WaveShot:
A Compact Portable Unmanned Surface Vessel for Dynamic Water Surface Videography and Media Production

Shijian Ma[*], ShiCong Ma[*], WeiZe Ma[1]
[*]project co-leads
https://wave-shot.github.io

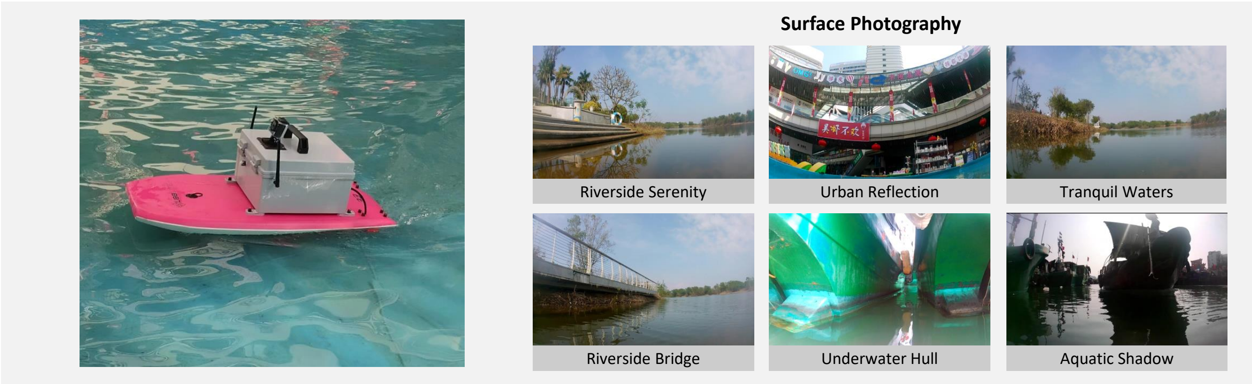


Figure 1: ***WaveShot*** . We introduce a compact portable unmanned surface vessel for dynamic water surface videography and media production. *Left:* WaveShot is sailing in the pool with a stable posture and recording videos of the water surface. *Right:* WaveShot can perform water surface filming tasks in various water environments.

Abstract

This paper presents WaveShot, an innovative portable unmanned surface vessel that aims to transform water surface videography by offering a highly maneuverable, cost-effective, and safe alternative to traditional filming methods. WaveShot is specially designed for the modern demands of film production, advertising, documentaries, and visual arts, equipped with professional-grade waterproof cameras and advanced technology to capture both static and dynamic scenes on waterways. We discuss the development and advantages of WaveShot, highlighting its portability, ease of transport, and rapid deployment capabilities. Experimental validation that is showcasing WaveShot's stability and high-quality video capture in various water conditions, and the integration of monocular depth estimation algorithms to enhance the operator's spatial perception. The paper concludes with an exploration of WaveShot's real-world applications, its user-friendly remote operation, and future enhancements such as gimbal integration and advanced computer vision for optimized videography on water surfaces.

1. Introduction

In film production, advertising shoots, documentary filmmaking, and various visual arts projects, the demand for shooting scenes on water is increasing. These scenes provide audiences with profound impressions due to their unique visual effects and emotional expression. However, water-based shooting faces numerous challenges, including the unpredictability of the shooting environment, the high demands on professional equipment and teams, and the complexity of ensuring the safety of all participants. Traditionally, these challenges have been overcome by using small boats, professional filming equipment, and experienced staff, which not only increases the cost of shooting but also limits its flexibility and accessibility.

With the advancement of technology and the push for innovative solutions, finding more efficient and flexible methods for water-based shooting has become particularly important. In this context, the development of portable water shooting boats has emerged. This new type of filming tool can provide a more flexible and cost-effective solution to meet the changing shooting demands and creative visions. Portable water shooting boats simplify the operation process, reduce reliance on professional personnel, and enhance the adaptability of the shooting environment, allowing film crews to capture the desired shots more freely in various aquatic environments. Additionally, the design of these boats aims to maximize safety during the shooting process while minimizing environmental impact. In this paper, we propose a portable water shooting boat capable of performing various water-based tasks in natural water bodies and expanding the shooting effects of water scenes by incorporating monocular depth estimation technology.

With the development of unmanned technology, small unmanned boat platforms have shown great potential in the field of water-based shooting, especially in hard-to-reach water areas or situations requiring low-cost operations. These platforms are usually equipped with professional waterproof cameras and theoretically can meet various water-based shooting needs, providing photographers and filmmakers with unprecedented shooting freedom and innovation opportunities. However, despite significant technological progress, unmanned boat platforms still face some key limitations in practical applications, particularly in terms of portability.

Portability refers to the convenience of carrying and transporting equipment, which is crucial in fast-paced and ever-changing shooting environments. An ideal portable water shooting platform should be easy to deploy and move between different shooting locations. However, many unmanned boat platforms currently available on the market fail to meet this requirement well. For example, some larger unmanned boats, such as catamarans with a length of up to 2.5 meters, while able to carry a large amount of equipment and provide a stable shooting platform, are limited in portability due to their size. Such boats cannot be easily fitted into ordinary cars and require specialized trucks for transportation, significantly increasing the complexity and cost of shooting.

