U-RME: Underwater Refined Motion Estimation in hazy, cluttered and dynamic environments*

Shilpi Gupta[1], Prerana Mukherjee[2], Santanu Chaudhury[3], and Brejesh Lall[1]

[1] Indian Institute of Technology, Delhi, India
Shilpi.Gupta@dbst.iitd.ac.in, brejesh@ee.iitd.ac.in
[2] Indian Institute of Information Technology, Sri City, Andhra Pradesh, India
prerana.m@iiits.in
[3] Indian Institute of Technology, Jodhpur, India
santanuc@iitj.ac.in

Abstract. Optical Flow is a popular method of computer vision for motion estimation. In this paper, we present a refined optical flow estimation method. Central to our approach is exploiting contour information as most of the motion lies on the edges. Further, we have formulated it as sparse to dense motion estimation. Proposed method has been evaluated on challenging real life image sequences of KITTI and Fish4Knowledge database. Results demonstrate that method performs well in case of low contrast, highly cluttered background, dynamic background, occlusion and illumination change.

Keywords: Optical flow · Marine Ecosystem · Dense Correspondence · Holistically Nested Edge Detection.

1 Introduction

Optical flow estimation is the primary block in multitude of computer vision applications involving motion information such as object segmentation, object detection and object tracking etc. Despite many research strides in this area, accurate estimation of optical flow is still an open problem due to challenges of real world videos. Traditional optical flow based approaches rely upon the energy minimization functions [9, 16]. Due to the recent advancements in deep learning paradigms, optical flow estimation utilizing supervised deep learning based methods such as FlowNet2.0[10], PWC-Net[24] etc have outperformed traditional approaches. However, deep learning based supervised approaches are heavily dependent on the availability on ground truth data. In order to handle this, the networks are trained on synthetic data due to unavailability of ground truth on real scene sequences. There is high differences in the real and synthetic images therefore existing approaches on synthetic data do not generalize on

* We would like to thank the Ministry of Electronics and Information Technology Government of India and the Media Lab Asia for financial assistance to carry out research.

real images. Motivated by above reasons, researchers explored the direction of unsupervised Optical Flow estimation techniques like DD Flow [15] etc. But these approaches could not surpass the accuracy of supervised methods.

In recent years, there has been growing research interest to study the behavior of marine species due to its potential applications. Due to the complexity of the underwater environment and the limitations of human divers, underwater scenario is mainly explored by submarines, remotely operated vehicles (ROVs) and autonomous underwater vehicles (AUVs). Marine video surveillance is highly preferred over photography by divers or net-casting methods, since it provides a large amount of continuous data without effecting the fish behaviour. Detecting objects in underwater video is highly challenging task. The challenge posed is due to poor quality of vision data due to high turbidity, appearance variation with depth, light attenuation, suspended particles in the medium and dynamic environments due to movement in water particles and coral reefs as shown in Fig. 1. Objects are highly deformable, identical or very similar in appearance.

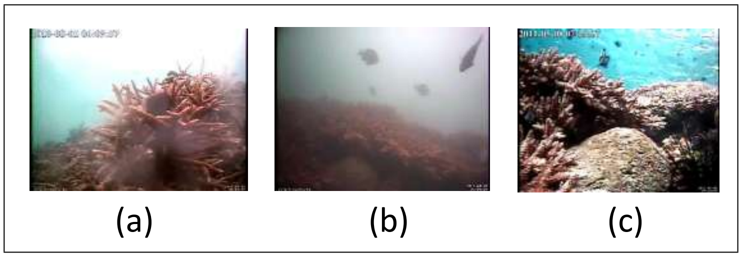

Fig. 1. Challenges in marine environment (a) Camouflage (b)Hazy environment (c) Clutter

