Limitations of Vision Guided Underwater Navigation

Maria E. Angelopoulou, Chourmouzios Tsiotsios and Maria Petrou

Department of Electrical and Electronic Engineering,
Imperial College London, London SW7 2AZ U.K.
E-mail: {m.angelopoulou, c.tsiotsios, maria.petrou}@imperial.ac.uk

Abstract: Underwater vehicles that are equipped with an on-board camera can utilize the captured frames to perceive the 3D structure of the ocean floor and thus perform safe underwater self-navigation. The focal length, frame rate and exposure time, which determine the frame capturing process, in association with the speed of the vehicle determine the capabilities of the vision-based navigation system. In this work, we quantify the effect of the above imaging parameters on the performance of the vision system, and the feasibility of vision-based navigation is assessed for various parameter values. Results demonstrate that a 1200×1600 camera with a $5mm$ lens and $15f/s$ frame rate allows the dense reconstruction of the ocean floor, thus enabling autonomous underwater navigation, provided there is enough processing power on the vehicle to perform the necessary image processing in real time.

Keywords: underwater vision, 3D reconstruction of the ocean floor, autonomous underwater navigation, focal length, frame rate, exposure time, motion blur.

1. INTRODUCTION

The safe underwater navigation of a multi-vehicle system can be considerably facilitated by employing on-board imaging and processing modules, which gradually sample and reconstruct the 3D surface of the ocean floor [Yoerger et al. (2007); Johnson-Roberson et al. (2010); Campos et al. (2011); Singh et al. (2007)]. Underwater multi-vehicle systems typically include both Remotely Operated Vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs) [Marques et al. (2007)]. The ROVs are cube-shaped vehicles, whose motion is relatively slow and whose movements are easy to control. On the contrary, the AUVs are long and narrow, designed to move fast at small altitudes from the ocean floor. In real-life underwater applications, the worst-case scenario for the forward motion of an AUV would involve a vehicle speed of $2m/s$ at altitudes as low as $0.5m$.

This paper assesses the effectiveness and feasibility of underwater vision-based navigation for various vehicle speeds and imaging parameters. These parameters are the focal length, frame rate and exposure time (shutter speed) of the on-board camera.

The target application is the navigation of a real-world underwater multi-vehicle system and imposes the following system specifications. The maximum frame rate of the camera that is mounted on the underwater vehicle is $15f/s$, which is the frame rate that we have considered, while the exposure time falls into the range of $1/10000$ to $1/15s$. Moreover, the image sensor of the camera has a resolution of 1200×1600 pixels, with each pixel being $4.40\mu m$ wide. Finally, the maximum speed of the underwater vehicle is $2m/s$, which corresponds to fast-moving AUVs, while its minimum distance from the ocean floor is $0.5m$.

⋆ The research leading to these results has received funding from the European Commission FP7-ICT Cognitive Systems, Interaction and Robotics under the contract #270180 (NOPTILUS).

2. LIMITATIONS OF THE VISION SYSTEM

In this section, we briefly discuss the imaging parameters and determine a strategy to perform effective underwater surface reconstruction, given the restrictions of the available hardware. A more detailed quantitative evaluation of the imaging parameters is presented in Section 3. To quantify the effect of these parameters on the performance of our vision system, the pin-hole camera model is employed.

2.1 The Pin-Hole Camera Model

Figure 1 shows schematically the pin-hole camera model we may use for simplicity, to work out the limitations the hardware imposes on the vision system. Let us assume that the camera is mounted at the bottom of the underwater vehicle, looking at the ocean floor. Let us say that this places the camera distance H from the ocean floor and that the focal length of the camera is F. If d is the linear dimension of the field of view of the camera along the direction of the motion of the vehicle, the corresponding length on the image plane will be d'. From the similar triangles in Fig. 1 we have:

$$\frac{F}{H} = \frac{d'}{d} \Rightarrow d' = \frac{Fd}{H} \quad \text{and } d = \frac{Hd'}{F} \tag{1}$$

The camera manufacturer, however, recommends the use of

$$\frac{F}{H} = \frac{d'}{d+d'} \Rightarrow d' = \frac{Fd}{H-F} \quad \text{and } d = \frac{(H-F)d'}{F} \tag{2}$$

instead [The Imaging Source Europe GmbH (2006)], perhaps in order to correct for the simplicity of the pin-hole camera model. In any case, as $d' << d$, there is a very small difference between the two models.

