Survey on Variational Scene Flow Estimation

Akash Yadav[1] and Rene Schuster[2]

[1] ayadav@rhrk.uni-kl.de
[2] rene.schuster@dfki.de

Abstract. The aim of the paper is to study various variational methods for scene flow estimation and classify them based on the camera setting and the energy formulation. The paper further focuses on to asses the optimization algorithms used in the literature and deduce a logical comparison in terms of deviation from the ground truth of different data sets.

Keywords: Scene flow, Optical flow, Disparity, dense flow, Variational Method, Optimization, Successive over smoothing SOR, Coarse-grain smoothing

1 Introduction

Scene flow estimation is a method to capture the motion vector of points corresponding to a surface in a three dimensional frame of reference between two time steps. Fundamentally, it has been evolved from the concept of optical flow which showcases the same behaviour but in two dimensions. It can also be inferred that optical flow is the projection of scene flow in image plane (2D).

Mathematically, scene flow is estimated from multiple projections of optical flows from different cameras. This also results in an ill-posed problem of non-consistent optical flows. This problem is solved using a regularization or smoothness term in the global energy function. However, unreasonable parameters for regularization in the energy function cause over-smoothing of scene flow which is a primary anomaly in the estimation.

It is important to note that the first noticeable contribution in the direction of scene flow estimation was done by Vedula et al. [8], who used multiple camera setting to process the optical flow through first-order approximation of the depth map to estimate the 3-D motion vector. Since then the applications of scene flow estimation have been vast ranging from image segmentation, image matching, feature extraction etc.

Although despite of its rising applications, scene flow estimation also inherits the problems concerned with optical flow calculation which is subjected to accuracy problems corresponding to occlusion, large displacement motion, varying illumination and insufficient texture.

This paper is an attempt to explore the variational methods of scene flow estimation. Section 2 illustrates the challenges and variational method associated with optical flow constraint. Section 3 further showcases the emergence of scene flow concept from underlying principle of optical flow and also discussing disparity and stereo motion. The formulation of energy framework from data terms has been discussed in Section 4. While approach of the researchers and their optimization algorithms has been discussed in Section 5. Section 6 states the evaluation of algorithms on data sets. While, Section 6 provides the closing argument for the paper.

2 Background

Scene flow estimation is the result of the summation of optical flow concepts and stereo matching. Therefore, in order to develop the algorithm for the evaluation of scene flow, it is important to consider the basics of optical flow and challenges associated with it.

2.1 Optical Flow

Optical flow is the motion vector corresponding to the apparent motion of the object in the image plane between two consecutive time steps. Mathematically, the image characteristics at a particular time can be illustrated as an intensity function $I(x,y,t)$ at image pixel position $\mathbf{x} = (x,y)^T$. Intensity function varies with the data source in grey scale or colored image sequences.

Challenges The calculation scheme for optical flow (and hence the scene flow estimation) is primarily based on the assumption that the brightness constancy(grey value) of a moving pixel between two consecutive time steps remains constant. However, the assumption has its limitation in scenes of varying illumination or on observing transparent object, which limits the usability for small displacements only. In addition, for complex scenes such as multiple object movement, occlusion leads to motion discontinuities and computation of erroneous flaws. Further the aperture problem caused by small aperture diameter makes it inaccessible to the structural information corresponding to the edges and corners of the object to generate two dimensional motion vector. This makes the optical flow calculation as ill posed, which is mainly solved by the introduction smoothness terms to predict the motion vector from previously available data.

Optical Flow Constraint The optical flow constraint as mentioned earlier, is the assumption that the brightness constancy of the object remains constant between two consecutive time steps. Mathematically, it can be represented as:

$$I(x,y,t) = I(x+u,y+v,t+1) \tag{1}$$

where $\mathbf{u}=(u,v)$ is the optical flow vector of a pixel $\mathbf{x}=(x,y)$ from time t to $t+1$. On linearizing the optical flow constraint for the image using first order Taylor approximation; we get:

$$I(x,y,t) \approx I(x,y,t+1) + \nabla I(x,y,t+1)^T \begin{pmatrix} u & v \end{pmatrix}^T \tag{2}$$

$$0 = \underbrace{I(x,y,t+1) - I(x,y,t)}_{I_t(x,y,t+1)} + \nabla I(x,y,t+1)^T \begin{pmatrix} u & v \end{pmatrix}^T \tag{3}$$

