

مروری بر روش‌های تخمین جریان نوری: از مدل‌های کلاسیک تا یادگیری عمیق، پیشرفت‌ها و چالش‌ها

ندا فرجی یامچلو³

مرتضی محمدی زنجیره²

پریسا نقدی¹

¹ دانشجوی کارشناسی ارشد، مهندسی برق-مخابرات سیستم، ² استادیار گروه مهندسی کامپیوتر، ³ استادیار گروه مهندسی برق-مخابرات

^{1,2,3} دانشکده فنی-مهندسی، دانشگاه بین‌المللی امام خمینی (ره)، قزوین، ایران

nfaraji@eng.ikiu.ac.ir³ zanjireh@eng.ikiu.ac.ir²parisaanaqdi@yahoo.com¹

چکیده - تخمین جریان نوری یکی از مفاهیم پایه بینایی کامپیوتر است که از طریق آن جابجایی پیکسل‌ها، بین دو فریم متوالی یک تصویر بدست می‌آید. این مقاله ضمن بررسی روش‌های کلاسیک تخمین جریان نوری و روش‌های مبتنی بر یادگیری عمیق، به چالش‌های موجود در این حوزه از جمله تغییرات شدید نور، الگوهای حرکتی پیچیده و انسداد می‌پردازد. با وجود عملکرد برتر مدل‌های یادگیری عمیق مانند FlowNet و RAFT در مقایسه با روش‌های کلاسیک در بسیاری از پایگاه‌های داده، چالش‌های موجود در مسائل دنیای واقعی، باعث کاهش دقت تخمین در این مدل‌ها می‌شوند. این مقاله ضمن بررسی چالش‌ها، مجموعه داده‌های موجود، روش‌های ارزیابی و کاربردهای جریان نوری به ارائه راهکارهایی برای بهبود عملکرد مدل‌ها طبق مقالات می‌پردازد. کلید واژه-انسداد، بینایی کامپیوتر، جریان نوری، یادگیری عمیق.

1- مقدمه

مدل‌های یادگیری عمیق قادرند جریان نوری را به طور مستقیم از دو تصویر متوالی تخمین بزنند و نیازی به استخراج ویژگی‌ها به صورت دستی ندارند، که اغلب استخراج آن‌ها هزینه‌بر است. بنابراین این مدل‌ها باعث پیشرفت‌های مهم در دقت و کارایی محاسباتی شده‌اند. با این حال این روش‌ها بی‌نقص نیستند. مدل‌های سنتی با محدودیت‌هایی مانند عدم توانایی در تخمین جابه‌جایی‌های بزرگ و هزینه محاسباتی بالا مواجه بودند. از طرفی مدل‌های یادگیری عمیق به حجم زیادی از داده‌های آموزشی نیاز دارند و در محیط‌های پیچیده دچار مشکل می‌شوند [3]. برای مثال داده‌های مصنوعی موجود مانند (MPI-Sintel) نمی‌توانند همه شرایط واقعی را پوشش دهند. تغییرات نور، انسداد اشیاء و صحنه‌های پویا، همچنان چالش‌های بزرگی برای این روش‌ها هستند. با این حال روش‌های کلاسیک همچنان جایگاه خود را دارند و در برخی موارد ترکیب آن‌ها با شبکه‌های عصبی نتایج بهتری ارائه می‌دهد.

یکی از چالش‌های اصلی در این حوزه، کمبود مجموعه

جریان نوری به عنوان یک ابزار پایه در بینایی کامپیوتر است که برای اندازه‌گیری حرکت نسبی اشیاء بین دو فریم متوالی استفاده می‌شود. این روش با محاسبه جابه‌جایی پیکسل‌ها بین دو فریم متوالی یک ویدئو، امکان درک حرکت اجسام را در صحنه‌های پویا فراهم می‌سازد. این مفهوم پیش‌نیازی برای کاربردهای مهم مانند ردیابی اشیاء، تقسیم‌بندی و بازسازی سه بعدی است.

مفهوم جریان نوری اولین بار توسط جیمز گیبسون James (J.Gibson) [1] در سال 1950 معرفی شد. با این حال اولین روش محاسباتی عملی توسط هورن شانک (Horn and Schunck) [2] در سال 1981 ارائه شد. در این روش از حداقل‌سازی یک تابع انرژی استفاده می‌شود تا همواری در میدان حرکت تخمین زده شده، تضمین شود. از آن زمان تاکنون روش‌های متعددی برای تخمین جریان نوری ارائه شده است، از جمله روش‌های مبتنی بر تغییرات، تطابق ویژگی‌ها و اخیراً شبکه‌های عصبی عمیق.

[4]، همچنین با تغییرات نور در اثر سایه، تابش نور خورشید و غیره دچار خطا می‌شود.

Lukas Kanade (LK): این روش در سال 1981 معرفی شد [5]. برخلاف روش هورن-شانک که فرض می‌کرد جریان نوری در کل تصویر هموار است، فرض هموار بودن جریان نوری را در یک ناحیه کوچک از تصویر (مثلاً یک پنجره 5×5 پیکسلی) دارد. در این روش مشتقات شدت روشنایی در دو جهت افقی و عمودی در یک ناحیه کوچک محاسبه می‌شود. سپس با استفاده از فرض ثابت بودن شدت روشنایی در طول حرکت، یک سیستم معادلات خطی تشکیل می‌شود که این معادلات به روش حداقل مربعات حل شده و در نهایت جریان نوری تخمین زده می‌شود. روش لوکاس کاناده برای جابجایی‌های کوچک مناسب است، اما برای جابجایی‌های بزرگ دقت آن کاهش می‌یابد.

Brox: این روش که در سال 2004 معرفی شد، از استراتژی‌های هرمی برای افزایش دقت تخمین جریان نوری استفاده می‌کند. در این روش، تصویر ابتدا در چندین مقیاس کوچک شده و جریان نوری در سطوح پایین‌تر تخمین زده می‌شود. سپس این تخمین به تدریج در مقیاس‌های بالاتر بهبود می‌یابد. مزیت اصلی این روش، دقت بالاتر در جابجایی‌های بزرگ است، زیرا در مقیاس‌های پایین، حرکت‌های وسیع کوچک‌تر و قابل کنترل‌تر به نظر می‌رسند [6].

2-2 روش‌های مبتنی بر یادگیری عمیق

روش‌های کلاسیک مانند لوکاس کاناده (LK) و هورن شانک (HS) به دلیل سادگی و وابستگی کمتر به داده‌های آموزشی همچنان در برخی کاربردها مورد استفاده قرار می‌گیرند. با این حال، این روش‌ها در جابجایی‌های بزرگ، تغییرات نور و انسداد عملکرد ضعیفی دارند. از این رو، محققان به سمت استفاده از روش‌های مبتنی بر یادگیری عمیق حرکت کرده‌اند که دقت بالاتری را ارائه می‌دهند.

