A variational optical flow approach using learned motion models for the determination of fluid flows

Julian Stapf*, Christoph Garbe

Interdisciplinary Center for Scientific Computing, University of Heidelberg, Germany
*correspondent author: Julian.stapf@iwr.uni-heidelberg.de

Abstract A new variational learning-based approach for determining fluid flow velocities and dominant motion patterns from particle images is proposed. It is similar to combined local and global optical flow approaches but applies linear spatio-temporal models, which have previously been obtained by methods of unsupervised learning using proper orthogonal decomposition (POD). The learned motion models are given by the first high energy POD modes and capture information about complex relations between neighboring flow vectors in spatio-temporal motion patterns. The estimated flow field is composed by the motion models and associated parameters. This ensures the solution to be restricted to the orthogonal space spanned by the learned POD modes. Additional information about local, dominant flow structures can be gained by the POD modes and related parameters. The method is easily tunable for different flow applications by choice of training data and, thus, is universally applicable. The estimation is done in a variational manner by minimizing a global energy functional consisting of a data term which includes the motion models and the associated parameters as well as an additional smoothness constraint for the parameters. This results in dense accurate flow fields given by a linear combination of the motion models and the estimated parameters. The approach was tested on synthetically generated image sequences and the estimated flow fields were compared to the known correct flow fields to quantify the quality of the method. It was compared to other optical flow techniques and correlation-based PIV.

1 Introduction

For measuring fluid dynamic transport processes, flow visualization (Merzkirch, 2007) is often indispensable. It yields an image sequence from which a velocity field, representing the fluid flow, can be computed. In order to make the flow visible, the fluid is seeded with some tracer material (e.g. small particles). In two dimensional setups the seeded fluid is illuminated at least twice within a plane by a laser light sheet and the light reflected by the particles is recorded. The velocity can be determined from the particle shift between subsequent image frames. The prevalent technique is particle image velocimetry (PIV) (Raffel et al., 2007) where the images are decomposed into small interrogation windows. Then the displacement vector is given by the maximum of the cross-correlation coefficient of associated windows from subsequent images.

One alternative technique is for instance optical flow which originates from computer vision and was originally developed to observe the motion of large object (Horn & Schunck, 1981). Many different methods exist (Barron et al., 1994; Beauchemin & Barron, 1995) but most common are so-called differential or gradient-based optical flow techniques which are based on spatio-temporal derivatives of the image intensity. Several applications have shown the potential of optical flow methods to determine fluid flow fields from PIV image sequences (Corpetti et al., 2006; Quénot et al., 1998; Ruhnau et al., 2005).

In this paper we propose a new, variational optical flow approach which uses previously learned motion models to solve the flow problem. It is an extension of the local learning-based optical flow approach proposed in (Stapf & Garbe, 2014). The spatio-temporal motion models are learned from appropriate training data by proper orthogonal decomposition (POD). The learning process is similar to the one used in (Nieuwenhuis et al., 2010). Learning-based motion estimation was previously
used by (Black et al., 1997) who employed learned spatial motion models to estimate non-rigid motion of human mouths.

In the proposed method the flow problem is solved within a global optimization approach applying a smoothness constraint on the flow coefficients. The solution is restricted to the space spanned by the learned basis flows given by the POD modes. The potential solution space can be controlled by the choice of the training data from which the motion models are learned. In this way prior knowledge can be introduced to the resulting flow field.

2 Optical flow basics

The proposed approach is based on optical flow methods that differ from correlation based techniques usually applied to PIV sequences. Optical flow is a two dimensional vector field representing the apparent velocities of brightness patterns in an image sequence. It is an approximation of the two dimensional motion field which itself is a projection of the three dimensional object motion relative to the image sensor onto the image plane. Motion field and optical flow only coincide exactly if the imaged objects do not change the irradiance on the image plane while moving in the scene (Jähne, 2005). However, in practice this is not always true and not every gray value change can be connected to a real world motion and vice versa. A demonstrative example for the disparity of optical flow and motion field is a homogeneous textured, spinning sphere together with a fixed illumination which leads, despite of the movement of the sphere, to zero optical flow. On the other hand a static sphere may lead to nonzero optical flow if the illumination is moving relative to the sphere.

