

# Advanced Data Terms for Variational Optic Flow Estimation

Frank Steinbrücker, Thomas Pock, Daniel Cremers

Institute for Applied Computer Science, Bonn University  
Institute for Computer Graphics and Vision, TU Graz

## Abstract

In this paper, we present optic flow algorithms which are based on a variety of increasingly sophisticated data terms. Such data terms allow to better identify correspondences between points in either image than the traditional intensity difference since they characterize the local image structure more uniquely. We present an algorithmic framework which allows to directly incorporate arbitrary data terms: In contrast to traditional approaches the minimization scheme is not based on local linearization. In particular, we quantitatively compare the classical intensity similarity with nonconvex truncated data terms, with a patch-based intensity difference and with a patch-based normalized cross correlation. Experiments demonstrate that at the expense of additional runtime more advanced data terms may help to improve flow estimates.

## 1 Introduction

### 1.1 Variational Optic Flow Estimation

Estimating a displacement field for consecutive images from an image sequence is among the central computational challenges in Computer Vision. It arises whenever one aims to identify correspondences between points in pairs of images. Examples include motion estimation, tracking, and medical multi-modal registration. The resulting dense correspondence between pairs of points in either image can subsequently be used for structure-from-motion algorithms, for object recognition, and other higher-level tasks.

The development of appropriate cost functionals and minimization algorithms for estimating displacement or optic flow fields has received much attention in the Computer Vision community, in particular because existing variational approaches typically give rise to nonconvex energies.



**Figure 1: Advantages of Patch-based Data Terms.**

Following the seminal work of Horn and Schunck [9], most state-of-the-art algorithms for estimating a motion field  $v : (\Omega \subset \mathbb{R}^2) \rightarrow \mathbb{R}^2$  given two consecutive images  $I_1, I_2 : \Omega \rightarrow \mathbb{R}$  are based on minimizing an energy of the form

$$E(v) = \lambda E_{data}(v) + E_{reg}(v), \quad (1)$$

where the data term favors a pointwise matching of locations that are *structurally similar* and the regularity term allows to impose some prior on which velocity fields  $v$  are more or less likely. Typical regularizers impose spatial smoothness of the flow field

$$E_{reg}(v) = \int_{\Omega} \rho(|\nabla v|) dx, \quad (2)$$

using quadratic penalizers  $\rho(s) = s^2$  [9] or robust discontinuity-preserving regularizers [3, 8, 11, 13, 12]. Alternatively, one can impose motion fields to be piecewise parametric and jointly minimize respective energies with respect to motion models and motion boundaries [7].

The data term for optic flow estimation typically imposes pointwise similarity of the intensity function:

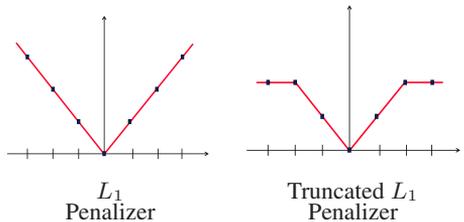
$$E_{data}(v) = \int_{\Omega} \Psi(|I_1(x) - I_2(x+v)|) dx, \quad (3)$$

again using either quadratic penalizers  $\Psi(s) = s^2$  or robust penalizers of the form  $\Psi(s) = |s|$  [3].

Unfortunately this data term is not very informative since for any given pixel in image  $I_1$  there generally exist many pixels in image  $I_2$  of the same intensity. While this ambiguity is alleviated by the regularity term, in practice having a more informative data term would allow respective algorithms to better identify correspondences without relying heavily on assumptions of spatial smoothness of the velocity field. The use of more sophisticated data terms has been limited due to the fact that energy minimization methods typically rely on local linearization schemes in order to compute optimal solutions. A correct representation of data terms such as patch-based normalized cross correlation is therefore not feasible. An exception is a recently developed variational framework which allows to additionally impose SIFT-based point matching into variational flow estimation [4]. The latter approach fundamentally differs from the approach developed in this paper in two ways: Firstly, SIFT features are only obtained for a sparse set of locations rather than for every point. Secondly, SIFT-based correspondences are *precomputed* and simply imposed in the energy minimization, i.e. the correction of respective feature points does not arise during optic flow estimation. As a consequence, incorrect correspondence estimates (based on matching wrong points) are likely to seriously deteriorate the subsequent motion estimation process.