On the other hand, there are also smaller unmanned boats available on the market, such as those around 1.1 meters in length, which are theoretically more portable. Although these small unmanned boats can solve the problem of transportation size, allowing them to be transported in ordinary cars, their weight is usually over 10 kilograms, meaning that at least two people are needed to carry them, thereby limiting their portability as well. Under the requirements of rapid deployment and shooting in multiple locations, such weight and the demand for carrying personnel become significant obstacles.

In the process of exploring solutions to the challenges faced by existing small unmanned boat platforms in terms of portability, we have developed an innovative solution, WaveShot. The design concept of WaveShot aims to fundamentally change the way water shooting is done by significantly reducing size and weight through optimized hardware design, thereby providing unprecedented portability. WaveShot's design takes into account the complexity and unpredictability of the actual shooting environment on the water. Users can easily lift WaveShot with one hand and place it in the water without complex preparation or teamwork. Operating WaveShot is equally simple, allowing users to achieve precise control and flexible shooting whether through ground stations or remote controllers. Through its differential propulsion system, onboard sensing and processing capabilities, remote operation functionality, and seamless integration with action cameras, WaveShot can achieve smooth and dynamic tracking shooting on the water surface in Figure 1. This not only improves the stability of capturing water scenes but also provides novel perspectives and dynamic effects for shooting.

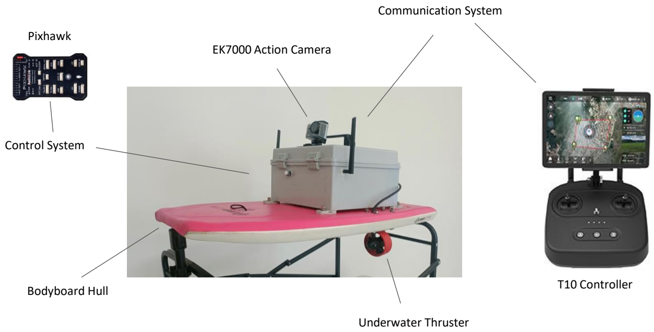
The main contribution of this paper is to introduce an innovative research and development outcome aimed at advancing water shooting technology, providing a new perspective and method for capturing dynamic and static water targets. Our work focuses on two main contributions:

1. Development of WaveShot: We designed and implemented WaveShot, an unmanned boat platform specifically designed for portable water target shooting. The design concept of WaveShot is based on improving the flexibility and convenience of the shooting process while ensuring that the shooting quality is not compromised. By optimizing its size and weight, WaveShot can be easily carried by individuals and deployed in various water environments, from calm lakes to dynamic rivers, quickly adapting to and executing shooting tasks.

2. Experimental validation and technological innovation: We validated the effectiveness of WaveShot in shooting static and moving water targets through a series of experiments. The experimental results show that WaveShot can provide stable and clear video capture under various water conditions, demonstrating its reliability and efficiency as a water shooting tool. Additionally, this paper introduces the use of the latest monocular depth estimation algorithms to enhance the videos shot by WaveShot, adding depth information to water scenes and providing a new perspective. This enhances the operator's spatial perception of water scenes.

2. Related Work

Unmanned Surface Vessels. Unmanned Surface Vessels (USVs) have attracted increasing research interest in recent years due to their potential advan-



Dimensions	100 x 35 x 20 cm
Weight	8.5 kg
Material	HDPE, PE
Color	Pink
Speed	Max 6 knots
Range	10 km
Wave Height	<0.5 m
Navigation	GPS + IMU
Comms Range	900 m

Figure 2: ***Hardware Details.*** *Left:* WaveShot consists of a bodyboard hull, control system, communication system, underwater thruster, and an action camera on the head. *Middle:* Manual operation is set up using the SKYDROID T10 Controller combined with a mobile ground station, allowing for precise navigation of WaveShot within a range of 900 meters. *Right:* Technical specifications of WaveShot.

tages in efficiency, cost, and safety. A significant amount of work has been done in the development of USV platforms and core technologies.

Many studies have focused on designing USV platforms for different applications. Bibuli et al. [2] developed an autonomous catamaran named Charlie for environmental monitoring tasks. Campbell et al. [3] proposed a USV called Measuring Dolphin for bathymetric measurements. The low-cost USV named WeMo, developed by Madeo et al. [13], performs real-time long-distance water quality monitoring tasks using a LoRa wireless network. For research on autonomous solar USVs as launch and recovery platforms for UAVs, Aissi et al. [1] proposed an innovative concept design that allows for autonomous launch and recovery of UAVs, as well as automatic battery charging in an unmanned mode, providing a viable solution for long-term maritime operations.