In general optical flow approaches relay on the brightness constancy condition which states that the brightness and therefore the gray value \( g(x, t) \) of a point \( x = (x, y)^T \) of an image pattern remains constant for a displacement of the pattern from one image to the successive image. This condition leads to the brightness constancy constraint equation (BCCE) which is given by

\[
(Vg)^T \cdot u + g_t = 0
\]  

(1)

where the optical flow vector is denoted by \( u = (u, v)^T \). The operator \( V = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right)^T \) is a two dimensional gradient vector and, therefore, \( Vg \) denotes the vector of spatial gradients of the gray value in x- and y-direction. The term \( g_t \) denotes the temporal derivative. For a single image location the problem given by equation (1) is ill-posed since it contains only one constraint equation with two unknowns. Therefore, an infinite number of solutions exists all of which lie along a line in the two dimensional velocity space. A special solution is the normal flow \( u_n \) which points in direction of the spatial gradient of the gray value function \( g \). The second component of the optical flow vector, however, remains unknown. This problem is commonly known as the aperture problem (Beauchemin & Barron, 1995). In order to explicitly solve the BCCE (1) additional constraints are needed. The different approaches to overcome the aperture problem can be classified into local and global methods.

Local approaches pool the constraint equations of small neighborhoods since there is one equation per image location. Therefore, several constraints are available to solve for the two unknowns within one neighborhood presuming that the neighborhood contains enough structure to provide different constraint equations. The optical flow is modeled within the neighborhoods by some parametric function. The simplest model assumes the optical flow to be constant in the local neighborhoods (Lucas & Kanade, 1981). This leads to an overdetermined system of linear equations which can easily be solved by the method of least squares. The neighborhood has to be
sufficiently large to contain enough information to circumvent the aperture problem. For large neighborhoods, however, the model of constant flow seems inadequate and more sophisticated models have to be chosen to describe the flow within the neighborhood correctly. One example of such a sophisticated model is the affine motion model (Haussecker & Spies, 1999). Global methods include global constraints restricting the solution space and render the problem solvable. Most common is the smoothness constraint which assumes that the velocities of neighboring image locations are similar and vary only gradually (Horn & Schunck, 1981). The problem is formulated as optimization problem and an energy functional consisting of a data term, which is usually the BCCE, and a regularization term, the smoothness constraint, has to be minimized. Thereby, the influence of the regularization term can be controlled by the user by choice of a regularization parameter. In order to minimize the energy functional the Euler-Lagrange equations are solved iteratively, for instance by the Gauss-Seidel method.

One major disadvantage of gradient based optical flow methods is the limitation to small displacements due to linearization procedures used to derive the underlying equations. Another factor limiting the technique to relatively small displacements is the usage of gradient images that require that particles of subsequent images stay in contact (Liu & Shen, 2008). To cope with large displacement usually a coarse-to-fine strategy is used. Therefore, the estimation of the optical flow is done successively on multiresolution pyramidal image structures which are generated by low-pass filtering and subsampling of the image sequence (Heitz et al., 2010; Mémin & Pérez, 1998).

3 Method

The principles of the learning-based approach are briefly sketched in Fig. 1. At first a set of motion models is learned from appropriate training data which consists of familiar flow fields. The estimated flow field is then constructed by a linear combination of the motion models and some
parameters which have to be determined. The problem is formulated, similar to global optical flow methods, as energy functional which has to be minimized in order to determine the parameters. As training flow fields computed vector fields or ground truth data can be applied. Also any other flow fields containing motion patterns which are typical for the application are suited as training data. By choice of the training data the motion models can easily be adapted to different flow applications. However, the flow typical motion patterns have to be present in the training data in order to yield motion models capable to represent all existing flows of the image sequence. For instance training data consisting of velocity fields of a laminar pipe flow is certainly not suited for turbulent flow applications since important flow structures such as curls are completely missing in the training data and, therefore, in the resulting motion models as well.

