A VARIATIONAL MODEL WITH ADAPTIVE REGULARIZATION BASED DENSE STEREO MATCHING

Manimala Mahato[1], Shirish S. Gedam[2] and Jyoti Joglekar[3]
[1,2]C.S.R.E., Indian Institute of Technology Bombay, Powai, Mumbai, India-400076
[1]Email: manimala.mahato@iitb.ac.in
[2]Email: shirish@iitb.ac.in
[3]DJSCOE, Vile Parle West, Mumbai, India-400056
[3]Email: jyotij1968@gmail.com

KEY WORDS: Dense disparity map, Stereo correspondence, Variational method, Adaptive regularization.

ABSTRACT:

An adaptive regularization based variational model is presented in this work, for obtaining dense disparity map of non-rectified stereo images. To overcome the problems such as to estimate accurate disparities near object boundaries, in repetitive texture regions or textureless regions and in occluded areas, we estimate disparity map by minimizing the global energy functional consists of data and regularizer terms, using variational model with coarse-to-fine pyramidal approach.The pyramidal approach is used to handle large disparities. To optimize the regularizer in the energy fuctional, we use spatially varying regularization parameter instead of a fixed value for the entire image which is common to any variational framework but unsuitable for remotely sensed stereo images because of various image characteristices such as texture. In this approach, we use the pixel wise image gradient and the estimated intermediate disparity gradient to initialize and update the regularization parameter at each pixel location. The initialization consists of K-means clustering in the image gradient space and assignment of a per-class value of regularization. This has impact on the required regularization factor for a group of pixels. Step wise updation is involved at all levels in the pyramid by calculating the disparity in scale space followed by computing the derivative of the disparity map.The proposed method is found to be effective in dealing with the limitation of fixed regularization of the core variational method for increasing the accuracy while estimating the dense disparity map. We evaluate the estimated disparity map quantitatively using bad pixel error with various threshold values comparing with the ground truth. Bad pixel error is calculated considering all the pixels of the input image as well as only nonoccluded pixels.

1. INTRODUCTION

In the field of computer vision and digital photogrammetry, the reconstruction of three-dimensional information is one of the key problems. One motivation that drives many researchers in this field is the goal to imitate the performance of the human visual system to understand the depth of the scene by a machine. Very large parts of the human brain are reserved only for the processing of the information provided by our eyes. The imitation of these processing capabilities by a machine is a challenging task.

Classical approaches for 3D reconstruction focus on image-based reconstruction i.e. estimating structure from stereo image pairs or from image sequences. Stereo image is the images of the same scene, captured from different viewpoints. Image of a scene is the projection of 3D scene onto a 2D plane. During this process, third dimension of an object i.e. depth is lost. To construct 3D from sequence of two-dimensional images, it is necessary to obtain the depth information. Depth map is computed using the disparity map obtained from the stereo matching process. Disparity is the difference in the projected positions of a point on the left and right images. Correspondence problem is to find these two projected positions in the stereo image pair. Since Marr and Poggio (Marr,1976), a variety of algorithms have been developed to solve correspondence problem for three decades starting from primitive area based technique and feature based technique to energy based technique. Accurate dense disparity map is the basic requirement for 3D reconstruction. Area based technique (Scharstein, 2002) generates dense disparity map but less accurate. Feature based technique (Faugeras,1993, Joglekar, 2014), gives more accurate but sparse disparity map. In energy based technique (Slesareva, 2005, Alvarez, 2002, Scharstein, 2002), deviation from data and regularization constraints are penalized by minimizing variational formulations. Variational method outperforms the other two techniques because they estimate disparities even at non-textured regions. As a result, always 100% dense disparity map is obtained.

Various attempts have been made to solve dense correspondence computation problem using variational method. It is a very well-known method used for detecting motion of feature points i.e. optical flow in computer vision. Optical flow is the displacement field that describes the pixel shift of same feature points between two frames.

In variational framework, data term minimizes the difference between the feature descriptors along the displacement. Regularization term guaranties that the displacement for neighbor pixels are similar. Tradeoff between the data term and the regularization term is decided by a parameter known as regularization parameter λ. Value of this λ plays very important role at object boundaries. To preserve object boundaries by respecting discontinuities in the image, regularizer constraint can be modeled *image-driven* (Alvarez 2002, Mansouri 1998) or *solution-driven* (Robert, 1996) or by selecting the value of λ appropriately.