## 1.2 Contribution

In this paper, we present and quantitatively compare a variety of more informative data terms which allow to better discriminate meaningful point correspondences. To this end, we show that a recently proposed alternative minimization scheme allows to exactly represent arbitrary data terms, including the  $L_1$ -distance computed over patches or the normalized cross correlation computed over patches. We will quantitatively analyze the performance of these alternative data terms, both regarding accuracy of estimated flow fields and regarding the increase in computation time associated with more sophisticated data terms. Specifically we will compare the standard pointwise  $L_1$ -distance of intensity to a truncated pointwise distance, an  $L_1$ -distance over local patches and a normalized cross correlation over local patches. Our experiments confirm that at the expense of additional computation time more informative data terms may give rise to bet-



**Figure 2: Schematic representation of a normal and truncated  $L_1$  penalty function. The truncated function does not penalize outliers as strongly as the normal one.**

ter optic flow estimates than traditionally considered data terms.

## 2 Advanced Data Terms in Optic Flow

In the following, we will study four different optic flow methods of the form (1) with a robust  $L_1$ -regularizer of the form  $\rho(s) = |s|$  and four different data terms

$$E_{data}(v) = \int_{\Omega} \Psi(I_1, I_2, x, v) dx \quad (4)$$

of increasing sophistication. Namely we will consider the following choices for the energy density  $\Psi$  in (4) (for simplicity we will abbreviate  $\Psi(I_1, I_2, x, v)$  with  $\Psi(x, v)$  and  $v(x)$  with  $v$ ):

1. The pointwise  $L_1$ -difference of intensity given by

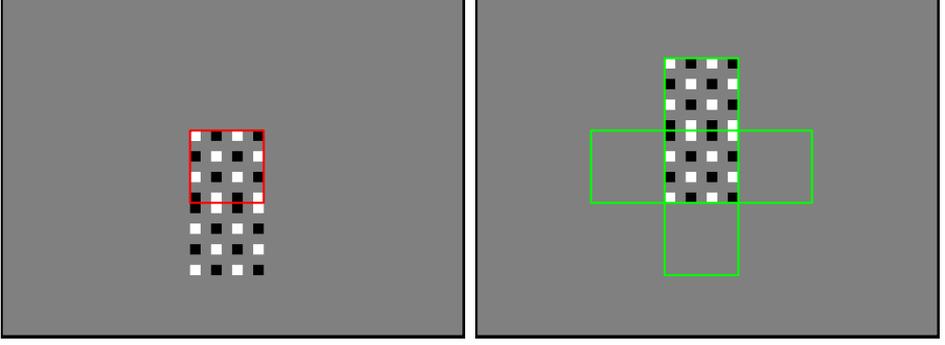
$$\Psi_{L_1}(x, v) = |I_1(x) - I_2(x + v)|. \quad (5)$$

2. A pointwise truncated linear penalizer of the intensity difference

$$\Psi_{trunc}(x, v) = \min \{|I_1(x) - I_2(x + v)|, T\}. \quad (6)$$

More than the  $L_1$  penalized intensity difference this should allow for outliers when computing a matching, because differences above the threshold value  $T$  will merely receive constant cost – see the schematic plots in Figure 2.

3. An  $L_1$ -distance of the intensity difference computed over a patch of size  $\sigma$ :



**Figure 3: Example for Patch-Based Optic Flow.** Optic flow methods based on linearization have problems determining the motion of the marked patch from frame 1 on the left to frame 2 on the right.

$$\begin{aligned} \Psi_{L_1P}(x, v) &= \int_{\Omega} G_{\sigma}(x-y) \\ &\quad \cdot |I_1(y) - I_2(y+v)| dy \\ &= G_{\sigma} * \Psi_{L_1}(x, v) \end{aligned} \quad (7)$$

defining a convolution of the intensity difference with a box function or truncated Gaussian of size  $\sigma$ . This data term favors a matching of points in either image if the local neighborhood structures are similar. In this sense it is more discriminative than the purely local intensity distance (5).

