

LOCAL SMOOTHNESS FOR GLOBAL OPTICAL FLOW

Lars Lau Rak t

Department of Computer Science, University of Copenhagen
Universitetsparken 5, 2100 Copenhagen, Denmark

ABSTRACT

We consider the problem of estimating the “smoothness parameter” that controls the tradeoff between data fidelity and regularity in optical flow estimation. We start by reviewing the problem of global estimation using the Optimal Prediction Principle (OPP) by Zimmer *et al.* [1]. Inspired by this technique and work on local-global optical flow we propose a simple method for fusing optical flow estimates of different smoothness by evaluating interpolation quality locally by means of L^1 block match on the corresponding set of gradient images. We illustrate the method in a setting where optical flows are estimated by a TV- L^1 energy. On average this procedure reduces the average endpoint error by 15% over flows estimated using the OPP, and gives flow fields that are consistently better than the single best flows with a fixed smoothness parameter.

Index Terms— optical flow, motion estimation, parameter estimation, TV- L^1

1. INTRODUCTION

The variational framework with competing data fit and regularity measures is widely adopted for optical flow estimation today [2, 3, 4, 1]. These types of methods typically have a parameter λ that determine the tradeoff between data fidelity and regularity of the estimated optical flow. We will refer to this parameter as the smoothness parameter. We consider the problem of choosing a good value of λ which is of crucial importance to the quality of the estimated flow.

We will first review the optimal prediction principle suggested by Zimmer *et al.* [1] as a method for automatically choosing a global smoothness parameter. We will consider variations of the principle, and then move on to the question of locally estimating smoothness. We will view this as a subproblem of the more general question of automatically determining which of two flows is best at a given spatial location. Several works have considered this problem from different angles. One notable example is the work of Lempitsky *et al.* [5], where simple flow candidates were fused to produce superior flows by minimizing a highly non-convex energy. Another interesting approach is presented by Mac Aodha *et al.* [6], where a random forest classifier is used to segment video

data such that the pixels belonging to each individual class has their flow computed by a specific optical flow algorithm. The method we present here is simple, elegant and easy to implement. The general setup is inspired by the work of Bruhn *et al.* [7], that first proposed to combine local and global approaches, but here the elements are combined in an alternative manner: A number of candidate flows are estimated with different values of the smoothness parameter. This is done by a global approach (TV- L^1). The computed candidates are then fused to a single superior flow using a local variant of the OPP, where we locally evaluate data fidelity (L^1 block match) on the corresponding pair of gradient images. While this setup is very simple, we show that we can substantially increase accuracy of the estimated flows.

The estimation principle is generic and can be used with any choice of optical flow algorithm for the candidate flows, as well as any choice of local evaluation method. Here we will restrict ourselves to grayscale images and use a variant of the TV- L^1 optical flow algorithm of Zach *et al.* [4] to keep the exposition clear and simple.

2. TV- L^1 OPTICAL FLOW

Optical flow refers to the apparent motion pattern between two images I_0 and I_1 . We calculate the displacement field v such that the difference $I_1(\mathbf{x} + v(\mathbf{x})) - I_0(\mathbf{x})$ is minimized in some sense, under the condition that v maintains sufficiently regularity. Here we will concentrate on the variational TV- L^1 formulation of the optical flow problem, where the optical flow is recovered as a minimizer of the energy E :

$$E(v) = \int \lambda |I_1(\mathbf{x} + v(\mathbf{x})) - I_0(\mathbf{x})| + \|\mathcal{D}v(\mathbf{x})\| d\mathbf{x}. \quad (1)$$

The first term in this energy is the L^1 part, i.e. an integral of the Euclidian norm of the difference $I_1(\mathbf{x} + v(\mathbf{x})) - I_0(\mathbf{x})$. This term builds on an assumption that image values are conserved along the motion over time. The second term is the vectorial total variation of the flow. This is not uniquely defined, and in this setting we have used the definition that \mathcal{D} is the 1-Jacobian of Goldluecke *et al.* [8], which means that the energy favors smoother displacement fields, but at the same time allows for discontinuities.

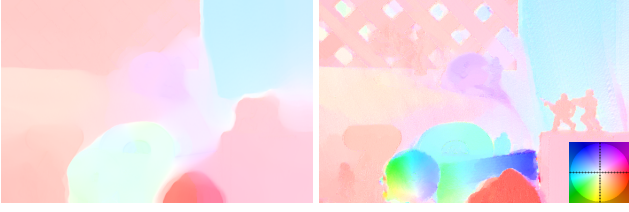


Fig. 1: Color coded TV- L^1 optical flows calculated for the *Army* sequence [11]. λ was set to respectively 2 and 30.

