On the Spatial Statistics of Optical Flow

Stefan Roth Michael J. Black
Department of Computer Science, Brown University, Providence, RI, USA
{roth,black}@cs.brown.edu

Abstract

We develop a method for learning the spatial statistics of optical flow fields from a novel training database. Training flow fields are constructed using range images of natural scenes and 3D camera motions recovered from hand-held and car-mounted video sequences. A detailed analysis of optical flow statistics in natural scenes is presented and machine learning methods are developed to learn a Markov random field model of optical flow. The prior probability of a flow field is formulated as a Field-of-Experts model that captures the higher order spatial statistics in overlapping patches and is trained using contrastive divergence. This new optical flow prior is compared with previous robust priors and is incorporated into a recent, accurate algorithm for dense optical flow computation. Experiments with natural and synthetic sequences illustrate how the learned optical flow prior quantitatively improves flow accuracy and how it captures the rich spatial structure found in natural scene motion.

Figure 1. Flow fields generated for an outdoor (top) and an indoor scene (bottom). The horizontal motion $\mathbf{u}$ is shown on the left, the vertical motion $\mathbf{v}$ on the right; dark/light means negative motion/positive motion (scaled to $0 \ldots 255$ for display).

1. Introduction

In this paper we study the spatial statistics of optical flow in natural imagery and exploit recent advances in machine learning to obtain a rich prior model for optical flow fields. This extends work on the analysis of image statistics in natural scenes and range images to the domain of image motion. In doing so we make connections to previous robust statistical formulations of optical flow smoothness priors and learn a new Markov random field prior over large neighborhoods using a *Field-of-Experts* model [24]. We extend a recent (and very accurate) optical flow method [7] with this new prior and provide an algorithm for estimating optical flow from pairs of images. We quantitatively compare the learned prior with more traditional robust priors and find that in our experiments the accuracy is improved by about 10% while removing the need for tuning the scale parameter of the traditional priors.

Natural image statistics have received intensive study [17], but the spatial statistics of optical flow are relatively unexplored because databases of natural scene motions are currently unavailable. One of the contributions of this paper is the development of such a database.

The spatial statistics of optical flow are determined by the interaction of 1) camera motion; 2) scene depth; and 3) the independent motion of objects. Here we focus on rigid scenes and leave independent motion for future work (though we believe the statistics from rigid scenes are useful for scenes with independent motion). To generate a realistic database of optical flow fields we exploit the Brown range image database [1], which contains depth images of complex scenes including forests, indoor environments, and generic street scenes. Given 3D camera motions and range images we generate flow fields that have the rich spatial statistics of natural flow fields. A set of natural 3D motions was obtained from both hand-held and car-mounted cameras performing a variety of motions including translation, rotation, and active fixation. The 3D motion was recovered from these video sequences using commercial software [2]. Figure 1 shows two example flow fields generated using the 3D motions and the range images.

The first-derivative statistics of optical flow exhibit highly kurtotic behavior as do the statistics of natural images. We observe that the first derivative statistics are well modelled by heavy tailed distributions such as the Student-t distribution. This provides a connection to previous robust statistical methods for recovering optical flow that modelled spatial smoothness using robust functions [6] and suggests that the success of robust methods is due to the fact that they capture the first order statistics of optical flow.

Our goal here is to go beyond such local (first derivative) models and formulate a Markov random field (MRF) prior that captures richer spatial statistics present in larger neighborhoods. To that end, we exploit a "Field of Experts" (FoE) model [24], that represents MRF clique potentials in terms of various linear filter responses on each clique. We model these potentials as a product of t-distributions and we learn both the parameters of the distribution and the filters themselves using contrastive divergence [15, 24].

We compute optical flow using the learned prior as a smoothness term. The log prior is combined with a data term and we minimize the resulting energy (log posterior). While the exact choice of data term is not relevant for the analysis, here we use the recent approach of Bruhn *et al.* [7], which replaces the standard optical flow constraint equation with a tensor that integrates brightness constancy constraints over a spatial neighborhood. We present an algorithm for estimating dense optical flow and compare the performance of standard robust spatial terms with the learned FoE model on both synthetic and natural imagery.

1.1. Previous work

There has been a great deal of work on modeling natural image statistics [17] facilitated by the existence of large image databases. One might expect optical flow statistics to differ from image statistics in that there is no equivalent of "surface markings" in motion and all structure in rigid scenes results from the shape of surfaces and the discontinuities between them. In this way it seems plausible that flow statistics share more with depth statistics. Unlike optical flow, direct range sensors exist and a time-of-flight laser was used in [18] to capture the depth in a variety of scenes including residential street scenes, forests, and indoor environments. Scene depth statistics alone, however, are not sufficient to model optical flow, because image motion results from the combination of the camera motion and depth. While models of self-motion in humans and cats [4] have been studied, we are unaware of attempts to learn or exploit a database of camera motions captured by a moving camera in natural scenes.

