

Complex Motion Models for Simple Optical Flow Estimation

Claudia Nieuwenhuis¹, Daniel Kondermann², and Christoph S. Garbe^{2*}

¹ Technical University of Munich, Germany
nieuwenhuis@in.tum.de

² IWR, University of Heidelberg, Germany

Abstract. The selection of an optical flow method is mostly a choice from among accuracy, efficiency and ease of implementation. While variational approaches tend to be more accurate than local parametric methods, much algorithmic effort and expertise is often required to obtain comparable efficiency with the latter. Through the exploitation of natural motion statistics, the estimation of optical flow from local parametric models yields a good alternative. We show that learned, linear, parametric models capture specific higher order relations between neighboring flow vectors and, thus, allow for complex, spatio-temporal motion patterns despite a simple and efficient implementation. The method comes with an inherent confidence measure, and the motion models can easily be adapted to specific applications with typical motion patterns by choice of training data. The proposed approach can be understood as a generalization of the original structure tensor approach to the incorporation of arbitrary linear motion models. In this way accuracy, specificity, efficiency and ease of implementation can be achieved at the same time.

1 Introduction

1.1 Keeping Optical Flow Estimation Simple

Optical flow refers to the displacement field between subsequent frames of an image sequence. Methods for its computation are in practice usually tradeoffs between speed, accuracy and implementation effort. Local methods such as the Lucas & Kanade approach [1] or the structure tensor approach by Bigün [2] are very fast and easy to implement but not very accurate due to the simplified assumption that neighboring pixels move in the same way. Global methods such as the method by Horn and Schunck [3] and today's state-of-the-art approaches are usually more accurate and can sometimes even be applied in realtime, yet with considerable implementation effort and expertise. Knowledge on adequate discretization, multigrid methods [5] and/or GPU implementation, variational calculus or Markov Random Fields, image and gradient filters [6], coarse to fine

* This work was funded by the "Heidelberg Graduate School of Mathematical and Computational Methods for the Sciences" (DFG GSC 220) and DFG grants GA1271/2-3 and CR250/6-1.

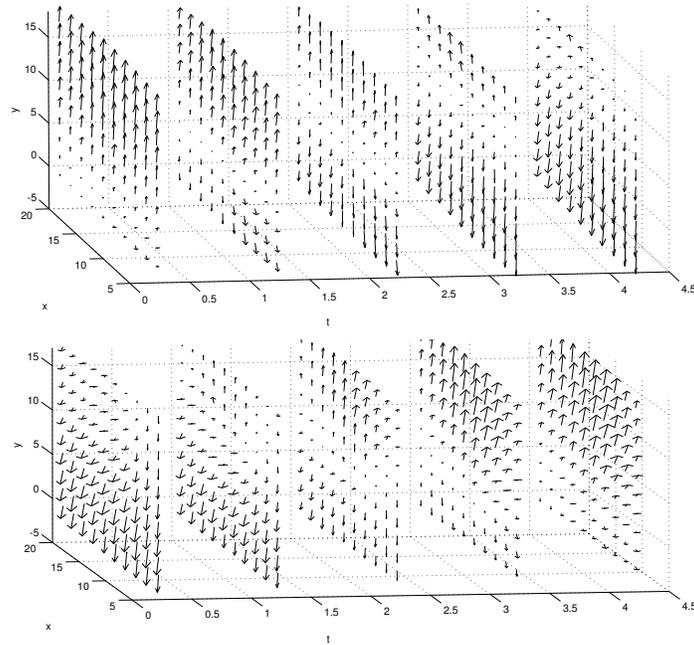


Fig. 1. Examples for learned motion models. The inclusion of temporal information allows for the representation of complex temporal phenomena, e.g. moving motion discontinuities (top) and moving divergences (bottom).

strategies for handling large motion, possibly in combination with image warping [7], as well as different norms and their characteristics for the preservation of motion boundaries is indispensable in order to obtain highly accurate and fast algorithms. Methods from the 1980s such as the Horn-Schunck approach can yield rather good results if a pyramid approach and bicubic interpolation are used to handle large motion combination with a multigrid solver. Yet, for industrial applications such high end knowledge is rare and expensive.

