Classification of Textures Distorted by Water Waves

Arturo Donate Gary Dahme Eraldo Ribeiro

Florida Institute of Technology
Department of Computer Sciences
Melbourne, FL 32901, U.S.A.
{adonate, gdahme, eribeiro}@fit.edu

Abstract

In this paper, we approach the novel problem of classifying images of underwater textures as observed from outside the water. Our main contribution is to combine a geometric distortion removal algorithm with a texture classification method to solve the problem of classifying images of submerged textures when the water is disturbed by waves. We show that by modeling the separate types of distortion, we can extract enough texture information to correctly classify textures using spatial statistical measurements on the texton representations. We evaluate our algorithm on both natural and artificial textures acquired in our laboratory. Results are promising and show the feasibility of our algorithm.

1 Introduction

In this paper, we approach the novel problem of classifying images of underwater textures as observed from outside the water. This is an interesting problem with a number of practical applications such as the search for objects after an aircraft accident or the automatic recognition of shallow water coral reefs.

Our main contribution is the development of a method for classification of textures distorted by water waves. We first present an algorithm for reconstructing images of submerged objects containing distortions caused by water waves. The method described in this paper is similar to the approach proposed by Efros et al. [4], but differs from theirs by separately addressing the distortions caused by refraction and motion blur. Additionally, we present a modified version of the histogram-based texture classification algorithm introduced in [8] by incorporating co-occurrence statistics onto the texture descriptor. In summary, our approach combines geometric distortion removal with texture classification to solve the problem of classifying images of submerged textures. We test our algorithm on both natural and artificial textures captured in our laboratory. Our results show that the undistortion algorithm retrieves enough texture information to achieve correct classification results using the co-occurrence based texture classifier.

Several authors have attempted to solve the problem of undistorting an image of an underwater object as observed from above the water [4, 9, 10]. Recently, Efros et al. [4] described an effective approach to this problem. They assume the temporal distortion distribution of local image subregions over time is approximately Gaussian. Efros et al. use an implementation of the shortest path algorithm guided by a normalized cross-correlation measure to determine the frame region containing minimal distortion (i.e., the local plane containing the region is fronto-parallel to the camera's image plane).

On the other hand, recent work on texture classification has focused on the modeling and recognition of surface materials [8, 11, 3]. Leung and Malik [8] introduced a filter bank-based discriminative method that models the essential local structures and attributes of textures. Leung and Malik classify textures by means of chi-squared similarity measurements between 1D frequency histograms of vector quantized texture images, producing excellent classification results. Extensions of Leung and Malik's work have been proposed to address problems regarding rotation and scale invariance [11, 3]. However, the loss of important spatial structural information during histogram generation may sometime result in erroneous texture classification. This is mainly the case for textures consisting of similar texture primitives (i.e., textons), but distinct spatial arrangements.

2 Refraction and Motion Blur

In this section, we discuss the effect of refraction and motion blur on images of a submerged planar object. The camera is located above the water and has the optical axis perpendicular to the water surface. We assume that the surface of the water is disturbed by waves of unknown magnitude and speed. Additionally, we also consider that both the

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

IEEE COMPUTER SOCIETY

observed object and the camera are stationary. Figure 1 illustrates the geometry of the problem. Refraction is driven

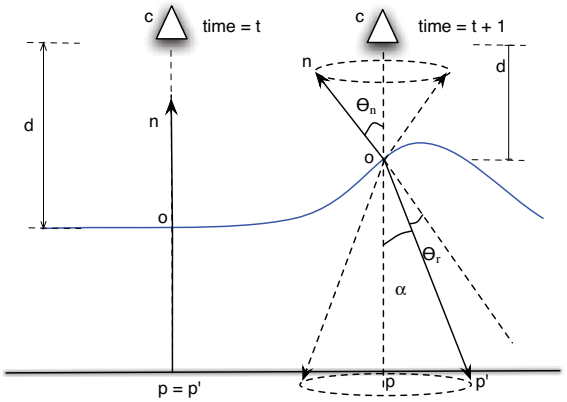


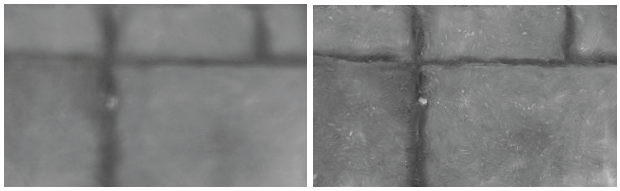
Figure 1. Refraction geometry with surface waves.

