LBP-SURF Descriptor with Color Invariant and Texture Based Features for Underwater Images

Prabhakar C J
Department of P.G. Studies & Research in Computer Science
Kuvempu University, Shankaraghatta - 577451
Karnataka, India
psajjan@yahoo.com

Praveen Kumar P U*
Department of P.G. Studies & Research in Computer Science
Kuvempu University, Shankaraghatta - 577451
Karnataka, India
praveen577302@gmail.com

ABSTRACT

In this paper, we introduce LBP-SURF, a local image descriptor for underwater environment, which is very efficient to extract color invariant and texture based features of underwater images. The current state-of-the-art feature descriptors viz. SIFT, SURF, DAISY, GLOH and variants are well known techniques for detecting and describing features of objects captured in the out-of-water environment. These standard descriptors have been proven to be the most robust to geometric variations. Nearly, all these geometrical invariant approaches avoid dealing with color images due to the color constancy problem. In underwater images, variation in color is very high compared to variations in geometrical properties due to propagation properties of light. The literature survey reveals that, the texture parameters that remain constant for the scene patch for the whole underwater image sequence. This motivated us to consider texture and color invariant features of underwater images, instead of using the gray-based geometrical invariant features. We normalize the color image using comprehensive color image normalization method to render the color values changed by the various radiometric factors of underwater environment. Our method uses Speeded Up Robust Features (SURF) to detect interest points from the normalized image. The texture features are extracted, and description is stored using Center-Symmetric Local Binary Patterns (CS-LBP) descriptor. The combination of SURF and CS-LBP, called LBP-SURF is evaluated extensively to verify its effectiveness with datasets acquired in underwater environment.

Keywords

Feature Detection, Feature Description, SIFT, SURF, CS-LBP

1. INTRODUCTION

*Corresponding author

ICVGIP '12, December 16-19, 2012, Mumbai, India

Feature detection, description and matching of stable and discriminative features are fundamental problems in computer vision. The feature detectors and descriptors play a key roles in many vision applications, such as image stitching, image registration, object detection and object recognition. For good object description, two criteria should be satisfied in the extracted features. The first one is the stability, i.e. the extracted features should be invariant to different photometric and geometric changes. The second one is the distinctiveness, which means that the extracted features should have minimum information to distinguish between the object which they describe and other objects [1]. The features are highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from many images. In an underwater environment, extracting the invariant features from the images is a challenging task due to optical properties of the water, which varies the same feature in sequence of images. When imaging underwater, even an image of the same object can be drastically different due to varying water conditions. The added complexities of impure water introduce issues such as turbulence, air bubbles, particles (such as sediments) and organic matter that absorb and scatter light, which can result in a very blurry and noisy image. Moreover, many parameters can modify the optical properties of the water, and underwater images show large variations in color rather than geometrical variations. As a result, descriptors of the same point on an object may be completely different between two underwater images taken under same imaging conditions.

Color is an important component for distinction between objects. If the color information in an object is neglected, a very important source of distinction may be lost, and many objects can be mismatched and misclassified [1]. Color invariance features provide high discriminative information cues, as color varies considerably with change in camera viewpoint, object pose and illumination. Underwater images usually suffers from radiometric variations and using raw color information leads in degradation in the performance of feature detection. Color-constancy algorithms often attempt to separate the illumination and the reflectance components on images similar to the human visual system. The most simple and commonly used color-invariant approach for Lambertian surface are the Chromaticity normalization and the gray world assumption. Chromaticity normalization is often used to removal of lighting geometry effects where as gray world assumption is used for removal of illuminant color effects. Neither of the two methods remove the dependency of both lighting geometry and illuminant color simultaneously [12]. Finlayson et al., [5][6] developed a method called comprehensive color image normalization method which removes lighting geometry effects and illuminant color effects both iteratively and non-iteratively.

Pattern recognition techniques using color often operate on the color distribution alone, ignoring the spatial, black and white (tonal) property of regions in the image which defines the texture [17]. Texture is an inherent property of entities or scenes, which has the characteristics of brightness, color, shape, scale, etc. Texture description is based on tone and structure. Tone is based on the properties of the pixel intensity of the primitive, while structure is the spatial relationship between primitives. Fine textures have higher spatial frequencies, and its texture primitives are small while coarse textures have larger primitives and lower spatial frequencies [17]. The captured underwater color images are difficult to process because of the blurring of the elements of the image with high clutter in the regions of interest and lack of distinct features. It is useful, to use texture parameters because it remains constant for the same scene patch for the whole image sequence, this helps in efficient way to extract invariant features from the underwater images [7].