The most similar work to ours also uses the Brown range image database to generate realistic synthetic flow fields [8]. The authors use a gaze tracker to record how people view the range images and then simulate their motion into the scene with varying fixation points. Their focus is on human perception of flow and consequently they analyze a retinal projection of the flow field. They also limit their analysis to first order statistics and do not propose an algorithm for exploiting these statistics in the computation of optical flow.

Previous work on learning statistical models of video focuses on the statistics of the changing brightness patterns rather than the flow it gives rise to. For example, adopting a classic sparse-coding hypothesis, video sequences can be represented using a set of learned spatio-temporal filters [27]. Other work has focused on the statistics of the classic brightness constancy assumption (and how it is violated) rather than the spatial statistics of the flow field [10, 25].

The lack of training data has limited research on learning spatial models of optical flow. One exception is the work by Fleet *et al.* [11] in which the authors learn local models of optical flow from examples using principal component analysis (PCA). In particular, they use synthetic models of moving occlusion boundaries and bars to learn linear models of the flow for these motion features. Local, non-overlapping models such as these may be combined in a spatio-temporal Bayesian network to estimate coherent global flow fields [12]. While promising, these models cover only a limited range of the variation in natural flow fields.

There is related interest in the statistics of optical flow in the video retrieval community; for example, Fablet and Bouthemy [9] learn statistical models using a variety of motion cues to classify videos based on their spatio-temporal statistics. These methods, however, do not focus on the estimation of optical flow.

The formulation of smoothness constraints for optical flow estimation has a long history [16], as has its Bayesian formulation in terms of Markov random fields [5, 14, 20, 22]. Previous work, however, has focused on very local models that are typically formulated in terms of the first differences in the optical flow (i. e., the nearest neighbor differences). This can model piecewise constant or smooth flow but not more complex spatial statistics. Other work has imposed geometric rather than spatial smoothness constraints on multi-frame optical flow [19].

Weiss and Adelson [28] propose a Bayesian model of motion estimation to explain human perception of visual stimuli. In particular, they argue that the appropriate image prior prefers "slow and smooth" motion. Their stimuli, however, are too simplistic to probe the nature of flow priors in complex scenes. We find these statistics are more like those of natural images in that the motions are piecewise smooth; large discontinuities give rise to heavy tails in the first derivative marginal statistics.

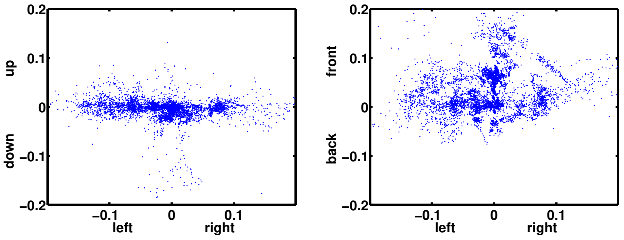

Figure 2. Scatter plot of the translational camera motion: *(left)* Left-right (x) and up-down motion (y). *(right)* Left-right (x) and forward-backward (z) motion (scale in meters).

2. Spatial Statistics of Optical Flow

2.1. Obtaining training data

One of the key challenges in learning the spatial statistics of optical flow is to obtain suitable training data. The issue here is that optical flow cannot be directly measured, which makes the statistics of optical flow a largely unexplored field. Synthesizing realistic flow fields is thus the only viable option for studying, as well as learning the statistics of optical flow. Our goal is to create a database of optical flow fields as they arise in natural as well as man-made scenes. It is unlikely that the rich statistics will be captured by any manual construction of the training data as in [11]. Instead we rely on range images from the Brown range image database [1], which provides accurate scene depth information for a set of 197 indoor and outdoor scenes. While this database captures information about surfaces and surface boundaries in natural scenes, it is completely static. A rigorous study of the optical flow statistics of independently moving objects will remain the subject of future work. While we focus on rigid scenes, the range of motions represented is broad and varied.

Apart from choosing appropriate 3D scenes, finding suitable camera motion data is another challenge. In order to cover a broad range of possible frame-to-frame camera motions, we used a database of 100 video clips of approximately 100 frames, each of which was shot using a hand-held or car-mounted video camera. The database is comprised of various kinds of motion, including forward walking and moving the camera around an object of interest. The extrinsic and intrinsic camera parameters were recovered using the *boujou* software system [2]. Figure 2 shows empirical distributions of the camera translations in the database. The plots reveal that left-right movements are more common than up-down movements and that moving into the scene occurs more frequently than moving out of the scene. Similarly the empirical distributions of camera rotation reveal that left-right panning occurs more frequently than up-down panning and tilting of the camera.

