Lecture 17:

Image Sequence Analysis II: Models for the Data Term

Contents

1. Introduction
2. Constancy Assumptions
3. Extensions to Colour Image Sequences
4. Motion Tensor Notation
5. More Robust Data Terms
6. Confidence Measure

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Introduction

- Variational optic flow methods obtain the displacement field

$$\mathbf{u} = \begin{pmatrix} u_1(x_1, x_2, x_3) \\ u_2(x_1, x_2, x_3) \\ 1 \end{pmatrix}$$

as minimiser of

$$E(\mathbf{u}) = \int_\Omega \Big(\underbrace{M(D^k f, \mathbf{u})}_{\text{data term}} + \alpha \underbrace{S(\nabla f, \nabla \mathbf{u})}_{\text{regulariser}} \Big)\, d\mathbf{x}$$

- Ω is either a spatial or a spatiotemporal domain.
- spatial case: $\mathbf{x} := (x_1, x_2)^\top$ and $\nabla := \nabla_2 := (\partial_{x_1}, \partial_{x_2})^\top$
  spatiotemporal case: $\mathbf{x} := (x_1, x_2, x_3)^\top$ and $\nabla := \nabla_3 := (\partial_{x_1}, \partial_{x_2}, \partial_{x_3})^\top$
- In Lecture 16, we have discussed different smoothness terms.
  Now let us analyse the options for the data term.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Constancy Assumptions

Brightness Constancy (Horn/Schunck 1981)

- Assuming that the brightness f does not change along path $(x_1(x_3), x_2(x_3))$ gives

$$\begin{aligned} 0 &= \frac{df(x_1(x_3), x_2(x_3), x_3)}{dx_3} \\ &= f_{x_1} u_1 + f_{x_2} u_2 + f_{x_3} \end{aligned}$$

where $f_{x_i} := \partial_{x_i} f$ and $u_i = \partial_{x_3} x_i$.

- Deviations from the brightness constancy assumption

$$0 = \mathbf{u}^\top \nabla_3 f$$

can be penalised in a least squares sense by the data term

$$M_1(D^1 f, \mathbf{u}) := (\mathbf{u}^\top \nabla_3 f)^2$$

M I A

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Gradient Constancy (Uras et al. 1988)

- Idea: Do not assume also constancy on the brightness, but on the orientation of the local structure.
- Imposing constancy of the (spatial) brightness gradient $\nabla_2 f = (f_{x_1}, f_{x_2})^\top$ gives

$$\begin{aligned} \mathbf{u}^\top \nabla_3 f_{x_1} &= 0, \\ \mathbf{u}^\top \nabla_3 f_{x_2} &= 0. \end{aligned}$$

- Squaring and summing up yields the data term

$$M_2(D^2 f, \mathbf{u}) := \sum_{i=1}^{2} (\mathbf{u}^\top \nabla_3 f_{x_i})^2.$$

- Since it assumes constancy on the orientation, it is good for translational and divergent motion.
- appears to be less suitable for significant rotational motion (orientation changes)

Constancy Assumptions (3)

Constancy of the Hessian

- constancy of the (spatial) Hessian of f:

$$\begin{aligned} \mathbf{u}^\top \nabla_3 f_{x_1 x_1} &= 0, \\ \mathbf{u}^\top \nabla_3 f_{x_1 x_2} &= 0, \\ \mathbf{u}^\top \nabla_3 f_{x_2 x_1} &= 0, \\ \mathbf{u}^\top \nabla_3 f_{x_2 x_2} &= 0. \end{aligned}$$

- corresponding quadratic penaliser:

$$M_3(D^3 f, \mathbf{u}) := \sum_{i=1}^{2} \sum_{j=1}^{2} (\mathbf{u}^\top \nabla_3 f_{x_i x_j})^2.$$

- same problem as for gradient constancy:
  good for translational and divergent motion, more problematic for significant rotational motion

Constancy Assumptions (4)

Constancy of the Gradient Magnitude

- The gradient magnitude $|\nabla_2 f| = \sqrt{f_{x_1}^2 + f_{x_2}^2}$ is invariant under rotations.