In this paper we reduce the necessary expertise and implementation effort to a minimum by learning statistically optimal (in the Gaussian sense) motion models from training data, which capture complex motion patterns, yet are still simple to compute. The complexity of the estimation task itself can be vastly reduced due to the linearity of the model. At the same time the learning based approach entails a strong adaptability to the specific optical flow estimation problem, e.g. in fluid dynamics or driver assistance systems (see Figure 1). Finally, a confidence measure directly inherent to the flow computation method can be applied to improve the result.

1.2 Motion Models

Because the optical flow problem is underconstrained, all estimation methods involve additional model assumptions on the structure of the motion.

The simplest motion model is a piecewise constant flow, which can be obtained by punishing the l_2 norm of the gradient of the flow vector $\|\nabla u\|^2$ [3] or by assuming constant motion within a small local neighborhood [1, 2]. This model is not adequate for most optical flow fields. Other methods assume more general models such as local planar models [8] or affine models [9–11]. General affine models have been integrated into a variational framework by Nir et al. [12]. In order to preserve motion boundaries in variational approaches, l_1 regularization in the space of bounded variations, the widely used TV-regularization, was introduced into denoising by Rudin, Osher and Fatemi [13] and was soon applied in optical flow estimation. Recently, a generalized version of the TV-regularizer, the total generalized variation (TGV) regularizer, allowing for piecewise affine motion with sharp boundaries has been introduced into variational approaches [14]. For physics based applications often different conditions apply in flow estimation, e.g. the Navier-Stokes equation for fluid flows [15, 16]. In the field of water or heat motion, patterns such as vortices or curls are common and can be handled by means of adequate regularization such as the div-curl regularizers [17] or model based approaches [18]. Flow fields based on such predefined models are often more accurate than those based on assumptions of constant flow. However, there are situations where even more complex and especially application specific models are necessary to compute accurate flow fields.

In these situations learning motion models from given sample motion data is a method to obtain superior results. Roth and Black [19] employed a general learning based approach using fields of experts, which is integrated into a global optical flow method. The approach is rather difficult to implement due to the fields of experts model. Black et al. [20] as well as Yacoob and Davis [21] integrated adapted, learning based models into a local optical flow method. To learn these models, principal component analysis (PCA) is used. This leads to a nonlinear energy functional, which is linearized and minimized by means of a coarse-to-fine strategy and coordinate descent. However, the models employed are either purely spatial [20] or purely temporal [21], or they are used for confidence estimation [22]. Our approach differs in four main aspects from these methods:

1. Instead of formulating a non-linear energy functional and applying gradient descent for its minimization we obtain an overdetermined system of equations which can be easily and globally solved by established least squares methods.
2. We employ spatio-temporal instead of purely spatial or temporal motion models, which can represent complex motion patterns over time.
3. We show results comparable to Farnebäck's (which are the best for local optical flow estimation), but with much less effort.
4. By means of a model-based confidence measure directly inherent to the flow computation method we can sparsify and reconstruct [23] the resulting flow field obtaining significantly lower errors.

Based on learned motion models for complex patterns, we end up with a simple, easily parallelizable and still accurate optical flow method. It is able to incorporate learned prior knowledge on special types of motion patterns and can be

adapted to all kinds of specific motion estimation problems. The goal of this paper is not to devise a method more accurate than any other method before, but to point out a simple to implement and efficient alternative to all the state-of-the-art optical flow approaches.

2 Motion Statistics

We use statistical methods to directly learn the motion model from sample motion data. For a given sample flow field $u : \Omega \rightarrow \mathbb{R}^2$ defined on the spatio-temporal image domain $\Omega \subset \mathbb{R}^3$ we randomly choose a specified number of locations (here 5000), where spatio-temporal sample patches of a fixed size $\omega \subset \Omega$ are drawn from. Such sample flow fields can be ground truth flow fields, computed flow fields or any other flow fields containing motion patterns typical for the application. The sample patches are vectorized (horizontal component stacked on top of vertical flow component) and stored as columns in a matrix M . To avoid bias towards any direction, we rotate the training samples four times by 90 degrees in order to obtain a zero sample mean.