In this paper, the objective is to find motion estimation of fishes in underwater videos using image processing and computer vision concepts. hence, we propose Underwater Refined Motion Estimation (U-RME) in hazy, cluttered and dynamic environments. Recent efforts towards this problem assume availability of annotated datasets [10, 19, 24]. Most popular approach for object detection is to train a model by supervised learning. To achieve desirable accuracy, these methods require a large amount of annotated data, which is highly time consuming and requires human expertise to recognize fishes in cluttered background and high camouflage based marine conditions. In practical scenario, the datasets do not span over all possible classes of fish, limiting their effectiveness in analyzing underwater ecosystem, tracking fish population etc. while at the same time dampening utilization of such techniques in exploratory research for new applications of underwater imaging. Moreover, it is also important to note that, a large amount of images available over the web is not part of standard datasets. The pre-trained object detectors process individual frames of videos while completely ignoring the temporal information. Human visual system does not receive static images, it receives continuous video streams. Appearance cues provide limited information when videos are recorded in low light and hazy conditions. In such cases, motion is an important factor to get significant information about moving object in videos. Gestalt principle also states that “grouping forms the basis of human perception”[11]. Points moving together can be grouped together and they often belong to the same object. Motion based grouping appears early in the stages of visual perception than static grouping.

To this end, we introduce a refined end-to-end motion estimation technique. The key contributions in this paper can be summarized as,

- We have updated the pipeline by including a pre-processing block to handle highly illumination varying and hazy environment. Holistically nested edge detection and a median filtering based objective function is adopted to get better motion boundaries in cluttered and dynamic environments.
- To the best of our knowledge, we are the first to utilize dense optical flow for detecting motion of fish in real-life dataset of fish4knowledge.
- We have evaluated the proposed approach with several flow based techniques over static and dynamic environments.

The paper is organized as follows. In Sec. 2, we provide the related work on optical flow and motion based object detection in marine environment. In Sec. 3, we discuss the proposed methodology. In Sec. 4, we give the experimental results and analysis followed by conclusion in Sec. 5.

2 Related Work

Optical flow estimation has been studied as an important topic in computer vision for long. Research work done in this area can be broadly classified into various categories. Traditional methods are based on variational approaches [9, 16, 1]. In such methods, the aim is to optimize the function of brightness constancy and spatial smoothness. These methods are suitable for small displacements but fail in case of large displacement flows. Coarse to fine approaches have been proposed to tackle large displacements [3]. Later approaches integrate feature matching to tackle this issue. Specially, they find sparse feature correspondences to initialize flow estimation and further refine it in a pyramidal coarse-to-fine manner. SIFT FLOW performs dense matching between the Scale Invariant Feature Trasform (SIFT) feature matching between two images [14]. The seminal work of EpicFlow [21] interpolates dense flow from sparse matches and has been widely used for scene flow estimation in dynamic environments.

Recently, the success of deep learning has inspired researchers to solve flow estimation problem as optical flow learning problem. The pioneering work is FlowNet 2.0 [10], which is based on supervised learning and generates a dense optical flow map with two consecutive frame and a trained model. SpyNet introduces a spatial pyramid network in order to handle large displacements [19]. Recently, PWC-Net [24] has been proposed to warp extracted features learned by CNNs instead of warping images over different scales. Although these approaches show promising performance but the problem is that these methods require a large amount of labeled training data, which is particularly difficult to obtain for optical flow particularly in case of underwater scenarios. Synthetic dataset is used for training such models, while real images are very different from synthetic images. Due to this gap, these models do not always perform well with real data.

Another promising direction is to develop unsupervised learning approaches [20]). The idea is to warp the target image according to the predicted flow, the difference between the reference image and the warped image is optimized using a photometric loss. Most recent work of DDFlow generate annotations on unlabeled data using a model trained with a classical optical flow energy function, and then retrain the model using those extra generated annotations [15]. There is still a large gap if we compare the performance with supervised methods.

Due to these limitations of supervised and unsupervised deep learning based approaches, in this paper we propose an unsupervised approach to estimate flow in hazy, cluttered and dynamic environment of marine videos. The closest work to ours is EpicFlow[21]. However, we introduce non-local median filtering [8] in the optimization function. This allows the noise suppression and introduces brightness constancy term in the energy optimization function to mitigate the illumination variations. Since, there is light dispersion in the medium leading to shadow effects we also perform preprocessing measures.