Some research has been dedicated to studying navigation, guidance, and control methods for USVs. Do et al. [5] modeled the motion dynamics of USVs and developed path-tracking controllers. Lee et al. [11] proposed a fuzzy logic-based navigation and collision avoidance method for USVs. Wei et al. [4] introduced a local obstacle avoidance method for USVs based on vector field histogram. Liu and Bucknall [12] optimized USV paths using variable spatio-temporal search while complying with maritime rules.

Despite extensive research and development in USVs, there is a noticeable gap in their application for professional filming tasks in aquatic environments. Current USV platforms mainly focus on environmental monitoring, measurements, and safe operations, with less attention to applications specifically for filming purposes using action cameras. To bridge this gap, we introduce WaveShot, a compact and portable USV specially designed for shooting on water using waterproof action cameras. WaveShot aims to leverage advancements in USV technology to provide a unique platform for capturing high-quality video footage in aquatic environments, marking a new application of USVs in the fields of cinematography and media production. By focusing on this domain, WaveShot aims to expand the utility of USVs beyond traditional applications and demonstrate the potential for innovative uses across various industries.

Monocular Depth Estimation. Monocular Depth Estimation (MDE) technology has achieved significant development in the field of computer vision and robotics research. Initially, this field relied on manually designed features and probabilistic graphical models for depth inference. Saxena et al.[15] adopted a supervised learning approach, using Markov Random Fields (MRF) to predict depth information from single images, training their model based on local and global features extracted from the images. Subsequently, with the rise of deep learning technology, Convolutional Neural Networks (CNNs) have gradually become the core technology for monocular depth estimation. Eigen and colleagues[6] developed a multi-scale CNN framework that improves depth prediction by learning from coarse to fine scales, and their approach achieved better results on standard datasets compared to traditional techniques.

To enhance the accuracy and efficiency of monocular depth estimation, researchers have been actively exploring the application of deep learning. For instance, Laina et al.[10] designed an advanced fully convolutional residual network (FCRN) that demonstrated outstanding performance across various scenes. They also introduced a novel reverse Huber loss function, which helps to accommodate a wide range of depth variations and promote smooth depth transitions. Unsupervised learning methods have also gained interest as they do not require true depth labels for training. Garg and his team[7] demonstrated an unsupervised learning framework that trains a CNN to predict depth by minimizing the photometric reconstruction error of stereo image pairs. Their method was strengthened by integrating left-right view consistency constraints[8] with temporal consistency constraints[17]. To improve the performance of monocular depth estimation across different scenes, researchers have also studied transfer learning and domain adaptation techniques. Godard et al.[9] proposed a self-supervised learning strategy that pre-trains the depth estimation model on a large stereo image dataset, then fine-tunes it on a target monocular image sequence, showing excellent performance in dealing with differences between datasets. Nonetheless, adapting MDE models to unknown domains remains a challenge.

The goal of zero-shot depth estimation is to train models on a diverse set of datasets so that they can predict the depth of any given image regardless of domain differences accurately. Innovative attempts in this direction have focused on expanding training datasets, guiding models through increasing the number of image samples and sparse supervision on limited key points. One of the milestone achievements in this area is the MiDaS [14] project, which introduced an affine-invariant loss function, effectively mitigating the issue of depth scale and shift differences between datasets, and achieving efficient cross-dataset joint training. In this work, we leverage the Depth Anything [16] zero-shot depth estimation model from MDE technology, combined with videos captured on the water surface by WaveShot, to generate multiple consecutive depth-estimated image sequences. This application breaks through the limitations of unmanned boats for water surface photography, showcasing the practical utility of advanced MDE models in real aquatic scenarios, providing new perspectives and capabilities for the field of water surface photography.

3. WaveShot Hardware

We have developed WaveShot, a portable waterborne filming boat, for capturing footage and executing filming tasks in various aquatic environments. WaveShot features low cost, flexibility, and simple automatic filming capabilities, integrating various AI visual technologies into the filmed videos. Specifically, we have incorporated four key design considerations:

1. **Maneuverability**: The system can move at approximately 2m/s surfing speed.
2. **Stability**: It maintains stability even under windy and wavy conditions.
3. **Full-body remote operation**: The boat body and camera, can be remotely controlled.
4. **Portability**: The design is ensuring easy transportation and quick deployment.

As illustrated in Figure 2, we chose a bodyboard as the hull for the boat. A bodyboard is a wave-riding board shorter than a surfboard, typically around 3 feet (1 meter) in length. Made of foam, it provides good buoyancy and stability on the water. Designed to support a person floating on the water, it ensures even distribution of buoyancy across its surface, maintaining stability even in complex water conditions or when carrying heavy loads. Lightweight materials such as high-density polyethylene (HDPE) and polyethylene (PE) provide buoyancy without adding excessive weight. The shape and contour of the bodyboard are designed to effectively cut through water, enhancing stability and maneuverability. We found that the bodyboard can withstand prolonged exposure to seawater and sunlight, which is advantageous for unmanned vessels operating in harsh environments. It is very lightweight, aiding in transportation and maneuverability, making efficient propulsion systems possible. Modifying and installing additional equipment on the hull is straightforward.