- Thus, assuming constancy of the gradient magnitude is better suited for rotational motion:

$$M_4(D^2 f, \mathbf{u}) := (\mathbf{u}^\top \nabla_3 |\nabla_2 f|)^2$$

Constancy of the Laplacian

- The Laplacian $\Delta_2 = f_{x_1 x_1} + f_{x_2 x_2}$ is the trace of the Hessian (and therefore the sum of its eigenvalues).
  It is invariant under rotations.

- Constancy of the (spatial) Laplacian $\Delta_2 f$:

$$M_5(D^3 f, \mathbf{u}) := (\mathbf{u}^\top \nabla_3 (\Delta_2 f))^2$$

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Constancy of the Determinant of the Hessian

- The determinant of the Hessian (product of its eigenvalues) is given by

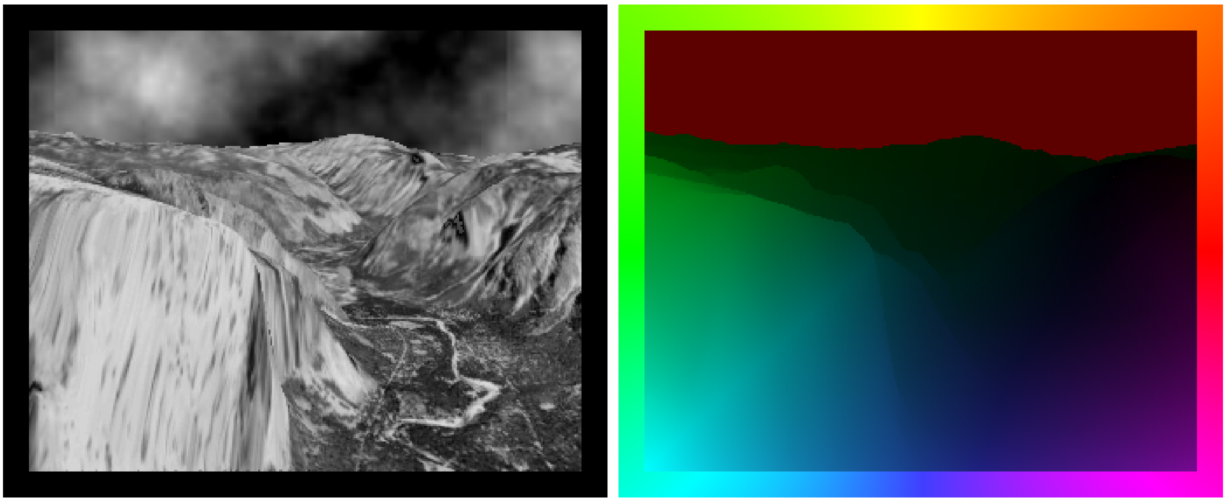
$$\det(\mathcal{H}_2 f) = f_{x_1 x_1} f_{x_2 x_2} - f_{x_1 x_2}^2 .$$

  It is invariant under rotations as well.