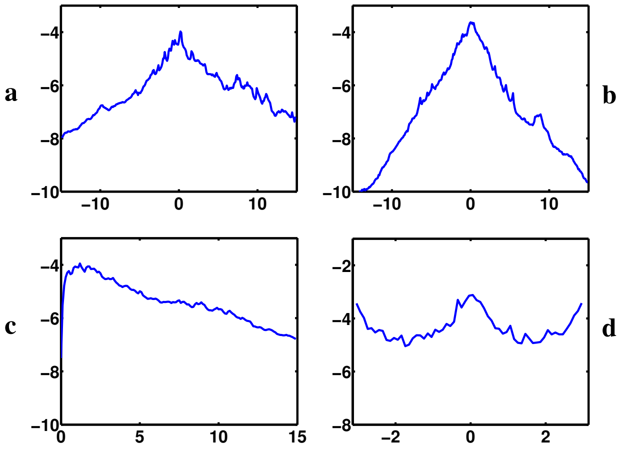

Figure 3. Simple statistics of our optical flow database: log-histograms of **(a)** horizontal velocity u, **(b)** vertical velocity v, **(c)** velocity r, **(d)** orientation θ.

To generate optical flow from this motion data we use the following procedure: We pick a random range image, and a random camera motion. A ray is then cast through every pixel of the first frame to find the corresponding scene depth from the range image. Each of these scene points is then projected onto the image plane for the second frame. The optical flow is simply given by the difference in image coordinates under which a scene point is viewed in each of the two cameras. We used this procedure to generate a database of 400 optical flow fields, each 360×256 pixels large[1]. The position of the camera and therefore the distance to the scene is determined by the position of the range finder. Nevertheless, in order to cover a range of possible scenarios, one can scale the magnitude of the translational motion. Figure 1 shows example flow fields from this database. Note that we do not explicitly represent the regions of occlusion or disocclusion in the database. A rigorous treatment of occlusions, which will most likely require a more sophisticated scene geometry model, will remain future work (c.f. [12, 23]).

2.2. Velocity statistics

Using the database, we study several statistical properties of optical flow. Figure 3 shows log-histograms of the image velocities in various forms. We observe that the vertical velocity in *(b)* is roughly distributed like a Laplacian distribution; however the horizontal velocity in *(a)* shows a broader histogram that falls off less quickly. This is consistent with our observations that horizontal camera motions

[1] The database is available to other researchers and can be obtained by contacting the authors.

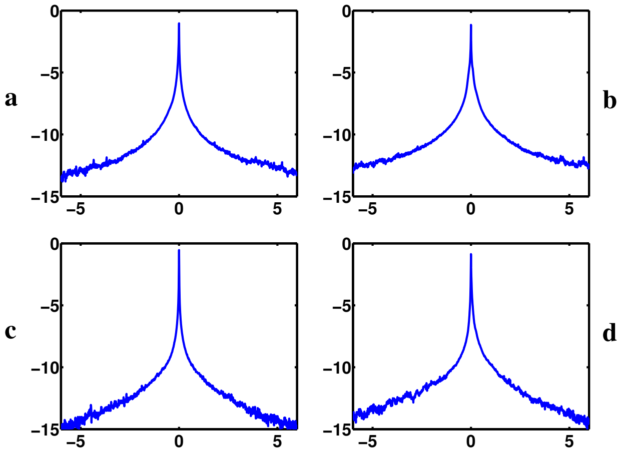

Figure 4. Derivative statistics of optical flow: log-histograms of **(a)** $\partial u/\partial x$, **(b)** $\partial u/\partial y$, **(c)** $\partial v/\partial x$, **(d)** $\partial v/\partial y$.

seem to be more common. Figure *(c)* shows the magnitude of the velocity, which falls off similar to a Laplacian distribution. We can also see that very small motions seem to occur rather infrequently, which suggests that the camera is rarely totally still. The orientation histogram in *(d)* again shows the preference for horizontal motion (the bumps at 0 and $\pm\pi$). However, there are also smaller spikes indicating somewhat frequent up-down motion (at $\pm\pi/2$).

2.3. Derivative statistics

Figure 4 shows the first derivative statistics of both horizontal and vertical image motion. The distributions are all heavy-tailed and strongly resemble Student t-distributions. Such distributions have also been encountered in the study of natural images, e. g., [17]. In natural images, the image intensity is often locally smooth, but occasionally shows large jumps at object boundaries or in fine textures, which give rise to substantial probability mass in the tails of the distribution. Furthermore, the study of range images [18] has shown similar derivative statistics. For scene depth the heavy-tailed distributions arise from depth discontinuities mostly at object boundaries. Because the image motion from camera translations is directly dependent on the scene depth, it is not surprising to see similar distributions for optical flow. The observed derivative statistics likely explain the success of robust statistical formulations for optical flow computation. The Lorentzian robust error function in particular, as used for example in [5], matches the empirical statistics very well.

