UNDERWATER VIDEO MOSAICING FOR SEABED MAPPING

Y. Rzhanov, L. M. Linnett

Center for Coastal and Ocean Mapping (C-COM)
University of New Hampshire
Durham 03824, USA

R. Forbes

International Centre for
Island Technology
Heriot-Watt University,
Edinburgh, UK

ABSTRACT

This paper presents improved techniques and applications in mosaicing of underwater video images. The applications are of use to many marine scientists. High resolution seabed maps are created. The improvements in the processing relate to removal of interframe variability and lighting, speed up of the mosaicing process and improved accuracy in the estimation of the transformation parameters. Results are presented for real data acquired under a variety of circumstances and scenes.

1. INTRODUCTION

Underwater imaging is usually severely restricted due to poor visibility conditions. In order to image relatively large areas of the seabed, many small areas have to be photographed in natural or artificial light. The need for such scenes typically occurs in underwater archaeology, marine biology studies and seabed/pipeline survey. At present this can be achieved only by means of video mosaicing, to provide a seamless combination of sequence of overlapping frames in a single high resolution image.

Video mosaicing is a rapidly growing area of research and development, dealing with greatly varying data, acquired by different means, of different quality and for different purposes. Currently there is no universal technique for the processing. There are two currently accepted techniques - one relying on feature extraction [1] and the other is based on image analysis in the frequency domain.

The present paper concentrates on processing techniques specifically for underwater images, addressing the problem of faster processing for real time mosaicing and data compression, and proposes several potential applications.

Images of the seabed (typical examples are shown in Fig. 1) are often lacking sharp pronounced features (like corners, edges or spots, which are used for feature extraction and mosaicing of aerial images) and this suggests the use of the Fourier-Mellin transform for co-registration of adjacent frames in a video sequence [3, 2]. Usually, underwater imaging is interested in planar views as opposed to panoramic views. Hence we need not perform projective transformations (considered in details in, say, [5]). Typically, video sequences are acquired by a diver, or an AUV (Autonomous Underwater Vehicle), moving at an almost constant height over a seabed in a straight line, with camera looking vertically down at the area to be imaged.

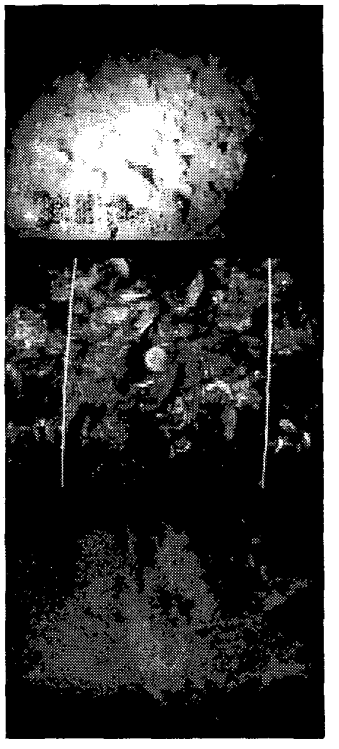

Fig. 1. Underwater images at various conditions.

Transformation relating two adjacent frames in our case is an affine transformation - it includes scaling, rotation and translation in two dimensions. Inevitably there are some perspective distortions - due to instability of the platform

0-7803-6297-7/00/$10.00 © 2000 IEEE

(diver or AUV) and the fact that the captured area is not flat. However, as the camera position is not calibrated, and orientation angles are constantly varying in an almost stochastic manner, this makes appropriate correction virtually impossible, so we ignore small local perspective distortions, assuming that they would lead to some local errors but will not spoil the overall appearance of the mosaic. Significant perspective distortions (diver moving across a prominent ridge, for example) would cause significant errors, and we will discuss the nature of these errors in the paper.

2. PROCESSING

The standard video frame rate (30 fps) is excessive for a typical diver's speed of 1 m/s. As the optimum overlap between frames is about 60-70 percent, and the area covered by a single frame is of the order of 2×2 square meters (based on the visibility in the Northern seas), the required frame rate for the processing is only about 2 fps. Video footage acquired by a diver is digitized on a computer (PC with a video capture card or Silicon Graphics workstation with a video board) with a frame rate of 2-3 frames per second. The first frame of the clip determines the orientation and scale of the mosaic - the rest of frames are adjusted to match the first one.