The tool which is used to learn the motion models is proper orthogonal decomposition (POD). POD is also known under different names, such as principle component analysis or Karhunen-Loève decomposition. In the field of fluid dynamics it is sometimes used to recover the most energetic structures of turbulent flows (Lumley, 1967). However, the true connection between the POD modes and coherent structures is still debated (Gordeyev & Thomas, 2000). Yet, the training flow fields are decomposed into several basis flows by POD. These POD modes form an optimal set of orthogonal basis functions and comprise our motion models. In this context optimal means that the largest part of the whole energy contained in the training data is included in the first basis function and the second largest part in the second basis function which is orthogonal to the first basis function and so on. Therefore, the basis functions are sorted due to their energy or information content. This means that the first few POD modes are the most important ones. Compared to that, later POD modes are less important and can be omitted practically without losing information. In Fig. 2 some POD modes of size $[31 \times 31 \times 1]$ are shown. It can be seen that the first modes contain large structures whereas later modes contain finer structures. The motion models are not restricted to be purely spatial as in the plot but can also contain a temporal dimension. To learn the motion models a defined number of several thousand spatio-temporal sample patches are chosen randomly from the training flow fields. These patches may for instance be of size $[15 \times 15 \times 7]$ pixels. In order to prevent any directional bias of the motion models the patches are rotated several times, the flow vectors in the patches are mirrored on the vertical and the horizontal axis, and the temporal direction of the flow field in the patches is reversed. Each of these transformed patches is stored and the number of sample patches is therefore enlarged. The set of sample patches has to be transformed into a two dimensional data matrix $A$ on which the POD can be performed. Therefore, each of the sample patches is transformed into a column vector by writing its entries in lexicographical order in one vector. Thereby, the horizontal velocity components are stored in the upper half and the vertical velocity components in the lower half of the vector. The final data matrix $A$ contains the vectorized sample patches in its columns and is of size $[2N \times M]$ with $N$ being the number of flow vectors in one patch and $M$ the number of sample patches including mirrored, reflected, and time reversed patches. The factor 2 in front of $N$ is due to the two components of the velocity vectors. Another advantage of performing the patch transformations is to yield a data matrix of zero mean.

To perform the POD an eigenvalue problem has to be solved which can easily be done by the use of singular value decomposition (SVD) which yields a factorization of the data matrix. For a real data matrix $A$ of size $2N \times M$ the SVD is defined as

$$ A = U \Sigma V^T $$

where $U$ and $V^T$ are orthogonal $2N \times 2N$ and $M \times M$ matrices, respectively. The diagonal matrix $\Sigma$ is of size $2N \times M$ and contains the singular values $\sigma_{1,2N}$ on its diagonal. Since in our case the number of components is smaller than the number of patches it is $2N < M$ and the total number of
singular values is $2N$. They are sorted in decreasing order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{2N} \geq 0$ along the diagonal of $\Sigma$. The column vectors of $U$ are the eigenvectors or POD modes, respectively. They form the system of orthogonal basis functions $\phi_{1,2N}$ and represent the motion models. The size of each basis function $\phi_k$ is $2N \times 1$. The value $\sigma_k$ is directly related to the variance or the information content of the associated POD mode $\phi_k$. One question is how many POD modes are needed to represent a certain fraction (e.g. 90%) of the total information content. In order to determine this

Fig. 2: Some sample motion models of size [31x31x1]. The first models contain large structures and later models contain fine structures.
number $K$ of POD modes the relative information content (RIC) can be used. Due to (Nobach et al., 2007) it is defined as

$$RIC(K) = \frac{\sum_{k=1}^{K} \sigma_k}{\sum_{k=1}^{2N} \sigma_k}$$

which is simply the normalized part of the whole information contained in the first $K$ basis functions. The fast decay of the singular values or rather the information content of the associated basis function can be seen in the left plot of Fig. 3 where the singular values are plotted versus the number $k$. On the right side of Fig. 3 the RIC is plotted against $K$ which is the number of used components at a time. Because the RIC approaches 1 relatively fast only the first part up to $K=300$ is shown for a better visibility. The plot indicates that approximately 200 of the 2000 components contain 99% of the whole information content.