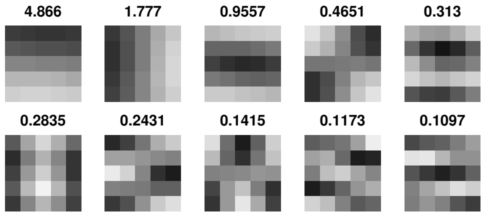

Figure 5. First 10 principal components of the horizontal velocity u in 5×5 patches. The numbers denote the variance of each principal component.

2.4. Principal component analysis

We also performed principal component analysis on small image patches of various sizes. Figure 5 shows the results for horizontal flow $\mathbf{u}$ in 5×5 patches. The principal components of the vertical flow $\mathbf{v}$ look very similar, but their variance is smaller due to the observed preference for horizontal motion. We can see that a large portion of the image energy is focused on the first few principal components which, as with images, resemble derivative filters of various orders (c.f. [11]).

3. Learning the Spatial Statistics

We learn the spatial statistics of optical flow using the recently proposed *Fields-of-Experts* model [24]. It models the prior probability of images (and here optical flow fields) using a Markov random field. In contrast to many previous MRF models, it uses larger cliques of for example 3×3 or 5×5 pixels, and allows learning the appropriate clique potential from training data. We argue here that spatial regularization of optical flow will benefit from prior models that capture interactions beyond adjacent pixels. For simplicity, we will treat horizontal and vertical image motions separately, and learn two independent models.

In the Markov random field framework, the pixels of an image or flow field are assumed to be represented by nodes V in a graph $G = (V, E)$, where E are the edges connecting nodes. The edges are typically defined through a neighborhood system, such as all spatially adjacent pairs of pixels. We will instead consider neighborhood systems that connect all nodes in a square $m \times m$ region. Every such neighborhood centered on a node (pixel) $k = 1, \ldots, K$ defines a maximal clique $\mathbf{x}_{(k)}$ in the graph. We write the probability of a flow field component $\mathbf{x} \in \{\mathbf{u}, \mathbf{v}\}$ under the MRF as

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{k=1}^{K} \psi(\mathbf{x}_{(k)}), \tag{1}$$

where $\psi(\mathbf{x}_{(k)})$ is the so-called potential function for clique $\mathbf{x}_{(k)}$ (we assume homogeneous MRFs here) and Z is a normalization term. Because of our assumptions, the joint probability of the flow field $p(\mathbf{u}, \mathbf{v})$ is simply the product of the probabilities of the two components $p(\mathbf{u}) \cdot p(\mathbf{v})$ under the MRF model. Most typical prior models for optical flow can be written this way, as they are based on first derivatives that can be approximated by differences of pixel neighbors.

When considering MRFs with cliques that do not just consist of pairs of nodes, finding suitable potential functions $\psi(\mathbf{x}_{(k)})$ and training the model on a database become much more challenging. In our experiments we have observed that linear filter responses on the optical flow database show histograms that are typically well fit by t-distributions. Based on this observation, the FoE model that we propose uses the Products-of-Experts framework [26], and models the clique potentials with products of Student t-distributions, where each expert distribution works on the response to a linear filter $\mathbf{J}_i$. The cliques' potential under this model is written as:

$$\psi(\mathbf{x}_{(k)}) = \prod_{i=1}^{N} \phi(\mathbf{J}_i^{\mathrm{T}} \mathbf{x}_{(k)}; \alpha_i), \tag{2}$$

where each expert is a t-distribution with parameter α_i:

$$\phi_i(\mathbf{J}_i^{\mathrm{T}} \mathbf{x}_{(k)}; \alpha_i) = \left(1 + \frac{1}{2}(\mathbf{J}_i^{\mathrm{T}} \mathbf{x}_{(k)})^2\right)^{-\alpha_i}. \tag{3}$$

The FoE optical flow prior is hence written as

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{k=1}^{K} \prod_{i=1}^{N} \phi(\mathbf{J}_i^{\mathrm{T}} \mathbf{x}_{(k)}; \alpha_i). \tag{4}$$

The filters $\mathbf{J}_i$ as well as the expert parameters α_i are jointly learned from training data using maximum likelihood estimation. Because there is no closed form expression for the partition function Z in case of the FoE model, maximum likelihood estimation relies on Markov chain Monte Carlo sampling, which makes the training process computationally expensive. To speed up the learning process, we use the contrastive divergence algorithm [15], which approximates maximum likelihood learning, but only relies on a fixed, small number of Markov chain iterations in the sampling phase. For our experiments, we trained FoE models with 8 filters of 3×3 pixels on randomly selected and cropped flow fields from the training database. We restrict the filters so that they do not capture the mean velocity in a patch and are thus only sensitive to relative motion. The use of larger patches or more filters may yield performance improvements, but this will remain future work. For more details about how FoE models can be trained, we refer the reader to [24].