3 Proposed Methodology

In this section, we detail the components of the proposed methodology. Fig. 3 shows the pipeline of the proposed methodology.

3.1 Preprocessing

The primary focus is to get accurate optical flow estimation for complex environment where videos are of poor resolution quality, hazy in nature with uneven and rapidly changing illumination changes. There are two major causes of haze in surveillance videos: (i) Fog or Smog in aerial videos and (ii) Turbidity of water, light scattering in water particle in underwater videos. In order to handle this, we require the pre-processing step in such videos. We can improve the quality of images either by image enhancement techniques or image restoration techniques. For image enhancement, we used DehazeNet [4]. This is a CNN based deep architecture for haze removal. Network takes a hazy image as an input, and output is a haze free image. This method outperforms the existing haze removal techniques which are based on many prior assumptions. Another approach we have exploited is inspired by underwater particle physics [6]. In this paper [6], authors proposed Simultaneous localization and mapping (SLAM) to do object detection. In this work, Light Scattering Model produces better results when compared with DehazeNet in case of marine videos. The qualitative comparative results obtained by both methods are shown in Fig. 3. As can be seen, light scattering is more in case of DehazeNet resulting in objects (fishes in this case) not being clearly visible.

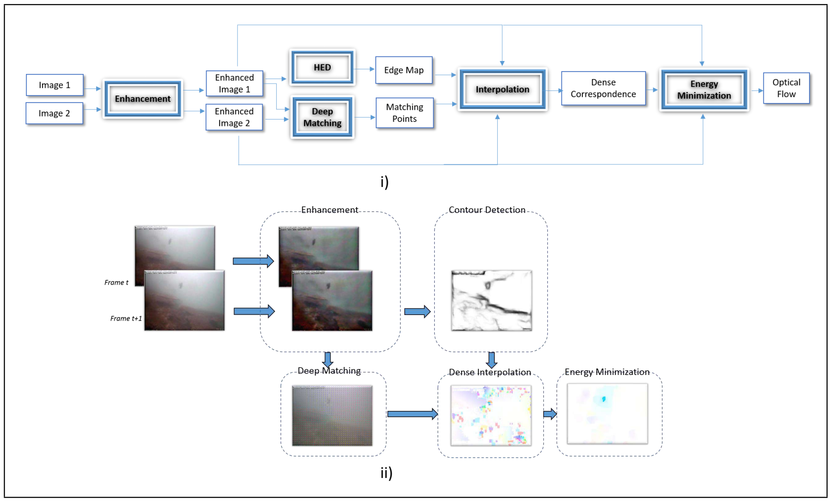

Fig. 2. i) Flowchart of proposed method. ii) Pipeline of the proposed flow method. Given two frames of the video, we first perform enhancement using Light scattering technique as proposed in [6]. Next, we use DeepMatching technique to get correspondence between two frames[22] and contours of the first frame (t) is computed using Holistic Nested Edges (HED) [26]. Finally, we combine these two cues to get dense interpolation image which enables computation of dense correspondence field. this is used to initialize the energy minimization framework of optical flow generation.

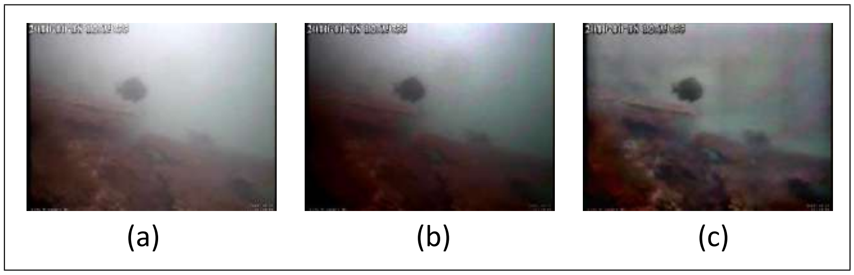

Fig. 3. (a) Original Image. Enhanced image by (b) DehazeNet[4] and (c) Light Scattering Model[6].

