Parameter Estimation in Image Processing and Computer Vision

Christoph S. Garbe and Björn Ommer

Abstract Parameter estimation plays a dominant role in a wide number of image processing and computer vision tasks. In these settings, parameterizations can be as diverse as the application areas. Examples of such parameters are the entries of filter kernels optimized for a certain criterion, image features such as the velocity field, or part descriptors or compositions thereof. Subsequently, approaches for estimating these parameters encompass a wide range of techniques, often tuned to the application, the underlying data and viable assumptions. Here, an overview of parameter estimation in image processing and computer vision will be given. Due to the wide and diverse areas in which parameter estimation is applicable, this review does not claim completeness. Based on selected key topics in image processing and computer vision we will discuss parameter estimation, its relevance, and give an overview over the techniques involved.

1 Introduction

In most image processing and computer vision tasks, one starts off with visual data such as an image or a sequence thereof. Of course, more general modalities are also becoming increasingly more common. These include spectral image sequences for example in satellite remote sensing or volumetric or even spectral volumetric time series, for example in state-of-the-art medical imaging devices. Cheap consumer devices are also transitioning beyond simple 2D capturing apparatuses. The Microsoft Kinect device, which captures intensity images and the scene depth at the same time, is an example of this. Also, first consumer grade cameras that capture light fields are on the horizon. All of these devices require adapted image processing

C.S. Garbe (✉) · B. Ommer
Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg,
Speyerer Str. 6, 69115 Heidelberg, Germany
{Christoph.Garbe,Bjoern.Ommer}@iwr.uni-heidelberg.de

H.G. Bock et al. (eds.), *Model Based Parameter Estimation*, Contributions in Mathematical and Computational Sciences 4, DOI 10.1007/978-3-642-30367-8_15,
© Springer-Verlag Berlin Heidelberg 2013

and computer vision algorithms. Common to all of them is the classical inverse problem they lead to: n-dimensional data is acquired and some model parameters need to be estimated to best describe the data. Of course, the model will strongly depend on the application, as will the metric in which ”best” is described by some sort of optimality condition. In this contribution, some common problems in image processing and computer vision will be described. Common methodologies for solving the resulting parameter estimation problem are presented.

This article is organized into two main parts, the first focusing on parameter estimation in low-level image processing, the second on high-level computer vision. In Sect. 2.1, parameter optimization for filter kernels will be introduced. The estimation of image motion or optical flow is outlined in Sect. 2.2. The reconstruction of images and optical flow fields is discussed in Sects. 2.4 and 2.5 respectively. The estimation of confidence and situation measures for optical flows is presented in Sect. 2.3. The segmentation of images based on their underlying motion is touched upon in Sect. 2.6. A combination of such approaches in a single functional is presented in Sect. 2.7 which is concerned with joint estimation of optical flow, segmentation and denoising of image sequences.

Parameter estimation in high level computer vision is presented in Sect. 3. Central to the analysis are key modeling decisions which are explained in Sect. 3.1. In Sect. 3.2 a compositional approach to object categorization is presented. The problem of object detection in cluttered scenes is discussed in Sect. 3.3.

2 Low-Level Image Processing

2.1 Optimization of Filter Kernels

A fundamental operation in image processing represents the filtering of intensity images. Such filtering is used for computing derivative filters of first order for motion estimating and for edge detection, and second or higher order for feature extraction of curvature information. Filter design is a well-established area in time-series signal processing and subject of standard textbooks [103, 107, 113]. The extension from 1-D signal processing to image processing is not trivial, however. This is largely due to uncertainty of design criteria for higher-dimensional signals and much more involved mathematical problems.

For edge detection and motion estimation, the computation of precise gradients of image intensities is vital. It can be show that the highly accurate computation of both, orientation and magnitude of gradients, are not feasible. Therefore, design choices have to be made. Very often, subspace problems, which are orthogonal to the gradient directions have to solved. Hence the precise direction of the gradients is more important that their magnitude. Making such a design choice leads to the formulation of a optimization problem, yielding the appropriate filter kernels. These filters are discretized by finite differences using convolution kernels optimized with respect to the model assumptions, scales and/or noise present in the data.

2.2 Optical Flow Estimation

For the estimation of motion from digital image sequences, a number of different techniques has been proposed. Generally, they can be categorized into four groups:

1. The class of *gradient based* techniques relies on computing spatio-temporal derivatives of image intensity, which can either be of first order [46,65] or second order [93,115].
2. *Region-based matching* may be employed when under certain circumstances (aliasing, small number of frames, etc.) it is inappropriate to compute derivatives of grey-values. In this approach the velocity is defined as a shift giving the best fit between image regions at different times [7,55,80].
3. The *energy-based methods* rely on the output energy of velocity-tuned filters. These methods are often referred to as *frequency-based methods* owing to their design in the Fourier domain [1,50,62].
4. Another class of methods is called *phase-based*, because velocity is defined in terms of phase behavior of band-pass filter output and phase information is used to estimate the optical flow [51,122].

Overviews of these estimators including error analysis can be found in [11, 12, 60]. One widely used technique to estimate the local optical flow $v(t,x)$ corresponding to an image sequence $u : [0,1] \times \Omega \to \mathbb{R}$ on an image domain $\Omega \subset \mathbb{R}^n$ $(n = 2,3)$ is the *first order gradient based* approach [46,65]. Together with phase based techniques [51], this approach offers the best performance with respect to accuracy [11,53]. Here, constancy of gray values $u(t, x(t))$ along trajectories $x(t)$ is assumed, leading to the constraint equation

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} u(t, x(t)) = \nabla u(t, x(t)) \cdot v(1, x(t)) + \partial_t u(t, x)\,. \tag{1}$$

This constraint equation is generally known as the brightness change constraint equation (BCCE). The BCCE gives us one constraint for two unknowns for image sequences or three unknowns for volume sequences. Thus it is an ill posed problem, which is also known as the aperture problem and only the component of the velocity orthogonal to gray-value structures can be computed from (1). Let us assume that v is at least locally constant. One approach to solve the aperture problem for locally constant v was presented by Guichard [59]. The aperture problem can also be solved *locally* with the Lucas–Kanade approach [82,83] or with the *structure tensor approach* [21] which minimizes the local energy functional

$$\int_0^1 \int_\Omega w(x-y, t-s)\,(\nabla u(s,y) \cdot v(s,y) + \partial_t u(s,y))^2\,\mathrm{d}y\,\mathrm{d}s\,. \tag{2}$$

Here $w(\cdot\,,\cdot\,)$ is a window function, indicating the local spatio-temporal neighborhood. These local approaches offer relatively high robustness with respect to noise

Fig. 1 On a traffic scene plotted motion vectors indicate regions, where a local flow estimater achieves reliable results. For significantly large regions the velocity can not be computed

and allow for a computation of confidence and type measures, which characterize the quality of estimates. Generally, they do not lead to dense flow fields (cf. Fig. 1). However, *global* variational estimators as discussed below lead by design to fully dense flow fields but are known to be more sensitive to noise.

2.2.1 Global, Variational Methods for Optical Flow Estimation

The study of variational methods in optical flow estimation started with the classical work of Horn and Schunk in 1984 [65]. They considered minimizers of the energy functional

$$E[v] = \int_{\Omega} [(\partial_t u, \nabla u) \cdot (1, v)]^2 + \alpha |\nabla v|^2 \delta x \tag{3}$$

acting on image intensities $u : [0, 1] \times \Omega \to \mathbb{R}$ at decoupled time steps $t \in [0, 1]$. Thereby they implicitly assume the optical flow field $v : [0, 1] \times \Omega \to \mathbb{R}^n$ to be spatially smooth. Here, the scalar α denotes the constant weighting parameter of the regularizer. Nagel and Enkelmann [94] replaced the second, regularizing term $\alpha |\nabla v|^2$ by a quadratic form

$$\int_{\Omega} \nabla v \cdot J^{\sigma}[u] \nabla v \delta x \tag{4}$$

which involves the local structure tensor $J^{\sigma}[u]$ of the image (σ indicates the involved filter width) and allows for a significant change in v across image edges indicated by steep image gradients. Alternatively, one can replace the quadratic regularization by a BV type regularization $\int_{\Omega} |\nabla u| \delta x$ as presented by Cohen [36] or other convex regularizers as considered by Weickert and Schnörr [123]. Connections to shape optimization have been exploited by Schnörr [108]. Weickert et al. [26] proposed a

combination of local flow estimation and global variational techniques to combine the benefits of robustness and dense field representation, respectively. A broader comparison of different regularization techniques involving quasi-convex functionals has been given by Hinterberger et al. [63]. In particular they consider $W^{1,p}$-approximation or BV type functionals. Applying multi-grid methods in the solution of the Euler Lagrange equations for the above combined global-local method was presented by Bruhn et al. [25]. In all these approaches the choice of the regularization terms basically determines the class of admissible flow fields. A rigorous analysis of assumptions on the flow field under which global minimizers of the non regularized variational problem can be found was given by Lefébure and Cohen [76].