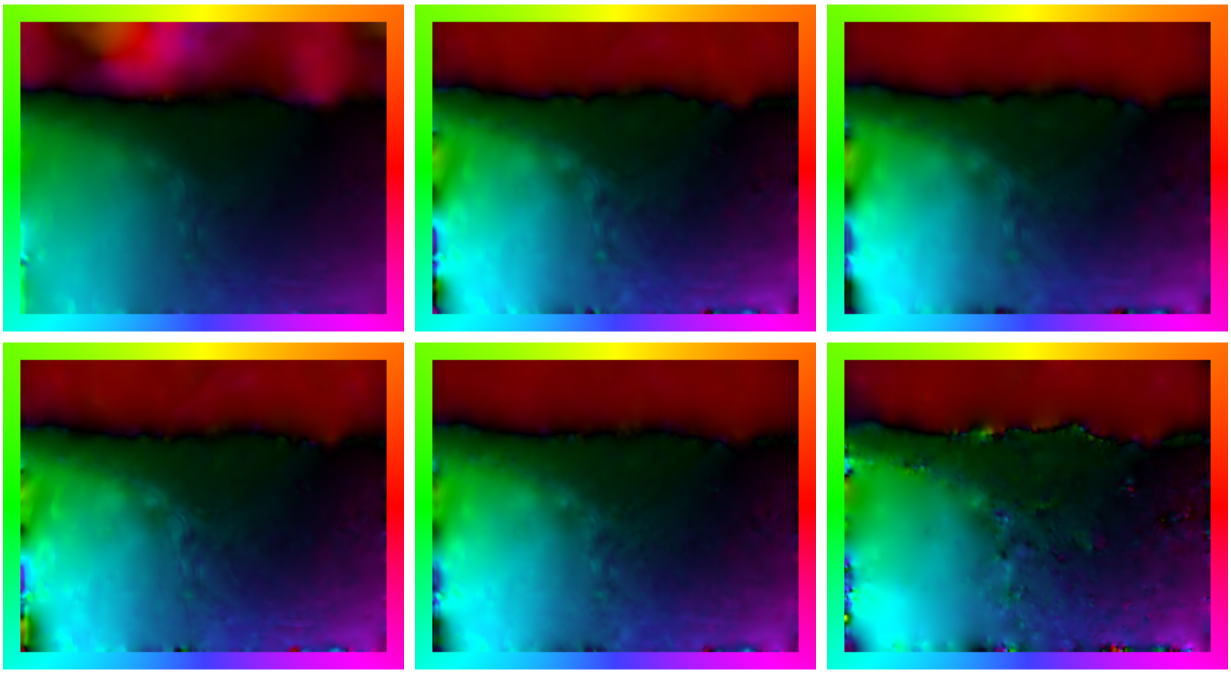
- This gives rise to the data term

$$M_6(D^3 f, \mathbf{u}) := (\mathbf{u}^\top \nabla_3 \det(\mathcal{H}_2 f))^2$$

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27



(a) **Left:** Frame 8 of the *Yosemite sequence with clouds*, depicting translational and divergent motion, 316×256 pixels. (b) **Right:** Ground truth flow field, colour-coded.



(a) **Top left:** Colour-coded flow field with the data term M_1 (brightness constancy) and a spatial Horn–Schunck regulariser. (b) **Top middle:** Data term M_2 (gradient constancy). (c) **Top right:** Data term M_3 (Hessian constancy). (d) **Bottom left:** Data term M_4 (constancy of gradient magnitude). (e) **Bottom middle:** Data term M_5 (constancy of Laplacian). (f) **Bottom right:** Data term M_6 (constancy of Hessian determinant). Author: A. Bruhn (2006)

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Quantitative Evaluation

- Embed the different data terms in an energy functional with a spatial homogeneous regulariser:

$$E(\mathbf{u}) \;=\; \int_\Omega \Big(M\left(D^k f, \mathbf{u}\right) + \alpha\,\left(|\nabla u_1|^2 + |\nabla u_2|^2\right) \Big)\, d\mathbf{x}.$$

- Compute the *average angular error*

$$\mathrm{AAE}(\mathbf{u}_\mathrm{e}, \mathbf{u}_\mathrm{c}) \;=\; \frac{1}{|\Omega|} \int_\Omega \arccos\left(\frac{\mathbf{u}_e^\top \mathbf{u}_c}{|\mathbf{u}_e||\mathbf{u}_c|} \right) d\mathbf{x}$$

  between the estimated flow field $\mathbf{u}_e = (u_{e1}, u_{e2}, 1)^\top$ and the correct flow field $\mathbf{u}_c = (u_{c1}, u_{c2}, 1)^\top$.

- Optimise the model parameters such that the average angular error is minimised.

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Constancy assumptions and corresponding average angular errors for Yosemite sequence with clouds.

