Sparsity Model for Robust Optical Flow Estimation at Motion Discontinuities

Xiaohui Shen and Ying Wu
Northwestern University
2145 Sheridan Road, Evanston, IL 60208
{xsh835,yingwu}@eecs.northwestern.edu

Abstract

This paper introduces a new sparsity prior to the estimation of dense flow fields. Based on this new prior, a complex flow field with motion discontinuities can be accurately estimated by finding the sparsest representation of the flow field in certain domains. In addition, a stronger additional sparsity constraint on the flow gradients is incorporated into the model to cope with the measurement noises. Robust estimation techniques are also employed to identify the outliers and to refine the results. This new sparsity model can accurately and reliably estimate the entire dense flow field from a small portion of measurements when other measurements are corrupted by noise. Experiments show that our method significantly outperforms traditional methods that are based on global or piecewise smoothness priors.

1. Introduction

Computing optical flow is a fundamental problem in computer vision. Most contemporary methods more or less originate from or are related to the two classical methods, i.e., Lucas & Kanade's method[18] based on the assumption that the set of pixels in a local region share the same motion, and Horn & Schunck's method[13] based on global smoothness regularization through variational methods.

These methods generally perform well when their assumptions hold. Unfortunately, in practice, these assumptions are largely violated at motion boundaries. When there are multiple motions in an image patch, motion boundaries and motion discontinuities exist, and thus the flow field is neither smooth nor reflecting a single motion. Because the motion boundaries are unknown, computing the optical flow over motion boundaries remains a very difficult problem.

To overcome this difficulty, many approaches have been investigated to modify these priors to fit the flow field better. Based on robust statistics, robust functions can be introduced to the least squares estimation[4], so that the dominant motion can be estimated and other motions treated as outliers can be identified. Based on a piecewise smoothness

model, the flow field and the separation of multiple motions can be done jointly by motion layer models[10, 24]. In addition, the motion discontinuity can be modeled in an inhomogeneous MRF so that the flow field can be inferred while preserving the discontinuity[12]. Motion segmentation is also adopted to differentiate multiple motions[20, 27]. Another interesting idea is to learn the statistics of the flow field and use it as the prior for inference[22, 23]. Despite these efforts, finding a simple and effective solution to handle motion discontinuities is still very attractive.
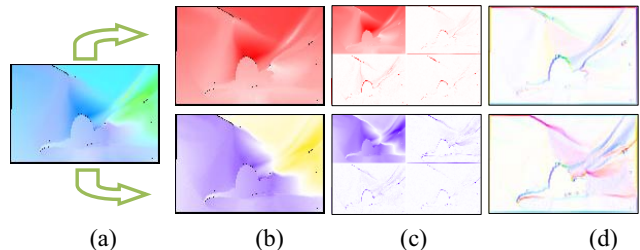

Figure 1. Sparsity model for optical flow. (a) The flow field, (b) the component of flow **u** and **v**, (c) 2-D Haar Wavelet decomposition, (d) the flow gradient field is also sparse. The values are mapped to color for display.

Different from any previous methods originated from Horn & Schunck or Lucas & Kanade, this paper introduces a new prior to the flow field, and presents a simple and elegant method to handle the complexity incurred by multiple motions and motion discontinuities. Instead of explicitly modeling the discontinuities and learning all sorts of priors, we only impose a simple assumption, i.e., the flow field has a sparse representation in other domains. This assumption is looser than the single motion assumption and the smooth flow field assumption, because it covers the case where the flow field exhibits multiple motions. This sparsity assumption is generally held in reality, because the flow field is much more structured than random Brownian motion and much simpler than natural images. See Fig.1(c) for example, most coefficients are zero after performing 2-D Wavelet decomposition to the flow field. Therefore the flow can be sparsely represented in the Wavelet domain. Based on this

assumption, we formulate flow computation in an underdetermined linear system, and its minimum L_1 norm solution leads to a very accurate estimate of the flow field. The proposed new approach is very simple and accurate.

The novelty of this work lies in three aspects: a) we validate that the flow field is sparse in other domain such as Wavelet and DCT, and such sparsity can well capture multiple motions and motion discontinuities in the flow field. Moreover, the sparsest fitting of the model can be obtained by the L_1-norm minimization, and yields very accurate flow estimation. b) A stronger additional gradient sparsity constraint is incorporated into the sparsity model, as shown in Fig.1(d), which distinguishes motion boundaries from measurement noises. c) Robust estimation techniques are performed to remove the outliers and reliably estimate the flow from a small portion of measurements.

2. Related Work

Ever since Horn & Schunck's and Lucas & Kanade's solutions, many approaches have been proposed to improve these methods to estimate high quality flow fields. Early work includes robust statistics[4] and MRF models for preserving motion discontinuities[12]. In [4], robust functions were introduced to the least squares estimation to identify the outliers, while in [12] multiple gradient-based and edge-based constraints were added and fused in an MRF to preserve motion discontinuities. All these methods are based on energy function optimization. Following this trend, more accurate energy functions (MRF models) as well as new optimization approaches were introduced recently[15, 16, 25]. [25] used rotation invariant convex regularizers to find the connection between the intensity constancy and smoothness regularization. Such a connection was further explored by [29], which introduced the complementarity of these two terms. [16] proposed a more accurate MRF energy as well as a discrete and continuous combined optimization method. [15] also proposed a new energy function based on discrete optimization.

