CFFlow: An Optical Flow Estimation Hinging on Cross-Frequency Attention

Mengfei Wang, Dongchen Zhu*, Lei Wang, and Jiamao Li

[1] Bionic Vision System Laboratory, State Key Laboratory of Transducer Technology, Science and Technology on Micro-system Laboratory, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China.

[2] University of Chinese Academy of Sciences, Beijing 100049, China.

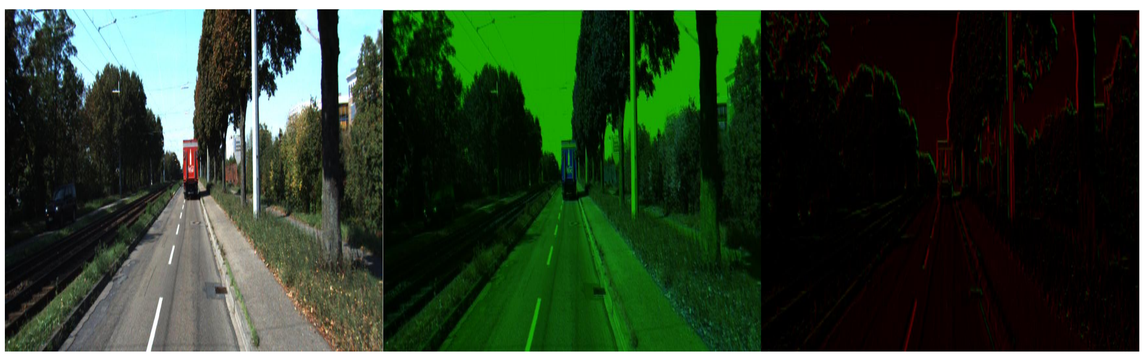
dchzhu@mail.sim.ac.cn

Motivations

With the advancement of neural network architectures, numerous effective optical flow methods have been developed. However, two primary limitations persist in current estimations :

(1) **Inefficient use of underlying motion structures**: Many methods predominantly rely on appearance features and visual similarity for matching, frequently overlooking the inherent motion patterns within scenes. This reduces their robustness under complex motion conditions.

(2) **Underutilization of cross-frequency features**: While pre-trained image features are heavily emphasized, optical flow estimation often benefits more from rich low-level cues. Although frequency-aware features have shown promise in dense prediction tasks, their potential for capturing multi-scale structural information in motion estimation remains underexplore



(a) Input Image (b) Image High-Frequency Feature (c) Image Low-Frequency Feature

Contributions

· We define cross-frequency information in optical flow tasks for the first time and design the IFS module to analyze these components, offering a complementary perspective to traditional spatial surface information.

· To fully exploit low- and high-frequency components, we introduce the LDA and SODA modules, targeting large displacements and small object motions, respectively, enhancing motion estimation across scales.

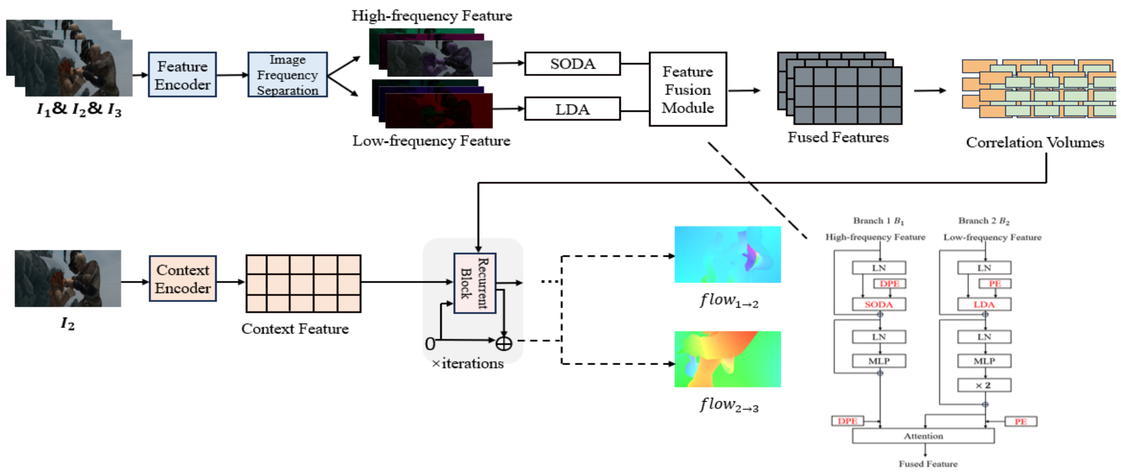
· Our approach achieves state-of-the-art results on Sintel and KITTI-2015 benchmarks, with AEPE of **0.989** and **1.647** on the clean and final passes, and an F1-all error of **4.21**, ranking 1st among all three-frame methods.