Different types of images might have different features or single image consists of different features at different area. A fixed regularization parameter for all the images or all pixels of the entire image is therefore an unsuitable factor to balance the data term and regularization term. The selection of an appropriate λ for each pixel rather than for each image is very much required. In this paper, we address this issue and present a variational method with adaptive regularization parameter (VMARP) approach to estimate the value of spatially varying λ depending upon the discontinuity in image features. Discontinuity in the image feature such as intensity, texture, color, and surface is marked by edges. Here we have marked edges based on intensity of pixels. To detect edges in any direction, two mutually perpendicular Sobel gradient detectors are used.

In this paper, we have focused on two objectives. First one is to estimate dense disparity map from stereo image pair using variational method with fixed λ based on (Sun, 2014). The second objective is to use adaptively changing λ in variational framework for each pixel based on image feature during the computation of dense disparity map. Our approach combines *image-driven approach* by respecting discontinuity in the image and *solution-driven* approach by respecting discontinuity in the gradually updated disparity.

Our paper is organized as follows: In Section 2, first we have discussed variational model with fixed λ. The quality of dense disparity map is affected by regularization parameter. This motivates us to propose a novel variational method with different regularization parameter estimated for each pixel. Quality measure used for the evaluation is defined in Section 3. In Section 4, the performance of our approach is evaluated on Middlebury stereo training datasets. Summary in Section 5 concludes this paper.

2. VARIATIONAL METHOD

Let us consider a rectified stereo image pair where $I_L(X)$ and $I_R(X)$ are the left image and right image respectively and X= (x, y) denotes the pixel location within the rectangular image domain Ω. The rectified stereo image pair has only horizontal disparity. The goal of this paper is to estimate the horizontal disparity d_X at each pixel X of $I_L(X)$ using proposed variational method with adaptive regularization parameter approach. The disparity d_X of a point X in the left image is the displacement field between X and its corresponding point X'in the right image i.e. $d_X = X - X'$. In the following, we present the variational model with adaptive regularization parameter approach.

A. The Variational Model

We compute d_X as minimizer of the energy functional in its spatially discrete form as

$$E(d) = E_{Data} + \lambda E_{Reg} \tag{1}$$

where the data term E_{Data} is given by

$$E_{Data} = \sum_{(x,y)} \left(I_L(x,y) - I_R(x + d_{x,y}, y)\right)^2 \quad (2)$$

Regularization term E_{Reg} is given by

$$E_{Reg} = \sum_{(x,y)} \left(d_{x,y} - d_{x+1,y}\right)^2 + \left(d_{x,y} - d_{x,y+1}\right)^2 \quad (3)$$

And λ is the regularization parameter. A standard incremental multi-resolution technique is used for optimization of the energy functional (Sun et al., 2014) based on coarse-to-fine pyramidal framework.

B. Adaptive Regularization Parameter

In our approach, spatially variant regularization parameter λ is estimated adaptively. It is done in two stages: Initialization stage and Updation stage.

1) **Initialization Stage:** In this stage, for every pixel $X = (x, y)$ of left image I_L, λ_X is initialized based on the magnitude of the gradient of that pixel. Gradient magnitude denotes the strength of the edges at discontinuity. It will be maximum at the edges and minimum in smooth textureless area. Steps for initialization of λ_X are given below:

a) Calculating the gradient magnitude: Left image I_L of the stereo image pair is convolved with two 3 X 3 mutually perpendicular Sobel gradient operator S_1 and S_2 as in (4). This computes $g_x(x, y)$ as x-directional gradient and $g_y(x, y)$ as y-directional gradient at each pixel (x, y).

$$S_1 = \begin{matrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{matrix} \qquad S_2 = \begin{matrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{matrix} \quad (4)$$

