Chapter 6

Detection and Matching of Features in Underwater Images

Chapter 6
Detection and Matching of Features in Underwater Images

In this chapter and its subsequent chapter, we focus on development of a novel approaches to solve our second research problem i.e extraction of features from underwater images for image matching and object recognition. An approach to match underwater images based on detection and matching of color and photometric invariant feature points is presented in this chapter. Feature points (interest points) are detected in the rectified left image using SIFT based feature detection method. Further, for the detected interest point in the left image, we find the corresponding matching point in the right image using window based correlation measure. The performance of the proposed approach is compared with SURF and KLT based feature detectors based on repeatability criteria and experimentally showed that our approach has very high repeatability rate.

6.1 Introduction

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 wide baseline matching, image stitching, image registration, image retrieval, robot localization, video data mining, building panoramas, texture recognition, object detection and object recognition. The significant point of feature point detection is stability and distinctiveness, i.e. the estimated feature point in the successive images should be stable with respect to geometric and photometric variations and it should be distinctive, which means that the extracted features should have minimum information to distinguish between the object which they describe and other objects [2]. Unmanned underwater vehicles (UUVs) are ideal tools to accomplish periodical inspections of jacket structures of offshore platforms [42]. A stereo vision based method is used to navigate
the vehicle to locate itself with the aid of environmental reference information. With the aid of the location information, motion of the vehicle can be estimated and based on the motion of the vehicle the pilot can navigate the vehicle in a more efficient and safer way. Implementation details include several steps such as feature extraction and tracking, stereo correspondence and robust motion estimation. In this chapter, we focus on performing one important step such as feature detection and matching in underwater stereo images in order to estimate the motion of the vehicle. Suppose the features have been successfully detected in the left image; the next step is to find the corresponding features in the right image.

The standard feature detectors and descriptors viz. SIFT, SURF and KLT technique address invariance in the spatial domain. They are not invariant when the considered image undergoes a destructive intensity transformation, where the image intensity values change substantially, inconsistently and irreversibly. Such transformations often significantly increase the complexity in discerning any underlying features and structures in the image, which varies the same feature in sequence of images. A typical example is intensity transformation introduced by underwater imaging. Light behaves differently in underwater, when imaging underwater, even an image of the same object can be drastically different due to varying water conditions [93]. The added complexities of impure water introduce issues such as turbulence, air bubbles, particles (such as sediments), and organic matter that can absorb and scatter light, which can result in a very blurry and noisy image. Since available feature descriptors are not invariant under such intensity transformations, matching the features detected from an underwater image and a clean out-of-water image, or the features detected from two underwater images taken in different underwater conditions, is a very challenging problem. Moreover, underwater images show large variations in color rather than geometrical variations.

Most of the existing approaches use gray geometric-based features. Nearly all geometrical invariant approaches avoid dealing with color images due to color constancy problems. 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 [2]. Color invariance features provide high
discriminative information cues, as color varies considerably with change in camera viewpoint, object pose and illumination. Underwater images usually suffer from radiometric variations using this 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 [54]. Finlayson et al., [34] developed a method called comprehensive color image normalization method, which removes lighting geometry effects and illuminant color effects both iteratively and non-iteratively.

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 [43]. 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 have been proposed in the literature. The most widely used detectors can be classified into two categories such as Harris-based and Hessian-based detectors [49] [86] [90] [91] [92]. The Harris corner detector [49] 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 [81] proposed the concept of automatic scale selection based on Hessian matrix as well as Laplacian. Mikolajczyk and Schmid [91] 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 [86] proposed to approximate the Laplacian of Gaussian (LoG) by a Difference of Gaussian (DoG) filter. SIFT proposed by David Lowe [86] is one of the widely used feature detector and descriptor, because
of its robustness, distinctiveness and efficient computation time in feature extraction and matching. SIFT uses Harris-based feature detector because it is more stable and repeatable than Hessian-based counterparts which are usually used in SURF [8]. It is also observed that approximations like the DoG used in SIFT can bring speed at a low cost in terms of high accuracy.

In this chapter, we introduce an approach to detect color invariant features from underwater left image and to find the corresponding feature points in the right image. We apply comprehensive color image normalization (Chapter 3) to normalize raw RGB color images in order to detect color and photometric invariant features. Color-invariant features are detected from the normalized left image using SIFT feature detection technique i.e. by using Difference-of-Gaussian (DoG) approach. We estimate the candidate matches in the right image for a given interest point in the left image by using window-based correlation measure. For a given, interest point in the left image and the possible candidate matches in the right image, we employ Euclidean similarity measure to find whether there exist true match or false match based on threshold.