In the context of optical flow estimation, the prior knowledge about the flow field is typically expressed in terms of an energy function $E(\mathbf{x}) = -\log p(\mathbf{x})$. Accordingly, we can express the energy for the FoE prior model as

$$E_{\text{FoE}}(\mathbf{x}) = -\sum_{k=1}^{K} \sum_{i=1}^{N} \log \phi(\mathbf{J}_i^{\mathrm{T}} \mathbf{x}_{(k)}; \alpha_i) + \log Z. \tag{5}$$

Note that that for fixed parameters α and $\mathbf{J}$, the partition function Z is constant, and can thus be ignored when estimating flow.

The optical flow estimation algorithm we propose in the next section relies on the gradient of the energy function with respect to the flow field. Since Fields of Experts are a log-linear model, expressing and computing the gradient of the energy is relatively easy. Following Zhu and Mumford [29], the gradient of the energy can be written as

$$\nabla_{\mathbf{x}} E_{\text{FoE}}(\mathbf{x}) = -\sum_{i=1}^{N} \mathbf{J}_i^{-} * \xi_i(\mathbf{J}_i * \mathbf{x}), \tag{6}$$

where $\mathbf{J}_i * \mathbf{x}$ denotes the convolution of image velocity component $\mathbf{x}$ with filter $\mathbf{J}_i$. We also define $\xi_i(y) = \partial/\partial y \; \log \phi(y; \alpha_i)$ and let $\mathbf{J}_i^{-}$ denote the filter obtained by mirroring $\mathbf{J}_i$ around its center pixel [29].

4. Optical Flow Estimation

In order to demonstrate the benefits of learning the spatial statistics of optical flow, we integrate our model with a recent, competitive optical flow method and quantitatively compare the results. As baseline algorithm, we chose the combined local-global method (CLG) as proposed by Bruhn *et al.* [7]. Global methods typically estimate the horizontal and vertical image velocities $\mathbf{u}$ and $\mathbf{v}$ by minimizing an energy of the form [16]

$$\begin{aligned} E(\mathbf{u}, \mathbf{v}) = \int_I & \rho_D(I_x \mathbf{u} + I_y \mathbf{v} + I_t) + \\ & \lambda \cdot \rho_S(\sqrt{|\nabla \mathbf{u}|^2 + |\nabla \mathbf{v}|^2}) \, \mathrm{d}x \, \mathrm{d}y. \end{aligned} \tag{7}$$

ρ_D and ρ_S are robust penalty functions, such as the Lorentzian [6]; I_x, I_y, I_t denote the spatial and temporal derivatives of the image sequence. The first term in (7) is the so-called data term that enforces the brightness constancy assumption. The second term is the so-called spatial term, which enforces (piecewise) spatial smoothness. Since in this model, the data term relies on a local linearization of the optical flow constraint equation (OFCE), such methods are usually used in a coarse-to-fine fashion, e. g., [6], which allows the estimation of large displacements. For the remainder, we will assume that large image velocities are handled using such a coarse-to-fine scheme.

The combined local-global method extends this framework through local averaging of the brightness constancy constraint by means of a structure tensor. This connects the method to local optical flow approaches that use locally smoothed image derivatives to estimate the image velocity. Using $\nabla I = (I_x, I_y, I_t)^{\mathrm{T}}$ we can define a spatio-temporal structure tensor as $J_\sigma(I) = G_\sigma * \nabla I \nabla I^{\mathrm{T}}$, where G_σ denotes a Gaussian convolution kernel with width σ. The CLG approach estimates the optical flow by minimizing

$$E_{\mathrm{CLG}}(\mathbf{w}) = \int_I \rho_D(\sqrt{\mathbf{w}^{\mathrm{T}} J_\sigma(I)\mathbf{w}}) + \lambda \cdot \rho_S(\sqrt{|\nabla \mathbf{w}|^2}) \, \mathrm{d}x \, \mathrm{d}y, \tag{8}$$

where $\mathbf{w} = (\mathbf{u}, \mathbf{v}, 1)^{\mathrm{T}}$. Experimentally, the CLG approach has been shown to be one of the best currently available optical flow estimation techniques. The focus of this paper are the spatial statistics of optical flow; hence, we will only make use of the 2D-CLG approach, i. e., only two adjacent frames will be used for flow estimation.