In general, numerical algorithms for the minimization tend to get stuck in local minima. Alvarez et al. [84] proposed to consider a scale space approach to solve this problem. Starting to correlate coarse representations of subsequent images in an image sequence via an optical flow field or a deformation, one proceeds on successively finer representation until the actual fine scale images are properly matched.

2.3 Confidence and Situation Measures for Optical Flows

In order to detect artifacts and erroneous flow vectors in optical flow fields, one can analyze confidence and situation measures. Confidence measures are used to estimate the correctness of flow fields, based on information derived from the image sequence and/or the displacement field. Based on proposed techniques, two kinds of confidence measures can be distinguish: situation and confidence measures. Situation measures are used to detect locations, where the optical flow cannot be estimated unambiguously. This is contrasted by confidence measures, which are suited for evaluating the degree of accuracy of the flow field based. Situation measures can be applied e.g., in image reconstruction [87], to derive dense reliable flow fields [110] or to choose the strength of the smoothness parameter in global methods (e.g., indirectly mentioned in [72]). Confidence measures are important for quantifying the accuracy of the estimated optical flow fields. A successful way to obtain robustness to noise in situation and confidence measures is also discussed in [70].

Confidence measures employed are generally chosen as innate to the flow estimation technique. By combining flow methods with non-inherent confidence measures [70] were able to show considerable improvements for confidence and situation measures. Altogether the results of the known measures are only partially satisfactory as many errors remain undetected and a large number of false positive error detections have been observed. Based on a derived optimal confidence map they obtain the results in Fig. 2 for Lynn Quam's Yosemite sequence [61], and the Street [90] test sequences.

For variational methods, the inverse of the energy after optimization has been proposed as a general confidence measure in [27]. For methods not relying on global smoothness assumptions, e.g., local methods, a new confidence measure based on linear subspace projections was proposed in [69]. The idea is to derive a

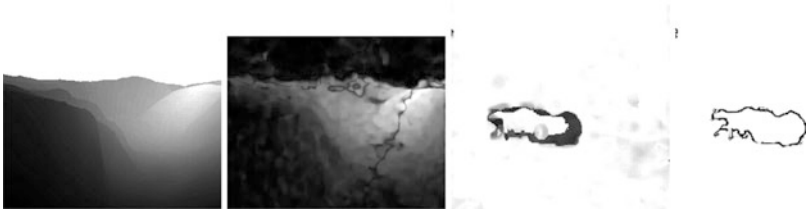

Fig. 2 Comparison of optimal confidence measure (*left*) to best known confidence measure (*right*) for Yosemite and Street sequences

spatio-temporal model of typical flow field patches using e.g., principal component analysis (PCA). Using temporal information the resulting eigenflows represent complex temporal phenomena such as a direction change, a moving motion discontinuity or a moving divergence. Then the reconstruction error of the flow vector is used to define a confidence measure.

2.4 Image Restoration

An active field of research is the restoration of damaged paintings or, in case of digital images, the reconstruction of blank image regions based on image information outside this area. It was first proposed by Masnou and Morel [87] and named "disocclusion." The term "inpainting" was introduced by Bertalmio et al. [17], Ballester et al. [9] proposed a variational approach based on the continuation of contours of equal luminance in the image, also called isophote lines. A variational approach based on level set perimeter and mean curvature was presented by Ambrosio and Masnou in [3]. Other approaches have been proposed for image inpainting, e.g., TV-inpainting and curvature-driven diffusion inpainting suggested by Chan and Shen [34, 35].

In general the problem of inpainting is stated as follows: Given an image $u_0 : \Omega \to \mathbb{R}$ and an inpainting domain $D \subset \Omega$, one asks for a restored image intensity $u : \Omega \to \mathbb{R}$, such that $u|_{\Omega \setminus D} = u_0$ and $u|_D$ is a suitable and regular extension of the image intensity u_0 outside D. The simplest inpainting model is based on the construction of a harmonic function u on D with boundary data $u = u_0$ on ∂D. This model is equivalent to the minimization of the Dirichlet functions $E_{\text{harmon}}[u] = \frac{1}{2}\int_D \|\nabla u\|^2 \delta x$ for given boundary data, as can be derived from Dirichlet's principle. Due to standard elliptic regularity the resulting intensity function u is smooth inside D. This means that edge information present on the boundary will not be restored in the inpainted area. An overview on first order variational functionals related to this problem has been given by Chan and Shen [31, 32]. To resolve this shortcoming, TV-type inpainting models have been proposed [29, 30]. They are based on the functional $E_{\text{TV}}[u] = \frac{1}{2}\int_D \|\nabla u\| \delta x$, which allows for steep transitions on some edge contour. The resulting image intensity is a Bounded Variation (BV)

function and thus characterized by jumps along rectifiable edge contours. In a weak sense, it solves the geometric PDE $h = 0$ with Neumann boundary conditions, where $h = \operatorname{div}(\|\nabla u\|^{-1}\nabla u)$ is the mean curvature on level sets or edge contours. Thus the resulting edges will be straight lines.

This issue of straight lines has been overcome in later work by [30,33]. Here, the functional of the energy is based on the curvature of the intensity level curves. This enforces a smooth transition between the level curves. The equations obtained from such models are highly nonlinear and of higher (fourth) order. Recently, Bredies et al. [24] proposed an approach for higher order TV which significantly improves on stair-casing effects often found in TV regularization. Such an approach has been applied to denoising depth images of Time-of-Flight (ToF) imagers [78].

In many applications the assumption of a sharp boundary ∂D of the corrupted region turns out to be a significant restriction. In fact the reliability of the given image intensity gradually deteriorates from the outside to the inside of the inpainted region. This can be reflected by a relaxed formulation of the variational problem. One considers the functional

$$\mathrm{e}^{\epsilon}[u] = \int_{\Omega} |u - u_0|\, H_{\epsilon} + \lambda(1 - H_{\epsilon})\|\nabla u\|^{p}\delta x\,, \tag{5}$$

where $\lambda > 0$, $p = 1$ or 2, and H_{ϵ} is a convoluted characteristic function χ_D and $\epsilon > 0$ indicates the width of the convolution kernel.

Frequently, one aims for a better continuation of image structures from outside the destroyed region. Mumford [92] phrased this problem in terms of a minimization of elastic energy of curves or surfaces $\mathcal{M}$ treated as elastic rods or shells and C^1 continuity conditions on ∂D. Morel and Masnou [86, 88] have further exploited the relation to a minimization of the Willmore functional

$$E[\mathcal{M}] = \frac{1}{2}\int_{\mathcal{M}} h^2\,\mathrm{d}a, \tag{6}$$

where h is the mean curvature of the manifold $\mathcal{M}$. They explicitly construct continuations of level lines in an occluded image region D as parameterized minimizers of the Willmore energy. Ambrosio and Masnou [4] revisited this problem in the context of geometric measure theory and derived minimizers of an implicit formulation of the Willmore energy

$$E[u] = \int_{D} \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right)^2 |\nabla u|\delta x. \tag{7}$$

General properties of such variational problems have been discussed by Bellettini et al. in [13]. Chen et al. [35] presented a finite difference relaxation algorithm for this Willmore functional and applied it to image inpainting. Esedoglu and Shen [44] proposed a phase field approximation of the Willmore energy and used a parabolic relaxation in their concrete minimization algorithm. Bertalmio et al. [16] and Ballester et al. [9] relaxed the Willmore functional and studied the simultaneous

extension of the image's normal field $\theta = \frac{\nabla u}{|\nabla u|}$ and the intensity u. This resulted to the minimization of the energy

$$E[\theta, u] = \int_D (|\nabla_{(t,x)} u| - \theta \cdot \nabla u) + (\operatorname{div} \theta)^p (a + b|\nabla G * u|) \delta x \,, \tag{8}$$

where G is a Gaussian smoothing kernel. For θ they impose the constraint $|\theta| \leq 1$. Obviously, the first energy integrant is zero if θ coincides with the image normal. Thus, for $p = 2$ the second term approximates the Willmore energy on the ensemble of level sets on D. Dirichlet boundary conditions for u and θ are assumed.