We will do the minimization of this energy using a method similar to the one proposed by Zach et al. [4]. The first step in the minimization is to linearize the optical flow constraint $I_1(\mathbf{x} + v(\mathbf{x})) - I_0(\mathbf{x})$ around the point $\mathbf{x} + v_0^x$ for each \mathbf{x} :

$$\underbrace{(v(\mathbf{x}) - v_0^x)^T \nabla I_1(\mathbf{x} + v_0^x) + I_1(\mathbf{x} + v_0^x) - I_0(\mathbf{x}))}_{=\rho(v)(\mathbf{x})}. \quad (2)$$

We then introduce an auxiliary flow variable u and split the minimization problem in two coupled problems

$$E(u, v) = \int \lambda |\rho(v)(\mathbf{x})| + \|\mathcal{D}u(\mathbf{x})\| d\mathbf{x} + \int \frac{1}{2\theta} \|v(\mathbf{x}) - u(\mathbf{x})\|^2 d\mathbf{x}. \quad (3)$$

This energy is then minimized iteratively in u and v in a coarse-to-fine image pyramid.

For u the energy is minimized by the methods presented in [8], and for grayscale images the minimization in v is easily achieved using the result of Zach et al. [4] which simply states that the minimizer is given by a pointwise projection to a line segment. In the setting of general vector valued images the minimizer can be found using the L^1 - L^2 minimization of Rakê et al. [9]. We will not go further in to detail about the exact procedures in this paper.

Remark. The optical flow parameters throughout the paper are set as follows: $\theta = 0.2$, 70 pyramid levels with a scale factor of 0.95. On each pyramid level we are performing 60 warps where we first solve (3) in v , and then do 10 inner iterations of the BM algorithm of [8] for minimization in u . \square

3. THE OPTIMAL PREDICTION PRINCIPLE

A question of great importance for the quality of the estimated flow is how to choose the smoothness parameter λ . An example of the effect of different λ values can be seen in Figure 1. Since different types of image sequences can have very different qualities when measured through a specific regularity or data fidelity measure, it is clear that a single choice of λ will hardly suffice for all sequences, and that methods for estimation of λ may be very useful, e.g. in video coding [10].

In this section we will consider the optimal prediction principle of Zimmer et al. [1] as a method for automatically

choosing λ . The OPP simply states that the optimal smoothness weighting λ is the one for which the corresponding flow predicts the next frame I_2 most precisely. This simple principle relies on the assumption that motion is nearly linear from I_0 through I_1 to I_2 . Instead of using the optical flow algorithm of Zimmer et al. we will evaluate the principle using the more generic and popular (although less sophisticated) TV- L^1 algorithm [4, 12, 9] and we will consider alternative prediction setups. The optimal prediction principle has previously been evaluated in an alternative setting [13], but with less convincing results than the ones originally obtained by Zimmer et al.

One can generalize the OPP to evaluate on any frame k . For the grayscale TV- L^1 optical flow algorithm, the natural way of determining the optimal smoothness is to determine the λ for which













$$P_k(\lambda) = \int_{\mathcal{T}} |I_k(\mathbf{x} + kv_\lambda(\mathbf{x})) - I_0(\mathbf{x})| d\mathbf{x}. \quad (4)$$

is minimized for some k , where v_λ is the TV- L^1 flow computed between I_0 and I_1 with smoothness parameter λ , and the non-pixel locations $\mathbf{x} + kv_\lambda(\mathbf{x})$ are evaluated using bicubic lookup. The most interesting cases are $k = -1$ corresponding to backward prediction, which means that we are assuming that the motion is linear through I_0 , $k = 1$ corresponding to choosing the flow with the best data fit and $k = 2$ which is the original OPP. Instead of using the same data fidelity measure as the optical flow algorithm, one can also choose a complementary measure, here we will also consider OPP measures on the corresponding gradient images. These measures will be denote by P_k^∇ .

Table 1 (a) holds average endpoint error (AEE) results for P_k and P_k^∇ , $k = 1, 2$ and the corresponding λ -values for the six Middlebury training sequences [11] with more than two frames, as well as the optimal λ s. We see that fixing $\lambda = 50$ will only increase the average AEE by 3% compared to the optimal values, while the original (P_2)-OPP estimate perform somewhat worse. As one would expect λ is consistently overestimated using P_1 , however interestingly P_1^∇ does not seem to cause overestimation issues at all, giving the overall best results at only 2% higher weighted average AEE than the optimal choices. We see that evaluation of prediction quality on the gradient image ∇I_2 does not work very well for the highly textured Urban sequences, which gives overall worse results of P_2^∇ than for the original OPP. This suggests that in the given setting a complementary measure of data fit may be a better indicator of quality than prediction ability of the flow.