FlowNet: اولین شبکه عصبی عمیق برای تخمین جریان نوری که در سال 2015 معرفی شد. این مدل براساس ساختار Encoder-Decoder و معماری U-Net طراحی شده است. در این ساختار، بخش Encoder وظیفه‌ی استخراج ویژگی‌های تصویر را بر عهده دارد، در حالی که بخش Decoder مسئول

داده‌های واقعی با برجسب دقیق است. که اغلب محققان را مجبور به استفاده از مجموعه داده‌های مصنوعی می‌کند. در این مقاله برخی از این مجموعه داده‌ها را نیز معرفی خواهیم کرد.

در ادامه این مقاله، ابتدا در بخش دو برخی روش‌های جریان نوری و مدل‌های یادگیری معرفی می‌شوند. در بخش سه مشکلات اصلی در محاسبه جریان نوری و راهکارهای موجود بررسی خواهند شد. در بخش چهارم مجموعه داده‌های پرکاربرد و روش‌های ارزیابی جریان نوری و در بخش پنجم کاربردهای مختلف این روش بررسی می‌شود.

2- روش‌های تخمین جریان نوری

در این قسمت انواع روش‌های تخمین جریان نوری را بررسی می‌کنیم. ابتدا روش‌های کلاسیک مانند روش‌های مبتنی بر تغییرات و تطابق ویژگی‌ها را معرفی کرده و سپس به معرفی روش‌های مدرن مبتنی بر یادگیری عمیق می‌پردازیم.

2-1 روش‌های کلاسیک

روش‌های کلاسیک تخمین جریان نوری بر پایه مفاهیم ریاضی و فیزیک هستند. این روش‌ها معمولاً از فرض ثابت بودن شدت روشنایی پیکسل‌ها در طول حرکت (Brightness Constancy Assumption) برای تخمین جریان نوری استفاده می‌کنند، یعنی فرض می‌کنند شدت روشنایی یک پیکسل در طول حرکت ثابت است و تغییر نمی‌کند. در این بخش به چند مورد از روش‌های کلاسیک اشاره می‌شود.

(HS) Horn and Schunck: این روش اولین مدل ارائه شده برای تخمین جریان نوری است که در سال 1981 معرفی شد [2]. در این روش، جریان نوری با کمینه سازی یک تابع انرژی تخمین زده می‌شود. این تابع شامل دو بخش اصلی است:

1- قید روشنایی: فرض ثابت بودن شدت روشنایی یک پیکسل در طول حرکت.

2- قید هموارسازی: فرض می‌کند که جریان نوری در کل تصویر ثابت است و تغییرات ناگهانی ندارد.

این روش برای جریان‌های هموار و کوچک مناسب است و در مرزهای حرکت (Motion Boundaries) دقت کمتری دارد

به کارگیری معماری ترنسفورمر و محاسبه تطابق در کل تصویر، بدون نیاز به حجم هزینه (Cost Volume) و مراحل بازگشتی متعدد، دقت بالایی را ارائه می‌دهد. برخلاف روش‌هایی مانند RAFT که به چندین مرحله اصلاح نیاز دارند، GMFlow تنها با یک مرحله پردازش، عملکردی قابل مقایسه یا بهتر را با سرعت پردازشی بالاتر فراهم می‌کند. این رویکرد، گامی مهم در جهت بهبود کارایی و کاهش پیچیدگی محاسباتی در تخمین جریان نوری محسوب می‌شود. [11]

2-3- ترکیب روش‌های کلاسیک و مدرن

SpyNet (2017): یک شبکه فشرده برای تخمین جریان نوری است که از روش سنتی رویکرد درشت به ریز (coarse_to_fine) (تخمین اولیه جریان در مدل‌های ساده و مقیاس پایین تصویر و بهبود در مقیاس‌های بالاتر) در ترکیب با تکنیک‌های یادگیری عمیق استفاده می‌کند. [12]

PWC-Net (2018): این شبکه از چارچوب‌های هرمی برای پردازش ویژگی‌های تصویر و تخمین جریان نوری استفاده می‌کند. این مدل با ترکیب یادگیری عمیق و تکنیک‌های سنتی تطابق استریو ویژگی‌های تصویر را استخراج می‌کند. استفاده از این داده‌ها باعث می‌شود جریان نوری دقیق‌تری تخمین زده شود. معماری این روش به مراتب سبک‌تر از معماری FlowNet2.0 است. [13]

DeepFlow (2013) و EpicFlow (2015): دو الگوریتم هستند که از روش‌های بهینه‌سازی انرژی (روش‌های کلاسیک) و تطابق ویژگی‌ها بین دو فریم متوالی برای تخمین جریان نوری استفاده می‌کنند. در این روش‌ها، ابتدا ویژگی‌های فریم‌های ورودی استخراج و تحلیل شده، سپس با تطابق این ویژگی‌ها در دو فریم متوالی، جریان نوری تخمین زده می‌شود. [3], [14], [15]

2-4- یادگیری بانظارت

در این روش از مدل‌های انتها به انتها (end-to-end) مانند FlowNet و PWC-Net برای تخمین جریان نوری استفاده می‌شود [7] و [13] این مدل‌ها بر روی داده‌های برچسب‌گذاری شده آموزش می‌بینند، یعنی تصاویر متوالی همراه با میدان‌های جریان نوری دقیق برای یادگیری مورد استفاده قرار می‌گیرند. یکی از چالش‌های اصلی این مدل‌ها،

بازسازی تصویر و تخمین جریان نوری است. این شبکه در دو نسخه ارائه شده است: FlowNetS (نسخه ساده) و FlowNetC (نسخه‌ای با قابلیت تطابق ویژگی‌ها). FlowNet از داده‌های مصنوعی مانند (FlyingChairs) برای آموزش استفاده کرده است، که شامل جفت تصاویر متوالی همراه با برچسب‌های جریان نوری واقعی (ground truth) می‌باشد. [7]

FlowNet2.0 (2017): این شبکه نسخه بهبودیافته FlowNet است که با ترکیب چندین زیر شبکه FlowNet، دقت تخمین جریان نوری را به میزان قابل‌توجهی افزایش داده است. [8]