Any flow vector $\mathbf{u}$ can be decomposed into a linear combination of the first $K$ learned motion models $\phi_{1..K}$ and some parameters $\alpha_{1..K}$. It is

$$\mathbf{u}(x,t) = \sum_{k=1}^{K} \alpha_k \phi_k.$$  

This equation can be substituted into the brightness constancy constraint equation (1) from Sec. 2 in order to restrict the solution to the space spanned by the motion models. Combining these two equations leads to a system of $N$ extended brightness constancy constraint equations

$$G_{x,y} \cdot \sum_{k=1}^{K} \alpha_k \phi_k + g_1 = 0$$

with

$$G_{x,y} = \begin{pmatrix}
g_{x1} & 0 & \cdots & 0 & g_{y1} & 0 & \cdots & 0 \\
0 & g_{x2} & 0 & \cdots & 0 & g_{y2} & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & g_{xN} & 0 & \cdots & 0 & g_{yN}
\end{pmatrix}$$

and $g_1 = \begin{pmatrix} g_{x1} \\ g_{x2} \\ \vdots \\ g_{xN} \\ g_{y1} \\ g_{y2} \\ \vdots \\ g_{yN} \end{pmatrix}$. 

Fig. 3: Example for the distribution of the singular values (left) and the RIC (right). The singular values are shown for all 2000 components. The RIC is shown only up to $K=300$ for a clear illustration of the first part.
Here, $g_{xi}$ denotes the gradient in x-direction at the $i$-th patch position, $g_{yi}$ the gradient in y-direction at the $i$-th patch position, and $g_t$ the time derivative at the $i$-th patch position.

The system of linear equations (5) can be solved for the parameters $\alpha_k$ similar to other local optical flow methods (see Sec. 2) with the method of least squares as it was done in (Stapf & Garbe, 2014). After solving for the parameters $\alpha_k$ the optical flow vector $u = (u, v)^T$ can be determined by the linear combination of Equation (4).

Within this contribution we present another solution strategy. The problem can also be solved similar to the local-global optical flow approach proposed by (Bruhn et al., 2005). Therefore, it is formulated as energy functional with an additional smoothness constraint that claims smooth variation of the parameters $\alpha_k$ by penalizing large gradients $\nabla \alpha_k$. This alternative formulation combines the advantages of local and global optical flow methods, which are robustness against noise and dense flow fields, respectively. In locations with no or only little structure local methods may fail to determine an optical flow vector. Global methods, however, fill in information from the surrounding and are able to state flow vectors in these areas. With the following notations

$$\tilde{\alpha} = (\alpha_1, \ldots, \alpha_K, 1)^T$$

$$l_k = G_{x,y} \cdot \phi_k$$

$$L = \begin{pmatrix} l_1 & \ldots & l_K & g_1 \end{pmatrix}$$

$$J = L^T L$$

the energy functional can be expressed by

$$E = \int_{\Omega} \tilde{\alpha}^T J \tilde{\alpha} + \sum_{k=1}^K |\nabla \alpha_k|^2 dxdy. \quad (7)$$

The influence of the smoothness term (second term) compared to the data term (first term) can be controlled by the user by choice of the weighting parameter $\gamma$. A large $\gamma$ leads to a stronger smoothing of the parameters $\alpha_k$. To minimize the energy functional the Euler-Lagrange equations are determined. Using the subscripts $i,j \in (1\ldots K)$ this yields a system of $K$ equations

$$\sum_{j=1}^K J_{ij} \alpha_j + J_{ik+1} - \gamma \Delta \alpha_i = 0. \quad (8)$$

The Laplacian of $\alpha_i$ is defined as $\Delta \alpha_i = \frac{\partial^2 \alpha_i}{\partial x^2} + \frac{\partial^2 \alpha_i}{\partial y^2}$. A finite difference approximation to the Euler-Lagrange equations (8) is given for pixel $n$ by

$$\sum_{j=1}^K J_{ij,n} \alpha_{j,n} + J_{ik+1,n} - \gamma \left( \sum_{m \in \mathcal{N}(n)} \frac{\alpha_{i,m} - \alpha_{i,n}}{h^2} \right) = 0. \quad (9)$$