Constancy Assumption	Average Angular Error
brightness	7.17^0
gradient	5.91^0
Hessian	6.46^0
gradient magnitude	6.37^0
Laplacian	6.18^0
Hessian determinant	8.10^0

- Gradient constancy can to be more useful than brightness constancy:
  It incorporates orientation information.
- Constancy of higher-order derivatives does not necessarily improve the result.
- Potential advantages of rotationally invariant expressions are absent when no pronounced rotational motion is involved.
- The data term of the Hessian determinant is too sparse to be useful.

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Extensions to Colour Image Sequences

- For a colour sequence

$$\mathbf{f}(x_1, x_2, x_3) = \begin{pmatrix} f_1(x_1, x_2, x_3) \\ f_2(x_1, x_2, x_3) \\ f_3(x_1, x_2, x_3) \end{pmatrix}$$

  the constancy assumptions are applied separately to each channel.

- Example: Brightness constancy assumption gives

$$\begin{aligned} \mathbf{u}^\top \nabla_3 f_1 &= 0, \\ \mathbf{u}^\top \nabla_3 f_2 &= 0, \\ \mathbf{u}^\top \nabla_3 f_3 &= 0. \end{aligned}$$

  Penalisation in a least squares sense leads to the data term

$$M(D^2 f, \mathbf{u}) = \sum_{i=1}^{3} (\mathbf{u}^\top \nabla_3 f_i)^2.$$

- The smoothness term is unaltered for flow-driven methods, since one searches for a joint displacement vector field $\mathbf{u} = (u_1, u_2, 1)^\top$ for all three channels.

- For image-driven regularisers, one replaces
  - in the isotropic case: $g(|\nabla f|^2)$ by $g(\sum_i |\nabla f_i|^2)$
  - in the anisotropic case: the projection matrix

$$D(\nabla f) \; := \; \frac{1}{|\nabla f|^2 + 2\lambda^2} \left(\nabla f^\perp \nabla f^{\perp\top} + \lambda^2 I\right)$$

  by

$$D(\nabla f) \; := \; \frac{1}{\sum_i |\nabla f_i|^2 + 2\lambda^2} \left(\sum_i \nabla f_i^\perp \nabla f_i^{\perp\top} + \lambda^2 I\right).$$

- Incorporating colour information in the optic flow estimation often gives slightly better results than transforming the sequence into a grey value sequence first.

Motion Tensor Notation

(Bruhn 2006)

- Let us now introduce a convenient general notation for the data term.

- We consider constancy assumptions on some image features $p_1,\ldots,p_n$.

- A weighted sum of the n linearised squared constraints gives

$$\begin{aligned} M &= \sum_{i=1}^{n} \gamma_i (\mathbf{u}^\top \nabla_3 p_i)^2 \\ &= \sum_{i=1}^{n} \gamma_i \, \mathbf{u}^\top \nabla_3 p_i \nabla_3 p_i^\top \mathbf{u} \\ &= \mathbf{u}^\top \left(\sum_{i=1}^{n} \gamma_i \nabla_3 p_i \nabla_3 p_i^\top \right) \mathbf{u} \end{aligned}$$

- Introducing the *motion tensor*

$$J := \sum_{i=1}^{n} \gamma_i \nabla_3 p_i \nabla_3 p_i^\top$$

allows to rewrite the data term in a compact way as a quadratic form:

$$M \;=\; \mathbf{u}^\top J \mathbf{u}.$$

- enables simple, modular implementations of different data assumptions
- Example: Assuming brightness and gradient constancy gives

$$J \;=\; \gamma_1 \, \nabla_3 f \, \nabla_3 f^\top + \gamma_2 \, \left(\nabla_3 f_{x_1} \nabla_3 f_{x_1}^\top + \nabla_3 f_{x_2} \nabla_3 f_{x_2}^\top \right)$$

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

More Robust Data Terms

There are several options to render a data term $M = \mathbf{u}^\top J \mathbf{u}$ more robust under noise and outliers (where the model assumptions are violated, e.g. at occlusion areas).