2.2 Focal Length

Compared with zoom lenses, fixed focus lenses are cheaper, lighter and smaller and, thus, very appropriate for low-cost underwater vision systems. We have therefore decided to employ a suitable fixed focus lens. The focal length of the lens remains to be determined.

As a short focal length implies a wider field of view, and as we want the successive frames to have as much an overlap as possible, a desirable lens to use could be the $5mm$ lens [The Imaging Source Europe GmbH (2006)]. Assuming that we are using this, we have $F = 0.005m$, $H = 0.5m$, and so $d' = d/99$. Since a pixel on the image plane has size $4.4\mu m = 4.4 \times 10^{-6}m$, we can find the physical length it corresponds to by setting $d' = 4.4 \times 10^{-6}m$. For the $5mm$ lens, this means that $d_{pixel} \simeq 0.44mm$ on the sea floor. Given that the camera has 1600 pixels along the long direction of its field of view, the physical length of its total field of view corresponds to $d = 1600\, d_{pixel} \simeq 0.7m$.

2.3 Frame Rate

The frame rate of the camera r in f/s determines the rate at which the scene of interest is sampled. The amount of overlap between successive frames determines the extent of the image region that can be reconstructed. A detailed discussion and evaluation of the above is presented in Section 3.

2.4 Exposure Time

Since the camera is mounted on a moving underwater vehicle, and its shutter remains open for some fraction of a second for each frame taken, significant motion blur may be caused. Knowing the speed v in m/s of the vehicle and the exposure time E in s of the camera, we can calculate the distance $x = v\,E$ in meters that the camera moved having the shutter open. Dividing the distance x with the pixel's physical length size d_{pixel}, we can calculate the number M of pixels that were recorded one on top of the other, causing the effect of motion blur:

$$M = \frac{x}{d_{pixel}} + 1. \tag{3}$$

When the vehicle is not moving, *i.e.* $x = 0$, $M = 1$ pixel will be recorded in each sensor cell. Therefore, no motion blur will be caused. In the worst-case scenario, the vehicle reaches its maximum speed of $2m/s$, and the camera's exposure time equals the time between two successive frames, *i.e.* $1/15s$, since a $15f/s$ frame rate is considered. In this case, the number of pixels of the scene that are recorded on a single pixel of the image sensor is $M \simeq 307$. This causes severe motion blur.

2.5 Targeting the Reconstruction of the Ocean Floor

The target application is the reconstruction of the ocean floor using vision techniques, such as structure from motion [Vidal and Hartley (2008)] and photometric stereo [Argyriou and Petrou (2009); Barsky and Petrou (2003)]. The limitations that are discussed above suggest us the following strategy.

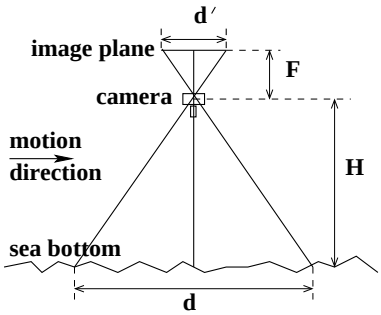


Fig. 1. The pin-hole camera model.

It is virtually out of the question to perform Photometric Stereo (PS) using a camera mounted on an AUV. This is for two reasons.
(i) For PS the camera has to be surrounded by at least 3 lights [Argyriou and Petrou (2009)], approximately symmetrically placed around it. This would be impossible if the camera were placed at the bottom of the vehicle, as there is no triangular base-line there due to the shape of the vehicle [Marques et al. (2007)]. It might be possible if the camera were placed on the side of the vehicle, looking sideways, because then we might be able to place three lights as shown in Fig. 2. This positioning of lights is far from ideal, but it might work. Placing the camera sideways has the additional advantage that the vision system will be able to co-operate with the sonar, which "sees" sideways too. Having the camera hanging underneath the vehicle and looking straight down at the ocean floor may not allow the use of PS, but the methodology of structure from motion may be used instead. In such a case, care should be taken so the successive frames captured have a significant part that overlaps, over which the structure from motion technique may be applied. However, placing the camera at the bottom of the AUV totally disassociates the sonar sensor system from the vision system, as they will not only look at different parts of the ocean floor, but the sonar will be totally unreliable within the field of view of the camera.(The sonar is not reliable for distances closer than $5m$.)
(ii) The second limitation of using PS with the AUVs is the motion blur, as discussed earlier. This will also be a problem with the structure from motion approach. However, as the AUV moves along a straight line with constant speed, one may be able to correct for this effect. An experimental analysis of that is presented in the next section.