A different approach has been considered by Bertalmio, Bertozzi and Sapiro [18] connecting fluid dynamics and image inpainting. A vorticity formulation of the stationary Navier–Stokes equations with zero viscosity can be written in terms of the stream function Ψ

$$\nabla^\perp \Psi \cdot \nabla \Delta \Psi = 0 \,, \tag{9}$$

where $\nabla^\perp$ denotes the orthogonal gradient and $v = \nabla^\perp \Psi$ the velocity field of the flow. The connection to image processing is drawn by replacing Ψ by the image intensity u. One tries to find a solutions of the above equation under boundary conditions for the intensity u and the direction of level lines $\nabla^\perp u$. In terms of physics, this can be thought of as the solution to the transport of the outside image into the hole D which solves the stationary Navier–Stokes equations. The corresponding algorithm is based on a third order, parabolic relaxation

$$\partial_t u - \nabla^\perp u \cdot \nabla \Delta u = 0 \,. \tag{10}$$

Recently, texture inpainting has attracted attention. Bertalmio et al. [19] proposed a technique to first decomposed the image into texture and structure and then propagated into the inpainting domain both these classes in different ways. This idea to decompose texture and structure is also applied in [58]. Some statistical approaches have been presented in [40] to perform a texture synthesis and structure propagation.

2.5 *Optical Flow Reconstruction*

In Sect. 2.2 a number of different approaches for estimating optical flow have been presented. Even the most advanced techniques cannot accurately estimate correct flow fields under all conditions. A powerful tool for detecting and removing incorrect flow vectors from the flow field are confidence measures [27, 70]. Discarting erroneous flow vectors results in accurate but sparse motion fields. However, many applications require dense flow fields. This reconstruction of missing vectors in optical flow fields is based on information from the surrounding areas is addressed in this section. The tasks is similar to that addresses in the previous section for image reconstruction, where it was called "inpainting."

The reconstruction of motion fields has lately been proposed in the field of video completion. In case of large holes with complicated texture previously used methods are often not suitable to obtain good results. Instead of reconstructing the frame itself by means of inpainting, the reconstruction of the underlying motion field allows for the subsequent restoration of the corrupted region even in difficult cases.

Hence, the reconstruction of motion fields called "motion inpainting" was first introduced for video stabilization by Matsushita et al. in [89]. The idea is to continue the central motion field to the edges of the image sequence where the field is lost due to camera shaking. This is done by a basic interpolation scheme between four neighboring vectors and a fast marching method. An extension of inpainting to higher dimensional surfaces has also been presented [10].

The reconstruction of optical flow fields can be accomplished by a simple extension of these inpainting functionals for images, e.g., TV-inpainting on two dimensional vector fields. However, these methods sometimes fail in situations where the course of the motion boundary is unclear, e.g., if round motion boundaries or junctions occur. Since image edges often correspond to motion edges the information drawn from the image sequence can be important for the reconstruction, especially in such cases where the damaged vector field does not contain enough information to uniquely determine the optical flow of motion boundaries.

Hence, in the special case of optical flow, the image sequence provides a source of information in addition to the corrupted vector field, which can be used to guide the reconstruction process in ambiguous cases. Optical flow fields have been used for the reconstruction of images in [15]. The resulting functional is nonlinear and can be minimized by means of the finite elements method. This techniques compares favorably to diffusion based and TV inpainting methods, see Fig. 3.

2.6 Motion Segmentation

A common task in image processing is the segmentation of images. Image segmentation in its typical form is the process of assigning a label to every image pixel in a way that pixels with the same label share certain visual characteristics. Image segmentation is closely related to the task of classification. For classification, one tries to assign to each pixel in the image a label or object class, where the classes are agreed in advance. These labels can also be probabilities of the pixel to belonging to certain objects. Depending on the cues that distinguish the object of interest from the background, segmentation can be based on features such as edge information, intensity, color, texture or motion. Well known approaches are:

- Variational Approach (Mumford and Shah functional [91], Geodesic active contours [68], Segmentation from motion[37–39]).
- Multi-resolution techniques (Pyramid linking [104], Wavelet coefficient analysis [73, 79, 116]).

The segmentation from motion can be achieved by *iterative algorithms* based on interleafed motion estimation steps and segmentation steps [111]. Well known

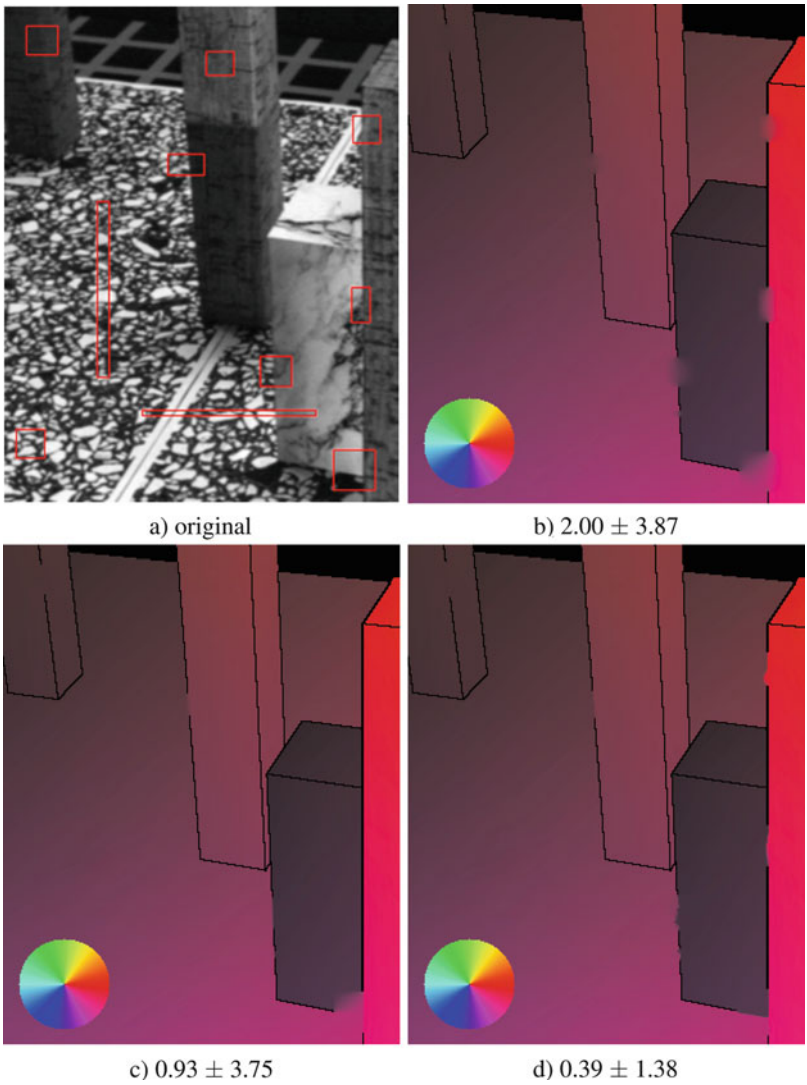

Fig. 3 Comparison of the proposed inpainting algorithm to diffusion and TV inpainting; the numbers indicate the average angular error within the corrupted regions after reconstruction; (**a**) Original corrupted Marble sequence, (**b**) Reconstruction result of diffusion based motion inpainting, (**c**) Reconstruction result of TV based motion inpainting, (**d**) Reconstruction result of image based motion inpainting. Taken from [15]

techniques are furthermore a *layered representation* of image sequences [119–121], *variational approaches*, for instance motion competition [38], and other techniques such as *tensor voting* [95,96] or *algebraic methods for multi-body motion models* [117]. The iterative algorithm introduced by [111] describes a probabilistic relaxation framework for robust multiple motion estimation and segmentation.

