TRACKING OF MULTIPLE MOVING FISHES IN MEDIUM AND HIGH DENSITY UNDERWATER SCENARIO

SHILPI GUPTA

BHARTI SCHOOL OF TELECOMMUNICATION TECHNOLOGY & MANAGEMENT

INDIAN INSTITUTE OF TECHNOLOGY DELHI

SEPTEMBER, 2024

TRACKING OF MULTIPLE MOVING FISHES IN MEDIUM AND HIGH DENSITY UNDERWATER SCENARIO

by

SHILPI GUPTA

BHARTI SCHOOL OF TELECOMMUNICATION TECHNOLOGY AND MANAGEMENT

Submitted

In fulfilment of the requirements of the degree of

Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

SEPTEMBER, 2024

Certificate

This is to certify that the thesis titled Tracking of Multiple Moving Fishesin Medium and High Density Underwater Scenario, submitted byShilpi Gupta, to the Indian Institute of Technology, Delhi, for the awardof the degree of Doctor of Philosophy, is a bonafide record of the researchwork done by her under our supervision. The contents of this thesis, in full orin parts, have not been submitted to any other Institute or University for theaward of any degree or diploma.

Professor Santanu Chaudhury
Dept. of Electrical Engineering
IIT-Delhi

Professor Brejesh Lall
Dept. of Electrical Engineering
IIT-Delhi

Acknowledgements

First and foremost, I would like to thank God, the Almighty, who has grantedme countless blessings, so that I have been finally able to accomplish the thesis.I would like to express my sincere gratitude to Prof. Santanu Chaudhury,Director, IIT Jodhpur for giving me an opportunity to work in the area ofunderwater computer vision. I am extremely grateful to Prof. Brejesh Lall, IITDelhi for his valuable guidance with understanding, continual encouragement,

and intellectual support during the completion of this thesis. There weremoments when I gave up it was Brejesh Sir who motivated me. Sir alwaysused to say “Life is too short to be sad”, this is going to be lifelong learningfor me. I am fortunate to work under his supervision.

I would like to thank my committee members, Prof. Prem Kalra, Prof.Tapan Gandhi, and Prof. Ranjan Bose for their precious time and valuablefeedback. I am thankful to Dr. Prerana Mukherjee, JNU for her unwaveringefforts, helping me during the late night working hours and inspiring me to realize my goals. Without her support, this thesis would not have been possible.

My Special thanks are extended to Dr. Chetan Ralekar, Postdoctoral ResearchAssociate at Massachusetts Institute of Technology for always supporting mewhen I needed help.I am thankful to my husband Mr. Praval Gupta for enriching my life andwidening my horizons with his patience, care, companionship, empathy, and love. I feel indebted to my daughter Prashi for giving me the necessary distractions from work, bringing balance and richness to my doctoral experience.

My Ph.D. journey will be incomplete if I do not mention my friends Parul,Jaya, Shweta, Devanshu, Preeti, and Anand, who always stood by me throughthick and thin. I would like to express my deepest gratitude to my parentsand sister for making me what I am today. I can never thank them enough.Last but not least, I am thankful for the support received from all my labmates and lab staff.

Shilpi Gupta

Abstract

Due to remarkable progress in deep learning techniques, computer vision hasgained increasing attention in research. The maximum work of this area is foron-ground scenarios which are either human-specific or vehicle specific. Whilewater covers around 70 percent of the earth's surface and Fishes are an inseparable part of water. In addition to that, Fish is a major food source forhumans, a living hood for many people belonging to the fishing industry, acontributor to the economy with fish farming, and an indicator of environmental change. Fish is of great importance for both the fish farming industry andmarine researchers. Sensor-based techniques and underwater acoustic information is used for underwater research. these methods are invasive, expensive,and difficult to deploy. However, underwater cameras are non-invasive, cost-effective, and can record continuous data. In view of this, I investigate in this thesis, computer vision for underwater videos focusing on fish.

In the first approach, we propose a refined optical flow estimation method.Estimating motion in underwater videos is challenging due to low contrast,highly cluttered background, dynamic background, and illumination change.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 motionestimation. The proposed method has been evaluated on challenging real-life image sequences of the Fish4Knowledge database. In the next approach, wesolved a highly challenging task of multiple fish tracking as fish movements arehighly erratic in nature. Trajectories of fishes convey critical information forthe analysis of fish behaviour. We have proposed DFTNet which incorporatesSiamese network for encoding the appearance similarity and attention-longshort term memory network to capture the motion similarity across subsequent frames. Finally, intersection over union matching score is computed toamalgamate spatial similarity cue in the final score. The proposed frameworkprovides a joint optimization score to maintain the tracklet information encoding appearance, motion, and spatial similarity cues. Motion detection andmultiple fish tracking are mid-level computer vision tasks, which can providethe basis for other high-level computer vision tasks like activity, intention, and behavior recognition of fish. Fishes often tend to show collective behavior. They prefer to move into a group. Identifying fish groups and their profilingis a major problem in underwater computer vision areas. In the existing literature, fish grouping behavior has been explained in many ways, but there isa lack of explanations supported by empirical research using video analyticsmethods. Analysis and comparison of work in this area is made difficult in the absence of data sets with ground truth of fish groupings. To

bridge thisgap, we have developed an Underwater Crowd Analysis Dataset UWCAD andgroup annotation tool GAT in chapter 5. We also propose a motion featureextraction method GGKLT and underwater temporal autoencoder for groupdetection UTAG by clustering the learned trajectory features by temporalautoencoder and global trajectory features.