The ROV vehicles are more controllable in terms of speed, and they also have the right shape for the placement of 4 lights around the camera, as shown in Fig. 3. To enable PS, the lights need to be synchronised with the camera, so that 4 successive frames are captured while the lights are switched on in turn. During a single frame capturing period, the ROV vehicle ideally remains stationary.

In summary, we propose the following scheme.
1. Use a camera mounted at the bottom of the AUV, motion deblurring and structure from motion to reconstruct

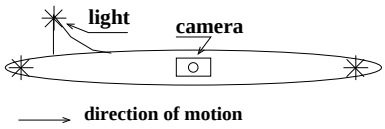


Fig. 2. A possible sideways mounting of the PS vision system on an AUV.

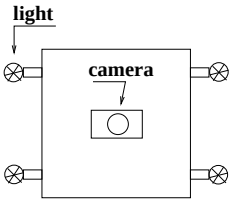


Fig. 3. Frontal view of an ROV with a PS system mounted on it.

parts of the ocean floor. How densely these parts will cover the ocean floor is a matter of how well one frame will overlap with the next. An experimental analysis for that is presented in a following section. In this approach the vision system and the sonar will be totally independent.
2. Use PS for the ROV vehicles, where sonar and vision systems may co-operate.

3. EVALUATION

The imaging parameters and the vehicle speed affect the intra-frame motion blur on each one of the captured frames and the inter-frame overlap between successive frames. As a result, the values of these parameters determine the accuracy of surface reconstruction. In this section, we assess the levels of motion blur and frame overlap, for a set of imaging and motion parameters that is of interest for our target application.

3.1 Modeling the Motion Blur

If the vehicle speed v, the exposure time E and the physical length of the camera pixel d_{pixel} are known, equation (3) renders the number of pixels M that are recorded one on top of the other due to motion blur.

Let us consider, for the exposure time, a minimum of $1/10000s$ and a maximum of $1/15s$, since the maximum frame rate of the camera is $15f/s$. Assuming that a $5mm$ lens is used, Fig. 4 demonstrates the number of pixels M for different exposure times and vehicle speeds. For $E = 1/10000s$, it is $M = 1$, and thus there is no motion blur. Such an E value may be too low for the subsea environment, as the incoming light may be insufficient for rendering valid pixel values, thus resulting in underexposed images. For $E = 1/15s$, $M = 78$ to $M = 307$ pixels of the real-world scene will be recorded by one camera pixel, depending on the vehicle's speed, resulting in significant motion blur.

Assuming that the degradation is linear and that we know M, the Point Spread Function (PSF) of the effect of motion blur can be worked out analytically [Petrou and Petrou (2010)]. Furthermore, a restored version of the original image can be estimated using inverse filtering methods. Figure 5 shows the modeled blurred versions of the original underwater image for different exposure times of the camera. The values of the Mean Square Error

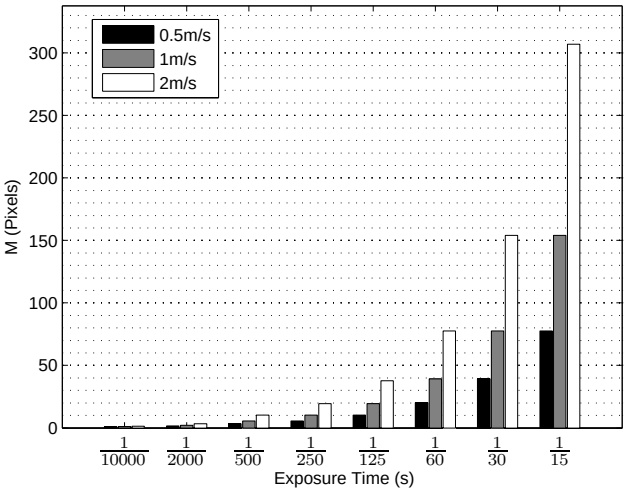
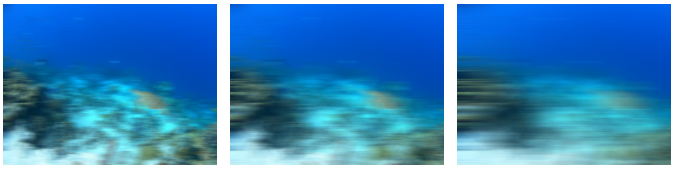


Fig. 4. The number of pixels M that are recorded one on top of the other due to motion blur, with respect to the exposure time E of the camera in seconds, for the indicated vehicle speeds.

