

## Contents

Chapter 1. Background	3
1. Optical Flow	3
2. Setup of the camera rig	6
3. Image Fusion	6
4. Lie Groups and the Noether Theorem	8
5. Total Variation	13
Chapter 2. Geometrical Prior	17
1. The Generalized Structure Tensor	18
2. Structure Tensor Based Prior	20
Chapter 3. Geometrical Optical Flow Model	23
1. Disparity	23
2. Localization	24
3. The solution algorithm	26
Chapter 4. Results	27
1. Uni-Modal Data	27
2. Eigenvalue analysis and the stabilization parameter $\lambda_2$	29
Chapter 5. Multimodal Optical Flow	37
1. Estimation of the resolution parameter $\sigma$	37



This reads more like the abstract to the thesis than the first paragraphs of the background chapter.

## CHAPTER 1

### Background

Presumably you will have an "Introduction" chapter coming before the background?

Carbon-Fiber reinforced polymer (CFRP) materials are becoming increasingly widespread in automotive and aerospace industries, but also in consumer goods, due to their adaptivity to different shapes, good rigidity and high strength-to-weight ratio. Improved fabrication techniques are reducing the production costs and time to manufacture. The properties of CFRP strongly depend on the processing of the material, thus defect detection and characterization are indispensable, especially for safety-relevant Active thermal measurement methods have become vital for the assessment of the quality of CFRP materials. These methods are based on the evaluation of a previously excited heat flow in the tested component and its disturbance by hidden defects, illustrated in Figure 1. The heat flow is generated with a heat pulse or through sinusoidal modulation, observed with a thermography camera, followed by a pixel-wise computation of the complex phase between the excitation signal and the reflected infrared signal. This phase information encodes the heat-loss within a penetration depth  $\delta$  of the probed material, with depths of 1mm to 2mm typical for CFRPs.

Current state-of-the-art thermography imagers possess resolutions of only  $640 \times 512$  pixels and a noise equivalent temperature difference of 20mK. Nevertheless, these cameras are very expensive, and the CFRP application domain requires the detection of defects at the noise limit. Figure 1b shows the amplitude part of the recorded thermographic image of a CFRP probe and figure 1c the phase part. The noise is clearly observable. On the other hand, cameras in the visible spectrum are inexpensive and easily deliver images of 10 mega-pixels per frame with very little noise (see figure 1d). Therefore our strategy is to enhance the resolution of thermography images by utilizing high spatial frequency information from a visible spectrum camera in an image fusion model which takes the disparity between the imagers into account without requiring stereo calibration.

#### 1. Optical Flow

The core topic of this thesis is optical flow. Optical Flow labels the task of densely measuring the motion between two or more frames captured by a camera, or the dense registration of two or more cameras on a pixel-by-pixel basis. Optical flow is a crucial step in many areas of computer vision. For instance optical flow estimation is a part of video compression (citation!!) used to detect areas of small brightness change, thus low motion. The pixels of such areas of a video are then grouped together and stored in an memory efficient manner. In recent years structure from stereography and structure from motion (video from a single camera) have gained popularity as a means to capture 3D models for film productions and also due to the availability of low cost 3D printing (citation!!). In both the stereography and the structure from motion pipelines optical flow is used for the triangulation of the dense point cloud, prior to generation of the final 3D mesh. In the case of optical flow between two cameras both cameras, also called modalities may be of different types. For instance in medical imaging multi-modal dense image registration is used to fuse image information from CT and MR modalities of the human brain [?] and of the human spine?? ?

Not clear what you're trying to say here.

Grammar / sentence structure problems.

Let the optical flow between the individual images be denoted by the *vector-field*  $d(x)$ . A typical optical flow model consists of a term which relates  $d(x)$  to the data, called

It feels odd having this figure here. It describes a lot of things which are unrelated to optical flow.

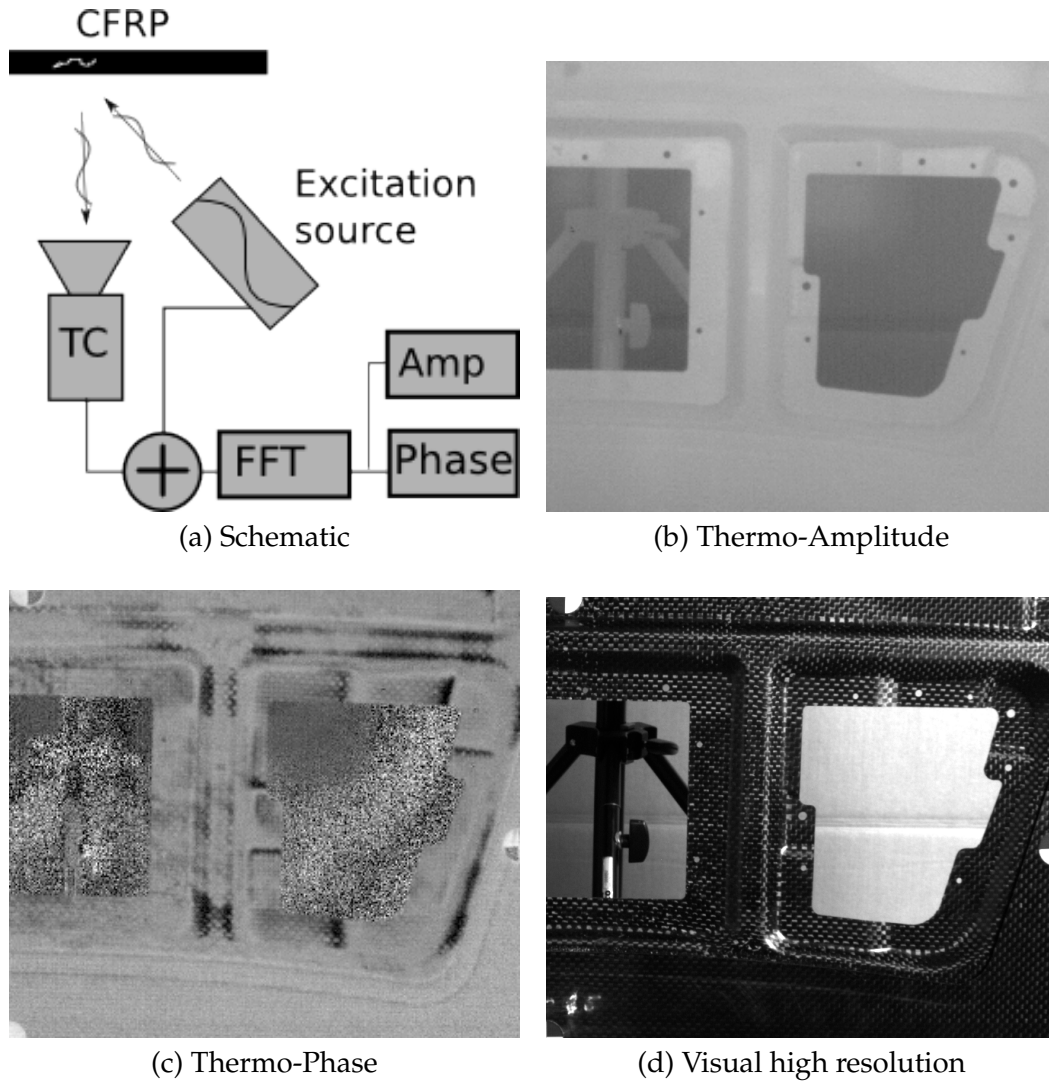


FIGURE 1. A sinusoidal excitation source emits a thermal wave onto a CFRP target. The target may contain a defect not visible on the surface, but which can be detected in the infrared (b,c). However a visual image, (d), can be acquired at a much higher resolution.

the likelihood. The other constituent is a term encoding the geometric properties of the vector field  $\mathbf{d}(\mathbf{x})$ , called the prior. The goal of this thesis is to introduce a likelihood which is suitable for the dense image registration of the thermo-graphic image in figure 1b with the higher resolved visual spectrum image in figure 1d. The other main contribution is the introduction of a new prior which borrows concepts from differential geometry to impose assumptions the geometry of the optical flow  $\mathbf{d}(\mathbf{x})$  to be estimated.

In the following we will give a short survey on the current types optical flow likelihoods and current state of the art priors. We will then introduce Lie algebras and the Noether Theorem which will play a vital role the definition of our geometrical prior.

Among the earliest methods for optical flow estimation are the methods described in the seminal papers of Horn and Schunck [?] and Lukas and Kanade [?]. In [?] the following model for computing the flow between two frames of a video was proposed

$$(1) \quad E(\mathbf{d}) = \int_{\Omega} (y(\mathbf{x}) - I(\mathbf{x} + \mathbf{d}(\mathbf{x})))^2 dx + \lambda \int_{\Omega} \sum_i \|\nabla d_i\|^2 dx$$

Completely out of context, the reader won't understand. Develop this problem more thoroughly later on.

No -- you haven't established the connection between optical flow and figure 1, so right now this doesn't make any sense.

In eq. (1) the frame  $I$  is warped back to the frame  $y$  by the field  $\mathbf{d}(\mathbf{x})$ . The second integral in eq. (1) imposes an isotropic smoothness constraint on the flow field  $\mathbf{d}$ . The likelihood in eq. (1) makes the assumption that the brightness of the scene recorded by the camera is constant from frame to frame. This is a very strong constraint, which is rarely met in real world multi-modal setups. For instance the intensities in figure 1b follow a completely different distribution than those in figure 1d. Thus we need a model that can bring both images onto a common intensity space. Furthermore the isotropic smoothness term in eq. (1) does not allow for discontinuities in  $\mathbf{d}$ . Several methods have been introduced which remove the assumption of isotropic flow [?, ?]. These Methods include (citation!!) TV-Regularization, anisotropic diffusion guided by directional operators like the structure tensor and level set methods of the Mumford-Shah type [?]. We will introduce a methodology for the geometrical characterization of anisotropic priors in section 4 following a review of the TV-Regularization prior in section 5. **Use notation!**

We will now discuss three statistical similarity measures (citation!!) for optical flow which avoid the assumption of brightness constancy. For this we will take the two images  $y$  and  $I$  to be random variables with the marginal distributions  $p(y)$  and  $p(I)$ . Then the mean and the variance are defined as

$$(2) \quad \mathbb{E}(X) = \int X \cdot p(X)$$

$$(3) \quad \text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

I find it awkward ending a section with an equation -- the discussion just seems to end, abruptly.

**1.1. Mutual Information.** Mutual Information (MI) is a popular similarity measure used mainly in medical imaging where images from different modalities including MR, CT and PET are registered against each other. For images  $y$  and  $I$  from two different modalities capturing the same scene, MI is defined with the joint distribution  $p(y, I)$  by

$$(4) \quad MI(y, I) = \int p(y, I) \ln \frac{p(y, I)}{p(y) \cdot p(I)} dy dI$$

MI measures how strong the images  $y$  and  $I$  statistically depend on each other. In the case that  $y$  and  $I$  are statistically independent,  $p(y, I) = p(y) \cdot p(I)$ , then by eq. (4) MI is zero. On the other side, MI is maximal when  $y$  completely determined by  $I$  or vice versa. However, as [?] puts it, MI does not explain the kind of dependency between images  $y$  and  $I$ , its maxima are statistically but not visually meaningful, since it disregards any spacial information, which is essential for optical flow. Thus optical flow likelihoods based on MI usually tend to have many local minima rendering MI too unconstrained for optical flow.

**1.2. Correlation Ratio.** To alleviate the problems with MI, [?] **argument-ed** that a better similarity measure would be one that measures the *functional* relation between the images  $y$  and  $I$ . The base key ingredient for their proposal is that the pixel values  $I(\mathbf{x})$  and  $y(\mathbf{x})$  are assumed to be the realizations of random variables, which by abuse of notation we denote by  $\hat{I}$  and  $\hat{y}$ . Then the normalized joint histogram of the images  $I$  and  $y$  can be interpreted as the joint probability distribution  $p(\hat{y}, \hat{I})$ , and the conditional distribution

$$(5) \quad p(\hat{y} | \hat{I} = I) = \frac{p(\hat{y}, \hat{I} = I)}{p(\hat{I} = I)}$$

Where is the common notation about similarity measure that connects the discussion?

I is the image, yes? You can't take the expectation over an image. Don't you mean I-hat?

encodes the spacial functional relationship between  $y$  and  $I$ . They introduced the Correlation Ratio (CR)

$$(6) \quad \eta(I|y) = \frac{\text{Var}(\phi^*(y))}{\text{Var}(I)} \quad \phi^*(y) = \underset{\phi}{\text{argmin}} \mathbb{E}_I (I - \phi(y))^2$$

The optimal function  $\phi^*$  was shown to be the expectation value of  $\hat{I}$ , conditioned on a realization of  $\hat{y}$

$$(7) \quad \phi^*(y) = \mathbb{E}(\hat{I} | \hat{y} = y) = \int I p(I|y) dI$$

The function  $\phi(\hat{y})$  maps any realization of  $\hat{y}$  to an expectation value of  $\hat{I}$ . As  $\hat{y}$  is a random variable,  $\phi(\hat{y})$  is also at random. Its variance measures how well  $I$  is functionally explained by a realization of  $\hat{y}$ . The measure in eq. (6) is bounded between 0 and 1, 0 indicating that  $y$  and  $I$  are independent, 1 indicating a functional relationship  $I = \phi^*(y)$ . The function  $\phi^*$  although not necessarily continuous, is measurable in the  $L_2$ -sense. Thus CR is a much stronger constraint than MI and has fewer, but more meaningful minima [?].

Meaning what?

**1.3. Cross Correlation.** Cross Correlation (citation!!) is the strongest constrained similarity measure. It is basically an additional constraint to CR, namely that the functional relationship in eq. 6 must be linear. Then  $\eta$  reduces to

$$(8) \quad \eta(I|y) = \frac{\text{Cov}(y, I)}{\text{Var}(I) \cdot \text{Var}(y)} \quad I = \lambda \cdot y$$

As we will see in section ?? a measure similar to eq. (8) will be computed based on the assumption that both  $y$  and  $I$  are Gaussian. The Gaussian assumption is valid when both cameras  $y$  and  $I$  produce Gaussian noise and the joint histogram is predominantly linear. Linearity in the joint histogram occurs when the recorded scene contains materials with uniform luminosity in the frequency bands of the cameras  $y$  and  $I$ .

Is this acceptable formatting for a thesis at Heidelberg? I have never seen a thesis formatted this way.

## 2. Setup of the camera rig

The data acquisition apparatus consists of a visible spectrum camera (VSC) mounted on top of a thermography camera (TC). The resolution of the VSC is  $1226 \times 1028$  pixels while that of the TC is  $640 \times 512$  pixels, both cameras with a focal length of 25 mm. We used a sinusoidal excitation source with a frequency of 0.1 Hz, which corresponds to a penetration depth of approximately 1.3 mm in the CFRP.

## 3. Image Fusion

Our camera setup not only consists of two cameras with differing spectral responses, the TC and the VSC also differ in spatial resolution. However the likelihoods given introduced above have in common that they do not directly model the difference in resolution. In figure 2a a model of the CCD of the low resolution TC is shown overlaid with a higher resolution grid representing the VSC. The gray region in figure 2a symbolizes one pixel of the TC and it can be seen that each pixel of the TC covers a group of pixels of the VSC. Since the TC pixel has a finite surface, we need to specify how this pixel absorbs photons landing at different points in its area in order to relate the covered pixels of the VSC to it. The response of each individual pixel in the TC is called the point spread function (PSF),  $W_\sigma(x, y)$ , the vector  $(x, y)$  being the location on the surface of the TC pixel with respect to the VSC coordinate frame. Figure 2b is the result of a theoretical model of a FLIR imager similar our TC. The model, obtained by Hardie et al. [?], combines absorption properties of the CCD pixel with physical properties of the camera lens. We can see that each TC pixel has a non uniform response to incoming photons. Using this information we can

The overall flow of the chapter is somewhat difficult to follow. The writing needs to be more systematic. This is a difficult thing to mark up, since the problem is not individual words or sentences, rather just that there are many places where notation is undefined, where context switches, where the flow of ideas is discontinuous.