Local Least Squares Fitting

- assumes that the flow field is constant within some neighbourhood of radius ρ
- weighted least squares fitting with a Gaussian K_ρ gives

$$\begin{aligned} M &= K_\rho * (\mathbf{u}^\top \, J_0 \, \mathbf{u}) \\ &= \mathbf{u}^\top \, (K_\rho * J_0) \, \mathbf{u} \end{aligned}$$

- This comes down to a Gaussian smoothing of the original motion tensor J_0:

$$J = K_\rho * J_0.$$

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

More Robust Data Terms (2)

Example

- Lucas and Kanade (1981) considered a local least squares fitting of the brightness constancy assumption. Using a Gaussian window function gives

$$J = K_\rho * (\nabla_3 f \, \nabla_3 f^\top).$$

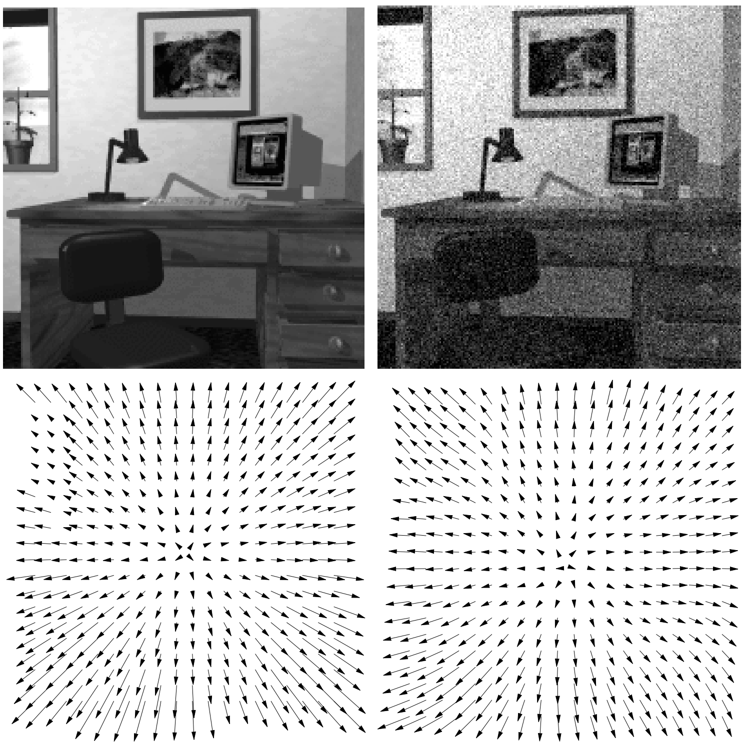
  This is nothing else but the spatiotemporal structure tensor.

- When combining the corresponding data term with a Horn-Schunck regulariser, one obtains the *combined local-global (CLG) method* of Bruhn et al. (2005):

$$E(\mathbf{u}) = \int_\Omega \Big(\mathbf{u}^\top \, \big(K_\rho * (\nabla_3 f \, \nabla_3 f^\top)\big) \, \mathbf{u} + \alpha \, \big(|\nabla u_1|^2 + |\nabla u_2|^2\big) \Big) \, d\mathbf{x}.$$

MI A 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

More Robust Data Terms (3)



(a) **Top left:** Frame 10 of the synthetic *office* sequence. (b) **Top right:** Degraded by Gaussian noise with $\sigma_n = 20$. (c) **Bottom left:** Ground truth optic flow field. (d) **Bottom right:** Computed optic flow field using the 2-D CLG method for the noisy sequence.

More Robust Data Terms (4)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Adaptive Averaging with Nonlinear Diffusion

(Brox et al. 2006)

- Local least squares fitting replaces a motion tensor J_0 by its Gaussian-smoothed variant $K_\rho * J_0$. This may be regarded as linear diffusion filtering of some tensor field $J_0(x_1, x_2, x_3)$; cf. Lecture 14.
- A natural modification would be to replace linear diffusion by some discontinuity-preserving nonlinear diffusion filter for tensor fields.
- can be regarded as adaptive averaging

More Robust Data Terms (5)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Nonquadratic Penalisation

(Black/Anandan 1991, Mémin/Pérez 1998, Hinterberger et al. 2002):

- The previous approaches are based on a data term that involves quadratic penalisation of deviations from constancy assumptions:

$$M = \mathbf{u}^\top J \mathbf{u}.$$

- One can gain robustness against outliers by substituting quadratic by nonquadratic penalisation:

$$M = \Psi(\mathbf{u}^\top J \mathbf{u})$$

- The penaliser is chosen as an increasing function $\Psi(s^2)$ that is convex in s, e.g.

$$\Psi(s^2) \;:=\; \varepsilon s^2 + \sqrt{\beta^2 + s^2}$$

  It approximates the L^1 penaliser $\Psi(s^2) = |s|$ for $\varepsilon, \beta \to 0$.

