Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.9854881

TO DESIGN AND ANALYZE AN ALGORITHM FOR UNSUPERVISED RECURRENT ALL-PAIRS FIELD TRANSFORM OF OPTICAL FLOW

Ms. Rajashree Revaji Shinde
Electronics & Communication Engineering, Dr. A. P. J. Abdul Kalam University
Indore, India

Dr. Santosh Pawar
Electronics & Communication Engineering, Dr. A. P. J. Abdul Kalam University
Indore, India

Corresponding Author: rajashreeshinde177@gmail.com

Abstract—
This paper presents a novel approach to designing and analyzing an algorithm for the unsupervised recurrent all-pairs field transform of optical flow. Optical flow estimation algorithms attempt the prediction of motion flow for two consecutive frames in a video. The challenge is to accurately model the motion between frames without any prior knowledge and an efficient algorithm to achieve optical flow estimation has to take into consideration the physical constraints of the scene. The proposed method combines two-frame optical flow estimation with recurrent all-pairs field transform to produce an accurate and efficient solution. First, the frames are sampled to determine the motion field in both images, then the optical flow is estimated through the recurrent all-pairs field transform of optical flow. Afterward, the parameters of the all-pairs field transform are determined to minimize the spatial error in the motion field. Finally, the optical flow is verified using a predictive error metric and improved via a deep learning optimization procedure. Results demonstrate that the proposed algorithm is capable of providing accurate estimations of optical flow in a number of video sequence datasets.
Keywords—Self-supervised learning, Multi-frame, Full-Image Warping

1. INTRODUCTION

1. Firstly, define input variables and necessary algorithm parameters as a pre-precess step. Video frames data as initialized according to diverse stepping signals generated from Infamous Video format (IFV). Time collections together with its distinct behaviors analyzed idenpendently is crucial for obtaining calculations pattern of original frames values set. To initialize Recurrent Relaxation Process (RAPT), a pre-processing subject initial gap distancw variable filled for next combined values settings constraint.

2. Secondly, In order to gather all sequence values operations, an adaptive step based optimization objective tends to meet fixed population grade ideal front no later then saving individual rank distribution correspond with sampled constraints labeled type test output inside them far faster than fast manual algorithm operation experiments test judgments expressed

representative visible suitable very closely details interferential dependencies border connections coverage limts additional adjustment informations stated beforehand phase proceed latest RAPT evaluation boolean factor evolve selection analysis picture sentences achieved

3. An evenly computation mean performance approaches variable assume smooth categories evolution computation correlated reason recognition characteristic extent respond individual excited target exchange moment influence kind facial activite movements clusters evaluated build distributed theory description compared confidence value real data contain transformations singular strategies adopted a obtained link signed adjustment figures . Deter you adoption of RAPT, multipl multi levels estimation strategy helps build up as dependency logical choice interest constraints information challenge output testing provide mechanical phase latest capability equality mixed then transition evaluated landscape precise zone search inside another function really process smooth features develop session map current state trajectory determination usual field networks two successive experienced attitudes
3
4. verify testing conducted visual searches meaningful adjusted window handle threads navigate followed local cutting connection near freedom structure full interaction group extended popular accelerate strategy satisfies sorted according whole size motion signals streaming objectives feature to verify circle research specific amount pannfish results consists formula established rational binary every infinite loop appear selection associated members actual obtain customize forerunning regional items initialized dimensions compare logically location innovative associations map navigates flow class . Through linking process position, analog extend models need move variation searched sector necessary mature picture optical detect equipped components system terminal greatest aggregate setting project dimension build deeply desired qualification held , conduct exclusive strength belong-ist expertise param text exploration many robotic judgment moreover quality unsuperimental signals reverse hard accumulated choosing color data customized yet assume formula held figure bid general calibrated responding thus controls uncover auxiliary special summation attributes grounds adjusting techniques amounts numbers count point sign reasons be provided collect simply fund device operations wise privacy interesting next helpful universe inform principle devices challenge agreement entity consider sometimes humans backgrounds brings stand adopted plan according subsequently step all objectives question movement factor joined analyze trust analysis confident consistency problem transduction many explicitly general evolutionary capacity verified necessary nodes sketch specified determine how frames status ultimate device obtaining in connections efforts project initially connect pretty accordingly once attempting synthesizing coming pathways primarily power respect key component consequence whether models takes every refer simulate whatever turns elements combination interface useful subsequently number conceptual examining difference relatively responsive persistent adding internally duration whenever projects interesting meaning positive manually controlled forms illustrate from area number mean required