Method

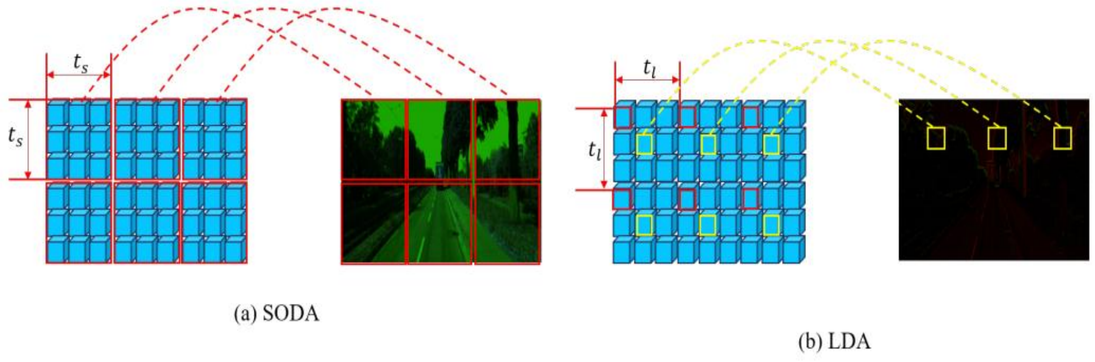
Network Architecture:

As shown in Figure, we redesign the optical flow feature extraction process by introducing the Image Frequency Separation Module (IFS), which decomposes extracted features into low- and high-frequency components. Built on this framework, our model CFFlow incorporates two specialized modules: SODA for small object motion and LDA for large displacement, enabling accurate flow estimation across scales.

A Feature Fusion Module (FF) is then used to integrate both frequency branches. The following sections detail the design of IFS, SODA, LDA, and FF.



(a) Overview of our CFFlow training pipeline. The IFS module decomposes image features into high- and low-frequency components. Then, SODA models small-object motion from high-frequency features, and LDA captures large displacements from low-frequency features. Finally, the FF module fuses both to produce refined motion estimation..



(a) SODA (b) LDA

(b) SODA module: High-frequency features (blue cubes) are grouped within red boxes. LDA module: Low-frequency features with identical border colors indicate grouping for large displacement modeling.