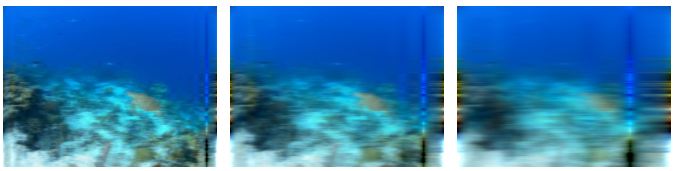
(a) $\frac{1}{10000}s, MSE = 0$ (b) $\frac{1}{500}s, MSE = 168$ (c) $\frac{1}{125}s, MSE = 406$



(d) $\frac{1}{60}s, MSE = 578$ (e) $\frac{1}{30}s, MSE = 786$ (f) $\frac{1}{15}s, MSE = 1032$

Fig. 5. Motion blurred images for the indicated exposure times, when the vehicle is moving with its maximum speed (2m/s).

(a) $\frac{1}{10000}s, MSE = 0$ (b) $\frac{1}{500}s, MSE = 92$ (c) $\frac{1}{125}s, MSE = 233$



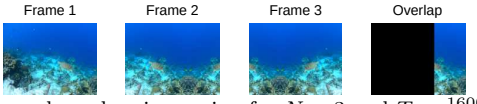
(d) $\frac{1}{60}s, MSE = 382$ (e) $\frac{1}{30}s, MSE = 607$ (f) $\frac{1}{15}s, MSE = 917$

Fig. 6. The restored versions of the degraded blurred images, using Wiener filtering. Image size: 1200×1600.

(MSE) between the original and the blurred images are given in the captions of Fig. 5. The image of Fig. 5a shows in fact the original image, as there is no motion blur for $M = 1$. The case of the worst case scenario, in terms of motion blur, is taken where the vehicle moves with its maximum speed of $2m/s$. The motion is considered to

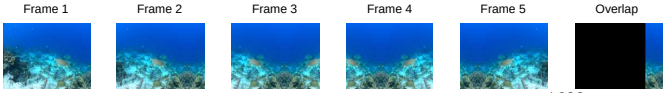
(a) The projection of the scene that is gradually captured.

(b) Frames and overlapping region for $N = 2$ and $T = \frac{1600}{2}$.



(c) Frames and overlapping region for $N = 3$ and $T = \frac{1600}{3}$.

(d) Frames and overlapping region for $N = 4$ and $T = \frac{1600}{4}$.



(e) Frames and overlapping region for $N = 5$ and $T = \frac{1600}{5}$.

Fig. 7. Visualizing the overlapping region for various N as the vehicle moves forward from left to right.

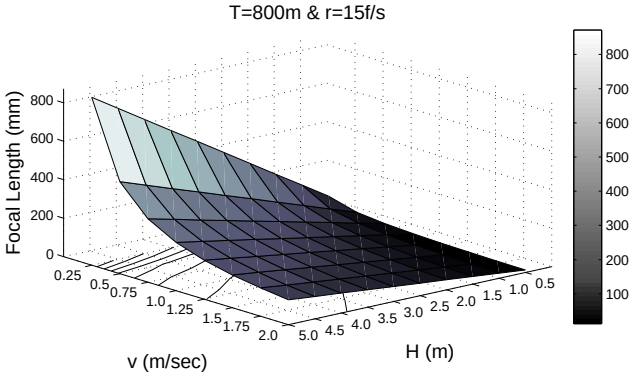
be in the x direction only. Figure 6 shows the respective restored images, after using a Wiener filter with $\Gamma = 0.03$ for restoration [Petrou and Petrou (2010)].