Texture classification, as one of the major problems in texture analysis, has been studied for several decades and shows striking improvements; however, the extraction of effective features for texture image representation is still considered as a challenging problem. One of the popular texture descriptor is Local Binary Pattern (LBP), which is fast and has been highly successful for various computer vision problems such as face recognition, video background subtraction and recognition of 3D textured surfaces. The LBP has a property that favors its usage in interest region description such as tolerance against illumination changes and computational simplicity. LBP operator is a statistical texture descriptor in terms of the characteristics of the local structure. LBP provides a unified description, including both statistical and structural characteristics of a texture patch, so that it is more powerful for texture analysis. The main drawback of LBP is that the operator produces a rather long histogram and is not too robust on flat image areas. To overcome the drawback found in LBP, Marko Heikkila et al. [11] modified the existing LBP to CS-LBP called Center-Symmetric Local Binary Patterns, which is robust on flat image areas usually found in underwater environment. CS-LBP descriptor outperforms the existing local descriptor for most of the test cases, especially for images with severe illumination variations. Instead of comparing every pixel with the center pixel, CS-LBP compares center-symmetric pairs of pixels. This reduces the number of comparisons to half of that of LBP.

The detection of interest points (keypoints) for a given image is a very challenging task due to variation in scale, orientation and illumination. The interest points are selected based on the distinguishable property of the locations, such as corners, blobs and T-junctions. The reliability of interest point detectors is evaluated based on repeatability criteria. The repeatability criteria finds the ability of feature detectors to consistently detect the same keypoints in two different images of the same scene under variations such as scale, illumination and view point changes. A wide variety of detectors has been proposed in the literature. The most widely used detectors can be classified into two categories such as Harris-based and Hessian-based detectors [10][16][18][19][20]. The literature survey reveals that, Hessian-based detectors are most stable and repeatable than Harris-based detectors. The Harris corner detector [10] proposed in 1988 is based on eigen values of the second moment matrix. This method is not scale-invariant due to the fact that the Harris corner detector is very sensitive to changes in image scale, so it does not provide a good basis for matching images of different sizes. To detect the blob–like structures, Lindeberg [15] proposed the concept of automatic scale selection based on Hessian matrix as well as Laplacian. Mikolajczyk and Schmid [19] have further extended this method to create robust and scale-invariant feature detector based on the combination of Harris-Laplacian and Hessian-Laplacian. They used a Harris measure to select the location and Laplacian to select the scale. Lowe [16] proposed to approximate the Laplacian of Gaussian (LOG) by a Difference of Gaussian (DOG) filter. SURF proposed by Herbert Bay et al. [3] is one of the widely used feature detector and descriptor, because of its robustness, distinctiveness and efficient computation time in feature extraction and matching. SURF uses Hessian-based feature detector because it is more stable and repeatable than Harris-based counterparts, which are usually used in SIFT and its variants. The determinant of the Hessian matrix rather than its trace seems advantageous, as it fires less on elongated, ill-localized structures. It is also observed that approximations like the DOG used in SIFT can bring speed at a low cost in terms of lost accuracy [3].

In this paper, we propose a technique called LBP-SURF to extract color invariant, and texture based features of underwater images. We apply comprehensive color image normalization to normalize the raw RGB color values, which helps in detection of reliable interest points. The interest points are detected using Hessian based approach used in SURF feature detection technique. The texture information of the interest point is extracted using CS-LBP descriptor. The robustness of the proposed method is evaluated based on repeatability criteria for feature detection and recall versus precision for descriptor. The experimental result reveals that LBP-SURF works well for underwater environment compared to other feature extraction techniques.

The paper is organized as follows: Section 2 describes a brief survey on existing feature detection and descriptors techniques that are applied on underwater images. The Comprehensive color image normalization is discussed in Section 3. In Section 4 we present the proposed LBP-SURF feature detection and description technique. The experimental results are demonstrated in Section 5. Finally, Section 6 draws the conclusion.

2. RELATED WORK

The researchers have developed feature descriptors, such as SIFT, SURF, GLOH, DAISY, etc., to detect and extract the geometric-based features from images for general purpose applications such as object recognition, classification and stereo matching. Over the past few years, the underwater vision is attracting researchers to investigate suitable feature descriptors for 3D reconstruction, mosaicing, image registration, object detection, localization and recognition of underwater objects. The literature survey reveals that the researchers have not been attempted to develop feature descriptors meant for underwater environment. Since there is no standard descriptor meant for underwater environment, the researchers have been using standard feature descriptors for underwater applications.

Trucco et al. [28] have adapted Kanade-Lucas-Tomasi (KLT) based feature detector for feature extraction and motion estimation in underwater environment. KLT is strongly based on Harris corner detector, where features are extracted using minimum eigenvalue of each 2×2 gradient matrix. Similarly, Plakas et al. [23] have adapted KLT based feature detector for extraction of features to 3D shape reconstruction from uncalibrated underwater video sequences. Anne Sedlazeck et al. [25] have adapted KLT based feature extraction method to reconstruct the 3D surface of a ship wreck using underwater monocular video sequence. Andrew Hogue and Michael Jenkin [13] also adapted KLT for 3D shape reconstruction of coral reefs using stereo image sequences. Brandou et al. [4] have adapted SIFT technique for feature extraction from underwater stereo images to reconstruct the 3D surface of coral reefs. They have captured images of coral reefs using two video cameras, which are aligned to capture stereo video sequences. To recognize the object in subsequent images, Jasiobedzki et al., [14] have used SIFT technique for feature extraction.