The remainder of the chapter is organized as follows: section 6.2, we present brief survey on feature detection technique applied for underwater images. The image matching technique is presented in section 6.3. The experimental results on underwater images are presented in section 6.4. Finally, section 6.5 draws the chapter summary.

6.2 Related Work

The researchers have developed feature detectors and descriptors such as SIFT, SURF, GLOH, DAISY, etc., to detect and extract the geometric-based features from out-of-water 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 detectors and descriptors meant for underwater environment. Since there is no standard detector and descriptor meant for underwater environment, the researchers have been using standard
feature detectors and descriptors for underwater applications.

Trucco et al. [134] have adapted Kanade-Lucas-Tomasi (KLT) based feature detector for feature extraction and motion estimation in underwater environment. KLT [87] is strongly based on Harris corner detector, where features are extracted using minimum eigenvalue of each 2 × 2 gradient matrix. Similarly, Plakas et al. [109] have adapted KLT based feature detector for extraction of features to 3D shape reconstruction from uncalibrated underwater video sequences. Anne Sedlazeck et al. [118] 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 [55] also adapted KLT for 3D shape reconstruction of coral reefs using stereo image sequences. Brandou et al. [17] 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., [66] have used SIFT technique for feature extraction.

6.3 Image Matching

The image matching can be carried out by finding corresponding feature points (keypoints) in both the images of interest. Begin with rectification of a given pair of normalized images to reduce the searching region along the horizontal axis, the stable and distinctive features of left rectified image are detected using DoG based procedure of SIFT technique. For a given interest point of left image, a set of possible matchings is established in the right image using window-based correlation measure. Finally, the true match is computed for a given interest point in the left image using point-point Euclidean distance measured between given point and its candidate matches based on threshold. The true match is selected as the one with minimum distance which lies within the threshold.
6.3.1 Uncalibrated Stereo Image Rectification

Given a pair of stereo images, rectification determines a transformation of each image plane such that pairs of conjugate epipolar lines become collinear and parallel to one of the image axes. The important advantage of rectification is that computing feature point correspondence is reduced to a 1-D search problem along the horizontal raster lines of the rectified images. We use Quasi-Euclidean epipolar rectification method for uncalibrated images proposed by Andrea Fusiello et al. [40].

6.3.2 Detection of Interest Points

Interest points (keypoints) are detected as the steps followed by SIFT feature detection technique proposed by David Lowe [86]. SIFT features are all natural features of images. They are favorably invariant to image translation, scaling, rotation, illumination, viewpoint, noise etc. Good speciality, rich in information, suitable for fast and exact matching in a mass of feature database. Relatively fast speed. The extraction of SIFT features from images can be done by applying sequentially steps such as scale-space extrema detection and keypoint localization.

**Scale-space extrema detection:** Interest points or keypoints are detected in this step. This is the first stage of computation searches over all scales and image locations. It is implemented efficiently by using a Difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation. A Gaussian pyramid is constructed from the input image by repeated smoothing and subsampling, and a Difference-of-Gaussian pyramid is computed from the differences between the adjacent levels in the Gaussian pyramid. Then, interest points are obtained from the points at which the Difference-of-Gaussian values assume extrema with respect to both the spatial coordinates in the image domain and the scale level in the pyramid.

Given a Gaussian-blurred image described as the formula

\[ L(x, y, \sigma) = G(x, y, \sigma) \ast I(x, y), \]  

(6.1)
where $\star$ is the convolution operation in $x$ and $y$, and

$$
G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}.
$$

(6.2)

To efficiently detect stable keypoint locations in scale space extrema in the Difference-of-Gaussian function convolved with the image, $D(x, y, \sigma)$, which can be computed from the difference of two nearby scales separated by a constant multiplicative factor $k$:

$$
D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) \star I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma).
$$

(6.3)

which is just be different from the Gaussian-blurred images at scales $\sigma$ and $k\sigma$.

The first step toward the detection of interest points is the convolution of the image with Gaussian filters at different scales, and the generation of Difference-of-Gaussian images from the difference of adjacent blurred images. The rotated images are grouped by octave (an octave corresponds to doubling the value of $\sigma$), and the value of $k$ is selected so that we can obtain a fixed number of blurred images per octave. This also ensures that we obtain the same figure of Difference-of-Gaussian images per octave (Figure 6.1)

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 (Figure 6.2). For each candidate keypoint: 1) Interpolation of nearby data is used to accurately determine its position; 2) Keypoints with low contrast are removed; 3) Responses along edges are eliminated; 4) The keypoint is assigned an orientation.