Gradient magnitude of each pixel is calculated using (5)

$$\nabla I_L(\mathrm{x}, \mathrm{y}) = \sqrt{g_x{}^2(\mathrm{x}, \mathrm{y}) + g_y{}^2(\mathrm{x}, \mathrm{y})} \quad (5)$$

b) **Classification of each pixel**: Each pixel X of I_L is classified into K clusters using iterative k-means classification algorithm based on its gradient magnitude. The steps are as follows:

i) The total number of clusters K, required to classify the given image, is determined based on the range of the gradient magnitude using (6).

$$K = M * \ln\left(\nabla I_{Lmax} - \nabla I_{Lmin}\right) \quad (6)$$

Where ∇I_{Lmax} and ∇I_{Lmin} are the computed maximum and minimum value of the gradient magnitude of the left image I_L respectively and M is a empirically selected multiplicative factor. Each cluster is numbered c_1, c_2, c_3, …, c_k.

ii) Mean of each cluster μ_k^0, k= 1, …, K is computed from the range of gradient magnitude of I_L which is ∇I_{Lmin} to ∇I_{Lmax} where superscript 0 represent the zeroth iteration for finding mean.

iii) Each pixel X is assigned to cluster k based on the minimum absolute difference of $\nabla I_L(\mathrm{x}, \mathrm{y})$ with μ_k^0.

iv) The mean of each cluster is recomputed. Existing cluster mean μ_k^0 is replaced by the updated cluster mean μ_k^{new}.

v) Step (iii) to (iv) is iterated till the following converging condition is satisfied.

$$max(|\mu_k^{new} - \mu_k^0|) > th \tag{7}$$

Where *th* is user-specified threshold and k = 1, …, K.

c) **Initialization of regularization parameter**

i) Range of λ is decided empirically and it is constant for the entire image. Minimum value of λ is λ_{min} and the maximum value of λ is λ_{max}.

ii) λ of cluster c_k is calculated using (9)

$$\lambda_{c_k}^0 = \lambda_{min} + (K - c_k) * \frac{(\lambda_{max} - \lambda_{min})}{K} \quad k = 1,2, \dots, K \tag{8}$$

All pixels belong to cluster c_k will have initial regularization parameter value $\lambda_{c_k}^0$.

iii) L_λ^0 is obtained where for each pixel *(x,y)*, $L_\lambda^0(x,y) = \lambda_{ck}^0$ if $(X=(x,y)^T \in c_k)$. Size of L_λ^0 is same as I_L. Disparity of each pixel of $I_L(x,y)$ is calculated using (1), (2) and (3) where $\lambda = L_\lambda^0(x,y)$.

d) **Pyramid construction***:* The disparity for each pixel is estimated using the coarse-to-fine pyramidal framework where at fine pyramid level disparity is determined with the initialization from the coarse pyramid level. Hence, there is a requirement of λ value for each pixel at every pyramid level. A λ-pyramid is constructed from L_λ^0, named P_λ^0. P_λ^0 is used for the computation of disparity map at every pyramid level. The size of P_λ^0 is same with the size of the image-pyramid. Consider the total number of levels in P_λ^0 is *PL*. The regularization parameter at each pixel $X = (x, y)^T$ at every pyramid level *pl* of P_λ^0 will be denoted by $L_\lambda^{pl}(x,y)$ where *pl= PL, … ,1* from coarser level to finer level.

2) **Updation Stage:** During the updation of disparity at every pyramid level *pl* for every pixel *(x, y)*, $L_\lambda^{pl}(x,y)$ is also updated. We consider pixels of two categories: pixels on the surface of any object where neighborhood pixels have similar disparity and pixels at the edge of the object where neighborhood pixels have dissimilar disparity. Regularization parameters for these two categories are taken care separately. Steps for updation stage are as follows:

a) The pixels on the surface of any object have similar rate of change disparity. Rate of change of disparity is computed for every pixel *(x,y)* at every pyramid level *pl* using following equation:

$$D_{pl}(x,y) = \frac{(d_{pl}(x,y) - d_{pl+1}(x,y))}{d_{pl+1}(x,y)} \tag{9}$$

Where d_{pl+1} is disparity map calculated at coarser pyramid level and d_{pl+1} is disparity map calculated at finer pyramid level.