By denoting I_t, I_u, I_v for the partial derivatives of the image function, we can summarize the optical flow as the function of apparent motion vector (u,v):

$$OFC(u,v) : 0 = I_t + I_u u + I_v v \tag{4}$$

However, it can be observed that the problem of aperture becomes visible with two variables to be determined from one equation. This becomes an ill posed problem and can only be solved by considering finite number of values for the variables and solving the equation for the other one.[9]

2.2 Variational Optimization for Optical Flow

The calculation scheme of optical flow algorithms primarily focuses on finding the solution for the ill-posed problem occurs due to aperture. Therefore, optical flow algorithms are divided into two classes depending upon the type of regularization: feature based methods and variational methods. Feature based methods computes the optical flow displacement for the pixel and its neighbouring pixels independently of the optical flow solution from the other pixels in the image. Lucas-Kanade in 1981 provided a solution based on the feature based method with an idea to introduce gradient descent method to minimize the grey value between the image pixel and the grey value of the next pixel. Variational approach considers the optical flow solutions for the surrounding pixels and later applying smoothness function on the optical flow field. It was introduced in 1981 by Horn and Schunck [4] to solve the problem of under-determined optical flow constraints by achieving the smoothness through penalizing the derivative of optical flow field.

$$\min_{u(x),v(x)} \int_{\Omega} (|\nabla u(\boldsymbol{x})|^2 + |\nabla v(\boldsymbol{x})|^2) d\Omega + \gamma \int_{\Omega} (I_t + I_x u(\boldsymbol{x}) + I_y v(\boldsymbol{x}))^2 d\Omega \tag{5}$$

where, $u(\boldsymbol{x}) \in \mathbb{R}$ and $v(\boldsymbol{x}) \in \mathbb{R}$ is the 2D displacement of motion vector for an image pixel $(\boldsymbol{x}) \in \mathbb{R}$; Ω is the image domain. In case of scene flow extra dimensions of depth gets introduced. Further, it can be observed that a quadratic (L2) regulariizer (the first part of the integral), penalizes the inconsistent optical flow fields to obtain smoothness. Also, the second integral is used to impose optical flow constraint with quadratic cost; and parameter γ balances the data and smoothness terms. The variational methods calculates the optical flow field for all the pixels hence giving the dense optical flow. However, the Horn and Schunck [4] variational method leads to abnormal behaviour for discontinuous optical flow fields, for example at occlusion boundaries. This problem can be further understood through a scenario where the dis-occluded surface region showcases different flow from the flow of the surface which was previously occluding it. To overcome this problem, many papers have adopted a much simpler approach by replacing the L2 regularizer in the Horn and Schunck [4] with L1 regularizer. The modified equation (without dependence on x)looks like:

$$\min_{u(x),v(x)} \int_{\Omega} (|\nabla u| + |\nabla v)|) d\Omega + \gamma \int_{\Omega} (I_t + I_x u + I_y v) d\Omega \tag{6}$$

The new method though looks simpler with computational perspective presents differentiable discontinuities between smoothness and data terms. Therefore, a differentiable regularizer or penalty $\Psi_\varepsilon(x^2)$ is introduced in place of absolute function x. Different papers have used different penalty function, the most popular among them is Charbonnier penalty function defined as: $\Psi_\varepsilon(x^2) = \sqrt{x^2 + \varepsilon^2}$. The energy optimization by adopting the modified approach is known as total variation method.

This paper discusses variational flow methods for scene flow estimation through different camera settings such as monocular (single camera), binocular (stereo camera) and also multi-camera setting; and data sources which are namely divided into RGB-D and Grey-scale.

3 Inception of Scene Flow Constraints into Motion Vectors

As mentioned earlier scene flow estimation is the cumulative optimization of energy function related to disparity and optical flow and disparity change. This section establishes the concept of disparity estimation at time t; using the disparity we can then calculate the optical flow and disparity change between time t and t+1.

3.1 Disparity Estimation

The introduction of binocular camera settings brings the disparity between the pixels associated to left and right image sequences at time t. To estimate disparity, a normal stereo epipolar geometry is assumed which with the help of rectification steps acts in the direction to coincide the pixel rows y for left and right images. For scene capturing, the stereo camera with focal length f_x,f_y and final pixel coordinates at (x_0, y_0) is focused on an object with (X,Y,Z) in world frame of reference; and the coordinates corresponding to the pixel in left and right frames are (x,y) and (x+d,y). The relation in the matrix form can be represented as:

$$\begin{pmatrix} x \\ y \\ \mathrm{d} \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} Xf_x \\ Yf_y \\ -bf_x \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \\ 0 \end{pmatrix} \tag{7}$$

The stereo algorithm determines the disparity d, to recreate the 3D scene and it can be achieved through two way. First, is to match the small array of pixel of the left image to the right image, second approach suggests energy minimization towards a global consistent solution.