To determine the keypoint orientation, a gradient orientation histogram is computed in the neighborhood of the keypoint (using the Gaussian image at the closest scale to the keypoint’s scale). The contribution of each neighboring pixel is weighted by the gradient magnitude and a Gaussian window with a $\sigma$ that is 1.5 times the scale of the keypoint. Peaks in the histogram correspond to dominant orientations. A separate keypoint is rerated for the direction corresponding to the histogram maximum, and any other direction within 80% of the maximum value. All the properties of the keypoint are measured relative to the
Figure 6.1: The blurred images at different scales, and the computation of the Difference-of-Gaussian images

Figure 6.2: Local extrema detection, the pixel marked $X$ is compared against its 26 neighbors in a $3 \times 3 \times 3$ neighborhood that spans adjacent DoG images
keypoint orientation, this provides invariance to rotation.

**Keypoint localization:** scale-space extrema detection produces too many keypoint candidates, some of which are unstable. A detailed fit to the nearby data for accurate location, scale, and ratio of principal curvatures is done. This information allows points to be rejected that have low contrast (and are therefore sensitive to noise) or are poorly localized along an edge.

At each candidate location, a detailed model is fit to determine scale and location. Keypoints are selected on basis of measures of their stability. The steps involved in selecting the suitable feature point (keypoint) is as follows:

1. Input a gray scale image.

2. Use a variable-scale Gaussian kernel $G(x, y, \sigma)$ to create scale space $L(x, y, \sigma)$.

3. Calculate Difference-of-Gaussian function as an approximate to the normalized Laplacian is invariant to the scale change.

4. Find the maxima or minima of Difference-of-Gaussian function value by comparing one of the pixels to its above, current and below scales in $3 \times 3$ regions.

5. Accurate the keypoints locations by discarding points below a predetermined value.

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{X}. \quad (6.4)$$

where $\hat{X}$ is calculated by setting the derivative $D(x, y, \sigma)$ to zero.

6. The extremas of Difference-of-Gaussian have large principal curvatures along edges, it can be reduced by checking

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r + 1)^2}{r}. \quad (6.5)$$

Evidently, $H$ is a $2 \times 2$ Hessian matrix, $r$ is the ratio between the largest magnitude and the smallest one.
6.3.3 Window-Based Correlation Measure

Correspondence between image pairs is a classical problem in the field of computer vision. Usually correlation measures are used for correspondence matching. We apply window-based correlation measure on the selected interest point in the left image to estimate the candidate matches in right image. Given an interest point \( p_l \) in the left image, with the coordinates \((x_l, y_l)\), the correlation window is centered at this point. Then, a search window is defined on the right image, also centered at \((x_l, y_l)\), and the correlation operation is performed. A correlation score is computed in the following way:

\[
corr(p_l, p_r) = \frac{\sum_{i=-w}^{w} \sum_{j=-w}^{w} (I_L(x_l+i, y_l+j) - \overline{I_L(x_l,y_l)}) \cdot (I_R(x_r+i, y_r+j) - \overline{I_R(x_r,y_r)})}{w^2 \sqrt{\sigma^2(I_L) \cdot \sigma^2(I_R)}}.
\]  

(6.6)

where \( w \) is the size of the window considered; \( I_L \) and \( I_R \) are the left and right images, respectively; \( \sigma^2(I) \) is the variance of the image computed in the correlation window; \( \overline{I(x,y)} \) is the mean of the correlation window in the image. The mathematical representation of \( \sigma^2(I) \) and \( \overline{I(x,y)} \) is given below:

\[
\sigma^2(I) = \frac{\sum_{i=-w}^{w} \sum_{j=-w}^{w} I(x+i, y+j)^2}{w^2} - \overline{I(x,y)}^2,
\]  

(6.7)

and

\[
\overline{I(x,y)} = \frac{\sum_{i=-w}^{w} \sum_{j=-w}^{w} I(x+i, y+j)}{w^2}.
\]  

(6.8)

We define a minimum threshold for estimating possible correspondence matching in the right image for a given interest point in the left image.