It is known that two-dimensional translation between two raster images can be determined using the phase shift theorem [4]. The procedure is as follows. The Fourier transform of each image is taken. Then the inverse Fourier transform is applied to a composite transform, consisting of the magnitude of the transform of one of input images, and of the phase which is equal to the phase difference of the transforms. The resulting image shows a sharp peak, offset of which is equal to translation vector between the original images.

The rotation and scaling between two video frames are first obtained using a Fourier transform and log-polar representation [2]. Fourier spectra of two images of the same area of a seabed are rotated and scaled replica of each other - translational information is not encoded in the magnitude spectrum. Thus, this information is left out on the first processing stage. "Polar" transform converts rotation (about center of the magnitude spectrum image) to translation in one direction (say, X), while taking "log" of the radius converts scaling to translation in another direction (Y) (magnitude spectrum of an image and its log-polar transform are shown in Fig. 2). Application of the Mellin transform [6] to the transformed images allows to determine rotation and scaling factors.

With these factors known, the second image is transformed (scaled and rotated) to the framework of the first image. The difference between images is now purely translational, and another application of Fourier-Mellin transform finds the corresponding translation.

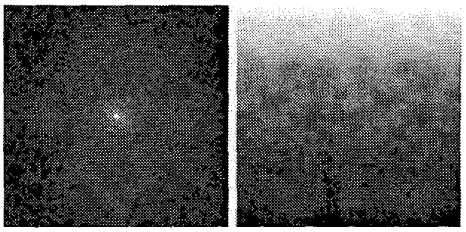

Fig. 2. Processing stages: Fourier magnitude spectrum (left) and its log-polar transform.

An example of a mosaiced image sequence is shown in Fig. 3. This represents a portion of a larger mosaic (5×5 meters) created to determine the horse mussel density in an area off the North coast of Scotland.

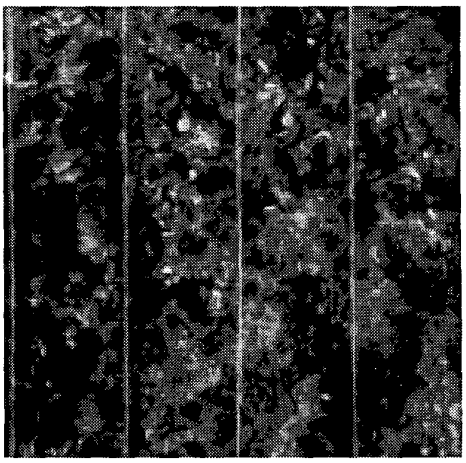

Fig. 3. Seabed image mosaic. Vertical lines are a rope grid used for area segregation.

3. SPECIFIC PROBLEMS AND SOLUTIONS

In this section we outline some processing difficulties and discuss methods to overcome them.

1. Underwater images are often acquired when the natural light is not sufficient for photography. The use of a light source attached to the camera (and hence moving with it) creates strong lighting artifacts - inhomogeneities, which are features of the frames, not of the captured imagery. The co-registration of such frames tends to suggest that there is no translation between them, as lighting inhomogeneities are the most prominent features in these images (Fig. 4, top image).

We have found an efficient way of eliminating these artifacts is through de-trending - in this case a surface is fit to the heightmap representing pixel values of the frame and then subtracted from the image. *A priori* knowledge about the nature of the light source may suggest the best shape of the surface, but usually it is sufficient to use a general low-order two-dimensional polynomial spline. Averaging of the spline coefficients over several frames provides further improvement in the removal of the lighting artifact.

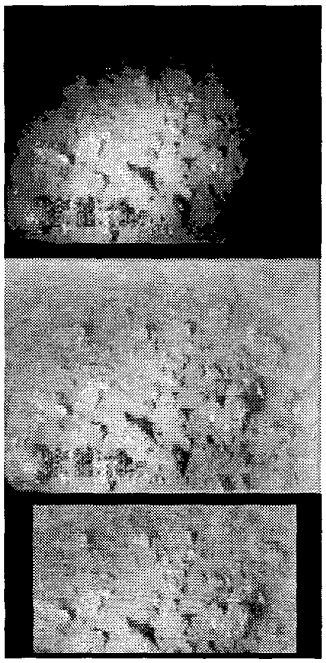

Fig. 4. Lighting artifacts removal. Original frame (top), detrended frame (middle), cut-off portion for mosaicing (bottom).

2. To limit windowing artifacts a Gaussian window is applied to each image. Secondly, an equalizing logarithmic transform is applied to the magnitude spectrum before log-polar transformation to enhance high frequency information.

