Decomposing and Regularizing Sparse/Non-sparse Components for Motion Field Estimation

Zhuoyuan Chen, Jiang Wang and Ying Wu
EECS Department, Northwestern University
2145 Sheridan Road, Tech L440, Evanston, IL
{zch318,jwa368,yingwu}@eecs.northwestern.edu

Abstract

Regularizing motion field is critical to achieve accurate estimation of the motion field. As the motion field may include discontinuity (e.g., at the motion boundaries), traditional smoothness regularization may not work well. Among many approaches to handling motion discontinuity, recent attempts pursued a sparse representation of the motion field for regularization, and achieved quite encouraging results. However, statistics show that these methods tend to over-sparsify the motion field, and thus confronted by the non-sparse noise in practice.

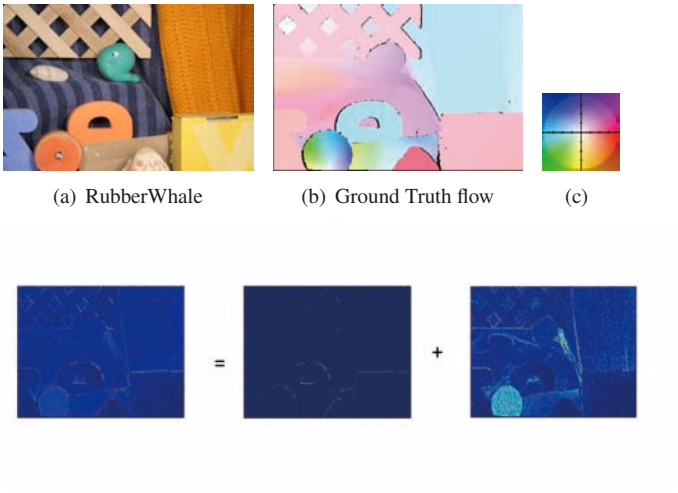
In this paper, we propose to decompose the motion field into sparse and non-sparse components for the motion boundaries and small universal noises, respectively. This separation approach regularizes these two sources differently. We propose a novel and efficient optimization algorithm to solve this problem. In addition, our study reveals the in-depth connection between this noise separation approach and the influence function approach in robust statistics. We validate and evaluate our new approach on the Middlebury benchmark, and have achieved outstanding testing performance.

1. Introduction

Motion estimation is a fundamental problem in many computer vision tasks. Originated from the classical methods [15, 19] in early 1980s, optical flow estimation has been advanced greatly and various techniques have been proposed. In general, these methods can be commonly viewed as an energy minimization problem:

$$E(\mathbf{u}) = E_D(\mathbf{u}) + E_S(\mathbf{u}), \tag{1}$$

where $\mathbf{u} = [u_x, u_y]$ denotes the horizontal and vertical components of the motion field to be estimated from consecutive images frames. E_D measures the matching of the visual features over time, or data consistency. E_S gives a regularization of the flow field, or the flow prior. In other words, E_D indicates matching error, and E_S indicates the deviation from the regularization.



(a) RubberWhale (b) Ground Truth flow (c)

(d) An illustration of our idea

Figure 1. Component decomposition. (a) "RubberWhale" [3]; (b) ground truth flow field; (d) an illustration of our idea: motion variation $|\nabla \mathbf{u}|$ can be decomposed into sparse and non-sparse components. Flow maps are visualized using color code in (c).

As an inverse problem, obtaining accurate flow field estimation is not a trivial task. Among various difficulties, motion boundary is a major hurdle confronting many flow estimation approaches. Because motion boundary gives the discontinuity in the motion field, it pretty much invalidates the commonly used spatial smoothness regularization for the motion field. It implies the need to solve the segmentation of the motion field, implicitly or explicitly.

There have been various approaches to address the motion discontinuity issue by proposing different prior models for the motion field, such as the layered motion model [28], discontinuous MRF model [22, 26], fusion model [17], etc. These methods either need to perform computationally intensive inference of latent variables, or need to differentiate

978-1-4673-1228-8/12/$31.00 ©2012 IEEE

motion boundaries from other edges. Of particular interest is a recent initial attempt to pursue a sparse representation for the motion field, even if the motion field has discontinuities at unknown places. Introducing a sparse prior for the motion field, e.g., requiring $\nabla \mathbf{u}$ to be sparse, in the regularization term E_S has produced quite encouraging results [25, 16].

Ideally, the motion field should be sparse. However, in reality simply imposing the sparsity prior on motion field [25] may not give very satisfactory results all the time. As described in later sections, imposing sparsity on motion field directly tends to over-sparsify it. This is mainly because of the influence from the matching term E_D, as in general the matching of the pixels contains errors. Such matching errors are induced from different factors and thus have different properties. One factor is the assumed matching criterion, e.g., the brightness constancy in optical flow. Complications in illumination do not make it true everywhere. In addition, the optical flow constraint is not held exactly, but as a first order approximation under a small motion. We can call it *model error*. Besides, there is another type of matching error induced by the motion boundary. At the motion boundary, some pixels that are visible in the previous frame may be occluded in the next frame. Because such pixels do not actually have their matching counterparts, forcefully and regardlessly matching them will result in a large error. We can call it *missing correspondences error*. The estimated motion is a balance between the matching error E_D and the deviation error from the regularization E_S. If the matching error in E_D is sparse, it can be absorbed by the sparse error in E_S, then the methods using sparse flow regularization work well. But when the matching error is not sparse and cannot be absorbed by E_S, the compromise between E_D and the sparsity-regularized E_S can not be made, and thus leading to an inaccurate flow estimation. Unfortunately, in reality, the matching error E_D can be non-sparse (in part due to the *model error*).