Next, we designed a waterproof control system on the bodyboard, capable of simultaneously controlling two external underwater thrusters and multiple sensors. This design choice is crucial for ensuring WaveShot's stability in water environments along predefined routes during filming. We installed a flip-top waterproof box on the boat's body to house the control board and sensor components. The selected control board is the open-source Pixhawk hardware, containing a high-performance microprocessor and sensors, including gyros, accelerometers, magnetometers, and barometers, ensuring precise navigation control and stability. We flashed Ardupilot firmware to support autonomous navigation of differential boats. To achieve autonomous navigation in various bodies of water, we integrated a GPS module onto Pixhawk, providing accurate geographic location information critical for route planning and heading control.

The communication system in the WaveShot control system ensures a stable connection between the boat and the ground station, allowing for transmission of control commands and reception of boat status information. This communication system includes radio transmission devices, antennas, and ground stations. We externally mounted two antennas on the control box of the boat to enhance signal reception and transmission, with a stable communication range of up to one kilometer during water tests. We used Mission Planner ground station software on the computer and QGroundControl ground station software on the phone to display real-time data of the boat, map positioning, flight planning, and monitor boat status, adjusting task parameters as needed. Additionally, we installed a receiver paired with a remote controller for manual operation.

Its power system is the core part of maneuvering and controlling the boat. We mounted two underwater thrusters on the bottom of the boat, using two underwater brushless thrusters as power sources, a configuration known as a differential propulsion system. The differential propulsion system achieves the boat's forward, backward, turning, and stopping movements by independently controlling the speeds of the two thrusters, providing good maneuverability and control accuracy. By simultaneously increasing or decreasing the speeds of the two thrusters, the boat can move forward or backward. The magnitude of the speed determines the boat's velocity. By changing the speed difference between the two thrusters, the boat can turn left or right. For example, increasing the speed of the right thruster and decreasing the speed of the left thruster will turn the boat to the left. The brushless thrusters operate fully submerged in water, with heat generated during operation dissipating through water cooling.

Above the control box, we installed an EK7000 action camera with a waterproof housing, capable of recording ultra-high-definition 4K videos and waterproof functionality in Figure 2. It also comes with a remote control and WiFi functionality for remote wireless video recording start-up and transmission, allowing us to collect video data accurately on demand and share it with others. Adjusting WaveShot's perspective and trajectory based on video feedback to improve filming quality. Some example tasks that WaveShot combined with computer vision technology can accomplish include:

- **Water scene filming:** WaveShot can capture stationary objects such as anchored boats, water buoys, or waterfront landscapes in natural water conditions. It is used to capture details of static objects such as textures of surrounding plants, floating objects on the water surface, or rocks by the water's edge. Additionally, WaveShot can identify and track rapidly moving objects such as kayaks, water scooters, or swimmers from a distance and follow the targets to obtain clear dynamic shots.
- **Human-robot interactions:** By processing images captured by the portable waterborne device, depth videos are generated through the "Depth Anything" project. This includes not only visual information of regular videos but also adds depth information to each pixel, providing additional dimensions and insights for various applications. It provides crucial spatial information to help identify potential obstacles and assess the navigability of rescue areas.

4. WaveShot Preprocessing

4.1. Environmental Assessment and Manual Mode Adjustment

Before undertaking underwater photography tasks, conducting a comprehensive environmental assessment of the target water body is crucial. This preliminary step involves careful observation to identify potential entanglements or obstacles (such as aquatic vegetation, debris) that may hinder the operation of the WaveShot equipment. Identifying these hazards is key to ensuring the safe operation and smooth functioning of the equipment. Once these obstacles are detected, the WaveShot is switched to manual mode. This allows for direct control of the device, enabling the operator to manually navigate the WaveShot to several pre-determined safe waypoints. These waypoints are carefully recorded for subsequent processing.

4.2. Navigation Planning

The recorded waypoints, indicating safe navigation paths, are inputted into the ground station's planning module. Through a series of processing steps within this module, a customized navigation route is designed for the WaveShot. This route is specifically tailored to optimize the waterborne photography mission, taking into account environmental conditions and the identified safe waypoints, ensuring smooth and obstacle-free operation throughout the photography mission.

