Abstract: This paper presents Image Fusion of two images that are captured based on Stereo vision. A Feature based algorithm coined as SURF (Speeded Up Robust Feature) Algorithm is used for Feature detection and Feature matching of images in MATLAB software. This SURF Algorithm relies its computation in integral images with high robustness and repeatability. The main interest in this approach is its pixels in images that are sorted representing image descriptors and are used in matching between two images to form rectified image.

Index Terms - Image Fusion, Stereo vision, SURF Algorithm, Feature detection, Feature matching, rectified image.

I. INTRODUCTION

Image Fusion is the task used to find out correspondences between two frames in a stereo vision. An Image is defined to be frames describing information in selected sample space. Achieving the analysis of various frames in same sample surface refers to the interpretation of Image Fusion. This is one of the promising research field in some of application like surveillance, military, robotics etc. And extension techniques are generated for creating a composite Image by details of the source Images taken over to retain features of Images recorded for same modality.

Image Fusion objective provides us to enhance Image in same platform that depends over the performance analysis of pixel level and feature level in merging state. The determination of Image fusion is done by pixel level image Fusion that takes up the value of pixel scale in initial stages of analysis. But this technique generates fused Image with reduced contrast in feature of Images. Hence analysis of Image is done by determining the depth of Image that is focused by describing qualitative analysis of Image. Over spectroscopic condition the visualization of frame needs a spatial congruence among platform over aspect of sample information. As in which contrast of data analysis is performed at different levels that is analyzed as low level, mid-level and high level. In the low level the Image is directly concatenated and processed. Whereas Mid-level Image fusion compressing of Images is done at the very initial step and later processed to fusion. And at high level image fusion images are fused by assigning certain pixel to particular class of sample [1].

In this project image fusion at high level is used in fusing images that are captured based on stereo vision. Stereo is one of the techniques used in computer vision to obtain 2D Images from videos sample. The processing of algorithm involved in computation is implemented by data-parallel programs. A three-dimensional vision occurs in human vision as left and right eye captures the world from different perspective view. And this results in a altitude shift in points of the same scene. The same mechanism is used in work where single object is captured by two cameras aligned at interval of time. In stereo matching of Images disparity corresponds to difference between point features over left camera and right camera over baseline as in Figure 1. The state of art in matching stereo Images are done by determining the disparity in Images. The baseline of camera could be differentiated as narrow baseline and broad baseline. The narrow base line has less spatial artifacts hence this tends to be providing a biased behavior over broad baseline due to less disparity whereas broad baseline stereo reserves more pixels that corresponds to high resolution in discriminated depth. But search range of pixel here may cause false matching in disparity as it includes missing information caused by pattern repeated region. This extraction of information in multiple 2D view is generally used in estimating actual distance of object in robot navigation.

![Figure 1: Stereo Vision System](image)

In Image processing this extracted information of images is used as feature description of Image is done by analysis of interest points. Features in images are referred to corners, blobs and T-junctions of Image. One of the important factors considered in interest point is repeatability. Repeatability is reliability in Image interest points under different viewing perspective. These are defined to be descriptors that contains information of the point with its neighboring surrounding. These descriptors must possess robust towards the noise. And matching of these descriptors are necessary for obtaining stereo Image rectified. Here detection of points is done for all the detected descriptors and later matched. There also includes many algorithms for matching the stereos taken by incorporating smoothness which suppress the estimation of disparity noise. In the calculation of disparity there is comparison between depth information of captured left and right Images of scene. The challenge here is generation of accurate information of depth in scene as pixel contains only color and spatial information that represents low level Image features. Hence there is need of developing efficient method for comparing the two Images of same scene with reasonable accuracy level in depth of frames by help of descriptors. A disparity map or depth map represents mapping of various pixels in Image where estimation is done by considering relative displacement of corresponding pixels. This disparity is proportional to baseline of the camera position [2]. The method corresponds in accumulating matching function in squared difference from stereo pair function instead of deciding on inherently used feature as depth of image.

Disparity and depth are inter-convertible based on stereo triangulation. Extraction of features to estimate depth from image is one of the important and hardest procedure having block and Image matching. Image matching is one of the commonly used technique to extract feature
of the sample Image from which we can analyze end to end estimation of depth including calculation of disparity. In work of W. Luo et al. [3] block match segments are used to match extracted features of the sample but this method is difficult to possess as the computational procedure is complex in nature. The disparity estimation over the corner detection of images is done under the work of N. Mayer et al. [4] by using Harris corner detection. But estimation of disparity over extraction of corner by this algorithm is time consuming. Many researches are done in improving the speed and time consumption but most them depends on performance of graphic processors. Depth estimation using hybrid patch-based Multiview stereo algorithm, over frames of stereos is done in work of Yasutaka et al. [5] and this method usually generates small patches from matched feature of stereo and this forms a gap between feature points over mesh model. The implemented method corresponds in accumulating matching function of SSD from stereo pair function instead of deciding on inherently used feature. When performing matching between two images there may exist some mismatched points in image and this existence of mismatching could be completely eliminated by adding RANSAC algorithm with SURF algorithm according to work done by Wei Chen et al. [6].