4. LOCAL FLOW EVALUATION

In the previous section we saw that estimation of a global λ -value in the presented setup gave good results with a cleverly chosen principle, however a spatially fixed λ seems inappropriate, when data fidelity in different image regions should

	best	best common	P_1	P_2	P_1^∇	P_2^∇	$P_{\cdot,\nabla}^{0.5}$	3×3	5×5	7×7
 Grove2	0.139 (38)	0.139 (50)	0.151 (192)	0.146 (142)	0.140 (77)	0.139 (50)	 Grove2	0.135	0.135	0.136
 Grove3	0.561 (150)	0.576 (50)	0.571 (195)	0.570 (156)	0.570 (63)	0.569 (64)	 Grove3	0.549	0.553	0.556
 Hydrangea	0.189 (13)	0.206 (50)	0.260 (198)	0.257 (184)	0.193 (26)	0.190 (17)	 Hydrangea	0.175	0.176	0.177
 RubberWhale	0.128 (58)	0.130 (50)	0.161 (197)	0.134 (104)	0.129 (26)	0.129 (52)	 RubberWhale	0.114	0.113	0.112
 Urban2	0.329 (76)	0.334 (50)	0.335 (100)	0.344 (194)	0.337 (52)	0.392 (21)	 Urban2	0.317	0.317	0.320
 Urban3	0.541 (77)	0.547 (50)	0.582 (186)	0.651 (34)	0.562 (41)	0.855 (21)	 Urban3	0.483	0.475	0.473
Weighted average	1.00	1.03	1.14	1.12	1.02	1.13	Weighted average	0.937	0.935	0.938

(a)

(b)

Table 1: (a) AEE results, and corresponding $\lambda \in \{0, 1, \dots, 200\}$ (in parenthesis) for the Middlebury optical flow database training sequences (> 2 frames). The weights used in the weighted average are the lowest observed AEE (“best” column) for the given sequence. (b) AEE results for locally fused flow candidates v_λ , $\lambda \in \{10, 20, \dots, 200\}$. The name of the column indicates the size of the used neighborhood. The final fused flow has undergone two times 5×5 median filtering. The weights used in the weighted average are the same as in (a).

clearly be weighted differently. A number of optical flow algorithms try to model this by using anisotropic regularization and/or spatially weighted data terms (see e.g. [14, 1]), and while these initiatives clearly increase the performance of the algorithms, they cannot account for all reasons of spatially varying smoothness.

We consider the problem of estimating local smoothness from the point of view of the more general problem of locally determining the best of two given flow candidates. This problem has previously been considered by Mac Aodha et al. [6] and Lempitsky et al.’s FusionFlow [5]. The FusionFlow approach has been taken one step further and integrated in the algorithm of Xu et al. [3] where several high quality candidate flows are fused at each pyramid level. In this work we will use the simple ideas from the optimal prediction principle to construct a straightforward and fast method for blending multiple flows to a single superior flow.

We suggest to localize the OPP by evaluating flow vectors in a surrounding neighborhood in a fashion similar to the classic Lucas-Kanade method [15]. Inspired by the results in the previous section we evaluate directly on the (gradient) target frame ∇I_1 . I.e. we propose to use best data fit on the gradient images for the estimation, which is given by

$$P_{x_0}^{\text{loc}}(v) = \sum_{x \in \mathcal{N}(x_0)} \|\nabla I_1^\sigma(x + v(x_0)) - \nabla I_0^\sigma(x)\|, \quad (5)$$

where $\mathcal{N}(x_0)$ is a local neighborhood of x_0 . For all flow candidates we calculate the measure at each pixel location x_0 and the flow vector from the candidate with the lowest value is then assigned at x_0 .

The locally varying smoothness is obtained by calculating n flows $(v_{\lambda_i})_{1 \leq i \leq n}$ with different values of λ_i , fuse them locally according to (5), and finally smooth the resulting flow by means of some spatial regularization to avoid problems caused in $P_{x_0}^{\text{loc}}$ by e.g. untextured image neighborhoods. Two applications of 5×5 median filters have been found to work well for this. The results of this procedure for $\lambda = 10, 20, \dots, 200$, where $\mathcal{N}(x_0)$ is respectively a 3×3 , 5×5 , and 7×7 neighborhood, are found in Table 1 (b).

We see that the AEEs are consistently lower than the ones corresponding to the best single λ s from Table 1 (a), which in turn is a significant improvement over the optimal prediction principle. In Figure 2 the estimated (non smoothed) λ -field for the RubberWhale sequence as well as the optimal combination (AEE 0.091) can be found. We see that, while both are quite noisy, the estimated λ -field does seem to capture some information about edges and textures, and weight the smoothness of the flow accordingly. It is worth to notice that the solution corresponding to the λ image depicted in Fig. 2 would differ slightly from the flow resulting from solving (3) with the same pointwise weights, as the flow vectors interact through the TV regularization and coarse-to-fine pyramid. Incorporating the estimation procedure into the algorithm, introducing anisotropy by maintaining a field of λ s throughout the pyramid levels (similar to the use of the fusion step in [3]) would probably increase quality of the solutions, in addition to giving a more consistent method for estimating the field of λ values.

