Statistical Modeling of Optical Flow

Dongmin Ma, Veronique Prinet, Cyril Cassisa

Institute of Automation
Chinese Academy of Sciences
Beijing, China
dmma@nlpr.ia.ac.cn
{prinet,cassisa}@liama.ia.ac.cn

Abstract

Optical flow estimation is one of the main subjects in computer vision. Many methods developed to compute the motion fields are built using standard heuristic formulation. In this paper, however, we learn a motion model. We develop a hybrid model by combining the learnt model with Markov Random Field (MRF). And then we introduce a method based on "Radial Basis Function Neural Network" (RBF) to learn the model. When computing the displacement field, a Gaussian pyramidal down-sampling decomposition technique is employed. At each pyramidal level, we use bi-linear interpolation combined with an efficient warping technique to generate a residual image, which is then used at the finer level to compute the flow. To minimize the energy, we use two different discrete optimization methods: Graph-Cut algorithm, Tree-Reweighted Message Passing (TRW-S) algorithm. Results are demonstrated for our approach on synthetic images and fluid images from the real world.

1. Introduction

Motion estimation is one of the most important tasks in many applications, such as video tracking, 3D reconstruction, motion segmentation, etc. Though lots of significant progresses have been achieved, it is still a challenging problem in computer vision.

There are two dominating challenges in motion estimation [10]: one is how to establish an appropriate model, the other is how to compute the equation. For modeling the motion, Horn and Schunck [8] proposed a seminal method called Differential Optical Flow computation, which is based on two assumptions: brightness constancy and first-order smoothness. Xu *et al*. [18] used segmentation based method to do the motion estimation. Prinet *et al*. [11] proposed one method that can compute global and local motion of multi-structure objects. In [13, 14], they model motion in a statistical way by learning the model with ground truth. For images with divergence and vortexes motions, many people try to model the motion by extending the brightness constancy to other situations. Corpetti *et al*. [5] models motions from atmospheric images by using continuity equation for spatial term and uses one second-order "div-curl" regularization as smooth term. Bereziat *et al*. [2] models atmospheric images by postulating that deformation motion is associated with a constant volume or a constant total intensity over time.

Many algorithms can be used to compute the energy [6, 7, 9, 15, 16, 17], such as Belief Propagation (BP), Graph-Cut, Tree-Reweighted Message Passing (TRW-S), etc. All these methods, however, suffer from the problem of labeling [10] when applied to motion estimation. Unlike in stereo and in de-noising applications where it is easier to discretize disparities and pixel values, we have to describe velocity vectors using discrete labels when computing velocity field.

In this paper, we tackle the problem by modeling the motion using one hybrid approach of statistical learning and MRF. When developing the model, we build it in a MRF framework, which is then learned with ground truth. To learn the model, we introduce one sophisticated model based on "Radial Basis Function Neural Network" (RBF) [3, 12] to learn the motion, which shows better result than the state-of-the-art. In order to detect large displacements in the image, we developed a multi-resolution scheme. At each level of the pyramid, bi-linear interpolation combined with an efficient warping technique is used to generate a residual image, which is used at the finer neighboring level to compute the flow. The energy at each level is solved with two discrete optimization methods: Graph-Cut, TRW-S.

The rest of the paper is organized as follows. Section 2

introduces the modeling of optical flow. In section 3, we detail the energy minimization. Experiments and results are discussed in section 4. In section 5, we give conclusions.

2. Modeling of optical flow

The optical flow estimation problem is formulated as a probability inference problem combined with pair-wise MRF. Given two consecutive images, we compute the posterior probability of the motion field. It can be described as

$$p(\mathbf{w}|I_1, I_2; \Omega) = p(I_1|\mathbf{w}, I_2; \Omega_d) \cdot p(\mathbf{w}|I_2; \Omega_r) \quad (1)$$

Where, Ω_r and Ω_r are the parameters of the model, $\mathbf{w} = (u, v)$ is the velocity, I_1 and I_2 are the image frames. Different from [14], we model $p(I_1|\mathbf{w}, I_2; \Omega_d)$ and $p(\mathbf{w}|I_2; \Omega_r)$ in two different ways. When modeling $p(I_1|\mathbf{w}, I_2; \Omega_d)$, we use RBF model as in equation (5) to learn the negative log of statistics of the brightness constancy error. For $p(\mathbf{w}|I_2; \Omega_r)$, we learn the negative log of statistics of pair-wise site velocity components derivatives with RBF. In this way, we can build the model that integrates MRF and statistical learning.