b) Weight w_{pl}^1 is calculated at every pixel of pyramid level *pl* as follows:

$$w_{pl}^1 = \frac{D_{pl}(x,y)}{D_{plmax}} \tag{10}$$

Where D_{plmax} denotes the maximum rate of change of disparity from coarser level *pl+1* to finer level *pl*.

c) Pixels at the edge of the object is identified by large disparity variation as compared to disparity of the pixels on same surface. Disparity variation at pyramid level *pl* is captured by gradient magnitude of rate of change of disparity at *pl* i.e. ∇D_{pl}.$\nabla D_{pl}$is obtained by taking the derivative on D_{pl}.

d) Another weight w_{pl}^2 is calculated for every pixel *(x,y)* at every pyramid level *pl* using the following equation:

$$w_{pl}^2(x,y) = \frac{\nabla D_{pl}(x,y)}{\nabla D_{plmax}} \tag{11}$$

Where ∇D_{plmax} is the maximum value of ∇D_{pl}which denotes maximum discontinuity point at pyramid level *pl.*

e) The disparity of the pixels within the object is found to be similar, and at the object boundary, it is substantially different. Hence, regularization parameter λ is required to be increased proportional to w_{pl}^1 and inversely proportional to w_{pl}^2 . A combined weight w_{pl}^c is calculated at every pyramid level *pl* for every pixel (x, y). It combines w_{pl}^1 and w_{pl}^2 using the following equation:

$$\mathrm{w}_{\mathrm{pl}}^{\mathrm{c}}(x,y) = k_1 \mathrm{w}_{\mathrm{pl}}^1(x,y) + \frac{1}{k_2 \mathrm{w}_{\mathrm{pl}}^2(x,y) + 1} \tag{12}$$

where $k_1$and k_2 is non-negative number. By selecting the suitable positive values of k_1 and k_2, the combined weight can be adjusted in (13). In our experiment, we have taken $k_1=k_2=1$.

f) Finally, $L_\lambda^{pl}(x,y)$ is updated using weight w_{pl}^c . The updated value will be:

$$\mathrm{L}_\lambda^{\mathrm{pl}}(x,y) = \mathrm{L}_\lambda^{\mathrm{pl}}(x,y) + \mathrm{w}_{\mathrm{pl}}^c(x,y) * \mathrm{L}_\lambda^{\mathrm{pl}}(x,y) \tag{13}$$

Where *pl= PL, …, 1* from coarser level to finer level.

3. Quality Measures

We compute Bad Pixel Error (BPE) as the quality measure for the computed disparity map *d(x,y)* based on the ground truth map *gt(x,y).* BPE [2] defines percentage of bad matching pixels having disparity error more than the acceptable tolerance threshold.

$$B_{\delta_d} = \frac{1}{N} \sum_{(x,y)} (|\mathrm{d(x,y)} - \mathrm{gt(x,y)}|) > \delta_d \tag{14}$$

Where δ_d is the tolerance threshold value, and N is the total number of pixels. For the experiments in this paper we use δ_d= 0.5 and 1.0.

4. Results and Discussion

The proposed algorithm has been evaluated on *Teddy* and *Piano* stereo image pair from Middlebury training datasets (Scharstein, 2014). Figure (1a) and (1c) shows the left image and right image of Teddy dataset and Figure (1b) and (1d) shows the same of Piano dataset. Ground truth of the input stereo pair is shown in Figure (1e) and (1f). Disparity maps estimated by the proposed method are shown in Figure (1g) and (1h). It is clearly shown from the disparity maps that proposed method is giving relatively sharp object boundaries. Disparities in connected areas such as walls and objects are estimated homogeneously.

We have tested the impact of adaptive regularization by comparing our proposed method with variational method with fixed regularization parameter (VMFRP) approach. For assessment of the performance of our approach, two different experiments are designed. In our first experiment, to evaluate the effectiveness of variational method on stereo matching, we have tested the performance of variational method using fixed regularization parameter

(VMFRP). In the second experiment, to evaluate the effectiveness of adaptive regularization in variational method, we have compared the performance of our VMARP (adaptive λ) approach with VMFRP (fixed λ) approach on the input datasets. Table I shows the comparison of VMFRP (fixed λ) and VMARP (adaptive λ) approach using Bad Pixel error (BPE) calculated using equation (14) with tolerance threshold 0.5 and 1.0 considering all pixels as well as only non-occluded pixels. The effect of adaptively adjusting the regularization in variational method on different image regions increases the BPE accuracy than using the best smoothness parameter for the entire image. The time complexity of our approach can be reduced by using parallel computing. This adaptive smoothness parameter approach in the variational framework can also be used for motion detection from video frames in computer vision.