RAFT (Recurrent All-Pairs Field Transforms): شبکه یکی از پیشرفته‌ترین الگوریتم‌های تخمین جریان نوری است که در سال 2020 معرفی شد. این شبکه با استفاده از همبستگی‌های همه‌جفتی (All-Pairs Correlation) و به‌روزرسانی‌های بازگشتی (Recurrent Updates) دقت بالایی در تخمین جریان نوری ارائه می‌دهد. معماری RAFT شامل سه بخش اصلی است: 1- استخراج ویژگی‌ها 2- محاسبه‌ی همبستگی چندمقیاسی 3- به‌روزرسانی بازگشتی میدان جریان نوری. برخلاف روش‌های سنتی، این شبکه با استفاده از یک گرادینان بهینه‌سازی شده و محاسبات تکرارشونده، تخمینی دقیق و هموار از جریان نوری ارائه می‌دهد. [9]

Flowformer (2022): این شبکه یک مدل پیشرفته برای تخمین جریان نوری است که بر پایه معماری ترنسفورمر و مکانیزم خودتوجهی (self-attention) طراحی شده است. این شبکه قادر است ویژگی‌های تصویر را در مقیاس‌های مختلف پردازش کند و با استفاده از self-attention ارتباطات بلند مدت، پیسکل‌ها را به طور موثر استخراج کند. یکی از مزایای کلیدی FlowFormer، عملکرد بهینه در شرایط پیچیده‌ای مانند تغییرات نور و سایه است که باعث می‌شود این مدل دقت بالاتری در تخمین جریان نوری در محیط‌های متغیر داشته باشد. [10]

GMFlow (2022): یک مدل پیشرفته برای تخمین جریان نوری است که به‌جای استفاده از همبستگی محلی، از مقایسه سراسری ویژگی‌ها در تصویر بهره می‌برد. این مدل با

محدودیت‌های روش‌های کلاسیک را کاهش داد و به بهبود تخمین حرکت کمک کرد.

3- استخراج ویژگی‌های تصویر برای تخمین حرکت: یو و همکاران (Yu et al) [18] رویکرد متفاوتی را پیشنهاد دادند. در این روش، ابتدا ویژگی‌های مهم تصویر شناسایی می‌شوند و سپس مدل، اختلاف شدت پیکسل‌ها را کاهش می‌دهد تا بتواند حرکت را دقیق‌تر تخمین بزند.

4- یادگیری سازگاری دوطرفه برای کاهش خطا: مایستر و همکاران (Meister et al) ایده‌ی جدیدی را معرفی کردند [19]. مدل باید هم‌زمان جریان نوری رو به جلو و معکوس را یاد بگیرد. با این تکنیک، خطاهای احتمالی در تخمین حرکت اصلاح می‌شوند و دقت افزایش می‌یابد.

5- آموزش فشرده برای بهبود یادگیری مدل‌های سبک‌تر: مدل‌هایی مانند SelFlow [20] و DDFlow [21] از آموزش فشرده (Knowledge Distillation) استفاده می‌کنند. ابتدا یک مدل پیچیده‌تر (مدل معلم) آموزش داده می‌شود و سپس مدل ساده‌تر (مدل دانش‌آموز) از آن برای یادگیری بهتر استفاده می‌کند. این روش، عملکرد مدل‌های سبک‌تر را در داده‌های واقعی بهبود می‌بخشد.

در مقایسه با یادگیری نظارت‌شده، روش‌های بدون نظارت معمولاً دقت کمتری دارند. اما با پیشرفت معماری‌های شبکه‌های عصبی، این روش‌ها بهبود یافته و در برخی شرایط، به عملکردی نزدیک به مدل‌های نظارت‌شده دست پیدا کرده‌اند.

3- چالش‌های کلیدی و راه‌حل‌ها

یکی از مشکلاتی که الگوریتم‌های تخمین جریان نوری در دنیای واقعی با آن روبه‌رو هستند، کاهش دقت در شرایط پیچیده است. تغییرات نور، انسداد و جابه‌جایی‌های بزرگ می‌توانند باعث ایجاد خطا در پیش‌بینی حرکت پیکسل‌ها شوند. در این بخش، به بررسی این چالش‌ها می‌پردازیم و راه‌حل‌هایی که برای آن‌ها پیشنهاد شده را معرفی می‌کنیم.

3-1- جابه‌جایی‌های بزرگ

حرکت سریع اشیاء، نرخ فریم پایین و فاصله‌ی کم اجسام از دوربین از جمله عواملی هستند که باعث ایجاد جابه‌جایی‌های

نیاز به مجموعه‌داده‌های بسیار بزرگ و برچسب‌گذاری شده است. علاوه بر این، مدل‌ها اغلب در انتقال به داده‌های دنیای واقعی دچار مشکل می‌شوند و ممکن است دقت مطلوب را نداشته باشند. برای حل این مشکلات، معمولاً دو راه‌حل رایج وجود دارد: 1- پیش‌آموزش روی داده‌های مصنوعی (Synthetic Pre-training) برای ایجاد یک پایه‌ی قوی‌تر در مدل. 2- تنظیم دقیق (Fine-tuning) روی داده‌های واقعی، به منظور بهبود عملکرد مدل در شرایط واقعی.

2-5- یادگیری نیمه نظارتی

این روش به صورت ترکیبی از داده‌های برچسب‌دار و بدون برچسب به صورت هم‌زمان استفاده می‌کند. ترکیب این دو نوع داده باعث می‌شود نیاز به داده‌های برچسب‌گذاری شده کاهش یابد و در عین حال، تعمیم‌پذیری مدل بهبود پیدا کند. یکی از روش‌های پرکاربرد در این زمینه، استفاده از شبکه‌های متخاصم مولد (GANs) است. [16] در این روش، از داده‌های بدون برچسب برای بهبود دقت مدل استفاده می‌شود، که باعث می‌شود بدون نیاز به حجم زیادی از داده‌های مصنوعی، عملکرد مدل در شرایط واقعی بهبود یابد.

2-6- یادگیری بدون نظارت

در یادگیری بدون نظارت، مدل‌ها بدون نیاز به داده‌های برچسب‌دار آموزش می‌بینند. برخلاف روش‌های نظارت‌شده که اطلاعات دقیقی درباره‌ی حرکت پیکسل‌ها دارند، این مدل‌ها تلاش می‌کنند تا با بررسی ویژگی‌های تصویر و روابط بین فریم‌های متوالی، الگوهای حرکتی را شناسایی کنند. روش‌های اصلی یادگیری بدون نظارت در تخمین جریان‌نوری عبارتند از:

1- روش‌های کلاسیک و اولیه: در روش‌های ابتدایی، فرض بر این بود که شدت روشنایی پیکسل‌ها در طول حرکت ثابت می‌ماند و جریان نوری هموار است. [2] این فرضیات کمک می‌کردند که مدل‌ها حرکت را تخمین بزنند، اما دقت آن‌ها در شرایط واقعی پایین بود.