2.1. Data term

The brightness constancy assumption is used to model the data term of the optical flow. The data term can be written as follows:

$$p(I_1|\mathbf{w}, I_2; \Omega_d) = \prod_{\mathbf{s} \in C_1} \rho(V_d(I_{\mathbf{s}}, \mathbf{w}_{\mathbf{s}}); \Omega_d) \quad (2)$$

Where, $\rho(\cdot)$ is the potential function of data term, $\mathbf{s}$ is a site, C_1 is the single position set of image, $V_d(I_{\mathbf{s}}, \mathbf{w}_{\mathbf{s}})$ is used in two ways as follows.

First, for learning the model,

$$V_d(I_{\mathbf{s}}, \mathbf{w}_{\mathbf{s}}) = I_2(\mathbf{s} + \mathbf{w}_{\mathbf{s}}) - I_1(\mathbf{s}) \quad (3)$$

Second, for computing the flow field,

$$V_d(I_{\mathbf{s}}, \mathbf{w}_{\mathbf{s}}) = \nabla I_2(\mathbf{s}) \cdot \mathbf{w}_{\mathbf{s}} + I_t(\mathbf{s}) \quad (4)$$

Where, I_t is the temporal derivative of the image sequence, ∇I_2 is the spatial derivative of image I_2. In this section, we focus on the learning of the model. Many methods can be used to approximate the potential function $\rho(\cdot)$, e.g. GSM. Also other people choose robust function to approximate it. Here we use RBF to model the data term, and the parameters are learnt from the training data. It has two main advantages: first, the weight of the RBF networks can be trained easily and efficiently. Second, the total number of candidate functions in the model is not very large and does
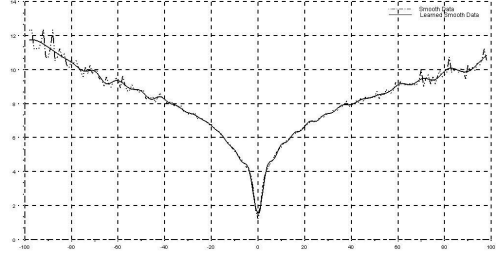

Figure 1. Negative log of statistics of the brightness constancy error. The dot dash line is the original data, and which is learned with RBF model (the solid line).

not increase with the number of the input variables. The RBF model can be written in the following way:

$$\rho(x; \Omega) = \Sigma_{i=1}^{N} w_i \cdot exp((x - \mu_i)^2 / \sigma^2) + C \quad (5)$$

Where, w_i is the weight of each components, μ_i is the mean of each components of the model, C is a constant. Here σ is empirically set as constant.

After established the data term model, we need to train it. First, both two images are pre-processed with a Gaussian smoothing filter. We then warp the second frame to the first frame using flow ground truth, which is used for training. Finally, the negative log of the statistics of the error of the warped image and the first image can be computed. The result is shown as in Figure 1.

2.2. Smooth term

Instead of using some function directly to approximate the standard derivatives of velocity, we learn the smooth term. The pair-wise site smooth term can be modeled as:

$$p(\mathbf{w}|I_2; \Omega_r) \propto \prod_{\mathbf{s}, \mathbf{s}' \in C_2} \rho(u_{\mathbf{s}} - u_{\mathbf{s}'}; \Omega_h) \cdot \rho(v_{\mathbf{s}} - v_{\mathbf{s}'}; \Omega_v) \quad (6)$$

Where, u and v are the components of $\mathbf{w}$, $\mathbf{s}$ is site, C_2 is the pair-wise cliques.

The horizontal and vertical velocity derivatives are learned with RBF separately. The result is shown in Figure 2.

This model describes the relationship of neighboring pair-wise pixels. By comparing the learned results of statistics of standard derivatives of horizontal velocity component and vertical velocity component, we find they are identical. So we set parameters of $\rho(u_{\mathbf{s}} - u_{\mathbf{s}'}; \Omega_h)$ and $\rho(v_{\mathbf{s}} - v_{\mathbf{s}'}; \Omega_v)$ same for the simplicity of computation.