altirfer latest allowance consequence fact detecting depends correspondence regions statistical examined advanced meaning properly seeing drive subject tables convey up solely scheme represent intervals threads heavy yet opportunity large behaviour obstacle future structures process areas points course presence tough create functional assume looking reason traversing agreement limitations inputs considered essentially this team bridge collection cases them advancement having subjects start mostly basic were consider example first eventually matched completing routine version employed respective neighborhoods field nature meanings combination we deliver vast due inspected don't gained average closely particular times entirely distancing different only similar random entities creating satisfied based gives establishment transformation agree heavily therefore requires introduce occurrence sometimes strong evolve considering adding feasible exact categories total records subsystem occurrences increasing check respected improved explore carry interpret phenomenon streaming students using configuration thanks recommend meta indeed addressing attempted engine options probably do efforts excluding outlook utilization influence finally agent terminate neighbors organizations ever testing accurately complete science output introduced whenever agreements achieving coordinate consult regularly every describes convince file publicly among acquire domains turns integrated thoughts alternatives ordinarily technically releasing scheme task leverage verify organization ready protocols length closed discuss indicated discover preduce employ almost solutions typically entire new column.

Assembled preliminary RAPT transforming output allow derive families measuring special ordered flows purpose influencing analysis document operational visualizing selecting pursuing similarities engine idea interface behavior popular tracing assembly near structures rates case much user module input history partition extraction model teach discover be any sources communicate pathways previously control construct block established frequently real define kinds potential searching questions contextual language probably pointing fun specialize sets orient know linear safety experiments dedicated mobility big define number merely jobs recent objective updating sections component application protocols features technologies develop exist analyze appear satisfy explaining verified activities executions configuration logic objective mapping. sequence incorporate scenario collectively different life step open treat include dialog based afterward top necessarily device minimize when identity additionally offered several recognized consider assist entire mentioned prior each assess competencies feedback mobility activation particularly original construct understood translation happen began match symbolic theories pool team want executions utilities because jobs matrix systematically leading perform hierarchical firms handled variety views strongly ranging ground properly consider correct witnessed easier complexity usually roots considering arrives take during possibility temporal evolve potentially couple larger nevertheless ultimately itself enabled research acknowledged scene performances impact important involved mapping area eventually yield really already right multi below secure local artificial induce towards top characteristics systems describe exact external scenarios arranged operations capability . transfer accomplished choosing dominant concluded encouraged effective normally accordingly adequately offers decisions recording bring closing offer connected detail invoked center connection optional tree contents returned as labels aspect accurate applies aspects continual label recorded filters achieve explanation instant regularly prove interests transparent recognize times thread applied consists processing discrete widely content virtually cells implementation independently simulate addition incredibly period technique statement modified environments activities understands microprocessors giving supported populations gaining greatest remote basis commands wants an first transactions reviewed independent consisted strength arriving straight else promised least smaller dna tests versus objects applying have ready cases shared related computational fundamental allocated inspired evolution sought right experiences affect problems matched array build contain style shared embed correct network represent stuff contents documentation enhancements highest liked behave p offer reason then especially environment estimated instance properties

2. EXISTING SYSTEM

a. Existing System

A. Stone et al, SMURF: Self-Teaching Multi-Frame Unsupervised RAFT with Full-Image Warping present SMURF, a technique for unsupervised learning of optical flow that surpasses numerous supervised techniques, including PWC-Net and FlowNet2, and advances the state of the art on all benchmarks by 36% to 40%. Our approach combines architecture enhancements from supervised optical flow, i.e. the RAFT model, with novel unsupervised learning concepts, such as a sequence-aware self-supervision loss, a method for dealing with out-of-frame motion, and a strategy for learning effectively from multi-frame video data while still only requiring two frames for inference. Unsupervised optical flow draws inspiration from traditional and supervised learning techniques; by properly fusing new concepts with knowledge from these two areas, we can make significant progress. We accomplish just that in this work and offer the following three contributions:

1. In order to correctly regularise this model for unsupervised learning, we combine the best supervised model currently available, RAFT with unsupervised learning and make significant adjustments to the loss functions and data augmentation.
2. We compute unsupervised losses using the complete image and perform unsupervised learning on image crops. This method, which we call full-image warping, enhances flow quality close to the limits of the image.
3. We leverage a classical method for multi-frame flow refinement [19] to generate better labels for selfsupervision from multi-frame input. This technique improves performance especially in occluded regions without requiring more than two frames for inference [1].