2- ترکیب شبکه‌های عصبی و روش‌های کلاسیک: احمدی و پاتراس (Ahmadi & Patras) [17] نشان دادند که با ترکیب روش‌های سنتی و شبکه‌های عصبی، می‌توان جریان نوری را بدون داده‌های برچسب‌دار نیز یاد گرفت. این مدل،

کلیدی تصویر بررسی می‌شود. مدل EpicFlow [15] از این تکنیک برای تخمین جریان نوری استفاده می‌کند. این روش در نواحی دارای ویژگی‌های متمایز عملکرد خوبی دارد، اما در مناطقی با بافت‌های تکراری یا یکنواخت ممکن است دچار خطا شود.

راهکارهایی نیز برای بهبود دقت در روش‌های تطابق به کار می‌رود، مانند استفاده از یک شبکه‌ی منظم برای افزایش سرعت در جستجوی تطابق و بهره‌گیری از اطلاعات پیکسل‌های مجاور برای اصلاح تخمین اولیه. [25]، [26]. البته مدل‌های مدرن یادگیری عمیق مانند PWC-Net و RAFT از روش‌های مبتنی بر تطابق ویژگی‌ها برای تعیین میزان جابه‌جایی بین پیکسل‌ها استفاده می‌کنند. این تکنیک‌ها باعث افزایش دقت در تخمین جریان نوری و مدیریت بهتر حرکت‌های پیچیده می‌شوند.

3-2- انسداد

انسداد زمانی رخ می‌دهد که یک شیء در یک فریم توسط شیء دیگری پوشانده شود یا یک پیکسل در یک تصویر، تطابقی در تصویر دیگر نداشته باشد. این پدیده باعث می‌شود تشخیص حرکت پیکسل‌های مسدود شده دشوار شود و الگوریتم‌های جریان نوری با چالش‌های اساسی روبه‌رو شوند.

3-2-1- سازگاری دوطرفه

در این روش، جریان نوری در دو جهت مستقیم و معکوس (جلو و عقب) محاسبه می‌شود. سپس، با مقایسه‌ی این دو جریان، نواحی دچار انسداد شناسایی شده و اصلاح می‌شوند. این تکنیک باعث بهبود دقت در تخمین حرکت و کاهش خطاهای ناشی از نواحی غیرقابل مشاهده می‌شود. روش VideoFlow [27] از این استراتژی برای بهبود تشخیص جریان نوری استفاده می‌کند.

3-2-2- MASKFLOWNET

MaskFlowNet [28] یک مدل مبتنی بر شبکه عصبی است که به‌طور هم‌زمان برای تخمین جریان نوری و شناسایی نواحی انسداد (Occlusion) و طراحی شده است. این مدل از معماری Encoder-Decoder استفاده می‌کند. ابتدا، ویژگی‌های تصویر توسط بخش Ecodoer استخراج می‌شوند.

بزرگ در تصاویر می‌شوند. این موارد باعث می‌شوند مدل‌ها نتوانند به راحتی مسیر حرکت پیکسل‌ها را دنبال کنند و دقت تخمین کاهش پیدا کند. این چالش به‌ویژه در مجموعه داده‌های KITTI [22] و MPI-Sintel [23] مشاهده شده است. روش‌های کلاسیک مانند الگوریتم لوکاس کاناده که برای جابه‌جایی‌های کوچک طراحی شده‌اند، در این شرایط عملکرد ضعیفی دارند. علاوه بر این، مدل‌های یادگیری عمیق نیز اگر به درستی آموزش نینند، دچار خطا خواهند شد. بنابراین، برای بهبود عملکرد در این شرایط، باید از روش‌هایی استفاده شود که بتوانند تغییرات بزرگ را بهتر مدیریت کنند.

3-1-1- روش‌های هرمی

یکی از راهکارهای موثر برای تخمین جریان نوری در شرایطی که جابه‌جایی‌های بزرگ وجود دارد، روش‌های هرمی هستند. در این روش‌ها، تصویر در چندین سطح مختلف پردازش می‌شود و مدل از وضوح پایین به وضوح بالا حرکت می‌کند تا تغییرات بزرگ را بهتر مدیریت کند. [6]. با این حال، یکی از محدودیت‌های این روش‌ها این است که ممکن است اشیای کوچک که جابه‌جایی‌های بزرگی دارند، در سطوح بالاتر (وضوح پایین‌تر) به درستی شناسایی نشوند. این مشکل به دلیل تأثیر تاری ناشی از کاهش وضوح تصویر رخ می‌دهد. روش‌های کلاسیک مانند مدل براکس از این تکنیک برای پردازش سلسله‌مراتبی تصویر استفاده کرده‌اند. همچنین، مدل‌های یادگیری عمیق پیشرفته مانند RAFT [9] و PWC-Net [13] نیز از این ایده برای مدیریت بهتر جابه‌جایی‌های بزرگ بهره می‌برند.

3-1-2- روش‌های تطابق در تخمین جریان نوری

جریان نوری را می‌توان به‌عنوان فرآیند تشخیص تطابق بین فریم‌های متوالی یک تصویر تفسیر کرد. دو رویکرد اصلی برای تطابق نقاط بین فریم‌ها وجود دارد: 1- روش مبتنی بر پیچ (Patch-based Matching) [24]: این روش، ویژگی‌های تصویر را در پیچ‌های (بخش‌های) مختلف جستجو می‌کند تا بتواند نقاطی را که بین فریم‌ها تغییر کرده‌اند، تطابق دهد. مدل DeepFlow [14] از این تکنیک استفاده کرده و با بهره‌گیری از اطلاعات پیکسل‌های اطراف، جابه‌جایی‌های بزرگ را بهتر مدیریت می‌کند. 2- روش مبتنی بر ویژگی (Feature-based Matching): در این روش، تطابق بین نقاط

3-3-2- استفاده از فضاهاى رنگى مستقل از نور

برخی فضاهاى رنگى مانند HSV، HSI، و RGB به گونه‌اى طراحی شده‌اند که شدت نور روی محاسبات تأثیر زیادی نداشته باشد.

3-3-3- استفاده از الگوی سنجش روشنایی (CENSUS TRANSFORM)

این روش، الگوی تغییرات روشنایی پیکسل‌هاى همسایه را به یک کد دودویی تبدیل می‌کند. در نتیجه، تصویر نسبت به تغییرات شدید روشنایی مقاوم‌تر خواهد شد.

