A Survey on: Underwater Video Processing for Detecting and Tracking Moving Objects

M. S. Srividya, Shobha G.

Abstract: Oceanographers need to study, analyse and interpret the biological and physical characteristics of marine organisms in the waterbed and sea floor. Images and Videos are important source of information and aids for their study. However, there are unique set of constraints in underwater environment that have limited our ability to process underwater images. Some of the important constraints in underwater images are associated with the physics of the light and attenuation of the electromagnetic spectrum. Processing issues also need to be dealt for the required application. Some of the research issues underwater have been the tasks associated with reconstructing three-dimensional information about the world from its two-dimensional Projections. In this paper the techniques of underwater video processing for detecting and tracking moving objects are discussed, analysed and compared.

Keywords: Video Processing, Detection, Tracking, Affine Transformation, Feature Extracion

I. INTRODUCTION

Traditionally, marine biologists determine the existence and analysis of different types of marine animals using several methods. Human manned photography and video-making, do not damage observed marine animals or their habitat, the collected samples are scarce or limited and is intrusive to the observed environment therefore do not capture normal organisms behaviours. In [4] they propose an automated Video processing system for detecting, tracking and counting fishes. Feature tracking is a key, underlying component in many approaches to object reconstruction, detection, localization and recognition of underwater objects. In [2], they propose to adapt Scale Invariant Feature Transform (SIFT) technique for feature tracking in underwater video sequences. The SIFT extracts features, which are invariant to scale, rotation and affine transformations. They have compared and evaluated SIFT with popular techniques. In underwater environment, extracting the invariant features from the sequence of images is a challenging due to optical properties of the water, which varies the same feature in sequence of images. In [5] some advances in colour restoration of underwater images, especially with regard to the strong and non-uniform colour cast which is typical of underwater images. The proposed colour correction method is based on Automatic Colour Equalization (ACE) model, an unsupervised colour equalization algorithm. ACE is a perceptual approach inspired by some adaptation mechanisms of the human visual system, in particular lightness constancy and colour constancy. It is unsupervised, robust and has local filtering properties that lead to more effective results.

In [6] they present an approach to video categorization aimed at facilitating a particular marine biology study. They develop vision algorithms that can address specific needs of marine biologists, such as fine-grained categorization of fish motion patterns. The approach consists of three steps. First, a fish detector is applied to identify and localize fish occurrences in each frame, under partial occlusion, and amidst dynamic texture patterns formed by whirls of sand on the sea bed. Then, tracking-by-detection is conducted. Given the similarity between fish detections, defined in terms of fish appearance and motion properties, fish tracking is formulated as transitively linking similar detections between every two consecutive frames, so as to maintain their unique track IDs. Finally, histograms of fish displacements along the estimated tracks are extracted.
Outlining the processing of underwater videos that is presented across most of the papers involves three important tasks. They are feature extraction, object detection and object tracking

II. FEATURE EXTRACTION

In [4], the aim of this subsystem is to detect the average texture and colour properties of each frame. The evaluated properties are: 1) Brightness: classified in Dark/Medium/ Bright; 2) Smoothness: classified in Blur/Clear; 3) Colour: identification of green colour tone, classified in Green/Not Green; Hue, Saturation and Value: classified in High/Medium/Low. The approach used for describing the image texture is based on analysing the statistical moments of the grey-level histogram. For colour analysis, they extract the Hue, Saturation and Value planes. For each plane, they compute the averaged percentage of the pixels whose value is larger than a suitable threshold in the video. To determine whether a video is *Green colour toned* or not, they determine the green plane of a frame and then count all the pixels in such plane which value is greater than 128. Next they aggregate those values to decide the overall green colour tone of the video.