3.2 Contour Detection

Motion discontinuity mainly appears on edges[21]. The proposed method is heavily dependent on the conservation of motion boundaries. Conventional edge detection methods like Canny [5] rely on local intensity change. This results into a lot of spurious edges being generated as shown in Fig. 4(b). Structured edge detection (SED) utilizes random forest ensemble which spans over all the oriented edge combinations[7]. However, such hand crafted features are heavily dependent on the nature of images. To preserve the optimal set of object boundaries, we have resorted to deep learning based Holistically Nested Edge Detector (HED)[26]. In this approach, a deeply supervised fully convolutional neural network is used for multi-scale and multi-level edge feature learning. Fig. 4 shows the comparative qualitative results of aforementioned methods. In case of hazy images, the object boundaries are not properly detected by SED as compared to HED. All the results have been generated before applying restoration techniques for fair evaluation and demonstrate the efficacy of HED over other compared edge detection methods even in hazy environments.

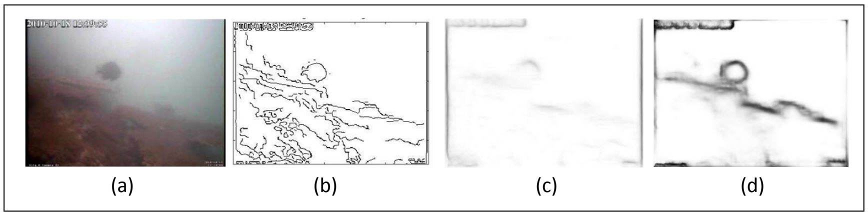

Fig. 4. (a) Original Image. Edges detected by (b) Canny[5], (c) SED [7] and (d) HED[26]

3.3 Energy minimization

Energy minimization in a coarse to fine manner is a popular technique to obtain dense flow field. However, this approach suffers with a drawback of error propagation. Error at coarser level can propagate across scales. Obtaining the initial set of matches is quite costly in this manner. This can be estimated directly utilizing state of art matching methods. We utilize DeepMatch [22] to obtain matching points between two consecutive frames. It works on a deep convolutional architecture designed for matching images. Fishes and humans are highly deformable objects. DeepMatch can efficiently handle such non-rigid deformation and determine dense correspondences between images. Next step is to get dense matching points from the sparse list of matches obtained by DeepMatch. We have adopted edge aware sparse-to-dense interpolation method of Epic Flow. Nadraya-Watson estimation [25] method is utilized for interpolation. This method uses Geodesic Distance (GD) instead of Euclidean distance and cost map is obtained by edge detector. In the proposed method, cost is estimated by Holistically nested edge detection [26]. Since, GD estimation among pixels is time consuming thus authors have proposed a graph based approximation for this in [21]. It provides a smart heuristic for initialization of Optical Flow. Further, we perform variational refinement of dense optical flow map. We minimize the energy term defined by data term (E_D), smoothness term (E_S), coupling (E_C) and median term (E_{med}). Dequin Sun et al [23] have justified that median filtering can significantly improve the result of optical flow field. They have incorporated the median filtering term in classical objective function and non-local coupling term to pertaining to the effect of data term. Flow updates are calculated by successive over relaxation method [27].

Given an image pair F_1 and F_2, such that $F_1, F_2 \epsilon R^{HXWX3}$ representing consecutive frames at time instants t and $t+1$. The goal is to estimate the optical flow $\mathbf{V} = (u, v), \mathbf{V} \epsilon R^{HXWX2}$. The energy is defined as the weighted sum of data term (E_D), smoothness term (E_S) and non-local term (E_{NL}). The non-local term consists of the coupling term (E_C) and median term (E_{med}) proposed by Li and Osher [13]. It can be calculated as,

$$E(\mathbf{u}, \mathbf{v}) = \rho_D E_D + \lambda_1 \rho_s E_S + \lambda_2 E_C + \lambda_3 E_{med} \tag{1}$$