by Snell's Law $\sin\theta_n = r_w \sin\theta_r$, where θ_n corresponds to the angle of the water's surface normal, θ_r is the angle of refraction between the surface normal and the refracted viewing ray $\overrightarrow{op}$, and r_w is the refraction coefficient of the water, equal to 1.33. In Figure 1, α is the angle that measures the amount of translation of the end point $\mathbf{p}$ of the viewing ray $\overrightarrow{op}$ due to refraction. The two main variables that drive the refraction distortion in Figure 1 are the angle of the water's surface normal (θ_n), and the distance between the camera and the point of refraction (d). This effect causes the end point of the viewing ray to translate by an amount $\|\mathbf{p}' - \mathbf{p}\| = d\tan[\theta_n - \arcsin(\frac{\sin\theta_n}{r_w})]$. As a result, when observing a submerged object over an extended period of time, a given image-frame subregion will often contain information from neighboring subregions. There are two key points to be considered. First, image distortion will vanish whenever the camera's optical axis is parallel to the surface normal of the water. Secondly, experimental evidence points out that the spatial distribution of location of the end point of the viewing ray is approximately Gaussian [4]. The dashed cone in Figure 1 describes the amplitude of the translation of $\mathbf{p}$.

Motion blur is the second main source of distortion in the scenario described above. It is caused by the interplay between the speed of oscillation of the surface waves and the camera's limited frame rate. Field and Brady [5] suggested that image blur can be quantified by measuring the amount of decay in high-frequency energy of the Fourier spectrum. Measuring blur by comparing high-frequency energy decay might not always work for images of different objects, as a simple reduction in the total number of edges would also cause a decay in high-frequency spectral energy with no apparent decrease in image sharpness [12]. Next, we describe an approach to remove both types of distortion.

3 Removing Distortions

We now briefly describe the details of our distortion removal approach. For a more detailed analysis of the algorithm, we refer the readers to [2]. A single frame in a video sequence of N frames will contain subregions with variable amounts of distortion due to both refraction and motion blur. As a result, every video frame is likely to present some degree of distortion. The goal of our algorithm is to obtain an undistorted image of the submerged scene and then classify the observed texture. We accomplish this by combining image subregions with minimum distortion, selected from a large number of video frames, to form a single undistorted image of the submerged object. We commence by subdividing the video frames into rectangular subregions of equal size. The result of this step is an ensemble of frames describing the information viewed at that local region over time. Our method works on these individual ensembles of subregions. In contrast with the method proposed by Efros et al. [4], our method considers the distortion caused by both refraction and motion blur. For each ensemble of images we conduct the following procedure. First, we remove images centered at points in which the viewing ray endpoint suffered a large amount of translation (i.e., large degree of refraction-induced translation). We accomplish this by using the K-means algorithm to cluster frames into 10 groups of similar level of refraction distortion. The initial number of clusters was chosen experimentally. Clustering is repeated until each cluster contains no less than 10% of the total number of frames. The ten percent threshold has been experimentally found to be the minimum number of frames required by our algorithm. In the worst-case scenario, the algorithm will converge to a single cluster. In the second step of the algorithm, we select the cluster containing the least amount of translation between frames. We take a statistical approach to this problem by selecting the cluster with the least overall temporal pixel variance (i.e., the cluster where the frames are the closest in appearance to each other). All remaining clusters are then discarded. The subregions in the selected



(a) Frame average (b) Our method

Figure 2. Reconstruction of a submerged tile texture.

cluster have very low levels of refraction induced distortion. We now address the distortion caused by motion blur. Here, we follow Field and Brady [5] and quantify the level of blur in each frame by measuring the degree of decay in

IEEE COMPUTER SOCIETY

high-frequency energy in the Fourier power spectrum. We express the power spectrum in polar coordinates [6] to produce a function $S(r,\theta) = \sum_{\theta=0}^{\pi} S(r,\theta)$, where r and θ are the variables of the polar coordinate system. The algorithm selects the frames with highest frequency content (i.e., above the mean value over frames). These are the images with the highest degree of sharpness. The outcome of this step is a reduced set of image frames with minimal amounts of refraction and motion blur. In the final step, we take an approach similar to [4] and use the normalized cross correlation (NCC) to calculate distances between frames. The aim of this step is to determine the frame that represents the center of the distribution of all frames. Our approach avoids the leakage problem in which blurred subregions may result in erroneous NCC [4] by considerably reducing the number of frames used in the final frame selection. Figure 2 shows an example of a reconstruction obtained by our algorithm.