In this approximation $\mathcal{N}(n)$ is the set of the 4 neighbors of pixel $n$. $J_{ij,n}$ denotes the $(ij)$-component of tensor $J$ in pixel $n$ and $\alpha_{j,n}$ denotes $\alpha_j$ in pixel $n$. The parameter $h$ denotes the size of the rectangular pixel grid (e.g. 1 in our case). This system of linear equations can be solved iteratively for instance by using the successive overrelaxation (SOR) method. With the iteration index $q$, the relaxation parameter $\omega \in (0,2)$, and the notations $\mathcal{N}^{-}(n) = \{ m \in \mathcal{N}(n) \mid m < n \}$ and $\mathcal{N}^{+}(n) = \{ m \in \mathcal{N}(n) \mid m > n \}$ the iterative formulation for the $K$ parameters $\alpha_{1\ldots K}$ is given by the system of $K$ equations
\[ \alpha_i^{n+1} = (1 - \omega) \alpha_i^n + \omega \left( \sum_{m \in \mathcal{N}(n)} \alpha_{i,m}^{n+1} + \sum_{m \in \mathcal{N}(n)} \alpha_{i,m}^n - \frac{h^2}{\gamma} \left( \sum_{j=1}^K \sum_{l=1}^J l_{ij,n} \alpha_{ij,n}^k + \sum_{j=1}^K \sum_{l=1}^J l_{ij,n} \alpha_{ij,n}^{k+1} + \sum_{l=1}^J l_{i+1,j,n} \right) \right) \]

(10)

For a relaxation parameter \( \omega = 1 \) this is similar to the Gauss-Seidel method. However, choosing \( \omega \) between 1.5 and 2 leads to a speed-up of the convergence compared to the Gauss-Seidel method.

After determining the parameters \( \alpha_k \) the optical flow vector \( \mathbf{u} = (u, v)^T \) is given by the linear combination of Equation (4). Thus, the solution is composed by the motion models and the belonging parameters \( \alpha_k \) and, therefore, the solution space is restricted to the space spanned by the learned basis functions. In this way prior knowledge in form of the motion models is incorporated to the flow problem.

4 Results

The approach was tested on synthetically generated image sequences which have the advantage that the true flow field is known and can be used to quantify the performance of the approach by comparing the true and the estimated flow field.

As error measure we used the angular error (AE) which is very common in the field of computer vision to evaluate optical flow techniques. It can be understood as the angular deviation of the true and the estimated flow vector each depicted as 3D orientation vectors in a spatio-temporal domain (Barron et al., 1994). The spatio-temporal domain consists of the three dimensions: displacement in x-direction, displacement in y-direction, and time difference between subsequent images (e.g. 1).

The strength of the AE is that it takes into account both, direction and magnitude differences of the estimated flow vector compared to the true vector which makes it a convenient all-in-one error measure. The average angular error (AAE) is given by the mean AE and yields one number for the whole flow field.

The first test case consists of a synthetic image sequence of the simulated, two dimensional flow over a backward facing step (BFS). The sequence contains some interesting flow aspects such as a sharp separation with a steep velocity profile in the free mixing layer and large rotations in the recirculation area. Due to its simple geometry and its easy reproducibility it is a traditional test case for many experimental techniques and numerical simulations and, therefore, well understood.

On the top left of Fig. 4 the flow field over the backward facing step is depicted. The magnitude of the flow is color coded and the black square indicates the region used in the later comparison of the angular error of the local and the global learning-based approach (see Fig. 6). The flow was simulated with the finite element software deal.II (Bangerth et al., 2007) for Reynolds number 1000 and maximum velocity 1.5 px/frame. The image sequence was constructed by warping a gray value pattern consisting of simulated particles of changing size, shape, and intensity with the simulated flow fields. One frame of the synthetically generated image sequence is shown on the top right of Fig. 4.

As second test case synthetic images of a self sustained two dimensional turbulent flow provided by (Carlier, 2005) were used. One flow field and one image frame of the sequence are depicted on the bottom of Fig. 4.