We will further refine the CLG approach by using a spatial regularizer that is based on the learned spatial statistics of optical flow. Many global optical flow techniques enforce spatial regularity or "smoothness" by penalizing large spatial gradients. In Section 3 we have shown how to learn higher order Markov random field models of optical flow, which we use here as spatial regularizer for flow estimation. Here, our objective is to minimize the energy

$$E(\mathbf{w}) = \int_I \rho_D(\sqrt{\mathbf{w}^{\mathrm{T}} J_\sigma(I)\mathbf{w}}) \, \mathrm{d}x \, \mathrm{d}y + \lambda \cdot E_{\mathrm{FoE}}(\mathbf{w}). \tag{9}$$

Since $E_{\mathrm{FoE}}(\mathbf{w})$ is non-convex, minimizing (9) is generally difficult. Depending on the choice of the robust penalty function, the data term may in fact be non-convex, too. We will not attempt to find the global optimum of the energy function, but instead perform a simple local optimization. At any local extremum of the energy it holds that

$$\nabla_{\mathbf{w}} E(\mathbf{w}) = 0. \tag{10}$$

We can discretize this constraint using (6) for the spatial term and the discretization from [7] for the data term. The discretized constraint can be written as

$$\mathbf{A}(\mathbf{w})\mathbf{w} = \mathbf{b}, \tag{11}$$

where $\mathbf{A}(\mathbf{w})$ is a large, sparse matrix that depends on $\mathbf{w}$, which is a vector of all x- and y-velocities in the image. In order to solve for $\mathbf{w}$, we make (11) linear by keeping $\mathbf{A}(\mathbf{w})$ fixed, and solve the resulting linear equation system using a standard technique such as GMRES [13]. This procedure is then iterated until a fixed point is reached.

4.1. Evaluation

To evaluate the proposed method, we performed a series of experiments with both synthetic and real data. The quantitative evaluation of optical flow techniques suffers from the problem that only a few image sequences with ground truth optical flow data are available. The first part of our evaluation thus relies on synthetic test data. To provide realistic image texture we randomly sampled 25 (intensity) images from a database [21], cropped them to 100×100 pixels, and warped the images with randomly sampled, synthesized flow that was generated in the same fashion as the training data to obtain 25 input frame pairs.

Method	AAE
(1) Quadratic	2.93°
(2) Charbonnier	1.70°
(3) Charbonnier + Lorentzian	1.76°
(4) FoE + Lorentzian	1.32°

Table 1. Results on synthetic test data: Average angular error (AAE) for best parameters.

We ran 4 different algorithms on all the test image pairs: *(1)* The 2D-CLG approach with quadratic data and spatial terms; *(2)* The 2D-CLG approach with Charbonnier data and spatial terms as used in [7]; *(3)* The 2D-CLG approach with Lorentzian data term and Charbonnier spatial term; *(4)* The 2D-CLG approach with Lorentzian data term and FoE spatial term. The Charbonnier robust error function has the form $\rho(x) = 2\beta^2\sqrt{1 + x^2/\beta^2}$, where β is a scale parameter. The Lorentzian robust error function is related to the t-distribution and has the form $\rho(x) = \log(1 + \frac{1}{2}(x/\sigma)^2)$, where σ is its scale parameter. For all experiments in this paper, we chose a fixed integration scale for the structure tensor ($\sigma = 1$). For methods *(2)* and *(3)* we tried $\beta \in \{0.05, 0.01, 0.005\}$[2] for both the spatial and the data term and report the best results. For methods *(3)* and *(4)* we fixed the scale of the Lorentzian for the data term to 0.5. For each method we chose a set of 5 different λ values (in a suitable range), which control the relative weight of the spatial term. Table 1 shows the average angular error in degrees (see [3]) averaged over the whole test data set. The error is reported for the λ value that gave the lowest average error on the whole data set, i. e., the parameters are not tuned to each individual test case. Table 1 summarizes the results and shows that the FoE flow prior improves the flow estimation error on this synthetic test database. In contrast to methods *(2)* and *(3)* the FoE prior does not require any tuning of the parameters of the prior. Only the λ value requires tuning for all 4 techniques.

In a second experiment, we learned an FoE flow prior for the Yosemite sequence [3] containing a computer generated image sequence (version without the cloudy sky). First we trained the FoE prior on the ground truth data for the Yosemite sequence, omitting frames 8 and 9 which were used for evaluation. Estimating the flow with the learned model and the same data term as above gives an average

[2]This parameter interval is suggested in [7].