Figure 1: Feature extracted video frame

For reconstruction of object [2] using Scale-Invariant Feature Transform (SIFT) technique for extracting and matching features in stereo video sequences is presented. They have captured coral reefs using

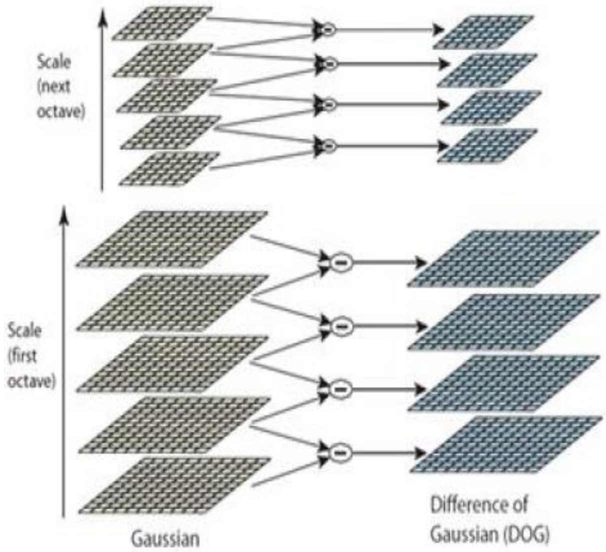
Manuscript Received March, 2014.

Prof. Srividya M. S., Asst. Prof, Dept of Computer Science &Engg., Global Academy of Technology, Bangalore, India
Dr. Shobha G., HoD, Dept of Computer Science &Engg.,RVCE, Bangalore, India.

two video cameras, which are aligned to capture stereo video sequences. [8] have adapted SIFT feature tracker for extraction of feature from images and recognizing the object in subsequent images.

In [2], they propose to adapt SIFT method for extraction of features from underwater monocular video sequences. SIFT is an algorithm in computer vision to detect and describe local features in images. The SIFT method is very suitable in the case where the interest points are invariant to image scaling and rotation, and partially invariant to changes in illumination. It is implemented efficiently by means of a Difference-of-Gaussian (DOG) function to identify potential interest points that are invariant to orientation and scale. Interest points for SIFT features correspond to local extrema of Difference-of-Gaussian filters at different scales.

Interest points (called keypoints in the SIFT framework) are identified as local maxima or minima of the DoG images across scales. Each pixel in the DoG images is compared to its 8 neighbors at the same scale, plus the 9 corresponding neighbors at neighboring scales. If the pixel is a local maximum or minimum, it is selected as a candidate keypoint.



<u>Figure 2:</u> The blurred images at different scales and DOG values

III. OBJECT DETECTION

One of the most common approaches in detecting and tracking targets in real time video applications is the temporal differencing (TD) technique. In this approach, video frames are separated by a constant time δt and compared to find regions which have changed. While this technique is fast, it has limitations. For instance, tracking is impossible if there is major camera motion, unless a proper image stabilization technique is employed. This approach also fails if the object becomes obstructed or terminates its motion. Template correlation matching is another approach that falls into the temporal differencing approach. The drawback of this approach is that it requires that the object of interest's appearance remains persistent and thus, it is not robust to changes in object size, orientation or even changes in the lighting conditions. There are many variants on the TD method but the easiest is to take consecutive video frames and define the absolute variance. A threshold function is then used to determine the change.

Other common methods are optical flow [9] and background subtraction [10] techniques. Two-dimensional(2D) image motion is the projection of the three-dimensional(3D) motion of targets from the world coordinates to the corresponding image plane [11].

In order to carry out foreground object detection and recognition even in a dynamic environment, two types of methods have been proposed: 1) using the motionless background model with an acceptable range of image differences at each pixel or local image area, and 2) using dynamically updated background model.

The method proposed in [2] is a robust background subtraction method for various illumination and motion conditions. This method has been developed to work for both indoor and outdoor scenes. The technique consists of two kinds of operation, one for removing the stationary background and the other is for removing shadows.

In general, shadows are of two classes: self and cast shadows. A self-shadow occurs in the portion of an object which is not illuminated by direct light. A cast shadow on the other hand is the area projected by the object in the direction of direct light [17]. There have been many approaches proposed to tackle this problem.