Kenton Oliver et al., [22] have evaluated quantitatively the combination of feature detectors and descriptors i.e. Harris and Hessian with SIFT and GLOH. They used the image captured out-of-water and convoluted with Point Spread Function (PSF) model to create a synthetic underwater images. The experimental result shows that major limitations are when dealing with photometric transformations introduced when imaging underwater environment and when the water gets murkier most detectors failed to extract interest point from images, which results in decrease in performance of descriptors. For underwater environment applications, Florian Shkurti et al. [26] have attempted to identify the suitable feature detector and descriptor. They evaluated different feature extraction techniques such as SURF, Shi-Tomasi, FAST and CenSurE. In matching stage to measure the similarity between image patches around the keypoints, SURF uses Approximate Nearest Neighbor search where as Shi-Tomasi, FAST and CenSurE uses normalized cross-correlation. They showed that SURF keypoints matched with Approximate Nearest Neighbor search, or Shi-Tomasi features matched via ZNCC are suitable for real-time processing applications. On the other hand, the FAST and CenSurE detectors were shown to be inaccurate for the image sequences with high levels of motion blur.

Piliang GONG et al., [8] have developed technique to estimate motion of underwater vehicle from stereo sequences. They used SURF algorithm to extract and track the features of stereo images. The experimental results showed that SURF is more robust feature detector and descriptor, because for images with large translation and rotation it still can extract the features and has better real-time performance. Even though underwater images usually have very noisy and the contrast ratio is low, SURF can still extract enough number of features under most of the conditions. Josep Aulinas et al. [2] have adapted SURF feature extraction and matching method for estimating the Region of Interest (RoI) and finding the corresponding point for landmark detection in sea floor and this helps in solving the Simultaneous Localization and Mapping (SLAM) problem. Prabhakar and Praveen Kumar [24] have quantitatively compared KLT, SIFT, SURF feature detectors and descriptors for underwater images. The authors have claimed that, the SURF detects more number of interest points compared to KLT and SIFT. Nuno Gracias and Shahriar Negahdaripour [9] used texture patches to provide more information on the interest points. They extracted texture patches from underwater video sequences at different altitudes to create mosaic of coral reefs in an underwater environment. Maricor Soriano et al., [27] have adapted Local Binary Pattern to extract texture information from underwater video images. They extracted color, texture and color-texture descriptors from coral images and measure recognition rates using each feature. They showed that texture was found to be more discriminating than using color or color-texture descriptors. Rafael Garcia et al. [7] used texture operators to improve image matching for underwater image sequences. They used texture energy filters as texture operators for extracting the interest points in the subsampled window. The authors have compared the texture vectors with the texture vector of the interest point of subsequent image by computing the point-to-point Euclidean distance.

3. COMPREHENSIVE COLOR IMAGE NORMALIZATION

The raw color recorded by the camera is not a reliable for color invariant feature detection and matching. In order to extract color invariant features, chromaticity normalization is commonly employed to removal of lighting geometry and gray world normalization is employed to remove illuminant color effect. The comprehensive color image normalization combines both chromaticity normalization and gray world assumption to remove both lighting geometry and illuminant color effects [6]. If we view a surface under a given light and then double the intensity of the light we expect a doubling in the recorded RGB values:

$$\begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} = \begin{pmatrix} \rho_i R_i \\ \rho_i G_i \\ \rho_i B_i \end{pmatrix} \tag{1}$$

where ρ_i is a simple scalar. The scalar has a subscript i indicating that all pixels can have their own individual brightness factors.

A change in light intensity modeled by Eq. 1 can be cancelled out by applying chromaticity normalization:

$$\begin{aligned} r_i &= \frac{\rho_i R_i}{\rho_i R_i + \rho_i G_i + \rho_i B_i}, \\ g_i &= \frac{\rho_i G_i}{\rho_i R_i + \rho_i G_i + \rho_i B_i}, \\ b_i &= \frac{\rho_i B_i}{\rho_i R_i + \rho_i G_i + \rho_i B_i} \end{aligned} \tag{2}$$

where the new co-ordinates r, g and b are independent of ρ_i.