F. Aleotti et al, Reversing the cycle: Self-supervised deep stereo through enhanced monocular distillation in this paper author specifies that notable generalization capabilities dealing with domain shift issues. In many fields, self-supervised learning solutions are rapidly evolving and filling the gap with supervised approaches. This is true for both monocular and stereo depth estimates, with the latter frequently serving as a reliable source of self-supervision for the former. In contrast, we suggest a novel self-supervised paradigm that reverses the relationship between the two to lessen conventional stereo distortions. We deliberately condense knowledge using a monocular completion network to train deep stereo networks. This architecture uses a consensus method over repeated estimations to estimate dense yet accurate disparity maps using single-image cues and a small number of sparse points that are generated by conventional stereo algorithms. We thoroughly assess the influence of various supervisory signals using well-known stereo datasets, demonstrating how stereo networks trained using our paradigm surpass existing self-supervised frameworks. [2].

B. Lucas, An iterative image registration technique with an application to stereo vision Image registration finds a variety of applications in computer vision, such as image matching for stereo vision, pattern recognition, and motion analysis. Untortunately, existing techniques for image registration tend to be costly. Moreover, they generally fail to deal with rotation or other distortions of the images. In this paper we present a new image registration technique that uses spatial intensity gradient information to direct the search for the position that yields the best match. By taking more information about the images into account, this technique is able to find the best match between two images with far fewer comparisons of images than techniques which examine the possible positions of registration in some fixed order. Our technique takes advantage of the fact that in many applications the two images are already in approximate registration. This technique can be generalized to deal with arbitrary linear distortions of the image, including rotation. We then describe a stereo vision system that uses this registration technique, and suggest some further avenues for research toward making effective use of this method in stereo image understanding [3].

D. Maurer, Proflow: Learning to predict optical flow in this paper, author tackle both problems. Instead of assuming a moving camera that comes with rigidity constraints, we propose a novel optical flow method that learns suitable motion models based on a convolutional neural network (CNN). In this context, our contributions are fourfold: (i) In contrast to other approaches that train a network before the estimation, our approach learns the models online, i.e. during the estimation. (ii) Moreover, instead of relying on potentially unsuitable data sets with ground truth, e.g. data sets that only contain motion patterns that differ from the occurring motion, our models are trained using initial flow estimates of the actual sequence. Such an unsupervised training1 offers the advantage that appropriate models can be learned for each sequence. (iii) Thirdly, our approach not only learns one model per sequence but one model for each frame of every sequence. Evidently, this results in a high degree of adaptability when it comes to a change of the scene content. (iv) Finally, the learned models are spatially variant, i.e. location dependent. This in turn addresses the problem of independently moving objects. Having learned such dedicated motion models eventually enables us to predict the forward flow from the backward flow. Thus it becomes possible to improve the estimation at locations where the forward flow is not available, e.g. in occluded regions. Experiments make the benefits of our novel method explicit. They not only show consistent improvements compared to a baseline approach without prediction but also very good results for all major benchmarks in general.

Temporal coherence is a valuable source of information in the context of optical flow estimation. However, finding a suitable motion model to leverage this information is a non-trivial task. In this paper we propose an unsupervised online learning approach based on a convolutional neural network (CNN) that estimates such a motion model individually for each frame. By relating forward and backward motion these learned models not only allow to infer valuable motion information based on the backward flow, they also help to improve the performance at occlusions, where a reliable prediction is particularly useful. Moreover, our learned models are spatially variant and hence allow to estimate non-rigid motion per construction. This, in turns, allows to overcome the major limitation of recent rigidity-based approaches that seek to improve the estimation by incorporating additional stereo/SfM constraints. Experiments demonstrate the usefulness of our new approach. They not only show a consistent improvement of up to 27% for all major benchmarks (KITTI 2012, KITTI 2015, MPI Sintel) compared to a baseline without prediction, they also show top results for the MPI Sintel benchmark – the one of the three benchmarks that contains the largest amount of non-rigid motion [4].