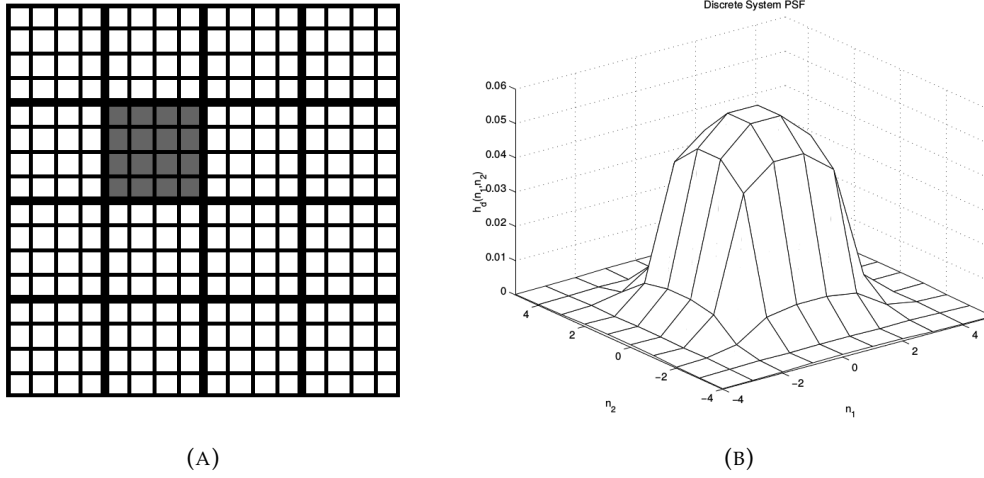


FIGURE 2. Figure 2a The thick grid depicts the CCD of the low resolution thermographic camera. The finer grid a virtual super-resolved version of the pixels in the TC. Figure 2b shows the point spread function  $W_\sigma(x, y)$  of the gray pixel in figure 2a, taken from Hardie et al. [?]. It shows that each pixel in the TC image has a non uniform response over its surface to incoming photons.

model a super-resolved version  $S$  of the TC image  $y$  with the help of the PSF  $W_\sigma$ , by stating that  $y$  is the result of the convolution of  $S$  with  $W_\sigma$

$$(9) \quad y = W_\sigma s + n \quad n \sim \mathcal{N}(0|C_n)$$

The problem of estimating  $S$  is that there is an infinite amount of high resolution TC images  $S^*$  which relate to  $y$  via eq. (9) since the high spacial frequency components of  $S$  are filtered out by  $W_\sigma$ . In [?] Hardie suggested use of a high resolution imager  $I_c$  whose camera center is co-aligned (hence the subscript  $c$ ) with the TC image  $y$  and correlated with  $S$ . The rationale behind their approach is to combine the desired features such as sharp edges and corners of  $I_c$  with the intensity spectrum of  $y$  into the super-resolved image  $S$ , while avoiding limitations such as the noise model of  $y$ . The limitation of their model is that the centers of the modalities  $y$  and  $I_c$  need to be co-linear. While this is the case in remote sensing applications, the model needs to be extended to the general case of two separated modalities. We will first outline the original model, and in chapter 3 we will introduce a new model for optical flow based on [?].

The key ingredient in the model of [?] is that the intensities of  $S$  and  $I_c$  are assumed to be samples drawn from the joint Gaussian  $p(S, I_c)$ . As  $I_c$  is already fixed as input data we can derive a conditional distribution for  $S$  via the Bayesian rule

$$(10) \quad p(S|I_c) = \frac{p(S, I_c)}{p(I_c)} \sim \mathcal{N}(\mu_{s|I_c}|C_{s|I_c})$$

$$(11) \quad C_{s|I_c} = C_{s,s} - C_{s,I_c}^2 \cdot C_{I_c,I_c}^{-1}$$

$$(12) \quad \mu_{s|I_c}(\mathbf{x}) = \mu_s + C_{s,I_c} \cdot C_{I_c,I_c}^{-1} (I_c(\mathbf{x}) - \mu_{I_c})$$

where the variances are computed globally

$$(13) \quad C_{u,v} = \int_{\Omega} (u(\mathbf{x}) - \mu_u) \cdot (v(\mathbf{x}) - \mu_v) dx$$

We see that the mean of  $S$  conditioned on  $I_c$ ,  $\mu_{s|I_c}$  is linear in the values of  $I_c$ , thus in this model the intensities of  $S$  are assumed to be *globally* linearly related to the intensities

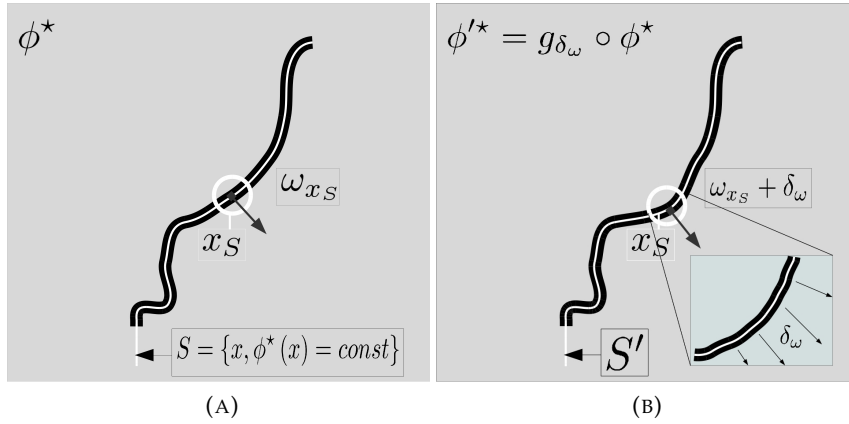


FIGURE 3. Local transformation of an image  $\phi$  with a level-set  $S$ . Figure 3a shows an image  $\phi(\mathbf{x})$  with a line  $S$  along which the intensity values are constant. At each point  $\mathbf{x}_S$  the vector  $\omega_S$  is the normal vector on  $S$ . Figure 3b shows the result of the local distortion of  $S$  under the action of the operator  $g_{\delta_\omega}$ .  $g_{\delta_\omega}$  acts on  $S$  by adding to  $\omega_S$  a spacial dependent vector  $\delta_\omega(\mathbf{x})$

of  $I_c$ . We combine eq. (10) with the Gaussian likelihood in eq. (9) to the posterior

$$(14) \quad p(S|y, I_c) \sim p(y|S) \cdot p(S|I_c) = \exp(-E(S))$$

with the associated energy

$$(15) \quad E(S) = \frac{1}{2} \int_{\Omega} \left( y(\mathbf{x}) - W_{\sigma} S(\mathbf{x}) \right)^2 \cdot C_n^{-1} dx + \frac{1}{2} \int_{\Omega} \left( S(\mathbf{x}) - \mu_{s|I_c}(\mathbf{x}) \right)^2 \cdot C_{s|I_c}^{-1} dx$$

The minimization of eq. (15) and thus maximization of (14) with respect to  $S$  gives the analytical solution [?]

$$(16) \quad \hat{s} = \mu_{s|I_c} + C_{s|I_c} \cdot W_{\sigma}^T \left( W_{\sigma} \cdot C_{s|I_c} \cdot W_{\sigma}^T + C_n \right)^{-1} \left( y - W_{\sigma} \mu_{s|I_c} \right)$$

Eq. (16) is intractable to compute due to the dense operator  $W_{\sigma}$  and the matrix-inverse operation. In [?] a computationally tractable approximation was introduced

$$\begin{aligned} \hat{s} &= \mu_{s|I_c} + C_{\tilde{s}|I_c} \cdot \left( C_{\tilde{s}|I_c} + C_n \right)^{-1} \left( y - \tilde{\mu}_{s|I_c} \right) \\ \tilde{I}_c &= W_{\sigma} I_c \quad \tilde{s} = W_{\sigma} s \approx y \end{aligned}$$

The key issue is that this method requires both modalities,  $I_c$  and  $y$ , to be co-registered. Since we are dealing with an optical flow problem  $y$  and thus  $S$  is shifted by a disparity  $\mathbf{d}(\mathbf{x})$  from  $I_c$ . This disparity has to be taken in to account by our model in chapter 3. The second issue is that the assumption that  $S$  and  $I_c$  are *globally* joint Gaussian is not supported by our data. However by computing  $C_{s|I_c}$  in *local* sub-domains of the space  $\Omega$  we can show that  $S$  and  $I_c$  are *locally* joint Gaussian. This will also be shown in chapter 3.

## 4. Lie Groups and the Noether Theorem

**4.1. Motivation.** In section 1 we had claimed that one of the problems in the model of Horn and Schunck [?] (see eq. (??)) is that the  $L_2$  prior

$$(17) \quad p_{L_2}(\nabla \mathbf{d}) = \exp \left( - \sum_i E_{L_2}(d_i) \right) \quad E_{L_2}(\phi) = \int_{\Omega} \|\nabla \phi\|^2$$



penalizes non-constant flow fields  $\mathbf{d}$ . This poses a problem since objects moving against each other generate a flow field  $\mathbf{d}$  which must be discontinuous at the object boundaries. In general the minimizers  $\phi^*$  of the energy  $E_{L_2}$  are the constant functions  $\phi = \text{const}$ . A slightly different description of the set of minimizers goes as follows: given  $u_0^*(\mathbf{x})^2 = c_0$  we can generate all other possible minimizers  $\phi_c^*$  by adding any  $c \in \mathbb{R}$  to  $u_0^*(\mathbf{x})$

$$(18) \quad A_c = \{\phi_c^*(\mathbf{x}) \mid \phi_c^* = \phi_0^* + c, \quad c \in \mathbb{R}\}$$

We can label the action of adding a real number  $c \in \mathbb{R}$  on to any function  $\phi(\mathbf{x})$  by  $g_c$ . Under the group of such actions  $\mathbb{G}_{\text{const}} = \{g_c\}$  our set of minimizers  $A_c$  is invariant

$$(19) \quad g_d \circ A_c = A_{c+d} = \{\phi_c^*(\mathbf{x}) \mid \phi_{c+d}^* = \phi_0^* + c + d, \quad c \in \mathbb{R}\} = A_c$$

as well as the  $L_2$  energy

$$(20) \quad g_d \circ E_{L_2}(\phi) = \int_{\Omega} \|\nabla(\phi + d)\|^2 = \int_{\Omega} \|\nabla\phi\|^2$$

In this sense we can state that the  $L_2$  prior  $p_{L_2}(\nabla\phi)$  is actually conditioned on the group of constant transformations

$$(21) \quad p_{L_2}(\nabla\phi) = p_{L_2}(\nabla\phi \mid \mathbb{G}_{\text{const}})$$

since it is invariant under the entire set  $\mathbb{G}_{\text{const}}$  but under no other set, this is why we call  $p_{L_2}$  *conditionally invariant*. The notion of conditional invariance serves as starting point for the following discussion. The idea is to introduce a methodology for designing  $p(\nabla\phi)$  whose set of maxima  $A$  respects certain geometric constraints related to a group of smooth transitions  $\mathbb{G}$  whose elements  $g(\mathbf{x}) \in \mathbb{G}$  are non-constant functions over  $\Omega$ . This methodology will us deduce explicit constraints on  $p(\nabla\phi)$  from expected geometric properties of  $A$ . The starting point is the idea that  $A$  can be seen as being generated by the group  $\mathbb{G}$

$$(22) \quad A = \{\phi^* \mid \phi^* = g \circ \phi_0^* \quad g \in \mathbb{G}\}$$

for some maximum  $\phi_0^*$ . In figure 3a one possible maximum  $\phi^* \in A$  is shown which contains a line  $S$  along which the intensities are constant

$$(23) \quad S = \{\mathbf{x} \mid \phi^*(\mathbf{x}) = 0, \mathbf{x} \in \Omega\}$$

If under a deformation of  $S$ , caused by the action of any operator  $g_{\delta\omega} \in \mathbb{G}$  (see figure 3b) the prior  $p(\nabla\phi)$  is invariant, it is clear the entire solution space  $A$  maximizes  $p(\nabla\phi)$ . Hence  $p(\nabla\phi)$  is conditioned on  $\mathbb{G}$

$$(24) \quad p(\nabla\phi) = p(\nabla\phi \mid \mathbb{G})$$

The power of this argument is that we can define the group  $\mathbb{G}$ , that is the specific types of geometrical transformations in order to specify the set of maxima  $A$  which should not be penalized by the prior  $p(\nabla\phi)$ . As we will see it is easier to specify  $\mathbb{G}$  then it is to directly specify  $A$  since the dimension of  $A$  is infinite, while  $\mathbb{G}$  can be reduced to a finite dimensional space called the Lie Algebra  $\mathcal{G}$ . This will lead us to a differential equation on  $p(\nabla\phi)$ .

**4.2. Lie Groups.** Lie groups are groups of objects which are smooth functions on a manifold at the same time being compatible with the group multiplication. For an  $n$  dimensional Lie group  $\mathbb{G}$  over a domain  $\Omega$  the map

$$\mathbb{G} \times \mathbb{G} \mapsto \mathbb{G} : (x, y) \rightarrow x \cdot y^{-1}$$

is smooth in both  $x$  and  $y$ . One of the remarkable properties of Lie group theory is that for every Lie group  $\mathbb{G}$  there exists a unique structure called a Lie algebra  $\mathcal{G}$ . The Lie algebra

is associated with the tangent space of  $\mathbb{G}$ ,  $\mathbb{T}\mathbb{G}$ . We introduce the Lie algebra by defining a one parameter Lie group  $\gamma(t)$

$$(25) \quad \gamma : \mathbb{R} \rightarrow \mathbb{G}_\gamma \in \mathbb{G}$$

$$(26) \quad \gamma(0) = e$$

$$(27) \quad \left. \frac{d}{dt} \gamma(t) \right|_{t=0} = X \in \mathcal{G}$$

The path  $\gamma$  is not unique, there are infinit many paths which have the same tangential vector  $X$  at  $t = 0$ . This defines an equivalence relation: the paths  $\gamma_1$  and  $\gamma_2$  are equivalent,  $\gamma_1 \sim \gamma_2$  if

$$(28) \quad \left. \frac{d}{dt} \gamma_1(t) \right|_{t=0} = \left. \frac{d}{dt} \gamma_2(t) \right|_{t=0}$$

By Eq. eq. (28) we can group the paths  $\gamma$  into equivalence classes  $[\gamma]$  For the vector  $X$  in eq. (27) we can find a descriptive formulation in form of a vector field  $\omega(\mathbf{z}_0)$ ,  $\mathbf{z}_0 \in \mathcal{M}$  where the space  $\mathcal{M}$  is a smooth manifold upon which  $\mathbb{G}$  acts. The vector field  $X$  induces a path  $\Gamma^X(\mathbf{z}_0, s)$  with the properties

$$(29) \quad \frac{d}{dt} \Gamma^X(\mathbf{z}_0, t) = X(\Gamma^X(\mathbf{z}_0, t))$$

$$(30) \quad \Gamma^X(\mathbf{z}_0, t)|_{t=0} = \mathbf{z}_0;$$

The coordinates of  $X$  relative to the space  $\mathcal{M}$  can be computed when we look at the space of smooth functions with support on  $\mathcal{M}$ ,  $\mathcal{F}(\mathcal{M})$ . The action of  $X$  on  $\mathcal{F}(\mathcal{M})$  can be computed by evaluating  $F \in \mathcal{F}(\mathcal{M})$  on the integral curve  $\Gamma^X(\mathbf{z}_0, s)$  and taking the derivative

$$(31) \quad \left. \frac{d}{ds} F(\Gamma^X(\mathbf{z}_0, s)) \right|_{s=0} = X F(\Gamma^X(\mathbf{z}_0, s))|_{t=0}$$

If  $F$  is one of the coordinate functions  $z_i$  then the vector field  $X$  has the coefficients  $\omega_i(\mathbf{z}_0)$  relative to the basis in  $\mathcal{M}$

$$(32) \quad X \cdot z_i|_{\mathbf{z}=\mathbf{z}_0} = \omega_i(\mathbf{z}_0)$$

The set of differential operators  $\{\partial_{z_i}\}$  can be interpreted as a basis of the Lie algebra and the vector field  $\omega(\mathbf{z}_0)$  is the coordinate representation of  $X$  at the point  $\mathbf{z}_0 \in \mathcal{M}$ . We would like build upon the discussion about the one dimensional Lie Group  $\mathbb{G}_\gamma$  and generalize it higher dimensionnal groups. An  $m$ -dimensional Lie Group  $\mathbb{G}$  is a set of elements which are parameterized by  $m$  parameters

$$(33) \quad g_{a_1 \dots a_m} \in \mathbb{G}$$

The parameters  $a_l$  define axis in the set  $\mathbb{G}$  which themselves are one parameter subgroups. For each parameter  $a_l$  there exists by extension of eq. (27) a vector field  $X_l$

$$(34) \quad \left. \frac{d}{da_l} g_{a_1 \dots a_m} \right|_{a_1 \dots a_m=0} = X_l$$