3.2 Stereo Motion Constraints

The formulation of stereo motion constraints for the image between two consecutive interval of time works analogous to the optical flow constraint. However, to accommodate the use of binocular camera settings, intensity function remains constant at time t and t+1 for both the left and right images. According to [9, 10] representation the scene flow field is defined by $(u, v, d')^T$, where u and v signifies the change from the initial position of the pixel i.e (x,y) at time t to $(x+u,y+v)$ at time t+1. It is also important to note that the other camera in the binocular camera setting accounts for the disparity estimation (in this case the right camera); therefore at time$t$, the right image pixel corresponds to $(x+d,y)$ which when moved further adds another disparity value d' to $x+d$ together with the change in position to achieve the new position of $(x+d+d',y+v)$. Also, it can be seen that a normal stereo epipolar geometry has been considered which constraints the movement of left and right image pixel in y direction at any given time. In addition, a cross term (12) for stereo matching between left view at time t and right image at time t+1 can also be considered. The scene flow constraints (OF:optical flow and ST:stereo) can be mathematically represented as:

$$E_{OF|left} : I(x,y,t)^l = I(x+u,y+v,t+1)^l \tag{8}$$

$$E_{OF|right} : I(x+d,y,t)^r = I(x+d+d'+u,y+v,t+1)^r \tag{9}$$

$$E_{st|t} : I(x,y,t)^l = I(x+d,y,t+1)^r \tag{10}$$

$$E_{st|t+1} : I(x+u,y+v,t+1)^l = I(x+d+d'+u,y+v,t+1)^r \tag{11}$$

$$I(x,y,t)^l = I(x+d+d'+u,y+v,t+1)^r \tag{12}$$

4 Formulation of Variational Framework for Scene Flow

The general taxonomy of the scene flow estimation showcases the formulation of the overall energy function from data terms and regularization (smoothness) term.

$$E(\vec{v}) = E_{Data}(\vec{v}) + \alpha E_{smooth}(\vec{v}) \tag{13}$$

where, E corresponds to overall Energy function, V encapsulates the change in position of image pixel i.e (u, v, d, d') between time t and t+1. Also E_{Data} and E_{smooth} represents Energy function associated to data source and smoothness respectively. The parameter α controls the amount of smoothness applied to a surface to showcase proper segmentation from other objects present in the scene.

While data terms are derived from the data sources which capture the scene and depends on factors like illumination, color, grey value, brightness constancy etc., the smoothness term deals with the variables in spatial scheme of things to achieve geometric consistency between different cameras. In this section, we will discuss and compare our primary literature [1, 3, 5, 6, 10] leading to the formulation of energy functions and optimization algorithm for the representation of Scene flow.

Over the years, the estimation of scene flow has been widely done using binocular (stereo) camera, multi-camera system, RGB-D camera (collects both color and depth information). However, it is the use of information which makes difference from one approach to another. Here, we will focus on the formulation of energy functions based on multi-view, binocular and RGB-D camera setting only.

Table 1: Typical methods for variational scene flow estimation

Literature	Year	Data Source	Description
Huguet et al. [5]	2007	Stereo camera	Basic framework for scene flow under binocular setting.
Wedel et al. [10]	2008	Stereo camera	Decoupling stereo and motion.
Basha et al. [1]	2013	Multiple camera	First to utilize point cloud representation.
Herbst et al. [3]	2013	RGB-D camera	Using color and depth
Jaimez et al. [6]	2015	RGB-D camera	Primal dual framework and regularization on 3D surface.
X.Xiang et al. [11]	2018	RGB-D camera	3D local rigidity assumption.

