Blind Wave Field Characterization from Fluid Lensing

EE368 Final Project Report

Ved Chirayath¹, Oscar Galvan-Lopez², Ronnie Instrella²
¹Department of Aeronautics & Astronautics, ²Electrical Engineering
Stanford University
Stanford, CA 94305 USA
ved@stanford.edu, ogalvan@stanford.edu, rinstre@stanford.edu

Abstract - Fluid Lensing is a theoretical model (Chirayath et al.) for fluid-optical interactions in turbulent flows as well as two-fluid surface boundaries that, when coupled with computational imaging and an unique image-processing pipeline, may be used to not only remove strong optical distortions along the line of sight, but also significantly enhance the angular resolution of an otherwise underpowered optical system. Fluid Lensing has recently been used to map coral reefs underwater from an aerial UAV-platform at sub-cm scale for the first time and enabled 3D reconstruction (Figure 1), which has proved a ground-breaking tool in modern marine research and for marine environment conservation. In our final project, we explore a novel, unsolved, problem in ocean optics - the inverse problem of Fluid Lensing. Namely, from a raw high-frame-rate video sequence, rather than just produce a distortion free image from Fluid Lensing, we explore uniquely characterizing the wave field at an instance in time. This information can not only provide an order of magnitude improvement to future Fluid Lensing implementations, but also a powerful remote sensing ability to detect features betraying ocean circulation patterns, flow properties, wave breaking and even the presence of underwater submersibles and marine fauna. We explore the application of two computer vision schemes, SIFT descriptors and Vocabulary Trees to match templates of surface wave fields coupled with optical simulations to experimental test setup data to identify the location of a dropped marble in a small water tank by observing how wave features distort a target template in the central region away from the marble drop location. Such a scheme can be used to determine surface wave shape, amplitude, wavelength and group velocity uniquely provided a large enough template library. We present promising results from SIFT descriptors to the application of this problem and discuss limitations and challenges encountered with real-world experiments.

1. Introduction

The main purpose of this project is to uniquely determine simulated ocean surface wave conditions exclusively by image processing on high frame rate video data. In Figure 2, we show a simple experimental demonstration of Fluid Lensing resolving a resolution test target. The test target is imaged from the bottom of a 12ft deep pool, while attached to a pole 7 feet above the surface (the inverse problem looking down is similar). A raw frame captured from a high-frame-rate video is pictured above and shows the significant optical distortion from waves in the pool interfering in the optical path. The middle image shows the result of passing one second of data (120 frames) to the Fluid Lensing scheme to reconstruct the target accurately, albeit with reduced contrast. Once resolved, we propose using this image as a reference for comparison to further raw image data showing distortions. However, further distortions will be tracked in an attempt to characterize the wave field responsible for the optical distortions. We hope to achieve the third image; namely, a unique picture of the wave field at an instance in time just

Figure 1 - Fluid Lensing results from UAV-based mapping of coral reefs

RAV video frame from UAV pass. Note the strong optical defects from wave distortion and surface reflection
120 frames (1 second) processed with Fluid Lensing to remove wave distortions and magnify targets
Distortion-free Fluid Lensed data are used to also compute parallax distance and create 3D map from same frame set. (Same target from different angle)
from the steps mentioned above, or at least the location of the marble drop in our experimental setup.

We implement and test a basic keypoint detection and template matching scheme to determine the direction of both simulated and recorded waves in a small test tank by comparing recorded video data to a library of simulated wave templates. Although a very limited test scenario, methods developed here may be eventually applied to larger real-life scenarios with ocean waves and more complex boundary conditions. In this way, our project aims to resolve the inverse problem and uniquely solve the fluid surface condition rather than relying on statistical models of wave properties. Such information could be used to provide an order of magnitude improvement to future Fluid Lensing implementations as well as a powerful remote sensing ability to measure ocean surface roughness, currently measured by Synthetic Aperture Radar or scatterometers.