[119–121] present a set of techniques for segmenting images into coherently moving regions in a layered representation. Cremers and Soatto [38, 39] gave an extension of the Mumford and Shah functional from intensity segmentation to motion based segmentation in terms of a probabilistic framework implemented by level sets. The geometric prior favors motion boundaries of minimal length and the likelihood term takes into account the orthogonality between the spatio-temporal gradient and the velocity vector. The classical pyramid linking segmentation, as discussed in [28, 105], is based on the Gaussian pyramid of the image. Several improvements were introduced by [104]. Instead of using a Gaussian pyramid, they propose a continuous scale-space. The segmentation by clustering in the feature space has been described by [116]. The feature extraction is based on local variance estimates of the wavelet coefficients of the image.

2.7 *Joint Estimation of Optical Flow, Segmentation and Denoising*

Rather than denoising images, computing optical flow and performing a segmentation step, all separately, all these components can be combined and computed concurrently [106, 114]. This approach is based on an extension of the well known Mumford–Shah functional which originally was proposed for the joint denoising and segmentation of still images. Given a noisy initial image sequence $u_0 : D \to \mathbb{R}$ on the space-time domain $D = [0, T] \times \Omega$ the following energy is considered

$$E_{\mathrm{MSopt}}[u, w, S] = \int_D \frac{\lambda_u}{2}(u - u_0)^2 \,\mathrm{d}\mathcal{L} + \int_{D\setminus S} \frac{\lambda_w}{2}\left(w \cdot \nabla_{(t,x)} u\right)^2 \,\mathrm{d}\mathcal{L}$$
$$+ \int_{D\setminus S} \frac{\mu_u}{2}\|\nabla_{(t,x)} u\|^2 \,\mathrm{d}\mathcal{L} + \int_{D\setminus S} \frac{\mu_w}{2}\|P_\delta[\zeta]\nabla_{(t,x)} w\|^2 \,\mathrm{d}\mathcal{L} + \nu\, \mathcal{H}^d(S)$$

for a piecewise smooth de-noised image sequence $u : D \to \mathbb{R}$, and a piecewise smooth motion field $w = (1, v)$ and a set $S \subset D$ of discontinuities of u and w. λ_u and λ_w are the weighting factors for the fidelity terms of u and w, while μ_u and μ_w are those for the smoothness terms respectively. The first term models the fidelity of the denoised image-sequence u, the second term represents the fidelity of the flow field w in terms of the optical flow equation (1). The smoothness of u and w is required on $D\setminus S$ and finally, the last term is the Hausdorff measure of the set S. A suitable choice of the projection $P_\delta[\zeta]$ leads to an anisotropic smoothing of the flow field along the edges indicated by ζ [106].

The model is implemented in [106, 114] using a phase-field approximation in the spirit of Ambrosio and Tortorelli's approach [5]. Thereby the edge set S is replaced by a phase-field function $\zeta : D \to \mathbb{R}$ such that $\zeta = 0$ on S and $\zeta \approx 1$ far from S. As in the original Ambrosio–Tortorelli model, a scale parameter ϵ controls the thickness of the region with small phase field values. The Euler–Lagrange equations of the corresponding parameters yield a system of three partial differential equations for the image-sequence u, the optical flow field v and the phase field ζ:

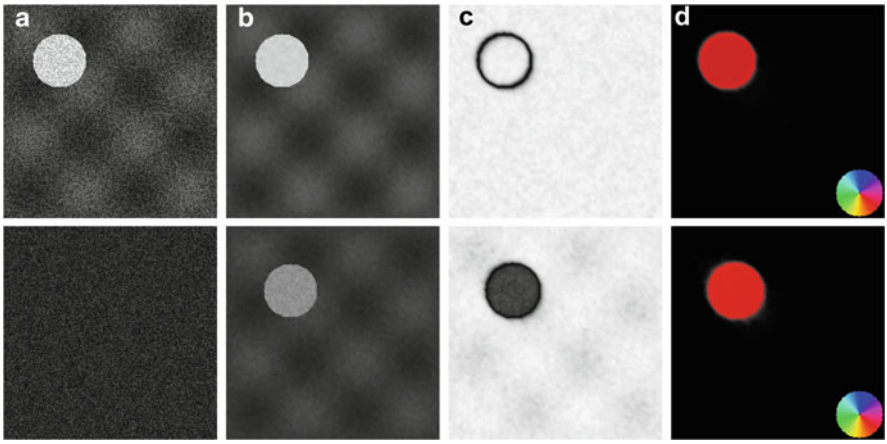

Fig. 4 Noisy test sequence: From top to bottom frames 9 and 10 are shown. (**a**) original image sequence, (**b**) smoothed images, (**c**) phase field, (**d**) estimated motion (*color coded*). Taken from [114]

Fig. 5 Pedestrian video: frames from original sequence (*left*); phase field (*middle*); optical flow, color coded (*right*) [114]

$$
\begin{aligned}
-\mathrm{div}_{(t,x)}\left(\frac{\mu_u}{\lambda_u}(\zeta^2+k_\epsilon)\nabla_{(t,x)}u+\frac{\lambda_w}{\lambda_u}w(\nabla_{(t,x)}u\cdot w)\right)+u &= u_0\\
-\epsilon\Delta_{(t,x)}\zeta+\left(\frac{1}{4\epsilon}+\frac{\mu_u}{2\nu}\|\nabla_{(t,x)}u\|^2\right)\zeta &= \frac{1}{4\epsilon}\\
-\frac{\mu_w}{\lambda_w}\mathrm{div}_{(t,x)}\left(P_\delta[\zeta]\nabla_{(t,x)}v\right)+(\nabla_{(t,x)}u\cdot v)\nabla_{(x)}u &= 0
\end{aligned}
\tag{11}
$$

Neumann boundary conditions are considered in this application. For details on this approximation and its discretization, we refer to [106].

In Fig. 4 we show results from this model on a noisy test-sequence where one frame is completely missing. But this does not hamper the restoration of the correct optical flow field shown in the fourth column, because of the anisotropic smoothing of information from the surrounding frames into the destroyed frame.

Furthermore, in Fig. 5 we consider a complex, higher resolution video sequence showing a group of walking pedestrians. The human silhouettes are well extracted

and captured by the phase field. The color–coded optical flow plot shows how the method is able to extract the moving limbs of the pedestrians.

3 High-Level Computer Vision

Object recognition in images and videos poses one of the long standing key challenges of computer vision. The problem itself is twofold since recognition involves localizing instances of an object class in novel, cluttered scenes (detection) and classifying these instances as belonging to one of several potential classes (categorization). Developing appropriate object models, which represent the appearance and geometry of each object class and thereby help to distinguish objects from another, constitutes the central problem of recognition. It is common practice to automatically learn these models from unsegmented training images [47, 56, 99], from bounding box segmentations [85, 100], from segmented training images [77], or from manually annotated training images [45]. Since the complexity of object models has to scale with the complexity of object classes, object detection and categorization become particularly challenging when object categories are featuring large intra-class variability. Additionally, the complexity of this problem depends on several other factors, such as the level of supervision during training, the between-class similarity, and the constraints that can be imposed on scenes (e.g., constraints on variation in scale or viewpoint).

3.1 Key Modeling Decisions

Visual recognition can be pursued on different levels of semantic granularity. One extreme strategy is exemplar detection (e.g., [81]), where exactly the same query object is sought in scenes with different environmental conditions such as background, lighting, occlusion, viewpoint etc. The other extreme is category-level object recognition where all instances of a category are to be recognized. Therefore, the granularity of the set of categories controls the complexity of the recognition task as it defines the within-class variability. Influential papers such as [6, 47] have focused research in the field of category-level object recognition on principled probabilistic object models with semi-local feature descriptors. The general goal is to represent objects by learning local appearance features and their spatial configuration and comprising both in a common model. Within this coarse fundamental modeling framework, the current approaches to object categorization can be characterized by the core modeling decisions they make.

1. *Local Descriptors:* A classical way to capture image region information are *appearance patches*, i.e., subsampled image patches that are vector quantized into a large codebook (e.g.,[2, 47, 77]). Complex edge histogram features such as *SIFT* features [81] are another popular choice. In [98] we have proposed a low

dimensional representation of image patches that is based on compact local edge and color histograms of subpatches. Another popular descriptor is *geometric blur* [14]. This feature weights edge orientations around a feature point using a spatially varying kernel. Moreover, edge contour based methods have been proposed in [48, 102]. Opelt et al. [102] extract curve fragments from training images and they apply Adaboost to learn strong object detectors.