3-3-4- مدل‌سازی فیزیکی تغییرات نور

برخی روش‌هاى پیشرفته، تأثیر نور را به‌صورت پارامترهاى خطی در مدل‌هاى جریان نوری وارد می‌کنند. مدل PBOFVI با در نظر گرفتن تغییرات نور به‌عنوان پارامترهاى ضرب و جمع، جریان نوری را به‌صورت دقیق‌تر تخمین می‌زند [31].

3-4- پردازش بلادرنگ (Real-Time-Processing)

برخی از کاربردهاى جریان نوری مانند ناوبری ربات‌ها یا رانندگی خودکار نیازمند پردازش سریع بیش از ۳۰ فریم در ثانیه هستند (FPS) هستند. مدل‌هاى پیچیده که مبتنی بر یادگیری عمیق هستند مانند (FlowNet2.0) اغلب به دلیل حجم محاسباتی بالا نمی‌توانند به این سرعت دست یابند. راهکار ارائه شده عبارت است از استفاده از شبکه‌هاى سبک‌وزن (Lightweight Networks). مدل‌هاى مانند LiteFlowNet3 با کاهش تعداد پارامترها (تا ۵.۲ میلیون) و حذف لایه‌هاى غیرضروری، باعث افزایش سرعت پردازش می‌شوند. [32]

در جدول 1، چالش‌ها و راهکارهاى مقابله با آنها در تخمین جریان نوری ارائه شده است.

4- ارزیابی و مجموعه داده‌ها

در این بخش برخی مجموعه داده‌ها و معیارهاى ارزیابی جریان نوری معرفی می‌شوند.

جدول 1- چالش‌ها و راهکارهاى مقابله با آنها در تخمین جریان نوری

سپس، بخش Decoder از این ویژگی‌ها برای بازسازی و محاسبه جریان نوری استفاده می‌کند.

در این مدل علاوه بر محاسبه جریان نوری، یک شاخه اضافی برای پیش‌بینی نواحی انسداد در نظر گرفته شده است. این شاخه یک ماسک انسداد تولید می‌کند که بخش‌هایی از تصویر را که جریان نوری در آنها دچار خطا می‌شود (به دلیل مخفی بودن نقاط یا انسداد) مشخص می‌کند. پس از تولید این ماسک، اطلاعات آن برای اصلاح جریان نوری به مدل بازگردانده می‌شود و به این ترتیب، مدل می‌تواند تخمینی دقیق‌تر و پایدارتر از جریان نوری ارائه دهد.

3-2-3- MIRRORFLOW

MirrorFlow [29] یک مدل پیشرفته برای تخمین جریان نوری است که از قید تقارن (Mirror Constraint) برای بهبود دقت تخمین در نواحی دچار انسداد (Occlusion) استفاده می‌کند. ایده‌ی اصلی این روش: هنگامی که یک شیء در یک فریم دچار انسداد می‌شود، ممکن است در فریم بعدی از زاویه‌ای دیگر قابل مشاهده باشد. این مدل از اطلاعات موجود در فریم‌هاى بعدی برای بازسازی و اصلاح جریان نوری در بخش‌هاى مسدود شده بهره می‌برد.

3-3- تغییرات نور

در بسیاری از روش‌هاى کلاسیک، فرض بر این است که حرکت یک شیء تنها بر اساس تغییر مسیر روشنایی محاسبه می‌شود [2]. اما در دنیای واقعی، شرایط نوری دائماً در حال تغییر است. عواملی مانند: تأثیرات نور خورشید و ابرها، انتقال از نواحی روشن به سایه و جابه‌جایی منابع نوری مصنوعی می‌توانند باعث خطا در تخمین جریان نوری شوند. بنابراین، لازم است که مدل‌هاى جریان نوری نسبت به این تغییرات مقاوم‌سازی شوند. راهکارهاى پیشنهادی برای مقابله با تغییرات نور در ادامه آورده می‌شود.

3-3-1- استفاده از توصیف‌گرهاى مقاوم در برابر تغییرات نورى

روش‌هایی مانند HOG (Histogram of Oriented Gradients) و SIFT (Scale-Invariant Feature Transform)، ویژگی‌هاى تصویر را بدون تأثیرپذیری از شدت نور استخراج می‌کنند. [30]

اجزای افقی و عمودی واقعی جریان نوری در پیکسل u_i, v_i به ترتیب اجزای عمودی و افقی بردار جریان نوری تخمینی در پیکسل $\hat{u}_i, \hat{v}_i$ هستند.

✓ خطای زاویه ای (Angular Error)

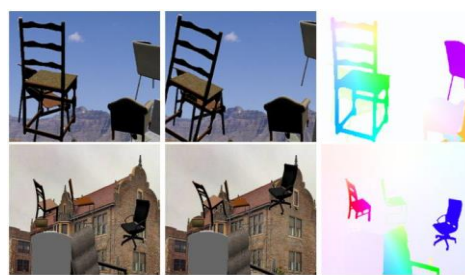
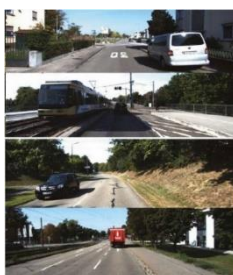
این معیار، اختلاف زاویه ای بین بردارهای جریان نوری تخمینی و مقدار واقعی را اندازه گیری می کند. مقدار AE به صورت زیر تعریف می شود:

$$AE = \arccos \left[\frac{f \cdot \hat{f} + 1}{\sqrt{(f \cdot \hat{f} + 1)(\hat{f} \cdot \hat{f} + 1)}} \right] \quad (2)$$

که در آن $f, \hat{f}$ به ترتیب بردارهای جریان نوری تخمینی و واقعی هستند و $f \cdot \hat{f}$ نمایانگر حاصل ضرب داخلی این دو بردار است. خطای زاویه ای (AE) برای جابجایی های کوچک مناسب است، اما در بردارهای بزرگ دقت کمتری دارد. در مقابل، میانگین خطای نقطه پایانی (EPE) برای بردارهای جریان نوری بزرگ مناسب تر است.

4-1-2- مجموعه داده ها

مجموعه داده های مختلفی برای آموزش و ارزیابی مدل های تخمین جریان نوری استفاده می شوند، که برخی از آنها به اختصار در جدول 2 بررسی شده اند. نمونه تصاویری از این مجموعه داده ها نیز در شکل 1 به تصویر کشیده شده است.



شکل 1- نمونه هایی از تصاویر مربوط به دیتاست های KITTI (سمت چپ)، MPI-Sintel (وسط) و Flying Chairs (سمت راست) است.