4.3. Optimization of Shooting Angles

A unique feature of the WaveShot is its multifunctional action camera, which can adjust its shooting angle from downward to upward relative to the device's head. Initial short-term test shots are taken to fine-tune this angle, aiming to achieve a balanced composition of footage. Adjustments are

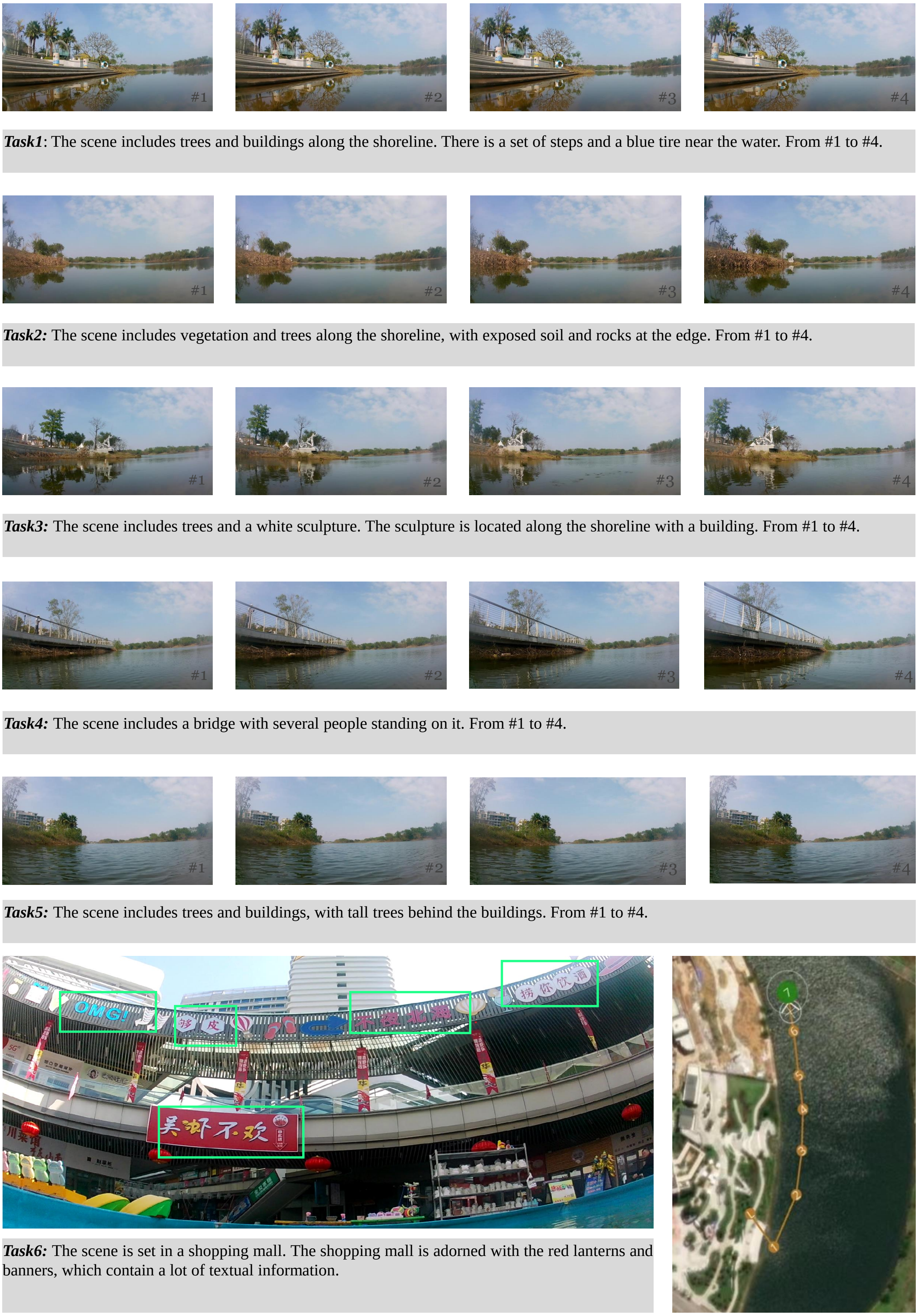


Figure 3: ***Task Definitions.*** We demonstrated six tasks designed to test WaveShot's ability to film static targets on the water. The waypoint map in the bottom right corner shows the routes WaveShot follows to execute each task according to preset waypoints. For each task, we described the objective in accordance with the sequence of images.

made based on the proportion of vessels and sky in the initial footage; if vessels dominate, the camera tilts backward, while if the sky predominates, it tilts forward. The goal is to achieve an angle that effectively captures the waterborne scenes, ensuring they are recorded effectively.

4.4. Speed Optimization for Video Shooting Efficiency

The speed of the WaveShot is another critical factor that needs careful consideration, especially when tracking moving waterborne targets. Excessive speed may cause the target to move out of range, thereby losing it from the camera's field of view, while too slow a speed, although advantageous for capturing clear and stable footage, may affect the overall efficiency of the waterborne photography mission. Therefore, a series of speed tests are conducted to find the optimal cruising speed that maintains clear and stable footage while efficiently covering the waterborne environment.

Through these preprocessing steps, the WaveShot is meticulously prepared to undertake its waterborne photography mission, ensuring that each operation is based on optimized settings considering environmental conditions, camera angles, and device speed. This preparation is crucial for achieving high-quality waterborne photography, capturing waterborne targets clearly and accurately.

5. Tasks

This study is designed to evaluate the capabilities of the WaveShot, a portable aquatic drone, in capturing video footage of both static and moving objects on water. The tasks are divided into two main categories: static object shooting and moving object shooting. Each category aims to test different aspects of the WaveShot's performance, including stability, clarity, tracking ability, and operational flexibility under various environmental conditions.