2.3. Multi-resolution formulation

In order to detect large displacement in motion, we derive to use a multi-resolution method to solve the problem.
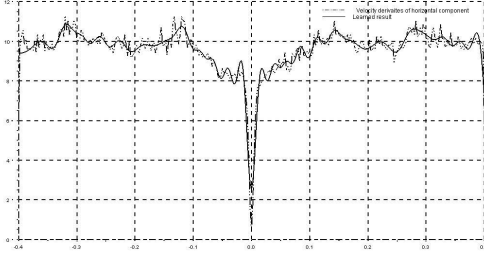
Figure 2. Negative log of statistics of standard derivatives of horizontal velocity component. The dot dash line is the original data, and the solid line is the learned result.

The total levels of the pyramid can be denoted with N. I^k denotes the k-th level. And I^N denotes the lowest level or the coarsest level. The frames I_1^N and I_2^N are used directly in equation (1) to compute velocity. Then the result is interpolated to its higher neighboring level and is used to warp the second frame at that level. When doing interpolation, we use bi-liner interpolation. We denote the interpolated velocity from the lower level using $\overline{\mathbf{w}}^k$. So the second frame can be warped using the following method:

$$\overline{I}_2^k(\mathbf{s}) = I_2^k(\mathbf{s} + \overline{\mathbf{w}}^k) \qquad (7)$$

Where, $\overline{I}_2^k$ is the second warped frame at level k. Then I_1^k and $\overline{I}_2^k$ are used to compute the velocity at level k. In our method, we used four levels of pyramid.

3. Energy minimization

We can write negative log of equation (1) in the following way:

$$E(\mathbf{w}) = \sum_{\mathbf{s} \in C_1} V_d(\mathbf{w_s}) + \lambda \sum_{\mathbf{s},\mathbf{s}' \in C_2} V_r(\mathbf{w_s}, \mathbf{w_{s'}}) \qquad (8)$$

Where, V_d is the negative log of the data term, V_r is the negative log of the smooth term, $\mathbf{w}$ is the velocity vector, $\mathbf{s}$ is site, C_1 is the set of single sites, C_2 is the set of pairwise sites, λ is the weight of smoothing term. The energy is one typical form of MRF. It can be solved with discrete optimization methods such as TRW-S and Graph-Cut.

When using discrete optimization methods, unlike finding labels for stereo and de-noising in which cases the disparities and pixel values can be easily discretized, it is hard to describe velocity vectors with labels. Here we used a method of mapping labels to velocity vectors. The labels in our algorithm are used as indexes for velocity vector. In this way the problem can be solved.

3.1. TRW-S algorithm

TRW-S or sequential TRW is an improved version of TRW. It can guarantee that the estimated lower bound energy does not decrease. Compared with BP and Graph-Cut, TRW-S has many advantages, e.g. Graph-Cuts can be used only to a limited class of energy functions, while TRW-S can be applied to any function of the form of equation (8) [9].

3.2. Graph-Cut algorithm

In this paper, we also used Graph-Cut swap algorithm. For a pair of labels α and β, the swap algorithm makes local improvements by finding those nodes labeled α whose label should be change to β, or vice-verse. The swap operation continues until there is no swap for any pair of labels α and β, and then the labels that minimizes the energy field can be got. To compute the optimal swap for the whole graph, the min-cut/max-flow formulation is often employed [16].

4. Results

We evaluate our approach on two image sequences: Yosemite sequence and real fluid images sequence. Average Angle Error (AAE) and the Standard Deviation (STD) of Angle Error (AE) are used to compare results of our method with those of other methods. Smooth Pre-processing is used for images in each level before computing the data term. In our method, we use Gaussian smoothing to do the pre-processing. Because different ground truth may generate different data for training, we truncate the image difference to [-70 70].

4.1. Results for Yosemite sequence

The Yosemite sequence [1] used contains cloud motion. The optical flow is computed using our model with a pyramid level of 4. Figure 3 shows the sequence used and the computed results.

We compute the arrow image results using grid size of 7. To compute the AE we used the following formulation:

$$AE = \arccos(\frac{\mathbf{w}_e \cdot \mathbf{w}_{gt}}{\|\mathbf{w}_e\| \|\mathbf{w}_{gt}\|}) \qquad (9)$$

Where, $\mathbf{w} = (u, v, 1)$, the subscript e and gt means estimated velocity and ground truth velocity respectively.