3.2 Visualizing the Overlapping Region

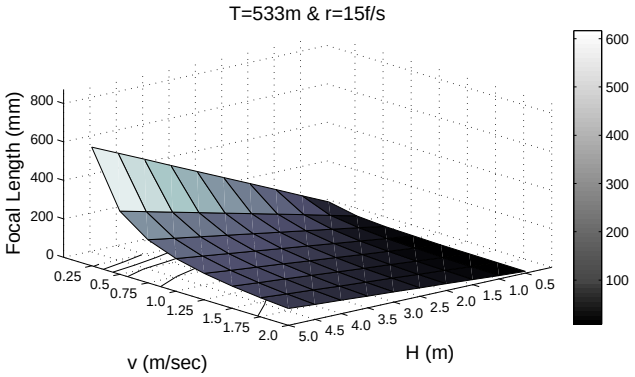
As mentioned in Section 2.5, the 3D reconstruction of the ocean floor is to be done using either structure from motion or PS vision techniques. At each reconstruction phase, successive frames share an overlapping region, which corresponds to the part of the ocean floor that can be reconstructed at this phase. If the overlapping regions that are associated with subsequent reconstruction phases also correspond to adjacent parts of the scene, then, due to the assumed motion linearity, placing the 3D reconstructed parts one after the other gives the continuous 3D surface.

Let R and C denote the number of rows and columns of pixels of the image sensor. In our analysis, it is $R = 1200$ and $C = 1600$. Let N denote the number of successive frames that participate in a single reconstruction phase, and let T denote the per frame translation on the image plane in pixels. If T is exactly equal to $T_N = C/N$, a perfect match of adjacent surface parts is rendered. Therefore, T_N is the optimal value of T for N input images, as it enables reconstructing the continuous surface with minimal computational cost and allows the AUV to move as much as possible, thus minimizing delays.

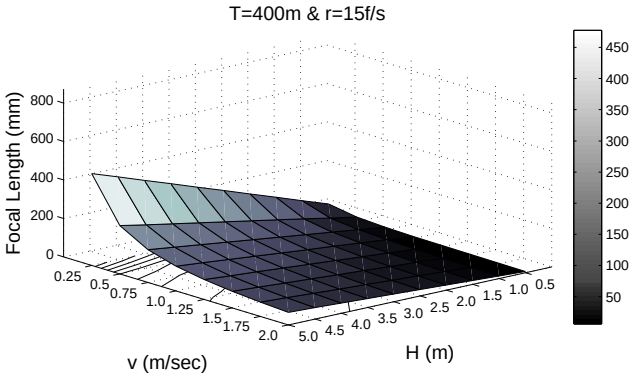
When $T < C/N$, the overlapping regions of subsequent reconstruction phases contain common parts, which are then included only once in the final 3D model of the ocean floor. When $T > T_N$, not all parts of the 3D surface are reconstructed, as successive reconstructed parts are not connected to each another. This is due to the existence of in-between parts, which are not included in any overlapping region. Thus, if the objective is to construct a dense 3D map of the ocean floor, *i.e.* a continuous 3D surface without any gaps, then the value of T cannot be larger than T_N. As it will be explained in the following



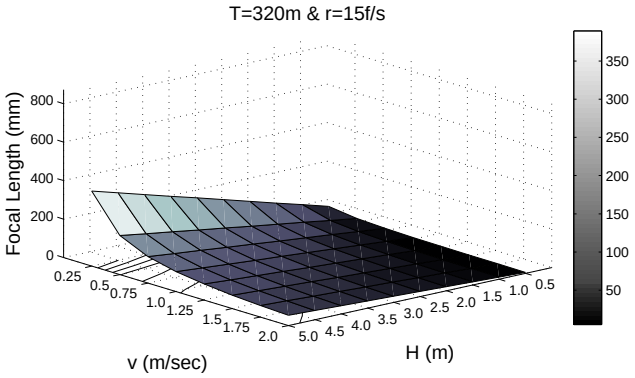
(a) $T = T_2$



(b) $T = T_3$



(c) $T = T_4$



(d) $T = T_5$

Fig. 8. Focal length F dictated by T_N for $N \in [2, 5]$, at frame rate $15f/s$, for various vehicle speeds and altitudes.

section, since our aim is to choose a fixed focus lens, it is impossible to keep the value of T fixed at T_N. Thus, the appropriate focal length should give $T \leq T_N$ for the entire range of system parameters, thus enabling the contiguous reconstruction of the ocean floor.