5.1. Static Object Shooting

For ***Dynamic Environment Static Object Shooting Task***, the WaveShot is required to capture video footage of a static object placed in natural water conditions, such as areas with slight waves or currents. The challenge is to maintain a stable framing of the target object while compensating for the dynamic changes in the water surface, ensuring the video's stability and clarity throughout the shooting process.

For ***Close-range Detail Shooting Task***, the task focuses on the WaveShot's ability to capture detailed video footage of a static object from a close range. It tests the drone's precision in controlling its distance and angle relative to the object, particularly when capturing specific parts of the object, like surface textures or color gradients. The task evaluates the drone's maneuverability and precision in capturing high-quality detailed shots.

For ***Multi-angle Shooting Task***, the WaveShot is tasked with performing multi-angle video shooting around a selected static target object on water. This involves capturing footage from various horizontal angles and, where possible, adjusting the drone's shooting height or angle to acquire top-down or angled views of the object. The task assesses the drone's ability to maintain video resolution and clarity from each angle, facilitating the use of the footage for subsequent 3D reconstruction work.

5.2. Moving Object Shooting

For ***Moving Water Object from Distance to Close Range Shooting Task***, the WaveShot must identify a moving target on the water from a distance and follow it as it moves closer, ultimately achieving clear, close-range footage. This task tests the drone's target recognition and dynamic tracking capabilities, as well as its ability to remain stable during the approach. The challenge lies in capturing the object's details and maintaining shooting continuity as the distance and possibly the speed of the object change.

For ***Moving Water Object from Close Range to Distance Shooting Task***, this task requires the WaveShot to start capturing footage of a moving target on the water from a close range and continue to maintain the target within the frame as it moves to a farther distance. This task similarly tests the drone's tracking and shooting stability, especially in adjusting the shooting angle to keep the moving target in a clear and central position as it moves away. The challenge is to maintain effective tracking of a moving object at a distance while ensuring the stability of the footage.

These tasks are designed to comprehensively evaluate the WaveShot's capabilities in a variety of shooting scenarios, highlighting its potential applications in fields such as environmental monitoring, aquatic research, and dynamic event documentation.

6. Experiments

In this study, we utilized the advanced capabilities of OpenAI's GPT-4V model to evaluate the performance of WaveShot across various tasks. The GPT-4V model is designed to respond to queries based on given text and images, which allows for a comprehensive assessment of WaveShot's capabilities through a series of designed questions and corresponding

Text	Clarity
"OMG!"	10
Text on the middle-sized banner in the center	8
Text on the small white banner in the top left	7
Text on the small white banner in the top right	7
Text on the bottom red banner	9
Average	**8**

Table 1: Text clarity evaluation

Fishing Boat: WaveShot captures a fishing boat moving from a distance to a closer position in low-light conditions.

Unmanned Boat: WaveShot captures an unmanned boat moving from a closer position to a distance under direct sunlight.

Figure 4: ***Task of Moving Object Capture on Water***.

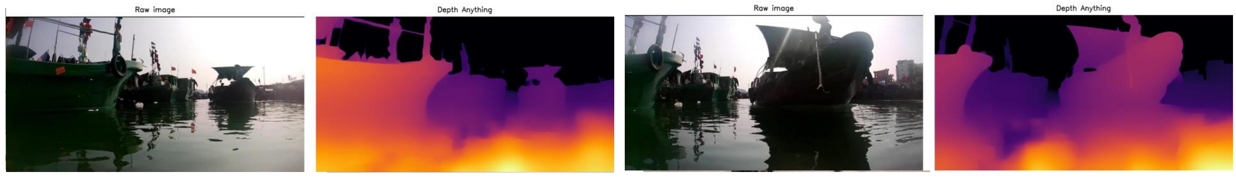
image sequences in Figure 3. This approach aims to provide a nuanced understanding of WaveShot's performance in capturing video details in water-based environments.

6.1. Static Object Detail Capture in Close Proximity

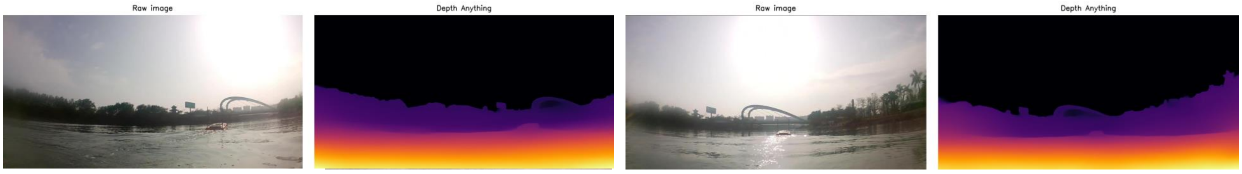
For Task 6, The first experimental setup involved a 15-meter-long and 6-meter-wide outdoor pool located at a shopping center to evaluate WaveShot's ability to capture static objects in detail at close range. The pool scene, surrounded by shops displaying various textual advertisements, served as an ideal setting for assessing the clarity with which WaveShot could capture text in images in Figure 3. With an overall text clarity score of 8, the results indicated that WaveShot could capture static object details with high clarity under near-distance and favorable lighting conditions in Table 1. This suggests that WaveShot is capable of meeting general photography needs, especially in static object scenarios.