The change in lighting color (assuming lighting geometry is fixed) is reasonably modeled of a change in illumination

$$\begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} \rightarrow \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \cdot \begin{pmatrix} \rho_i R_i \\ \rho_i G_i \\ \rho_i B_i \end{pmatrix} \tag{3}$$

To cancel out dependence on illumination color, as modeled by Eq. 3. First, estimate mean of each channel independently i.e. the mean of red pixel value for an image is:

$$\mu(R) = \frac{\sum_{i=1}^{N} R_i}{N} \quad (4)$$

where N is the number of pixels in the image: Under a change in illuminant color (Eq. 4) the mean becomes:

$$\mu(R)^{'} = \frac{\sum_{i=1}^{N} aR_i}{N} = a\mu(R) \quad (5)$$

That is, the mean changes by the same scale factor a, thus to cancel the effect of light color on RGBs we can apply the following transformation:

$$R_i^{'} = \frac{aR_i}{a\mu(R)}, G_i^{'} = \frac{aG_i}{a\mu(G)}, B_i^{'} = \frac{aB_i}{a\mu(B)} \quad (6)$$

Eq. 6 is commonly referred to as a gray-world normalization. Eq. 2 and 6 remove the effects of lighting geometry and illumination color respectively but neither, by itself, suffices to remove the effect of both factors. Therefore, Finlayson et al. [5] defined a third normalization called comprehensive normalization, which can remove both dependencies. It is defined as:

1. $I_0 = I$ initialization
2. $I_{i+1} = G(C(I_i))$ iteration step
3. $I_{i+1} = I_i$ termination condition

where $G()$ and $C()$ function stands for gray-word and chromaticity normalization. That is, chromaticity and gray-world normalization are applied successively and repeatedly to an image until the resulting image converges to a fixed point.

4. LBP-SURF FRAMEWORK

In the LBP-SURF framework, first the interest points are detected using Hessian based feature detection technique used in SURF. In the second stage, for each interest point, a texture descriptor is built using CS-LBP technique to distinctively describe the local region around the interest point. The final stage is matching the descriptors to decide if this point belongs to the object of interest or not. The matched points can be used for further processing such as performing underwater object recognition or scene reconstruction. We employ Nearest Neighbor Distance Ratio (NNDR) to measure the similarity between the feature descriptors.

4.1 SURF: Detection of Interest Points

SURF interest point detection uses a very basic Hessian-matrix based approximation [3]. This lends itself to the use of integral images, which reduces the computation time drastically. In the case of SURF the Hessian matrix is approximated roughly using simple box filters of size 9×9 in constant time. SURF techniques uses integral images way of representation which allow for fast computation of box type convolution filters. The integral image $I_{\Sigma}(\mathbf{x})$ at a location $\mathbf{x} = (x, y)^T$ represents the sum of all pixels in the input image I within a rectangular region formed by the origin and $\mathbf{x}$.

$$I_{\Sigma}(\mathbf{x}) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j). \quad (7)$$

Given a point $\mathbf{x} = (x, y)$ in an image I, the Hessian matrix $H(\mathbf{x}, \sigma)$ in $\mathbf{x}$ at scale σ is defined as follows

$$H(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix} \quad (8)$$

where $L_{xx}(\mathbf{x}, \sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$ with the image I in point $\mathbf{x}$, and similarly for $L_{xy}(\mathbf{x}, \sigma)$ and $L_{yy}(\mathbf{x}, \sigma)$.

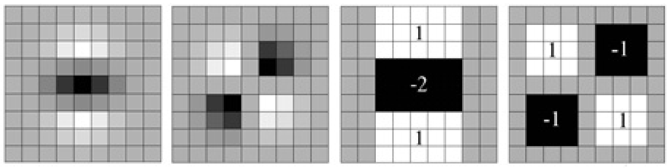

Figure 1: The first two images represent the discretised Gaussian second order partial derivative in y (Lyy) and xy direction (L_{xy}) respectively. The last two images represent the approximation for the second order Gaussian partial derivative in y (D_{yy}) and xy direction (D_{xy}). The gray regions assumed to be zero.

The box filters in Fig. 1 are approximations of a Gaussian with $\sigma = 1.2$ and represent the lowest scale (i.e. highest spatial resolution) for computing the blob response maps. The weights applied to the rectangular regions are kept simple for computational efficiency. This yields

$$\det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2. \quad (9)$$

The relative weight w of the filter responses is used to balance the expression for the Hessian's determinant. This is needed for the energy conservation between the Gaussian kernels and the approximated Gaussian kernels,

$$w = \frac{|L_{xy}(1.2)|_F |D_{yy}(9)|_F}{|L_{yy}(1.2)|_F |D_{xy}(9)|_F} = 0.912... \simeq 0.9, \quad (10)$$

where $|x|_F$ is the Frobenius norm.

A scale-space representation is needed to detect features in various sizes. As with Difference-of-Gaussian (DOG), the scale-space is divided into octaves. An octave is a set of images filtered with increasing kernel size. In each octave, the filter is scaled up by the factor of 2. The fast filter response calculation allows SURF to scale up the filters instead of downscaling the image. The filter size doubles on each octave i.e. on the first octave the filter sizes are increased by 6 in each step, in the second octave by 12 in each step and finally in third octave by 24 in each step (Fig. 2).

Finally local features are selected as local maxima in $3 \times 3 \times 3$ neighborhood in the scale space. A fast method for non-maximum suppression proposed by Neubeck and Van Gool is used to locate these extrema points [21].