We use principal component analysis (PCA) to obtain motion models from the sample data. PCA means that we compute a new orthogonal basis system $B := [b_1, \dots, b_p] \in \mathbb{R}^{p \times p}$, $p = 2|\omega|$ (each column represents one eigenvector containing a horizontal and a vertical flow component at each pixel of the patch ω), within which the original sample fields in M are decorrelated. This is simply done by computing the eigenvectors of the matrix M . Examples for such typical motion patterns are presented in Figure 1.

Let the basis components, i.e. the eigenvectors b_j of M , be sorted according to decreasing eigenvalues. Then the first $k \leq p$ eigenvectors with the largest eigenvalues contain most of the variance of the sample data, whereas the eigenvectors with small eigenvalues usually represent noise or errors in the sample flow fields and should be removed from the set. Hence, the first k basis components span a linear k -dimensional subspace of the original sample data space preserving most of the sample data information. In order to select the number of eigenvectors containing the fraction $\delta \in (0, 1)$ of the information of the original dataset we choose the value k based on the eigenvalues λ_i of the eigenvectors b_i , $i \in \mathbb{N}_p$ such that

$$k := \operatorname{argmin}_{j \in \mathbb{N}_p} \frac{\sum_{i=1}^j \lambda_i}{\sum_{i=1}^p \lambda_i} \geq \delta . \quad (1)$$

Within the resulting k -dimensional subspace a vectorized flow field patch $u(\omega)$ can be approximated by a linear combination of the first k principal components:

$$u(\omega) \approx \sum_{j=1}^k \alpha_j b_j . \quad (2)$$

The linear subspace restricts possible solutions of the flow estimation problem to the subspace of typical motion patterns statistically learned from sample data.

3 Parameter Estimation

Optical flow computation is based on the assumption that the brightness of a moving pixel remains constant over time. If $x : [0, T] \rightarrow \mathbb{R}^2$ describes the trajectory of a point of an object we can model a constant brightness intensity $I : \Omega \rightarrow \mathbb{R}$ as $I(x(t), t) = \text{const}$. A first order approximation yields the brightness constancy constraint equation (BCCE), $\nabla_{x,y}$ is the spatial gradient operator

$$\frac{dI}{dt} = 0 \quad \Leftrightarrow \quad (\nabla_{x,y} I)^T \cdot u + \frac{\partial I}{\partial t} = 0 \quad , \quad (3)$$

Instead of estimating the optical flow u itself we want to estimate the coefficients $\alpha_j, j \in \mathbb{N}_k$, in (2) which implicitly define the displacement field for the neighborhood of the current pixel. In this way, the resulting optical flow is restricted to the learned subspace spanned by the k principal components. In the sense of Nir et al. [12] one could speak of an over-parameterized model, since we estimate k coefficients to obtain a two-dimensional flow vector, which is chosen from the center of the reconstructed patch.

To estimate the coefficients α_j we make two assumptions:

1. the current flow field patch can be approximated by a linear combination of principal components (2),
2. each of the flow vectors within the patch fulfills the BCCE (3).

Substituting u in (3) by the linear combination of model vectors in (2), we obtain the following energy to be minimized for a whole patch located at $\omega \subset \Omega$

$$E(\alpha) = \left| \left(\begin{array}{cccccc} I_{x1} & 0 & \dots & 0 & I_{y1} & 0 & \dots & 0 \\ 0 & I_{x2} & 0 & \dots & 0 & I_{y2} & 0 & \dots & 0 \\ \vdots & & \ddots & & & & \ddots & & \vdots \\ 0 & \dots & 0 & I_{xn} & 0 & \dots & 0 & I_{yn} \end{array} \right) \left(\sum_{j=1}^k \alpha_j b_j \right) + \frac{\partial I(\omega)}{\partial t} \right|^2 \rightarrow \min, \quad (4)$$