جدول 2- مجموعه داده ها

چالش ها	راه کارها
جابجایی ها ی بزرگ	روش های هرمی/تطابق ویژگی
انسداد	MaskFlowNet/MirrorFlow/ دو طرفه
تغییرات روشنایی	توصیفگرهای مقاوم (HOG/SIFT)/الگوی سنجش (CnsusTransform)/مدل PBOFVI/ف ضاهای رنگی (HSV, HSI)
پردازش بلادرنگ	شبکه های سبک وزن (Little FlowNet3)

4-1-1- ارزیابی

از جمله مهم ترین معیارها برای ارزیابی دقت و عملکرد الگوریتم های تخمین جریان نوری، می توان به دو معیار زیر اشاره کرد:

✓ میانگین خطای نقطه پایانی (Average Endpoint Error)

این معیار، میانگین فاصله ای اقلیدسی بین بردارهای تخمینی و بردارهای واقعی جریان نوری را محاسبه می کند. مقدار EPE به صورت زیر تعریف می شود:

$$EPE = \frac{1}{N} \sum_{i=1}^N \sqrt{(u_i - \hat{u}_i)^2 + (v_i - \hat{v}_i)^2} \quad (1)$$

که در آن N تعداد کل پیکسل ها در تصویر، u_i, v_i به ترتیب

دیتاست	تعداد فریم	داده واقعی	نوع سناریوها	روش تولید Ground Truth	کاربردها
KITTI	600	بله	محیط‌های شهری، خودروها، پیاده‌روها	اسکنر LiDAR و برچسب‌گذاری دستی	ارزیابی مدل‌های جریان نوری در خودروهای خودران
MPI-Sintel	2,194	خیر	حرکت‌های بزرگ، تغییرات نور و انسداد	شبیه‌سازی سه‌بعدی	ارزیابی الگوریتم جریان نوری در شرایط پیچیده
Dark Flow	1,200	خیر	شرایط نوری کم، تصاویر تاریک با نویز	مدل‌سازی نویز سنسور دوربین و ثبت حرکت پیکسل‌ها در نور کم	بهبود الگوریتم‌ها در شرایط نور کم
Flying Chairs	22,872 جفت	خیر	حرکات دوبعدی صندلی‌ها در پس‌زمینه‌های تصادفی	شبیه‌سازی CAD مصنوعی	آموزش مدل‌های یادگیری عمیق
Middlebury	72	بله	ترکیب تصاویر واقعی و مصنوعی	نورپردازی ساختاریافته برای ثبت اختلاف پیکسل‌ها	ارزیابی و مقایسه مدل‌های مختلف جریان نوری

جدول مقایسه روش‌های تخمین جریان نوری 4، عملکرد مدل‌های مختلف را بر اساس معیار میانگین خطای نقطه پایانی (EPE) در چهار مجموعه داده (MPI-Sintel (Clean)، MPI-Sintel (Final)، KITTI 2015 و KITTI 2012) نشان می‌دهد. این معیار میزان اختلاف بین جریان نوری تخمین زده شده و مقدار واقعی را محاسبه می‌کند. مجموعه داده Sintel (Clean) شامل صحنه‌های ساده‌تر و بدون اغتشاش است، در حالی که Sintel (Final) شامل افکت‌های نوری، سایه و جزئیات پیچیده‌تر است که شرایط واقعی‌تری را شبیه‌سازی می‌کند. مجموعه داده‌های KITTI نیز از صحنه‌های واقعی خیابانی تشکیل شده‌اند و چالش‌هایی مانند انسداد و جابه‌جایی‌های بزرگ را شامل می‌شوند.

بررسی این جدول نشان می‌دهد که روش‌های کلاسیک مانند لوکاس کاناده (LK) و هورن شانک (HS) عملکرد ضعیفی دارند، زیرا در برابر تغییرات نوری و حرکت‌های پیچیده دقت لازم را ندارند.

همچنین در جدول 3 مقایسه روش‌های تخمین جریان نوری بر اساس معیارهای F1-bg (خطای ناحیه پس‌زمینه)، F1-fg (خطای ناحیه پیش‌زمینه) و F1-all (خطای کلی) آورده شده است. مدل‌های مبتنی بر یادگیری عمیق (FlowFormer، PWC-Net،RAFT) دقت بالایی دارند (خطای کلی کمتر) و نسبت به روش‌های سنتی بسیار سریع‌تر اجرا می‌شوند. مدل‌های سنتی (LK و HS) دارای خطای بالا و سرعت پردازش پایین هستند، که آن‌ها را برای کاربردهای بلادرنگ نامناسب می‌کند. مدل‌های ترکیبی (MirrorFlow، EpicFlow، DeepFlow) با ترکیب یادگیری عمیق و روش‌های سنتی، دقت تخمین را بهبود داده‌اند. با این حال، زمان اجرای آن‌ها همچنان بیشتر از مدل‌های کاملاً مبتنی بر CNN است. مدل‌های پیشرفته مانند FlowFormer و RAFT نه تنها دقت بالایی دارند، بلکه از نظر سرعت پردازش نیز بهینه‌تر هستند. این نشان‌دهنده پیشرفت‌های اخیر در تخمین جریان نوری با استفاده از یادگیری عمیق است.

جدول 3- مقایسه روش‌های تخمین جریان نوری

روش	معماری	زمان اجرا	Density	F1-all	F2-fg	F1-bg
FlowFormer	Transformer-based+self_Atention	0.3s	100%	4.68%	6.18%	4.37%
RAFT	Recuurent All-Pairs+GRU	0.2s	100%	5.10%	6.87%	4.74%
MaskFlowNet	CNN-based with MsKing	0.06s	100%	6.11%	7.70%	5.79%
LiteFlowNet3	Lightweight CNN	0.07s	100%	7.34%	7.75%	7.26%
LiteFlowNet2	Lightweight CNN	0.0486s	100%	7.62%	7.64%	7.62%
FlowNet2	Deep Leraning(Stacked FlowNet)	0.1s	100%	10.41%	8.75%	10.75%
PWC-Net	Hybrid(Traditional Stereo Matching+CNN)	0.03s	100%	9.60%	9.31%	9.66%
MirrorFlow	Hybrid(Classic+Deep Learning)	11m	100%	10.29%	17.07%	8.93%
EpicFlow	Hybrid(variational+Feature Matching)	17.3s	100%	26.29%	28.69%	25.81%
DeepFlow	Hybrid((variational+Feature Matching)	17s	100%	28.48%	31.06%	27.96%
SpyNet	Lightweight CNN+Coarse-to-Fine	150s	100%	35.07%	43.62%	33.36%
HS	Classic Optical Flow	2.6m	100%	41.81%	51.39%	39.90%
Pyramid_LK	Classic Optical Flow	2.5s	100%	72.67%	76.82%	71.84%