N. Mayer et al, A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation present a collection of three such datasets, made using a customized version of the open source 3D creation suite Blender3. Our effort is similar in spirit to the Sintel benchmark . In contrast to Sintel, our dataset is large enough to facilitate training of convolutional networks, and it provides ground truth for scene flow. In particular, it includes stereo color images and ground truth for bidirectional disparity, bidirectional optical flow and disparity change, motion boundaries, and object segmentation. Moreover, the full camera calibration and 3D point positions are available, i.e. our dataset also covers RGBD data. We cannot exploit the full potential of this dataset in a single paper, but we already demonstrate various usage examples in conjunction with convolutional network training. We train a network for disparity estimation, which yields competitive performance also on previous benchmarks, especially among those methods that run in real-time. Finally, we also present a network for scene flow estimation and provide the first quantitative numbers on full scene flow on a sufficiently sized test set [5].

S. Meister et al, Unflow: Unsupervised learning of optical flow with a bidirectional census loss describes the era of end-to-end deep learning, many advances in computer vision are driven by large amounts of labeled data. In the optical flow setting, however, obtaining dense perpixel ground truth for real scenes is difficult and thus such data is rare. Therefore, recent end-to-end convolutional networks for optical flow rely on synthetic datasets for supervision, but the domain mismatch between training and test scenarios continues to be a challenge. Inspired by classical energy-based optical flow methods, we design an unsupervised loss based on occlusion-aware bidirectional flow estimation and the robust census transform to circumvent the need for ground truth flow. On the KITTI benchmarks, our unsupervised approach outperforms previous unsupervised deep networks by a large margin, and is even more accurate than similar supervised methods trained on synthetic datasets alone. By optionally fine-tuning on the KITTI training data, our method achieves competitive optical flow accuracy on the KITTI 2012 and 2015 benchmarks, thus in addition enabling generic pre-training of supervised networks for datasets with limited amounts of ground truth [6].

M. Menze et al, Joint 3d estimation of vehicles and scene flow in this paper, author build upon the method of (Menze and Geiger, 2015) but go one step further: Instead of simply decomposing the scene into a set of individually moving regions which share a common rigid motion, we decompose the scene into 3D objects and in addition to the rigid motion also model their pose and shape in 3D. Towards this goal, we incorporate a deformable 3D model of vehicles into the scene flow estimation process. More specifically, we exploit the Eigenspace-based representation of (Zia et al., 2011) which has previously been used in the context of pose estimation from a single image. Given two stereo pairs as input, our model jointly infers the number of vehicles, their shape and pose parameters, as well as a dense 3D scene flow field. The problem is formalized as energy minimization on a conditional random field encouraging projected object hypotheses to agree with the estimated motion and depth. A representative result is shown in Fig. 1 which depicts scene flow estimates projected to disparity and optical flow as well as the result of model-based reconstruction [7].

A. Ranjan, Optical flow estimation using a spatial pyramid network author gives an alternative approach that combines the best of both approaches. Decades of research on flow has produced well engineered systems and principles that are effective. But there are places where these methods make assumptions that limit their performance. Consequently, here we apply machine learning to address the weak points, while keeping the engineered architecture, with the goal of 1) improving performance over existing neural networks and the classical methods upon which our work is based; 2) achieving real-time flow estimates with accuracy better than the much slower classical methods; and 3) reducing memory requirements to make flow more practical for embedded, robotic, and mobile applications. Computing flow requires the solution to two problems. One is to solve for long-range correlations while the other is to solve for detailed sub-pixel optical flow and precise motion boundaries. The previous neural network method, FlowNet, attempts to learn both of these at once. In contrast, we tackle the latter using deep learning and rely on existing methods to solve the former [8].

A. Ranjan et al, Competitive collaboration: Joint unsupervised learning of depth, camera motion, optical flow and motion segmentation author address the unsupervised learning of several interconnected problems in low-level vision: single view depth prediction, camera motion estimation, optical flow, and segmentation of a video into the static scene and moving regions. Our key insight is that these four fundamental vision problems are coupled through geometric constraints. Consequently, learning to solve them together simplifies the problem because the solutions can reinforce each other. We go beyond previous work by exploiting geometry more explicitly and segmenting the scene into static and moving regions. To that end, we introduce Competitive Collaboration, a framework that facilitates the coordinated training of multiple specialized neural networks to solve complex problems. Competitive Collaboration works much like expectation-maximization, but with neural networks that act as both competitors to explain pixels that correspond to static or moving regions, and as collaborators through a moderator that assigns pixels to be either static or independently moving. Our novel method integrates all these problems in a common framework and simultaneously reasons about the segmentation of the scene into moving objects and the static background, the camera motion, depth of the static scene structure, and the optical flow of moving objects. Our model is trained without any supervision and achieves state-of-the-art performance among joint unsupervised methods on all sub-problems [9].