Just like in eq. (31) the vectorfields  $X_l$  each have a coordinate representation relative to the space  $\mathcal{M}$

$$(35) \quad X_l = \sum_i \xi_i^l(\mathbf{z}_0) \cdot \partial_{z_i}$$

The vector fields  $X_l$  constitute a basis for the Lie Algebra  $\mathcal{G}$ . All one parameter subgroups  $\gamma$  (see eq. (27)) can be represented as paths in the parameterspace of  $\mathbb{G}$

$$(36) \quad \gamma(t) = g_{a_1(t) \dots a_m(t)}$$

and the vectorfield  $X$  from eq. (27) is computed from the derivatives of the parameters

$$(37) \quad \left. \frac{d}{dt} \gamma(t) \right|_{t=0} = \sum_l \frac{d}{dt} a_l \cdot \left. \frac{d}{da_l} g_{a_1 \dots a_m} \right|_{t=0} = \sum_l \alpha_l X_l$$

When we combine eq. (37) and eq. (35) we get a coordinate expression for the coefficient vectorfield  $\omega$  in the basis of the Lie Algebra

$$(38) \quad \omega^i(\mathbf{z}) = \sum_l \alpha_l \cdot \xi^i(\mathbf{z})$$

The Lie algebra is connected to the infinitesimal Lie group  $\mathbb{U}_r \subset \mathbb{G}$  via

$$(39) \quad g_{\mathbb{U}} = \mathbb{1} + X \in \mathbb{U}_r$$

For the rest of this work we will focus on the space  $\mathcal{M} = \Omega \times J^k(\mathcal{C}^\infty(\Omega))$ . The space  $\Omega \subset \mathbb{R}^n$  is an open subset and  $J^k(\mathcal{C}^\infty(\Omega))$  is the set of smooth differentiable functions with compact support in  $\Omega$  and their derivatives up to order  $k$ . The points  $\mathbf{z} \in \mathcal{M}$  are vectors of the *independent* variables  $\mathbf{x}$ , the *dependent* variable  $\phi(\mathbf{x})$  and its derivatives  $\phi_{,K}$  where  $K$  is a multiindex

$$(40) \quad \mathbf{z} = (\mathbf{x}, \phi(\mathbf{x}), \phi_{,K}(\mathbf{x}))$$

For this work we will focus only on first order derivatives,  $k = 1$  The vector field  $X$  then has the form

$$(41) \quad X = \omega_\phi(\mathbf{z}_0) \partial_\phi + \sum_{i=1}^n \omega_i(\mathbf{z}_0) \partial_{x_i}$$

Under the action of  $g_{\mathbb{U}}$  points in  $\mathcal{C}^\infty(\Omega) \times \Omega$  transform like

$$(42) \quad \begin{pmatrix} \mathbf{x}' \\ \phi'(\mathbf{x}') \\ \phi'_{,j}(\mathbf{x}') \end{pmatrix} = \begin{pmatrix} \mathbf{x} + \omega(\mathbf{x}) \\ \phi(\mathbf{x}) + \omega_\phi(\mathbf{x}) + \sum_i \omega_i(\mathbf{x}) \partial_{x_i} \phi(\mathbf{x}) \\ \phi_{,j}(\mathbf{x}) + \omega_{\phi,j}(\mathbf{x}) - \sum_i \partial_{x_i} \phi \partial_{x_i} \omega_j(\mathbf{x}) \end{pmatrix}$$

### 4.3. Noether's Theorem.

4.3.1. *Motivation.* In this section we will focus on geometric properties of the prior in our model. We will assume the prior  $P(\nabla\phi)$  which is only depends first order gradients of the field  $\phi(\mathbf{x})$ . In general, priors have infinit sets of maxima  $\phi^*$ . For instance the maximas of the prior  $P_{L_2}(\nabla u) = -\ln \|\nabla u\|_{L_2}$  in eq. (65) form the set of constant fields in the domain  $\Omega$

$$(43) \quad A = \{u^* | u^*(\mathbf{x}) = \text{const } \mathbf{x} \in \Omega\}$$

It is trivial to see that there exists a one parameter Lie group of elements  $g_c$  which take  $A$  into its self

$$(44) \quad g_c : A \rightarrow A, \quad u^* \rightarrow u^* + c, \quad c = \text{const in } \Omega$$

and that  $P_{L_2}(\nabla u)$  is invariant under the action of  $\{g_c\}$ . Since  $c$  is constant in  $\Omega$ ,  $\{g_c\}$  is the only Lie group under which  $P_{L_2}(\nabla u)$  is invariant. Knowledge of the solution set  $A$  allows us to parameterize the solutions of the complete model in eq. (65) by

$$(45) \quad u(\mathbf{x}) = m + h(\mathbf{x}) \quad m = \|u^0\|_{L_2}$$

It is known that the global mean value of the solution  $u^*$  in the model in eq. (65) is equal to that of the data  $u^0$  Thus the parameterization allows any gradient algorithm to converge faster to the solution  $u^*(\mathbf{x})$  given the initial guess

$$(46) \quad u_{init}(\mathbf{x}) = m, \quad m = \|u^0\|_{L_2}$$

The solution space  $A$  of  $P_{L_2}(\nabla\phi)$  is too trivial for most real applications, since  $P_{L_2}(\nabla\phi)$  penalizes any other solution  $u^* \notin A$  which contains structure. We want to assume that a

more general prior  $P(\nabla\phi)$  which has maxima  $\phi^*(\mathbf{x}) \neq \text{const}$ . The level-sets of each field  $\phi^*(\mathbf{x})$  are taken to be the integral curves  $\Gamma^X(s)$  corresponding to a vectorfield  $X$  of some Lie algebra,  $X \in \mathcal{G}$ . The space of the maxima of  $P(\nabla\phi)$  is then entirely determined by the Lie algebra  $\mathcal{G}$

$$(47) \quad A_{\mathbb{G}} = \left\{ \phi^* \left| \frac{d}{ds} \phi^*(\Gamma^X(s)) = 0 \forall X \in \mathcal{G} \right. \right\}$$

The action of the Lie group on elements of its own Lie algebra preserves the algebra

$$(48) \quad g \cdot X \cdot g^{-1} \in \mathcal{G} \quad \forall X \in \mathcal{G}, g \in \mathbb{G}$$

and by construction the set maxima in eq. (47) it is apparent that  $\mathbb{G}$  takes  $A_{\mathbb{G}}$  onto itself

$$(49) \quad g \cdot A_{\mathbb{G}} = A_{\mathbb{G}} \quad \forall g \in \mathbb{G}$$

Since we have characterized the prior  $P(\nabla\phi)$  by the set of its maxima in eq. (47) which is explained by the algebra of the group  $\mathbb{G}$  we denote  $P(\nabla\phi)$  as being conditioned on  $\mathbb{G}$ ,  $P(\nabla\phi|g \in \mathbb{G})$ . But due to the invariance the maxima set in eq. (47) under the action of  $\mathbb{G}$ ,  $P(\nabla\phi)$  is invariant under  $\mathbb{G}$

$$(50) \quad P(\nabla\phi|g \in \mathbb{G}) = \text{const} \quad \text{w.r.t } g \in \mathbb{G}$$

This property we call *Conditional Invariance*. The most important aspect of the above discussion is that given geometrical assumptions on the solutions  $\phi^*$  in terms of  $\mathbb{G}$  and  $A_{\mathbb{G}}$  the condition in eq. (50) must be fulfilled, and thus it serves as a guidance in the design of the prior  $P(\nabla\phi|g \in \mathbb{G})$ .

4.3.2. *Noethers First Theorem*. We are now going to make eq. (50) more precise by considering the negative log-prior energy

$$(51) \quad I = -\ln P(\nabla\phi) = \int_{\Omega} \mathcal{E}(x, \nabla\phi) dx$$

we are interested in the action of  $\mathbb{G}$  (see eq. (42)]. The energy in eq. (51) is said to be preserved under the Lie group  $\mathbb{G}$  if the following relation holds

$$(52) \quad I' = \int_{\Omega} \mathcal{E}'(x', \nabla\phi') dx' = \int_{\Omega} \{ \mathcal{E}(x, \nabla\phi) + \partial_i \delta Q^i \} dx$$

where the vectorfield  $\delta Q^i$  is some arbitrary smooth function. If eq. (52) holds then the resulting Euler-Lagrange equations  $[I]$  remain unchanged and thus  $\mathbb{G}$  is a symmetry of the Euler-Lagrange equations. In [?, ?] it was reasoned that the knowledge of the symmetries of the Euler-lagrange equations  $[I]$  can be used to make assumptions on the form of the solutions  $\phi^*$  and thus narrow down the solution space. To be more precise, the first Noether Theorem states that if the energy integral in eq. (52) is preserved under the transformations eq. (42) then the Euler-lagrange equations must fulfill

$$(53) \quad [I] \omega_{\phi} = \partial_{\mu} (W^{\mu} - \delta Q^{\mu})$$

where

$$(54) \quad [I] = \frac{\delta I}{\delta \phi} - \frac{d}{dx^{\nu}} \frac{\delta I}{\delta \phi_{,\nu}}$$

are the Euler-Lagrange equations of  $I$  and the field  $W^{\mu}$  is defined by

$$(55) \quad W^{\mu} = -\frac{\delta I}{\delta \phi_{,\mu}} \omega_{\phi} + \omega_i \left( \frac{\delta I}{\delta \phi_{,\mu}} \phi_{,i} - \delta^{\mu,i} I \right)$$

When eq. (53) is evaluated at the solution  $\phi^*$  of the Euler-Lagrange equation  $[I] = 0$  then  $W^{\mu}$  must be divergence-free. The form of the divergence free vector field  $W^{\mu}$  dictates the

form of the geometry of the level-sets of  $\phi^*$ . We will now show an example where knowledge of the symmetry and thus the divergence-free  $W^\mu$  fields makes basic assumptions on the solution space of the corresponding Euler-Lagrange equations possible.

4.3.3. *Kepler's Two Body Problem.* Keplers two body problem is the problem of calculating the problem of estimating the trajectory of a body of mass  $m_e$  (the earth) which is moving within the vicinity of another body with mass  $m_s$  (the sun). According to Newton there exists a gravitational force between the masses coming from the energy  $V(r)$  of the gravitational field surrounding the mass  $m_s$  at the origin in  $\mathbb{R}^3$

$$(56) \quad V(\mathbf{r}_e(t)) = -\frac{m_e \cdot m_s}{r} \quad r = \|\mathbf{r}_e - \mathbf{r}_s\|$$

The kinetic energy of the mass  $m_e$  is  $\frac{1}{2}m_e\dot{r}^2$  so that the Lagrangian of the path  $\mathbf{r}_e(t)$  is

$$(57) \quad L(\mathbf{r}_e(t)) = \frac{1}{2}m_e\dot{r}_e^2 + \frac{1}{2}m_e\dot{r}_s^2 - V(\mathbf{r}_e(t))$$

The Euler-Lagrange equations are easily computed

$$(58) \quad \ddot{r}_e + \frac{m_s + m_e}{r^2} = 0$$

The parameter  $t$  is the time parameter of the two body system. The Kepler Lagrangian in eq. (57) exhibits a symmetry under four different one parameter Lie group actions, namely the action of time shift and rotations around the three spacial axis (the group  $SO(3) \times \mathbb{R}$ )

$$(59) \quad t' = t + \delta t$$

$$(60) \quad \mathbf{r}' = \mathbf{r} + \partial_{\theta_i} \mathbf{r}' \delta \theta_i \quad i = x, y, \text{ or } z$$

where  $\theta_i$  are rotation around the  $x$ -,  $y$ - or  $z$ -axis. From Noethers theorem there exist four corresponding conserved quantities:

$$(61) \quad \mathcal{H} = \frac{1}{2}m_e\dot{r}^2 + V(\mathbf{r}_e(t)) \quad \text{time shift}$$

$$(62) \quad l_x = z\dot{y} - y\dot{z} \quad \text{Rotation around } x\text{-axis}$$

$$(63) \quad l_y = z\dot{x} - x\dot{z} \quad \text{Rotation around } y\text{-axis}$$

$$(64) \quad l_z = x\dot{y} - y\dot{x} \quad \text{Rotation around } z\text{-axis}$$

The conserved quantity  $\mathcal{H}$  in eq. (61) is the *Hamiltonian Energy* of the two body system. It constant time and thus manifests that the total energy of the two body system does not disipate away since there are no external forces interacting with the two masses  $m_e$  and  $m_s$ , that is the two body system is a *closed system*. The vector  $\mathbf{l} = (l_x, l_y, l_z)$  (eqs. eq. (62) to eq. (64)) is called the *angular momentum* of the masses  $m_e$  and  $m_s$  as they rotate around eachother. The solutions to the Euler-Lagrange equations in eq. (58) are elliptic curves in the plane orthorgonal to  $\mathbf{l}$ . The constancy of  $\mathbf{l}$  with respect to the special orthogonal group  $SO(3)$  comes the fact the plane is embedded in the eucleadian coordinate space with unit metric, rather some general riemanian space.

## 5. Total Variation

The earliest attempts to optimization in computer vision all had in common, the use of isotropic priors for the regularization of the unknowns to be estimated. For example one of the earliest attempts for image denoising invloves minimizing the functional ([?])

$$(65) \quad E(u) = \int (u - u^0)^2 dx + \frac{\lambda}{2} \int |\nabla u|^2 dx$$

The first term in eq. eq. (65) is the likelyhood which states the minimizer  $u^*$  must be close in its intensity distribution to the given data  $u^0$ . The second term, the prior energy

imposes smoothness on the minimizer  $u^*$ . Both terms are quadratic in  $u$  and thus the Euler-Lagrange equations for  $E(u)$  are linear in  $u$  making them computationally easy to solve. The problem with the prior  $\frac{\lambda}{2} \int |\nabla u|^2 dx$  is that it does not allow the solutions  $u^*$  to have discontinuities. Different approaches for anisotropic priors exist, for instance [?] introduced a quadratic prior

$$(66) \quad E_{prior} = \int (\nabla u)^T D (\nabla u)$$

The operator  $D$  is a local  $2 \times 2$  symmetric valued matrix with eigenvectors tangential to the level-sets of  $u^0$ . This is why  $D$  steers the direction of the gradients in eq. (66) in tangential direction of the level-sets, and thus also of the discontinuities of  $u$  and  $u^0$ . The upside is that the prior in eq. (66) combined with the likelihood in eq. (81) still lead to Euler-Lagrange equations linear in  $u$ . The downside of the prior in eq. (66) is that the operator field  $D$  must be precomputed on the data  $u^0$ , e.g. with an eigenvalue analysis of the structure tensor.