Object identification algorithms are commonly based on single image analysis, such as the extraction of a single video frame from a sequence. This mismatch of, in particular, temporal processing paradigms means that most object analysis algorithms are not well suited to the data with which they are presented. In order to bridge this gap, [1] investigate the temporal preconditioning of video data through a biologically-inspired vision model, based on multi-stage processing analogous to the vision systems of insects. In doing so, they argue that such an approach can lead to improved object identification through the enhancement of object perimeters and the amelioration of lighting and compression artefacts such as shadows and blockiness.

IV. OBJECT TRACKING

In [4] use a combination of two algorithms for tracking: the first one is based on the matching of blob shape features and the second one based on the histogram matching. They use a feature vector represented by the parameters of the object such as the centroid of the object, the area of the object and the orientation (in degrees) of the object.

Tracking by “colour matching” has been carried out by applying the Continuously Adaptive Mean Shift Algorithm (CamShift), which is an adaptation of the Mean Shift algorithm. Given a probability density image, Camshift finds the mean (mode) of the distribution by iterating in the direction of maximum increase in probability density. The Probability Distribution Function (PDF) of images adopted in our work is the Histogram Back-Projection, which associates the pixel values in the image with the value of the corresponding histogram bin.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

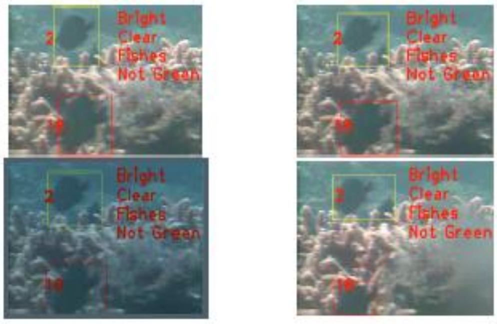


Figure 3: Tracking example with four frames grabbed at four consecutive times

In [2], the extraction of SIFT features from video sequences can be done by applying sequentially steps such as Scale-space extrema detection, Keypoint localization, Orientation assignment and finally Generation of keypoint descriptors.
Typically, a video sequence includes a static background and moving object like fish translating and warping along consecutive frames. In [5] they break up an image into regions of interest representing the object like fish. Simple background subtraction is used to segment the region of interest and then the segmented regions are labelled to track across the frames.

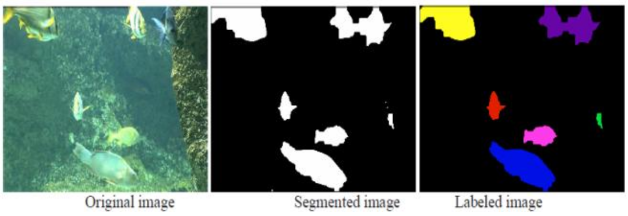


Figure 4: Segmentation and Tracking

V. OPEN AREAS

New algorithms for detection and tracking will be implemented in order to investigate improved efficiency. Furthermore, the algorithms developed to perform the video analysis, (such as pre-processing, detection, tracking and counting) could be integrated into a more generic architecture so that the *best* algorithm for each step will be selected. The performance level for the algorithms will be determined by a measure such as processing time or certain user provided requirements. Thus a combination of optimal algorithms to perform the video analysis could be utilized.

VI. CONCLUSION

The purpose of this survey is to provide an overview of the functionality of underwater video processing for detection and tracking of moving objects. Most systems use colour and texture features, few systems use shape feature, and still less use layout features. Various approaches has been merged and used in various areas to improve the performance of the system and to achieve better results in different applications.