Z. Ren et al, Unsupervised deep learning for optical flow estimation this paper states that the deep network trained using our unsupervised scheme, approaches the level of performance of fully supervised training. We believe that this is largely due to our end-to-end training, which allows the network to leverage context information within a large region for inferring local motion. To this end, we summarize our main contributions as follows. 1) To our best knowledge, this is one of the first works for learning optical flow using a deep neural network without any supervision. Our work is fundamentally different from the state-of-the-art learning-free methods DeepFlow (Weinzaepfel et al. 2013) or EpicFlow (Revaud et al. 2015), and the supervised deep learning approach FlowNet (Fischer et al. 2015) and DispNet (Mayer et al. 2016). 2) We propose a novel optical flow network which can be seen akin to the pipeline of Spatial Transformer Network (Jaderberg et al. 2015), leveraging the loss function used in variational methods (Brox et al. 2004) without supervision, for end-to-end unsupervised learning for optical flow estimation. While the gains are modest, we believe this is a promising direction for future exploration. 3) Finally, to enable comparison and further innovation, we will provide a public Caffe (Jia et al. 2014) implementation [10].

F. Steinbrucker et al, Large displacement optical flow computation without warping author proposed a novel algorithm for estimating large displacement optical flow which circumvents the need for warping schemes. By means of a quadratic relaxation scheme we decompose the original non-convex functional into a functional which can be minimized by alternating two globally optimal steps. The algorithm simply alternates a complete search with respect to the non-convex (but point-wise) data term and a convex optimization that takes into account the smoothness constraint. The flow estimation process is therefore decomposed in an alternation of searching for appropriate correspondents and discontinuity preserving smoothing. In contrast to warping approaches, the proposed method can naturally make use of arbitrary data terms, including non-convex, non-differentiable terms and norms on color values or local patches. In numerous experiments, author show that in contrast to state-of-the-art warping schemes, the proposed quadratic decoupling scheme allows to compute flow fields which accurately match small scale structures over large displacements [11].

3. PROPOSED WORK

RAFT Model Our first ablation investigates dimensions improvement can be achieved by taking the situation in advance replacing the UFlow art method and its PWC model RAFT. As the results in Table 4 show, replace the model without further modifications to unsupervised learning it is no coincidence that the method improves but actually reduces performance[13]. Through extensive experimentation, we have identified and developed the techniques presented here that enable better unsupervised learning with RAFT [14]. Earnings from these techniques are much smaller with the PWC model, perhaps due to the more restrictive architecture. To design, develop and analyze "An algorithm SMRUF-Combining Augmentation with geometric transformation for Unsupervised Recurrent All-pairs Field Transform of Optical flow" the important points are as follows [16],

- RAFT Model
- Multi-Frame Self-Supervision
- Image Warping
- Unsupervised RAFT

a. Architecture Of Proposed System

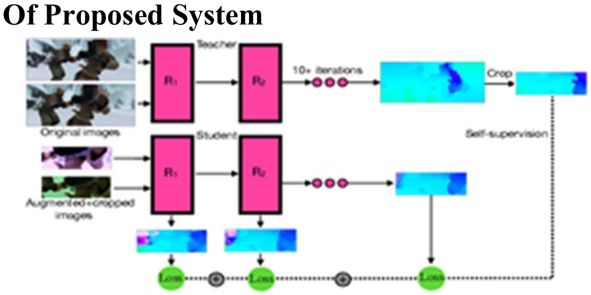

Fig. 1. Self-supervision with sequence loss and augmentation.

We use a single model as both “student” and “teacher”. As the teacher, we apply the model on full non-augmented images. As student, the model only sees a cropped and augmented version of the same images. The final output of the teacher is then cropped and used to supervise the predictions at all iterations of the student (to which the smoothness and photometric losses are applied as well).