4. The normalized cross correlation (NCC) computed over a local patch of size  $\sigma$ :

$$\begin{aligned} \Psi_{ncc}(x, v) &= 1 - \int_{\Omega} P_1(x, y) P_2(x+v, y) dy \\ &= 1 - \langle P_1(x), P_2(x+v) \rangle, \end{aligned} \quad (8)$$

where the local signature  $P_i$  is computed for image  $I$  by removing the local average and normalizing the resulting vector of intensities:

$$P_i(x, y) = \frac{J_i(x, y) - \overline{J_i(x)}}{\sqrt{\int_{\Omega} (J_i(x, z) - \overline{J_i(x)})^2 dz}}. \quad (9)$$

Hereby,  $J_i(x, y)$  denotes the image  $I_i$  restricted to the neighborhood of  $x$  and  $\overline{J_i(x)}$  denotes its average:

$$J_i(x, y) = I_i(y) \cdot G_{\sigma}(x-y), \quad (10)$$

$$\overline{J_i(x)} = \frac{\int_{\Omega} J_i(x, y) dy}{\int_{\Omega} G_{\sigma}(x-y) dy} \quad (11)$$

$J_i(x, y)$  and  $P_i(x, y)$  denote the value in image  $i$  at location  $y$  in the patch around  $x$ .

In contrast to other robust data terms, such as the image gradient or Hessian in [13, 12], the NCC is invariant to multiplicative illumination changes.

## 2.1 From Patch Comparison to the CLG Method

Interestingly, the proposed patch comparison is related to the combined local global (CLG) method of Bruhn et al. [5]. The CLG method was originally introduced as a fusion of Lucas and Kanade [10] with Horn and Schunck [9]. In the following, we will prove that local linearization applied to the squared  $L_2$ -distance over local patches leads to the CLG method:

Replacing the  $L_1$ -norm in equation (7) with the squared  $L_2$ -norm and linearizing the data term yields the following equation:

$$\begin{aligned} \Psi_{L_2P} &= G_{\sigma} * (I_1(y) - I_2(y+v))^2 \\ &\approx G_{\sigma} * (I_1(x) - \nabla I(x)^{\top} v - I_2(x))^2 \\ &= G_{\sigma} * (v^{\top} \nabla I \nabla I^{\top} v + (I_2 - I_1) \nabla I) \\ &=: \hat{v}^{\top} J_{\sigma}(\nabla_3 I) \hat{v}, \end{aligned} \quad (12)$$

where  $\hat{v} = \begin{pmatrix} v \\ 1 \end{pmatrix}$  is the velocity vector in homogeneous coordinates and  $J_{\sigma}(\nabla_3 I)$  is the smoothed

spatio-temporal structure tensor  $(\partial_x, \partial_y, \partial_t)^\top$ :

$$J_\sigma(\nabla_3 I) = G_\sigma * \nabla_3 I \nabla_3 I^\top. \quad (13)$$

This yields the well-known CLG-method.

## 2.2 Drawbacks of Linearization

The CLG method can be extended to nonlinear data terms by applying an additional function  $\Phi$  to the motion tensor:

$$J_\sigma(\nabla_3 I) = \begin{cases} \Phi(G_\sigma * \nabla_3 I \nabla_3 I^\top), \text{ or} \\ G_\sigma * \Phi(\nabla_3 I \nabla_3 I^\top) \end{cases} \quad (14)$$

However, the motion tensor is still based on the linearized constancy assumption, yielding several drawbacks. Figure 3 illustrates some of these drawbacks. The marked patch in frame 1 on the left clearly moves upward in frame 2. However, a linearized data term as in equation (12) does not lead to this decision:

Assuming  $G_\sigma$  to be radial symmetric and the origin to lie in the center of the patch, for every image gradient at point  $\begin{pmatrix} p \\ q \end{pmatrix}$ , there is an orthogonal image gradient of equal magnitude at point  $\begin{pmatrix} -q \\ p \end{pmatrix}$ . Summed up in  $G_\sigma * \nabla I \nabla I^\top$ , these gradients create a multiple of the identity matrix:

$$\begin{pmatrix} p \\ q \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}^\top + \begin{pmatrix} -q \\ p \end{pmatrix} \begin{pmatrix} -q \\ p \end{pmatrix}^\top = (p^2 + q^2) I_d \quad (15)$$

Moreover, for every gradient at point  $\begin{pmatrix} p \\ q \end{pmatrix}$  there is an opposing gradient of equal magnitude at point  $\begin{pmatrix} -p \\ -q \end{pmatrix}$  with an equal temporal derivative  $(I_2 - I_1)$  in the patch. In  $G_\sigma * ((I_2 - I_1) \nabla I)$  these sum up to 0, leaving

$$\Psi_{L_{2p}} \approx \gamma v^\top, \quad \gamma \in \mathbb{R}_+ \quad (16)$$

Without the regularity term, this suggests the patch does not move at all, which is exactly the worst choice in this example.