6.2. Dynamic Environment Static Object Shooting

In a natural water setting next to a residential area in Beihai, Guangxi, WaveShot was tested for its performance in capturing static targets under unknown wind and wave conditions. The experiment included five different static targets from Task 1 to Task 5 in Figure 3, revealing WaveShot's ability to maintain focus and capture reflections and colors effectively, even in slight water surface movements. The first static target shows clear and sharp static objects such as trees, buildings, and ground decorations. This indicates that WaveShot maintains good focus even in the presence of slight disturbances on the water surface during the shooting process. The reflection on the water surface is well captured while preserving the colors and details of the scenery. This suggests that WaveShot has excellent light management capabilities, enabling it to capture high-quality videos under different lighting conditions. The colors appear natural without noticeable color shifts or excessive saturation, indicating good color reproduction. For the second static target, the image details are clearly visible with good focus, although there is slight blurring in certain areas of the frame. In the case of the third static target, the footage appears relatively smooth without significant movements or shaking. There are slight fluctuations in reflections and shadows, likely caused by the natural movement of the water surface. Both the nearby sculpture and the distant trees exhibit high clarity, with well-preserved



Left: long-range fishing boats and their corresponding depth maps. ***Right***: short-range fishing boats and their corresponding depth maps.



Left: short-range unmanned boats and their corresponding depth maps. ***Right***: long-range unmanned boats and their corresponding depth maps.

Figure 5: ***Task of Water Scene Depth Estimation***.

details and no blurring.Regarding the fourth static target, WaveShot maintains stability without any noticeable shaking or jumping. For the last static target, the details of the buildings and the natural environment are clear, and the reflections on the water surface are also relatively crisp. However, there is slight blurring in distant details, suggesting that the lens may not focus as well on distant objects compared to nearby ones. WaveShot demonstrates impressive stability and video quality in water surface environments, indicating its suitability for capturing static objects in slightly dynamic water conditions.

6.3. Moving Object Capture on Water

For the experimental shooting of moving objects on water, we selected two objects to capture: an unmanned boat and a fishing boat in Figure 4. When it came to filming the fishing boat up close, WaveShot was positioned slightly lower. Its close-up lens captured some details of the boat, such as the color of the hull, its structure, and items on board. The main parts of the boat, such as the bow and the sides, remained in focus and clear. However, due to the gloomy weather or low lighting conditions during the filming, the brightness and contrast of the content were affected. The video displayed WaveShot's capabilities in both stationary and dynamic shots, particularly in maintaining stable filming.

In another experiment, WaveShot captured the process of a small unmanned boat moving away from the camera. Each frame captured it at a further distance. During the filming, WaveShot appeared to successfully track the moving unmanned boat and kept it centered in the frame as it moved away. This was beneficial in maintaining focus on the subject. However, the camera was facing a strong light source, possibly the sun, which resulted in overexposed images. This caused the loss of detail in the sky and surrounding scenery. In the initial frames, the boat remained relatively clear, but as the distance increased, the clarity decreased, especially when filming a moving subject. Overall, WaveShot performed well in terms of stability and subject tracking. However, it faced challenges in exposure control and clarity, particularly in long-distance shots and under varying lighting conditions.

6.4. Water Scene Depth Estimation

Using video material from the moving object capture experiments, relative depth estimation for a fishing boat scene and an unmanned boat scene was conducted in Figure 5. The depth estimation utilized color coding to indicate the distance of objects from the observer, with warm colors for closer objects and cool colors for distant objects. While the depth estimation maps generally reflected the depth information accurately, some inaccuracies were noted at edges and transition areas. The depth maps showed effective distance estimation between the unmanned boat and its surroundings, although finer details like the antennas on the unmanned boat were not captured, indicating limitations in detail retention in multi-object water scenes.

Overall, the experiments conducted provide valuable insights into WaveShot's capabilities in capturing details and estimating depths in water-based environments, highlighting its strengths and areas for improvement in both static and dynamic conditions.

7. Conclusion, Limitations and Future Directions

In this work, we introduced WaveShot, a novel compact and portable unmanned surface vehicle designed for filming and media production in aquatic environments. By integrating differential propulsion, onboard sensing and processing, remote operation, and action camera functionality, WaveShot enables smooth tracking shots and stable image capture on the water surface. It provides a flexible and accessible platform for aquatic photography and extends the toolkit for film and media production. This research demonstrates the potential of unmanned surface vehicles in creative arts and other emerging applications. Addressing current limitations and advancing the concept of WaveShot can lead to more dynamic image capture and novel visual perspectives in aquatic environments.