where I_{cq} denotes the image derivative with respect to c at patch pixel index q and $I(\omega)$ denotes the vectorized intensities within the image patch ω containing $n := |\omega|$ pixels. Denoting the image gradient matrix in 4 as A , we obtain the following system of equations depending on the parameters α for each patch ω

$$(A \cdot B)^T \cdot \alpha = -I_t, \quad (5)$$

which can be solved by a simple least squares approach (or by the more sophisticated least median of squares, which can handle outliers e.g. at motion boundaries [24]). The resulting parameter vector α represents the estimated flow for the whole patch. To obtain the flow at the central pixel, we compute the linear combination and choose the central flow vector.

The proposed approach can be understood as a generalization of the original structure tensor approach to the incorporation of arbitrary linear motion models. In case we choose only two model vectors, a normalized horizontal and vertical one, we obtain exactly the original structure tensor approach.

4 Confidence Estimation

To improve the resulting optical flow field, we apply the confidence measure proposed in [22], which assigns a reliability value in the interval $[0, 1]$ to each flow vector. Here, 0 stands for no, 1 for full reliability. Based on a confidence map computed flows can be sparsified in order to reduce the average error. In this way the accuracy of further processing steps can be increased or errors can be removed by means of reconstruction algorithms, e.g. inpainting. According to [22] we assume that all correct flow field patches can be described in terms of the learned motion model, i.e. lie in the space of eigenflows. Thus, the inverse of the difference between the projection into the model space spanned by the principal components and the original patch $u(\omega)$ at location $\omega \in \Omega$ centered at x indicates the confidence

$$c(x) = \frac{1}{1 + |u(\omega) - (B \cdot B^T u(\omega))|}. \quad (6)$$

Since the basis system has already been computed, the application of the confidence measure is trivial and effective.

5 Results

In this section we present results on the accuracy, efficiency and adaptability of the proposed optical flow method. For the implementation we use the filters optimized for optical flow by Scharr [6] to estimate image derivatives. We presmoothed all sequences by a Gaussian filter with spatial $\sigma = 0.8$.

Accuracy Our goal is not to devise an optical flow method yielding accuracy as high as the top ranking state-of-the-art methods on the Middlebury database. Instead, we want to demonstrate that the learned motion models can capture much of the complexity of optical flow fields and allow for an efficient and simple to implement flow computation method. To test our approach we use five different test sequences: the famous Marble sequence, the Yosemite and the Rubber Whale sequence from the Middlebury dataset [4] as well as the Street and Office sequence [25]. Learning is performed on a set of various computed flow fields. Table 1 shows a comparison to results obtained by the original structure tensor approach [2] based on the same patch size. Furthermore, we show error values and chosen parameters for all test sequences for different densities after sparsification based on the confidence measure proposed in section 4. Figure 2 displays the corresponding HSV coded motion fields, which show improvements especially for large moving planes and near motion boundaries. Here oversmoothing due to large patch sizes as in the traditional approach is avoided since the complex relations between neighboring vectors are already contained in the model.

Results					
Density (%)	Yosemite	Marble	Rubber Whale	Street	Office
100	1.53 ± 1.69	2.55 ± 4.25	7.85 ± 15.95	4.99 ± 13.72	3.83 ± 4.98
90	1.37 ± 1.43	1.87 ± 2.73	5.24 ± 10.43	3.65 ± 8.38	3.35 ± 3.85
80	1.24 ± 1.37	1.49 ± 2.05	4.36 ± 9.45	3.04 ± 6.05	3.01 ± 3.38
70	1.15 ± 1.38	1.27 ± 1.65	4.12 ± 9.75	2.44 ± 4.52	2.75 ± 3.25
traditional	3.42 ± 10.01	5.25 ± 6.49	19.30 ± 17.23	5.75 ± 16.92	5.55 ± 11.82
ω, k	$19 \times 19 \times 3, 10$	$19 \times 19 \times 7, 6$	$19 \times 19 \times 1, 2$	$19 \times 19 \times 3, 2$	$21 \times 21 \times 1, 5$

Table 1. Comparison of the angular error and standard deviation of the model based approach to the traditional approach. The density refers to the flow field after sparsification based on the confidence measure [22]. ω denotes the spatio-temporal patch size, k the number of eigenvectors.