TABLE I. COMPARISON OF BAD PIXEL ERROR (BPE) ERROR ESTIMATED BY VARIATIONAL METHOD WITH FIXED REGULARIZATION PARAMETER (VMFRP) AND VARIATIONAL METHOD WITH ADAPTIVE REGULARIZATION PARAMETER (VMARP) APPROACH ON TEDDY AND PIANO STEREO IMAGE PAIR

Image name	Considered pixels	Technique	BPE 0.5	% Improvement	BPE 1.0	% Improvement
Teddy	All pixels	VMFRP	17.14	--	13.22	--
		VMARP	16.02	6.5%	11.5	13%
	Non-Occluded pixels	VMFRP	16.67	--	12.99	--
		VMARP	15.05	9.72%	11.04	15.01%
Piano	All pixels	VMFRP	39.26	--	31.9	--
		VMARP	36.01	8.27%	28.06	12.04%
	Non-Occluded pixels	VMFRP	36.03	--	28.91	--
		VMARP	33.06	8.24%	25.52	11.73%

5. CONCLUSION

The most important input for 3D reconstruction is accurate dense disparity map. Conventionally, area based and feature based approach is used for stereo matching. But it is prone to errors caused by distortion in the imaging process. We have proposed a method to estimate dense disparity map using variational framework in which adaptive regularization parameter balances the two main constituents: data term and regularization term in an excellent manner at object boundaries leading to sharper boundaries in disparity map, and at textureless smooth region. As a result, more accurate and 100% dense disparity map is estimated. The methodology is proposed by analyzing the characteristics of remotely sensed stereo images with minimal input from the user where ground truth is not available. The ultimate objective of our present study is to estimate dense disparity map for remotely sensed images.