Besides the MRF models, another natural way to depict piecewise smoothness and motion discontinuities is to segment the motion. In [20] a motion segmentation model was first generated by energy function optimization, and then used to introduce another mixture of local smoothness and region-wise parameterizations. In [27], the images were partitioned to segments in which the flow was constrained by an affine motion. The final results were obtained by minimizing the error in segmentation. More recently, along with the establishment of Middlebury Dataset[3], more flow data with ground truth were available and accordingly were used to learn the statistics of flow field[22, 23]. In [17], 3D structure sensors was used to construct a parametric model for optical flow. Despite such great efforts, most of them are still based on global or piecewise smoothness constraints.

On the other hand, the initial application of sparsity to the image domain can be traced back to Olshausen's image sparse coding[21]. The direct way to seek sparsity is L_0-norm minimization, which is a NP-hard problem. Therefore no global optimum is guaranteed, which limited its application. Since the theory of compressive sensing[11, 6] was proposed, which proved the equivalence of L_0-norm minimization and L_1-norm minimization for sparsity under certain conditions, such limitation has been relieved. Immediately after the emergence of compressive sensing, it has been applied to many vision related areas. In face recognition[26], a face image is considered as a linear combination of the training images of the same person with some measurement noises, and thus can be sparsely represented by the image training dataset. [26] also proposed a method to handle the noise problem when the noises are sparse. Similarly, in the application of target tracking[19], each object can also be sparsely represented by templates with sparse noises caused by lighting change or occlusion. [28] designed an overcomplete codebook by randomly choosing small image patches, and used it to sparsely represent other image patches for image super-resolution. In [9], a compressive sensing method was described to directly recover background subtracted images. When the objects of interest(foreground) occupy a small portion of the camera view, the background images can be sparsely represented in spatial domain. A low dimensional sparse representation of the background is then learned and adapted.

To the best of our knowledge, this paper is a first attempt in exploring the sparsity for optical flow estimation. By imposing sparsity, smoothness in local regions and motion discontinuities can be encoded at the same time. Therefore we don't need other constraints before accurately recovering the flow field. In addition, dense noise problem in sparse representation is still an open problem in general. The method in [26] can work when only some portion of the measurements have noises. In this paper we address this problem in optical flow by adding gradient sparsity constraint and performing robust estimation. In Section 3 we will validate that optical flow field can be sparsely represented. Accurately estimated flow based on sparsity will also be shown. Section 4 introduces our robust flow estimation method against noise in detail. Experiments are shown in Section 5, and Section 6 draws the conclusions.

3. Formulation

3.1. Sparsity model for the flow field

A basic assumption for most differential methods in optical flow estimation is that the intensities across the images remain constant over time. When the displacements are small, we can perform a first-order Taylor expansion

and obtain the optical flow constraint:

$$I_x u + I_y v + I_t = 0 \tag{1}$$

where I_x, I_y and I_t are the first-order partial derivatives of the image intensities $I(x,t)$, while (u,v) are the components of motion in the horizontal and vertical direction respectively. By stacking the constraints of all pixels in the image together, we have the following representation:

$$-\mathbf{I}_t = \begin{bmatrix} \mathbf{I}_x & \mathbf{I}_y \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \tag{2}$$

where $\mathbf{I}_t$, $\mathbf{u}$ and $\mathbf{v}$ are all vectors of length n, the number of pixels in image. $\mathbf{I}_x$ and $\mathbf{I}_y$ are n-dimensional diagonal matrix. Since it is a highly underdetermined system, various additional constraints or priors are imposed to regularize the problem, such as the global smoothness[13] and local velocity constancy[18]. However, they fail at some local regions in the flow field, especially those at motion discontinuities.

In this paper, we consider the optical flow problem from a distinct perspective. If we divide the whole images into small blocks (*e.g.* 16×16, or 32×32), the flow field in each block, although discontinuities and large variations may still be present, always has some simple structures and exhibits homogeneity in each structure. Thus we can find a basis to represent the flow field such that it is **sparse** in another domain. Actually if we look at $\mathbf{u}$ and $\mathbf{v}$ respectively, each can be considered as a simple image. Therefore commonly used bases for sparse representation of natural images can be adopted here, such as Wavelet, DCT or Curvelet. Fig.2 gives us an illustration. Although the flow field in Fig.2(a) seems complex, most Haar Wavelet coefficients of $\mathbf{u}$ and $\mathbf{v}$ are nearly zero, which means the flow in the Haar Wavelet domain is sparse.
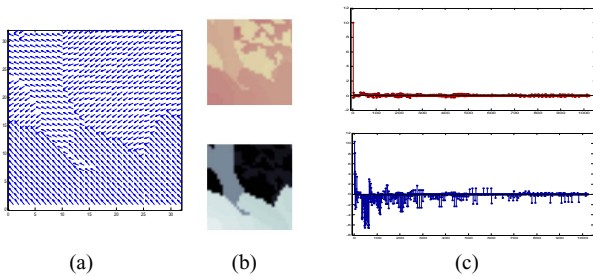

(a)   (b)   (c)

Figure 2. Sparse representation of the flow field. (a) The flow field, (b) the color map of $\mathbf{u}$ and $\mathbf{v}$ respectively, and (c) the Wavelet coefficients of $\mathbf{u}$ and $\mathbf{v}$, most of which are small values near zero.