جدول 4- مقایسه روش های تخمین جریان نوری براساس معیار EPE

روش	سال	MPI-Sintel(Clean)	MPI-Sintel(Final)	KITTI 2012	KITTI 2015	پارامترها(میلیون)
Horn-schunck	1981	8.73	9.61	21.14	23.93	-
Brox et al	2004	5.1	7.2	-	-	-
DeepFlow	2013	5.38	7.21	14.12	15.88	-
EpicFlow	2015	4.12	6.29	12.95	13.84	-
FlowNetS	2015	7.28	8.81	18.79	21.07	-
FlowNet2	2017	3.96	6.02	10.41	11.11	162.4
SpyNet	2017	6.69	8.43	17.33	19.84	1.2
PWC-Net	2018	5.04	5.47	9.60	10.03	8.75
LiteFlowNet2	2019	3.48	4.69	7.10	7.95	6.42
LiteFlowNet3	2020	2.99	4.45	5.12	6.35	5.20
RAFT	2020	1.61	2.86	4.66	5.04	5.3
FlowFormer	2022	1.16	2.09	-	-	18.2
GMFlow	2022	1.29	2.19	-	-	7.5
VideoFlow	2023	0.991	1.649	-	-	13.5
DDVM	2023	1.01	2.40	-	-	-

بهینه‌سازی حجم ویدئو: جریان نوری در فشرده‌سازی ویدئو هم کاربرد دارد. این روش با بررسی تغییرات بین فریم‌ها، از ذخیره‌ی داده‌های تکراری جلوگیری می‌کند و باعث کاهش حجم فایل بدون افت کیفیت می‌شود.

تصویربرداری پزشکی و تحلیل حرکت در بدن: در MRI و سونوگرافی، جریان نوری می‌تواند حرکت اندام‌ها، تغییرات بافتی و حتی جریان خون را تحلیل کند. این قابلیت به تشخیص سریع‌تر مشکلات پزشکی کمک می‌کند. [33]

6- نتیجه‌گیری

تخمین جریان نوری یکی از مسائل اساسی در حوزه بینایی کامپیوتر محسوب می‌شود که در کاربردهای متعددی از جمله پردازش ویدئو، سامانه‌های ناوبری خودکار و تحلیل حرکت مورد استفاده قرار می‌گیرد. در این مقاله، روش‌های مختلف تخمین جریان نوری مورد بررسی قرار گرفت و مزایا و محدودیت‌های هریک تحلیل شد. روش‌های کلاسیک، به دلیل سادگی و عدم نیاز به داده‌های گسترده برای آموزش، همچنان در برخی زمینه‌ها کاربرد دارند. با این حال، این روش‌ها در شرایطی مانند تغییرات نوری شدید، انسداد و جابجایی‌های بزرگ، دقت کافی را ارائه نمی‌دهند. در مقابل، روش‌های مبتنی بر یادگیری عمیق، با بهره‌گیری از شبکه‌های عصبی پیچیده، بهبود چشمگیری در دقت تخمین جریان نوری ایجاد کرده‌اند، اما این دسته از مدل‌ها به داده‌های آموزشی حجیم و منابع محاسباتی بالا نیاز دارند که محدودیت‌هایی را در استفاده عملی آن‌ها ایجاد می‌کند.

یکی از چالش‌های اساسی در این حوزه، افزایش تعمیم‌پذیری مدل‌ها به داده‌های دنیای واقعی است. بسیاری از الگوریتم‌های یادگیری عمیق، در محیط‌های کنترل‌شده و بر روی مجموعه داده‌های مصنوعی آموزش می‌بینند که ممکن است منجر به کاهش دقت در شرایط واقعی شود. برای رفع این چالش، ترکیب روش‌های یادگیری نظارتی، نیمه‌نظارتی و بدون نظارت می‌تواند به توسعه مدل‌هایی با تعمیم‌پذیری بالاتر منجر شود. همچنین، اجرای الگوریتم‌های تخمین جریان نوری در زمان واقعی نیازمند بهینه‌سازی معماری‌های شبکه و کاهش پیچیدگی محاسباتی است. توسعه مدل‌های سبک‌تر و کارآمدتر می‌تواند به افزایش سرعت پردازش و کاهش هزینه‌های محاسباتی کمک کند.

روش‌های مبتنی بر بهینه‌سازی مانند Brox et al (2004) و EpicFlow دقت بیشتری ارائه داده‌اند، اما هزینه پردازشی آن‌ها بالاست. با ورود مدل‌های یادگیری عمیق مانند FlowNet، RAFT و FlowFormer دقت تخمین جریان نوری بهبود یافته است. به‌ویژه GMFlow و FlowFormer که از معماری ترنسفورمر استفاده می‌کنند، توانسته‌اند تعادل بهتری بین دقت و هزینه محاسباتی برقرار کنند. مدل‌های جدیدتر مانند VideoFlow کمترین میزان خطا را نشان می‌دهند، که نشان‌دهنده پیشرفت این حوزه در کاهش خطای تخمین و بهبود تعمیم‌پذیری مدل‌ها است. با وجود پیشرفت‌های حاصل‌شده، همچنان چالش‌هایی مانند مدیریت انسداد، تغییرات نوری شدید، جابه‌جایی‌های بزرگ و کاهش هزینه‌های محاسباتی نیاز به بهبود دارند. توسعه روش‌هایی که بتوانند تعادل مناسبی میان دقت، سرعت پردازش و تعمیم‌پذیری در شرایط مختلف ایجاد کنند، همچنان از موضوعات مهم پژوهشی در این حوزه محسوب می‌شود.

5- کاربردهای جریان نوری

جریان نوری یکی از روش‌های پرکاربرد در پردازش تصویر است که در حوزه‌های مختلف مانند بینایی کامپیوتر، تحلیل ویدئو، واقعیت افزوده و تصویربرداری پزشکی استفاده می‌شود. برخی کاربردها عبارتند از:

ردیابی حرکت در ویدئوها: بسیاری از سیستم‌های هوشمند برای تشخیص و دنبال کردن مسیر حرکت اشیاء از جریان نوری استفاده می‌کنند. برای مثال، در یک مسابقه فوتبال، این روش می‌تواند حرکت بازیکنان را تحلیل کند یا در سامانه‌های کنترل ترافیک، مسیر خودروها را ردیابی کند.