4.1 Multi-view stereopsis

The first work on scene flow estimation was done by Vedula and Sundar [8]. They used multiple cameras to compute multiple optical flow and later regenerating them to obtain a scene flow. However, arguably the first evidence of muti-view camera setting for variational approach has been depicted by Pons et al. [7]. This approach does not follow any conventional approach of scene flow estimation and has not compared its results with respect to other approaches; hence to achieve better cohesiveness, has been discarded from the survey. An another approach which illustrates the use of multiple view has been presented by Basha et al. [1]. It collects spatial (different view points) and temporal (3D motion) data for maintaining brightness consistency using N calibrated static cameras. The data term consists of three brightness constancy penalizers and obtained by summing them on image domain. "BC_m penalizes deviation from the brightness constancy assumption before and after 3D displacement; BC_{s_1} and BC_{s_2} penalize deviation from the brightness constancy assumption between the reference view and each of the other views at time t and t + 1, respectively" [1].

$$BC_m = \sum_{i=0}^{N-1} c_m^i \Psi(|I_i(p_i, t) - I_i(\hat{p}_i, t+1)|^2) \tag{14}$$

$$BC_{s_1} = \sum_{i=0}^{N-1} c_{s_1}^i \Psi(|I_0(p_0, t) - I_i(\hat{p}_i, t)|^2) \tag{15}$$

$$BC_{s_2} = \sum_{i=0}^{N-1} c_{s_2}^i \Psi(|I_0(p_0, t+1) - I_i(\hat{p}_i, t+1)|^2) \tag{16}$$

where p, $\hat{p} = (x, y)^T$ is the projection of 3D object for position P(X,Y,Z) displaced position $\hat{P} + \vec{V}$, '$\vec{V}$' being the 3D displacement (U, V, W) between time t and t+1 respectively; $\Psi(s^2) = \sqrt{s^2 + \varepsilon^2}$ ($\varepsilon = 0.0001$) represent penalty function discussed previously. The 3D point 'P' can be estimated from Eq(7).

The smoothness term S_m and S_s accounts for the change in the motion field and surface change w.r.t camera. The smoothness term is modelled using TV regularizer.

$$S_m(\vec{v}) = \Psi(|\nabla u(x, y)|^2 + |\nabla v(x, y)|^2 + |\nabla w(x, y)|^2) \tag{17}$$

$$S_s(Z) = \Psi(|\nabla Z(x, y)|^2 \tag{18}$$

Finally, the overall energy function can be written as:

$$E(Z, \vec{v}) = \int_\Omega [\underbrace{BC_m + BC_{s_1} + BC_{s_2}}_{Data} + \alpha \underbrace{(S_m + \mu S_s)}_{Smooth}] dxdy \tag{19}$$

and μ balances the motion and surface smoothness.

4.2 Binocular setting

Binocular camera (or stereo camera) setting makes things simpler as compared to multiple view approach since the stereo cameras gives a rectified epipolar constraint. Most stereo based methods like [5] use coupling of stereo motion constraints and disparity to estimate scene flow. Also, few methods such as [9, 10] has also considered decoupling of stereo motion and disparity and later coupling them together using semi-global matching (SGM). The decoupling in the later case allows for the separate optimization on each factors to achieve maximum benefits in terms of results and computational efficiency. Huguet and Devernay (2007)[5] adopted a joint framework approach in their paper [5]; further considering an additional gradient constancy assumption to achieve robustness against changes in illumination (inspired from Brox et al. [2] along with the grey value constancy assumption discussed in Eq(1). The difference in he intensity and illumination between two pixels can be illustrated as:

$$\Delta(I, x; I', y) = |I'(y) - I(x)|^2 + \gamma|\nabla I'(y) - \nabla I(x)|^2 \tag{20}$$

Therefore in [5] motion constraints for equation 8,9,10,11 thus become $\Delta E_{OF|left}$, $\Delta E_{OF|right}$, $\Delta E_{ST|t}$, $\Delta E_{ST|t+1}$.

$$E_{data} = \int_{\Omega} (\beta_{OF|left}\Delta E_{OF|left} + \beta_{OF|right}\Delta E_{OF|right} + \beta_{ST|t}\Delta E_{ST|t} + \beta_{ST|t+1}\Delta E_{ST|t+1})dxdy \tag{21}$$

where, β function accounts for occlusions and ranges between 1 and 0 for non occluded and fully occluded image respectively. Similarly, the data term used in Wedel et al. (2008, 2011) however consider only flow $E_{OF|left}$, and disparity change $E_{OF|t+1}$, also mentioned previously.