4.2 CS-LBP Descriptor

Center-Symmetric Local Binary Patterns is one of the commonly used texture descriptor [11]. CS-LBP is a modified version of Local Binary patterns in which instead of comparing each pixel with the center pixel, CS-LBP compares center-symmetric pairs of pixels. It will halves the

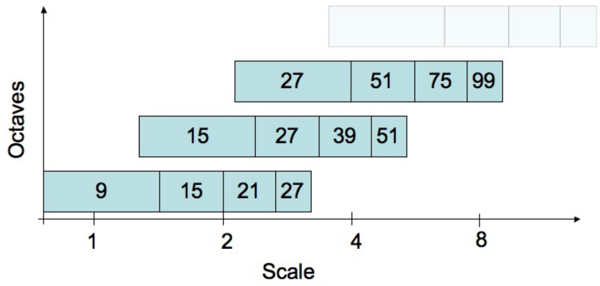

Figure 2: Graphical representation of the filter side lengths for three different octaves. The logarithmic horizontal axis represents the scales. Note that the octaves are overlapping in order to cover all possible scales seamlessly

number of comparisons for the same number of neighbors. The robustness on flat image regions is obtained by thresholding the graylevel difference with a small value T. The mathematical representation of CS-LBP is

$$CS - LBP_{R,N,T}(x, y) = \sum_{i=0}^{(N/2)-1} s(n_i - n_{i+(N/2)})2^i, \quad (11)$$

$$s(x) = \begin{cases} 1 & x > T \\ 0 & \text{otherwise} \end{cases}$$

where n_i and $n_{i+(N/2)}$ correspond to the grayvalues of center-symmetric pairs of pixels of N equally spaced pixels on a circle of radius R. It should be noticed that the CS-LBP is closely related to gradient operator, because like some gradient operators it considers graylevel differences between pairs of opposite pixels in a neighborhood. We employed uniform weighting scheme for the region around detected interest points.

In order to incorporate spatial information into the descriptor, the input region is divided into cells with cartesian grid. We use 4×4 (16 cells) cartesian grid. For each cell a CS-LBP histogram is built. The resulting descriptor is a 3D histogram of CS-LBP feature locations and values. The number of different feature values ($2^{N/2}$) depends on the neighborhood size (N) of the chosen CS-LBP operator. The final descriptor is built by concatenating the feature histograms computed for the cells to form $M \times M \times 2^{N/2}$ - dimensional vector, where M and N are the grid size and CS-LBP neighborhood size, respectively.

4.3 Descriptor Matching

After extracting stable, color invariant and distinctive texture features, similarity between feature descriptors of the left image and the right image is computed. The most commonly used approaches are Nearest Neighbor (NN) and Nearest Neighbor Distance Ratio (NNDR). Lowe [16] proposed to obtain matches through a NN approach, by using the Euclidean distance between a left and right image descriptor by considering a suitable threshold of T:

$$NN = \sqrt{\sum_{i=1}^{128}(d_R(i) - d_L(i))^2} < T \quad (12)$$

where d_R and d_L are feature descriptors of Right and Left images. This metric measure the descriptors distance in their dimensional space, and it is thresholded to avoid matching between descriptors with low similarity.

In the last few years, the NNDR (also called ambiguity distance) has gained more popularity because of its enhanced performance. The ambiguity distance compares the distance between the closest and the second closest descriptor. The distance ratio is thresholded to avoid matching when the nearest neighbors are very similar. Considering $d_{L,1}$ the closest descriptor and $d_{L,2}$ the second closest descriptor, we can define the ambiguity distance as:

$$NNDR = \frac{\sqrt{\sum_{i=1}^{128}(d_R(i) - d_{L,1}(i))^2}}{\sqrt{\sum_{i=1}^{128}(d_R(i) - d_{L,2}(i))^2}} < T \quad (13)$$

We compute the similarity between the descriptors using NNDR based approach. The threshold T is set empirically.

5. EXPERIMENTAL RESULTS

We have performed several experiments in order to validate our approach for underwater video images captured in a small water body in which the object was kept at a depth of 5 feet from the surface level of the water. The image capturing setup consists of a waterproof camera which is Canon-D10 and objects (pipes and valves). The camera is moved around an object at a distance [1m, 2m], near the corner of the water body to capture the color video sequence. We have captured three video sequence in three different water conditions with turbidity levels. From each video sequence, we select two test frames which are not consecutive (frame 1 and frame 20) for experimentation (Fig. 3). The captured video images are suffered from non-uniform lighting, low contrast, blurring and diminished colors. The variations in corresponding color values are very high due to various factors such as, attenuation and scattering of light in underwater environment. The attenuation and scattering of light cause variations in the measured color of each scene point at its projections on the two images. In order to render the color values, we apply comprehensive color image normalization for raw RGB color images. Fig. 4 shows the detected interest points using fast Hessian method for color corrected images.