Overall, the significance of this project is its potential to decrease the computational costs associated with image reconstruction using Fluid Lensing as well as to provide a novel remote sensing capability for existing airborne and space borne instruments. If the wave library is large enough to accurately determine wave parameters from recorded video, then it may be possible to use direct optical coupling, deconvolution and raytracing to provide an order of magnitude increase in Fluid Lensing performance and resolution.

2. Methodology

Simulated template videos of surface field waves traveling in various directions are generated in MATLAB and used to characterize wave field characteristics in query videos of an experimental setup (Figure 3) through SIFT keypoint detection. An estimate of the dominant wave direction based on distortions in the underlying image template is generated, and the results are compared to the actual wave direction observed in the query video to determine where a marble was dropped in the test tank (x,y coordinates) by observing only wave distortions over template image far away from the drop location in the center of the test tank. Our simulation library only included an array of test drop locations to localize the marble drop location for this project; however a future larger library can account for additional variables such as surface wave shape, amplitude, wavelength and group velocity which would correspond to different marble drop heights, sizes and densities.

2.1 - Wave simulation

The first step in our project seeks to simulate the wave field of a small pool of water perturbed by a droplet striking the surface. The surface of water is approximated with a modified 2D Helmholtz wave equation (Eq. 1) with Dirichlet boundary conditions of no movement at the boundary. To better mimic the real-world physical system involving surface tension and the Navier Stokes equations, we add an additional energy dissipation term to decay waves over time. Single gaussian perturbations to the surface act as droplets to create ripples. As time evolves, this simulation resembles wave behavior in a small pool qualitatively well. The more complicated case of multiple perturbations is simulated with a series of random gaussian perturbations using the above scheme.

\[
\frac{\partial^2 u}{\partial t^2} = c^2 \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) - p \frac{\partial u}{\partial t}
\]

Equation 1 - 2D Helmholtz wave equation

\[
D = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}
\]

Equation 2 - Matrix D for finite difference solution
It can be shown\(^5\) that a simple finite difference scheme solution to this wave equation can be solved by a 2D convolution with the matrix \(D\) (Eq. 2) and a symmetric stencil. We modified open-source work by Maxim Vedenyov, license attached for reference, to couple the optical problem to this wave solution; namely the physical phenomenon of refraction by the air-water boundary and surface waves. Initial conditions and wave parameters were chosen such as to replicate observed experimental results. Please refer to MATLAB code, Part 1 for complete description of relevant parameters.

2.2 - Optical coupling
The optical interaction arising from the fluid surface deformation is assumed to arise solely from refraction at the air-water boundary and the shape of surface waves generated in section 2.1. The wave perturbations are assumed to be small angle deformations at the fluid surface relative to the mean fluid surface and refractive indices for water and air at STP are used. This allows for the small angle approximation to be used for Snell’s law of refraction:

\[
n_1 \sin(\theta_1) = n_2 \sin(\theta_2)
\]

becomes

\[
n_1 \theta_1 = n_2 \theta_2
\]

Equation 3 - Snell’s law of refraction for small angle perturbations

This simplification means that an \(x,y\) shift of a pixel in the reference image, when distorted by a surface wave, is directly proportional to the wave surface angle. Thus, it is sufficient to simulate the optical coupling to the wave solution by calculating the derivative of the wave field at a point \((x,y,t)\). Such derivatives can be computed by convolution with a Sobel operator. However, it should be noted that for large wave amplitudes, and hence large angle surface conditions, this small angle approximation results in large errors. Thus, for our simulations and experimental setup, we limit our experiments to small amplitude wave perturbations.

2.3 - Wave simulation library
We use the above surface wave field and optical coupling simulation to generate a time-evolving video. A series of arrays serve as test target images (Figure 3 shows the nine-disc array pattern as it is distorted by surface waves; refer to MATLAB code, Part 1 for full set). We can observe how these shapes get morphed by waves arising from single perturbations in nine equally spaced corners of the image. In the case of the sequence in Figure 3, we began with a wave coming from the bottom center. We repeat this process for a grid of different perturbation starting locations to generate a video library of wave simulations over a time step large enough to capture multiple reflections off the boundaries of the test tank, as we are interested in being able to determine perturbation location from a complicated wave field. Boundary conditions and test tank size were chosen to mimic those of experimental setup as closely as possible.