Consider $\mathbf{W}$ as the bases, then the flow can be sparsely represented as $\mathbf{u} = \mathbf{W}\mathbf{s}_\mathbf{u}$ and $\mathbf{v} = \mathbf{W}\mathbf{s}_\mathbf{v}$, where $\mathbf{s}$ are the sparsity coefficients. Accordingly, (2) can be rewritten as

$$-\mathbf{I}_t = \begin{bmatrix} \mathbf{I}_x & \mathbf{I}_y \end{bmatrix} \begin{bmatrix} \mathbf{W} & 0 \\ 0 & \mathbf{W} \end{bmatrix} \begin{bmatrix} \mathbf{s}_\mathbf{u} \\ \mathbf{s}_\mathbf{v} \end{bmatrix} \tag{3}$$

The optical flow estimation problem therefore can be formulated as sparse signal recovery based on highly shorted measurements.

3.2. Computing optical flow using the sparsity model

The theory in compressive sensing has already validated that when the signal is sparse, it can be accurately recovered from much fewer measurements with high probability, by minimizing its L_1-norm[8, 11]. Let $\mathbf{y} = -\mathbf{I}_t$, $\mathbf{A} = \begin{bmatrix} \mathbf{I}_x & \mathbf{I}_y \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} \mathbf{W} & 0 \\ 0 & \mathbf{W} \end{bmatrix}$, and $\mathbf{s} = \begin{bmatrix} \mathbf{s}_\mathbf{u} & \mathbf{s}_\mathbf{v} \end{bmatrix}^T$. The problem can be casted as

$$\mathbf{s}^* = \arg min \|\mathbf{s}\|_1 \ s.t. \ \mathbf{y} = \mathbf{A}\mathbf{B}\mathbf{s} \tag{4}$$

and the estimated flow can be computed by

$$\mathbf{f}^* = \begin{bmatrix} \mathbf{u}^* & \mathbf{v}^* \end{bmatrix}^T = \mathbf{B}\mathbf{s}^* \tag{5}$$

This is a convex optimization problem that can be conveniently converted to a linear program like basis pursuit[11]. Such a sparsity model is quite simple, but powerful. Consider a noise-free case in Fig.3, in a 16×16 flow block, assuming the measurements $\mathbf{y}$ have no noise and the constraints in (1) strictly hold, and the Haar Wavelet is used for sparse representation. The estimated flow by this basic sparsity model is almost perfect, preserving crisp motion boundaries and notably outperforming the results from traditional methods based on smoothness priors[13, 4].

Notice that when solving (5), we solely maximized the sparsity of $\mathbf{s}$ under the intensity constancy constraints without any other assumed priors. This is the most different point for our method from others, and is also the main reason that the method performs well at discontinuities in Fig.3. Meanwhile, as a non-parametric method, it can accurately estimate the motion generated by various parametric models, such as affine transformation and rotation(Fig.4).

Moreover, even if the measurements are not fully sampled, our method can still work effectively. In Fig.5, the partial derivatives are calculated on a limited number of pixels. We can accurately estimate the whole dense flow field, even if only 60% of the partial derivatives are known. That means, by our sparse model, we can estimate $2n$ flow signals from no more than $0.6n$ observed measurements.

Up to now we assume that the constraint in (1) strictly holds. However, it is a first-order approximation of intensity constancy. In addition, there inevitably exist noises when calculating the derivatives I_x, I_y and I_t, which violate the constraints and impair the performance of our method. In the next section, we will thoroughly discuss our robust flow estimation when the constraint in (1) is not satisfied at some pixels.
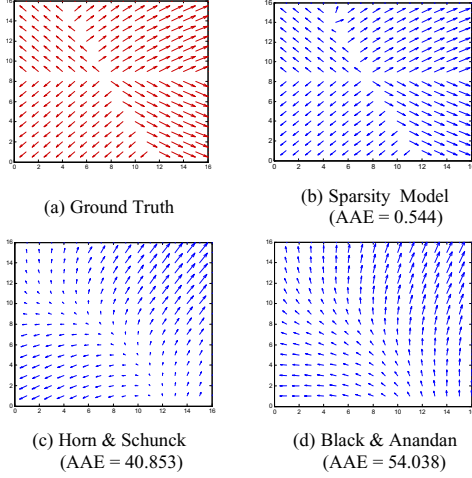
(a) Ground Truth

(b) Sparsity Model
(AAE = 0.544)

(c) Horn & Schunck
(AAE = 40.853)

(d) Black & Anandan
(AAE = 54.038)

Figure 3. Computing optical flow without measurement noise. The flow field is 16×16. AAE = Average Angular Error. No pyramid is used in all methods.