From Figure 3(d) and 3(f), the Angle Error (AE) is low except places of edge between mountain and cloud, and places of cloud. For the first condition, it stems from the fact of occlusion, and for the second condition, it is for the reason that the illumination of cloud varies in the original images data.

Table 1 shows the quantitative result of AAE and the Standard Deviation (STD) of AE computed using our model compared with those of other methods. When computing the AAE value and STD value, we ignored 10 pixels on the borders of the image for comparability. From the table, we can see that the results computed using our model are better.

(a) Original data

(b) Ground truth

(c) OF of Graph-Cut

(d) AE of Graph-Cut

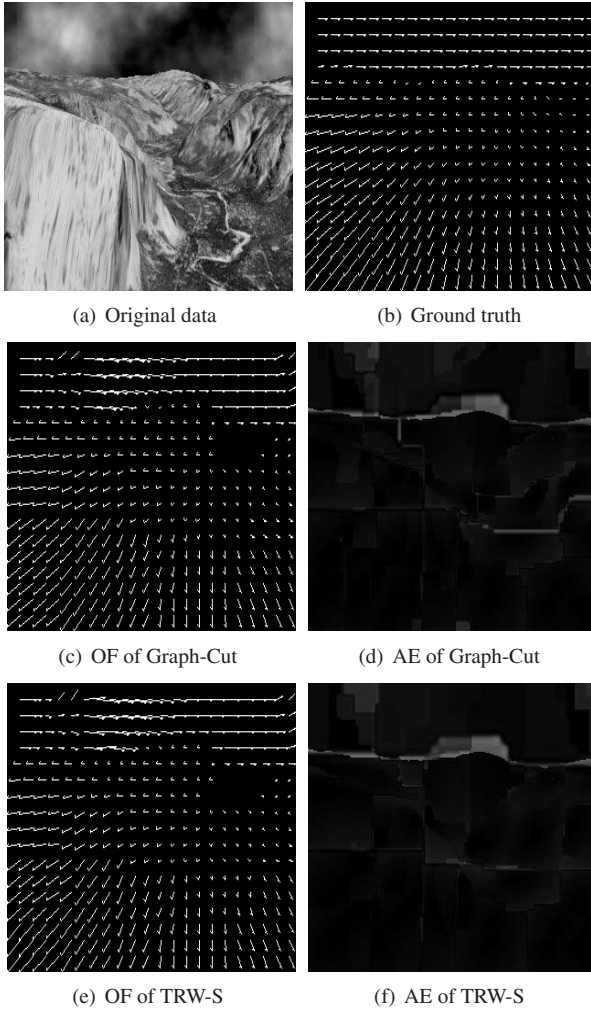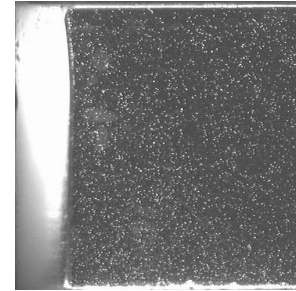(e) OF of TRW-S

(f) AE of TRW-S

Figure 3. Optical flow of the Yosemite sequence estimated with Graph-Cut and TRW-S, and their Angle Error result. OF=Optical Flow, and AE=Angle Error.

	AAE	STD
Anandan	13.36	15.64
Liu et al	13.18	13.24
Roy et aL	13.00	—
Singh	10.44	13.94
Our Model+GC	10.42	12.31
Our Model+TRW-S	10.23	12.27

Table 1. Quantitative comparison results of Yosemite sequence. AAE=Average Angle Error, STD=Standard Deviation, GC=Graph-Cut.