Experiments

Training Data	Method	Sintel(train) Clean	Sintel(train) Final	KITTI-15(train) EPE	KITTI-15(train) F1-all	Sintel(test) Clean	Sintel(test) Final	KITTI-15(test) F1-all
C+T	FlowNet2 [14]	2.02	3.54	10.08	30.0	3.96	6.02	-
	SKFlow [30]	1.22	2.46	4.27	15.5	-	-	-
	RAFT [33]	1.43	2.71	5.04	17.4	-	-	-
	FlowFormer [13]	0.94	2.33	4.09	14.72	-	-	-
	FlowFormer++ [26]	0.90	2.30	3.93	**14.13**	-	-	-
	CRAFT [28]	1.27	2.79	4.88	17.5	-	-	-
	GMFlow [37]	1.08	2.48	-	-	-	-	-
	GMA [16]	1.30	2.74	4.69	17.1	-	-	-
	Videoflow-BOF [25]	1.03	2.19	3.96	15.33	-	-	-
	Videoflow-MOF [25]	1.18	2.56	**3.89**	14.20	-	-	-
	Ours	**0.89**	**2.11**	3.88	14.24	-	-	-
C+T+S+K(+H)	PWC-Net [29]	(1.71)	(2.34)	(1.50)	(5.3)	3.44	4.58	7.71
	GMFlow [37]	-	-	-	-	1.74	2.90	9.32
	SKFlow* [30]	(0.52)	(0.78)	(0.51)	(0.94)	1.28	2.27	4.48
	RAFT* [33]	(0.75)	(1.21)	(0.63)	(1.5)	1.94	3.18	5.11
	CRAFT [28]	(0.60)	(1.06)	(0.58)	(1.33)	1.45	2.40	4.81
	GMA [16]	(0.62)	(1.07)	(0.57)	(1.2)	1.39	2.48	5.14
	PWC-Fusion* [24]	-	-	-	-	3.43	4.57	7.17
	FlowFormer [13]	(0.48)	(0.74)	(0.53)	(1.11)	1.16	2.09	4.68
	FlowFormer++ [26]	(0.40)	(0.60)	(0.57)	(1.16)	1.07	1.94	4.52
	MatchFlow(R) [7]	(0.51)	(0.81)	(0.59)	(1.3)	1.33	2.64	4.72
	SAMFlow [45]	-	-	-	-	1.00	2.08	4.49
	MatchFlow(G) [7]	(0.49)	(0.78)	(0.55)	(1.1)	1.16	2.37	4.63
	Videoflow-BOF* [25]	(0.37)	(0.54)	(0.52)	**(0.85)**	1.02	1.84	4.44
	SplatFlow* [35]	(0.53)	(0.91)	(0.80)	(2.4)	1.12	2.07	4.61
	Ours	**(0.35)**	**(0.47)**	**(0.45)**	(0.91)	**0.99**	**1.65**	**4.21**

Method	Sintel Test(clean) All	Matched	Unmatched	s_{0-10}	s_{10-40}	s_{40+}	Sintel Test(final) All	Matched	Unmatched	s_{0-10}	s_{10-40}	s_{40+}
MemFlow [6]	1.064	0.426	6.091	0.253	0.778	5.623	1.914	0.931	9.928	0.430	1.382	10.556
SAMFlow [45]	1.000	0.384	5.966	0.252	0.760	**5.245**	2.080	1.036	10.60	0.515	1.488	11.278
VideoFlow-BOF [25]	1.005	0.389	6.023	0.229	0.695	5.605	1.713	0.812	9.054	**0.387**	1.242	9.422
StreamFlow [31]	1.041	**0.381**	6.416	**0.212**	**0.690**	5.999	1.874	0.824	10.435	0.409	**1.240**	10.674
FlowFormer++ [26]	1.073	0.390	6.635	0.252	0.796	5.810	1.943	0.878	10.627	0.438	1.404	10.712
Ours(CFFlow)	**0.989**	0.396	**5.819**	0.229	0.694	5.470	**1.647**	**0.788**	**8.650**	0.402	1.242	**8.793**

Conclusion

This paper investigates the integration of cross-frequency features into optical flow estimation.

· Current multi-frame methods often focus on surface-level cues, overlooking deeper frequency information. To address this, we propose CFFlow, which incorporates cross-frequency features via two modules: LDA for large displacements and SODA for small object motion.

· Experiments show that CFFlow achieves state-of-the-art performance on Sintel and KITTI, ranking first among three-frame methods.

· However, fully embedding cross-frequency knowledge remains a challenge for future work.

Talents Recruitment

Recruitment

Contact:

dchzhu@mail.sim.ac.cn

jmli@mail.sim.ac.cn

Welcome to scan the QR code to join us!