In view of this, special treatment needs to be taken for flow regularization rather than simply over-sparsify the flow field. In this paper, we propose a new regularization approach that imposes both smoothness and sparsity regularization for the flow field. In this approach, the regularization deviation error in E_S is decomposed into a sparse component and a non-sparse component. The sparse component regularizes the large deviation (e.g., due to missing correspondences) via a L_0 norm penalty; and the non-sparse component regularizes the small universal error via a L_2 norm penalty. We also use the same error-decomposition strategy for E_D. This paper presents two algorithms for the optimization task. One is a non-convex solver, and the other one is a more efficient convex solver. Our experiments on the Middlebury benchmark dataset verifies that our approach can achieve very satisfactory performance, without using many heuristics. In addition, our further study reveals a very interesting connection between this error decomposition approach and the influence function approach in robust statistics. This is an original attempt to bridge the two seemingly different approaches.

The rest of the paper is organized as follows. Section 2 reviews related work. In Section 3, we study the statistics on flow fields and introduce our new regularization model. An efficient optimization method is proposed in Section 4 to solve the problem efficiently. Section 5 reveals the connection between this model and robust function. We validate and evaluate our approach in extensive experiments in Section 6 and conclude in Section 7.

2. Related Work

Estimating motion field is a fundamental problem in computer vision and it has been intensively studied in the past several decades. This section does not intend to give a comprehensive survey of the literature, but only describes the methods that are related to our work in this paper. We categorize these methods based on the data consistency term E_D and the motion regularization term E_S.

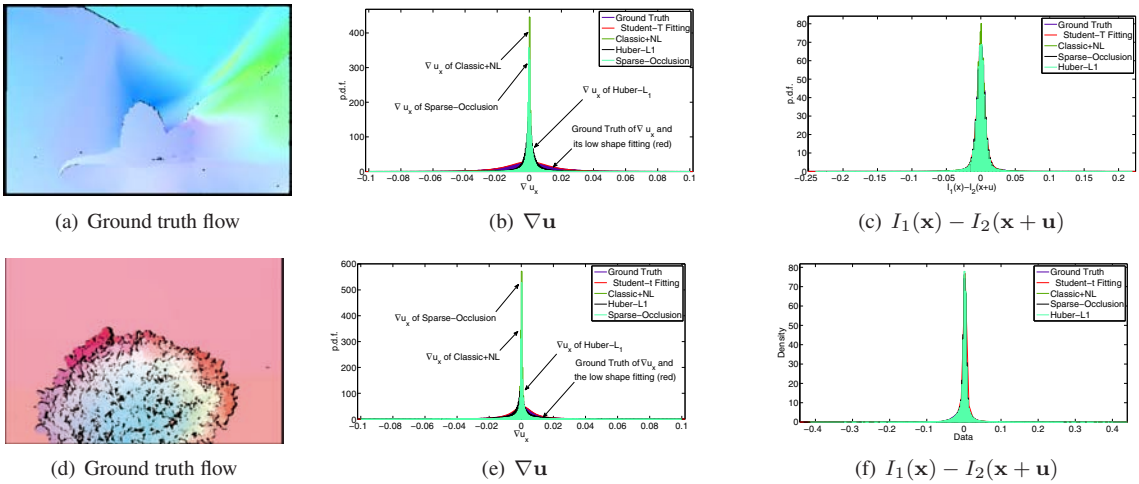
2.1. Data consistency term

Defining the matching between image frames is critical, especially for motion field estimation which needs fine matching at the pixel level. A general treatment is to assume brightness (or color) consistency between two consecutive frames I_1 and I_2 as $I_1(\mathbf{x}) = I_2(\mathbf{x} + \mathbf{u})$, where $\mathbf{x}$ is the pixel location and $\mathbf{u}$ is the motion. In view of this, the data consistency term evaluates the matching error by:

$$E_D(\mathbf{u}) = \sum_{\mathbf{x}} \phi(I_1(\mathbf{x}) - I_2(\mathbf{x} + \mathbf{u})), \qquad (2)$$

where $\phi(\cdot)$ measures the matching error. For example, it can be a commonly used L_2 norm [15, 19] by assuming the matching errors bear a Gaussian distribution. However, as illumination and occlusion invalidate this brightness consistency, there exist outliers that lead to large matching errors and ruin motion estimation.

To handle these outliers, $\phi(\cdot)$ can be chosen as a robust influence function to suppress the large matching errors induced by the outliers [4, 5]. Various forms of influence function have been proposed and compared [3], including the use of L_1 norm $\phi(s) = |s|$ as in [32], the non-convex $L_{0.9}$ norm $\phi(s) = |s|^{0.9}$ as in [16], the regularized TV penalty $\phi(s^2) = \sqrt{s^2 + \epsilon^2}$ (also called "Charbonnier penalty") as in [8, 7, 9], the non-convex Lorentzian function [5, 26, 22], and the Geman-McClure function [17], etc. The introduction of robust influence function generally improves the performance of optical flow methods, although we yet to find a principled way to choose the best form and tune the parameters of the influence function.



(a) Ground truth flow (b) $\nabla \mathbf{u}$ (c) $I_1(\mathbf{x}) - I_2(\mathbf{x}+\mathbf{u})$

(d) Ground truth flow (e) $\nabla \mathbf{u}$ (f) $I_1(\mathbf{x}) - I_2(\mathbf{x}+\mathbf{u})$

Figure 2. A statistical comparison of ground truth data with various high-performing algorithms including Classic+NL [27], Sparse-occlusion[2] and Wedel's improved $TV-L_1$-improved[33] on $\nabla \mathbf{u}$ (b)(e) and $I_1(\mathbf{x}) - I_2(\mathbf{x}+\mathbf{u})$ (c)(f). The red line is a low-shape fitting of ground truth data.

Besides the brightness consistency, other consistencies can also be imposed, for example the image gradient consistency [7, 8]: $\nabla I_1(\mathbf{x}) = \nabla I_2(\mathbf{x}+\mathbf{u})$. Gradient consistency is generally less sensitive to illumination, but it does not constrain smooth regions that do not have significant image gradients. A good practice is to combine the brightness and image gradient consistency [34, 7].

A very recent approach different from the use of robust function was to combine L_2 and L_0 norm [2]. The matching error can be decomposed and measured by two terms: an L_2 norm term for the small but dense error, and an L_0 norm term for the large but sparse error. Identifying the sparse errors may lead to the detection of occlusion [2].