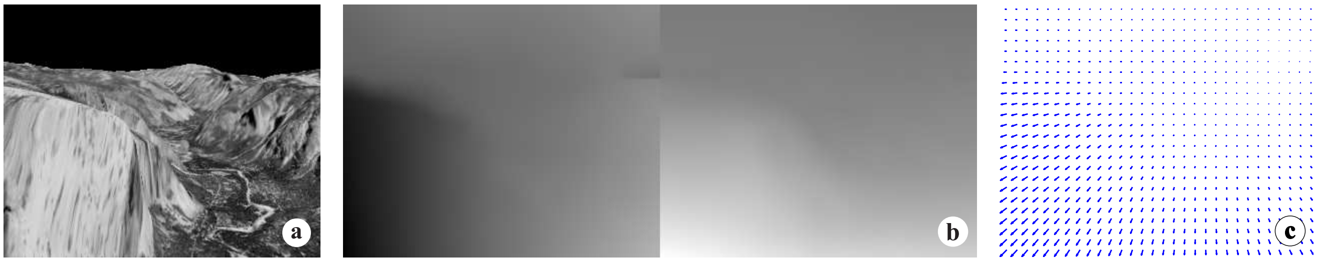

Figure 6. Optical flow estimation: Yosemite fly-through. *(a)* Frame 8 from image sequence. *(b)* Estimated optical flow with separate $\mathbf{u}$ and $\mathbf{v}$ components. Average angular error $1.47°$. *(c)* Estimated optical flow as vector field.

angular error of $1.47°$ (standard deviation $1.54°$). This is $0.15°$ better than the result reported for the standard two frame CLG method (see [7]), which is as far as we are aware the current best result for a two frame method. While training on the remainder of the Yosemite sequence may initially seem unfair, most other reported results for this sequence rely on tuning the method's parameters so that one obtains the best results on a particular frame pair. Figure 6 shows the computed flow field. We can see that it seems rather smooth, but given that the specific training data contains only very few discontinuities, this is not very surprising. In fact, changing the λ parameter of the algorithm so that edges start to appear leads to numerically inferior results.

An important question for a learned prior model is how well it generalizes. To evaluate this, we used the model trained on our synthetic flow data to estimate the Yosemite flow. Using the same parameters described above, we found that the accuracy of the flow estimation results decreased to $1.82°$ average angular error (standard deviation $1.61°$). This suggests that our training database is not representative for the kinds of geometries or motions that occur in the Yosemite sequence. Furthermore this suggests that particular care must be taken when designing a representative optical flow database.

In a final experiment, we evaluated the FoE flow prior on a real image sequence. Figure 7 shows the first frame from this "flower garden" sequence. The sequence has two dominant motion layers, a tree in the foreground and a background, with different image velocities. We applied the FoE model as trained on the synthetic flow database and used the parameters as described above for model *(4)*. Figure 7 shows the obtained flow field, which qualitatively captures the motion and object boundaries well.

5. Conclusions and Future Work

We have presented a novel database of optical flow as it arises when realistic scenes are captured with a hand-held or car-mounted video camera. This database allowed us to study the spatial statistics of optical flow, and furthermore enabled us to learn prior models of optical flow using the *Fields-of-Experts* framework. We have integrated the FoE flow prior into a recent, accurate optical flow algorithm and obtained moderate accuracy improvements. In our current work, we are training models with larger filters such as 5×5 and joint models of horizontal and vertical motion which we hope will further increase accuracy. While our experiments suggest that the training database may not yet be representative of the image motions in certain sequences, we believe that this is an important step towards studying and learning the spatial statistics of optical flow.

There are many opportunities for future work that build on the proposed prior and the database of flow fields. For example, Calow *et al.* [8] point out that natural flow fields are inhomogeneous; for example, in the case of human motion, the constant presence of a ground plane produces quite different flow statistics in the lower portion of the image than in the upper portion. Here we proposed a homogeneous flow prior but it is also possible, with sufficient training data, to learn an inhomogeneous FoE model.

It may be desirable to learn application-specific flow priors (e. g., for automotive applications). This suggests the possibility of learning multiple categories of flow priors and using these to classify scene motion for applications in video databases.

A natural extension of our work is the direct recovery of structure from motion. We can exploit our training set of camera motions to learn a prior over 3D camera motions and combine this with a spatial prior learned from the range imagery. The prior over 3D motions may help regularize what is often a difficult problem given the narrow field of view and small motions present in common video sequences.

Future work must also consider the statistics of independent, textural, and non-rigid motion. Here obtaining ground truth is more problematic. Possible solutions involve obtaining realistic synthesized sequences from the film industry or hand-marking regions of independent motion in real image sequences.