5.1 Evaluation of Feature Detector

The performance of our approach for feature detection is evaluated based on repeatability criteria and compared with other popular feature detection methods such as Harris and Difference-of-Gaussian (DOG) used in SIFT. The repeatability measurement is computed as a ratio between the number of point-to-point correspondences that can be established for detected points, and the mean number of points detected between two images:

$$r_{1,2} = \frac{C(I_1, I_2)}{mean(fp_1, fp_2)} \quad (14)$$

where $C(I_1, I_2)$ denotes the number of corresponding matches

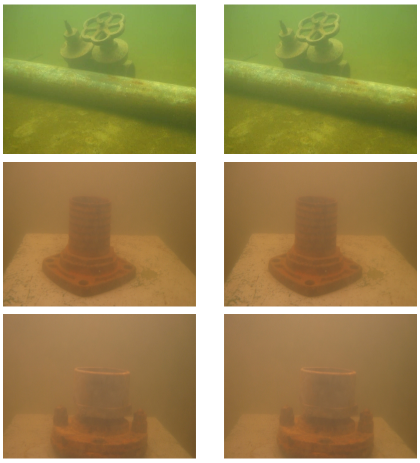

Figure 3: Underwater video test images: first row - dataset1; second row - dataset2; third row - dataset3;

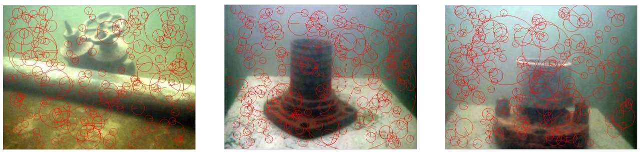

Figure 4: Feature detected using our approach for color corrected first frame belongs to three datasets

between first and second frame considered, $mean(fp_1, fp_2)$ represents the mean value of the number of the feature points detected in the first and second frame.

5.1.1 *Normalized V/S Un-normalized Images*

The experiments are conducted to evaluate the performance of repeatability criteria for normalized v/s un-normalized images and it is observed that the repeatability rate for normalized images of dataset1 is 0.5723, whereas for un-normalized images is 0.4354. Similarly, the improved performance is achieved for dataset2 and dataset3. From this, it is found that normalization increases the performance of repeatability criteria.

5.1.2 *Blur Changes*

Normally, the underwater images suffer from blurring due to attenuation and scattering effects on illuminated light in an underwater environment. Many existing feature detection techniques, which are used for general-purpose applications may not be suitable for detecting features in underwater environment. We present experimental results carried out using dataset1 to evaluate effects of our approach by considering images with variation in blurring. The radius of blur is varied from 5 to 40 in length using MATLAB. Fig. 5 shows the first image and second image of dataset1 with radius of the blur 15. The repeatability score of each feature detection technique is calculated by using Eq. (14). The Fig. 6 shows the repeatability score of our approach, Harris and DOG against varying radius of blur. The good repeatability score of our approach is 0.55 when the radius of blur is 5. Similarly, the repeatability score of Harris is 0.48 and DOG is 0.42 for radius of blur 5. Our approach performs good compared to Harris and DOG when the radius of blur increases (radius of blur 5).

Figure 5: Dataset1 : (a) Left frame, (b) Right frame with Blur radius 15

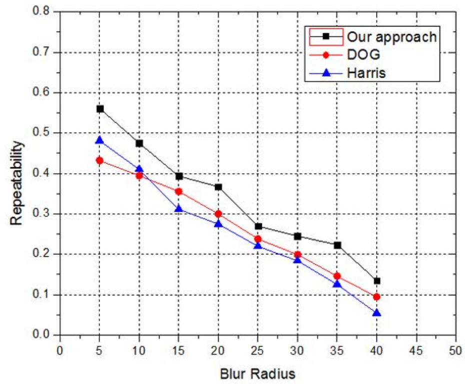

Figure 6: Variation of Blur Parameters versus Repeatability rate

5.1.3 *Illumination Changes*

In the underwater environment, the natural illumination typically varies spatially and temporally. The incident light on the water surface is refracted into the water by waves in a spatio-temporal varying manner. This effect leads to variation in illuminated light in the underwater. We have conducted experiments using Dataset1 with varying illumination. To show the effects of our approach for illumination variation, the illumination of first frame (LF) is not varied and the illumination is varied for second frame (RF) using Photoshop software. The Fig. 7 shows the pair of images with illumination variation of dataset1 used in the experiments. From Table 1 we can see that our approach is best compared to Harris and DOG based on repeatability score. The number of matched feature points is 981 using our approach compared to matched feature points 412 and 65 obtained using Harris, DOG respectively. This due to the fact that our approach detect most illumination invariant features from the video sequence where the disparity between the frames is very high. The underwater video sequence has such property that the lighting variations between the frames are very high due to optical properties of the water. Therefore, the our approach is the most suitable feature detector for underwater video sequence.