2. *Spatial Model:* A second choice concerns the model that combines all local features with their spatial distribution to represent object shape. It should be emphasized that this notion of shape is not based on the object boundary but on the geometry of object parts that are distributed all over the object. Object models have to deal with two problems, simultaneously. On the one hand individual local appearance descriptors in a test image are to be matched against those from a learned model. On the other hand the co-occurrence and the spatial relations between individual features have to be taken into account to represent the global object geometry. The simplest approach is, therefore, to histogram over all local descriptors found in an image (e.g., [41]) and to categorize the image directly based on the overall feature frequencies. On the one hand such *bag of features* methods offer robustness with respect to alteration of individual parts of an object (e.g., due to occlusion) at low computational costs. On the other hand they fail to capture any spatial relations between local image patches and they often adapt to background features. By making the restricting assumption that the spatial structure of objects is limited in its variation with respect to the image, Lazebnik et al. [75] can improve the performance of the bag of features approach using a spatially fixed grid of feature bags. At the other end of the modeling spectrum we find *constellation models*: Originally, Fischler and Elschlager [49] have proposed a spring model for coupling local features. Inspired by the *Dynamic Link Architecture* for cognitive processes, Lades et al. [71] followed the same fundamental idea when proposing their face recognizer. Lately increasingly complex models for capturing part constellations have been proposed [47]. However the complexity of such a joint model of all parts causes only small numbers of parts to be feasible. To incorporate larger numbers of parts, [2, 77] use a simpler object model and a comparably large codebook of distinctive parts. Leibe and Schiele [77] use a probabilistic Hough voting strategy to distinguish one category from the background. In [99] we advance the idea of large numbers of parts by grouping parts prior to spatially coupling the resulting compositions in a graphical model. Conflicting categorization hypotheses proposed by compositions and the spatial model are then reconciled using probabilistic inference in the underlying Bayesian network. The processing pipeline for this automatic scene analysis is presented in Fig. 6. Finally, Berg et al. [14] describe and regularize the spatial distortion resulting from matching an image to a training sample using thin plate splines.

3. *Hierarchies:* For a long time, research on object recognition has aimed at building hierarchical models [52]. Despite this effort, many popular current methods such as [41, 47, 77] are single layered. Recently, *probabilistic latent semantic analysis* (pLSA) [64] has become popular (e.g.,[109]), where a hidden

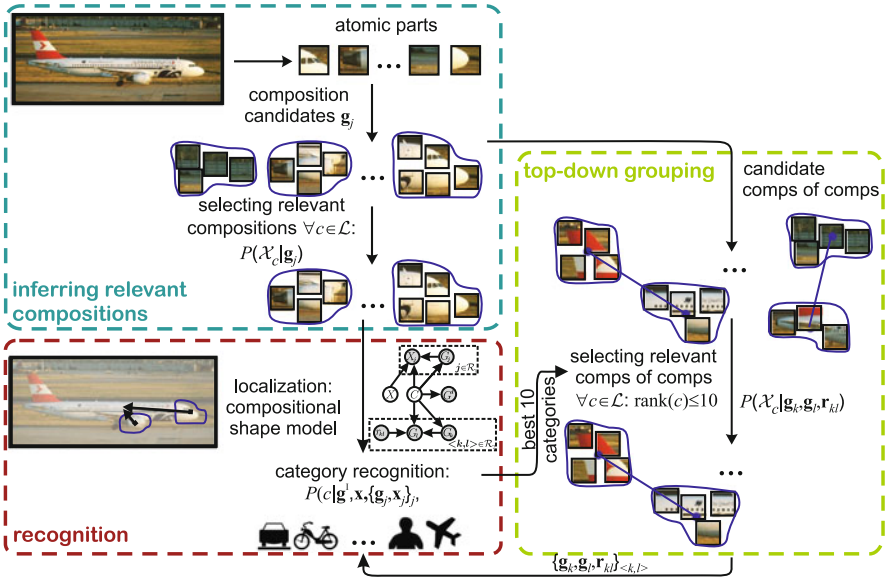

Fig. 6 Processing pipeline for automatic scene analysis. Key steps: Feature extraction, perceptual grouping to form compositions, selection of relevant compositions, object localization and recognition, and top-down grouping to form compositions of compositions which then yield an update of object hypotheses

representation layer of abstract concepts is introduced. Other examples for hierarchical approaches are the feature hierarchies of [43], the hierarchical parts and structure model of [23] or the deep compositional hierarchies of [101].

4. *Learning Paradigm:* Another modeling decision is related to the learning paradigm, i.e.,pursuing a generative versus a discriminative approach. Generative models have been very popular in the vision community, e.g., [2,14,22,47,77,98, 112]. They naturally establish correspondences between model components and image features. Discriminative approaches are for instance [41] and [118]. To recognize faces in real-time Viola and Jones [118] use boosting to learn simple features which are based on local intensity differences.
5. *Degree of Supervision:* Similar to the influential paper by Fergus et al. [47] several other approaches (e.g.,[2, 41, 99]) have been proposed that only need training images (showing objects and even background clutter) and the overall category label of an image. The restriction of user assistance is desirable for scaling methods up to large numbers of categories with large training sets. A system that can be trained in an unsupervised manner is for instance that of [56], whereas Felzenszwalb and Huttenlocher have taken a supervised approach to object detection in [45]. Furthermore, Jin and Geman [67] present a compositional architecture with manually built structure for license plate reading. In their conclusion they emphasize the complexity of the future challenge of learning such a compositional model. In [97] we have addressed exactly this problem in the even less constraint case of large numbers of natural object classes.

3.2 A Compositional Approach to Object Categorization

Despite the complexity of the recognition challenge, learning of object representations from a small number of samples is possible due to the *compositional nature* of our (visual) world. As Attneave [8] points out, the visual stimulus is highly redundant in the sense that there exist significant spatial interdependencies in visual scenes. *Compositionality* (cf. Geman's work [54]) serves as a fundamental principle in cognition and especially in human vision [20] that exploits these dependencies. It refers to the prominent ability of perception to represent complex entities by means of comparably few, simple, and widely usable parts. Additional information that is missing in the individual parts is added by incorporating relations between them. The compositional approach presented in [97] automatically learns characteristic compositions of atomic parts. A visualization of relevant compositions by clustering is shown in Fig. 7. Perceptual grouping is applied to obtain candidate compositions. The statistics of relevant compositions that are both reliable and discriminative are then learned from the training data. To avoid overfitting to spurious compositions, cross-validation is performed.

3.3 Object Detection in Cluttered Scenes

Object detection in cluttered natural scenes requires matching object models to the observations in the scene. Since the objective function for matching is a highly non-convex function over scale space, the task of finding good matches is an extremely challenging optimization problem. The two leading approaches to this problem are sliding windows, e.g.,[42, 118], and voting methods, which are based on the Hough transform [66]. Sliding windows scan over possible locations and scales, evaluate a binary classifier, and use post-processing such as non-max suppression to detect objects. The computational burden of this procedure is daunting although various techniques have been proposed to deal with the complexity issue, e.g.,cascaded evaluation [118], interest point filtering, or branch-and-bound [74]. In contrast to this, Hough voting [66] parametrizes the object hypothesis (e.g.,the location of the object center) and lets each local part vote for a point in hypothesis space.