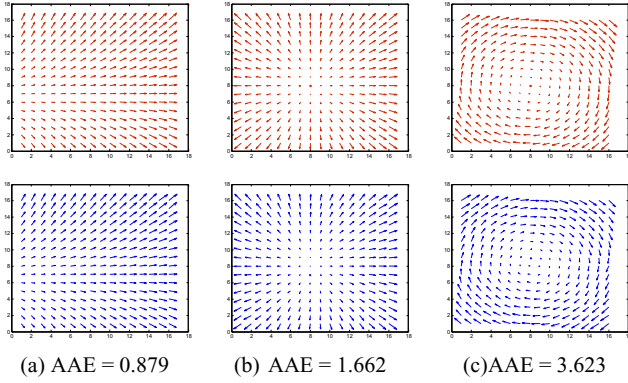(a) AAE = 0.879   (b) AAE = 1.662   (c)AAE = 3.623

Figure 4. Estimating different parametric flow using non-parametric sparsity model. First row: ground truth, second row: estimated results. AAE = Average Angular Error.
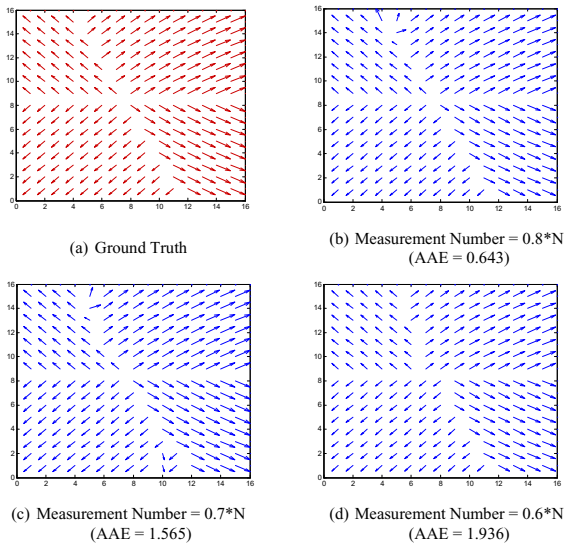

(a) Ground Truth

(b) Measurement Number = 0.8*N
(AAE = 0.643)

(c) Measurement Number = 0.7*N
(AAE = 1.565)

(d) Measurement Number = 0.6*N
(AAE = 1.936)

Figure 5. Optical flow estimation from small portion of measurement. N = 256.

4. Robust Optical Flow

4.1. The noise problem

The noises may come from 1) the residuals after first-order approximation, 2) the measurement noises in the calculation of I_t, which are reflected by $\mathbf{y}$ in (5), and 3) the noises in computing of I_x and I_y, which is contained in $\mathbf{A}$. The constraint $\mathbf{y} = \mathbf{ABs}$ hence can be modified as

$$
\begin{aligned}
\mathbf{y} &= (\mathbf{A} + \mathbf{N_A})\mathbf{Bs} + \mathbf{e_y} \\
&= \mathbf{ABs} + \mathbf{e_A} + \mathbf{e_y} = \mathbf{ABs} + \mathbf{e}
\end{aligned}
\tag{6}
$$

Accordingly, our sparsity model in (5) can be reformulated as

$$
\mathbf{s}^* = \arg min \|\mathbf{y} - \mathbf{ABs}\|_2 + \lambda \|\mathbf{s}\|_1
\tag{7}
$$

where λ is the Lagrangian multiplier. (7) relaxes the constraint $\mathbf{y} = \mathbf{ABs}$, which is more reasonable than (5) when measurement noises need to be considered.

Unlike the noises in other applications of compressive sensing to the computer vision problems (*e.g.* face recognition[26] and human tracking[19]), the noise $\mathbf{e}$ here is always dense with large variations. We looked into that property by comparing the differences between the estimated noise-free measurements and the ground truth. Fig.6(a) shows the statistical distribution of the noise values. The noise distribution can be considered as a zero-mean Gaussian. We tested the performance of (7) by manually adding dense zero-mean Gaussian noises with different variations. The relationship of the the reconstructed errors and the measurement noises is shown in Fig.6(b).
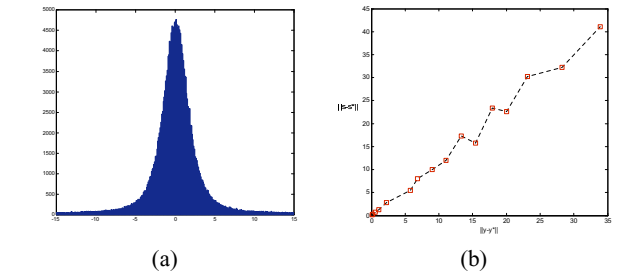

(a)   (b)

Figure 6. The property of noise in optical flow.