5.2 Evaluation of Feature Descriptor

The performance of our approach for feature description is compared with popular descriptors such as SIFT [16] and

Figure 7: Dataset1 : (a) Left frame, (b) Right frame with illumination varied using Photoshop

Table 1: Comparison of Feature Detector with Illumination Variation for dataset1

Method	LF	RF	Matches	Repeatability
Harris	924	945	412	0.4409
DOG	185	154	65	0.3835
Our approach	1909	1974	981	0.5053

SURF [3] based on the number of correct and false matches between a pair of video images. We have used dataset1, dataset2 and dataset3 for experimentation purpose. The descriptors are matched using NNDR approach i.e. the descriptors having a value below a threshold is selected as the matched point. The threshold is empirically set. The descriptor of the first image is compared with each descriptor of the second image to count the number of correct matches as well as the number of false matches. The results are presented with *recall* versus $1 - precision$ graph. Recall is the number of correctly matched points with respect to the number of corresponding points between the two images of the same scene. It is given by:

$$recall = \frac{\#\text{correct matches}}{\#correspondences} \tag{15}$$

The number of false matches relative to the total number of matches is represented as $1 - precision$

$$1 - precision = \frac{\#\text{false matches}}{\#\text{correct matches} + \#\text{false matches}} \tag{16}$$

The experiments are conducted for both normalized v/s un-normalized images, and it is observed that recall rate for normalized images are high compared to un-normalized images for all three datasets. Fig. 8 shows the performance of proposed LBP-SURF, SIFT and SURF for normalized images of dataset1, dataset2 and dataset3 respectively. The axis of all the graphs is scaled between 0 and 1. The X-axis of the graph plot 1- precision and the Y-axis of the graph plots recall. The graphs show that the our approach gives highest recall rate compared to SIFT and SURF descriptors for underwater images. This shows that texture information used in our approach remains constant in frames of same scene compared to geometrical information used by SIFT and SURF.

6. CONCLUSIONS

In this paper, we present an approach to extract color-invariant and texture based features for underwater video sequence. Since there is no benchmark database is available for underwater video sequence, we have conducted experiments using underwater video images captured in different water conditions with turbidity characteristics. In underwater images, it is very difficult to extract feature points due to poor quality of images. Moreover, the raw RGB images are not suitable for detecting color-invariant feature points; therefore, in order to obtain color-invariant features, we applied comprehensive color image normalization to normalize the raw RGB values. The interest points are detected using fast Hessian feature detection method. The experimental results show that, for detecting interest points our approach outperforms compared to Harris and DOG detectors. Further, proposed LBP-SURF descriptors outperforms other geometric based feature descriptors. The LBP-SURF is robust to illumination, blur variation and computationally efficient than SIFT and SURF descriptor. The proposed LBP-SURF can be applied to various underwater computer vision problems such as object recognition and scene reconstruction.

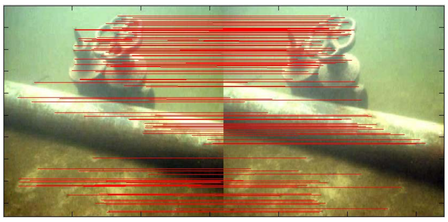

Figure 9: Feature points matching using proposed LBP-SURF method for normalized images of dataset1

7. ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. This research was supported by Naval Research Board (Grant No.158/SC/2008-09), DRDO, New Delhi, India.

8. REFERENCES

[1] A. E. Abdel-Hakim and A. A. Farag. CSIFT: A SIFT descriptor with color invariant characteristics. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 1978–1983, 2006.

[2] J. Aulinas, M. Carreras, X. Llado, J. Salvi, R. Garcia, R. Prados, and Y. Petillot. Feature extraction for underwater visual SLAM. In *OCEANS, 2011 IEEE - Spain*, pages 1–7, June 2011.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[4] V. Brandou, A. Allais, M. Perrier, E. Malis, P. Rives, J. Sarrazin, and P. Sarradin. 3D reconstruction of natural underwater scenes using the stereovision system IRIS. In *OCEANS 2007 - Europe*, pages 1–6, June 2007.

[5] G. Finlayson, B. Schiele, and J. L. Crowley. Comprehensive colour image normalization. In *ECCV '98*, volume 1406, pages 475–490. 1998.

[6] G. Finlayson and R. Xu. Illuminant and gamma comprehensive normalisation in logRGB space. *Pattern Recognition Letters*, 24(11):1679–1690, 2003.