This problem comes from the fact that linearized data terms can only be used to compute small displacements, where the higher-order derivatives in the Taylor-series approximating the displaced point do not preponderate. Therefore, displacements larger than the patch to be displaced cannot be computed with linearized data terms. If only small displacements are being considered, the problem of

matching the center patch to one of the patches surrounding it cannot arise.

The standard solution to the problem of computing large displacements with linearized data terms is embedding the approach into a coarse-to-fine hierarchy in a Gaussian scale-space. On coarser scales, the displacements are sufficiently small for the linearized data term to be a good approximation of the nonlinear one. However, for this example this method will fail as well, since on coarser scales of the Gaussian scale-space the intensity of a patch converges to the average gray value of the patch it covers on finer scales. The average gray values of all the patches in frame 2 of Figure 3 are the same, rendering each patch an equally good match for the center patch.

This mutual annihilation of point information in the patch suggests to avoid linearization for patch-based data terms in optic flow computation.

## 2.3 Optimization of Arbitrary Data Terms

Recently Steinbrücker et al.[14] showed that one can entirely avoid the linearization of the data term in optic flow estimation. To this end, they reformulate the original optimization problem (1) using a quadratic relaxation scheme [6, 1, 15] and solve the resulting optimization problem

$$E(v, u) = \int_{\Omega} \lambda \Psi(x, v) + \frac{1}{2\theta} (v - u)^2 + \rho(|\nabla u|) dx. \quad (17)$$

This energy functional has the nice property, that if either one of the fields  $u$  or  $v$  is kept fixed, the functional can be globally minimized with respect to the other field.

If  $v$  is fixed, the functional is convex in  $u$ . Therefore, a global minimizer  $u$  can be computed efficiently, for example by gradient descent. If  $u$  is fixed, the functional has only a pointwise dependency on  $v$ . Therefore, it can be minimized globally with respect to  $v$  by a complete search.

Moreover, this search can use any data term proposed in Section 2, as well as any other data term that can be sampled. In the following section, we will demonstrate that the use of patch-based data terms can yield a great deal of improvement with respect to the quality of the computed flow field. The minimizer of the functional in equation (17) is computed by alternating the two globally optimal minimization steps described above for  $u$  and  $v$ ,

AEE [pixel]	$\Psi_{L_1}$	$\Psi_{trunc}$	$\Psi_{L_1p}$	$\Psi_{ncc}$
$\lambda$	50	50	30	10
Rubberwhale	0.1735	0.1724	0.1658	<b>0.0836</b>
Hydrangea	0.2104	0.2107	0.1864	<b>0.1353</b>
Dimetrodon	0.1709	0.1704	<b>0.1645</b>	0.2219
Grove2	0.1852	0.1852	<b>0.1785</b>	0.1924
Urban2	2.6375	2.6015	<b>1.7914</b>	3.9976
Venus	0.3971	0.3964	<b>0.3171</b>	0.4356

**Table 1: Comparison of the data terms regarding the average end point error on different Middlebury sequences.**

keeping the respective other one fixed. While this approach does not guarantee to find the global optimum of the whole functional, it delivers very good results compared to other methods, as described in [14]. To achieve subpixel-accuracy in the data terms, they are sampled on a regular grid of subpixel positions, using bilinear interpolation of the images.

The parameter  $\theta$ , which can be interpreted as an annealing parameter in equation (17), is decremented in each iteration, letting  $u$  and  $v$  converge in the end.

### 3 Experimental Evaluation

The focus of this paper is a comparison of different data terms for variational optic flow and not of different algorithms. Therefore we constrain our quantitative analysis to sequences of the Middlebury benchmark for optic flow [2] with known ground truth, namely ‘‘Rubberwhale’’, ‘‘Hydrangea’’, ‘‘Dimetrodon’’, ‘‘Grove2’’, ‘‘Urban2’’, and ‘‘Venus’’, where for all data terms we can determine optimal parameter values. For each sequence the optic flow was computed from frame 10 to frame 11 and evaluated against the provided ground truth flow by means of the average endpoint error (AEE), defined as

$$AEE(v) = \frac{\sum_{x \in \Omega_d} |v(x) - v_{truth}(x)|}{|\Omega_d|} \quad (18)$$

on the discrete pixel grid  $\Omega_d$ .