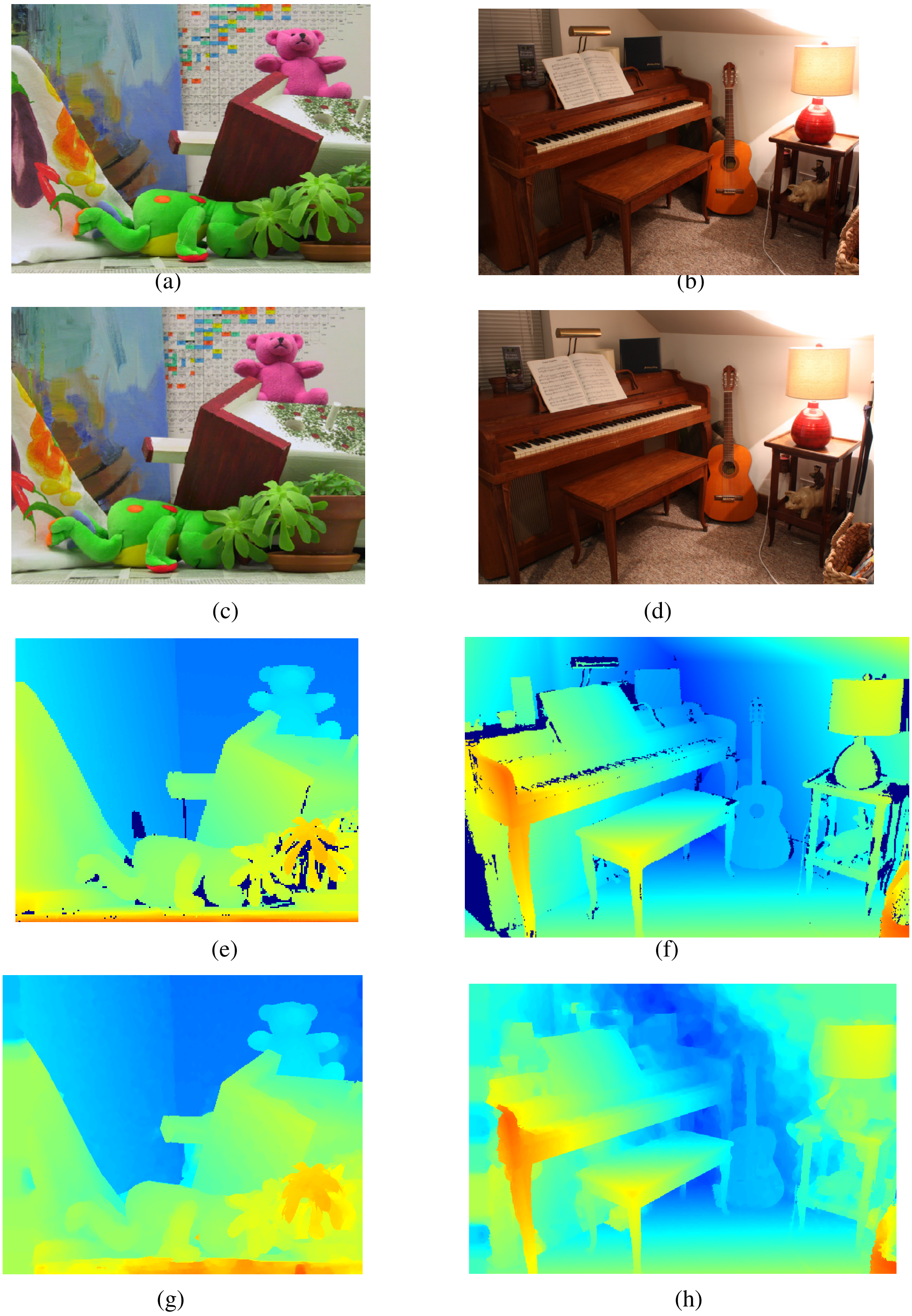

Figure 1. Stereo dataset: (*Column 1*): *Teddy* dataset, (*Column 2*): *Piano* dataset, (*Row 1*) : Left Image, (*Row 2):* Right Image, (*Row 3*): Ground Truth, (*Row 4*): Disparity map estimated by proposed method.

REFERENCES

Alvarez L., Deriche R., Sanchez J., and Weickert J., 2002 Dense disparity map estimation respecting image discontinuities: A pde and scale-space based approach, Journal of Visual Communication and Image Representation, vol. 13(1), pp. 3-21.

Faugeras O., 1993, “Three-dimensional computer vision: a geometric viewpoint,” MIT press.

Joglekar J., Gedam S. S., and Mohan B. K., 2014, Image matching using sift features and relaxation labeling technique-A constraint initializing method for dense stereo matching, IEEE Transactions on Geoscience and Remote Sensing, vol. 52(9), pp. 5643-5652.

Marr D. and Poggio T., 1976, Cooperative computation of stereo disparity, In From the Retina to the Neocortex, pp. 239-243. Springer.

Mansouri A. R., Mitiche A., and Konrad J., 1998, Selective image diffusion: application to disparity estimation. In Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on, pp. 284-288. IEEE.

Raket L. L.. 2012, Local smoothness for global optical flow, In 2012 19th IEEE International Conference on Image Processing, pp. 57-60.

Robert L. and Deriche R.., 1996, Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities, In European Conference on Computer Vision, pp. 439-451. Springer.

Scharstein D. and Szeliski R., 2002, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,International journal of computer vision, vol. 47(1-3), pp. 7-42.

Scharstein D., Hirschmuller H., Kitajima Y., Krathwohl G., Nesic N., Wang X., and Westling P., 2014, High-resolution stereo datasets with subpixel-accurate ground truth, In German Conference on Pattern Recognition, pp. 31-42. Springer.

Slesareva N., Bruhn A., and Weickert J., 2005, Optic flow goes stereo: A variational method for estimating discontinuity-preserving dense disparity maps, In Pattern Recognition, pp. 33-40. Springer.

Sun D., Roth S., and Black M. J., 2014, A quantitative analysis of current practices in optical flow estimation and the principles behind them, International Journal of Computer Vision, vol. 106(2), pp. 115-137.