In the context of shock-filtering ([?, ?, ?]) it was shown that the functional

$$(67) \quad E_{L_1}(u) = \int |\nabla u| dx$$

has the appealing property that it does not penalize large discontinuities. However its functional derivative with respect to  $u$  is ill conditioned in the case  $\nabla u \approx 0$ . To alleviate the case, [?] chose the approximative prior

$$(68) \quad E_{L_1approx}(u) = \int \sqrt{|\nabla u|^2 + \epsilon} dx$$

which is well behaved for  $\epsilon > 0$ . They were able to achieve good results with relatively sharp preserved discontinuities with data  $u^0$  having low SNRs. Never the less in the limit  $\epsilon \rightarrow 0$  the Euler-Lagrange equations become more and more computationally instable. A theoretically more well conditioned form of TV is needed which we will outline, following ([?]). To do this we need to explore the function space the minimizers of eq. (67) might belong to. Smooth functions  $u_{smooth}$  are functions for which  $\nabla u$  exists everywhere, thus they may be minimizers of eq. (67). But functions  $u_{discont}$  containing discontinuities do not have finite  $L_1$  norm of their gradients,  $E_{L_1}(u_{discont}) = \infty$  since the gradient  $\nabla u_{discont}$  does not exist at the discontinuities. A generalization of eq. (67) is possible if one assumes  $\nabla u$  to be a distribution, more precisely a radon measure in the space  $\mathcal{M}(\Omega)$ . If there exists a radon measure  $\mu \in \mathcal{M}(\Omega)$ , such that for every  $\phi \in \mathcal{C}_0(\Omega)$  with compact domain, the following equality holds

$$(69) \quad \int_{\Omega} u \cdot \text{Div} \phi dx = - \int \phi d\mu < \infty$$

then  $\mu$  is called the weak derivative of  $u$  and we can identify  $\nabla u = \mu$ . It is then possible to define the function space of bounded variation

$$(70) \quad BV = \{u \in L_1(\Omega) \mid \nabla u \in \mathcal{M}(\Omega)\}$$

Now it is possible to define a norm on  $BV$ . By virtue of the Hölder relation there exists a scalar  $C$  for which we can determine the upper bound of eq. (69)

$$(71) \quad \int_{\Omega} u \cdot \text{Div} \phi dx \leq C \|\phi\|_{\infty}$$

The scalar  $C$  is the norm of the radon measure  $\nabla u$  and is called the total variation of  $u$

$$(72) \quad TV(u) = \sup \left\{ \int_{\Omega} u \cdot \text{Div} \phi dx \mid \|\phi\|_{\infty} \leq 1 \right\}$$

As was discussed in [?] the functions  $u$  are geometrically piecewise smooth, meaning there exists a partitioning  $\{\Omega_k\}$  of  $\Omega$  such that  $(\nabla u)_{\Omega_k}$  are  $L_1$  integrable. If  $dl_{mk}$  is a line segment in the intersection  $\Omega_m \cap \Omega_k$  then  $TV(u)$  can be written in the form

$$(73) \quad TV(u) = \sum_k \|\nabla u_{\Omega_k}\|_{L_1} + \sum_{k < m} \int_{\Omega_k \cap \Omega_m} |u_k - u_m| dl_{km}$$

where  $u_k$  the value of  $u$  on the portion of  $\partial\Omega_k$  which is interfacing with  $\Omega_m$  and vice versa for  $u_m$ . The first term in eq. (73) penalizes the smooth parts of  $u$  (the gradients  $(\nabla u)_{\Omega_k}$ ) while the second term penalizes the length of the section  $\Omega_m \cap \Omega_k$  while maintaining the values  $u_{k,m}$  and thus the *jump*  $|u_k - u_m|$ . It essentially penalizes the curvature of the line interfacing with both  $\Omega_k$  and  $\Omega_m$ . We will make this point clear in the following section.

**5.1. The Mean Curvature of Total Variation.** In this section we will discuss the geometrical properties of the TV norm in eq. (72). The subgradient of eq. (72) is equal to the set

$$(74) \quad \partial TV(u) = \left\{ -\text{Div}\sigma \mid \sigma \cdot \nu = 0 \text{ on } \partial\Omega, \sigma = \frac{\nabla u}{|\nabla u|} \text{ if } |\nabla u| \neq 0 \right\}$$

This set defines the set of lines  $L(v) = TV(u) + \langle \text{Div}\sigma | v - u \rangle$  tangential to  $TV$  at a point  $u \in BV$ . We define a one parameter Lie group  $\gamma(t)$ , such that its vectorfield  $X$  fullfills the condition

$$(75) \quad X \cdot u(\Gamma^X(\mathbf{x}_0, t)) = 0$$

then its integral curves  $\Gamma^X(t) = (x(t), y(t))$  are the level sets of  $u$ . The level sets  $\Gamma^X$  have a curvature  $\kappa$  and the standard formula for  $\kappa$  is

$$(76) \quad \kappa = \frac{1}{\|\dot{\Gamma}^X\|_{L_1}^3} (\dot{x} \cdot \ddot{y} - \dot{y} \cdot \ddot{x})$$

If the vector field  $X$  is expressed by the coordinate vector  $\xi(\mathbf{x}_0)$  then it can be shown  $\kappa$  is a function of the laplacian relative to the coordinate vector  $\xi(\mathbf{x}_0)$ .

$$(77) \quad \kappa(\mathbf{x}_0) = \frac{\Delta_{\xi\xi} u(\mathbf{x}_0)}{|\nabla u(\mathbf{x}_0)|}$$

This form can easily be transformed into a divergence quantity

$$(78) \quad \kappa = \text{Div} \left( \frac{\nabla u}{|\nabla u|} \right)$$

This shows us that the subgradient in eq: eq. (74) is equal to the curvature of the levelsets  $\Gamma^X(t)$

$$(79) \quad \kappa = -\partial TV(u)$$

The eq. eq. (79) exposes the capital geometrical property of the TV norm: The TV norm penalizes the curvature of the level-sets of an image. As  $\kappa$  is an invariant of the Lie group  $SE(2)$ , the group of rotations and translations,  $TV$  is also an invariant of that group.

**5.2. Image Denoising.** Image denoising is the problem of estimating a *clean* image  $u^*$  given a noisy image  $u^0$ . The image  $u^0$  is connected to  $u^*$  via

$$(80) \quad u^0 = u^* + n \quad n \sim \mathcal{D}$$

where  $\mathcal{D}$  is some distribution and  $n$  is a noise term drawn from  $\mathcal{D}$ .  $u^*$  is estimated from the family of functionals

$$(81) \quad F(u) = \frac{1}{q} \int_{\Omega} |u - u^0|^q dx + \lambda TV(u)$$

The degree  $q$  of the data term must be matched to the form of the distribution  $\mathcal{D}$ . Using the subgradient in eq. (74) the Euler-Lagrange equations can be calculated

$$(82) \quad [F](u) = \begin{cases} |u^* - u^0|^{q-2} (u^* - u^0) - \lambda \text{Div} \phi & \text{in } \Omega \\ \phi \cdot \nu = 0 & \text{on } \partial\Omega \\ \phi = \frac{\nabla u}{|\nabla u|} & |\nabla u| \neq 0 \text{ - nearly everywhere} \end{cases}$$

As was discussed before the term  $\text{Div} \phi$  is equal to the curvature  $\kappa$  on the level-set of  $u$ . Thus the parameter  $\lambda$  controls how strong the curvature of the level-sets are penalized. Never the less the functional eq. (81) still permits jumps in the image  $u^*$ .



## CHAPTER 2

### Geometrical Prior

We shall now proceed to introduce a prior based on the considerations made in chapter 4. For this we review the structure tensor. Consider an image  $\phi(\mathbf{x})$ . We would like to characterize the dominant strength and the orientation of  $\nabla\phi$  within a subdomain  $A \subset \Omega$ . In [?] it was suggested that the orientation vector  $n$  of the level sets in  $\phi_A$  are constrained by

$$(83) \quad \mathbf{n}^T \cdot \nabla\phi(\mathbf{x})|_{\mathbf{x} \in A} = 0$$

The vector  $\mathbf{n}$  is assumed to be constant within the domain  $A$ . It can be computed by minimizing the energy

$$(84) \quad J(\mathbf{n}) = \frac{1}{2} \int_A w(\mathbf{x}) \mathbf{n}^T \cdot (\nabla\phi(\mathbf{x}) \nabla^T\phi(\mathbf{x})) \mathbf{n} = \frac{1}{2} \mathbf{n}^T S \mathbf{n}$$

The matrix  $S$  is called the structure tensor. Since  $S$  is a symmetric matrix there exists an orthogonal decomposition

$$(85) \quad S = V^T D V \quad D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \quad V = (\mathbf{v}_1, \mathbf{v}_2)$$

The eigenvalues give of the squared strength of the gradient in the basis defined by the columns of  $V$ . They characterize the structure in  $A$  in the following way

- $\lambda_1 > \lambda_2$ : Strong linear level set with normal vector  $\mathbf{v}_1$
- $\lambda_1 \approx \lambda_2 \approx 0$ : No strong gradient, image is approximately constant
- $\lambda_1 \approx \lambda_2 \gg 0$ : No *linear* level sets, level sets have strong curvature

The constraint in eq. (83) can be explained with the method in Chapter 4.2 as the action of the translational group  $\mathbb{T}$  and the associated algebra  $\mathcal{T}$ . The basis of  $\mathcal{T}$  are the cartesian derivative operators  $\partial_x$  and  $\partial_y$  and the elements of  $\mathcal{T}$  have the form

$$(86) \quad X = \omega_x \partial_x + \omega_y \partial_y \in \mathcal{T}$$

The integral curves of  $X$ ,  $\Gamma^X(s)$  (see eq. (29)) are the straight lines

$$(87) \quad \Gamma^X(s) = \mathbf{x}_0 + \omega s$$

If  $s$  is the arc length of  $\Gamma^X$  then the coefficient vector  $\omega$  is the normalized vector  $n$  and the constraint eq. (??) is written in terms of the vector field  $X$

$$(88) \quad \left. \frac{d}{ds} \phi(\Gamma^X(s)) \right|_{s=0} = X\phi(\mathbf{x}_0) = 0$$

Since the basis operators  $\partial_x$  and  $\partial_y$  commute,  $\{\partial_x, \partial_y\} = 0$  the vector field  $X$  is translation invariant and as a consequence of eq. (88), the structure tensor  $S$  is also translation invariant. Under the rotation group  $SO(2)$  the structure tensor is *not* invariant. Nonetheless it has an important transformation property: the transformed structure tensor  $S'$  may be written in terms of the old matrix  $S$  and the rotation matrix  $R_\theta \in SO(2)$

$$(89) \quad S' = R_\theta^T S R_\theta$$

## 1. The Generalized Structure Tensor

In [?] a generalisation of the structure tensor was introduced. The generalisation is based on the introduction of the canonical coordinates  $\xi(\mathbf{x})$  and  $\eta(\mathbf{x})$  which pose a deformation of the cartesian coordinate space  $\Omega$ . The prime example is the transformation from cartesian to polar coordinates  $(x, y) \rightarrow (r, \theta)$ . The gradient with respect the new coordinates can be expressed with the cartesian coordinates via the jacobian matrix  $J$

$$(90) \quad \begin{pmatrix} \partial_\xi \\ \partial_\eta \end{pmatrix} = J^{-1} \cdot \begin{pmatrix} \partial_x \\ \partial_y \end{pmatrix} \quad J = \begin{pmatrix} \xi_x & \eta_x \\ \xi_y & \eta_y \end{pmatrix}$$

The differential operators  $\{\partial_\xi, \partial_\eta\}$  also form the basis for the algebra  $\mathcal{H}$  of the general Lie group  $\mathbb{H}$ , that is  $[\partial_\xi, \partial_\eta] = 0$  if and only if the following conditions hold

$$(91) \quad \partial_x \xi = -\partial_y \eta$$

$$(92) \quad \partial_y \xi = \partial_x \eta$$

The eqs. eq. (92) are the famous Cauchy-Riemann differential equations and their combination give the separate wave equations

$$(93) \quad \Delta \xi = 0$$

$$(94) \quad \Delta \eta = 0$$

A solution of eq. (93) implies that there must also exist a solution for eq. (94). This is why one calls the pair  $\{\xi, \eta\}$  a pair of *conjugate functions*. Within the coordinate frame  $(\xi, \eta)$  the operators  $\{\partial_\xi, \partial_\eta\}$  obey the natural conditions

$$(95) \quad \begin{aligned} \partial_\xi \xi &= 1, & \partial_\xi \eta &= 0 \\ \partial_\eta \xi &= 0, & \partial_\eta \eta &= 1 \end{aligned}$$

The integral curves  $\Gamma^X$  generated by the operators  $\{\partial_\xi, \partial_\eta\}$  can be written as an exponential taylor series

$$(96) \quad \Gamma^X(s) = \exp(s \cdot X) \quad X = \omega_\xi \partial_\xi + \omega_\eta \partial_\eta$$

Level set functions  $\phi$  satisfying

$$(97) \quad \frac{d}{ds} \phi(\Gamma^X(s)) = 0$$

may exist if and only if ([?]) the exponential series in eq. (96) seperates which according to the Baker-Hausdorff-Campbell formula is only the case when the operators  $\{\partial_\xi, \partial_\eta\}$  commute

$$(98) \quad \exp(s \cdot X) = \exp(s \cdot \omega_\xi \partial_\xi) \cdot \exp(s \cdot \omega_\eta \partial_\eta) \Leftrightarrow [\partial_\xi, \partial_\eta] = 0$$

On the otherside if the coordinate functions  $(\xi, \eta)$  satisfy the Cauchy-Riemann equations (eqs. (??)) then one verifys that  $[\partial_\xi, \partial_\eta] = 0$

**The Group  $O(2)$ .** We will now show an example on the group  $O(2)$ , the group of rotations and dilations. The transformation of the cartesian coordinate system to the polar coordinates in 2 dimensions is given by the equations

$$(99) \quad x = r \cdot \cos(\theta)$$

$$(100) \quad y = r \cdot \sin(\theta)$$

Using the expression in eq. (90) the jacobian  $J$  is easily calculated

$$(101) \quad J = \frac{1}{r} \begin{pmatrix} x & -y \\ y & x \end{pmatrix}$$

so that the derivative operators  $\{\partial_\xi, \partial_\eta\}$  may be expressed in the cartesian domain

$$(102) \quad \partial_\xi = x \partial_x + y \partial_y \quad \partial_\eta = -y \partial_x + x \partial_y$$

The coordinates  $\xi$  and  $\eta$  that satisfy eq. (95) with the derivative operators in eq. (102) are functions of the cartesian coordinates

$$(103) \quad \xi(x, y) = \log(r) \quad r = \sqrt{x^2 + y^2}$$

$$(104) \quad \eta(x, y) = \arctan\left(\frac{y}{x}\right)$$

The function  $\eta(x, y)$  is easily recognisable as the angle  $\theta$  to the  $x$ -axis while the function  $\xi(x, y)$  is not the radius  $r$ . This is because the operators  $\{\partial_\xi, \partial_\eta\}$  must both have the same dimension, that means they must be invariant dilations  $r \rightarrow \lambda \cdot r$ .

The level sets of the algebra spun by the operators  $\{\partial_\xi, \partial_\eta\}$  are linear with respect to the coordinates  $(\xi, \eta)$  by virtue of eq. (95). Thus by arguments similar those following eq. (??), the authors in [?] introduced the *generalized structure tensor* (GST)

$$(105) \quad S_{\xi, \eta} = \int \begin{pmatrix} (\partial_\xi \phi)^2 & \partial_\xi \phi \partial_\eta \phi \\ \partial_\xi \phi \partial_\eta \phi & (\partial_\eta \phi)^2 \end{pmatrix} d\xi d\eta$$

As  $S_{\xi, \eta}$  is a symmetric matrix there exists a decomposition

$$(106) \quad S_{\xi, \eta} = V^T D V \quad D = \begin{pmatrix} \lambda_\xi & 0 \\ 0 & \lambda_\eta \end{pmatrix} \quad V = (\mathbf{v}_1, \mathbf{v}_2)$$

The rotation matrix  $V$  acts in the  $(\xi, \eta)$  coordinate space. It does not necessarily correspond to rotations in the cartesian coordinates  $(x, y)$ . See [?] for a discussion on the steerability of the GST.

**1.1. The Transformation Properties of the GST.** Within the generalized coordinate frame  $(\xi, \eta)$  the action of the group  $\mathbb{G}$  manifests itself as a translation

$$(107) \quad g_{\epsilon_1, \epsilon_2} \circ \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} \xi + \epsilon_1 \\ \eta + \epsilon_2 \end{pmatrix}$$

As mentioned before the basis operators  $\{\partial_\xi, \partial_\eta\}$  and as a consequence all elements of the Lie algebra  $\mathcal{G}$  commute with  $\mathbb{G}$  (expand on left invariance!!)

$$(108) \quad g_{\epsilon_1, \epsilon_2} \circ X = X \circ g_{\epsilon_1, \epsilon_2} \quad \forall X \in \mathcal{G}$$

Another important fact is that under the transformation in eq. (107) the volume element  $d\xi d\eta$  is invariant. The consequence is that the GST in eq. (105) is invariant with respect to the generalized translation in eq. (107)

$$(109) \quad g_{\epsilon_1, \epsilon_2} \circ S_{\xi, \eta} = S_{\xi, \eta}$$

The GST has another interesting transformation property. As eq. (106) indicates, there exists an action of the rotation group  $SO(2)$  on the generalized coordinate frame  $(\xi, \eta)$

$$(110) \quad \xi' = \cos(\theta) \xi + \sin(\theta) \eta$$

$$(111) \quad \eta' = -\sin(\theta) \xi + \cos(\theta) \eta$$

The action of the transformation in eq. (111) yields a basis transformation of the Lie algebra  $(\partial_\xi, \partial_\eta)$

$$(112) \quad \partial'_\xi = \cos(\theta) \partial_\xi + \sin(\theta) \partial_\eta$$

$$(113) \quad \partial'_\eta = -\sin(\theta) \partial_\xi + \cos(\theta) \partial_\eta$$

Under the change of basis in eq. (113) the GST transforms like a tensor

$$(114) \quad S'_{\xi, \eta} = R^T S_{\xi, \eta} R \quad R = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

It is important to note (see [?]) that the transformation in eq. (111) is not necessarily connected to the rotations on the cartesian space  $\Omega$  the functions  $\xi(\mathbf{x})$  and  $\eta(\mathbf{x})$  are embedded in. In fact it deforms the level sets corresponding to  $(\partial_\xi, \partial_\eta)$  in a highly non linear manner.