$$E_D = \sum_{i,j} (F_1(i,j) - F_2(i + u_{i,j}, j + v_{i,j})) \tag{2}$$

$$E_S = \sum_{i,j} ((u_{i,j} - u_{i+1,j}) + (u_{i,j} - u_{i,j+1}) + (v_{i,j} - v_{i+1,j}) + (v_{i,j} - v_{i,j+1})) \tag{3}$$

$$E_C = (\|\mathbf{u} - \hat{\mathbf{u}}\|^2 + \|\mathbf{v} - \hat{\mathbf{v}}\|^2) \tag{4}$$

$$E_{med} = \sum_{i,j} \sum_{(i',j') \epsilon N_{i,j}} (\|\hat{u}_{i,j} - \hat{u}_{i',j'}\| + \|\hat{v}_{i,j} - \hat{v}_{i',j'}\|) \tag{5}$$

Eq. 2 indicates the data term and ρ_D is the data penalty function. due to color constancy, we do not need to consider the change in RGB values between two images in the data term. λ_1, λ_2 and λ_3 are the regularization parameters. Eq. 3 indicates the smoothness term and ρ_S is the smoothness/spatial penalty function. Charbonnier penalty function is used to penalize data and smoothness term . Eq. 4 and eq. Eq. 5 denote the coupling and median filtering term respectively. $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ denote the auxiliary flow field. $N_{i,j}$ denote the set of neighbors of pixel (i, j) [23].

4 Experimental results and analysis

In this section, we present the results obtained in evaluation of the proposed method. First, we describe the implementation and dataset details. Later, we discuss the empirical evaluation and analyze the test results.

4.1 Experimental setup and parameter settings

All evaluations has been carried out on a machine with 32 GB RAM, Intel Core i7Xeon 1650 processor and Ubuntu 16 operating system. MATLAB 2017b was used as the programming platform. The values used for the different regularization parameters are $\lambda_1 = 1, \lambda_2 = 0.10$ and $\lambda_3 = 1$. Neighbourhood pixels $N_{i,j} = 5 * 5$ window.

4.2 Dataset

In the past, most of the research work has focused on flow estimation on synthetic images and other high quality datasets. While working on the proposed approach, our motive was to work with challenging images of real life underwater scenario. As per the best of our knowledge there does not exist any underwater dataset with ground truth. We have evaluated our method on complex and challenging image sequence of Fish4knowledge [12] dataset. This dataset does not have ground truth of optical flow. Due to lack of ground truth we will present qualitative results on this dataset. For quantitative evaluation, we have also tested our proposed method on popular optical flow dataset of KITTI [17].

KITTI Dataset: This dataset has been extensively used by researchers for quantitative evaluation of various optical flow methods. KITTI 2015 dataset consists of 200 training and 200 test scenes with moving camera and moving objects.

The Fish4knowledge Dataset: We have evaluated our results on this data set because this dataset comprises of real life data of ocean having challenges like moving background of coral reefs and movement in water, highly varying illumination due to light scattering in water particles, very low quality of videos, crowded scenes due to randomly moving fishes and camouflage. Videos are 10 minutes long with a resolution of 320x240 and a 24-bit color depth at a frame rate of 5 fps. Limitation of this data is that it does not have any ground truth of optical flow. For fair evaluation, we demonstrate qualitative results obtained on this data.

4.3 Quantitative Analysis

Most popular performance measure metrics for optical flow is the Angular Error [2] and Endpoint Error [18]. We have compared results with three kind of flow methods. First one is Feature based flow calculation methods like LDOF[3], SIFTFLOW[14]. Second one is sparse to dense method Epic Flow [21] and last one is Unsupervised deep learning based method like DD Flow [15]. DD Flow has already outperformed the other unsupervised flow methods. Focus of our proposed method is Underwater scenario. In most of the cases Angular Error is less than other flow computation methods, while End-point Error has marginal or no improvement

Table 1. Comparison between Average Angular Error(AAE) and Average Endpoint Error (AEE) for different optical flow methods on KITTI 2015 dataset