Our proposed approach is inspired by this error separation idea as well as the classical "Mumford-Shah" variational model[20], but has different treatments. First, we also decompose the motion field regularization error, in addition to the matching error as done in [2]. This has the benefit of not over-sparsifying the motion field regularization. Second, our approach uses a dynamic combination of these two error norm terms rather than the fixed combination as done in [2]. This dynamic weighting scheme is based on an annealing strategy, and it is able to alleviate the problem of local minima. As a result, it significantly improves the performance. Third, we give a new solution to the optimization task. Our new convex optimization algorithm converges faster, is less sensitive to parameter settings and it is highly parallel. More importantly, our study gives the first attempt to reveal the very interesting connection between this error separation approach and the robust influence function approach.

2.2. Motion Prior Regularization Term

Motion field estimation is an inverse problem. The data consistency term alone cannot well constrain the solution, because it treats the motion at individual pixels independently. It is critical to regularize the solution by imposing the constraints for the motion field (or the motion prior). A commonly used regularization term is to assume smooth total variation of the motion field:

$$\int \phi(\sqrt{||\nabla \mathbf{u}_x(\mathbf{x})||^2 + ||\nabla \mathbf{u}_y(\mathbf{x})||^2}) d\mathbf{x} \qquad (3)$$

Original Horn-Schunck model [15] takes ϕ as the L_2 norm. Non-convex robust cost function can also be introduced such as the Charbonnier penalty in [9]. In fact, the motion field may not be smooth everywhere, but observes discontinuity at the motion boundary. Considering the image content, edge preserving filters can be used to penalize motion-discontinuity more on the uniform regions and less on the non-smooth regions [34, 2, 14].

The prior of the motion field can actually be modeled by Markov random fields. To include high-order interactions in the field, the field of expert (FOE) model has been proposed [23], where each expert is a student-t distribution. In addition, the Gaussian scale mixture model [31] was also investigated in motion analysis [26]. The endeavor along this line leads to more complicated generative model for the motion prior [18].

Different from the above approaches, a recent study showed that the motion field prior may have a sparse representation, even if the motion field includes motion discontinuities [25]. Using a learned sparsity of the motion field for regularization has led to quite encouraging results [16]. Our proposed method also uses sparsity regularization, but it is

different from these existing methods. Statistics show that these existing methods tend to impose an over-sparsified regularization. And our approach decomposes the motion field for both sparse and non-sparse regularization.

3. Spatial Prior of Optical Flow

3.1. Over-sparsify or Over-smooth?

As it is generally accepted, a quadratic spatial regularization term $||\nabla \mathbf{u}_x||^2 + ||\nabla \mathbf{u}_y||^2$ [19][15] produces over-smoothed flow fields on motion boundaries. Various robust penalty functions are proposed [4] [5] to tackle this problem by assuming piecewise smoothness. According to the theoretical development of compressive sensing [12] [10], a L^q-norm with $0 < q \leq 1$ enables us to pursue a sparse model for better results.

However, the flow field is not as sparse as we expect. In a recent work of sparse coding flow fields [25], it showed that the motion field is nearly sparse but with non-sparse small dense errors. The reason, in our opinion, is two-fold: (1) the optical flow equation $I_x\mathbf{u}_x + I_y\mathbf{u}_y + I_t = 0$ is an approximation under the small motion assumption; (2) natural image sequences contain deformation, non-translational motion, illumination changes and other small noises.

In Fig. 2, we show a statistical comparison of motion variation and color deviation of the ground truth and of those generated from various high-performing algorithms, including $TV - L^1$-improved [33], Sparse Occlusion [2], Classic+NL [27]. We exclude both occluded and discontinuous motion pixels with large motion variations as well as large color deviation. It is obvious that the deviation error of flow fields is over-sparsified at $\nabla \mathbf{u} \to 0$, although the matching error is as sparse as we expect.

3.2. Sparse/Non-Sparse Error Separation

Different from many existing methods trying to find one universal fitting model for the motion field, we assume that $\nabla \mathbf{u}_x$ contains two components:

$$\nabla \mathbf{u}_x(\mathbf{x}, t) \triangleq \begin{cases} n_{\mathbf{u}_x}(\mathbf{x}, t) & \mathbf{x} \in D \setminus \Omega \\ e_{\mathbf{u}_x}(\mathbf{x}, t) & \mathbf{x} \in \Omega \end{cases} \quad (4)$$

where Ω denotes the motion discontinuity regions. The same notations of $n_{\mathbf{u}_x}$ and $e_{\mathbf{u}_x}$ are defined for $\mathbf{u}_y$. Therefore, we use $n_{\mathbf{u}_*}$ to refer to $n_{\mathbf{u}_x}$ and $n_{\mathbf{u}_y}$, and $e_{\mathbf{u}_*}$ to $e_{\mathbf{u}_x}$ and $e_{\mathbf{u}_y}$. In this error separation model, $e_{\mathbf{u}_*}$ describes the variation at motion discontinuity regions, and it tends to be sparse. $n_{\mathbf{u}_*}$ covers the rest regions, showing universal small variations.

In Sparse-Occlusion [2], a Gaussian assumption of small bias is made for color deviation from Lambertian reflection. But in the case of $\nabla \mathbf{u}_x$ and $\nabla \mathbf{u}_y$, it is more complicated: if only translational motion exists, a Gaussian model is a good approximation; but in real motion, affine motion and deformation will complicate the case. Our statistical results reveal that the universal motion variation $n_{\mathbf{u}_*}$ is a Gaussian-like distribution with a heavier tail. Therefore, for $n_{\mathbf{u}_*}$, we assume a generalized Student's t-distribution:

$$n_{\mathbf{u}_*} \sim \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}(\frac{1}{\pi\nu\sigma})^{\frac{1}{2}}[1 + \frac{n_{\mathbf{u}_*}^2}{\nu\sigma}]^{-\frac{\nu+1}{2}}$$

or a scale-mixture-of-Gaussian:

$$n_{\mathbf{u}_*} \sim \sum_i \pi_i \frac{1}{\sigma_i} e^{-\frac{n_{\mathbf{u}_*}^2}{2\sigma_i^2}}$$

In the case of small deviation $n_{\mathbf{u}_*} \to 0$, this can be safely approximated by a Gaussian distribution.