## 2. Structure Tensor Based Prior

Our objective is to construct a prior  $P(\nabla\phi)$  which is invariant to the transformations in eq. (114). We want to define a finite set of Lie groups  $\mathbb{G}_i$  for which the classes of level sets  $A_{\mathbb{G}_i}$  are minimizer sets for  $P(\nabla\phi)$  (see eq. (47)). The methodology goes as follows: We define energy  $E(\nabla\phi)$  (the negative log of  $P(\nabla\phi)$ ) as a *product of the determinants* of the corresponding GSTs  $S_{\xi_i, \eta_i}$

$$(115) \quad E(\nabla\phi) = \prod_i \text{Det}(S_{\xi_i, \eta_i})$$

The energy in eq. (115) inherits the translation invariance (only when  $\phi \in A_{\mathbb{G}_i}$ ??) of the GST in eq. (109), which we will show now. Due to the rotation invariance of the determinants in eq. (115) we can write the individual determinants in terms of their eigenvalues

$$(116) \quad \text{Det}(S_{\xi_i, \eta_i}) = \lambda_{\xi_i} \lambda_{\eta_i}$$

We can write the eigenvalues  $\lambda_{\xi_i}$  and  $\lambda_{\eta_i}$  as the squares of the orthorgonal operators  $X_{\xi_i}$  and  $X_{\eta_i}$  which constitute a rotation of the basis  $(\partial_{\xi_i}, \partial_{\eta_i})$

$$(117) \quad \text{Det}(S_{\xi_i, \eta_i}) = (X_{\xi_i}(\phi))^2 (X_{\eta_i}(\phi))^2$$

Under the adjoint action of the group  $\mathbb{G}_i$  the operators  $X_{\xi_i}$  and  $X_{\eta_i}$  are invariant

$$(118) \quad \left. \frac{d}{d\epsilon_1} (g_{\epsilon_1, \epsilon_2} \circ X_\xi \circ g_{\epsilon_1, \epsilon_2}^{-1}) \right|_{\epsilon_1, \epsilon_2=0} = [X_\xi, \partial_\xi] = 0$$

$$(119) \quad \left. \frac{d}{d\epsilon_2} (g_{\epsilon_1, \epsilon_2} \circ X_\xi \circ g_{\epsilon_1, \epsilon_2}^{-1}) \right|_{\epsilon_1, \epsilon_2=0} = [X_\xi, \partial_\eta] = 0$$

eq. (119) also holds for  $X_{\eta_i}$ . It is evident that under the adjoint action the determinant  $\text{Det}(S_{\xi_i, \eta_i})$  remains invariant. Since the determinant  $\text{Det}(S_{\xi_i, \eta_i})$  vanishes when  $\phi \in A_{\mathbb{G}_i}$  the whole energy in eq. (115) is invariant to any of the adjoint actions of the  $\mathbb{G}_i$  if  $\phi$  is locally in any of the sets  $A_{\mathbb{G}_i}$

The open question which remains to be answered is, how does the energy eq. (115) transform when  $\phi$  is not locally contained in any of the  $A_{\mathbb{G}_i}$ . The basis elements of two different  $\mathcal{G}_i$  and  $\mathcal{G}_j$  do not necessarily commute

$$(120) \quad [\partial_{\xi_i}, \partial_{\xi_j}] \neq 0$$

$$(121) \quad [\partial_{\xi_i}, \partial_{\eta_j}] \neq 0$$

$$(122) \quad [\partial_{\eta_i}, \partial_{\eta_j}] \neq 0$$

$$(123)$$

The question arises that when  $\phi$  is locally within the vicinity of a particular  $\phi_i^* \in A_{\mathbb{G}_i}$ ,  $\|\phi - \phi_i^*\| \leq \delta$  will it be brought further away from  $\phi_i^*$  under the action of  $\mathcal{G}_j$ , that is  $\|g_j \circ \phi - \phi_i^*\| \leq \delta'$ ,  $\delta' > \delta$ ? Or can the groups  $\mathcal{G}_i$  and  $\mathcal{G}_j$  be related to each other such that  $\delta' < \delta$ ? To answer this we look at the product algebra  $\mathcal{G}_i \times \mathcal{G}_j$  spun by the basis elements  $X_l$ . These operators may not be commutative but may be in involution

$$(124) \quad [X_i, X_j] = \sum_l C_{i,j}^l X_l$$

If this is the case then the commutator  $[X_i, X_j]$  is also an element of the joint algebra  $\mathcal{G}_i \times \mathcal{G}_j$ . (if  $\phi$  is roughly linear, then a dilation  $\partial_r$  will stretch the level sets to a line thus resulting in a levelset of the translation group)

**2.1. Analysis of the Eigenvalues of the Rotation Dilation Group.** We will now turn our focus on the eigenvalues of the Rotation Dilation GST. We use the polar coordinates from eq. (100). The integration window of the GST is

$$(125) \quad \xi_0 - \epsilon_\xi < \xi_0 < \xi_0 + \epsilon_\xi$$

$$(126) \quad \phi_0 - \epsilon_\phi < \phi_0 < \phi_0 + \epsilon_\phi$$

where  $\xi_0 = \ln(r_0)$  so that eq. (126) translates to a region around the curvature radius  $r_0$  and the angle  $\phi_0$ . The level sets parameterised by the polar coordinates in the region in eq. (126) are the sectional curves of constant curvature  $k_0^{-\epsilon_\xi} < k_0 < k_0^{\epsilon_\xi}$ ,  $k_0 = \frac{1}{r_0}$ . The Rotation Dilation GST from eq. (105) can be written in cartesian coordinates

$$(127) \quad S_{\xi,\eta} = \int_{y_0-\epsilon}^{y_0+\epsilon} \int_{x_0-\epsilon}^{x_0+\epsilon} \begin{pmatrix} (\partial_\xi \phi)^2 & \partial_\xi \phi \partial_\eta \phi \\ \partial_\xi \phi \partial_\eta \phi & (\partial_\eta \phi)^2 \end{pmatrix} \frac{1}{r^2} M(x, y) dx dy$$

The mask  $M(x, y)$  enforces the conditions eq. (126). The orientations of the level sets in this domain are fixed and thus This is why the minimum eigenvalue of the GST is zero only for level sets matching the curvature  $k_0$  and the orientation  $\phi_0$ . Since the integration space of  $S_{\xi,\eta}$  is centered around  $\phi_0$ ,  $S_{\xi,\eta}$  is not rotation invariant. In figure 1a the image function  $I(x, y) = -r^2$  is depicted. The GST was evaluated for  $r_0 = 30$  and  $\phi_0 = \frac{\pi}{4}$  (Figures 1b to 1d) and  $\phi_0 = \frac{\pi}{2}$  ((Figures 1e to 1g)). The eigenvalue corresponding to  $X_\eta$ , the derivative in angular direction is denoted by  $\lambda_2$  (Figures 1d to 1g). It is observed that  $\lambda_2$  has a minimum at the radius  $r_0 = 30$  and the angles  $\phi_0 = \frac{\pi}{4}$  (figure 1d) and  $\phi_0 = \frac{\pi}{2}$  (figure 1g). As a result the determinant of the GST is only minimal at exactly those values (see figures 1b and 1e).

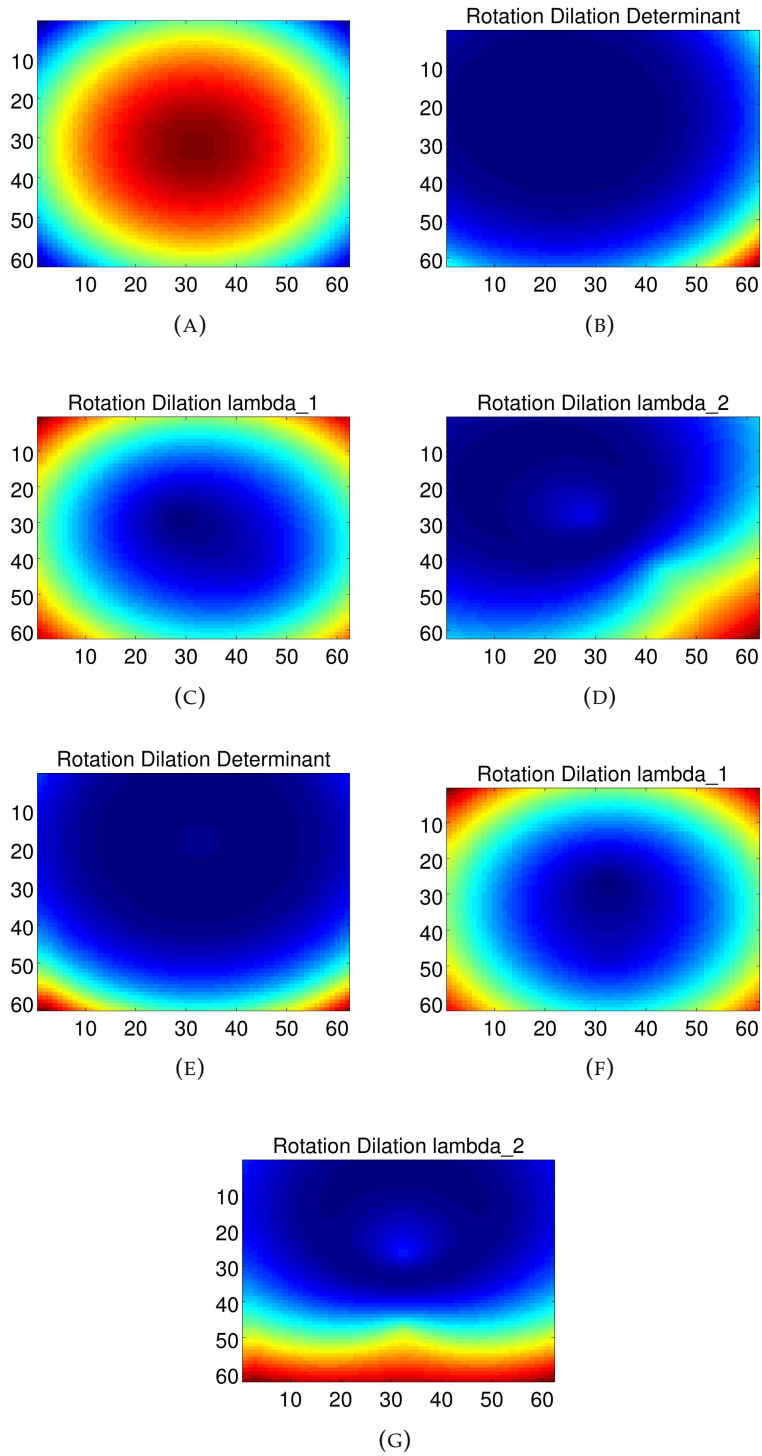


FIGURE 1. Figure 1a is the image  $I(r) = -r^2$ , figures 1b to 1d are the determinant, higher and lower eigenvalue of the GST for  $r_0 = 30$  and  $\phi_0 = \frac{\pi}{4}$ . Figures 1e to 1g are same for  $\phi_0 = \frac{\pi}{2}$ .

## Geometrical Optical Flow Model

Not clear what you are referring to.

In this chapter we will introduce our new model optical flow based on the image fusion algorithm from Hardie et al. [?]. We will address the two issues outlined in section 3, namely that the images  $y$  and  $I$  (Figures 1b and 1d) are *not* co-aligned and *not* joint Gaussian.

Problem from where? Discussed / defined where? Give equation?

### 1. Disparity

To solve the problem that the images  $y$  and  $I$  are not co-aligned we need a mapping between the pixels of  $y$  and those of  $I$ . However the problem exists that the modalities producing the images  $y$  and  $I$  have different native resolutions so that it is not clear how the pixels of  $I$  should be mapped to pixels in  $y$ . In section [?] a method was introduced which produces a super-resolved image  $S$  given co-aligned data  $y$  and  $I_c$ . In that model pixels in  $S$  and in  $I_c$  have a one-on-one relationship, so that it is natural to map pixels in  $I$  to  $S$  rather than to  $y$  directly. We model the disparity between the images  $S$  and  $I$  by setting the co-aligned VSC image  $I_c$  to be the result of the original VSC  $I$ , warped by an unknown optical flow field  $\mathbf{d}(\mathbf{x})$ ,

Vague and too far away. Maybe repeat the relevant equations here?

$$(128) \quad I_c(\mathbf{x}) = I(\mathbf{x} + \mathbf{d}(\mathbf{x})) = I_d(\mathbf{x})$$

Substituting eq. (128) into eq. (14) and following, we obtain the likelihood

$$(129) \quad p(s|y, I, \mathbf{d}) = p(s|y, I_d)$$

with the energy

$$(130) \quad E_{data}(S, \mathbf{d}) = \frac{1}{2} \int_{\Omega} (y(\mathbf{x}) - W_{\sigma} S(\mathbf{x}))^2 \cdot C_n^{-1} dx + \frac{1}{2} \int_{\Omega} (S(\mathbf{x}) - \mu_{s|I_d}(\mathbf{x}))^2 \cdot C_{s|I_d}^{-1} dx$$

We can insert the simplified closed form expression for  $\hat{S}$  from eq. (??) into  $E_{data}$

$$(131) \quad E_{data}(\mathbf{d}) = \frac{1}{2} \int_{\Omega} (y(\mathbf{x}) - f \cdot \tilde{I}_d(\mathbf{x}))^2 \cdot C_{s|\tilde{I}_d} \left( C_{s|\tilde{I}_d} + \lambda C_n \right)^{-2} \quad f = C_{y, \tilde{I}_d} C_{\tilde{I}_d, \tilde{I}_d}^{-1}$$

The data term  $E_{data}$  defines a likelihood for  $\mathbf{d}$

Don't just state the properties! State the rationale / insight / reason why the property is relevant or interesting.

$$(132) \quad p(\mathbf{d} | I) = \exp(-E_{data}(\mathbf{d}))$$

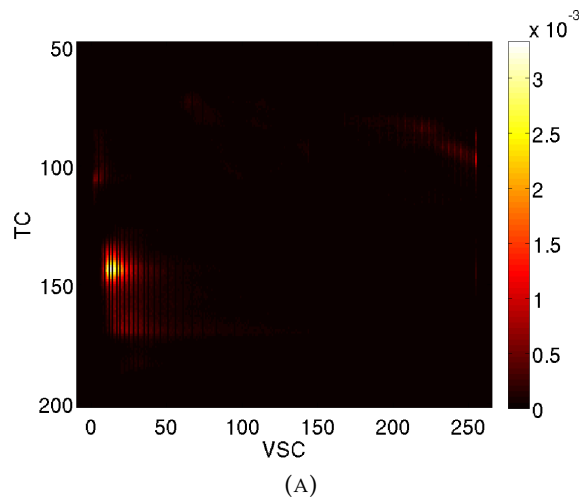
and has the following interesting properties: the factor  $f$  transforms the intensity range of the image  $I_d$  to a range similar to that of  $y$  so that  $E_{data}$  is a measure for the similarity between  $y$  and  $f \cdot I_d$ . Furthermore as  $f \cdot I_d$  and  $y$  become more similar the conditional variance converges  $C_{s|I} \rightarrow 0$ . Thus in the vicinity of the optimal solution  $\hat{\mathbf{d}}$  the posterior distribution eq. (137) becomes sharper.

The minimization of eq. (137) with respect to  $\mathbf{d}(\mathbf{x})$  is an ill-posed problem. For regions  $\mathcal{V}_I$  of  $I$  and  $\mathcal{V}_y$  of  $y$  in which  $I$  and  $y$  contain little structural information, a pixel  $\mathbf{x}_I \in \mathcal{V}_I$  can be mapped to any pixel in  $\mathcal{V}_y$ , meaning that  $\mathbf{d}(\mathbf{x}_I)$  is ambiguous. This problem is referred to as the aperture-problem of optical flow (refs...). We therefore apply a smoothness prior for  $\mathbf{d}(\mathbf{x})$ . A common choice for such a prior would be the total variation (TV) norm, [?, ?], . However we will use our structure tensor based prior from

Equation numbering should restart in each chapter. It is a bit unusual to have the numbering just continue. Make sure you are using the official thesis style.