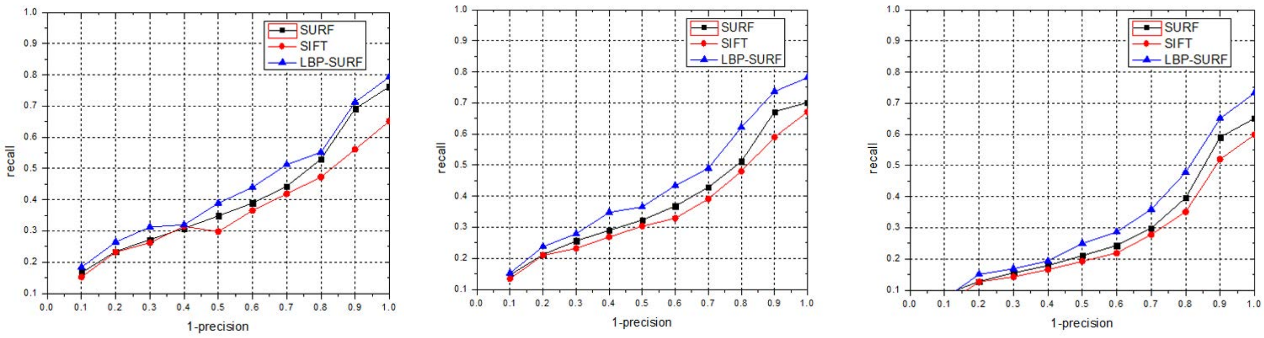

Figure 8: Comparison of proposed LBP-SURF with SIFT and SURF for dataset1 (left), dataset2 (center) and dataset3 (right) respectively

[7] R. Garcia, C. Xevi, and J. Battle. Detection of matchings in a sequence of underwater images through texture analysis. In *International Conference on Image Processing, ICIP '2001*, volume 1, pages 361–364, 2001.

[8] P. Gong, Q. Zhang, and A. Zhang. Stereo vision based motion estimation for underwater vehicles. In *Second International Conference on Intelligent Computation Technology and Automation, ICICTA '09*, volume 3, pages 745–749, October 2009.

[9] X. Gracias and S. Negahdaripour. Underwater mosaic creation using video sequences from different altitudes. In *Proceedings of MTS/IEEE OCEANS, 2005*, volume 2, pages 1295–1300, September 2005.

[10] C. Harris and M. Stephens. A Combined Corner and Edge Detection. In *Proceedings of The Fourth Alvey Vision Conference*, pages 147–151, 1988.

[11] M. Heikkila, M. Pietikainne, and C. Schmid. Description of interest regions with center-symmetric local binary patterns. In *Computer Vision, Graphics and Image Processing, ICVGIP 2006*, volume 4338, pages 58–69, 2006.

[12] Y. S. Heo, K. M. Lee, and S. U. Lee. Robust stereo matching using adaptive normalized cross-correlation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):807–822, April 2011.

[13] A. Hogue and M. Jenkin. Development of an underwater vision sensor for 3D reef mapping. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006*, pages 5351–5356, October 2006.

[14] P. Jasiobedzki, S. Se, M. Bondy, and R. Jakola. Underwater 3D mapping and pose estimation for ROV operations. In *OCEANS 2008*, pages 1–6, September 2008.

[15] T. Lindeberg. Scale-space theory: a basic tool for analyzing structures at different scales. *Journal of Applied Statistics*, 21(1–2):225–270, 1994.

[16] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[17] M. S. A. Marcos, M. Soriano, and C. Saloma. Low-level color and textre feature extraction of coral reef components. *Science Diliman*, 15(1):45–50, 2003.

[18] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *Proceedings of Eighth IEEE International Conference on Computer Vision, ICCV '2001*, volume 1, pages 525–531, 2001.

[19] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60:63–86, 2004.

[20] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65:43–72, 2005.

[21] A. Neubeck and L. Van Gool. Efficient non-maximum suppression. In *18th International Conference on Pattern Recognition,ICPR '2006*, volume 3, pages 850–855, 2006.

[22] K. Oliver, W. Hou, and S. Wang. Image Feature Detection and Matching in Underwater Conditions. In *Proceedings of SPIE. Ocean Sensing and Monitoring II, Orlando, FL*, volume 7678, 2010.

[23] K. Plakas, E. Trucco, and A. Fusiello. Uncalibrated vision for 3-D underwater applications. In *OCEANS '98*, volume 1, pages 272–276, September 1998.

[24] C. J. Prabhakar and P. U. P. Kumar. Quantitative Evaluation of Feature Trackers for Underwater Video Sequences. In *Proceedings of The International Conference on Signal, Image and Video Processing*, pages 186–191, January 2012.

[25] A. Sedlazeck, K. Koser, and R. Koch. 3D reconstruction based on underwater video from ROV kiel 6000 considering underwater imaging conditions. In *OCEANS 2009 - EUROPE*, pages 1–10, May 2009.

[26] F. Shkurti, I. Rekleitis, and G. Dudek. Feature tracking evaluation for pose estimation in underwater environments. In *Canadian Conference on Computer and Robot Vision (CRV' 11)*, pages 160–167, May 2011.

[27] M. Soriano, S. Marcos, C. Saloma, M. Quibilan, and P. Alino. Image classification of coral reef components from underwater color video. In *MTS/IEEE Conference and Exhibition OCEANS, 2001*, volume 2, pages 1008–1013, 2001.

[28] E. Trucco, Y. Petillot, I. Ruiz, K. Plakas, and D. Lane. Feature tracking in video and sonar subsea sequences with applications. *Computer Vision and Image Understanding*, 79(1):92–122, 2000.