To demonstrate the combined effect of T and N on the extent of the overlapping region, Fig. 7 visualizes the capturing of a set of N images, for various values of N. The AUV is assumed to have a constant speed and heading from left to right, along the imaged continuous scene of Fig. 7a. The orientation of the camera is assumed to be such that the longer dimension of its image sensor is along the motion direction. This orientation is preferable, as it allows larger values of T. In Figures 7b, 7c, 7d and 7e, N is equal to 2, 3, 4 and 5, respectively, and $T = T_N$. For each N, the scene region that is captured at each frame period is identified, along with the overlapping region. The overlapping region is in fact the part of Frame 1 that is also included in the $N-1$ subsequent frames.

3.3 Determining the Focal Length

In Fig. 8, equation (2) is employed to calculate F for vehicle speeds within the range of 0.5 and $2m/s$ and altitudes within the range of $0.5m$ and $5m$. Altitudes beyond $5m$ are not within our range of interest, as the large volume of water between the camera and the scene would result in images of insufficient quality.

In Fig. 9, a selected set of vehicle speeds and altitudes is considered, and the F values that are dictated by T_N for $N \in [2, 5]$ are given. The worst-case scenario corresponds to $v = 2m/s$ and $H = 0.5m$, which is shown in Fig. 9b. In order to achieve $T \leq T_N$, an increased angle of view is required. As F increases, the angle of view decreases according to the pin-hole camera model of Fig. 1. For the given set of motion parameters, F cannot be larger than the value shown in Fig. 9. Thus, when $N = 5$, the $5mm$ lens suggested by the camera manufacturer [The Imaging Source Europe GmbH (2006)] is required. To safely use the $8mm$ lens instead, N should always be 3 or less (Fig. 9b). If structure from motion is employed, an N value of 3 is typically effective [Vidal and Hartley (2008)]. However, it should be noted that using an $8mm$ instead of a $5mm$ lens would decrease the size of the field of view from $0.7m$ (Section 2) to $0.43m$. Therefore, between the lenses that are suggested by the camera manufacturer [The Imaging Source Europe GmbH (2006)], we can choose either the $5mm$ or the $8mm$ model, with the former option rendering a larger field of view and allowing larger N values, thus increasing the flexibility as far as the algorithmic options are concerned.

3.4 The Effect of the Frame Rate

Apart from the focal length of the lens, the frame rate of the camera also significantly affects the value of the per frame translation T. As a reference, Fig. 10 presents the T values derived for the indicated frame rates, for various vehicle speeds, altitudes and for the suggested focal length of $5mm$. Frame rates up to $r = 35f/s$ have been considered. However, for the given camera, the frame rate cannot exceed $r = 15f/s$, and thus, for our system, Fig. 10b gives the best possible performance.

At a given point of the 3D graphs of Fig. 10, contiguous $3D$ reconstruction is feasible if at that point $T \leq T_N$. Therefore, after determining the number of input frames N, and thus the value of T_N, the graphs of Fig. 10 can be used as a reference to assess the feasibility of contiguous reconstruction for the given frame rate.

4. CONCLUSIONS

In this paper, the feasibility of subsea surface reconstruction is evaluated, with respect to the vehicle speed and the three critical imaging parameters, namely the focal length of the lens, the frame rate of the camera and the exposure time.

To enable contiguous vision-based 3D reconstruction of the ocean floor, the per frame translation in pixels T should not exceed the $T_N = C/N$ value that corresponds to the given number of reconstruction inputs. As demonstrated in Figures 8 and 9, for a frame rate of $15f/s$, a focal length of $5mm$ enables contiguous reconstruction for $N \leq 5$, at a $2m/s$ speed and at altitude as low as $0.5m/s$. If a faster camera is employed, the higher frame rate redetermines the parameter space, as demonstrated in Fig. 10.

Finally, the exposure time of the camera should be equal to its minimum value that allows sufficient light to reach the CCD sensor so as to render valid pixel intensities. This would be the optimal point of the motion blur and light integration trade-off. This point will be experimentally determined based on real underwater images that correspond to various exposure times.

REFERENCES

Argyriou, V. and Petrou, M. (2009). Photometric Stereo: An Overview. *Advances in Imaging and Electron Physics*, 156, 1–54.

Barsky, S. and Petrou, M. (2003). The 4-source Photometric Stereo Technique for Three-Dimensional Surfaces in the Presence of Highlights and Shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10), 1239–1252.