Since both minimization steps for computing the minimizing  $u$  and  $v$  for equation (17) can be easily parallelized, we implemented our methods

Time [s]	$\Psi_{L_1}$	$\Psi_{trunc}$	$\Psi_{L_1p}$	$\Psi_{ncc}$
Rubberwhale	381	404	4366	5587
Hydrangea	381	404	4366	5587
Dimetrodon	381	404	4366	5587
Grove2	481	512	4919	7266
Urban2	481	512	4919	7266
Venus	263	280	2581	3936

**Table 2: Comparison of the data terms regarding the their runtime on different Middlebury sequences.**

on NVIDIA’s ‘‘Tesla C1060’’ graphics card to significantly speed up the computation. The intensity values of the images were rescaled to the interval  $[0, 1]$  and for  $\Psi_{trunc}$  the truncation threshold  $T$  was set to  $\frac{1}{3}$ . For the patch-based  $L_1$  difference and normalized cross correlation we used patches consisting of 9 samples on a rectangular  $3 \times 3$  pixel grid.

Table 1 lists the AEE values for the proposed data terms and the standard  $L_1$  image intensity difference. For each data term, the regularity value  $\lambda$  was optimized with respect to the AEE summed over all images.

It can be seen that while some of the proposed data terms perform worse than the standard  $L_1$  image intensity difference on some images, on other images they were able to reduce the error to less than one half compared to the  $L_1$  image intensity difference.

Figures 4 and 5 show visualizations of the corresponding flow fields. While there is hardly any visual difference between the flow fields of  $\Psi_{L_1}$  and  $\Psi_{trunc}$ , the two patch-based data terms show a significantly discriminative potential. For example, the normalized cross correlation is able to correctly detect the hole in one of the objects in the ‘‘Rubberwhale’’ sequence (also depicted in Figure 1), while the other data terms fail in that regard. Something similar can be observed in the ‘‘Venus’’ sequence: The patch-based data terms describe the gap in the middle of the image much better than the other ones do. However, the ‘‘Venus’’ sequence also shows very well that the patch-based data terms tend to produce more outliers than the other data terms.

Table 2 lists the computation time for each image and data term. The times are proportional to the image size and patch size, with a small additional com-

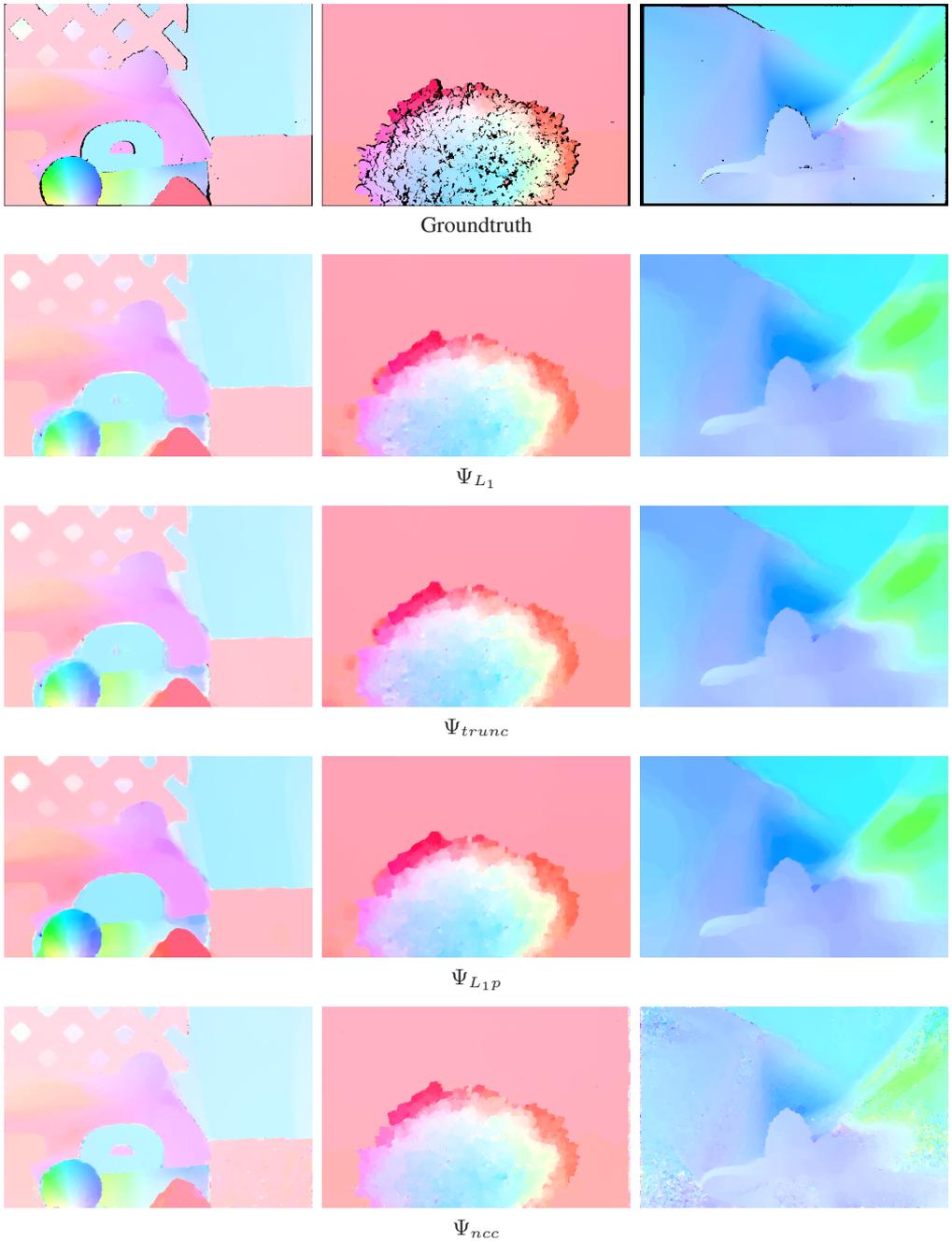
putation and caching overhead for the patch-based data terms.