2.4 - Experimental Setup
A series of query videos are used to test the accuracy of our proposed algorithm. All of the tested images use a 9x9 grid of black dots as an image template or the Brain Coral target. Two types of test videos are used in this project: the first set of query videos display simulated waves with perturbations originating from unknown locations within a container, and the second set contains...
waves generated in two water-filled transparent containers captured using a digital camera. A printed version of the 9x9 grid of black dots is placed under the container with a camera pointed orthogonally to the water’s surface. A small marble is dropped into the tank at various locations, generating waves across the fluid surface. The visual distortions to the black dots are recorded and passed through the scorekeeping algorithm. An illustration of the experimental set up is shown in Figure 5.

**Experimental Set up**

![Experimental Set up](image)

Figure 5 - A simple experimental set up is used to record waves moving across a small transparent container, which are generated by dropping a marble into a water bath

2.5 - SIFT Descriptor Library

Our proposed algorithm utilizes a library of image descriptors generated from eight simple wave simulations. Each wave is generated through a single perturbation to the fluid surface at a specific point within square (512x512 pixels) and rectangular (513x653 pixels) containers. In this study, the waves originate near the container boundary at one of the eight main compass directions, generating a wavefront that travels through the entire container over multiple time points. The optical effects of these waves are observed through a grid of 9x9 black dots centered on a white background. These dots are visually distorted as the wave travels through the container. The Scale Invariant Feature Transform (SIFT) is used to generate feature descriptors for each frame in all of the wave simulation videos. This set of descriptors form a reference library, which we use to detect, match and characterize waves from either the test videos or wave simulations. In this study, our generated library is used to determine the general direction of moving waves in a video.

2.6 - Wave Detection and Matching: SIFT Scorekeeping

Query videos are matched to a particular library entry using a simple SIFT scorekeeping algorithm, which is outlined in Figure 7. Each frame from the test video is converted from RGB (8-bit unsigned int) to grayscale and thresholded to segment and isolate the desired comparison template, which in this case is a 9x9 grid of black dots. SIFT keypoints are detected from each frame, and matched to the descriptors in the image library. The library entry with the greatest number of descriptor matches is considered to be the closest match. The entries with the closest match from a particular query frame is awarded a score of +1. This process is repeated for every frame in the query image, and a tallied score of the total number of closest query frame matches for each library entry is recorded. The library entry with the highest score is considered the closest matching video.

![SIFT Scorekeeping](image)

Figure 7 - Implementation of the SIFT scorekeeping method

In this implementation of SIFT Scorekeeping, if two or more library entries share the same number of detector
matches for a particular frame in the queried image, neither score is incremented. Ties in the total score are weighted equally among separate library entries; if two wave simulations have the same final score, the waves are assumed to have travelled equally in both directions. Moreover, a stopping criterion is used to terminate the matching process early if a single entry’s score is significantly greater than the rest. In simulated test cases under controlled conditions, this criterion should reduce the required runtime for simple test cases.

2.7 - Alternate Vocabulary Tree Method for Wave Detection and Matching

Building the Tree
Whereas in the previous method SIFT descriptors were used as a reference library, in the Vocabulary Tree Method SIFT Descriptors were used to build a vocabulary tree. This tree was subdivided into various nodes and leaves allowing us to arrange the large reference library of SIFT descriptors into some ordered format. The importance of the tree method is that its goal is to speed up an algorithm which is very computationally intensive when used on images. The primary motivation then, was to lower the computation time necessary to allow the use of the method on video. Once the vocabulary tree was built, SIFT descriptors from each individual reference frame for each reference video were processed through the tree. This resulted in a probability mass function of visit counts over each leaf node which was stored in a m x n x l matrix, where m corresponds to the number of leaves in the tree, n corresponds to the number of frames chosen for analysis, and l corresponds to the number of videos in the reference library.