In [100] we have shown that object scale is an inherently global property, which makes local scale estimates unreliable and, thus, leads to a *scale-location-ambiguity*. An illustration of this approach is presented in Fig. 8. When the Hough transform is extended to provide hypotheses for location *and* scale, each local feature casts votes that form lines through scale space rather than just a single point as in current voting methods [48, 77, 102]. Since all points on a voting line are statistically dependent, they should agree on a single object hypothesis rather than being treated as independent votes. Ideally, all points on an object would yield lines that intersect in a single point. Due to intra-category variation, and background clutter, the points of intersection are, however, degraded into scattered clouds. Finding these clusters

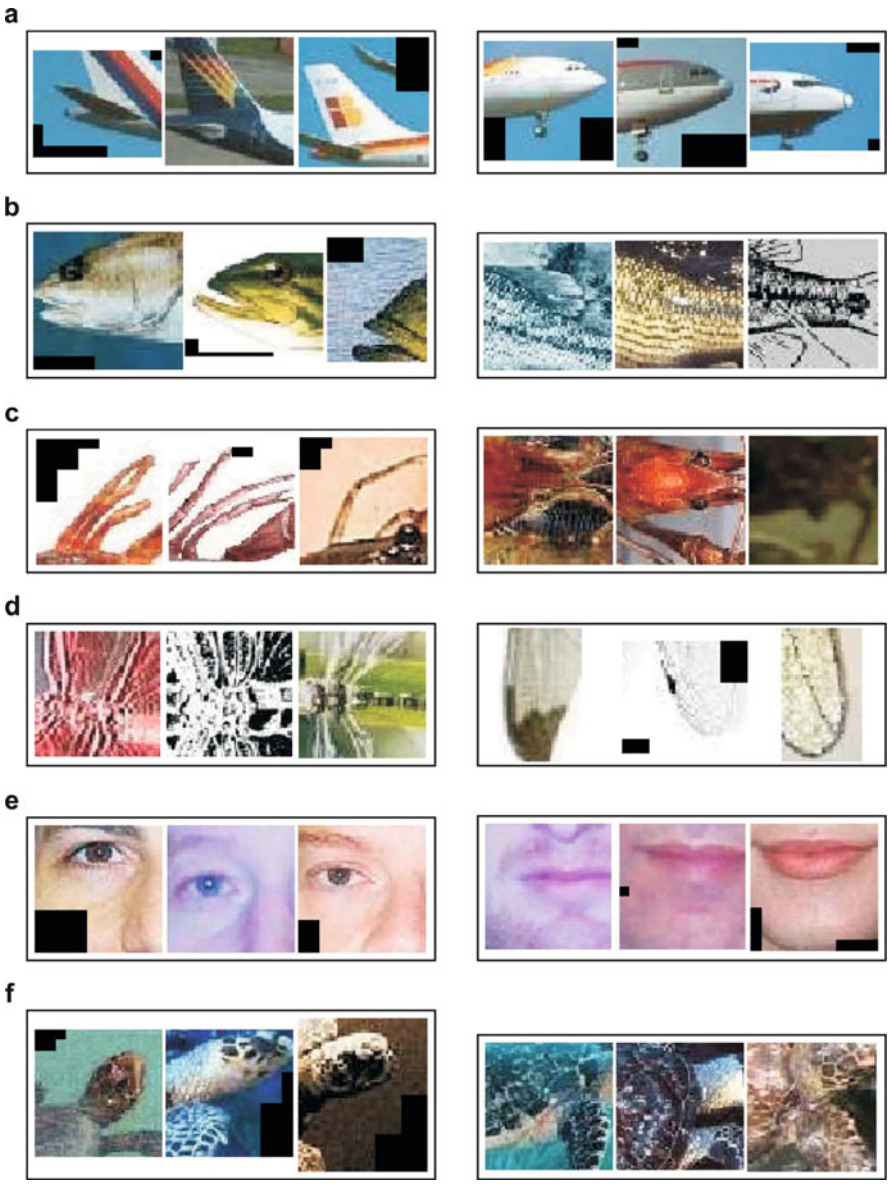

Fig. 7 Visualization of relevant compositions by clustering. For each category, the two prototypical compositions with highest relevance are illustrated by visualizing the closest compositions to that prototype. (**a**) airplanes, (**b**) bass, (**c**) crayfish, (**d**) dragonfly, (**e**) faces, and (**f**) hawksbill

becomes difficult since their number is unknown (it is the number of objects in the scene) and because the assignment of votes to objects is not provided (segmentation problem). To address these issues we frame the search for globally consistent object hypotheses as a weighted, pairwise clustering of local votes without scale

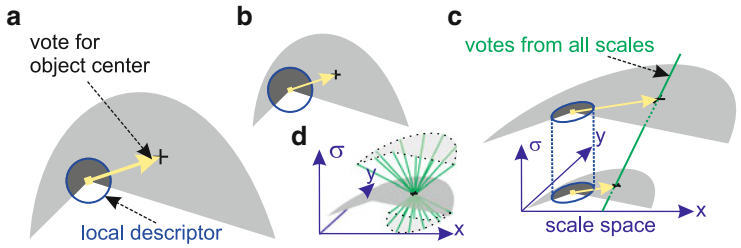

Fig. 8 Voting in scale space with scale-location-ambiguity. The *circle* indicates the spatial support of a local feature. Based on the descriptor, the difference of object scale between (**a**) and (**b**) is not detectable. Thus, a local feature casts votes for the object center on all scales, (**c**). These votes lie on a line through scale space, since the position of the center relative to a feature varies with object scale. Without noise, all voting lines from an object intersect in a single point in scale space, (**d**). For all other scales this point is blurred as indicated by the *dotted outline*

estimates. Clustering avoids a local search through hypothesis space [77] and the pairwise setting circumvents having to specify the number of objects ahead of time. Moreover, clustering voting lines deals with the large number of false positives [57] which point votes produce and that hamper the commonly used local search heuristics such as binning [77].

4 Conclusion

In this contribution, we have given a brief overview of some areas of image processing and computer vision in which parameter estimation plays a central role. We have not attempted to give a complete and exhaustive overview. However, we have outlined some of the approaches and techniques commonly used in the area and demonstrated them on typical applications. From our overview it becomes apparent that most parameter estimation problems in image processing and computer vision are highly non-linear. Inherent are often large amounts of image data and the requirement of fast computation times. Therefore, the presented applications would profit from algorithmic and conceptual advances.

References

1. Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. Journal of the Optical Society of America A 2(2):284–299
2. Agarwal S, Awan A, Roth D (2004) Learning to detect objects in images via a sparse, part-based representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(11):1475–1490
3. Ambrosio L, Masnou S (2003a) A direct variational approach to a problem arising in image reconstruction. Interfaces and Free Boundaries 5:63–81

4. Ambrosio L, Masnou S (2003b) A direct variational approach to a problem arising in image reconstruction. Interfaces Free Bound 5(1):63–81
5. Ambrosio L, Tortorelli VM (1992) On the approximation of free discontinuity problems. Boll Un Mat Ital B 6(7):105–123
6. Amit Y, Geman D (1998) A computational model for visual selection. Neural Computation 11(7):1691–1715
7. Anandan P (1989) A computational framework and an algorithm for the measurement of visual motion. International Journal of Computer Vision 2:283–319
8. Attneave F (1954) Some informational aspects of visual perception. Psychological Review 61(3):183–193
9. Ballester C, Bertalmio M, Caselles V, Sapiro G, Verdera J (2001a) Filling-in by joint interpolation of vector fields and gray levels. IEEE Transactions on Image Processing 10(8):1200–1211
10. Ballester C, Caselles V, Verdera J (2003b) Disocclusion by Joint Interpolation of Vector Fields and Gray Levels. Multiscale Modeling & Simulation 2(1):80, DOI 10.1137/S1540345903422458
11. Barron JL, Fleet DJ, Beauchemin S (1994) Performance of optical flow techniques. International Journal of Computer Vision 12(1):43–77
12. Beauchemin SS, Barron JL (1995) The computation of optical flow. ACM Computing Surveys 27(3):433–467
13. Bellettini G, Dal Maso G, Paolini M (1993) Semicontinuity and relaxation properties of a curvature depending functional in 2d. Ann Scuola Norm Sup Pisa Cl Sci 20:247–297
14. Berg AC, Berg TL, Malik J (2005) Shape matching and object recognition using low distortion correspondence. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 26–33
15. Berkels B, Kondermann C, Garbe C, Rumpf M (2009) Reconstructing optical flow fields by motion inpainting. In: Energy Minimization Methods in Computer Vision and Pattern Recognition, Springer Verlag, vol LNCS 5681, pp 388–400, DOI 10.1007/978-3-642-03641-5_29
16. Bertalmio M, Sapiro G, Caselles V, Ballester C (2000a) Image inpainting. In: Computer Graphics (SIGGRAPH '00 Proceedings), pp 417–424
17. Bertalmio M, Sapiro G, Randall G (2000b) Morphing active contours. PAMI 22(7):733–743
18. Bertalmio M, Bertozzi A, Sapiro G (2001) Navier-Stokes, fluid dynamics, and image and video inpainting. Proceedings of the International Conference on Computer Vision and Pattern Recognition, IEEE I:355–362
19. Bertalmio M, Vese L, Sapiro G, Osher S (2003) Simultaneous structure and texture image inpainting. IEEE Transactions on Image Processing 12(8):882–889, DOI 10.1109/TIP.2003.815261
20. Biederman I (1987) Recognition-by-components: A theory of human image understanding. Psychological Review 94(2):115–147
21. Bigün J (1988) Local symmetry features in image processing. PhD thesis, Linköping University, Linköping, Sweden
22. Borenstein E, Sharon E, Ullman S (2004) Combining top-down and bottom-up segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, *Workshop Percept Org in Comp Vision*
23. Bouchard G, Triggs B (2005) Hierarchical part-based visual object categorization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 710–715
24. Bredies K, Kunisch K, Pock T (2010) Total generalized variation. Siam Journal On Imaging Sciences 3(3):492, DOI 10.1137/090769521
25. Bruhn A, Weickert J, Feddern C, Kohlberger T, Schnörr C (2003) Real-time optic flow computation with variational methods. In: Petkov N, Westenberg M (eds) Computer Analysis of Images and Patterns, Lecture Notes in Computer Science, vol 2756, Springer Berlin/Heidelberg, pp 222–229
26. Bruhn A, Weickert J, Schnörr C (2005) LucasKanade meets HornSchunck: Combining local and global optic flow methods. Int J Computer Vision 61(3):211–231