Stability In the following, we examine the stability of our method for different parameter choices, i.e. patch sizes and numbers of eigenvectors. Table 2 exhibits average error and standard deviation for different patch sizes based on 7 principal components and for different numbers of principal components for fixed patch size $\omega = 21 \times 21 \times 3$. The results suggest that large patch sizes are favorable for lower errors, whereas the number of principal components has less influence.

space \ time	Model Size			Principal components			
	1	3	7	k	$21 \times 21 \times 3$	k	$21 \times 21 \times 3$
5×5	7.12 ± 12.76	4.72 ± 7.62	3.01 ± 3.55	2	1.93 ± 2.07	6	1.40 ± 1.53
9×9	3.93 ± 6.46	2.69 ± 3.49	2.12 ± 2.27	3	1.86 ± 2.06	7	1.35 ± 1.45
15×15	2.39 ± 3.01	1.81 ± 1.97	1.50 ± 1.70	4	1.53 ± 1.53	8	1.36 ± 1.40
21×21	1.79 ± 1.87	1.66 ± 1.57	1.35 ± 1.45	5	1.44 ± 1.56	9	1.35 ± 1.41

Table 2. Left: angular error and standard deviation for different spatio-temporal sizes ω for the Yosemite sequence using 7 principal components. Right: angular error for different numbers of principal components for a fixed patch size of $21 \times 21 \times 3$.

Adaptability The proposed algorithm is adaptable to all kinds of scenes where typical, complex motion patterns need to be computed such as in fluid dynamics or driver assistance systems in vehicles. Figure 3 shows spatial principal components computed on particle image velocimetry (PIV) test data, on the Yosemite sequence and on a motion boundary. The examples show that for very different kinds of flow fields we obtain very different motion models.

Efficiency The proposed local optical flow method can be implemented efficiently due to several reasons. First, the algorithm only takes into account a limited local image region in order to estimate the displacement vector for each pixel. Hence, it takes only limited space and can be easily parallelized for the computation on graphics hardware. Second, the computation of the PCA model can be carried out once before the estimation of the optical flow and can be used for all kinds of sequences and the confidence estimation later on.

Computation times grow linearly with the number of pixels contained in the patch, and almost linearly with the number of eigenvectors. Figure 4 shows computation times for a single pixel on a standard CPU.

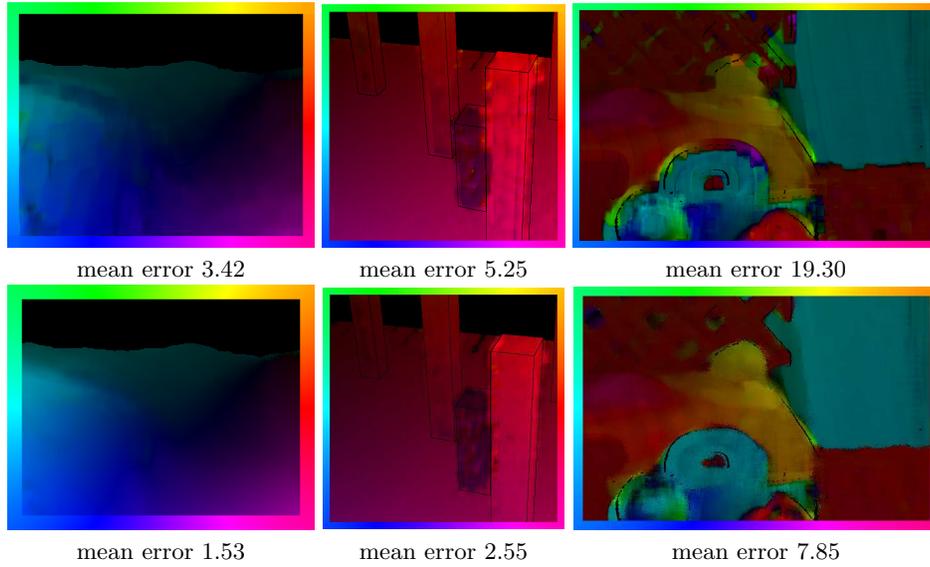


Fig. 2. Comparison of original structure tensor approach (top) and the motion model based structure tensor approach (bottom), HSV-coded flow fields and mean angular error for the Yosemite, Marble and the Rubber Whale sequence.