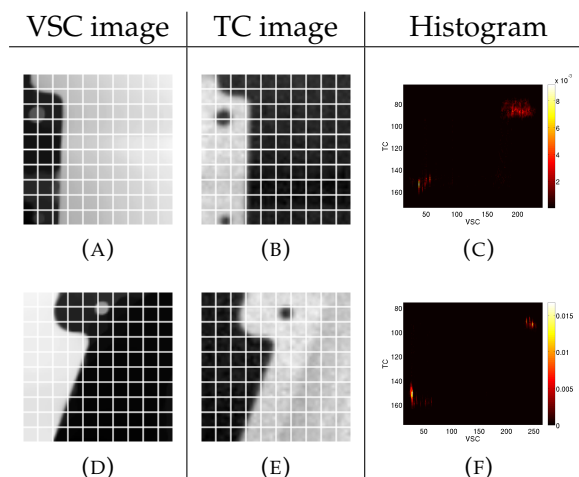
Not clear where this comes from.

Poor figure numbering. Should be something like 3.1



Wrong caption

FIGURE 1. Joint Histogram of the TC image figure 1b and the VSC image figure 1d. We observe that there is no linear relationship between the TC and the VSC



I don't really see how you reach this conclusion from the figure.

FIGURE 2. Different roi's and their joint histograms. A grid is shown in the VSC and the TC image to emphasize the disparity between them. The gridsize is 10 pixels. In the histograms we see there is a linear relationship between the VSC and the TC roi's

chapter 2 which is well suited for linear displacements which occur often in multimodal setups. Combining (132) and  $p_{ST}(\nabla \mathbf{d})$  we obtain the posterior

$$(133) \quad p(\mathbf{d}|y, I) = p(y, I|\mathbf{d}) \cdot p_{ST}(\nabla \mathbf{d})$$

## 2. Localization

The assumption that the intensities of the images  $y$  and  $I$  are globally linear related is a very strong constraint that can hold in most cases only unimodal data. In the case of the VSC and TC data in figure 1 the assumption of linearity fails. In figure 1 the global joint histogram of the VSC and the TC image is shown. We can see that the distribution in the joint histogram lacks a linear relationship between the TC and the VSC. However in figure 2 we have evaluated the histogram within local region of interests. The histograms



So what is the overall significance or conclusion? Not clear how this is relevant.

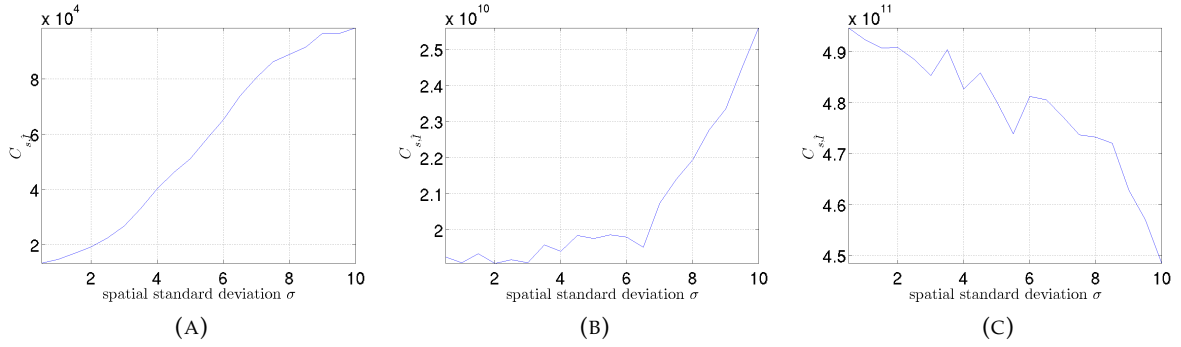


FIGURE 3. Median conditional variance  $\hat{C}_{s|I}^{\sigma,a}$  for  $a = 5$  (figure 3a),  $a = 23$  (figure 3b) and  $a = 33$  (figure 3c). We can see that for small values of  $a$   $\hat{C}_{s|I}^{\sigma,a}$  has a minimum at  $\sigma < 2$ , and for larger values of  $a$  the profile changes so that the minimum of  $\hat{C}_{s|I}^{\sigma,a}$  is at  $\sigma \geq 10$

Always capitalize acronyms

in figures 2c and 2f show that within the **FOI** the assumption of linearity between the intensities of the TC and the VSC is well supported. Therefore we propose a local version of the variance in eq. (13)

$$(134) \quad C_{u,v}(\mathbf{x}_0) = \int_{\Omega} \omega(\mathbf{x} - \mathbf{x}_0) (u(\mathbf{x}) - \mathbb{E}(u, \mathbf{x}_0)) \cdot (v(\mathbf{x}) - \mathbb{E}(v, \mathbf{x}_0))$$

where  $\omega$  is a window function which we take to be constant within a subset  $W \subset \Omega$

$$(135) \quad \omega(\mathbf{x}) = \begin{cases} \frac{1}{|W|-1} & 0 \leq x, y \leq a \\ 0 & \text{else} \end{cases}$$

Then  $C_{s|I}^{\sigma,a}(\mathbf{x})$  becomes a local measure that measures how linear the intensities of  $y$  and  $I$  are within the sub domain  $W$ . The problem that arises is how large to set the window size  $a$ . If it is set too small the signal to noise ratio will be too small so that not enough information of the features in the TC and the VSC image are captured to robustly register them. On the other hand if  $a$  is set too large we eventually loose the local linearity between the TC and the VSC image. In figure 3 we have plotted the median conditional variance

$$(136) \quad \hat{C}_{s|I}^{\sigma,a} = \text{median} \left( C_{s|I}^{\sigma,a}(\mathbf{x}) \right)$$

as a function of  $\sigma$  for three fixed values of the window size  $a$ . In figure 3a ( $a = 5$ )  $\hat{C}_{s|I}^{\sigma,a}$  has a minimum for  $\sigma < 2$ , and in figure 3c ( $a = 33$ ) it is minimal for  $\sigma \geq 10$ . The profile of  $\hat{C}_{s|I}^{\sigma,a}$  changes from monotonic increasing to monotonic decreasing for small to large values of  $a$ . Since we know the value for the scale parameter  $\sigma$ ,  $\sigma^* = 2$  from the ccd resolutions of the cameras, the idea find the optimal value  $a^*$  such that  $\hat{C}_{s|I}^{\sigma^*,a^*}$  is minimal at  $\sigma = \sigma^*$ . For  $a = 23$  this is the case as we see in figure 3b. Thus for the data in figure 1  $a^* = 23$  is the optimal value so that  $\hat{C}_{s|I}^{\sigma^*,a^*}$  has physically meaningful minimum  $\sigma^* = 2$ . The local data term  $E_{data}$  now has the form

$$(137) \quad E_{data}(\mathbf{d}) = \frac{1}{2} \int_{\Omega} \left( y(\mathbf{x}) - f(\mathbf{x}) \cdot \tilde{I}_{\mathbf{d}}(\mathbf{x}) \right)^2 \cdot C_{s|\tilde{I}_{\mathbf{d}}}^{\sigma^*,a^*}(\mathbf{x}) \left( C_{s|\tilde{I}_{\mathbf{d}}}^{\sigma^*,a^*}(\mathbf{x}) + \lambda C_n \right)^{-2}$$

$$(138) \quad f(\mathbf{x}) = C_{y,\tilde{I}_{\mathbf{d}}}(\mathbf{x}) C_{\tilde{I}_{\mathbf{d}},\tilde{I}_{\mathbf{d}}}^{-1}(\mathbf{x})$$

and together with our prior from chapter 2 the energy for the complete optical flow model is

$$(139) \quad E(\mathbf{d}) = E_{data}(\mathbf{d}) + \frac{\lambda}{2} \left( \sum_i \text{Det}(S(d_i)) \right)$$

The matrix  $S(d_i)$  is the structure tensor (see eq. (84)) acting on each component of the optical flow  $\mathbf{d}$ . In this model we are making the assumption that the motion boundaries are locally linear. This assumption is valid for object boundaries with small curvature but as we will see in chapter ?? this assumption fails at junction points in the optical flow field, since those are where objects are partially occluding each other and moving in opposite directions.

### 3. The solution algorithm

To minimize 139 and obtain the optimum flow field  $\hat{\mathbf{d}}$  we deploy a simple newton scheme with a nested linearization of 139. The linearized model is solved by a conjugate gradients algorithm with block Jacobi preconditioning. The problem with this approach is that the regularizer is quartic in the flow field components and thus the linearization becomes numeric instable for the initial steps of the algorithm.

---

#### Algorithm 1 Optical Flow with Structure Tensor prior

---

```

Initialize  $\mathbf{d}_0 = 0$ 
Set  $\mathbf{r}_0 = \frac{\delta E(\mathbf{d})}{\delta \mathbf{d}}(\mathbf{d}_0)$ 
scale  $s = s_{\text{Max}}$ 
while  $s > 1$  do
  downsample  $y_s = G_s \star y_0, I_s = G_s \star I_0$ 
  while  $\|\mathbf{r}\| > \epsilon$  or  $k < N$  do
    set  $\mathbf{d}_{k+1} = \mathbf{d}_k + \alpha \delta$ 
    expand  $E(\mathbf{d}_{k+1}) = E(\mathbf{d}_k) + \alpha \mathbf{b}_k^T \delta + \frac{\alpha^2}{2} \delta^T Q_k \delta$ 
    solve  $Q_k \delta = \mathbf{b}_k$  for  $\delta$  with conjugate gradients and suitable preconditioning
    compute  $\mathbf{d}_{k+1} = \mathbf{d}_k + \alpha \delta, k \rightarrow k + 1$ 
  end while
  upsample  $d_N$ , set  $d_0 = d_N, k = 0$ 
   $s = s - 1$ 
end while

```

---

The problem arises in step 9 of the iterative algorithm. The second functional derivative  $Q_k$  of the energy model 139 consists of one part coming from the likelihood and one part coming from the prior,  $Q_k = Q_k^{data} + \lambda Q_k^{reg}$ . The matrix  $Q_k^{reg}$  is the second derivative of the prior in 139 with respect to  $\mathbf{d}$ . At small  $k$  its eigenvalues are small due to the initial guess  $\mathbf{d}_0 = 0$ . The matrix  $Q_k^{data}$  is the second derivative of the likelihood in eq. (139). In regions where there is no motion the eigen values of  $Q_k^{data}$  are also small. This makes the linearized solution in step 9 numerically instable. Our solution to this problem is to extend 139 to include an  $L_2$  prior on the flow field  $\mathbf{d}$  but with a small lagrange multiplier  $\lambda_2$

$$(140) \quad E(\mathbf{d}) = \int (y - \hat{s}_{I,\mathbf{d}})^2 \cdot C_{s|I_d} + \frac{\lambda}{2} \left( \sum_i (\text{Det}(S(d_i)) + \lambda_2 \|\nabla d_i\|) \right)$$

With the  $L_2$  prior in 140 the linearized solution in step 9 becomes numerically stable.

## CHAPTER 4

# Results

### 1. Uni-Modal Data

We will now discuss the results of our optical flow method on the middleburry data set for which there exists ground truth (GT). As the GT is the true flow field for the data we use it to assess the quality of the computed optical flow. To do this we define the Endpoint error (EPE) and the angular error (AE) as

$$(141) \quad e_{EPE} = \|\mathbf{v} - \mathbf{v}_{gt}\|$$

$$(142) \quad e_{AE} = \cos(\angle(\mathbf{v}, \mathbf{v}_{gt})) \in \{-1, 1\}$$

The EPE  $e_{EPE}$  measures how well the computed optical flow  $\mathbf{v}$  fits the true optical flow  $\mathbf{v}_{gt}$ . In cases where  $\mathbf{v}$  does not match  $\mathbf{v}_{gt}$  well, we would still like to check how both vectors are aligned. This alignment is depicted by the AE values ranging between  $-1$ , for minimal alignment (worst case), and  $1$  for maximal alignment (best case).

Figure	Filtersize	Median, Min, Max EPE	Median, Min, Max AE
figure 3a	7	2.36, 0.01, 7.24	0.42, $-1.00$ , 1.00
	9	1.32, 0.00, 6.02	0.87, $-1.00$ , 1.00
	11	1.15, 0.00, 6.45	0.91, $-1.00$ , 1.00
figure 3f	7	0.84, 0.01, 13.35	0.87, $-1.00$ , 1.00
	9	0.46, 0.01, 8.23	0.97, $-1.00$ , 1.00
	11	0.40, 0.00, 8.25	0.98, $-1.00$ , 1.00
figure 4a	7	0.47, 0.01, 5.22	0.97, $-0.96$ , 1.00
	9	0.28, 0.00, 3.71	0.99, $-1.00$ , 1.00
	11	0.25, 0.00, 2.50	0.99, $-1.00$ , 1.00
figure 4f	7	0.44, 0.00, 2.73	0.98, $-1.00$ , 1.00
	9	0.34, 0.00, 2.65	0.99, $-1.00$ , 1.00
	11	0.30, 0.00, 3.12	0.99, $-1.00$ , 1.00

TABLE 1. EPE and AE analysis

EPE and AE values for different region of interests and filter sizes (Figures 3a to 4f). The second column shows the median, minimum and maximum EPE per roi. The third column shows the median, minimum and maximum AE per roi. The table shows that the EPE gets better with increasing filtersize. Despite this the values for roi's with non-linear geometry (figure 3) generally have higher EPE values than the roi's with linear or constant geometry (figure 4)

**1.1. Middleburry Dataset.** In figure 1 the rubber whale sequence of the middleburry data set is shown, and in figure 1b the corresponding ground truth  $\mathbf{v}_{gt}$ . In figure 1d the computed flow-field  $\mathbf{v}$  is shown for a filter size of 11, while in figure 1c the resulting flow for the TV model is shown. Figures 3 and 4 show different region of interrests (roi) for which the EPE and AE are shown on a pixel basis for the structure tensor model and Figures 5 and 6 show the same for the TV model. We can observe from the comparison between figures 1d and 1c that the TV model produces smoother results which are closer

Figure	Median, Min, Max EPE	Median, Min, Max AE
figure 5a	1.38, 0.00, 5.83	0.92, -1.00, 1.00
figure 5f	0.20, 0.00, 3.34	1.00, -1.00, 1.00
figure 6a	0.04, 0.00, 2.08	1.00, -1.00, 1.00
figure 6f	0.09, 0.00, 2.06	1.00, -1.00, 1.00

TABLE 2. EPE and AE analysis

EPE and AE values for different region of interests for the TV model (Figures 5a to 6f). The first column shows the median, minimum and maximum EPE per roi. The second column shows the median, minimum and maximum AE per roi. Compared to table 1 the median EPE is lower for nearly all roi's, while the median AE do not differ that much

Figure	Filtersize	Median, Min, Max EPE	Median, Min, Max AE
figure 7a	7	0.73, 0.00, 6.80	0.99, -1.00, 1.00
	9	0.60, 0.00, 7.29	0.99, -1.00, 1.00
	11	0.96, 0.01, 15.60	0.98, -1.00, 1.00
figure 7f	7	0.36, 0.00, 7.00	1.00, 0.00, 1.00
	9	0.27, 0.00, 6.79	1.00, 0.00, 1.00
	11	0.41, 0.01, 6.55	1.00, 0.00, 1.00

TABLE 3. EPE and AE analysis

EPE and AE values for different region of interests and filter sizes (Figures ?? to ??).

Since the motion boundaries in figure 2a are all curvilinear there is no correlation between the filtersize and the EPE.

Figure	Median, Min, Max EPE	Median, Min, Max AE
figure 8a	0.44, 0.00, 6.12	1.00, -1.00, 1.00
figure 8f	0.12, 0.01, 7.38	1.00, 0.00, 1.00

TABLE 4. EPE and AE analysis

EPE and AE values for different region of interests for the TV model (Figures ?? to ??).