Campos, R., Garcia, R., and Nicosevici, T. (2011). Surface Reconstruction Methods for the Recovery of 3D Models from Underwater Interest Areas. In *OCEANS*, 1–10.

Johnson-Roberson, M., Pizarro, O., Williams, S.B., and Mahon, I. (2010). Generation and Visualization of Large-Scale Three-Dimensional Reconstructions from Underwater Robotic Surveys. *Journal of Field Robotics*, 27(1), 21–51.

Marques, E., Pinto, J., Kragelund, S., Dias, P., Madureira, L., Sousa, A., Correia, M., Ferreira, H., Goncalves, R., Martins, R., Horner, D., Healey, A., Goncalves, G., and Sousa, J. (2007). AUV Control and Communication using Underwater Acoustic Networks. In *OCEANS*, 1–6.

Petrou, M. and Petrou, C. (2010). *Image Processing: The Fundamentals*. Wiley, 2 edition.

Singh, H., Roman, C., Pizarro, O., Eustice, R., and Can, A. (2007). Towards High-resolution Imaging from Underwater Vehicles. *The International Journal of Robotics Research*, 26(1), 55–74.

The Imaging Source Europe GmbH (2006). Lenses - Selection and Setup. www.theimagingsource.com (website last visited in January 2012).

Vidal, R. and Hartley, R. (2008). Three-View Multibody Structure from Motion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2), 214–227.

Yoerger, D.R., Jakuba, M., Bradley, A.M., and Bingham, B. (2007). Techniques for Deep Sea Near Bottom Survey Using an Autonomous Underwater Vehicle. *The International Journal of Robotics Research*, 26(1), 41–54.