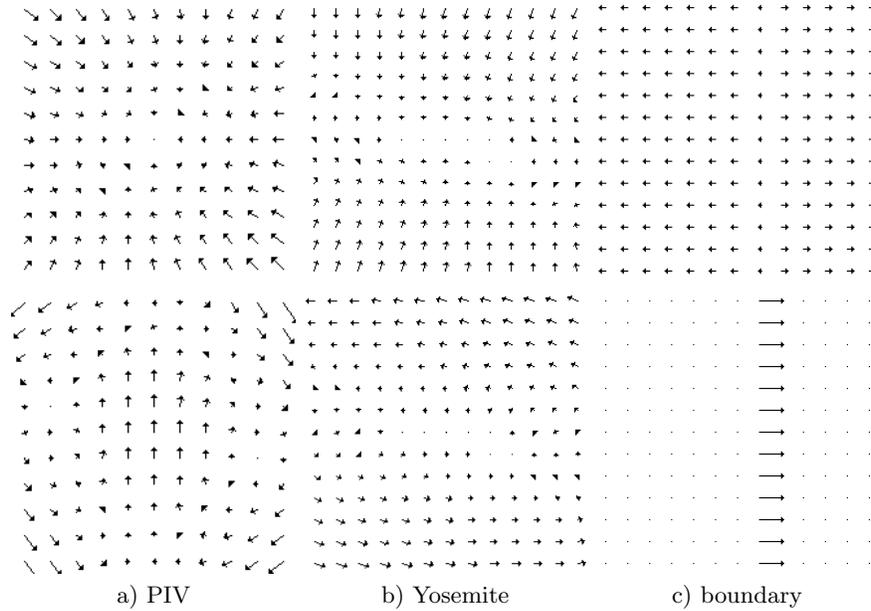


Fig. 3. Principal components based on training data from totally different application fields, a) PIV data consisting of fluid motion patterns, b) Yosemite consisting mostly of translations, c) a motion boundary.

6 Summary and Conclusion

In this paper we proposed a generalization of the traditional algorithm by Bigün for optical flow estimation to the incorporation of complex motion models. This approach has four advantages: 1) The resulting method yields errors approximately half the value of the traditional approach due to the use of motion models, which can capture complex spatio-temporal motion patterns. In this way, we incorporate prior knowledge on regular flow field patches without the need for explicit regularization which would drastically increase computation times and implementation complexity. And the results are improved especially near motion boundaries and for planar motion. 2) The algorithm boils down to the simple task of carrying out principal component analysis on training data and solving an overdetermined linear system of equations at each pixel location by means of least squares. Thus, the implementation effort and necessary expertise are strongly reduced, which makes our approach especially interesting for industrial applications. 3) The learned motion models are adaptable to all kinds of specific applications, where special motion patterns occur, e.g. in the field of fluid dynamics or driver assistance systems. 4) The approach is stable with respect to parameter variations.

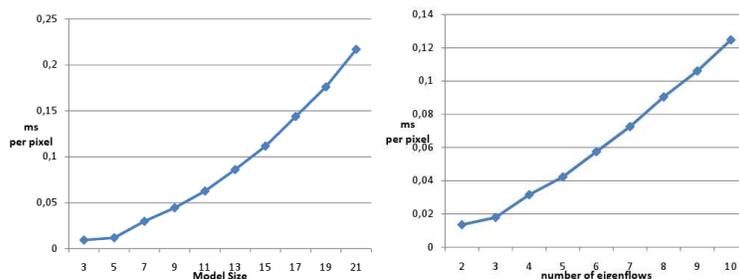


Fig. 4. Computation times per pixel for increasing patch sizes (left) and increasing numbers of eigenvectors (right) on the CPU.