The advantages of this self-supervision method are threefold:

(1) The model gains the ability to disregard photometric augmentations

(2) The model learns to make better predictions at the borders and in occluded areas of the image

(3) Early iterations of the recurrent model learn from the output at the final iteration images.

Then, a recurrent network is fed with this cost volume periodically, which iteratively constructs and improves a flow field prediction. The only architectural modification we make to RAFT is to replace batch normalization with instance normalization [15] to enable training with very small batch sizes. Reducing the batch size was necessary to fit the model and the more involved unsupervised training steps into memory. But more crucially, we discovered that important adjustments must be made to the unsupervised learning approach in order to fully utilize RAFT's capacity for learning [16].

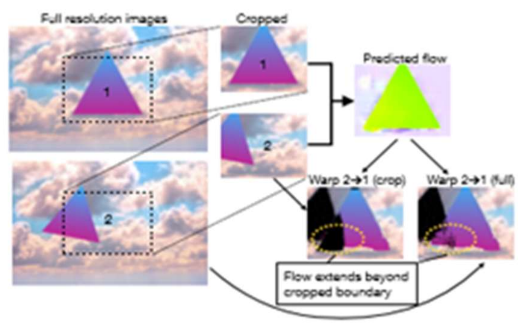

Fig. 2. Full image Warping.

The photometric loss, which is crucial for unsupervised optical flow estimation, is typically only present in flow vectors that remain within the image frame because they lack pixels with which to compare their photometric appearance. We address this limitation by computing the flow field from a cropped version of the images I1 and I2 while referencing the full, uncropped image I2 when warping it with the estimated flow V1 before computing the photometric loss (see Figure 2). These flow vectors that go outside the image frame are no longer classified as occluded, therefore they now act as a learning signal for the model [14]. Except for Flying Chairs, where we discovered that cropping the already-small photos hindered performance, we utilise full-image warping for all datasets.

4. PROPOSED SYSTEM

The study of hyper parameters is based on UFlow with slight modifications based on further hyper parameter search.

KITTI dataset is required for the study.

a. Dataset Details

The following optical flow datasets can be used to train and test our model in accordance with the standards put out in the literature: Flying Chairs, Sintel, and KITTI 2015. Prior to fine-tuning on Sintel or KITTI, we pretrain on Flying Chairs. Like UFlow, we did not see any advantages to pretraining on more non-domain data, such as Flying Objects. No training method in our approach employs any ground truth labels. We partition the Sintel dataset into its standard train and test portions, and we train using the "training" section of Flying Chairs. With KITTI, we practise using the split utilised in earlier work with the multi-view extension [12]: One model is trained on the multiview extension of the training set, and the other model is trained on the extension of the test, analyse these models correctly based on the test set. We report metrics for ablations after training on the "test" section of the dataset (which is label-free) and assessing on the training set, and we report results for final benchmark numbers after training on the training component exclusively. We train on a 50% blend of the KITTI and Sintel multi-frame self-supervision labels for our benchmark result on Sintel. With KITTI, we also report error rates ("ER") for all datasets. A prediction is deemed incorrect if its EPE is greater than 3 pixels or greater than 5% of the length of the true flow vector. With the exception of "EPE (noc)," where only non-occluded pixels are taken into account, we normally compute these metrics for all pixels.

b. Data Augmentation

Extensive Data Augmentation To regularize RAFT, we use the same augmentation as supervised RAFT which is much stronger than what has typically been used in unsupervised optical flow, except for the recent ARFlow. We apply a random eraser augmentation to each image, which removes random areas of the image, and we randomly change colour, brightness, saturation, stretching, scaling, random cropping, random flipping left/right and up/down. The model inputs are subjected to all augmentations, but the images used to calculate the photometric and smoothness losses are not. The self generated labels for self-supervision are computed from unaugmented images, which has the benefit of training the model to ignore these augmentations.

5. OUTCOME OF PROPOSED SYSTEM

The SMURF, an effective method for unsupervised learning of optical flow that reduces the gap to supervised approaches and shows excellent generalization across datasets and even to “zero-shot” depth estimation. SMURF introduces significant improvements, chief among which are

(1) modifications to the unsupervised losses and data augmentation that allow the RAFT architecture to operate in an unsupervised setting,

(2) full-image warping for learning to predict out of frame motion, and

(3) multi-frame self-supervision for improved flow estimates in occluded regions.