A Confidence Measure

Motivation

- Variational optic flow methods are global methods that give dense flow fields. However, the flow estimates do not have the same reliability at all locations.
- The gradient magnitude $|\nabla f|$ has been proposed as a confidence measure for such global methods, but does not work well (Barron et al. 1994).
- Local methods have natural confidence measures that help to avoid computing flow values at locations where there is not enough information for a reliable estimate.
- Is there a good and natural confidence measure for global methods based on energy functionals?

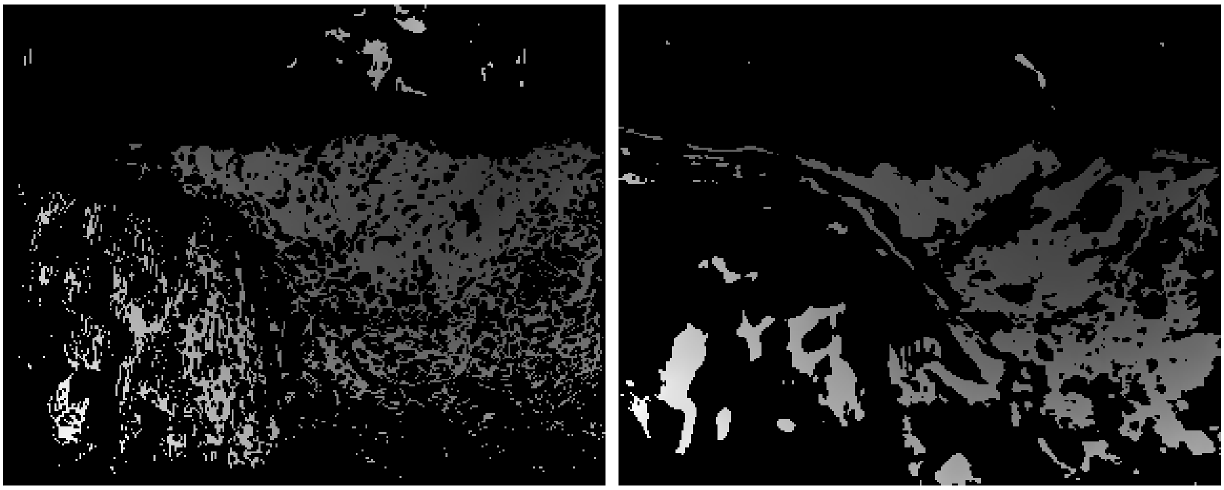
A Simple Confidence Measure for Global Energy Functionals

(Bruhn/W. 2006)

- The global energy functional E penalises deviations from model assumptions by summing up the local deviations E_i from all pixels i.
- use E_i as a local confidence measure:
  - Locations where E_i is large violate the model assumptions and are regarded as unreliable.
  - Small values for E_i indicate high reliability.
- can be used for sparsifying the flow field:
  For finding the 30 % locations with the highest reliability, discard the 70 % of all pixels where E_i has the largest values.
- may work surprisingly well in spite of its simplicity

A Confidence Measure (3)

M I A 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27



Confidence criterion for the *Yosemite* sequence. (a) **Left:** Locations with the lowest contributions to the energy (20 % quantile). The non-black grey values depict the optic flow magnitude. (b) **Right:** Locations where the angular error is really lowest (20 % quantile).

A Confidence Measure (4)

M I A 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Comparison between the “nondense” results from Barron *et al.* and the sparsified linear 3-D CLG results for the *Yosemite* sequence with cloudy sky. AAE = average angular error. CLG = average angular error of the CLG method with the same density.