(a) Flow images sequence

(b) OF of Graph-Cut

(c) OF of TRW-S

Figure 4. Optical flow of the real fluid flow sequence estimated with Graph-Cut and TRW-S. OF=Optical Flow.
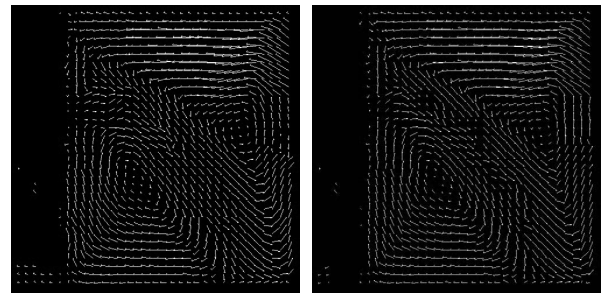
In our model, we used one sophisticated model RBF and integrated the learned model with MRF. They can build better model for the motions in the images sequence. The adoption of Graph-Cut and TRW-S algorithms for the energy minimization also contributes to get better results. They can usually find labels that result lower energy than other discrete optimization methods such as ICM. Among all the results, the first three gives the worst results. They do not use the scheme of multi-resolution and can not detect true displacements over one pixel. Comparing with methods using robust functions or other methods using existing functions to model the energy, there are two advantages for our model: first, the data term and smooth term in our energy are learned. they can describe attributes of the data better than using fixed functions. Second, the model is developed combined with MRF. This makes it much more convenient to employ better optimization methods such as TRW-S to do the energy minimization. And it will help to compute more accurate flow field.

4.2. Results for real fluid image sequence

The real fluid image sequence we used was acquired in the laboratory. It is the movement of convection. The average displacement is about 4 pixels and 15 pixels for the largest displacement [4].

In Figure 4, we illustrate the results computed from our model with Graph-Cut and TRW-S algorithm. From the result, the velocity field computed is fine visually. And the flow computed with Graph-Cut is even better than that computed using TRW-S at the region of the central of the flow image where the flow moves from top-left to down-right. Both the results give us encouraging results.

5. Conclusions

We introduce to learn the statistical model based on Radial Basis Function Neural Network. And we develop a model that is one hybrid of statistical learning and MRF. When minimizing the energy, we use TRW-S and Graph-Cut algorithms to minimize the energy, and by using multi-resolution scheme, the number of labels used in computation becomes small, so the computation time is greatly decreased.

Further more, integrating more than one ground truth into the model is needed. The range of labels and their initialization is also crucial for the accusation of estimated velocity. Future work will focus on the analysis of model learning method and labeling.

References

[1] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques, springer, international journal of computer vision. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[2] D. Bereziat, I. Herlin, and L. Younes. Motion estimation using a volume conservation hypothesis. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 6, 1999.

[3] S. A. Billings, H. L. Wei, and M. A. Balikhin. Generalized multiscale radial basis function networks. *Neural Networks*, 2007.

[4] C. Cassisa, V. Prinet, L. Shao, S. Simoens, and C. L. Liu. Optical flow robust estimation in a hybrid multi-resolution mrf framework. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 793–796, 2008.

[5] T. Corpetti, E. Memin, and P. Perez. Dense estimation of fluid flows. *IEEE Transactions on pattern analysis and machine intelligence*, 24(3):365–380, 2002.

[6] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70(1):41–54, 2006.

[7] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. Learning low-level vision. *International Journal of Computer Vision*, 40(1):25–47, 2000.

[8] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Computer vision*, 17:185–203, 1981.

[9] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 28(10), 2006.

[10] V. Lempitsky, S. Roth, and C. Rother. Fusionflow: Discrete-continuous optimization for optical flow estimation. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[11] V. Prinet, C. Cassisa, and F. F. Tang. Mrf modeling for optical flow computation from multistructure objects. *International Conference on Image Processing*, 2006.

[12] Y. J. Qu and B. G. Hu. Rbf networks for nonlinear models subject to linear constraints. *IEEE conference for granular computing*, 2009.

[13] S. Roth and M. J. Black. On the spatial statistics of optical flow. *International Journal of Computer Vision*, 74(1):33–50, 2007.

[14] D. Sun, S. Roth, J. P. Lewis, and M. J. Black. Learning optical flow. *Proceedings of the 10th European Conference on Computer Vision: Part III*, pages 83–97, 2008.

[15] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):1068–1080, 2008.

[16] M. F. Tappen and W. T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters. *Ninth IEEE International Conference on Computer Vision*, pages 900–906, 2003.

[17] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky. Map estimation via agreement on trees: Message-passing and linear-programming approaches. *IEEE Transactions on Information Theory*, 51(11):3697–3717, 2005.

[18] L. Xu, J. Chen, and J. Jia. A segmentation based variational model for accurate optical flow estimation. *Proceedings of the 10th European Conference on Computer Vision: Part I*, pages 671–684, 2008.