6.3.4 Similarity Measure

We employ Euclidean similarity measure to find whether there exist true match or false match between the pair of feature points based on suitable threshold. Setting the threshold too high or too low results in false matches, i.e. incorrect matches. The mathematical representation is:

\[
E(p_l, p_r) = \sqrt{(x_l - x_r)^2 + (y_l - y_r)^2}.
\]  

(6.9)
where $E(p_l, p_r)$ is the Euclidean distance between $p_l$ and $p_r$ i.e. left and right image points respectively. $(x_l, y_l)$ and $(x_r, y_r)$ are the coordinate values of interest points $p_l$ and $p_r$ respectively.

### 6.4 Experimental Results

We have conducted experiments using our underwater image database to verify the efficacy of our approach for detection and matching of feature points for underwater images, which are captured with various turbidity levels and different water conditions. The experiments are carried out for three pairs of underwater color stereo images (dataset 1, dataset 2 and dataset 3). In order to normalize the color values which are changed by various factors such as attenuation and scattering of light in the underwater environment, we employed comprehensive color image normalization for raw RGB color images. The normalization is an important step prior to feature detection, because, normalization equalizes the color values for the whole image at once and helps in finding discontinuity regions which is essential for feature detector. Normally, the feature detectors for out-of-water images are compared with other standard feature detectors using benchmark image databases to test there performance according to specific criteria. To our knowledge, there is no underwater benchmark image database exist at present. Therefore, the performance of our approach for feature detection and matching is compared and evaluated with other feature detectors such as SURF and KLT for our underwater color image datasets based on repeatability criteria and recall and $1 - precision$ values.

Uncalibrated stereo rectification technique is applied on the normalized stereo images (Figure 6.3). Interest points are detected using SIFT (DoG) based feature detection technique to extract stable and distinctive features. The Figure 6.4 shows the detected interest points in normalized and rectified left image of dataset 1, dataset 2 and dataset 3. The corresponding candidate feature points are computed using window-based correlation measure; we employ Euclidean similarity measure to find whether there exist true match or false match between the pair of feature points based on threshold value 0.4. The detected feature points in the left image and its corresponding candidate points are shown in Figure 6.5. The result of true match is computed by employing Euclidean distance measure on
dataset1, dataset2 and dataset3 respectively is shown in Figure 6.6.

6.4.1 Performance Evaluation for Feature Detection

The performance of proposed feature detection method i.e. SIFT (DoG) is evaluated quantitatively and compared with two standard methods such as SURF [8] and KLT [122] [131] for dataset1, dataset2 and dataset3 based on the repeatability criteria and processing time. 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 in two images:

\[
R_{ir} = \frac{C(I_l, I_r)}{\text{mean}(P_l, P_r)}.
\]  

(6.10)

where \(C(I_l, I_r)\) denotes the number of corresponding interest points, \(P_l\) and \(P_r\) denotes number of interest points found in left and right images respectively. The processing time and repeatability evaluation is carried out on three feature detection methods, which shows the tendency of the three methods’ time cost and accuracy of detection. The results are influenced by the factors such as the size and quality of the image types, and the parameters of the algorithm. The results were obtained on an Intel Pentium Core i5-2500 at the speed 3.30 GHz and 4 GB of RAM. The algorithms are implemented in MATLAB environment. The results in Table 6.1 clearly show that repeatability of SIFT is relatively high compared to SURF and KLT based feature detector. It is also observed that SIFT feature detection method extract more number of feature points. Even though the processing time of SURF is low compared to SIFT and KLT, the SIFT can efficiently detect more stable, distinctive and repeatable feature points, with increase in the repeatability rate.

6.4.2 Blur Variation

Normally, the underwater images suffer from blurring due to attenuation and scattering effects of illuminated light in an underwater environment. In this section, we present experimental results carried out using three datasets (dataset1, dataset2 and dataset3) to evaluate effects of our approach for feature detection and matching. For each dataset, the
Figure 6.3: Underwater normalized and rectified stereo images: first row - dataset1; second row - dataset2; third row - dataset3

Figure 6.4: Detected interest points in rectified left image of dataset1, dataset2 and dataset3 respectively using SIFT (DoG)

Figure 6.5: Correspondences computed for dataset1, dataset2 and dataset3. The red colored ‘*’ marks correspond to the interest points in the left image, while blue colored ‘*’ marks are their matchings in the right image
Figure 6.6: Matched feature points for dataset1, dataset2 and dataset3 respectively

Table 6.1: Comparison of feature detection methods for comprehensive color normalized images of dataset1, dataset2 and dataset3