In Fig.6(b) the x-axis is the L_2-norm of measurement noise $\|\mathbf{e}\|_2$, and the y-axis is the L_2-norm of reconstructed error $\mathbf{e_s} = \|\mathbf{s} - \mathbf{s}^*\|_2$. We can see that the reconstructed errors increase linearly with respect to the measurement noises. This linear relationship is also stated and proved theoretically in [7]. Moreover, when the noises are large, the solution by L_1-norm minimization may not converge to the ground truth. Look at a simple 2-D space illustration in Fig.7. In Fig.7(a), S is the constraint, x^* has the minimum L_1 norm among all points falling onto S, and therefore is the optimal solution. Due to the existence of noises, the

original constraint becomes S' in Fig.7(b), and the optimal solution becomes x', which is far away from x^*.
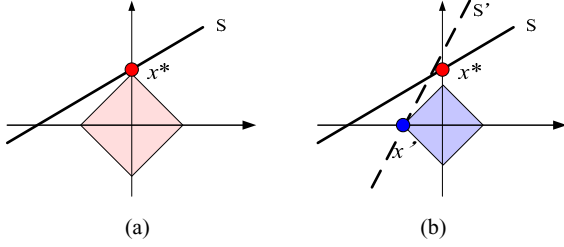

(a)   (b)

Figure 7. The influence of noise on L_1-norm minimization.

Therefore, traditional way to solve (5) or (7) is not good enough to handle noises. Meanwhile, the error correction method for sparse noises would also fail when the noises are dense. To this end, we propose a new way to handle the noise problem: add a new sparsity constraint, and use robust estimation techniques to reject the outliers.

4.2. Additional sparsity constraint

In Section 3.1 we stated that the flow field is sparse in the Wavelet domain. Now we argue that not only the flow field is sparse, but the gradient field of the flow is also sparse. It is straightforward to understand. As we stated in 3.1, the flow field usually has some structures. In each structure the motion tends to change smoothly, the gradients of the flow are also small; while at the discontinuities between structures, the gradients are large. Therefore only a small number of gradients should have large values, and the flow gradient field has to be sparse in the image domain. Here we want to emphasize on two points:

1. This is a strong complement to our original sparsity constraint, as a sparse flow field does not necessarily have sparsity on its gradient field.

2. This constraint is clearly different from other first-order constraints like global smoothness. It allows large values on some derivatives, and imposes no constraints on their spatial relationships.

Similarly, we want to minimize $\|\nabla \mathbf{f}\|_1$. Normally it is difficult to compute $\|\nabla \mathbf{f}\|_1$ accurately. However, if we calculate the horizontal and vertical components respectively, and use simple kernel $[1, -1]$ and $[1, -1]^T$ to convolute $\mathbf{u}$ and $\mathbf{v}$, each component can be linearly approximated as

$$\begin{aligned} \mathbf{g_u} &= \begin{bmatrix} \mathbf{D}_x & \mathbf{D}_y \end{bmatrix}^T \mathbf{u} \\ \mathbf{g_v} &= \begin{bmatrix} \mathbf{D}_x & \mathbf{D}_y \end{bmatrix}^T \mathbf{v} \end{aligned} \qquad (8)$$

where $\mathbf{g_u}$ and $\mathbf{g_v}$ have length $2n$, $\mathbf{D}_x$ and $\mathbf{D}_y$ denote linearized first order derivative operators. Hereby we can represent the flow gradient field as a linear combination of the flow field:

$$\mathbf{g} = \mathbf{D}\mathbf{f} \qquad (9)$$

where $\mathbf{g} = \begin{bmatrix} \mathbf{g_u} & \mathbf{g_v} \end{bmatrix}^T$, $\mathbf{f} = \begin{bmatrix} \mathbf{u} & \mathbf{v} \end{bmatrix}^T$. Combining with the original sparsity model (7), we propose a new sparsity model:

$$\begin{aligned} (\mathbf{s}^*, \mathbf{g}^*) &= \arg\min \|\mathbf{y} - \mathbf{ABs}\|_2 + \lambda_1 \|\mathbf{s}\|_1 \\ &\quad + \mu(\|\mathbf{y} - \mathbf{AD}^+\mathbf{g}\|_2 + \lambda_2 \|\mathbf{g}\|_1) \\ \mathbf{f}^* &= \arg\min \|\mathbf{f} - \mathbf{Bs}\|_2 + \mu \|\mathbf{f} - \mathbf{D}^+\mathbf{g}\|_2 \end{aligned} \qquad (10)$$

where μ is the parameter to balance these two constraints. $\mathbf{D}^+$ is the pseudo-inverse of $\mathbf{D}$. We can modify the derivative operators on the boundaries of the flow field to make $\mathbf{D}_x$ and $\mathbf{D}_y$ full rank respectively without influencing the whole sparse representation.