References

1. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision (DARPA). In: Proceedings of the DARPA Image Understanding Workshop. (1981) 121–130
2. Bigün, J., Granlund, G., J. Wiklund: Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Journal of Pattern Analysis and Machine Intelligence* **13**(8) (1991) 775–790
3. Horn, B., Schunck, B.: Determining optical flow. *Artificial Intelligence* **17** (1981) 185–204

4. Baker, S., Roth, S., Scharstein, D., Black, M., Lewis, J., Szeliski, R.: A database and evaluation methodology for optical flow. In: Proceedings of ICCV. (2007) 1–8
5. Bruhn, A., Weickert, J., Feddern, C., Kohlberger, T., Schnörr, C.: Real-time optic flow computation with variational methods. *IEEE Transactions in Image Processing* **14**(5) (2005) 608–615
6. Scharr, H.: Optimal filters for extended optical flow. In: Complex Motion, Lecture Notes in Computer Science. Volume 3417., Springer (2004)
7. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: ECCV. (2004) 25–36
8. Black, M., Jepson, A.: Estimating multiple independent motions in segmented images using parametric models with local deformations. *IEEE Workshop on Motion of Non-Rigid and Articulated Objects* (1994)
9. Ju, S.X., Black, M.J., Jepson, A.D.: Skin and bones: Multi-layer, locally affine, optical flow and regularization with transparency. In: CVPR. (1996)
10. Black, M., Yacoob, Y.: Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In: Proceedings of the International Conference on Computer Vision (ICCV). (1995)
11. Farneback, G.: Fast and accurate motion estimation using orientation tensors and parametric motion models. In: ICPR. Volume 1. (2000) 135–139
12. Nir, T., Bruckstein, A.M., Kimmel, R.: Over-parameterized variational optical flow. *International Journal of Computer Vision* **76**(2) (June 2006) 205–216
13. Rudin, L., Osher, S.: Total variation based image restoration with free local constraints. In: ICIP. Volume 1. (1994) 31–35
14. Bredies, K., Kunish, K., Pock, T.: Total generalized variation, techn. rep. (2009)
15. Vlasenko, A., Schnörr, C.: Physically consistent and efficient variational denoising of image fluid flow estimates. *IEEE Transact. Image Process.* **19**(3) (2010) 586–595
16. Haussecker, H., Fleet, D.: Computing optical flow with physical models of brightness variation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* **23**(6) (2001) 661–673
17. Gupta, S., Gupta, E.N., Prince, J.L.: Stochastic models for div-curl optical flow methods. *IEEE Signal Processing Letters* **3** (1996) 32–35
18. Cuzol, A., Hellier, P., Mmin, E.: A low dimensional fluid motion estimator. *Int. J. Comp. Vision* **75** (2007)
19. Roth, S., Black, M.: On the spatial statistics of optical flow. In: International Conference on Computer Vision, Proceedings. Volume 1. (2005) 42–49
20. Black, M., Yacoob, Y., Jepson, A., Fleet, D.: Learning parameterized models of image motion. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR). (1997)
21. Yacoob, Y., Davis, L.: Learned temporal models of image motion. In: International Conference on Computer Vision, Proceedings. (1998)
22. Nieuwenhuis, C., Kondermann, D., Jähne, B., Garbe, C.: An adaptive confidence measure for optical flows based on linear subspace projections. In: Pattern Recognition. Volume 4713 of LNCS., Springer (2007) 132–141
23. Nieuwenhuis, C., Kondermann, D., Garbe, C.: Postprocessing of optical flows via surface measures and motion inpainting. In: Pattern Recognition. Volume 5096 of LNCS., Springer (2008) 355–364
24. Suter, D.: Motion estimation and vector splines. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, IEEE (1994) 939–942
25. McCane, B., Novins, K., Crannitch, D., Galvin, B.: On benchmarking optical flow. *Computer Vision and Image Understanding* **84**(1) (2001) 126–143