4. Joint Estimation of Motion and Errors

4.1. Our Motion Regularization Term

Under the assumption discussed above, we seek to estimate a flow field by minimizing the number of nonzero elements of $e_{\mathbf{u}_*}$ as well as the negative log-likelihood of $n_{\mathbf{u}_*} = \nabla \mathbf{u}_* - e_{\mathbf{u}_*}$:

$$E_S(\mathbf{u}, e_{\mathbf{u}_*}) \triangleq \sum_{\mathbf{x}} \lambda_1 \log(1 + \frac{||\nabla \mathbf{u}_* - e_{\mathbf{u}_*}||^2}{\sigma}) + \lambda_2 ||e_{\mathbf{u}_*}||_{L^0}$$

$$\approx \sum_{\mathbf{x}} \frac{\lambda_1}{\sigma} ||\nabla \mathbf{u}_* - e_{\mathbf{u}_*}||^2 + \lambda_2 ||e_{\mathbf{u}_*}||_{L^0}$$

where $* \in \{x, y\}$ denotes both x and y directions, and $e_{\mathbf{u}_*}$ is the unknown sparse component of $\nabla \mathbf{u}_*$.

4.2. Color consistency

For color consistency, we also adopt the separated penalty function of data term proposed in [2]. In their work, the source of color matching error is assumed to come from two components: a sparse occlusion $e_D(\mathbf{x})$ and a non-sparse Lambertian reflection $n_D(\mathbf{x})$ as:

$$I_1(\mathbf{x}) = \begin{cases} I_2(\mathbf{x} + \mathbf{u}) + n_D(\mathbf{x}), \ \mathbf{x} \in D \setminus \hat{\Omega} \\ e_D(\mathbf{x}), \ \mathbf{x} \in \hat{\Omega} \end{cases} \quad (5)$$

where $\hat{\Omega}$ denotes the occlusion regions. Then we have:

$$E_D(\mathbf{u}, e_D) = \sum_{\mathbf{x}} ||I_1(\mathbf{x}) - I_2(\mathbf{x}+\mathbf{u}) - e_D||_{L^2}^2 + \mu ||e_D||_{L^0} \quad (6)$$

4.3. Optimization Methods

For a small motion, we make the first-order approximation for linearization $I_2(\mathbf{x}) = I_1(\mathbf{x}) + \nabla I_1(\mathbf{x})\mathbf{u}(\mathbf{x})$ and infer the flow fields $\mathbf{u}$ as well as the sparse e_D, $e_\mathbf{u}$ simultaneously as:

$$\{\hat{\mathbf{u}}, \hat{e_D}, \hat{e_\mathbf{u}}\} = \arg \min_{\{\mathbf{u}, e_D, e_\mathbf{u}\}} E_D(\mathbf{u}, e_D) + E_S(\mathbf{u}, e_\mathbf{u}) \quad (7)$$

Optical Flow Optimization
repeat
Initialize rL^1 coefficients in eq(7) as 1;
repeat
compute $d\mathbf{u}, e_D, e_\mathbf{u}$ to optimize eq(7)
reweighting the L^1 norm coefficients;
end
$\mu \to \mu/2, \lambda_2 \to \lambda_2/2$;
until $\mu < \mu_{\min}$ and $\lambda_2 < \lambda_{2,\min}$

Table 1. Algorithm for flow estimation.

In practice, the optimization of L_0-norm is NP-hard. We relax it to a consequential reweighted L_1 method [11] to iteratively solve the non-convex programming problem and provide two optimization algorithms here to solve Equation (7) as shown in Table 1.

4.3.1 A Non-convex Solver

Similar to the method in [2], we use Nesterov's gradient-based energy function minimization scheme [21] to solve the problem. At each iteration we use a weighted combination of current descent direction $\nabla E^k(\mathbf{u}, e_D, e_\mathbf{u})$ with previous iteration $\sum_{i=1}^{k-1} \nabla E^i(\mathbf{u}, e_D, e_\mathbf{u})$, appropriately scaled by a Lipschitz constant. We refer readers to [2, 21] for a detailed description.

4.3.2 A Convex Solver

We propose a more efficient solution here. The quadratic approximation of equation(7) enables us to apply a fast convex optimization strategy:

$$\{\hat{d\mathbf{u}}, \hat{e}_D^k, \hat{e}_S^k\} = \arg\min \sum_{\mathbf{x}} ||I_x^k d\mathbf{u}_x + I_y^k d\mathbf{u}_y + I_t^k - e_D^k||^2 + \mu||e_D^k||_{rL^1} + \frac{\lambda_1}{\sigma}||\nabla\mathbf{u} - e_\mathbf{u}^k||^2 + \lambda_2||e_\mathbf{u}^k||_{rL^1}$$

where $||.||_{rL^1}$ denotes a reweighted L^1 norm and $\mathbf{u} = \mathbf{u}^0 + d\mathbf{u}$. We employ a coordinate-descent algorithm to solve the problem:
(1) Fix $\mathbf{u}$, we optimize e_D and $e_{\mathbf{u}_*}$:

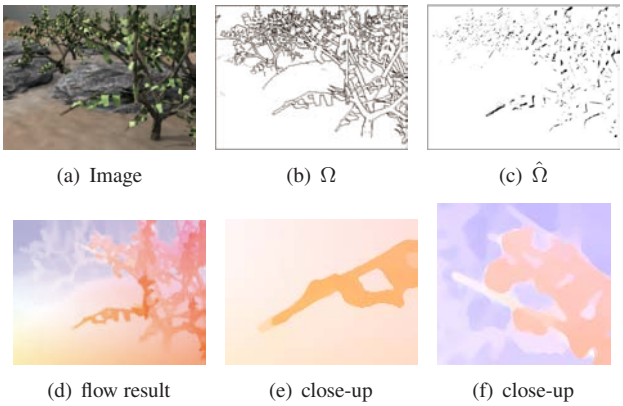
$$\hat{e}_D = \arg\min_{e_D} ||I_x^k d\mathbf{u}_x + I_y^k d\mathbf{u}_y + I_t^k - e_D^k||^2 + \mu||e_D||_{rL^1}$$

$$\hat{e}_{\mathbf{u}_*} = \arg\min_{e_{\mathbf{u}*}} \frac{\lambda_1}{\sigma}||\nabla\mathbf{u}_* - e_{\mathbf{u}_*}||^2 + \lambda_2||e_{\mathbf{u}_*}||_{rL^1}$$

This can be solved by the shrinkage formula [13] very efficiently and highly parallel in nature.
(2) Fix e_D and $e_{\mathbf{u}_*}$ and solve $\mathbf{u}$:

$$d\mathbf{u} = \arg\min_{d\mathbf{u}} ||I_x^k d\mathbf{u}_x + I_y^k d\mathbf{u}_y + I_t^k - e_D^k||^2 + \frac{\lambda_1}{\sigma}||\nabla\mathbf{u}_* - e_{\mathbf{u}_*}||^2$$

which is a quadratic programming problem and has a closed form solution.



(a) Image (b) Ω (c) $\hat{\Omega}$ (d) flow result (e) close-up (f) close-up

Figure 3. Optical flow estimation. (a) Original image; (b) inferred support of $e_\mathbf{u}$; (c) inferred support of e_D; (d) final result; (e)(f) close up show of our results.

4.4. Model Refinement

4.4.1 Edge-Preserving Spatial Regularization

Edge-preserving regularization [32] [34] can also be naturally integrated into our model to further improve the performance. Let $w(\mathbf{x}) \triangleq \exp(-||\nabla I_1(\mathbf{x})||^\kappa)$:

$$E_S(\mathbf{u}) = \sum_{\mathbf{x}} w(\mathbf{x})(\frac{\lambda_1}{\sigma}||(\nabla\mathbf{u}_* - e_{\mathbf{u}_*})||^2 + \lambda_2||e_{\mathbf{u}_*}||_{L^0})$$

4.4.2 Boosted-joint Sparsity Pursuit Refinement

In our experiments, we find the spatial distribution of motion discontinuity Ω and occlusion region $\hat{\Omega}$ are strongly correlated. The supports of e_D and $e_\mathbf{u}$ can be strong clues to each other. Inspired by the recent work on joint-sparse recovery in statistics society [29], we use $\Omega \cup \hat{\Omega}$ as the support candidates of $e_\mathbf{u}$ and e_D to further boost the performance.

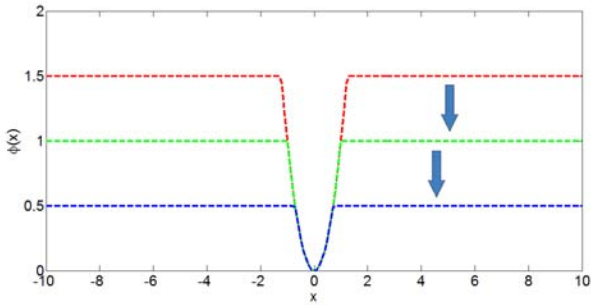
In Figure(3), we show the optimization process. (b) and (c) indicate the support Ω and $\hat{\Omega}$. (e) and (f) show close-ups in final flow result (d).

5. Robust Sparse/Non-sparse Error Separation

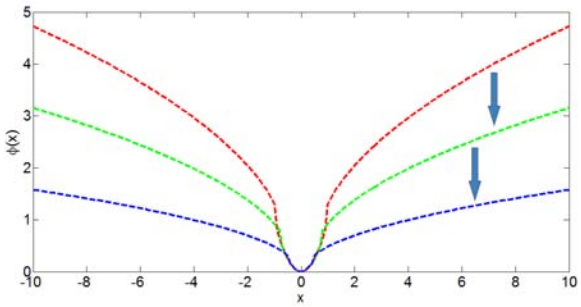
Using fixed separation coefficient λ_1, λ_2 and μ does not always produce good results. However, in our experiments, we find that a dynamic weighting scheme, i.e., gradually decreasing the coefficients μ and λ_2, yields a leap in performance. Intuitively, it is an annealing method to avoid local minima. For a further exploitation, we investigate the error-separated cost function:

$$\phi(x) \triangleq \inf_e(\lambda_1||x - e||^2 + \lambda_2||e||_{L^0}) \quad (8)$$

by using different λ_1, λ_2 as shown in Fig(4(a)). We also tried different mathematical forms of error-separation and give another example in Fig(4(b)).



(a) $\phi(x) \triangleq \inf_e(\lambda_1||x-e||^2 + \lambda_2||e||_{L^0})$



(b) $\phi(x) \triangleq \inf_e(\lambda_1||x-e||^2 + \lambda_2||e||_{L^{0.8}})$

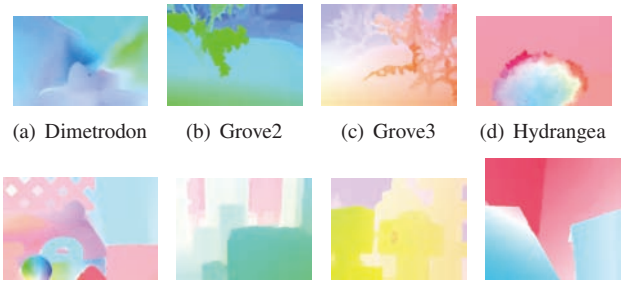
Figure 4. The shape of the error-separation cost function appears very similar to the influence function in robust statistics.