The first column shows the median, minimum and maximum EPE per roi. The second column shows the median, minimum and maximum AE per roi. Compared to table 1 the median EPE is lower for nearly all roi's, while the median AE do not differ that much

to the ground truth (figure 1b). In tables 1 and 4 the median values for the EPE and AE in various region of interrests are listed. Indeed we can observe that the EPE for the TV is approximately half the value of that of the structure tensor model. We chose the median as opposed to the mean EPE as it is robust outlier values of the EPE at single pixel locations and thus gives a better assessment of the quality of the flow within the roi.

Table 1 shows also how the EPE and the AE vary with increasing filtersize: The EPE decreases while the AE increases. In figure 4 the roi's have mostly a constant motion field or a motion field with linear boundary, thus according to table 1 they have lower EPE values then the roi's in figure 3. The roi with the largest discrepancy from the group of linear motions is figure 3a which depicts a rotating wheel. Since the wheel is largely free of texture, the motion field (figure 3d) is penalized by the structure tensor prior in such a way that it acquires spurious linear motion boundaries. This is the reason for its high EPE value. The roi in figure 3f shows another case of a motion field violating the assumption of linear motion boundaries. In the ground truth roi in figure 3j there are two junction points where three objects are occluding and moving against each other. This type of motion is penalized by the structure tensor prior so that the flow at these points is oversmoothed. The TV model (ref!) like the structure tensor model penalizes

non linear motion boundaries. figure 5d shows the result of the TV model for the wheel roi. Just like in the structure tensor model, the flow on the circumference of the wheel is heavily penalized resulting in high EPE values and wrong AE values (see table 4). figure 5i shows the resulting flow of the TV model at the two junctions in figure 5f. Similar to our proposed prior the flow is oversmoothed at the junctions resulting in high EPE values (see table 4).

On the otherside both models are faithful to roi's with constant motion or linear motion boundaries (see figures 4 and 6). In figure 4d we see that the structure tensor model inflicts more of the texture from the underlying data (figure 4a) on the estimated flow than the TV model (see figure 6d for the result of the TV model) thus leading to a slightly higher EPE value (table 1). Figure 4i shows an example of an roi with a linear motion boundary for the structure tensor model. Comparing it to the corresponding result for the TV model figure 4i, we see that the structure tensor model produces sharper lineat motion boundaries.

In figure 2 another sequence of the middleburry data set is shown. In this sequence the camera is rotating around a hydrangea. As the ground truth shows there are no linear motion boundaries, thus only the constant motion of the background is correctly detected (upto some artifacts in the upper left corner in figure 2d), see the EPE and AE values in figure 7 and table 3.

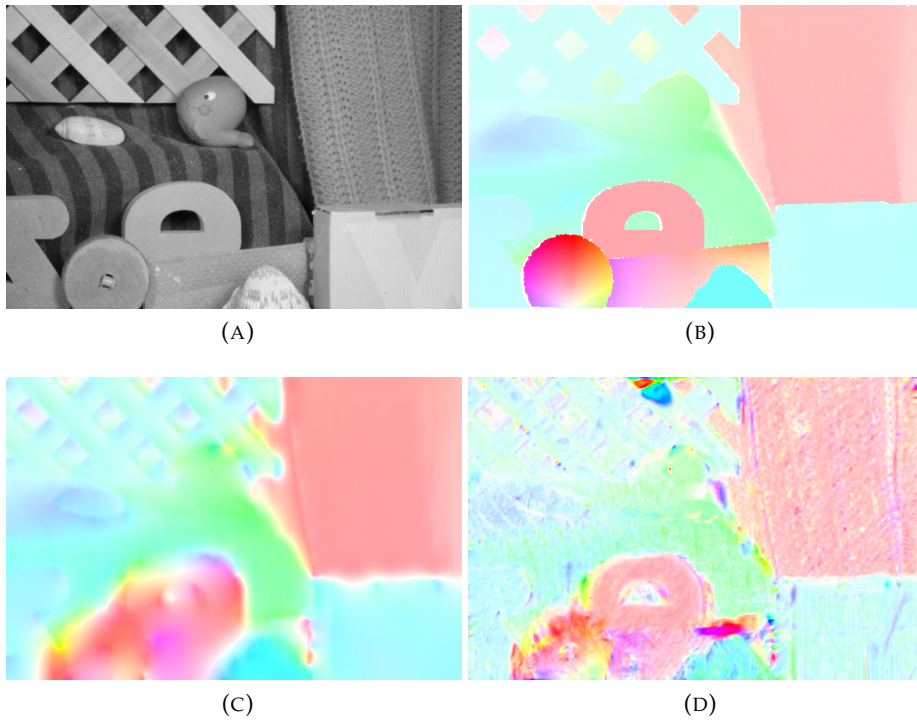


FIGURE 1. Rubberwhale Sequence

Figure 1a shows one frame of the sequence. figure 1d shows the estimated optical flow, figure 1c the result of the TV model and figure 1b shows the provided ground truth

## 2. Eigenvalue analysis and the stabilization parameter $\lambda_2$

In chapter 3 we stated that the  $L_2$  term in eq. (140) is needed to support the numerical stability of the model. We will back this statement now. Figures 9, 10 and 11 show the largest eigenvalue of  $Q_{reg}^i, \sigma_Q^i$  at each iteration on the coarsest scale of the pyramid for

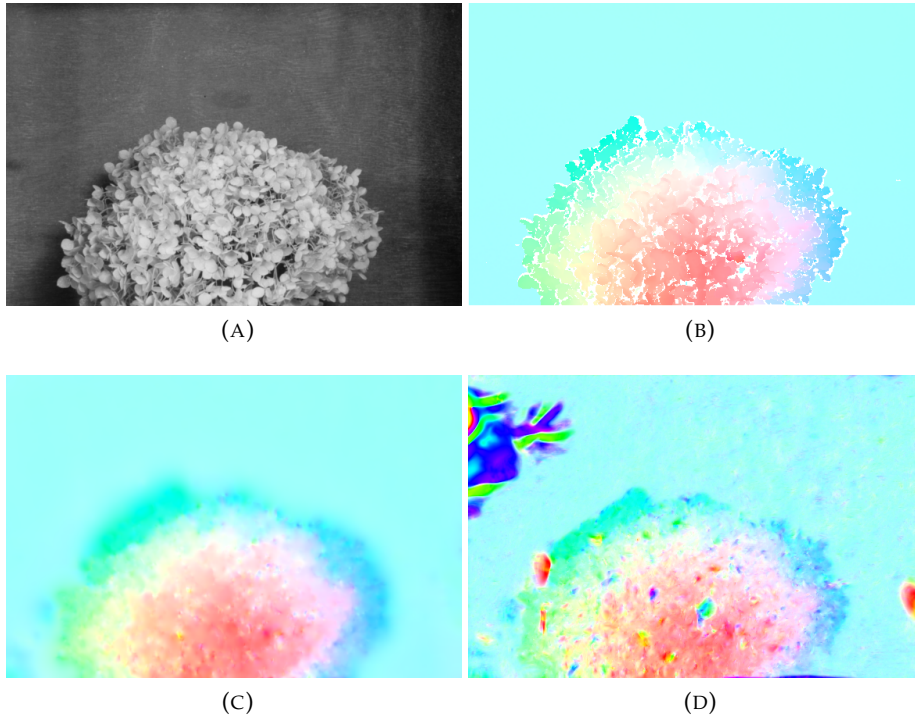


FIGURE 2. Hydrangea Sequence

Figure 2a shows one frame of the sequence. figure 2d shows the estimated optical flow, figure 2c the result of the TV model and figure 2b shows the provided ground truth


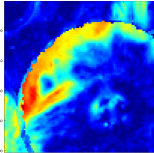
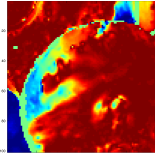
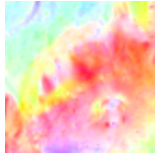


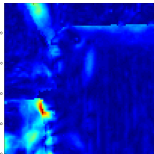
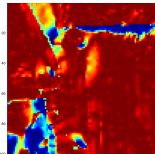
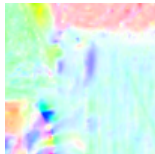

Region of interest	EPE	AE	Resulting Flow	Ground Truth $v_{gt}$
 (A)	 (B)	 (C)	 (D)	 (E)
 (F)	 (G)	 (H)	 (I)	 (J)

FIGURE 3. Error Analysis ST model: This figure shows two examples of motion field with nonlinear boundaries. In figure 3c we see that along the circumference of the wheel the EPE has the largest values and in figure 3h the is largest the junction point where three objects ar moving against each other.

different values of  $\lambda_2$ . They all show that  $\sigma_Q^N$  rises to a maximum after which it decreases and converges. The initial value of  $\sigma_Q^i$  is of the order of  $\lambda_2$  indicating that in the initial steps the  $L_2$  term in eq. (140) governs the regularization. As the number of iterations increases the structure tensor determinant gets more weight, until the point where its

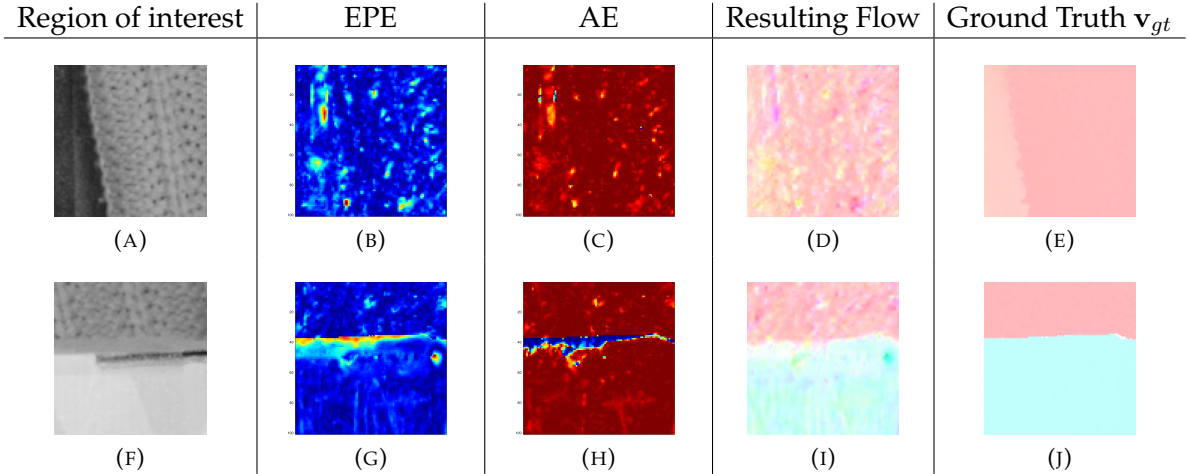


FIGURE 4. Error Analysis ST: This figure shows two examples of motion fields with linear boundaries. In figures 4d and 4i we can see that the resulting flow with texture inflicted from the data. Nevertheless the EPE values are nearly homogenous and small (see figures 4c and 4h)

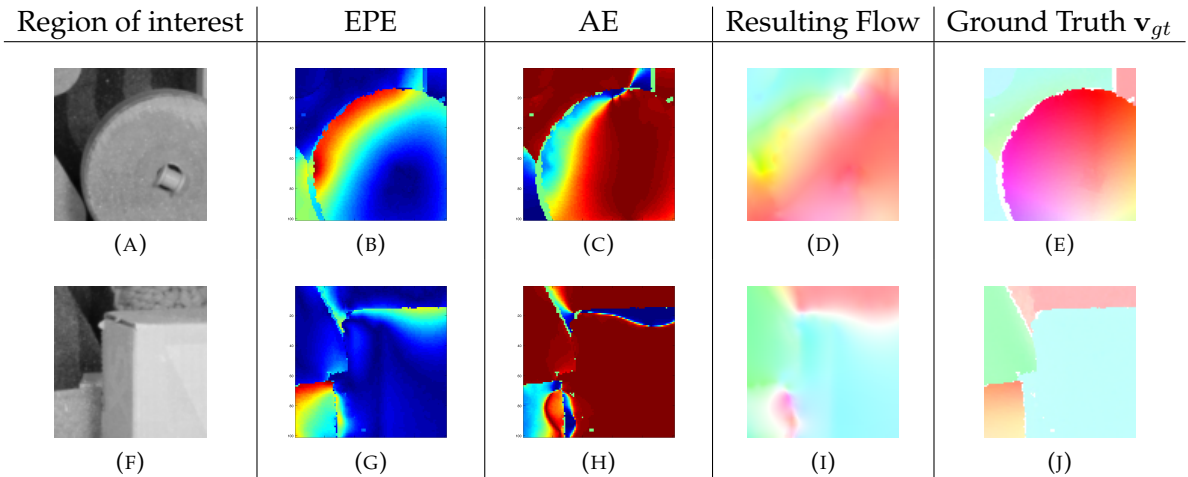


FIGURE 5. Error Analysis TV model: This figure shows two examples of motion field with nonlinear boundaries. In figure 5c we see that along the circumference of the wheel the EPE has the largest values and in figure 5h the is largest the junction point where three objects ar moving against each other.

influence over weighs that of the  $L_2$  term As can be seen this point comes after fewer iterations the smaller  $\lambda_2$  is set. On the other side Figures 13, 14 and 15 show the vector  $\mathbf{b}$ , that is the Euler-Lagrange equation vector for different values of  $\lambda_2$ . Comparing the magnitude of  $\mathbf{b}$  in Figures 13, 14 and 15 we see that for  $\lambda_2 = 10^{-9}$   $\mathbf{b}$  is several orders of magnitude larger then the other cases, which leads to longer convergence rates or numerically instable solution. This means we have a tradeoff between

- $\lambda_2 \sim 10^{-3}$ : Faster convergence but less influence of structure tensor (need  $i > 40$  iterations for ST to act)
- $\lambda_2 \sim 10^{-9}$ : slower convergence but more influence of structure tensor (need only  $i > 1$  iterations for ST to act)

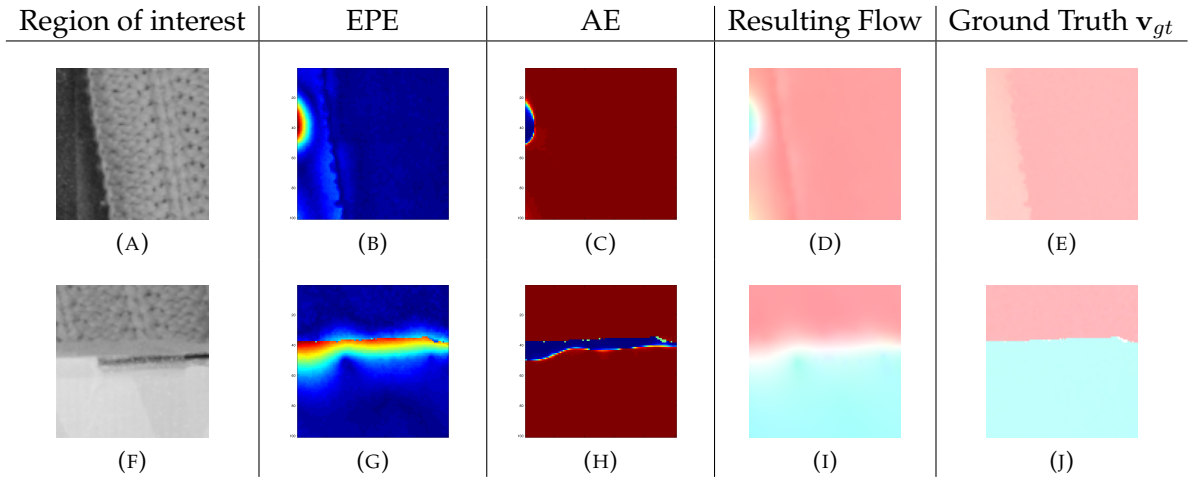


FIGURE 6. Error Analysis TV: This figure shows two examples of motion fields with linear boundaries. In figures 6d and 6i we can see that the resulting flow with texture inflicted from the data. Nevertheless the EPE values are nearly homogenous and small (see figures 6c and 6h)

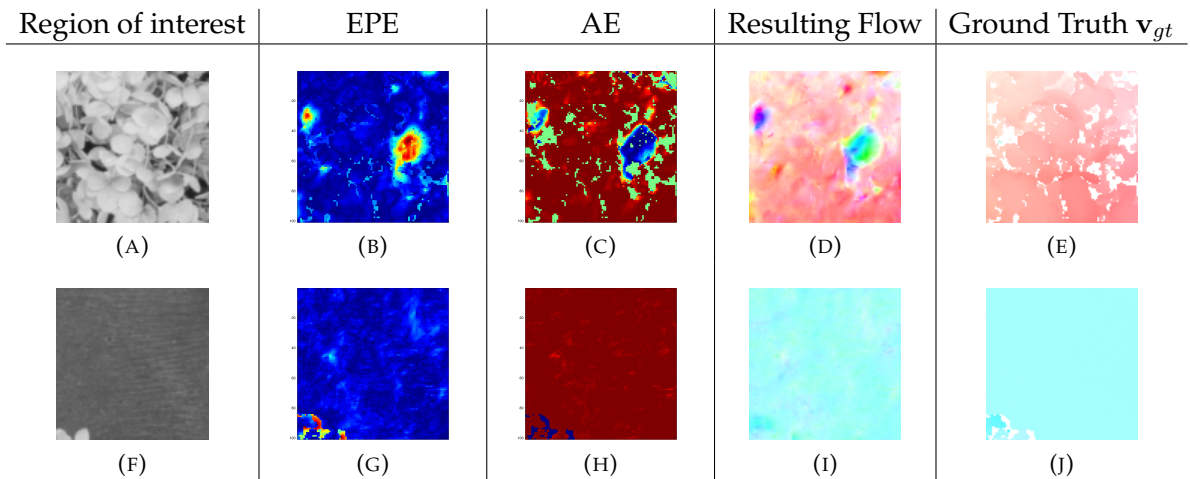


FIGURE 7. Error Analysis:  
Second Column: Endpoint Error, Third Column: Angular Error.

