. Fast cost-volume filtering for visual correspondence and beyond. In: *CVPR*, pp. 3017–3024.

Shulman, D. and Hervé, J.-Y., 1989. Regularization of discontinuous flow fields. *Proceedings Workshop on Visual Motion*.

Tomasi, C. and Kanade, T., 1991. Detection and tracking of point features. Technical report, *IJCV*.

Vaudrey, T., Wedel, A., Chen, C.-Y. and Klette, R., 2009. Improving optical flow using residual and sobel edge images. In: *ArtsIT*, pp. 215–222.

Wedel, A., Pock, T., Zach, C., Cremers, D. and Bischof, H., 2008. An improved algorithm for TV-L1 optical flow. In: *Proc. of the Dagstuhl Motion Workshop*, Springer.

Yoon, K.-J., Member, S. and Kweon, I. S., 2006. Adaptive support-weight approach for correspondence search. *IEEE Trans. PAMI* 28, pp. 650–656.

Zach, C., Pock, T. and Bischof, H., 2007. A duality based approach for realtime tv-l1 optical flow. In: *DAGM*, pp. 214–223.