Finally we note that the presented method improves the AEEs on the six training sequences by almost 10% compared

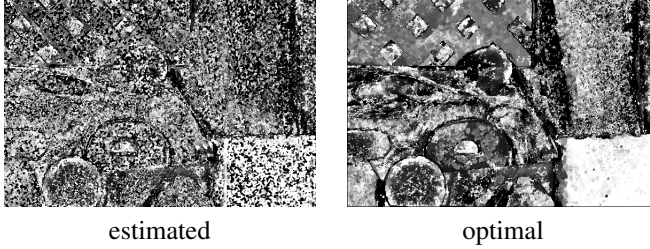


Fig. 2: The estimated λ values (3×3 neighborhood) and the optimal choice for the *RubberWhale* sequence. Dark means lower λ -value and light corresponds to a higher value.

to the results of Mac Aodha et al. [6], which fuses high quality solutions calculated among others by the very optimized TV- L^1 -improved algorithm of Wedel et al. [12] and Werlberger et al.'s Anisotropic Huber- L^1 flow algorithm.

5. CONCLUSION AND FUTURE WORK

In this work we have presented a method where local data fidelity is used to fuse flows of different smoothness. This provides a method for locally estimating the smoothness parameter of the optical flow algorithm. We have demonstrated that the proposed method consistently produces flow candidates of better quality than the single best global flow they were built from, while the computational workload is similar to that of the optimal prediction principle. Another great advantage of the proposed setup is that the estimation is done in terms of a local evaluation, and as such there is no limit on how complex a measure we can use, since we do not need to minimize a complex model.

The proposed method is generic, and can also be used to fuse flows that vary in other respects than smoothness (other parameters, different algorithms, etc.). One example could be flows calculated with different data fidelity functionals (brightness constancy vs. gradient constancy cf. [3]), which would however require a more complex local evaluation.

6. REFERENCES

- [1] H. Zimmer, A. Bruhn, and J. Weickert, "Optic flow in harmony," *International Journal of Computer Vision*, vol. 93, pp. 368–388, 2011.
- [2] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *ECCV*, 2004, vol. 4, pp. 25–36.
- [3] L. Xu, J. Jia, and Y. Matsushita, "Motion detail preserving optical flow estimation," in *CVPR'10*, 2010, pp. 1293–1300.
- [4] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime TV- L^1 optical flow," in *In Ann. Symp. German Association Patt. Recogn*, 2007, pp. 214–223.
- [5] V. Lempitsky, S. Roth, and C. Rother, "Fusionflow: Discrete-continuous optimization for optical flow estimation," in *CVPR'08*, 2008, pp. 1–8.
- [6] O. Mac Aodha, G. J. Brostow, and M. Pollefeys, "Segmenting video into classes of algorithm-suitability," in *CVPR'10*, 2010, pp. 1054–1061.
- [7] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods," *International Journal of Computer Vision*, vol. 61, pp. 211–231, 2005.
- [8] B. Goldluecke, E. Strekalovskiy, and D. Cremers, "The natural total variation which arises from geometric measure theory," *SIAM Journal on Imaging Sciences*, vol. 5, pp. 537–563, 2012.
- [9] L. Rakêt, L. Roholm, M. Nielsen, and F. Lauze, "TV- L^1 optical flow for vector valued images," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, Y. Boykov, F. Kahl, V. Lempitsky, and F. Schmidt, Eds., vol. 6819 of *Lecture Notes in Computer Science*, pp. 329–343. Springer, 2011.
- [10] X. Huang, L. Rakêt, H. Van Luong, M. Nielsen, F. Lauze, and S. Forchhammer, "Multi-hypothesis transform domain Wyner-Ziv video coding including optical flow," in *Multimedia Signal Processing (MMSP)*, 2011, pp. 1–6.
- [11] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *International Journal of Computer Vision*, vol. 31, no. 1, pp. 1–31, 2011.
- [12] A. Wedel, C. Zach, T. Pock, H. Bischof, and D. Cremers, "An improved algorithm for TV- L^1 optical flow," in *Dagstuhl Motion Workshop*, 2008.
- [13] O. Demetz, J. Weickert, A. Bruhn, and H. Zimmer, "Optic flow scale space," in *Scale-Space and Variational Methods in Computer Vision*, A. M. Bruckstein and B. ter Haar Romeny, Eds., *Lecture Notes in Computer Science*. Springer, 2011.
- [14] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof, "Anisotropic Huber- L^1 optical flow," in *BMVC*, 2009.
- [15] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*, April 1981, pp. 674–679.