$$E_{data} = \int_{\Omega} (\Psi((E_{OF|left})^2) + c(x, y)\Psi((E_{OF|right})^2) + c(x, y)\Psi((E_{ST|t+1})^2))dxdy \tag{22}$$

where, c(x,y):0,1 is a indicator function for disparity (0 for no disparity). In addition, the smoothness terms in Huguet and Devernay (2007) [5] takes optical flow (u,v), disparity (d) and disparity flow (d-d') into account. While, Wedel et al. (2008) considers only optical flow and disparity change (d') only. The smoothness energy term for [5] and [10] in mention in Eq(23,24). It is interesting to note that the energy formulation for both [5] and [10] use Charbonniers penalty function $\Psi(x^2)$ separately to each data terms to respond to occlusion discrepancies, which leads to a robust energy with L1 minimization.

$$E_{smooth} = \int_{\Omega} \Psi(|\nabla u|^2 + |\nabla v|^2 + \lambda|\nabla(d - d')|^2 + \mu|\nabla d|^2)dxdy \tag{23}$$

$$E_{smooth} = \int_{\Omega} \Psi(\delta|\nabla u|^2 + \delta|\nabla v|^2 + \gamma|\nabla d'|^2)dxdy \tag{24}$$

4.3 RGB-D Camera

The data collection process through RGB-D cameras becomes simple as the color (or brightness) and depth information is easily collected. The scene flow estimation from RGB-D camera setting involves projection and back-projection of 3D scene and 2D image for consecutive time steps. Considering the translation $(X_1 : (X, Y, Z) \to X_2 : (X, Y, Z))$ in 3D space to be $\vec{V} = (U\ V\ W)$ and projecting on the image plane, the projection $\vec{v} = (u\ v)$ can be expressed as:

$$\vec{v} = (1/Z)(M - x_1^T S)(\vec{V}) \tag{25}$$

where $M = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}$ is the camera intrinsic parameter matrix; $S = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$ and $x_1 = \begin{pmatrix} x & y & 1 \end{pmatrix}$ is the point on image plane which is projected from X_1 on 3D scene. In this section we have discussed papers from Herbst et al. [3], Jaimez et al.[6], X.Xiang et al. [11]. While [3, 6] represents the scene in point cloud, [11] uses patch to establish energy terms. The data terms in [3, 6, 11] focuses on 3-D motion estimation from color (or brightness) and depth change constancy (or geometric constancy). The foundation of brightness constancy in [3, 6, 11] can be represented using Eq(1). However, Herbst et al. [3] has also considered gradient constancy while the other two has ignored it. The data terms corresponding to depth change constancy are similar in [3, 6, 11].

$$E_{DCC} : Z = (x + u, y + v) - Z(x, y) - W(x, y) \tag{26}$$

Moreover, energy data term corresponding to [3, 11] showcases Charbonnier penalty to normalize the L1 energy function to achieve convexity; while Jaimez et al. [6] uses exact L1 norm which makes its data term non convex and non linear with multiple local minima. Jaimez et al. [6] further refines data terms (color constancy and depth change constancy separately) using coarse-to-fine scheme on the linearized version of Eq(1) (i.e Eq(3) at each level of the pyramid to achieve a global minima.

For regularization, Herbst et al. [3] and X.Xiang et al. [11] uses anisotropic regularizer which takes into account the local gradient signals: image gradient, depth difference and surface orientation. Along with smoothing energy function, [3] also includes a flow magnitude penalty to balance pixel against meter. Contrary to [3, 11], [6] uses total variational (TV) regularizer.

5 Energy Optimization

The calculation of variations deals with the finding of minima for the energy function. This can be achieved by using Euler-Lagrange equation following $\nabla E = 0$. The number of equations obtained through Euler-Lagrange implementation depends on the components of flows (u, v, d, d') collaborating in the formulation of energy equation. The solution of partial differential equation for these unknown functions u, v, d, d' are highly coupled; and will lead to regeneration of scene flow.

It can be noted that the energy function is not convex because of the existing non-linearity in data terms and diffusion terms of Euler-Lagrange equation. To solve these complex non-linear algorithm, an incremental multi-resolution algorithm with two nested iterations are used which is majorly inspired from [2] are being implemented in [1, 3, 5, 9, 6, 11]. The first iteration computes the small increments of scene flow parameters (say $\delta u, \delta v, \delta d, \delta d'$ for stereo setting) through the successive over-relaxation (SOR) internally. This is followed by the external iteration corresponding to down sampling strategy (also known as coarse-to- fine). The parameters through coarse-to-fine undergo warping to achieve a new result $u_0 + \delta u, v_0 + \delta v, d_0 + \delta d, d'_0 + \delta d'$ for each level, finally converging to (u, v, d, d') for the full resolution.