Scoring the Videos
For each frame of the query video being analyzed, the L1 distance was calculated between each reference frame. This resulted in a n x l matrix that characterized that frame’s score against our reference library. For each reference video, the minimum score was taken (indicating the best frame match) and added to it’s corresponding place in a row vector of length l which kept track of the cumulative score for all frames versus each individual video. This process was repeated for a predefined number of frames. At the end of this process, the reference video which most closely resembles the query video would be that which had the lowest score.

3. Experimental Results

3.1 - SIFT Scorekeeping Method
The results of the SIFT Scorekeeping method are visualized using arrows positioned at each of the 8 possible wave locations. The length of each arrow is proportional to the relative score of a particular library entry; a longer arrow indicates a higher relative score for a particular video within the library. These arrows are superimposed on the frames of the original query video, and are updated at each time step.

The results from a number of test videos are presented in this project, with 6 test cases shown in Figure 7. In the basic simulation data with a centered 9x9 grid of black dots of equal size to those found in the library, the wave direction is detectable with varying levels of error across the videos. Little error is observed when the wave originates at a location that closely matches one of videos within the library; in this case, a single template video’s score is significantly higher than the rest. The estimated direction is not as apparent when the wave starts at a location between two of the simulated waves within the library; for instance, if the wave perturbation occurs between the bottom right corner and the rightmost boundary of the container, multiple arrows of similar length indicate some level of ambiguity in the estimated direction of the wave.

In all of our simulated test cases, the scorekeeping algorithm correctly determined the general direction of the wave by interrogating up to 100 frames per test video. However, applying this algorithm on captured video did not produce accurate results. In this case, the same rectangular and square simulated wave SIFT descriptor libraries are used to determine keypoint matches. While a few of the experimental test videos estimated the direction correctly, many were completely off and unreliable. Overall, the test footage from the experimental setup had a significantly higher level of error than the simulated data, although in certain cases the wave direction was properly resolved. Both a successful and unsuccessful test case from captured video footage are shown in Figure 8.

3.2 - Vocabulary Tree method
There were some mixed results from our experiments using the Vocabulary Tree method. The algorithm was run once with the query video being Video 1. As you can see from the scores in Table 1, test 1, the algorithm works perfectly. The problem arose when the query video was
Table 1 - Vocabulary Tree Results

<table>
<thead>
<tr>
<th>Drop location</th>
<th>Score, Test 1</th>
<th>Score, Test 2</th>
<th>Score, Test 3</th>
<th>Score, Test 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottom</td>
<td>1.875</td>
<td>134.921</td>
<td>138.548</td>
<td>30.000</td>
</tr>
<tr>
<td>Bottom Left</td>
<td>146.502</td>
<td>86.480</td>
<td>161.720</td>
<td>30.000</td>
</tr>
<tr>
<td>Bottom Right</td>
<td>145.190</td>
<td>79.535</td>
<td>161.628</td>
<td>30.000</td>
</tr>
<tr>
<td>Center</td>
<td>158.125</td>
<td>156.977</td>
<td>154.692</td>
<td>30.000</td>
</tr>
<tr>
<td>Left</td>
<td>81.050</td>
<td>134.963</td>
<td>139.780</td>
<td>30.000</td>
</tr>
<tr>
<td>Right</td>
<td>74.488</td>
<td>135.108</td>
<td>138.382</td>
<td>30.000</td>
</tr>
<tr>
<td>Top</td>
<td>98.462</td>
<td>136.415</td>
<td>140.275</td>
<td>30.000</td>
</tr>
<tr>
<td>Top Left</td>
<td>149.525</td>
<td>71.456</td>
<td>161.036</td>
<td>30.000</td>
</tr>
<tr>
<td>Top Right</td>
<td>149.884</td>
<td>79.249</td>
<td>161.425</td>
<td>30.000</td>
</tr>
</tbody>
</table>