4 Spatial Statistics for Texture Classification

This section describes our texture classification method. Here, we extend the original method proposed by Leung and Malik [8] by incorporating spatial statistics of textons. Our current implementation is based on measurements on the gray-level co-occurrence matrix [7] calculated on texton maps. For more analysis and experiments on this method, we refer the readers to [1]. Leung and Malik's method models the basic repetitive elements of texture images as cluster centers of responses of a bank of convolution filters. The center of each cluster will represent a texton (i.e., basic texture element). The textons are then grouped to form a dictionary that is used to "explain" the appearance of general textures. Classification is accomplished by comparing 1D frequency histograms of vector quantized maps of the original textures based on the texture dictionary. Histogram comparison is achieved using the chi-square similarity measure $\chi^2 = \sum_i \frac{(p_i - q_i)^2}{p_i + q_i}$ where p_i and q_i are 1D histograms.

While the chi-square similarity is a suitable means for classification of textures exhibiting statistically stationary features, the case may arise where two textures that share similar frequency statistics will nevertheless differ greatly in terms of spatial properties. An effective way of measuring the spatial interaction between elements of each texture is by calculating spatial statistics on the texton map. In our implementation, we use the gray-level co-occurrence matrix (GLCM) [7]. We use 4 of the 16 originally proposed texture descriptors for the GLCM [7]: contrast, correlation, energy, and homogeneity. Each measurement is calculated along four orientations (i.e., 0^o, 45^o, 90^o, and 135^o) with a distance of 10 pixels to account for interactions among textons. As a result, we make 16 measurements for each texture and use them to build our texture descriptor.

Figure 3 illustrates the importance of quantifying the spatial relationship between textons. Each row in the figure shows the calculated texture descriptors for pollen grain surfaces. The first column shows the original images. The second column shows the corresponding texton maps generated using the texton-based modeling proposed by Leung and Malik [8]. The third column displays the resulting texton frequency histograms for each texton map. Finally, the last column displays a histogram representation of our texton co-occurrence descriptor. The two textures are clearly very distinct. The texture in the top row is composed of elongated texture patterns while the one in the bottom row has large blobs as texture elements. Yet, both textures produce very similar texton frequency histograms. Our method produces significantly distinct measurements both in shape and amplitude. This is an example when the frequency-based method does not always provide enough information for the discrimination of certain textures. It is worth noting that an increase in the number of textons (i.e., bins in the histogram) is likely to increase the discrimination power of the texton frequency histograms.

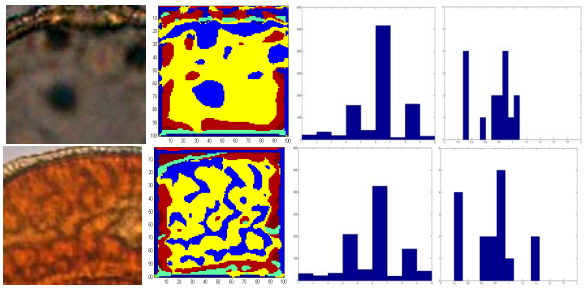


Figure 3. Similar frequency histograms of distinct textures. (a) Texture images; (b) Texton maps; (c) Texton frequency histograms; (d) Texton spatial co-occurences.

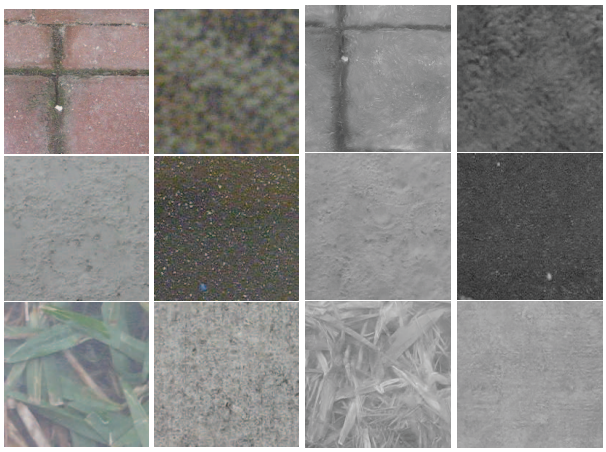
5 Experiments

In this section, we demonstrate the effectiveness of our algorithm for classification of submerged textures observed from outside the water. Our experimental setup consists of a glass water tank and a digital camera on a tripod placed inside of the tank. The water tank was placed on top of each texture (i.e., grass, concrete, sidewalk), and the camera captured a few seconds of video while the water surface was manually disturbed to generate waves. A set of undistorted texture images were first used to train our texture classifier. The videos are then processed by the image recovery algorithm to obtain the reconstructed images. We classify each texture with respect to the ones in the training set.