<table>
<thead>
<tr>
<th></th>
<th>Methods</th>
<th>SIFT (DoG)</th>
<th>SURF (Fast Hessian)</th>
<th>KLT (Harris)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dataset1</td>
<td>Number of Feature Points Detected</td>
<td>840</td>
<td>755</td>
<td>571</td>
</tr>
<tr>
<td></td>
<td>Number of Matches</td>
<td>513</td>
<td>383</td>
<td>267</td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>0.6</td>
<td>0.5</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>Processing Time (seconds)</td>
<td>11.85</td>
<td>6.4</td>
<td>15.2</td>
</tr>
<tr>
<td>dataset2</td>
<td>Number of Feature Points Detected</td>
<td>620</td>
<td>537</td>
<td>491</td>
</tr>
<tr>
<td></td>
<td>Number of Matches</td>
<td>448</td>
<td>352</td>
<td>235</td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>0.69</td>
<td>0.61</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>Processing Time (seconds)</td>
<td>9.06</td>
<td>4.7</td>
<td>11.9</td>
</tr>
<tr>
<td>dataset3</td>
<td>Number of Feature Points Detected</td>
<td>186</td>
<td>154</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>Number of Matches</td>
<td>122</td>
<td>83</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>0.62</td>
<td>0.51</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>Processing Time (seconds)</td>
<td>6.16</td>
<td>2.6</td>
<td>8.3</td>
</tr>
</tbody>
</table>
right image blur length is varied about 20 using MATLAB and left image is kept unaltered. Figure 6.7 shows the blur variation for dataset1, dataset2 and dataset3 and corresponding repeatability score is illustrated in the Table 6.2.

Figure 6.7: Stereo images with blur variation; first column - left image, second column - right image with blur variation

6.4.3 Illumination Variation

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 spatiotemporal varying manner. This effect leads to variation in illuminated light in the underwater. We have conducted experiments for three datasets (dataset1, dataset2 and dataset3) with varying illumination. To show the effect of our approach for feature detection and matching for illumination variation, the illumination for right image is varied using Adobe Photoshop software and left image is kept unaltered. The Figure 6.8 shows the pair of images with illumination variation of dataset1, dataset2 and dataset3 used in the experiments and corresponding repeatability score is illustrated in the Table 6.3.
Table 6.2: Comparison of feature detection methods for blur variation

<table>
<thead>
<tr>
<th></th>
<th>Methods</th>
<th>SIFT (DoG)</th>
<th>SURF (Fast Hessian)</th>
<th>KLT (Harris)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dataset1</td>
<td>Number of Feature Points Detected</td>
<td>840</td>
<td>755</td>
<td>571</td>
</tr>
<tr>
<td></td>
<td>Number of Matches</td>
<td>435</td>
<td>296</td>
<td>213</td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>0.51</td>
<td>0.39</td>
<td>0.37</td>
</tr>
<tr>
<td>dataset2</td>
<td>Number of Feature Points Detected</td>
<td>620</td>
<td>537</td>
<td>491</td>
</tr>
<tr>
<td></td>
<td>Number of Matches</td>
<td>424</td>
<td>287</td>
<td>214</td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>0.66</td>
<td>0.52</td>
<td>0.43</td>
</tr>
<tr>
<td>dataset3</td>
<td>Number of Feature Points Detected</td>
<td>186</td>
<td>154</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>Number of Matches</td>
<td>93</td>
<td>65</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>0.5</td>
<td>0.42</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Figure 6.8: Stereo images with illumination variation; first column - left image, second column - right image with illumination varied
Table 6.3: Comparison of feature detection methods for illumination variation

<table>
<thead>
<tr>
<th></th>
<th>Methods</th>
<th>SIFT (DoG)</th>
<th>SURF (Fast Hessian)</th>
<th>KLT (Harris)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dataset1</td>
<td>Number of Feature Points Detected</td>
<td>840</td>
<td>755</td>
<td>571</td>
</tr>
<tr>
<td></td>
<td>Number of Matches</td>
<td>471</td>
<td>334</td>
<td>252</td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>0.56</td>
<td>0.44</td>
<td>0.43</td>
</tr>
<tr>
<td>dataset2</td>
<td>Number of Feature Points Detected</td>
<td>620</td>
<td>537</td>
<td>491</td>
</tr>
<tr>
<td></td>
<td>Number of Matches</td>
<td>451</td>
<td>307</td>
<td>193</td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>0.72</td>
<td>0.57</td>
<td>0.4</td>
</tr>
<tr>
<td>dataset3</td>
<td>Number of Feature Points Detected</td>
<td>186</td>
<td>154</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>Number of Matches</td>
<td>103</td>
<td>72</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>0.55</td>
<td>0.46</td>
<td>0.44</td>
</tr>
</tbody>
</table>