सारांश

गहन शिक्षण तकनीकों में उल्लेखनीय प्रगति के कारण, कंप्यूटर विज़न ने अनुसंधान में अधिक ध्यान आकर्षित किया है। इस क्षेत्र का अधिकतम कार्य जमीनी परिदृश्यों पर है जो या तो मानव-विशिष्ट या वाहन-विशिष्ट हैं। जबकि पानी पृथ्वी की सतह का लगभग 70 प्रतिशत भाग कवर करता है और मछलियाँ पानी का एक अविभाज्य हिस्सा हैं। इसके अलावा, मछली मनुष्यों के लिए एक प्रमुख खाद्य स्रोत है, मछली पकड़ने के उद्योग से जुड़े कई लोगों के लिए आजीविका है, मछली पालन के साथ अर्थव्यवस्था में योगदानकर्ता है, और पर्यावरण परिवर्तन का एक संकेतक है। मछली पालन उद्योग और समुद्री शोधकर्ताओं दोनों के लिए मछली का बहुत महत्व है। पानी के नीचे अनुसंधान के लिए सेंसर-आधारित तकनीकों और पानी के नीचे ध्वनिक जानकारी का उपयोग किया जाता है। ये विधियां आक्रामक, महंगी और लागू करने में कठिन हैं। हालाँकि, पानी के नीचे के कैमरे गैर-आक्रामक, लागत प्रभावी हैं और निरंतर डेटा रिकॉर्ड कर सकते हैं। इसे ध्यान में रखते हुए, मैं इस थीसिस में मछली पर ध्यान केंद्रित करने वाले पानी के नीचे के वीडियो के लिए कंप्यूटर विज़न की जांच करता हूं

पहले दृष्टिकोण में, हम एक परिष्कृत ऑप्टिकल प्रवाह अनुमान विधि का प्रस्ताव करते हैं। कम कंट्रास्ट, अत्यधिक अव्यवस्थित पृष्ठभूमि, गतिशील पृष्ठभूमि और रोशनी परिवर्तन के कारण पानी के नीचे के वीडियो में गति का अनुमान लगाना चुनौतीपूर्ण है। हमारे दृष्टिकोण का केंद्र अधिकांश गति के रूप में समोच्च जानकारी का दोहन करना है किनारों पर पड़ा है. इसके अलावा, हमने इसे विरल से सघन गति अनुमान के रूप में तैयार किया है। प्रस्तावित विधि का मूल्यांकन चुनौतीपूर्ण वास्तविक जीवन पर किया गया है

फिश4नॉलेज डेटाबेस के छवि अनुक्रम। अगले दृष्टिकोण में, हमने एकाधिक मछली ट्रैकिंग के अत्यधिक चुनौतीपूर्ण कार्य को हल किया क्योंकि मछली की गतिविधियाँ प्रकृति में अत्यधिक अनियमित होती हैं। मछलियों के प्रक्षेप पथ मछली के व्यवहार के विश्लेषण के लिए महत्वपूर्ण जानकारी देते हैं। हमने डीएफटीनेट का प्रस्ताव दिया है जिसमें उपस्थिति समानता को एन्कोड करने

के लिए सियामी नेटवर्क और बाद के फ्रेमों में गति समानता को पकड़ने के लिए ध्यान-दीर्घकालिक मेमोरी नेटवर्क शामिल है। अंत में, संघ मिलान स्कोर पर प्रतिच्छेदन की गणना अंतिम स्कोर में स्थानिक समानता क्यू को समाहित करने के लिए की जाती है। प्रस्तावित ढांचा ट्रैकलेट सूचना एन्कोडिंग उपस्थिति, गति और स्थानिक समानता संकेतों को बनाए रखने के लिए एक संयुक्त अनुकूलन स्कोर प्रदान करता है। मोशन डिटेक्शन और मल्टीपल फिश ट्रैकिंग मध्य-स्तरीय कंप्यूटर विज़न कार्य हैं, जो मछली की गतिविधि, इरादे और व्यवहार की पहचान जैसे अन्य उच्च-स्तरीय कंप्यूटर विज़न कार्यों के लिए आधार प्रदान कर सकते हैं।

मछलियाँ अक्सर सामूहिक व्यवहार दिखाती हैं। वे समूह में रहना पसंद करते हैं। पानी के भीतर कंप्यूटर दृष्टि क्षेत्रों में मछली समूहों की पहचान करना और उनकी रूपरेखा बनाना एक बड़ी समस्या है। मौजूदा साहित्य में, मछली समूहन व्यवहार को कई तरीकों से समझाया गया है, लेकिन वीडियो विश्लेषण विधियों का उपयोग करके अनुभवजन्य अनुसंधान द्वारा समर्थित स्पष्टीकरण की कमी है। मछली समूहों की जमीनी सच्चाई के साथ डेटा सेट के अभाव में इस क्षेत्र में काम का विश्लेषण और तुलना करना मुश्किल हो गया है। इस अंतर को पाटने के लिए, हमने अध्याय 5 में एक अंडरवाटर क्राउड एनालिसिस डेटासेट यूडब्ल्यूसीएडी और ग्रुप एनोटेशन टूल जीएटी विकसित किया है। हम टेम्पोरलऑटोएनकोडर और वैश्विक प्रक्षेपवक्र सुविधाओं द्वारा सीखे गए प्रक्षेपवक्र सुविधाओं को क्लस्टर करके ग्रुपडिटेक्शन यूटीएजी के लिए एक मोशन फीचरएक्सट्रैक्शन विधि जीजीकेएलटी और अंडरवाटर टेम्पोरल ऑटोएनकोडर का भी प्रस्ताव करते हैं।

Contents

List of Figures

List of Tables