REFERENCES

[1] SamanPoursoltan, Russell Brinkworth, Matthew Sorell "Biologically-inspired Video Enhancement Method For Robust Shape Recognition," University of Adelaide, Australia, IEEE, 2013.
[2] Prabhakar C J & Praveen Kumar P U. "Feature Tracking of Objects in Underwater VideoSequences.", Kuvempu University, India, ACEEE 2012.
[3] SaeedVahabiMashak, BehnamHosseini, S.A.R. Abu-Bakar."Background Subtraction for Object Detection under Varying Environments."UniversitiTeknologi Malaysia, IJCISIMA, 2012
[4] M. Weber, M. Welling, and P. Perona. "Detecting ,Tracking And Counting Fish In Low Quality Unconstrained Underwater Videos" University of Edinburgh, Edinburgh, UK.
[5] "Underwater ColorConstancy :Enhancement of Automatic Live Fish Recognition" M. Chambah, D. Semani, A. Renouf, P. Courtellemont, A. Rizzi, Université de La Rochelle, France, Dept. of Information Technology - University of Milano/Italy
[6] CodrutaOrnianaAncuti, CosminAncuti, Tom Haber and Philippe Bekaert "FUSION-BASED RESTORATION OF THE UNDERWATER IMAGES," Belgium, IEEE, 2011.
[7] MohamedAmer, Emil Bilgazyev "Fine-grained Categorization of Fish Motion Patterns in Underwater Videos", Oregon State University, ICCV, 2011
[8] PiotrJasiobedzki, Stephen Se, Michel Bondy, and Roy Jakola,"Underwater 3D mapping and pose estimation for ROVoperations", OCEANS 2008, pp. 1-6, September 2008.
[7] Li Ma, KaihuaWu,L. Zhu, "Fire smoke detection in videoimages Using Kalman filter and Gaussian Mixture Colormodel", International Conference on Artificial Intelligence andComputational Intelligence, 2010
[8] Li Xu, Feihu Qi, Renjie Jiang, YunfengHao, Guorong Wu,"Shadow Detection and Removal in Real Images: A Survey"Shanghai JiaoTong University, P.R. China, June 2006.
[9] J Shin, S Kim, et all, "Optical flow-based real-time object tracking using non-prior training active feature model"Computer Vision and Image Understanding, June 2005,Volume 98, Issue 3, Pages 462-490.
[10] B.Hosseini, SaeedVahabiMashak and A.S.R Abu Bakar,"Human Movement Based on Rule based classifier" Asia Modelling Symposium, May 2010.
[11] [14] S.S. Beauchemin and J.L. Barron, "The Computation of Optical Flow", ACM Computing Surveys, 1995 Vol. 27,No.3, PP 43-467.
[12] S.Y. Chien, S.Y. Ma and L.G. Chen ," Efficient moving objectsegmentation algorithm using background registrationtechnique" IEEE Trans Circuits Syst. Video Technol, JULY2002
[13] Hu Fuyuan, Zhang Yanning, et all, "A New Method of MovingObject Detection and Shadow Removal" JOURNAL OFELECTRONICS (CHINA), July 2007
[14] Dongxiang Zhou, Hong Zhang and NilanjanRay,"TextureBased Background Subtraction", Proceedings of theInternational Conference on Information and Automation,Zhangjiajie China, IEEE Conference, June 20 -23, 2008.
[15] YunlongGuo, Bo Yang, Yangyang Ming, Aidong Men, "AnEffective Background Subtraction Under the Mixture ofMultiple Varying Illuminations",Second InternationalConference on Computer Modeling and Simulation, IEEEConference, 2010.
[16] V. Brandou, A. G. Allais, M. Perrier, E. Malis, P. Rives, J.Sarrazin, and P. M. Sarradin, "3D Reconstruction of NaturalUnderwater Scenes Using the Stereovision System IRIS", OCEANS 2007 - Europe , pp. 1-6, 2007.
[17] David G Lowe, "Distinctive image features from scale-invariantkeypoints", *International Journal of Computer Vision*, vol. 60(2),pp. 91-110, 2004.
[18] David G Lowe, "Object recognition from local scale-invariantfeatures", Proceedings of the International Conference on ComputerVision, vol. 2, pp. 1150-1157