We choose  $\lambda_2 = 10^{-6}$  since in this case  $\mathbf{b}$  is of the same order of magnitude as for  $\lambda_2 = 10^{-3}$  but as we see in figure 10 the structure tensor only needs 4 iterations until its eigenvalues overweigh the eigenvalues of the  $L_2$  term. We also choose  $N = 10$  for the number of iterations per pyramid scale, since according to figure ?? the update vector  $\delta$  gets unstable after 15 iterations.




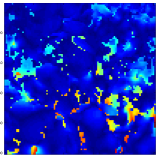
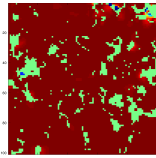


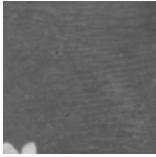
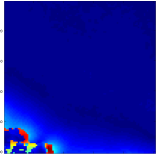
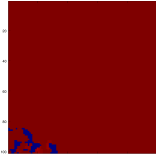


Region of interest	EPE	AE	Resulting Flow	Ground Truth $\mathbf{v}_{gt}$
 (A)	 (B)	 (C)	 (D)	 (E)
 (F)	 (G)	 (H)	 (I)	 (J)

FIGURE 8. Error Analysis:  
Second Column: Endpoint Error, Third Column: Angular Error.

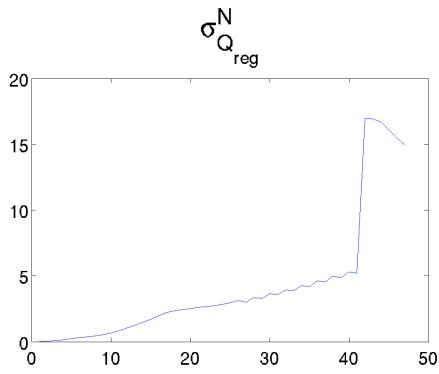


FIGURE 9.  $\lambda_2 = 10^{-3}$

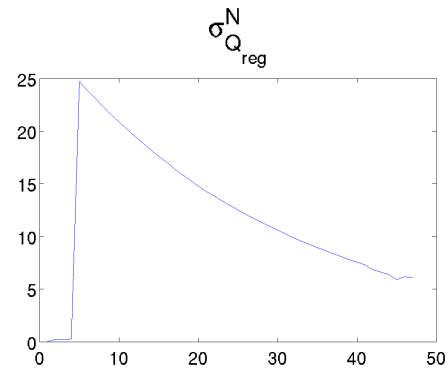


FIGURE 10.  $\lambda_2 = 10^{-6}$

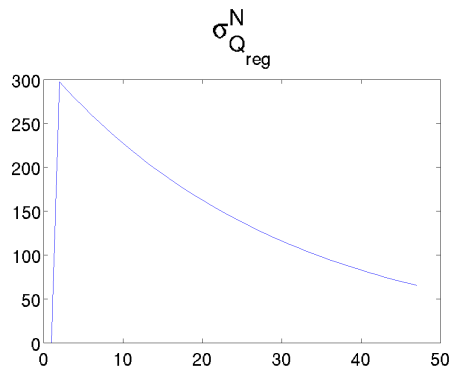


FIGURE 11.  $\lambda_2 = 10^{-9}$

FIGURE 12. Analysis of the largest eigenvalue  $\sigma_Q^i$  of  $Q^{reg}$

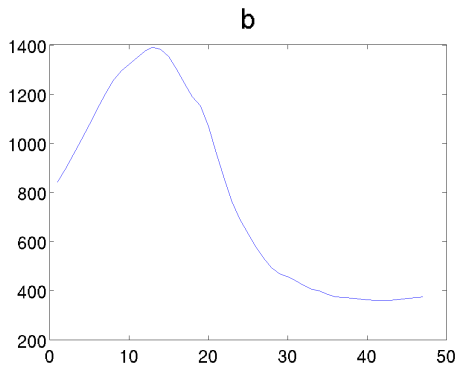


FIGURE 13.  $\lambda_2 = 10^{-3}$

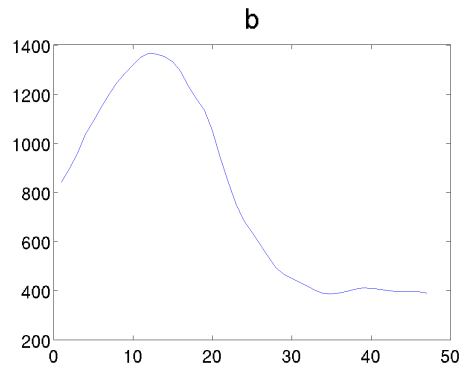


FIGURE 14.  $\lambda_2 = 10^{-6}$

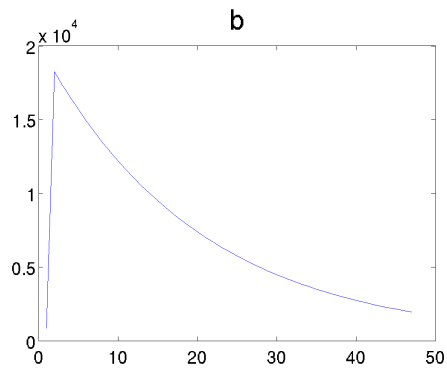


FIGURE 15.  $\lambda_2 = 10^{-9}$

FIGURE 16. Analysis of the Euler-Lagrange vector **b**

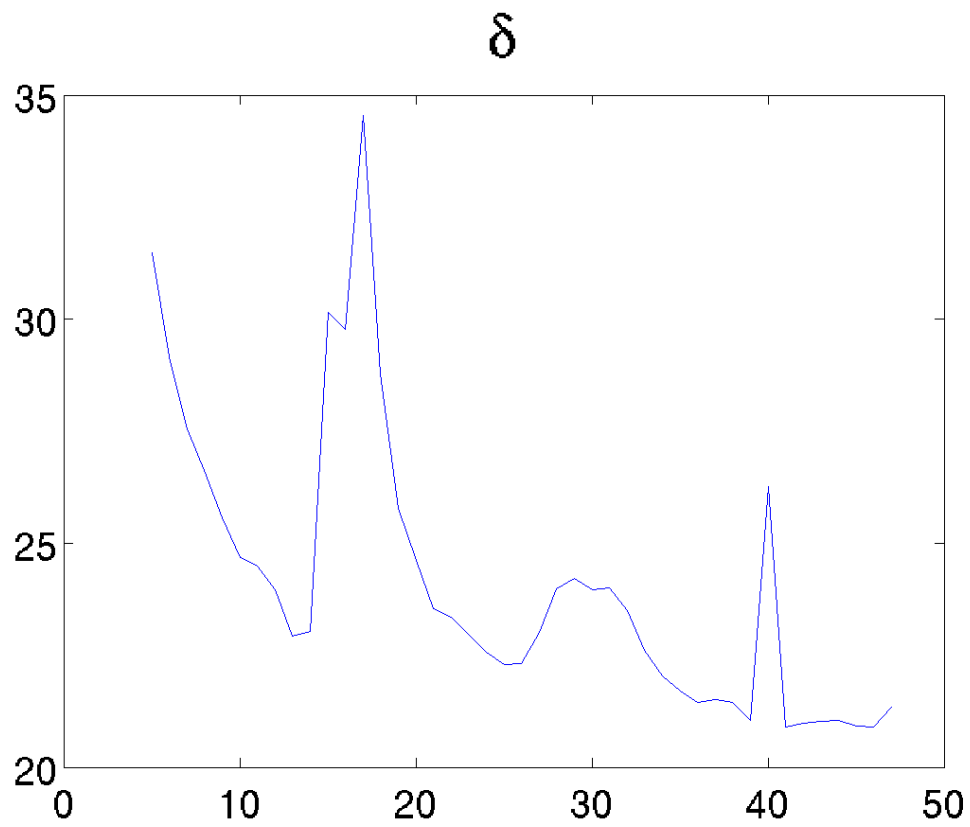


FIGURE 17. Analysis of the Euler-Lagrange vector  $\delta$  eq. (??)



## Multimodal Optical Flow

### 1. Estimation of the resolution parameter $\sigma$

Our optical flow model eq. (??) is based on the assumption that the modalities to be registered have a linear relationship in their intensity spectrum. This is not the case for TC images and VSC images of arbitrary objects. However in the case of bare CFRPs the linearity assumption holds. CFRPs are black bodies when in thermal equilibrium at 30 deg  $C$  since the emmissivity of carbon is approximately 0.98 (see [?]). It is in this case that in the amplitude image in figure 1b the CFRP has a uniform amplitude. In the visual spectrum domain (figure 1d) the CFRP is not a perfect black body due to the reflective nature of the epoxy coating, however the epoxy coating is uniformly distributed so that the reflections do not cause image gradients, which are not correlated to geometric features. Since the TC and the VSC have different resolutions we must take the difference in resolution into account. This difference in resolution is encoded in the scale parameter  $\sigma$  of our local likelihood model in eq. (9). The local conditional variance  $C_{s|\tilde{I}}(\mathbf{x})$  in eq. (??) is a measure for the similarity of the TC image  $y$ , and thus  $s$  and the VSC image  $I$  with a local subdomain  $W \subset \Omega$ . The local conditional variance  $C_{s|\tilde{I}}(\mathbf{x})$  has two parameters we need to estimate: the scale parameter  $\sigma$  from the likelihood in eq. (9) and the window size  $a$  of the window function  $\omega$ . Since  $C_{s|\tilde{I}}(\mathbf{x})$  is varies spacially we compute its median value  $\hat{C}_{s|\tilde{I}}$ . In figure ?? we have plotted for various window sizes  $a$  the median conditional variance  $\hat{C}_{s|\tilde{I}}$  over the filter size  $\sigma$ . We can see that for window sizes  $a \leq 23$   $\hat{C}_{s|\tilde{I}}$  has minima at  $\sigma \approx 0$  while for larger window sizes  $a \geq 31$  it tends to be minimal at filtersizes  $\sigma > 6$ . Figure ?? show their optimum  $\sigma^*$  plotted over the window size  $a$ . We see that window sizes  $a < 21$  and  $A > 31$  lead to unrealistic scale differences  $\sigma^* \approx 0$  and  $\sigma^* \geq 6$ , since the actual difference in scale must be  $\sigma \approx 2$  judged by the resolutions of the VSC and the TC. This value is produced only at  $a = 23$  and  $a = 27$  and we choose  $a = 23$  since  $C_{s|\tilde{I}}(\mathbf{x})$  is smaller compared to the case  $a = 27$ .

In figure ?? we show the resulting optical flow for different region of interests (roi). Figures 2a and 2f show the resulting optical flow  $d$  which match the corresponding VSC image  $I$  and TC image  $y$  in the table. Fow each roi we computed the joint histogram  $p(y, I)$  (Figures 2b and 2g). In figure 2b  $p(y, I)$  has two isolated maxima which is sufficient for for a linear relationship between  $y$  and  $I$ . In figure 2g the linearity is obstructed to a minor degree since the TC image in figure 2j has a slight structural difference in the lower left corner compared to figure 2h.

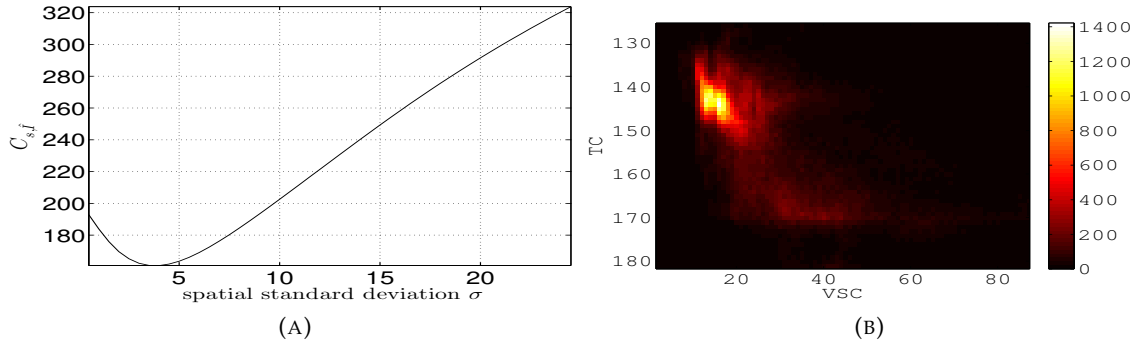


FIGURE 1. 1a: Dependence of  $C_{s|\tilde{I}}$  on the scaling parameter  $\sigma$ . 1b: Joint Histogram  $p(y, I)$  of the TC and smoothed VSC image pair  $y$  and  $\tilde{I}$  at the optimum  $\sigma^* = 4$ , the scale at which  $y$  and  $\tilde{I}$  are maximally linear.

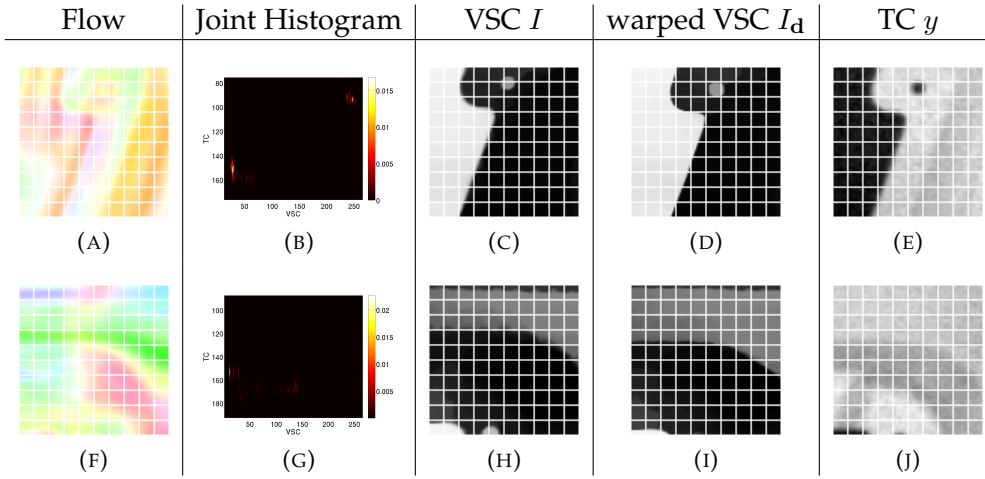


FIGURE 2. Multimodal Optical Flow: The resulting flow  $d$ , VSC image  $I$ , the warped VSC  $I_d$ , the TC image  $y$  as well as the joint histogram  $p(y, I)$  are shown for different region of interests. We can observe that the boundaries of the flow are blurred. This comes from the window function  $\omega$  in the local likelihood. The joint likelihood  $p(y, I)$  was evaluated only for the roi's. It has at most two maxima, which suffices to constitute a linear relationship between  $y$  and  $I$ . A grid is overlaid on the roi's for  $I$ ,  $I_d$  and  $y$  with 10 pixels per element to visually asses the quality of the flow. We can see the larger features are correctly matched, while smaller features are matched in a suboptimal fashion

??