3. Due to symmetry properties of the Fourier transform, a log-polar transform has to be applied to only half of the transform (the other half contains the complex conjugate of the first half). However, in this case, the resulting image will also have edge discontinuities, as in (3) above. Use of the whole image eliminates discontinuities on the vertical edges, while Gaussian masking in vertical direction reduces discontinuities on the horizontal edges.

4. SPEED IMPROVEMENTS

A single peak appearing as a result of the Fourier-Mellin transform, is caused by the fact that the Fourier transforms of two spatially translated images differ by a constant phase shift. Consider two one-dimensional images (signals) with the same localized pulse of arbitrary shape, positioned in different locations on a black (zero) background. These images are rasterised (sampled) with a given frequency. Magnitude spectra of their FT are the same. Phase spectra differ by the same amount: phase, corresponding to DC level is the same; phase of the lowest frequency differs by $\delta\phi$; the next - by $2\delta\phi$; etc. The value of $\delta\phi$ is related to a spatial shift between the pulses: if the total number of samples is N, pulses are located at samples k_1 and k_2 respectively, then:

$$\delta\phi = 2\pi\frac{k_1 - k_2}{N} = 2\pi\frac{\delta k}{N}$$

This actually implies that to calculate the translation, we don't have to perform an FFT, subtract phases, perform an inverse FFT - and then to have to search for a spike! All that is needed, is to use a DFT and compute the single m-th coefficient of the Fourier transform for both images. The translation δk between images is then:

$$\delta k = \frac{\delta\phi}{2\pi}\frac{N}{m}$$

where $\delta\phi$ is the phase difference for the m-th coefficient. This guarantees at least $2N$ times speed up! (And no calculation involving the magnitude spectrum.)

Reality is not that simple, unfortunately. Images are not just translated replica of each other. They have parts that are unique and do not have counterparts. These give rise to non-coherent contributions to the phase distribution (while similar parts of the images contribute to phase coherence).

Assuming that overlap is significant, we thus compute not a single, but a few (J) Fourier coefficients m_j and average the result:

$$\delta k = \frac{1}{J}\sum_{j=0}^{J-1}\frac{\delta\phi}{2\pi}\frac{N}{m_j}$$

With $J \ll N$, this gives significant savings in processing time.

Although translation refers to sampled data (rasterised image), the translational vector is not necessarily integer. If this is the case, the combination of two frames can be done in different ways. First, translation can be rounded to the nearest integer, and the composite image is then created by per-pixel averaging. Another approach is to use an interpolation technique (bi-linear, for example) and to treat translation as non-integer. Both techniques produce a smoothing

effect (low-pass filtering), and the mosaic loses some of its sharpness.

5. CONCLUSIONS

This paper has presented a technique for video mosaicing of underwater images. The method has been applied to underwater pipeline survey data, shipwreck data and for marine biological measurements. Specifically a robust technique has emerged which allows for removal of lighting artifacts and mosaicing of image sequence data at rates matching real time acquisition. The technique shows much promise for the gathering of new types of information from the seabed using video as opposed to sonar. Video having the advantage of being WYSIWYG, whereas sonar is time-dependent. With the increasing use of highly mobile AUV's this type of data gathering will increase and scientists will gain further information from this relatively new form of imagery.

6. REFERENCES

[1] N. Gracias and J. Santos-Victor, 1998, Automatic mosaic creation of the ocean floor. OCEANS'98 Conference Proceedings, vol. 1, pp. 257-262.

[2] J. Davis, 1998, Mosaics of scenes with moving objects. IEEE Comput. Soc. Proceedings, vol. 1, pp. 354-360.

[3] B. S. Reddy and B. N. Chatterji, 1996, An FFT-based technique for translation, rotation and scale-invariant image registration. IEEE Transactions on Image Processing, vol. 5, pp. 1266-1271.

[4] C. D. Kuglin and D. C. Hines, 1975. The phase correlation image alignment method. IEEE Conference on Cybernetics and Society, pp.163-165.

[5] S. Mann and R. W. Picard, 1997, Video orbits of the projective group: a simple approach to featureless estimation of parameters. IEEE Transactions of Image Processing, vol. 6, pp. 1281-1295.

[6] D. Casasent and D. Psaltis, 1997, Position oriented and scale invariant optical correlation. Applied Optics, vol. 15, pp. 1793-1799.