Methods	**Proposed**	**DD Flow**	**EPIC FLow**	**SIFT FLow**	**LDOF**
Average Angular Error (AAE)	5.411669e+01	5.078307e+01	5.397299e+01	6.331552e+01	5.200370e+01
Average End Point Error(AEPE)	1.870108e+01	1.904110e+01	1.816318e+01	4.396911e-01	1.590453e+01

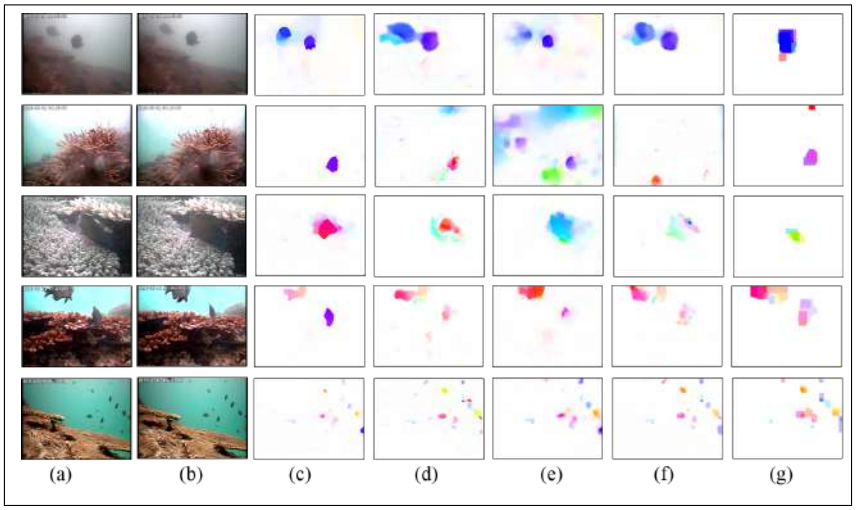

Fig. 5. Success cases: (a) Frame t (b) Frame t+1 (c) Proposed (d) DDFlow (e) EF (f) LDOF (g) SIFTFlow.

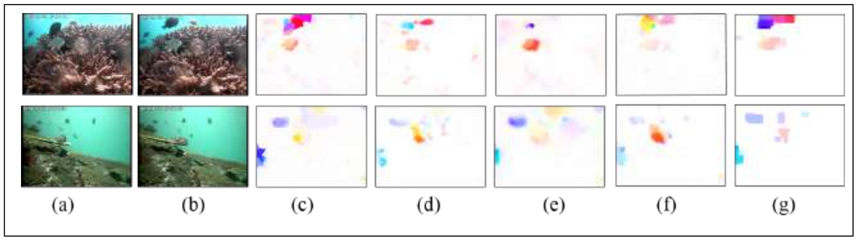

Fig. 6. Failure cases: (a) Frame t (b) Frame t+1 (c) Proposed (d) DDFlow (e) EF (f) LDOF (g) SIFTFlow.

4.4 Qualitative Analysis

In the absence of ground truth information, we present visual results to compare our method with competing methods on Fish4knowledge dataset. In Fig. 5 results of proposed technique is compared with existing methods, i.e., DDFlow [15], EPIC Flow [21], LDOF [3], and SIFT Flow [14].

These are the cases where our refined technique delineates the moving object accurately. We want to highlight that our approach is more robust to challenges such as occlusions, cluttered background, large illumination change, Low contrast, high water turbidity, crowded and fast moving and deformable objects like fish. In Fig. 5, row 1 shows the case of low contrast and high water turbidity. Middle rows presents the result under cluttered back-ground, deformable object, occluded fish and illumination change. Crowded and small fishes scenario is in the last row. Fig. 6 has results of a few cases where our proposed method could not perform well. When we have analysed the intermediate results of these frames we found that there is need to refine the interpolation method to get more accurate results.

5 Conclusion

We proposed a refined optical flow estimation for challenging underwater video sequences. Motion information of objects is crucial for such low quality videos. We demonstrate how to effectively exploit the edge information of image to capture motion information. We have shown significant improvement for underwater videos and comparative results for dynamic environments on road sequences. The proposed flow estimation technique can be further extended to segment and track the objects in hazy, cluttered and dynamic environments.