27. Bruhn A, Weickert J, Kohlberger T, Schnörr C (2006) A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. International Journal of Computer Vision 70(3):257–277
28. Burt PJ, Adelson EH (1983) The laplacian pyramid as a compact image code. IEEE TransCOMM 31:532–540
29. Caselles V, Morel JM, Sbert C (1998) An axiomatic approach to image interpolation. Image Processing, IEEE Transactions on 7(3):376–386, DOI 10.1109/83.661188
30. Chan T, Shen J (2001a) Mathematical models for local nontexture inpaintings. SIAM Journal on Applied Mathematics 62(3):1019–1043, DOI 10.1137/S0036139900368844
31. Chan T, Shen J (2001b) Non-texture inpainting by curvature-driven diffusions (ccd). J Visual Comm Image Rep 12:436–449
32. Chan T, Shen J (2002) Mathematical models for local non-texture inpaintings. SIAM J Appl Math 62:1019–1043
33. Chan TF, Tai XC (2003) Level set and total variation regularization for elliptic inverse problems with discontinuous coefficients. Journal of Computational Physics 193(1):40–66, DOI 10.1016/j.jcp.2003.08.003
34. Chan TF, Osher S, Shen J (2001) The digital tv filter and nonlinear denoising. IEEE Transactions on Image Processing 10(2):231–241
35. Chan TF, Kang SH, Shen J (2002) Euler's elastica and curvature-based inpainting. SIAM Appl Math 63(2):564–592
36. Cohen I (93) Nonlinear variational method for optical flow computation. In: Proc. of the Eighth Scandinavian Conference on Image Analysis, vol 1, pp 523–530
37. Cremers D, Schnörr C (2002) Motion competition: Variational integration of motion segmentation and shape regularization. In: Van Gool L (ed) Pattern Recognition - Proc. of the DAGM, Lecture Notes in Computer Science, vol 2449, pp 472–480
38. Cremers D, Schnörr C (2003) Statistical shape knowledge in variational motion segmentation. Image and Vision Computing 21:77–86
39. Cremers D, Soatto S (2005) Motion competition: A variational approach to piecewise parametric motion segmentation. International Journal of Computer Vision 62:249–265
40. Criminisi A, Pérez P, Toyama K (2004) Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on Image Processing 13(9):1200–1212, DOI 10.1109/TIP.2004.833105
41. Csurka G, Dance CR, Fan L, Willamowski J, Bray C (2004) Visual categorization with bags of keypoints. In: ECCV, *Workshop on Stat Learn in CV'04*
42. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: CVPR, pp 886–893
43. Epshtein B, Ullman S (2005) Feature hierarchies for object classification. In: ICCV, pp 220–227
44. Esedoglu S, Jianhong S (2002) Digital inpainting based on the Mumford-Shah-Euler image model. Euro Jnl Appl Math 13:353–370
45. Felzenszwalb PF, Huttenlocher DP (2005) Pictorial structures for object recognition. IJCV 61(1):55–79
46. Fennema C, Thompson W (1979) Velocity determination in scenes containing several moving objects. Computer Graphics and Image Processing 9:301–315
47. Fergus R, Perona P, Zisserman A (2003) Object class recognition by unsupervised scale-invariant learning. In: CVPR, pp 264–271
48. Ferrari V, Fevrier L, Jurie F, Schmid C (2008) Groups of adjacent contour segments for object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(1):36–51
49. Fischler MA, Elschlager RA (1973) The representation and matching of pictorial structures. IEEE Transactions on Computers c-22(1):67–92
50. Fleet DJ (1992) Measurement of Image Velocity. Kluwer Academic Publishers, Dordrecht, The Netherlands

51. Fleet DJ, Jepson AD (1990) Computation of component image velocity from local phase information. International Journal of Computer Vision 5:77–104
52. Fukushima K (1980) Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics 36(4):193–202
53. Galvin B, McCane B, Novins K, Mason D, Mills S (1998) Recovering motion fields: an evaluation of eight optical flow algorithms. In: BMVC 98. Proceedings of the Ninth British Machine Vision Conference, vol 1, pp 195–204
54. Geman S, Potter DF, Chi Z (2002) *Composition Systems*. Quarterly of Applied Mathematics 60:707–736
55. Glazer F, Reynolds G, Anandan P (1983) Scene matching through hierarchical correlation. In: Proc. Conference on Computer Vision and Pattern Recognition, Washington, pp 432–441
56. Grauman K, Darrell T (2006) Pyramid match kernels: Discriminative classification with sets of image features. Tech. Rep. MIT-2006-020, MIT
57. Grimson W, Huttenlocher D (1990) On the sensitivity of the hough transform for object recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on 12(3):255 – 274, DOI 10.1109/34.49052
58. Grossauer H (2004) A combined pde and texture synthesis approach to inpainting. In: Pajdla T, Matas J (eds) Computer Vision - ECCV 2004, Lecture Notes in Computer Science, vol 3022, Springer-Verlag, pp 214–224, DOI 10.1007/978-3-540-24671-8_17
59. Guichard F (1998) A morphological, affine, and galilean invariant scale–space for movies. IEEE Transactions on Image Processing 7(3):444–456
60. Haußecker H, Spies H (1999) Motion. In: Jähne B, Haußecker H, Geißler P (eds) Handbook of Computer Vision and Applications, vol 2, Academic Press, chap 13
61. Heeger D (1987) Model for the extraction of image flow. Journal of the Optical Society of America 4(8):1455–1471
62. Heeger DJ (1988) Optical flow using spatiotemporal filters. International Journal of Computer Vision 1:279–302
63. Hinterberger W, Scherzer O, Schnörr C, Weickert J (2001) Analysis of optical flow models in the framework of calculus of variations. Tech. rep., Numerical Functional Analysis and Optimization, Revised version of Technical Report No. 8/2001, Computer Science Series, University of Mannheim, Germany
64. Hofmann T (2001) Unsupervised learning by probabilistic latent semantic analysis. Machine Learning 42(1):177–196
65. Horn B, Schunk B (1981) Determining optical flow. Artificial Intelligence 17:185–204
66. Hough P (1962) Method and means for recognizing complex patterns. *U.S. Patent 3069654*
67. Jin Y, Geman S (2006) Context and hierarchy in a probabilistic image model. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 2145–2152
68. Kass M, Witkin A, Terzopoulos D (1987) Snakes: Active contour models. In: ICCV87, pp 259–268
69. Kondermann C, Kondermann D, Jähne B, CS Garbe (2007a) An adaptive confidence measure for optical flows based on linear subspace projections. In: Hamprecht F, Schnörr C, Jähne B (eds) Pattern Recognition, Springer Verlag, vol LNCS 4713, pp 132–141, DOI 10.1007/978-3-540-74936-3_14
70. Kondermann C, Mester R, Garbe C (2008) A statistical confidence measure for optical flows. In: Forsyth D, Torr P, Zisserman A (eds) Computer Vision - ECCV 2008, Springer Verlag, vol LNCS 5304, pp 290–301, DOI 10.1007/978-3-540-88690-7_22
71. Lades M, Vorbrüggen JC, Buhmann JM, Lange J, von der Malsburg C, Würtz RP, Konen W (1993) Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers 42:300–311
72. Lai S, Vemuri B (1995) Robust and efficient algorithms for optical flow computation. In: Proc. of Int. Symp. Comp. Vis., pp 455–460
73. Laine A, Fan J (1996) Frame representations for texture segmentation. IEEE Transactions on Image Processing 5(5):771–780, URL citeseer.ist.psu.edu/article/laine96frame.html