As we can see, the shape of the functional penalty $\phi(x)$ looks very similar to the robust cost function dated back to 1990s. If we date back even earlier to the classical variational "Mumford-Shah" model [20, 6], proposed in 1989 to treat inner-region R_i and boundary Γ differently as:

$$E(f,\Gamma) = \mu^2 \iint_R (f-g)^2 dxdy + \iint_{R-\Gamma} ||\nabla f||^2 dxdy + \nu|\Gamma| \tag{9}$$

while our regularized term is actually a variation with a discrete space. Interestingly, the robust functions mentioned above have always been applied to solve the "Mumford-Shah" model.

In [4] different forms of robust function $\phi(x)$ were studied for matching error and spatial regularization including truncated quadratic function, Geman-McClure and Lorentzian. A generalized regularized TV penalty is used in Classic+NL [27]. All these functions have $\phi(x) = O(x^2)$ at $x \to 0$ and a sub-linear increase when x becomes larger. Let's take a truncated quadratic function $\phi(x) = \min\{x^2, \sigma\}$ as an example: the influence of parameter σ controls the shape around $x \approx 0$, which plays a similar role as selecting different λ_1 and λ_2 in our approach. As shown in Figure4(a), our dynamic weighting gradually changes the "shape" of the related robust function: at the first few initial steps, we use large λ_2 and μ for a warm start; then as the flow estimation is becoming more accurate, we decrease these weights to reflect better precision and introduce sparsity gradually into the estimation.



(a) Dimetrodon (b) Grove2 (c) Grove3 (d) Hydrangea

(e) RubberWhale (f) Urban2 (g) Urban3 (h) Venus

Figure 6. Results of our algorithm on Middlebury training set [3].

This annealing scheme has also been practiced in some state-of-the-art methods and achieved good performance. As done in Learning-Flow [26], a series of mixed cost function $E_C(\mathbf{u}) = \alpha E_Q(\mathbf{u}) + (1-\alpha)E(\mathbf{u})$ is employed to vary the convexity, where $E_Q(\mathbf{u})$ and $E(\mathbf{u})$ stand for quadratic and non-convex penalty functions. Classic+NL [27] well studied the secrets of optical flow and used a quadratic cost function in optimization for initialization and gradually changed the penalty form to non-concave. As shown in this section, the connection between error-separation and robust function provides more insights to this practice.

5.1. A Natural Selective Mechanism

Many studies [25, 27, 28, 22] suggest that: simply combining different priors will only lead to an average performance, while selection and fusion [34, 1, 17] bring improvement. The early attempt of using robust function is itself a natural selective process: different components of $\nabla\mathbf{u}$ decouples the middle-level segmentation and low-level flow estimation. As in Figure5, the spatial support of $e_{\mathbf{u}}$ indicates a motion segmentation boundary of the scene.

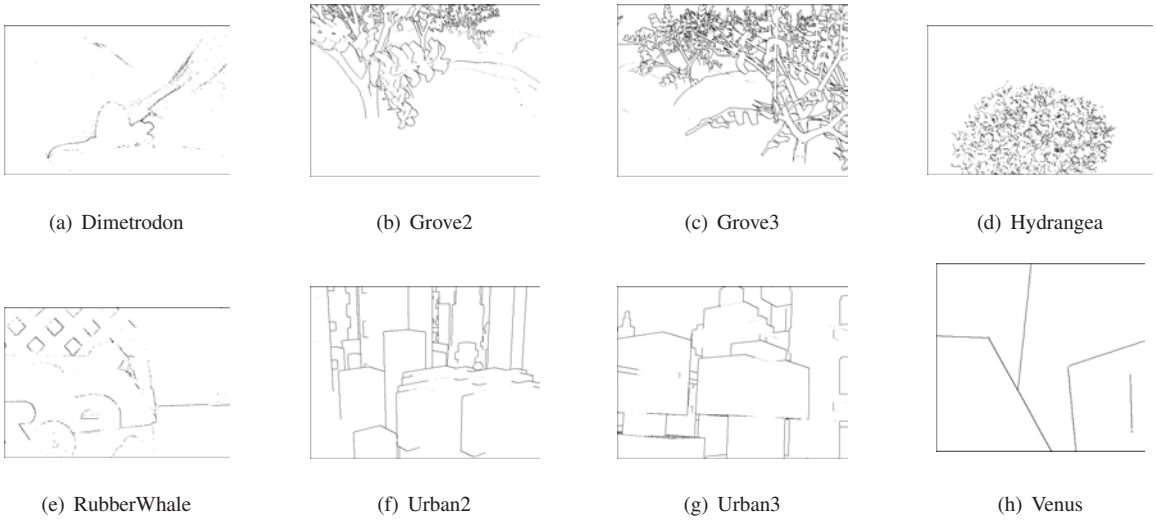
6. Experimental Results

To evaluate our approach, we use the Middlebury benchmark dataset. To allow for illumination changes between frames, we pre-process images with the structure-texture decomposition [24]. We implement our approach in a coarse-to-fine/warping strategy to tackle large motion.

6.1. Validation

Our first experiment is to validate our sparse/non-sparse error separation approach. In Table 2, we compare our method with l_1-TV-improved[33] and sparse-occlusion[2]. The improvement is obvious and significant.

In Figure. 6, we show the flow estimation results on the Middlebury benchmark sequences.



(a) Dimetrodon (b) Grove2 (c) Grove3 (d) Hydrangea

(e) RubberWhale (f) Urban2 (g) Urban3 (h) Venus

Figure 5. The spatial support of $e_{\mathbf{u}}$ on Middlebury benchmark.

AEPE	Venus	RubberWhale	Hydrangea	Grove2	Grove3	Urban2	Urban3	Dimetrodon
(ours)	**0.23**	**0.07**	**0.14**	**0.10**	**0.48**	**0.21**	**0.37**	**0.12**
[2]	0.30	0.18	0.19	0.16	0.59	0.39	0.84	0.15

Table 2. Performance comparison on Middlebury train set.