Our experimental results demonstrate the effectiveness of WaveShot in filming both static and dynamic objects on water. The intuitive control interface, allowing even beginners to quickly learn complex operations. Qualitative and quantitative analyses suggest that WaveShot has advantages in stability, detail capture, tracking, and depth estimation using state-of-the-art monocular models.

However, the current WaveShot prototype and methods have some limitations. The wireless transmission range is currently limited to approximately 1 kilometer, restricting the operational radius. Additionally, factors such as weather, lighting, and water flow turbulence may negatively impact image quality and model performance. Occlusions, reflections, and the absorbance properties of water still pose challenges for accurate depth estimation, especially over longer distances.

This work opens up many promising directions for future development. We hope to integrate waterproof gimbal stabilization systems in future work to further enhance filming stability in adverse conditions. On the software side, the application of more powerful control and computer vision techniques tailored for aquatic filming can optimize the filming process in water scenes.

References

[1] Mohammed Aissi, Younes Moumen, Jamal Berrich, Toumi Bouchentouf, Mohammed Bourhaleb, and Mohammed Rahmoun. Autonomous solar usv with an automated launch and recovery system for uav: State of the art and design. In *2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*, pages 1–6, 2020. doi: 10.1109/ICECOCS50124.2020.9314415. 3

[2] Marco Bibuli, Gabriele Bruzzone, Massimo Caccia, Giovanni Indiveri, and Alessandro Antonio Zizzari. Line following guidance control: Application to the charlie unmanned surface vehicle. *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3641–3646, 2008. URL https://api.semanticscholar.org/CorpusID:3009769. 3

[3] S. Campbell, W. Naeem, and G.W. Irwin. A review on improving the autonomy of unmanned surface vehicles through intelligent collision avoidance manoeuvres. *Annual Reviews in Control*, 36(2):267–283, 2012. ISSN 1367-5788. doi: https://doi.org/10.1016/j.arcontrol.2012.09.008. URL https://www.sciencedirect.com/science/article/pii/S1367578812000430. 3

[4] Yikang Chen, Jingjing Wang, and Yu Zhang. A local obstacle avoidance method for unmanned surface vehicle based on acceleration and angular velocity. In *2023 2nd International Symposium on Sensor Technology and Control (ISSTC)*, pages 243–246, 2023. doi: 10.1109/ISSTC59603.2023.10280968. 3

[5] K.D. Do, Z.P. Jiang, and J. Pan. Robust adaptive path following of underactuated ships. *Automatica*, 40(6):929–944, 2004. ISSN 0005-1098. doi: https://doi.org/10.1016/j.automatica.2004.01.021. URL https://www.sciencedirect.com/science/article/pii/S0005109804000482. 3

[6] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network, 2014. 3

[7] Ravi Garg, Vijay Kumar BG, Gustavo Carneiro, and Ian Reid. Unsupervised cnn for single view depth estimation: Geometry to the rescue, 2016. 4

[8] Clément Godard, Oisin Mac Aodha, and Gabriel J. Brostow. Unsupervised monocular depth estimation with left-right consistency, 2017. 4

[9] Clément Godard, Oisin Mac Aodha, Michael Firman, and Gabriel Brostow. Digging into self-supervised monocular depth estimation, 2019. 4

[10] Iro Laina, Christian Rupprecht, Vasileios Belagiannis, Federico Tombari, and Nassir Navab. Deeper depth prediction with fully convolutional residual networks, 2016. 3

[11] Sang-Min Lee, Kyung-Yub Kwon, and Joongseon Joh. A fuzzy logic for autonomous navigation of marine vehicles satisfying colreg guidelines. *International Journal of Control Automation and Systems*, 2:171–181, 2004. URL https://api.semanticscholar.org/CorpusID:18996157. 3

[12] Yuanchang Liu and Richard Bucknall. Efficient multi-task allocation and path planning for unmanned surface vehicle in support of ocean operations. *Neurocomputing*, 275:1550–1566, 2018. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2017.09.088. URL https://www.sciencedirect.com/science/article/pii/S092523121731617X. 3

[13] Dario Madeo, Alessandro Pozzebon, Chiara Mocenni, and Duccio Bertoni. A low-cost unmanned surface vehicle for pervasive water quality monitoring. *IEEE Transactions on Instrumentation and Measurement*, 69(4):1433–1444, 2020. doi: 10.1109/TIM.2019.2963515. 3

[14] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer, 2020. 4

[15] Ashutosh Saxena, Sung Chung, and Andrew Ng. Learning depth from single monocular images. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems*, volume 18. MIT Press, 2005. URL https://proceedings.neurips.cc/paper_files/paper/2005/file/17d8da815fa21c57af9829fb0a869602-Paper.pdf. 3

[16] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data, 2024. 4

[17] Tinghui Zhou, Matthew Brown, Noah Snavely, and David G. Lowe. Unsupervised learning of depth and ego-motion from video, 2017. 4