74. Lampert CH, Blaschko MB, Hofmann T (2008) Beyond sliding windows: Object localization by efficient subwindow search. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
75. Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
76. Lefébure M, Cohen LD (2001) Image registration, optical flow and local rigidity. Journal of Mathematical Imaging and Vision 14:131–147
77. Leibe B, Leonardis A, Schiele B (2004) Combined object categorization and segmentation with an implicit shape model. In: ECCV, *Workshop Stat Learn'04*
78. Lenzen F, Schäfer H, CS Garbe (2011) Denoising Time-of-Flight data with adaptive total variation. In: Bebis G, Boyle R, Koracin D, Parvin B (eds) Advances in Visual Computing, ISVC 2011, vol LNCS 6938, pp 337–346, DOI 10.1007/978-3-642-24028-7_31
79. Liapis, S S, Tziritas, G (2004) Colour and texture segmentation using wavelet frame analysis, deterministic relaxation, and fast marching algorithms. Journal of Visual Communication and Image Representation 15:1–26
80. Little JJ, Verri A (1989) Analysis of differential and matching methods for optical flow. In: IEEE Workshop on Visual Motion, Irvine, CA, pp 173–180
81. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2):91–110
82. Lucas B, Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: DARPA Image Understanding Workshop, pp 121–130
83. Lucas BD (1984) Generalized image matching by the method of differences. PhD thesis, Carnegie-Mellon University, Pittsburgh, PA
84. Luis Alvarez JS Joachim Weickert (1999) A scale-space approach to nonlocal optical flow calculations. Proceedings of the Second International Conference on Scale-Space Theories in Computer Vision pp 235–246
85. Maji S, Malik J (2009) Object detection using a max-margin hough transform. In: CVPR
86. Masnou S (2002) Disocclusion: A variational approach using level lines. IEEE Transactions on Image Processing 11(2):68–76
87. Masnou S, Morel J (1998a) Level lines based disocclusion. In: Proc. of ICIP, vol 3, pp 259–263
88. Masnou S, Morel JM (1998b) Level lines based disocclusion. In: 5th IEEE International Conference on Image Processing (ICIP), Chicago, vol 3, pp 259–263
89. Matsushita Y, Ofek E, Ge W, Tang X, Shum HY (2006) Full-frame video stabilization with motion inpainting. Pattern Analysis and Machine Intelligence, IEEE Transactions on 28(7):1150–1163, DOI 10.1109/TPAMI.2006.141
90. McCane B, Novins K, Crannitch D, Galvin B (2001) On benchmarking optical flow. Computer Vision and Image Understanding 84(1):126–143
91. Mumford D, Shah J (1989) Optimal approximation by piecewise smooth functions and associated variational problems. Communications on Pure and Applied Mathematics 42:577–685
92. Müller-Urbaniak S (1994) Eine Analyse des Zweischritt-[Theta]-Verfahrens zur Lösung der instationären Navier-Stokes-Gleichungen. Tech. rep.
93. Nagel HH (1983) Displacement vectors derived from second-order intensity variations in image sequences. Computer Graphics and Image Processing 21:85–117
94. Nagel HH, Enkelmann W (1986) An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. IEEE Trans Pattern Anal Mach Intell 8(5):565–593
95. Nicolescu,M, Medioni,G (2002) 4d voting for matching, densification and segmentation into motion layers. In: Proceedings of the International Conference on Pattern Recognition, vol 3
96. Nicolescu,M, Medioni,G (2003) Layered 4d representation and voting for grouping from motion. Pattern Analysis and Machine Intelligence 25(4):492–501
97. Ommer B, Buhmann J (2010) Learning the compositional nature of visual object categories for recognition. PAMI 32(3):501–516

98. Ommer B, Buhmann JM (2005) Object categorization by compositional graphical models. In: Energy Minimization Methods in Computer Vision and Pattern Recognition, LNCS 3757, pp 235–250
99. Ommer B, Buhmann JM (2006) Learning compositional categorization models. In: ECCV, pp 316–329
100. Ommer B, Malik J (2009) Multi-scale object detection by clustering lines. In: Computer Vision, 2009 IEEE 12th International Conference on, pp 484 –491, DOI 10.1109/ICCV.2009.5459200
101. Ommer B, Sauter M, Buhmann JM (2006) Learning top-down grouping of compositional hierarchies for recognition. In: CVPR, *Workshop POCV*
102. Opelt A, Pinz A, Zisserman A (2006) Incremental learning of object detectors using a visual shape alphabet. In: CVPR, pp 3–10
103. Oppenheim AV, Schafer RW (2009) Discrete-Time Signal Processing, 3rd edn. Prentice Hall
104. Petrovic,A E, Vanderheynst,P (2004) Multiresolution segmentation of natural images: From linear to nonlinear scale-space representation. IEEE Transactions on Image Processing 13(8):1104–1114
105. Pietikäinen M, Rosenfeld A (1981) Image segmentation by texture using pyramid node linking. Systems, Man and Cybernetics, IEEE Transactions on 11(12):822–825, DOI 10.1109/TSMC.1981.4308623
106. Preußer, Droske M, Garbe CS, Telea A, Rumpf M (2007) A phase field method for joint denoising, edge detection and motion estimation. SIAM Appl Math, 68(3): 599–618, DOI 10.1137/060677409
107. Proakis JG, Manolakis DK (2006) Digital Signal Processing, 4th edn. Prentice Hall
108. Schnörr C (1992) Computation of discontinuous optical flow by domain decomposition and shape optimization. International Journal Computer Vision 8(2):153–165
109. Sivic J, Russell BC, Efros AA, Zisserman A, Freeman WT (2005) Discovering objects and their localization in images. In: ICCV
110. Spies H, Garbe CS (2002) Dense parameter fields from total least squares. In: Van Gool L (ed) Pattern Recognition, Springer-Verlag, Zurich, CH, Lecture Notes in Computer Science, vol LNCS 2449, pp 379–386
111. Strehl,A, Aggarwal,JK (2000) A new Bayesian relaxation framework for the estimation and segmentation of multiple motions. In: Proceedings of the 4th IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI 2000), 2–4 April 2000, Austin, Texas, USA, IEEE, pp 21–25, URL citeseer.ist.psu.edu/strehl00new.html
112. Sudderth EB, Torralba AB, Freeman WT, Willsky AS (2005) Learning hierarchical models of scenes, objects, and parts. In: ICCV
113. Tan L (2007) Digital Signal Processing: Fundamentals and Applications. Academic Press
114. Telea A, Preusser T, Garbe C, Droske M, Rumpf M (2006) A variational approach to joint denoising, edge detection and motion estimation. In: Proc. DAGM 2006, pp 525–353
115. Tretiak O, Pastor L (1984) Velocity estimation from image sequences with second order differential operators. In: Proc. 7th International Conference on Pattern Recognition, pp 20–22
116. Unser M (1995) Texture classification and segmentation using wavelet frames. IEEE Transaction on Image Processing 11:1549–1560
117. Vidal R, Ma Y (2004) A unified algebraic approach to 2-d and 3-d motion segmentation. In: ECCV (1), pp 1–15
118. Viola PA, Jones MJ (2001) Rapid object detection using a boosted cascade of simple features. In: CVPR, pp 511–518
119. Wang JYA, Adelson EH (1993) Layered representation for motion analysis. In: Wang JYA, Adelson EH (eds) Proceedings CVPR'93, New York City, NY, Washington, DC, pp 361–366
120. Wang JYA, Adelson EH (1994) Representating moving images with layers. IEEE Transaction on Image Processing 3(5):625–638
121. Wang,JYA,Adelson,EH (1994) Representing moving images with layers. The IEEE Transactions on Image Processing Special Issue: Image Sequence Compression 3(5):625–638

122. Waxman AM, Wu J, Bergholm F (1988) Convected activation profiles and receptive fields for real time measurement of short range visual motion. In: Proc. Conf. Comput. Vis. Patt. Recog, Ann Arbor, pp 771–723
123. Weickert J, Schnörr C (2001) A theoretical framework for convex regularizers in PDE–based computation of image motion. Int J Computer Vision 45(3):245–264