Finally, a more detailed analysis of motion boundaries is warranted. In particular, our current flow prior does not explicitly encode information about the occluded/unoccluded surfaces or the regions of the image undergoing dele-

Figure 7. Optical flow estimation: Flower garden sequence. *(a)* Frame 1 from image sequence. *(b)* Estimated optical flow with separate $\mathbf{u}$ and $\mathbf{v}$ components.

tion/accretion. Future work may also explore the problem of jointly learning motion and occlusion boundaries using energy-based models such as the Fields of Experts.

Acknowledgements We would like to thank David Capel (2d3) and Andrew Zisserman for providing camera motion data, Andrés Bruhn for providing additional details about the algorithm from [7], John Barron for clarifying details about the Yosemite sequence, and Frank Wood for helpful discussions. This research was supported by Intel Research and NIH-NINDS R01 NS 50967-01 as part of the NSF/NIH Collaborative Research in Computational Neuroscience Program.

References

[1] Brown range image database. `http://www.dam.brown.edu/ptg/brid/index.html`.

[2] boujou, 2d3 Ltd., 2002. `http://www.2d3.com`.

[3] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *IJCV*, 12(1):43–77, 1994.

[4] B. Y. Betsch, W. Einhäuser, K. P. Körding, and P. König. The world from a cat's perspective - Statistics of natural videos. *Biological Cybernetics*, 90(1):41–50, 2004.

[5] M. J. Black and P. Anandan. Robust dynamic motion estimation over time. *CVPR*, pp. 296–302, 1991.

[6] M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *CVIU*, 63(1):75–104, 1996.

[7] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *IJCV*, 61(3):211–231, 2005.

[8] D. Calow, N. Krüger, F. Wörgötter, and M. Lappe. Statistics of optic flow for self-motion through natural scenes. *Dynamic Perception*, pp. 133–138, 2004.

[9] R. Fablet and P. Bouthemy. Non parametric motion recognition using temporal multiscale Gibbs models. *CVPR*, vol. 1, pp. 501–508, 2001.

[10] C. Fermüller, D. Shulman, and Y. Aloimonos. The statistics of optical flow. *CVIU*, 82(1):1–32, 2001.

[11] D. J. Fleet, M. J. Black, Y. Yacoob, and A. D. Jepson. Design and use of linear models for image motion analysis. *IJCV*, 36(3):171–193, 2000.

[12] D. J. Fleet, M. J. Black and O. Nestares, Bayesian inference of visual motion boundaries. *Exploring Artificial Intelligence in the New Millennium*, Lakemeyer, G. and Nebel, B. (Eds.), Morgan Kaufmann, July 2002.

[13] G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins University Press, 1996.

[14] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using markov random fields. *PAMI*, 15(12):1217–1232, 1993.

[15] G. E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Comp.*, 14:1771–1800, 2002.

[16] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1–3):185–203, 1981.

[17] J. Huang. *Statistics of Natural Images and Models*. PhD thesis, Brown University, 2000.

[18] J. Huang, A. B. Lee, and D. Mumford. Statistics of range images. *CVPR*, vol. 1, pp. 1324–1331, 2000.

[19] M. Irani. Multi-frame optical flow estimation using subspace constraints. *ICCV*, vol. 1, pp. 626–633, 1999.

[20] J. Konrad and E. Dubois. Multigrid Bayesian estimation of image motion fields using stochastic relaxation. *ICCV*, pp. 354–362, 1988.

[21] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *ICCV*, vol. 2, pp. 416–423, 2001.

[22] D. W. Murray and B. F. Buxton. Scene segmentation from visual motion using global optimization. *PAMI*, 9(2):220–228, 1987.

[23] M. G. Ross and L. P. Kaelbling. A systematic approach to learning object segmentation from motion. *NIPS 2003 Workshop on Open Challenges in Cognitive Vision*, 2003.

[24] S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. *CVPR*, vol. 2, pp. 860–867, 2005.

[25] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. *CVPR*, pp. 310–315, 1991.

[26] Y. W. Teh, M. Welling, S. Osindero, and G. E. Hinton. Energy-based models for sparse overcomplete representations. *JMLR*, 4(Dec):1235–1260, 2003.

[27] J. H. van Harteren and D. L. Ruderman. Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proc. R. Soc. Lon. B*, 265(1412):2315–2320, 1998.

[28] Y. Weiss and E. H. Adelson. Slow and smooth: A Bayesian theory for the combination of local motion signals in human vision. MIT AI Memo 1624, 1998.

[29] S. C. Zhu and D. Mumford. Prior learning and Gibbs reaction-diffusion. *PAMI*, 19(11):1236–1250, 1997.