By adding this additional sparsity constraint, the performance can be significantly improved, as shown in Fig.8. We crop an image region of size 32×32 from the *Dimetrodon* sequence in the Middlebury Dataset[3]. The first frame is shown in Fig.8(a), and the ground truth is shown in Fig.8(b). Fig.8(c) is the result by solving (7), and the result by solving (10) is shown in Fig.8(d). Obviously combing two constraints achieved better performance, the average angular error decreased from 8.145 to 5.342.
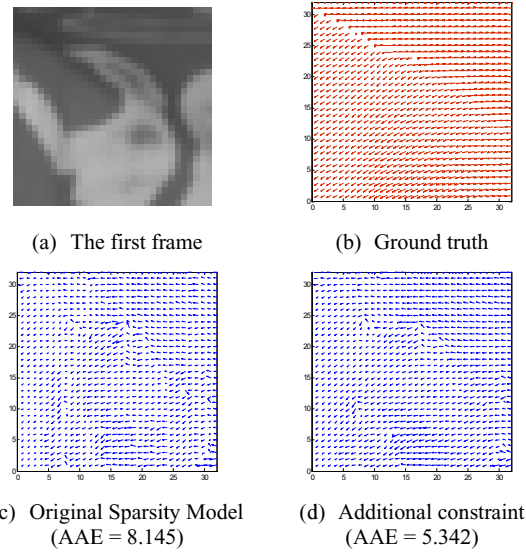

(a) The first frame   (b) Ground truth

(c) Original Sparsity Model   (d) Additional constraint
     (AAE = 8.145)   (AAE = 5.342)

Figure 8. The improvement by performing the additional sparsity constraint on the flow gradient field.

4.3. Robust estimation

In Section 3.2 we already showed that our sparsity model can recover the flow field from a small portion of measurements, while in Section 4.1 we analyzed that the measurement noises in flow estimation obey quasi-Gaussian distribution. Most measurements contain small errors while
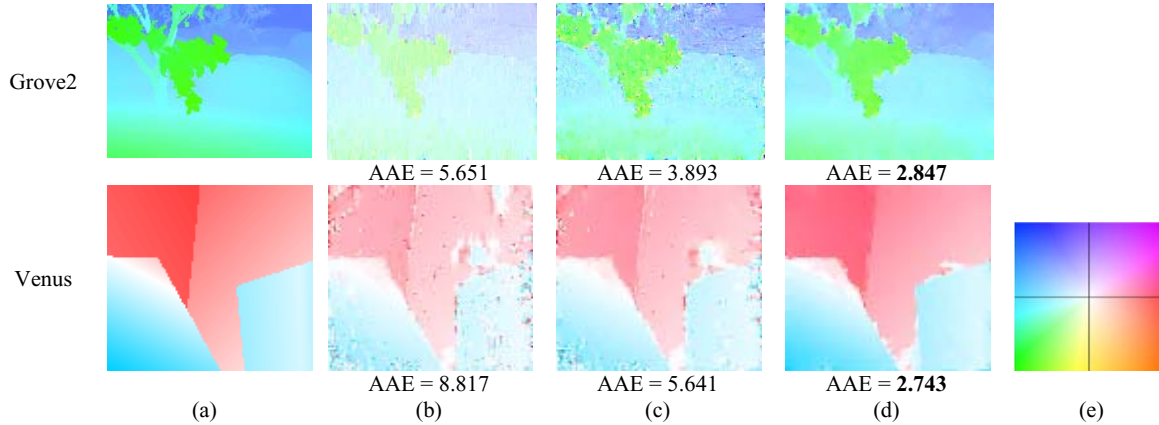
Figure 9. The improvement by additional sparsity constraint and robust estimation. (a) Ground truth, (b) results by Wavelet sparsity, (c) combing derivative sparsity, (d) robust estimation, (e) color map for flow.

some may be completely corrupted and become outliers. Therefore, if we can reject those outliers and use the remaining samples to estimate the entire flow, the results shall be improved.

Thus we can use robust estimation techniques to remove the outliers and refine the estimated flow. Among those techniques, RANSAC is the most commonly used one. The algorithm of robust optical flow estimation using RANSAC can be described in three steps:

a) **Initialization.** Let n be the number of all measurements, thus the number of variables in flow field is $2n$. In the beginning of each iteration, randomly select $\beta n (0 < \beta < 1)$ samples as the measurements $\mathbf{y}$. Estimate the entire flow field by solving (10).

b) **Model fitting.** For the remaining $(1-\beta)n$ samples, each sample has its estimated flow u_i and v_i from Step 1. Compute SSD as its matching score:

$$S = \sum_{(x,y) \in \mathcal{N}(i)} (I_0(x,y) - I_1(x + u_i, y + v_i))^2$$

(11)

where $\mathcal{N}(i)$ are the neighbors of sample i. Normally the neighborhood is 3×3 or 5×5. If S is small, it means that u_i and v_i are accurately estimated. Therefore if $S < T$, we consider sample i fits the model, otherwise not. T here is a pre-defined threshold.

c) **Refining.** If more than $\varepsilon n (\beta < \varepsilon < 1)$ samples are fitted, the model is considered good. We then use all the fitted samples to estimate the flow again. Compare the average SSD of all flows estimated by this model with the initial model in Step 1, and choose the one of a smaller SSD. If less than εn samples are fitted, the model is considered bad and is discarded.

Repeat the above three steps until we find a model with a sufficiently good matching score. The model with the best score then gives us the final result. In our experiments we set $\beta = 0.6$ and $\varepsilon = 0.8$. T is initially set to be a small value. If after many iterations we still cannot find a model with enough fitted samples, T can be increased sequentially.