Similarly, Jaimez et al. introduced a primal-dual algorithm into a coarse-to-fine framework to achieve RGB-D scene flow estimation. While energy formulation is based on linearization of data term; the energy function for smoothness consists of a convex TV regularizer in [6]. Therefore, the global energy function tends to show a dual nature which is the basis for the formulation of primal-dual algorithm. The iterative solver also favours the real-time implementation of scene flow because the primal and dual pixel-wise updates can be computed in parallel GPU.

6 Evaluation

Evaluation of optimization algorithm for the development for scene flow is achieved through testing the algorithm through standard data sets (real data set and synthetic data set) and later comparing the error with the ground truth. The popular data sets include Middleburry (Teddy, Cones and Venus), KITTI, EITSATS. These data sets are mainly used for stereo setting, which consists of optical flow ground truth and disparity ground truth information. Due to the lack of data set for RGB-D, disparity map of the available data set is converted to depth map. However, for RGB-D

The error types are: Mean square error (MSE) Root mean square error (RMSE), Average Angular Error (AAE), Normalized root mean square error (NRMSE), Average end-point error (EPE), Mean Average Error (MAE) . Evaluation also includes performance as run-time. A tabular view of optimization techniques and evaluation of literature is depicted below.

Table 2: Typical methods for variational scene flow estimation

Literature	Data set	Error Type	Error
Huguet et al. [5]	Sphere Teddy Cones	OF(RMSE, AAE)	0.69, 1.75 1.25, 0.55 1.11, 0.69
Wedel et al. [10]	EITSATS	OF(RMSE, AAE)	0.59, 4.13
Basha et al. [1]	Sphere	SF(NRMSE, AAE)	9.71, 3.39
Herbst et al. [3]	-	-	-
Jaimez et al. [6]	Sphere	(NRMS-V, AAE) PD-flow with TVg	0.068, 6.653
X.Xiang et al. [11]	Teddy Cones	(EPE, MSE, MAE)	0.49, 0.61, 0.81 0.41, 0.55, 0.76

7 Conclusion

This paper presents an outlook of variational scene flow estimation approaches through different data sources and discusses the methodology to deal with complex ill-posed problems. There is still opportunity to improve in the quest to evaluate data sets and analysis of computational performance.

References

1. Tali Basha, Yael Moses, and Nahum Kiryati. Multi-view scene flow estimation: A view centered variational approach. *International Journal of Computer Vision*, 101(1):6–21, Jan 2013.
2. Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. In *European conference on computer vision*, pages 25–36. Springer, 2004.
3. E. Herbst, X. Ren, and D. Fox. Rgb-d flow: Dense 3-d motion estimation using color and depth. In *2013 IEEE International Conference on Robotics and Automation*, pages 2276–2282, May 2013.
4. Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. *Artif. Intell.*, 17:185–203, 1981.
5. Frédéric Huguet and Frédéric Devernay. A variational method for scene flow estimation from stereo sequences. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–7. IEEE, 2007.
6. M. Jaimez, M. Souiai, J. Gonzalez-Jimenez, and D. Cremers. A primal-dual framework for real-time dense rgb-d scene flow. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 98–104, May 2015.
7. J. P. Pons, R. Keriven, and O. Faugeras. Modelling dynamic scenes by registering multi-view image sequences. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 822–827 vol. 2, June 2005.
8. Sundar Vedula, Simon Baker, Peter Rander, Robert Collins, and Takeo Kanade. Three-dimensional scene flow. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 722–729. IEEE, 1999.
9. Andreas Wedel and Daniel Cremers. *Optical Flow Estimation*, pages 5–34. Springer London, London, 2011.
10. Andreas Wedel, Clemens Rabe, Tobi Vaudrey, Thomas Brox, Uwe Franke, and Daniel Cremers. Efficient dense scene flow from sparse or dense stereo data. In *European conference on computer vision*, pages 739–751. Springer, 2008.
11. Xuezhi Xiang, Mingliang Zhai, Rongfang Zhang, Wangwang Xu, and Abdulmotaleb El Saddik. Scene flow estimation based on 3d local rigidity assumption and depth map driven anisotropic smoothness. *IEEE Access*, 2018.