The global variational learning-base approach (LBA-glob) was compared to the local learning-based approach (LBA-loc) proposed in (Stapf & Garbe, 2014) and to other optical flow and correlation-based PIV algorithms. The structure tensor approach (STA) (Bigtin et al., 1991) uses a simple flow model of constant flow and is similar to the local approach propose by (Lucas & Kanade, 1981) but applies total least squares instead of simple least squares. The affine model
Fig. 4: Top row: Simulated flow field of the BFS (left) and one frame of the synthetic image sequence (right). Bottom row: Flow field of the 2D turbulent sequence (left) and one frame of the image sequence (right). In the areas marked by the black squares the angular error of the variational and the local learning-based approach is compared (see Fig. 6).

Fig. 5: Average angular error (AAE) for different noise ratios and different methods for the backward facing step (left) and the two dimensional turbulent sequence (right).

approach (AMA) (Haussecker & Spies, 1999) uses an affine model with six parameters. The approach from Horn and Schunck (HSA) (Horn & Schunck, 1981) is the standard global optical flow approach and uses a smoothness constraint as regularization term. To perform correlation-
Fig. 6: Angular error in the areas marked in Fig. 4 for the backward facing step sequence (top row) and the 2D turbulent sequence (bottom row). On the left side the results from the local learning-based approach are shown and on the right side the results of the variational learning-based approach are depicted.

Based PIV the software package fluere 1.3 by Kyle P. Lynch, based on an iterative image deformation algorithm proposed in (Scarano & Riethmuller, 1999) was used. The results are shown in Fig. 5 for the BFS (left) and the 2D turbulence (right). Depicted is the average angular error of the approaches for growing noise levels of the image sequences. In order to obtain the noisy images Gaussian noise with varying standard deviation was added to the sequences. Therefore, the noise-to-signal ratio is given by $\text{NSR} = \sigma/\mu$ with standard deviation $\sigma$ and signal mean $\mu$.

For both learning-based approaches the correct flow fields were used as training data. However, the results presented in (Stapf & Garbe, 2014) show that the influence of the training data on the resulting flow fields is only marginal as long as important flow features are present in the training data. Therefore, using correct flow fields from other sequences or flow fields computed with PIV as training data yields similar results.

Both learning-based approaches perform better than the optical flow and correlation-based PIV methods for the test cases. For each method and each noise level all parameters such as model size,
The performance of the approach is very good compared to other methods by choice of training data. Smoothing parameters etc. were optimized individually. Since spatio-temporal motion models were applied within the optical flow approaches PIV was also utilized in multiframe correlation mode, where a sliding ensemble correlation is used over a small kernel of images.

Fig. 5 indicates that the flow fields estimated with the variational learning-based approach are slightly better than the flow fields estimated with the local learning-based approach. This can also be seen in Fig. 6 where the angular error of both methods is compared visually. Due to the applied smoothness constraint the angular error of the global approach is much smoother than from the local approach. Therefore, large spikes of the angular error received from the local approach are smoothed down by the global approach. This means that in areas were the estimation of the optical flow is somewhat problematic, e.g. due to a lack of gray value structures, information is interpolated from the surrounding which leads to a lower error.

Additional information about the underlying flow structures is provided by the motion models together with the parameters. The flow field is composed of typical flow structures given by the learned motion models. The parameters indicate the weight of the flow structures at every location. Fig. 7 shows the first five motion models together with the appropriate parameter images.

The third motion model for instance displays a rotation flow and the parameter image shows where rotations are present. To check if rotations are present at the indicated locations, the parameter image of the rotation mode was compared to the curl of the estimated flow field given by \( \nabla \times u = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \) in Fig. 8.

5 Conclusion

A novel variational learning-based optical flow approach for the estimation of fluid flow fields was presented. The estimated flow fields are restricted to the space spanned by previously learned typical motion models. These motion models comprise complex spatio-temporal flow patterns and are represented by the first high-energy POD modes. In this way prior knowledge is incorporated on local flow field patches. The approach is quite general since it can easily be adapted to all types of flow applications by choice of training data. The tests on two synthetically generated particle image sequences showed that the performance of the approach is very good compared to other optical flow and PIV methods. Error measures, comparing the true and the estimated velocity field, show significantly lower values for the learning-based approach. The method is also more
robust to noise than all the other tested methods. The flow field is decomposed into typical flow structures, the motion models, and associated parameters. By this decomposition, additional information about the type of flow is provided. Location and class of dominant flow structures are revealed.

6 References