In the experiments, we chose 6 textures for the classification task: brick, carpet, concrete, tile, grass, and sidewalk. For each texture, 5 undistorted subregions of images were used to train the classifier. Subregions were 200×200 pixels

0-7695-2521-0/06 $20.00 © 2006 IEEE

IEEE COMPUTER SOCIETY



(a) Test images (b) Classification images

Figure 4. Original and reconstructed images

in size. Each video sequence of the distorted textures is first processed by our algorithm to remove the distortion from the data. The resulting images are then classified against the training set. Figure 4 shows sample subregions used by the algorithm. In the figure, column (a) and column (b) show the training images and the reconstructed images used for classification, respectively. We used 12 undistorted images for classification. Table 1 shows rates of correct texture classification. It is important to note that color was not used for classification in order to test the ability of the image recovery algorithm to recover sufficient texture information in order to to perform correct classification.

Table 1. Correct Classification Results

Number	Texture	Percentage (%)
1	Brick	100
2	Carpet	92
3	Concrete	92
4	Tile	83
5	Grass	83
6	Sidewalk	17

Our algorithm achieved a correct classification rate of 80% for most textures in this experiment. The features of the brick texture are largely different from all other textures in our experiments, hence the classification rate of 100%. We would expect the correct classification percentage for the brick texture to decrease slightly if we had included other textures with similar features (e.g., small tiles). On the other hand, the sidewalk texture provided the worst classification results. In this case, the algorithm was only able to correctly classify 2 out of the 12 subregions of the undistorted image. We attribute this to the fact that the algorithm was unable to reconstruct many of the sidewalk's small, distinguishing features.

6 Conclusions

Recognizing underwater textures from videos with surface waves is a novel and challenging problem. Potential applications include automatic object detection in shallow water as well as video-based coral reef recognition.

Our contribution in this paper is to combine a geometric distortion removal method with a texture classification method to solve the problem of classifying images of submerged textures when the water is disturbed by waves. We also suggest an improvement on texture classification methods based on texton frequency histograms by adding spatial statistics measurements on the texton maps. Results are promising and show the feasibility of the idea.

Future work includes extending the method to add neighborhood consistency among local overlapping subregions. We are currently working on the development of algorithms for video-based coral reef recognition in shallow water.

References

[1] G. Dahme and E. Ribeiro. Spatial Statistics of Textons. *International Conference on Computer Vision Theory and Applications*, pages 13–19, 2006.

[2] A. Donate and E. Ribeiro. Improved Reconstruction of Images Distorted by Water Waves. *International Conference on Computer Vision Theory and Applications*, pages 228–235, 2006.

[3] J. Dong and M. J. Chantler. Capture and synthesis of 3D surface texture. *International Journal of Computer Vision*, 62:177–194, 2005.

[4] A. Efros, V. Isler, J. Shi, and M. Visontai. Seeing through water. In *Advances in Neural Information Processing Systems 17*, pages 393–400. MIT Press, Cambridge, MA, 2005.

[5] D. Field and N. Brady. Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research*, 37(23):3367–3383, 1997.

[6] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley, Reading, MA, 1992.

[7] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3:610–621, 1973.

[8] T. Leung and J. Malik. Recognising surfaces using three-dimensional textons. In *IEEE International Conference on Computer Vision*, pages 1010–1017, 1999.

[9] H. Murase. Surface shape reconstruction of a nonrigid transparent object using refraction and motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(10):1045–1052, October 1992.

[10] R. Shefer, M. Malhi, and A. Shenhar. Waves distortion correction using cross correlation. `http://visl.technion.ac.il/projects/2000maor/`, 2001.

[11] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62:61–81, 2004.

[12] Z. Wang and E. P. Simoncelli. Local phase coherence and the perception of blur. In *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

0-7695-2521-0/06 $20.00 © 2006 IEEE

IEEE COMPUTER SOCIETY