## 4 Conclusion

In this paper, we introduced novel more informative data terms for variational optic flow estimation which allow to better discriminate meaningful point correspondences. To this end we build up on a recently suggested alternative minimization strategy which does not require local linearization of the data term. We show that the alternative minimization method allows to represent dissimilarity measures like patch-based intensity distance or patch-based normalized cross correlation accurately without approximations that arise through the traditional linearization. A quantitative comparison of various data terms of increasing sophistication demonstrates that at the expense of increased runtime these more informative data terms lead to improvements in optic flow estimation.

## References

- [1] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition - modeling, algorithms, and parameter selection. *Int. J. of Computer Vision*, 67(1):111–136, 2006. 4
- [2] S. Baker, D. Scharstein, J.P. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *IEEE Int. Conf. on Computer Vision*, 2007. 5
- [3] M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *IEEE Int. Conf. on Computer Vision*, pages 231–236, 1993. 1
- [4] T. Brox, C. Bregler, and J. Malik. Large displacement optical flow. In *Proc. IEEE Conf. on Comp. Vision Patt. Recog. (CVPR'09)*, 2009. 2
- [5] A. Bruhn, J. Weickert, and C. Schnoerr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods - updated version with errata. *Int. J. of Computer Vision*, 61(3):211–231, 2005. 3
- [6] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Im. Vis.*, 20(1-2):89–97, 2004. 4
- [7] D. Cremers and S. Soatto. Motion Competition: A variational framework for piecewise parametric motion segmentation. *Int. J. of Computer Vision*, 62(3):249–265, May 2005. 1
- [8] R. Deriche, P. Kornprobst, and G. Aubert. Optical flow estimation while preserving its discontinuities: A variational approach. In *Asian Conf. on Computer Vision*, volume 2, pages 290–295, Singapore, 1995. 1
- [9] B.K.P. Horn and B.G. Schunck. Determining optical flow. *A.I.*, 17:185–203, 1981. 1, 3
- [10] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc.7th International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, 1981. 3
- [11] H.H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 8(5):565–593, 1986. 1
- [12] 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*, 67(2):141–158, April 2006. 1, 3
- [13] C. Schnörr. Segmentation of visual motion by minimizing convex non-quadratic functionals. In *12th Int. Conf. on Pattern Recognition*, Jerusalem, Israel, October 1994. 1, 3
- [14] F. Steinbrücker, T. Pock, and D. Cremers. Large displacement optical flow computation without warping. In *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan, 2009. 4, 5
- [15] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime TV-L1 optical flow. In *Pattern Recognition (Proc. DAGM)*, LNCS, pages 214–223, Heidelberg, Germany, 2007. Springer. 4



**Figure 4: From Left to Right: Flow Images of the “Rubberwhale”, “Hydrangea”, and “Dimetrodon” sequences.**



Groundtruth



$\Psi_{L_1}$



$\Psi_{trunc}$



$\Psi_{L_1P}$



$\Psi_{ncc}$

Figure 5: From Left to Right: Flow Images of the “Grove2”, “Urban2”, and “Venus” sequences.