6.4.4 Normalization V/S Un-Normalization

In the previous section, we conclude that comprehensive color image normalization boost the performance of feature detection method (repeatability rate). The literature survey reveals that, normalization is not mandatory step prior to detect features in the out-of-water images. In this section, we experimentally verified the importance of normalization prior to detection of feature points. The experiments are conducted to evaluate the performance of repeatability rate for normalized (using comprehensive color image normalization) versus un-normalized images and it is observed that the repeatability rate for normalized images of dataset1 is 0.6, whereas for un-normalized images is 0.45. Similarly, the improved performance is achieved for dataset2 and dataset3. Therefore, the normalization is essential prior to feature detection, otherwise the performance of feature detection decreases.

6.4.5 Performance Evaluation for Feature Matching

The performance of proposed method for feature matching is compared with SURF [8] and KLT [122] [131] based on the number of correct and false matches between a pair of stereo images. In order to evaluate accuracy of the matching procedure, we have used two step
approach: first, the interest points have been detected in the left image, and the window-based correlation measure with Euclidean distance has been applied to the right image. We set threshold value empirically to find the true match i.e. the value below a threshold is selected as the matched point. The threshold value is 0.4. Next, a human operator has marked all the visually incorrect matches, in order to know which correspondences have been incorrectly established. The results are tabulated with \textit{recall} and \textit{precision} values (Table 6.4). 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:

\[
\text{recall} = \frac{\# \text{correct matches}}{\# \text{correspondences}}
\]  

(6.11)

The number of false matches relative to the total number of matches is represented as \textit{precision}:

\[
1 - \text{precision} = \frac{\# \text{false matches}}{\# \text{correct matches} + \# \text{false matches}}
\]  

(6.12)

Table 6.4: \textit{Recall} and \textit{Precision} values for feature matching of dataset1, dataset2 and dataset3

<table>
<thead>
<tr>
<th>Methods</th>
<th>Recall</th>
<th>1 – Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>dataset1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIFT (DoG)</td>
<td>0.73</td>
<td>0.65</td>
</tr>
<tr>
<td>SURF (Fast Hessian)</td>
<td>0.64</td>
<td>0.62</td>
</tr>
<tr>
<td>KLT (Harris)</td>
<td>0.54</td>
<td>0.52</td>
</tr>
<tr>
<td>dataset2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIFT (DoG)</td>
<td>0.75</td>
<td>0.71</td>
</tr>
<tr>
<td>SURF (Fast Hessian)</td>
<td>0.65</td>
<td>0.61</td>
</tr>
<tr>
<td>KLT (Harris)</td>
<td>0.59</td>
<td>0.55</td>
</tr>
<tr>
<td>dataset3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIFT (DoG)</td>
<td>0.71</td>
<td>0.67</td>
</tr>
<tr>
<td>SURF (Fast Hessian)</td>
<td>0.68</td>
<td>0.66</td>
</tr>
<tr>
<td>KLT (Harris)</td>
<td>0.62</td>
<td>0.58</td>
</tr>
</tbody>
</table>

It is observed that recall rate and precision is very high for proposed method compared to SURF and KLT methods for all three datasets. Hence, the proposed method is efficient technique to extract and match feature points for underwater images.
6.5 Chapter Summary

In this chapter, we present a novel approach to detect and match color invariant features in underwater stereo images. Since, there is no benchmark dataset is available for underwater images, we conducted experiments using our underwater image database. In underwater environment, the problem of finding correspondences in stereo images is specific step in order to estimate the motion of an underwater vehicle. 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 in the rectified left image using SIFT feature detection technique. The corresponding point in the right image is established for a given interest point in the left image using window-based correlation measure with Euclidean distance. We evaluated our approach for the purpose of suitability for underwater environments and compared with other feature detection methods such as SURF and KLT.

The experimental results show that proposed method for feature detection and matching is more suitable for real-time application in underwater environment. Since, it obtains more number of feature points and correspondence matches, compared to SURF and KLT based feature detection methods. Even though the processing time of SIFT based feature detection method is high compared to SURF, but the repeatability of SIFT is high compared to SURF. The SIFT method can even efficiently detect more stable and distinctive feature points with illumination and blur changes.