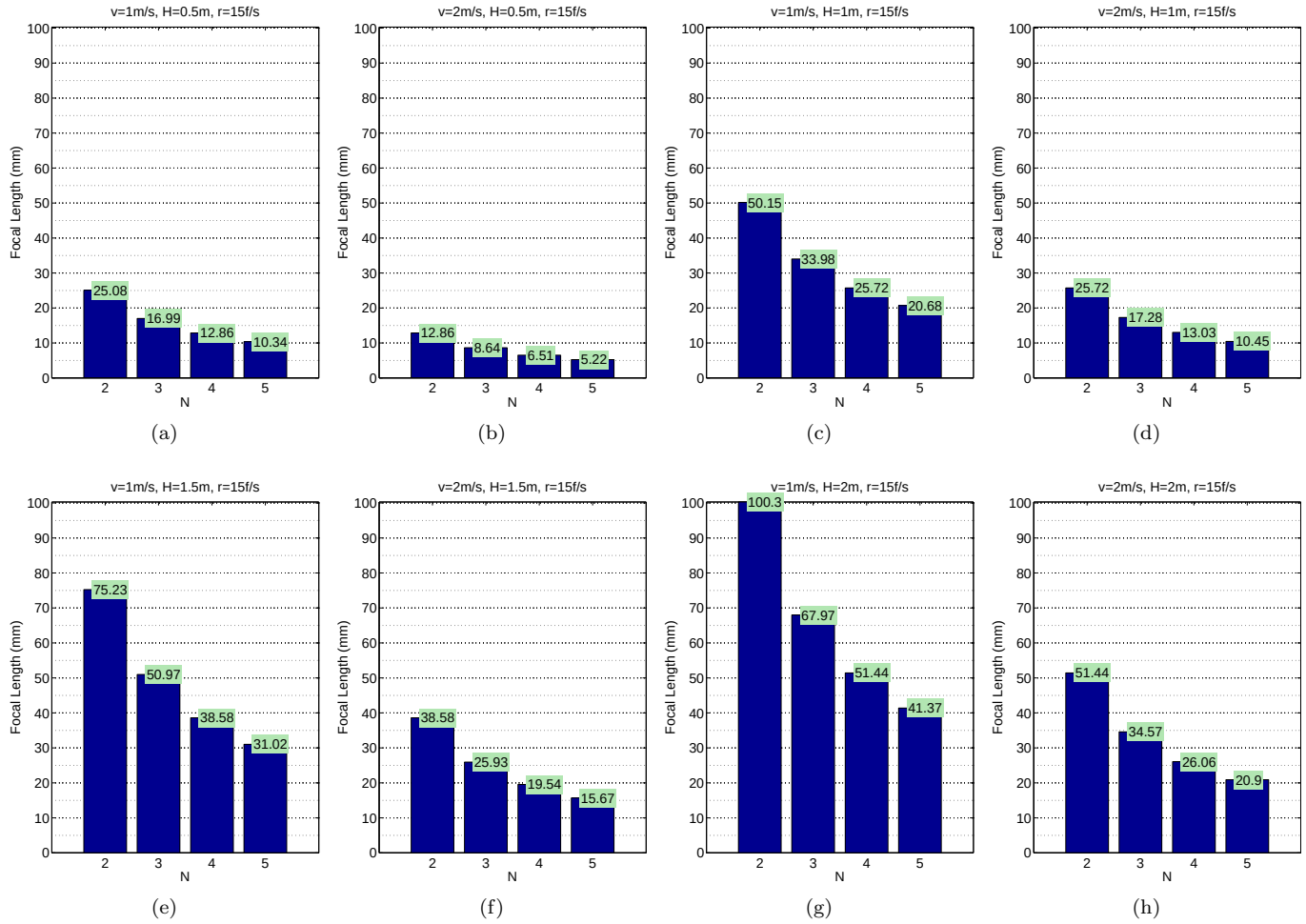


Fig. 9. Focal length F for $N \in [2, 5]$, at frame rate $15f/s$, for selected vehicle speeds and altitudes (as indicated).

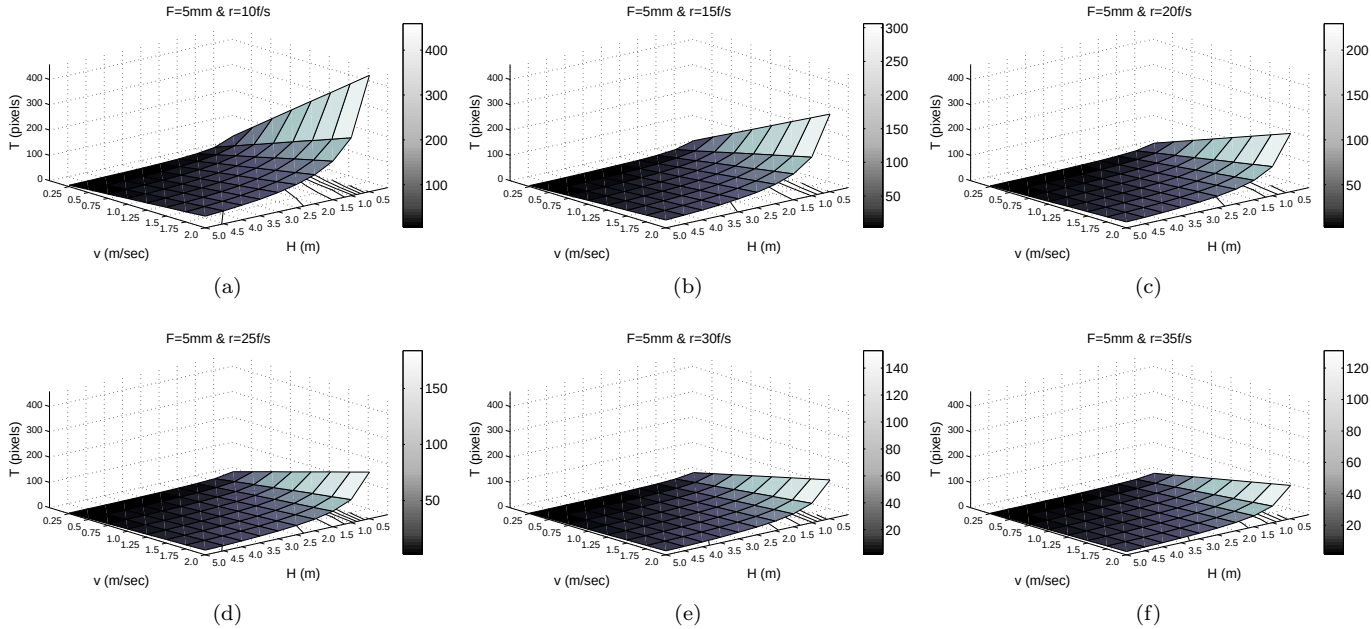


Fig. 10. Per frame translation T at the indicated frame rates, for various vehicle speeds and altitudes, when $F = 5mm$.