We use this robust estimation algorithm to refine the result in Fig.8, the average angular error is reduced from 5.342 to 4.442. It is a distinctive property for the sparsity model that it can reconstruct the whole signal from a small portion of the measurements, which allows us to estimate the flow field using robust estimators such as RANSAC. The robust estimation for the sparsity model is not confined in optical flow, but can be generalized to most applications of sparse representation and compressive sensing.

5. Experiments

We performed our algorithm to estimate dense optical flow of two-frame gray images on the Middlebury Dataset[3](*Dimetrodon*, *Venus*, *Hydrangea*, *Grove2* and *Grove3*). As we assume the first-order approximation of intensity constancy (1) holds, which requires small $\mathbf{u}$ and $\mathbf{v}$, we downsampled all the images to 1/4 of their original heights and widths. The images are then divided by overlapping blocks of size 16×16. The sparsity model is used for each image block to generate the final results. The Haar Wavelet is used as the codebook for sparse representation. In L_1-norm minimization, we used the method introduced in [14] and the code in [2]. As for the regularization parameters, there are no guidelines to select them at this moment. Instead, we tried different parameters and used the one with the best performance. In particular, we set $\lambda_1 = \lambda_2 = 0.8$, and $\mu = 2$ in (10). We observed that the results only have insignificant differences in a range around these parameters. The Horn & Schunck's method and Black & Anandan's method is based on the implementation of [1]. Since the images are already downsampled, the pyramid level in

these two methods is set to be 1.

Fig.9 gives us two examples that the additional sparse constraint and robust estimation helped improve the performance of sparsity model. We display the flow field according to the color map in Fig.9(e). Fig.9(b) are the results by solely seeking the sparsity in the Wavelet domain. As we can see there are many noises in the estimated flow. The flow field of the *Grove2* sequence looks pale. It is because the colors are mapped from the relative values of the flow. There are some large noises which make other flow values relatively small. In the *Venus* sequence, the flows in the middle region as well as the upper right corner are fully corrupted. By performing additional gradient sparsity constraint, the influence of measurement noises is attenuated, and the flows present stronger homogeneity in local regions, as shown in Fig.9(c). Large noises are removed in *Grove2*, and the upper right corner of the flow field in *Venus* is recovered well. However, the middle part is still very noisy.

After robust estimation, the outliers in the measurements are removed, and accordingly the performance is significantly improved, especially in those noisy regions. Notice that the flows in a small region of the middle part of *Venus* are still not very accurate. It is in part because that the image derivatives in this region are very small and most measurements are severely corrupted by noise. As for now we only estimate the flows block by block based on their own information, if too many measurements in a block have large noise, robust estimation would also fail. A possible way to handle this problem is to enlarge the block size and adjust the flows in each block according to the results in the surrounding blocks.

We also compared the results with Horn & Schunck's method[13], and Black & Anandan's method[4]. The reason we choose these two methods to compare is that we aim to investigate the effectiveness of different assumptions in optical flow under the same condition, i.e., the sparsity assumption *vs.* global and piecewise smoothness assumption. Therefore we didn't choose other multiple layer methods with coarse-to-fine strategies. Meanwhile, the reported performance of these two methods in the Middlebury benchmark is comparable with others.

The average angular error and endpoint error on each sequence have been listed in Table1. The errors of our method are much lower than that of the other two methods on every sequence. The estimated flow of the remaining sequences other than *Grove2* and *Venus* (which have already been shown in Fig.9) are shown in Fig.10. We can see from Fig.10 that our results are more crisp at the motion discontinuities and meanwhile also smooth in homogeneous regions. It validated that our method is powerful to handle motion discontinuities without any other priors or using methods like model selection.

Currently we have not submitted our results to the Mid-

Sequences		H&S	B&A	Sparsity
Dimetrodon	AAE	7.292	5.142	**3.874**
	Endpoint	0.152	0.109	**0.081**
Venus	AAE	12.963	7.938	**2.743**
	Endpoint	0.364	0.204	**0.073**
Hydrangea	AAE	9.385	6.08	**2.472**
	Endpoint	0.271	0.195	**0.069**
Grove2	AAE	9.614	4.524	**2.847**
	Endpoint	0.231	0.109	**0.067**
Grove3	AAE	17.81	9.843	**5.749**
	Endpoint	0.539	0.316	**0.188**

Table 1. Average angular error and endpoint error of three methods on 5 sequences.Sparsity model achieved better performance.

dlebury Benchmark for evaluation, as we performed our model in downsampled images with small motion. However, our results can be refined by any coarse-to-fine methods such as warping[5] and motion segmentation. Actually simply interpolating our results to higher resolution level can still provide comparable performance. Since this is not the focus of this paper, and it is hard to clarify whether the final high resolution results benefit from our approach or the coarse-to-fine techniques, we choose to present the performance in the initial coarse level. And the comparison results under the same condition has already proved its effectiveness.

6. Conclusion

This paper introduced a new sparsity prior in optical flow estimation. Unlike traditional smoothness constraints, sparsity can handle motion homogeneities and motion discontinuities at the same time, which leads to accurate estimation of the flow field. A stronger sparsity constraint on flow gradient field and robust estimation techniques are also introduced to guarantee the robustness of the sparsity model in handling the measurement noises.