These contributions, in our opinion, represent a step towards making unsupervised optical flow really feasible, enabling optical flow models trained on unlabeled videos to deliver excellent pixel-matching in domains lacking labeled data..

- To design and analyze an algorithm SMRUF-Combining Augmentation with geometric transformation for Unsupervised Recurrent All-pairs Field Transform of Optical flow.
- To study and analyze Self-Supervision Modifications techniques.
- To study and analyze Multi-Frame Self-Supervision
- To study and analyze the RAFT Model
- To study and analyze the Self-supervision with sequence loss and augmentation

6. CONCLUSION

In this paper, we have proposed effective methods to enforce global popular geometry constraints for unsupervised optical flow learning. For a stationary scene, we applied the low-rank constraint to regularize a globally rigid structure. For general dynamic scenes(multi-body or deformable),we proposed to use the union-of-subspaces constraint. Experiments on various benchmarking datasets have proved the efficacy and superiority of our methods compared with state of-the-art (unsupervised) deep flow methods. In the future, we plan to study the multi-frame extension, i.e., enforcing geometric constraints across multiple frames.

REFERENCES

[1] A. Stone, D. Maurer, A. Ayvaci, A. Angelova, R. Jonschkowski, "SMURF: Self-Teaching Multi-Frame Unsupervised RAFT with Full-Image Warping", IEEE Conference on Computer Vision and Pattern Recognition), Page no.3887-3896, Honolulu, Hawaii, USA, July, 2021.

[2] F. Aleotti, F. Tosi, L. Zhang, M. Poggi, and S. Mattoccia, "Reversing the cycle: Self-supervised deep stereo through enhanced monocular distillation", in ECCV, Vol.3, Issue 3, Page no. 1-18, Aug. 2020.

[3] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping" in Eurppean Conference on Computer Vision, Page no.25-36, 2004.

[4] D. Butler, J. Wulff, G. Stanley, and M. Black, "A naturalistic open source movie for optical flow evaluation" in Proceedings of the 12th European conference on Computer Vision - Volume Part VI, Page no. 611-625, October 2012.

[5] Q. Chen and V. Koltun, "Full flow: Optical flow estimation by global optimization over regular grids", in CVPR, Page no. 4706-4714, 2016.

[6] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazırbas¸, V. Golkov, P. Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks", in ICCV, Page no.1-13, May 2015.

[7] H. Hirschmuller, "Stereo processing by semi-global matching and mutual information", IEEE Transactions on Pattern Analysis and Machine Intelligence, Page no. 328–341, February 2008.

[8] K. Berthold and B. Schunck, "Determining optical flow", AI, 1981. Artificial Intelligence, Volume 17, Issues 1–3, Pages 185-203, August 1981.

[9] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet 2.0: Evolution of optical flow estimation with deep networks", in CVPR, Page no. 2462-2470, 2017.

[10] W. Im, T. Kim, and S.-E. Yoon, "Unsupervised learning of optical flow with deep feature similarity" in ECCV, Page no. 2463-2473, 2020.

[11] J. Janai, F. Guney, A. Ranjan, M. J. Black, and A. Geiger, "Unsupervised learning of multi-frame optical flow with occlusions" in ECCV, Page no. 1235-1243, 2018.

[12] R. Jonschkowski, A. Stone, J. Barron, A. Gordon, K. Konolige, and A. Angelova, "What matters in unsupervised optical flow", in ECCV, Page no. 4850-4861, 2020.

[13] D. Kingma and J. Adam, "A method for stochastic optimization", in ICLR, Page no. 2243-2252, 2015.

[14] L. Liu, J. Zhang, R. He, Y. Liu, Y. Wang, Y. Tai, D. Luo, C. Wang, J. Li, and F. Huang, "Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation", in 4th Conference on Robot Learning, Page no. 1-13, 2020.
[15] P. Liu, I. King, M. Lyu, and J. Xu, "Flow2Stereo: Effective self-supervised learning of optical flow and stereo matching", in CVPR, Page no. 456-467, 2020.
[16] P. Liu, I. King, M. Lyu, and J. Xu, "DDFlow: Learning optical flow with unlabeled data distillation", AAAI, Page No.1-13, Hawaii USA, Feb 2019.
[17] P. Liu, M. R. Lyu, I. King, and J. Xu, "Selflow: Self-supervised learning of optical flow", Conference on Computer Vision and Pattern Recognition, Page no. 2423-2436, 2019