References

1. Anandan, P.: A computational framework and an algorithm for the measurement of visual motion. International Journal of Computer Vision **2**(3), 283–310 (1989)
2. Barron, J.L., Fleet, D.J., Beauchemin, S.S.: Performance of optical flow techniques. International journal of computer vision **12**(1), 43–77 (1994)
3. Brox, T., Malik, J.: Large displacement optical flow: descriptor matching in variational motion estimation. IEEE transactions on pattern analysis and machine intelligence **33**(3), 500–513 (2010)
4. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing **25**(11), 5187–5198 (2016)
5. Canny, J.: A computational approach to edge detection. In: Readings in computer vision, pp. 184–203. Elsevier (1987)
6. Cho, Y., Kim, A.: Visibility enhancement for underwater visual slam based on underwater light scattering model. In: 2017 IEEE International Conference on Robotics and Automation (ICRA). pp. 710–717. IEEE (2017)
7. Dollár, P., Zitnick, C.L.: Structured forests for fast edge detection. In: Proceedings of the IEEE international conference on computer vision. pp. 1841–1848 (2013)

8. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. Multiscale Modeling & Simulation **7**(3), 1005–1028 (2008)
9. Horn, B.K., Schunck, B.G.: Determining optical flow. Artificial intelligence **17**(1-3), 185–203 (1981)
10. Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T.: Flownet 2.0: Evolution of optical flow estimation with deep networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2462–2470 (2017)
11. Johansson, G.: Visual perception of biological motion and a model for its analysis. Perception & psychophysics **14**(2), 201–211 (1973)
12. Kavasidis, I., Palazzo, S., Di Salvo, R., Giordano, D., Spampinato, C.: A semi-automatic tool for detection and tracking ground truth generation in videos. In: Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications. p. 6. ACM (2012)
13. Li, Y., Osher, S.: A new median formula with applications to pde based denoising. Communications in Mathematical Sciences **7**(3), 741–753 (2009)
14. Liu, C., Yuen, J., Torralba, A.: Sift flow: Dense correspondence across scenes and its applications. IEEE transactions on pattern analysis and machine intelligence **33**(5), 978–994 (2010)
15. Liu, P., King, I., Lyu, M.R., Xu, J.: Ddflow: Learning optical flow with unlabeled data distillation. arXiv preprint arXiv:1902.09145 (2019)
16. Lucas, B.D., Kanade, T., et al.: An iterative image registration technique with an application to stereo vision (1981)
17. Menze, M., Geiger, A.: Object scene flow for autonomous vehicles. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
18. Otte, M., Nagel, H.H.: Optical flow estimation: advances and comparisons. In: European conference on computer vision. pp. 49–60. Springer (1994)
19. Ranjan, A., Black, M.J.: Optical flow estimation using a spatial pyramid network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4161–4170 (2017)
20. Ren, Z., Yan, J., Ni, B., Liu, B., Yang, X., Zha, H.: Unsupervised deep learning for optical flow estimation. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)
21. Revaud, J., Weinzaepfel, P., Harchaoui, Z., Schmid, C.: Epicflow: Edge-preserving interpolation of correspondences for optical flow. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1164–1172 (2015)
22. Revaud, J., Weinzaepfel, P., Harchaoui, Z., Schmid, C.: Deepmatching: Hierarchical deformable dense matching. International Journal of Computer Vision **120**(3), 300–323 (2016)
23. Sun, D., Roth, S., Black, M.J.: Secrets of optical flow estimation and their principles. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 2432–2439. IEEE (Jun 2010)
24. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8934–8943 (2018)
25. Wasserman, L.: All of statistics: a concise course in statistical inference. Springer Science & Business Media (2013)
26. Xie, S., Tu, Z.: Holistically-nested edge detection. In: Proceedings of the IEEE international conference on computer vision. pp. 1395–1403 (2015)
27. Young, D.M.: Iterative solution of large linear systems. Elsevier (2014)