We believe that this sparsity model is a new fundamental assumption in optical flow, and can be well combined with other existing methods in flow estimation. It worths more research efforts to further explore the sparsity in optical flow estimation.

Acknowledgements

References

[1] *http://www.cs.brown.edu/ dqsun/research/software.html.*

[2] *http://www.stanford.edu/ boyd/l1_ls/.*

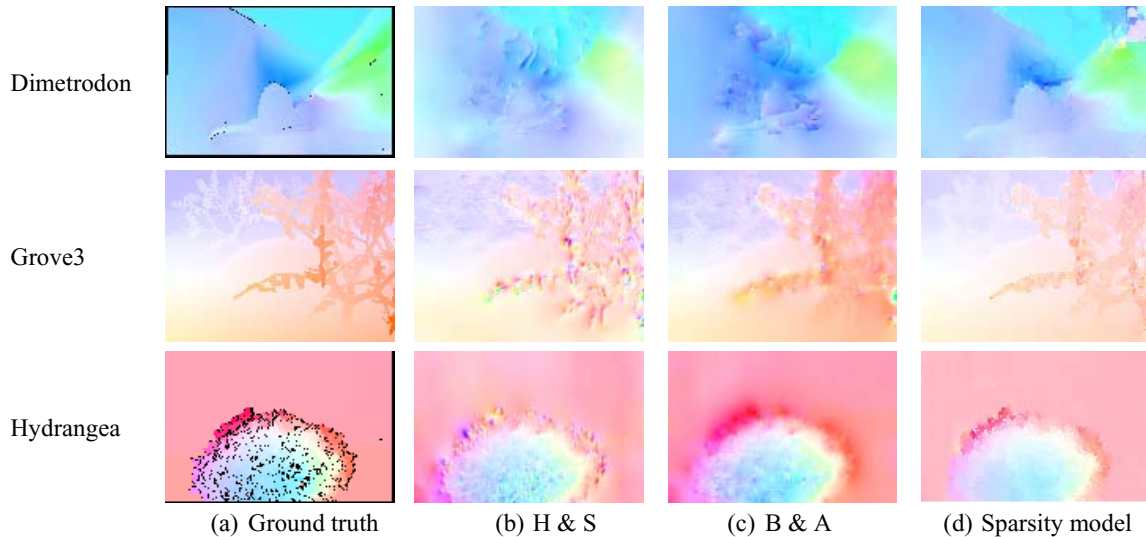|			
Dimetrodon			
Grove3			
Hydrangea			
(a) Ground truth	(b) H & S	(c) B & A	(d) Sparsity model

Figure 10. Comparison Results. (a) Ground truth, (b) Horn & Schunck's method, (c) Black & Anandan's method, (d) Sparsity model.

[3] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *ICCV*, 2007.

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

[5] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. *ECCV*, 2004.

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

[7] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math*, 59(8):1207–1223, 2006.

[8] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008.

[9] V. Cevher, D. Reddy, M. Duarte, A. Sankaranarayanan, R. Chellappa, and R. Baraniuk. Compressive sensing for background subtraction. In *ECCV*, 2008.

[10] T. Darrell and A. Pentland. Robust estimation of a multilayered motion representation. In *CVPR*, pages 296–302, 1991.

[11] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52:1289–1306, 2006.

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

[13] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[14] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. A method for large-scale l_1-regularized least squares. *IEEE Selected Topics in Signal Processing*, 1, 2007.

[15] C. Lei and Y. H. Yang. Optical flow estimation on coarse-to-fine region-trees using discrete optimization. In *ICCV*, 2009.

[16] V. Lempitsky, S. Roth, and C. Rother. Discrete-continuous optimization for optical flow estimation. In *CVPR*, 2008.

[17] H. Liu, R. Chellappa, and A. Rosenfeld. Accurate dense optical flow estimation using adaptive structure tensors and a parametric model. *TIP*, 12:1170–1180, 2003.

[18] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI*, pages 674–679, 1981.

[19] X. Mei and H. Ling. Robust visual tracking using l1 minimization. In *ICCV*, 2009.

[20] E. Mémin and P. Pérez. Hierarchical estimation and segmentation of dense motion fields. *IJCV*, 45:129–155, 2002.

[21] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

[22] S. Roth and M. J. Black. On the spatial statistics of optical flow. In *ICCV*, 2005.

[23] D. Sun, S. Roth, J. Lewis, and M. Black. Learning optical flow. In *ECCV*, 2008.

[24] J. Wang and E. Adelson. Representing moving images with layers. *TIP*, 3:625–638, 1994.

[25] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in pde-based computation of image motion. *IJCV*, 45:245–264, 2001.

[26] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *PAMI*, 31(2), 2009.

[27] L. Xu, J. Chen, and J. Jia. Segmentation based variational model for accurate optical flow estimation. In *ECCV*, 2008.

[28] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. In *CVPR*, 2008.

[29] H. Zimmer, A. Bruhn, J. Weickert, L. Valgaerts, A. Salgado, B. Rosenhahn, and H. Seidel. Complementary optic flow. *EMMCVPR*, 2009.