6.2. Overall Performance

Our algorithms are highly competitive with the state-of-the-art as shown in Figure7. The top-ranked MDP-flow [34] uses extended flow initialization to preserve small-scale motion structures; Layers++ [28] uses a graphical model to address depth order; TC [30] makes use of multiple frame temporal coherency; LSM [16] learns a sparse regional model to code spatial flow field. We have not addressed these issues in our work. We mainly aim to show the effectiveness of sparse/non-sparse error separation.

7. Conclusion and Future Work

In this work, we introduce a novel generalized sparse/non-sparse error separation approach to optical flow fields estimation. By exploiting the statistics on optical flow fields data set, we find that the gradients of flow fields come from two sources: a sparse large motion-discontinuity component and a small dense Gaussian component. Based on that, we propose a novel regularization function to explicitly model the decomposition. We propose two efficient algorithms to solve the optimization problem. Then, we exploit how the separation coefficients influence the performance, which reveals the in-depth connection between the noise separation approach and the robust penalty function.

Our experiments on the Middlebury benchmark dataset clearly show that our approach can achieve very satisfactory performance, without using many heuristics.

8. Acknowledgement

This work was supported in part by National Science Foundation grant IIS-0347877, IIS-0916607, US Army Research Laboratory and the US Army Research Office under grant ARO W911NF-08-1-0504, and DARPA Award FA 8650-11-1-7149.

References

[1] O. M. Aodha, G. J. Brostow, and M. Pollefeys. Segmenting video into classes of algorithm-suitability. *Computer Vision and Pattern Recognition (CVPR)*, 2010.

[2] A. Ayvaci, M. Raptis, and S. Soatto. Occlusion detection and motion estimation with convex optimization. In *Advances in Neural Information Processing Systems*. December 2010.

[3] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal on Computer Vision*, 2011.

[4] M. Black and P. Anandan. Robust dynamic motion estimation over time. *IEEE Conf. on Computer Vision and Pattern Recog.*, 1991.

[5] M. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, 1996.

[6] A. Blake and A. Zisserman. *Cambridge, MA: MIT Press*. John Wiley and Sons Inc, 1987.

[7] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. *European Conf. on Computer Vision*, 2004.

Average angle error	avg. rank	Army (Hidden texture) GT	Army im0	Army im1	Mequon (Hidden texture) GT	Mequon im0	Mequon im1	Schefflera (Hidden texture) GT	Schefflera im0	Schefflera im1	Wooden (Hidden texture) GT	Wooden im0	Wooden im1	Grove (Synthetic) GT	Grove im0	Grove im1	Urban (Synthetic) GT	Urban im0	Urban im1	Yosemite (Synthetic) GT	Yosemite im0	Yosemite im1	Teddy (Stereo) GT	Teddy im0	Teddy im1
		all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext
Layers++ [38]	6.9	3.11 2	8.22 2	2.79 7	2.43 5	**7.02** 1	2.24 11	**2.43** 1	**5.77** 1	2.18 12	**2.13** 1	**9.71** 1	**1.15** 1	**2.35** 1	**3.02** 1	**1.96** 1	3.81 13	11.4 10	3.22 16	2.74 22	4.01 26	2.35 20	1.45 2	**3.05** 1	1.79 7
MDP-Flow2 [40]	7.0	3.32 8	8.76 6	2.85 9	**2.18** 1	7.47 3	1.85 4	2.77 3	6.95 3	2.06 9	3.25 15	17.3 19	1.59 13	2.87 7	3.73 5	2.32 5	3.15 3	11.1 9	2.65 3	2.04 7	3.64 12	1.60 6	1.88 7	4.49 8	1.49 2
[60]	7.8	3.14 4	8.75 5	2.76 5	3.02 17	10.6 14	2.43 16	3.45 9	8.96 9	2.36 15	2.66 3	13.7 4	1.42 2	2.85 6	3.75 7	2.33 6	3.28 5	9.40 3	2.73 6	2.42 15	3.31 4	2.69 24	1.47 3	3.07 2	1.66 4
LSM [41]	8.5	3.12 3	8.62 3	2.75 4	3.00 16	10.5 13	2.44 17	3.43 7	8.85 8	2.35 14	2.66 3	13.6 3	1.44 3	2.82 4	3.68 3	2.36 8	3.38 7	9.41 4	2.81 10	2.69 21	3.52 10	2.84 28	1.59 4	3.38 4	1.80 8
TC-Flow [48]	9.1	**2.91** 1	**8.00** 1	2.34 2	**2.18** 1	8.77 7	**1.52** 1	3.84 14	10.7 15	**1.49** 1	3.13 11	16.6 15	1.46 5	2.78 3	3.73 5	**1.96** 1	3.08 2	11.4 10	2.66 4	1.94 6	3.43 6	3.20 41	3.06 16	7.04 16	4.08 34
Classic+NL [31]	10.1	3.20 6	8.72 4	2.81 8	3.02 17	10.6 14	2.44 17	3.46 10	8.84 7	2.38 17	2.78 5	14.3 5	1.46 5	2.83 5	3.68 3	2.31 4	3.40 8	**9.09** 1	2.76 9	2.87 27	3.82 16	2.86 32	1.67 6	3.53 5	2.26 12