Technique	Density	AAE	CLG
Singh, step 2, $\lambda_1 \leq 0.1$	97.7 %	10.03°	6.02°
Heeger, level 0	64.2 %	22.82°	2.89°
Weber/Malik	64.2 %	4.31°	2.89°
Horn/Schunck, original, $\|\nabla f\| \geq 5$	59.6 %	25.33°	2.61°
Heeger, combined	44.8 %	15.93°	1.98°
Lucas/Kanade, $\lambda_2 \geq 1.0$	35.1 %	4.28°	1.62°
Fleet/Jepson, $\tau = 2.5$	34.1 %	4.63°	1.59°
Horn/Schunck, modified, $\|\nabla f\| \geq 5$	32.9 %	5.59°	1.55°
Nagel, $\|\nabla f\| \geq 5$	32.9 %	6.06°	1.55°
Fleet/Jepson, $\tau = 1.25$	30.6 %	5.28°	1.48°
Heeger, level 1	15.2 %	9.87°	1.13°
Uras *et al.*, $\det(H) \geq 1$	14.7 %	7.55°	1.11°
Singh, step 1, $\lambda_1 \leq 6.5$	11.3 %	12.01°	1.05°
Waxman *et al.*, $\sigma_f = 2.0$	7.4 %	20.05°	0.94°
Heeger, level 2	2.4 %	12.93°	0.76°

A Confidence Measure (5)

A Surprising Result

- The success of the confidence measure shows that locations with *low (!)* gradient magnitude belong to the most reliable locations for global methods:
  They give the smallest values for E_i.

Reason for this Behaviour

- Locations with high gradient magnitudes may lack reliability:
  - Noise may cause high gradients.
  - Occlusions may also result in high gradients.
- Locations with low gradient magnitudes can be more reliable:
  - They benefit from the filling-in effect that propagates information from *all* surrounding locations where the gradients are high.
  - This averages the local errors and leads to higher reliability.

Summary

Summary

- Besides assuming brightness constancy, alternative constancy assumptions can be useful, e.g.
  - constancy of the various partial derivatives (gradient, Hessian, ...)
  - constancy of expressions that are invariant under rotations (gradient magnitude, Laplacian, determinant of the Hessian)
- Colour image sequences apply constancy assumptions separately for each channel. The data term sums over all channels.
- The motion tensor allows a compact notation of various constancy assumptions.
- Additional robustness can be achieved by
  - integrating the constancy assumption over a neighbourhood
  - adaptive averaging by nonlinear tensor-valued diffusion
  - replacing quadratic penalisation by nonquadratic one
- A simple confidence measure exists that is based on the local contribution to the energy. Pixels with high contributions are less reliable.

References

- J. Weickert, A. Bruhn, T. Brox, N. Papenberg: A survey on variational optic flow methods for small displacements. In O. Scherzer (Ed.): *Mathematical Models for Registration and Applications to Medical Imaging*. Springer, Berlin, pp. 103–136, 2006. (`http://www.mia.uni-saarland.de/publications.shtml`). *(discussion of different data terms)*
- A. Bruhn, J. Weickert, C. Schnörr: Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, Vol. 61, No. 3, 211–231, 2005. (`http://www.mia.uni-saarland.de/publications.shtml`). *(CLG method)*
- T. Brox, J. Weickert, B. Burgeth, P. Mrázek: Nonlinear structure tensors. *Image and Vision Computing*, Vol. 24, No. 1, 41–55, Jan. 2006. (`http://www.mia.uni-saarland.de/publications.shtml`). *(motion tensor smoothed with nonlinear diffusion)*
- A. Bruhn, J. Weickert: A confidence measure for variational optic flow methods. In R. Klette, R. Kozera, L. Noakes, J. Weickert (Eds.): Geometric Properties from Incomplete Data. Springer, Dordrecht, pp. 283–297, 2006. (`http://www.mia.uni-saarland.de/publications.shtml`). *(detailed evaluation of the confidence measure)*