بهبود پایداری ویدئو: هنگام ضبط ویدئو، لرزش‌های دوربین می‌توانند باعث کاهش کیفیت تصویر شوند. الگوریتم‌های جریان نوری، این حرکات ناگهانی را تشخیص داده و آن‌ها را اصلاح می‌کنند تا ویدئو پایدارتر و روان‌تر نمایش داده شود.

فناوری‌های واقعیت افزوده (AR): در برنامه‌های واقعیت افزوده، حرکت دوربین یا کاربر باید به‌درستی شناسایی شود تا اشیاء مجازی به‌صورت همگام با محیط واقعی نمایش داده شوند. این فناوری در بازی‌های موبایلی، فیلترهای تصویری و راهنمایی‌های مجازی استفاده می‌شود.

[9] Z. Teed and J. Deng, "RAFT: Recurrent all-pairs field transforms for optical flow," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020

[10] Z. Huang, X. Shi, C. Zhang, Q. Wang, K. C. Cheung, H. Qin, J. Dai, and H. Li, "FlowFormer: A transformer architecture for optical flow," in *Proc. 17th Eur. Conf. Comput. Vis. (ECCV)*, Tel Aviv, Israel, Oct. 2022, pp. 668–685

[11] H. Xu, J. Zhang, Z. Zhang, C. Lu, W. Luo, "GMFlow: Learning Optical Flow via Global Matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 8121–8130

[12] A. Ranjan and M. J. Black, "Optical flow estimation using a spatial pyramid network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017

[13] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018.

[14] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "DeepFlow: Large displacement optical flow with deep matching," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2013, pp. 1385–1392.

[15] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "EpicFlow: Edge-Preserving Interpolation of Correspondences for Optical Flow," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015.

[16] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2016.

[17] A. Ahmadi and I. Patras, "Unsupervised convolutional neural networks for motion estimation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2016.

[18] J. J. Yu, A. W. Harley, and K. G. Derpanis, "Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2016.

[19] S. Meister, J. Hur, and S. Roth, "UnFlow: Unsupervised learning of optical flow with a bidirectional census loss," in *Proc. AAAI Conf. Artif. Intell.*, 2018.

[20] P. Liu, M. R. Lyu, I. King, and J. Xu, "SelfFlow: Self-supervised learning of optical flow," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019.

[21] P. Liu, M. R. Lyu, I. King, and J. Xu, "DDFlow: Learning optical flow with unlabeled data

با وجود پیشرفت‌های اخیر، مسیرهای متعددی برای تحقیقات آینده در این حوزه وجود دارد. ترکیب روش‌های کلاسیک و یادگیری عمیق می‌تواند منجر به ارائه مدل‌هایی شود که ضمن حفظ دقت، کارایی بالاتری در محیط‌های واقعی داشته باشند. علاوه بر این، توسعه مجموعه داده‌های جامع‌تر و متنوع‌تر که طیف وسیع‌تری از سناریوهای دنیای واقعی را پوشش دهند، می‌تواند در بهبود عملکرد مدل‌ها تأثیر بسزایی داشته باشد. پیشرفت در این زمینه نه تنها موجب ارتقای دقت روش‌های تخمین جریان نوری خواهد شد، بلکه در حوزه‌های مرتبط همچون سامانه‌های خودران، تحلیل ویدئو و فناوری‌های پردازش تصویر نیز تأثیر قابل توجهی خواهد داشت.

مراجع

[1] J. J. Gibson, *The Perception of the Visual World*, Boston, MA, USA: Houghton Mifflin, 1950.

[2] B. K. P. Horn, "Determining optical flow," *Artif. Intell.*, vol. 17, no. 1-3, pp. 185-203, 1981.

[3] T. S. Shah and S. Haider, "Traditional and modern strategies for optical flow: An investigation," *SN Appl. Sci.*, vol. 3, no. 742, 2021.

[4] A. Singh, "Enhancement of the Horn and Schunk optic flow algorithm by adaptive smoothing and region merging," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 1991.

[5] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artif. Intell. (IJCAI)*, Vancouver, BC, Canada, Aug. 1981, pp. 674–679.

[6] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2004, pp. 25-36.

[7] A. Dosovitskiy, P. Fischer, E. Ilg, P. Haeusser, C. Hazirbas, V. Golkov, P. Van der Smagt, D. Cremers, and T. Brox, "FlowNet: Learning optical flow with convolutional networks," *arXiv:1504.06852[cs.CV]*, 2015

[8] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2462–2470

- [33] A. Alfarano, L. Maiano, L. Papa, and I. Amerini, "Estimating optical flow: A comprehensive review of the state of the art," *Comput. Vis. Image Underst.*, vol. 249, p. 104160, 2024.
- distillation," in *Proc. 33rd AAAI Conf. Artif. Intell. (AAAI)*, Honolulu, HI, USA, Jan. 2019, pp. 8770–8777
- [22] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2012.
- [23] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open-source movie for optical flow evaluation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2012.
- [24] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "PatchMatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph. (SIGGRAPH)*, vol. 28, no. 3, p. 24, 2009.
- [25] C. Bailer, B. Taetz, and D. Stricker, "Flow Fields: Dense correspondence fields for highly accurate large displacement optical flow estimation," arXiv:1508.05151 [cs.CV], 2015.
- [26] Q. Chen and V. Koltun, "Full Flow: Optical flow estimation by global optimization over regular grids," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 4706–4714.
- [27] X. Shi, Z. Huang, W. Bian, D. Li, M. Zhang, K. C. Cheung, S. See, H. Qin, J. Dai, and H. Li, "VideoFlow: Exploiting temporal cues for multi-frame optical flow estimation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2023, pp. 12469–12480.
- [28] S. Zhao, Y. Sheng, Y. Dong, E. I.-C. Chang, and Y. Xu, "MaskFlowNet: Asymmetric feature matching with learnable occlusion mask," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 6278–6287.
- [29] J. Hur and S. Roth, "MirrorFlow: Exploiting symmetries in joint optical flow and occlusion estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 312–321.
- [30] D.-H. Trinh and C. Daul, "On illumination-invariant variational optical flow for weakly textured scenes," *Comput. Vis. Image Underst.*, vol. 179, pp. 1–18, 2019.
- [31] X. Liao, Z. Cai, J. Chen, T. Liu, J. H. Lai, and H. Li, "Physics-based optical flow estimation under varying illumination conditions," *Signal Process. Image Commun.*, vol. 117, p. 117007, 2023.
- [32] T.-W. Hui, C. C. Loy, and X. Tang, "LiteFlowNet3: Resolving correspondence ambiguity for more accurate optical flow estimation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 169–185