(a) Average Angle Error

Average endpoint error	avg. rank	Army (Hidden texture) GT	Army im0	Army im1	Mequon (Hidden texture) GT	Mequon im0	Mequon im1	Schefflera (Hidden texture) GT	Schefflera im0	Schefflera im1	Wooden (Hidden texture) GT	Wooden im0	Wooden im1	Grove (Synthetic) GT	Grove im0	Grove im1	Urban (Synthetic) GT	Urban im0	Urban im1	Yosemite (Synthetic) GT	Yosemite im0	Yosemite im1	Teddy (Stereo) GT	Teddy im0	Teddy im1
		all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext
MDP-Flow2 [40]	5.5	0.09 7	0.23 3	0.07 3	0.16 2	**0.52** 1	0.13 3	0.22 2	0.46 3	0.17 4	0.17 12	0.93 20	0.09 5	0.65 4	0.98 5	0.43 4	**0.29** 1	**0.91** 1	0.26 3	0.11 7	0.13 10	0.17 8	0.51 8	1.11 10	0.72 7
Layers++ [38]	6.2	0.08 2	**0.21** 1	0.07 3	0.19 5	0.56 3	0.17 13	**0.20** 1	**0.40** 1	0.18 8	**0.13** 1	**0.58** 1	**0.07** 1	**0.48** 1	**0.70** 1	0.33 2	0.47 12	1.01 3	0.33 12	0.15 26	0.14 19	0.24 24	**0.46** 1	**0.88** 1	0.72 7
[60]	8.7	0.08 2	0.23 3	0.07 3	0.22 13	0.73 15	0.18 15	0.28 8	0.64 8	0.19 13	0.14 2	0.71 4	0.08 2	0.67 7	0.99 7	0.48 7	0.49 15	1.06 5	0.32 9	0.14 21	**0.11** 1	0.28 36	0.49 3	0.98 2	0.73 10
LSM [41]	9.4	0.08 2	0.23 3	0.07 3	0.22 13	0.73 15	0.18 15	0.28 8	0.64 8	0.19 13	0.14 2	0.70 2	0.09 5	0.66 5	0.97 4	0.48 7	0.50 16	1.06 5	0.33 12	0.15 26	0.12 5	0.29 39	0.50 5	0.99 4	0.73 10
TC-Flow [48]	9.5	**0.07** 1	**0.21** 1	**0.06** 1	**0.15** 1	0.59 4	**0.11** 1	0.31 14	0.78 15	**0.14** 1	0.16 7	0.86 13	0.08 2	0.75 12	1.11 13	0.54 11	0.42 6	1.40 17	0.25 2	0.11 7	0.12 5	0.29 39	0.62 14	1.35 15	0.93 26
Classic+NL [31]	10.4	0.08 2	0.23 3	0.07 3	0.22 13	0.74 17	0.18 15	0.29 11	0.65 11	0.19 13	0.15 4	0.73 5	0.09 5	0.64 3	0.93 3	0.47 5	0.52 18	1.12 8	0.33 12	0.16 32	0.13 10	0.29 39	0.49 3	0.98 2	0.74 13

(b) Average Endpoint Error

Figure 7. Experimental results of our algorithm on Middlebury test set [3]. Our method is rank the 3rd for both Average Angle Error (AAE) and Average End-point Error (AEPE).

[8] T. Brox and J. Malik. Large displacement optical flow: Descriptor matching in variational motion estimation. *submitted to Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2010.

[9] A. Bruhn and J. Weickert. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 2005.

[10] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, Feb. 2006.

[11] E. J. Candès. Enhancing sparsity by reweighted l_1 minimization. *Journal of Fourier Analysis and Applications*, 14(5-6):877–905, Dec. 2008.

[12] D. Donoho. Compressed sensing. *IEEE Trans. on Information Theory*, 52:1289–1306, 2006.

[13] E. Hale, W. Yin, and Y. Zhang. Fixed-point continuation for l1-minimization: Methodology and convergence. *SIAM: Journal on Optimization*, 19(3):1107–1130, 2008.

[14] A. B. Henning Zimmer and J. Weickert. Optic flow in harmony. *International Journal of Computer Vision (IJCV)*, 93(3):368–388, 2011.

[15] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 16:185–203, 1981.

[16] K. Jia, X. Wang, and X. Tang. Optical flow estimation using learned sparse model. *International Conf. on Computer Vision*, 2011.

[17] V. Lempitsky, S. Roth, and C. Rother. Fusionflow: Discrete-continuous optimization for optical flow estimation. *Computer Vision and Pattern Recognition (CVPR)*, 2008.

[18] C. Liu, W. T. Freeman, and E. H. Adelson. Analysis of contour motions. *Advances in Neural Information Processing Systems*, 2006.

[19] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Intl Joint Conf on Artificial Intelligence*, pages 674–679, 1981.

[20] D. Mumford and J. Shah. Optimal approximation of piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.*, 42:577–685, 1989.

[21] Y. Nesterov. Smooth minimization of nono-smooth functions. *Mathematical Programming*, 103(1):127–152, 2005.

[22] S. Roth and M. Black. On the spatial statistics of optical flow. *International Conf. on Computer Vision*, 2005.

[23] S. Roth and M. Black. Fields of experts: A framework for learning image priors. *IEEE Conf. on Computer Vision and Pattern Recog.*, 2007.

[24] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60:259–268, 1992.

[25] X. Shen and Y. Wu. Sparsity model for robust flow estimation at motion discontinuities. *IEEE Conf. on Computer Vision and Pattern Recog.*, 2010.

[26] D. Sun, J. P. Lewis, S. Roth, and M. Black. Learning optical flow. *European Conf. on Computer Vision*, 2008.

[27] D. Sun, S. Roth, and M. Black. Secrets of optical flow estimation and their principles. *IEEE Conf. on Computer Vision and Pattern Recog.*, 2010.

[28] D. Sun, E. Sudderth, and M. Black. Layered image motion with explicit occlusions, temporal consistency and depth ordering. *Advances in Neural Information Processing Systems*, 2010.

[29] E. van den Berg and M. P. Friedlander. Joint-sparse recovery from multiple measurements. 2009.

[30] S. Volz, A. Bruhn, L. Valgaerts, and H. Zimmer. Modeling temporal coherence for optical flows. *IEEE Conf. on Computer Vision*, 2011.

[31] M. Wainwright and E. Simoncelli. Scale mixtures of gaussians and the statistics of natural images. *Advances in Neural Information Processing Systems*, 1999.

[32] A. Wedel, D. Cremers, T. Pock, and H. Bischof. Structure- and motion-adaptive regularization for high accuracy optic flow. *International Conf. on Computer Vision*, 2009.

[33] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers. An improved algorithm for tv-l1 optical flow. *Dagstuhl Motion Workshop*, 2008.

[34] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. *IEEE Conf. on Computer Vision and Pattern Recog.*, 2010.