![Figure 9 - Real-world test coin drop locations and region of interest](image)

not an exact match. In Table 1, test 2, a mystery video which was not an exact match for any of the reference videos was queried. To the ordinary human, the video clearly shows a disturbance emanating from the bottom right hand corner of the video. Our algorithm gave us a more or less four-way tie. The candidates for the four-way tie were: the top left corner, top right corner, bottom left corner, and bottom right corner. This was a limitation in our experiment, because our image was rotationally symmetric, it was almost impossible for the algorithm to determine which of the four corners was nearest to the disturbance. The results were even less promising when querying an actual video that was filmed. As can be seen from Table 1, test 3, the algorithm was unable to distinguish between the videos at all. To compensate for this limitation in the experiment, video footage was shot of a coral image (shown below) beneath a plastic container. A coin was dropped in several spots (blue stars, Figure 9) surrounding an area being filmed (red rectangle, Figure 9). The previously described process was repeated and

the with another query video being shot along the bottom left side. These results can be seen in Table 1, test 4. As can be seen in those results, there was no distinction between any of them.

4. Discussion & Conclusions

Our basic task of detecting the location of a marble dropping into a small water tank by observing wave distortions and matching to simulated results was successful provided the real-world experiment closely mimicked simulation conditions. Despite using a simple 2D wave simulation model with basic optical coupling due to refraction, our SIFT scorekeeping algorithm produced promising results from a small template library with only 9 drop locations. We were able to successfully detect marble drop locations occurring away from our template locations. However, as our real-world experiment grew in complexity and deviated from the simulation templates, detection errors grew rapidly. The idea of using SIFT descriptors to match images is a well established process in digital image processing. The extension of this process for use with video and wave perturbations, however, is non-trivial due to the complicated relationship between surface waves and boundary conditions of the tank.

One issue we encountered was the template reference image of 9 disks arranged in a 3x3 square, was vertically, horizontally, and rotationally symmetric. This made it problematic when attempting to identify which of the 9 disks was closest to the disturbance. This resulted in degeneracies in detection, which were partially ameliorated by using the coral reference image, which is fractal in nature and rotationally and translationally asymmetric. Another issue that may have contributed to poor robustness was the limited number of videos in our reference library. Ideally we would like to be able to have a reference video describing a wave originating at every single pixel of our video. This was a bit impractical given the time constraints and computational cost, but can be pursued in future iterations.

Suggestions for further improvements and changes

Both proposed scorekeeping algorithms rely heavily on basic SIFT keypoint detection and matching, which leaves room for a number of further improvements for blind wave field characterization applications. For example, it could be advantageous to assign higher scores to library videos, match a series of consecutive frames, or use homography matrices to locate and isolate only the regions that contains the background template (9 black dots or a reference background already solved). Moreover, generating a new library of SIFT descriptors from recorded data could prove to better handle the experimental query videos, as opposed to relying solely on wave simulations. Lastly, we hope to extend our methods to more complicated cases and generate a larger library to account for additional variables such as
surface wave shape, amplitude, wavelength and group velocity which would correspond to different marble drop heights, sizes, densities and surface tension properties.

Acknowledgments

We would like to acknowledge our graduate advisor, André Araujo, Professor Bernd Girod, David Chen and Matt Yu for their support and insight throughout this project.

Work Breakdown

Ved wrote and integrated the final report and the project proposal, designed the poster and was responsible for creating an air-water surface wave simulation with optical refractive coupling to approximate a real-world ocean optics problem. He produced a library of simulation videos with waves coming from simulated marble drops at nine locations for two types of test tank shapes and a variety of background image templates including disc arrays and an actual Brain Coral image to compare to experimental data.

Ronnie contributed the basic framework and implementation of the SIFT Library and SIFT scorekeeping algorithm, as well as all of the figure s associated with this method. He also produced the query experimental videos used to test the algorithm.

Oscar was responsible for developing the vocabulary tree method for video analysis and filming of the coral experimental videos.

References