The Image fusion of stereo frames is proposed in this paper under usage of SURF algorithm for various application. The usage of image fusion is guided to user for producing stereo contains without prior knowledge of stereo Images. In shooting guide function of stereo proposed algorithm is used to obtain dynamic depth of Image without inclusion of artifacts. On contrary, there is no need of preserving spatial details like edge detection of images hence, this is one of the effective algorithms that could be addressed in various applications.

The remaining paper is organized as follows: In Section II we represent proposed work for fusing two images. In Section III, we describe experimental results followed by conclusion in Section IV.

II. PROPOSED WORK

![Diagram](image.png)

Figure 2: Process Flow Diagram

According to process flow chart as in Figure 2, image sample is captured by the two cameras aligned as Left camera and Right camera. Then Surf Algorithm is used in feature detection of the points to estimate and match points to get rectified image. Surf is the algorithm used for capturing images and representing them over invariant Image comparison. By the computation of Images using SURF algorithm we can analyze the Images over different application like object tracking and object recognition. The task involved in searching of discrete correspondence in Images is done in various levels. Initially all the interest points over the two Images of camera are considered to find the repeatability of points in different viewing conditions used for analyzing. In the next step geometrical deformation of feature vectors is used as descriptors to find robust towards noise and in detecting errors of vectors. Finally, descriptor vector is matched among two different Images over Euclidean distance of two cameras.

Surf Algorithm is usually applied to the Image initially for finding Feature Extraction and later feature matching of Image is estimated. This is partially inspired by image transformation done by SIFT (Scale-Invariant Feature Transform) algorithm. To detect interest points in image SURF uses scale-normalized determinant of Hessian matrix is used as SURF descriptor for image matching. Detecting of interest points is computed with help of integral image. Transforming of image data into set of features present in image is feature extraction. In the feature extraction the approach uses approximation computing the sum of values of pixels in the image that helps in calculating the average intensity of the Image in two perspective by utilizing integral Images and weiner filter estimation. By using this over the descriptor interest points, second order gaussian derivatives is evaluated at low computational cost that is independent on the size of Image. Feature extraction by this approach helps in providing context information at multiscale features [7]. By using this over the descriptor interest points, second order gaussian derivatives is evaluated at low computational cost that is independent on the size of Image. Extraction of descriptors by this is done using Gaussian filter scale σ at computational window position (x, y) at pixel P and defined by hessian matrix as

\[ H(x, y) = \begin{pmatrix} L_{xx}(x, y) & L_{xy}(x, y) \\ L_{xy}(x, y) & L_{yy}(x, y) \end{pmatrix} \]

where \( L_{xx}(x, y), L_{xy}(x, y) \) and \( L_{yy}(x, y) \) represents approximation of pixels at second order partial derivative with response \( \frac{\partial^2 g(\sigma)}{\partial x^2}, \frac{\partial^2 g(\sigma)}{\partial x \partial y} \) and \( \frac{\partial^2 g(\sigma)}{\partial y^2} \) with pixel P of image taken at values are found for value of L at window position (x, y) and (y, y). This method helps in finding box filters as proposed by Bay et al. [8] in image segmentation by using Gaussian function considered as \( g(\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \).

For this Haar wavelet response is applied at points on region of interest and usage of this algorithm provides us more efficiency than SIFT. By this discrete second order derivatives required descriptor is obtained as Blob of the images as described in work done by Shambavi Jain et al. [9].
Features detected as blobs of images using hessian matrix are matched between two images by method of feature matching. Matching corresponding Images are matched based on the camera calibration over the same scene. From the Image data obtained detection of set of interest points is extracted to match preliminary features of the Image considered. Later in the normalized region is local descriptors is found using the key points in the normalized region to match associated points. A global approach over the disparity algorithm is applied to formulate difference between pair of pixels in process of the disparity values. Mapping of disparity values is done by assigning similar depth of the neighboring pixel by estimating difference between them. Hence, assigning of disparity map is achieved over the epipolar line located over coordinate systems. Finally, by squared difference of pixels the two Images are combined together to form the rectified output of the two Images. This method helps us in obtaining more efficient fused image when compared to SIFT algorithm as per work by WenTao et al. [10].

In the implementation of work process a novel approach is made here in converting the input images to gray scale image as this helps in restoration of low illuminance image at various color shifts that could be supervised as proposed by Zou Muchun et al. [11]. Also, the conversion of gray scale related work is demonstrated efficiently at [12], that is further used in processing of images as per the work done in [13].

III. RESULTS AND DISCUSSION

For stereo vision of two Images proposed algorithm is implemented by MATLAB software. Here in Figure 3 object captured by two different cameras are named as Left camera Image and Right camera Image and are imported for applying SURF algorithm. For analysis purpose these two Images are used in obtaining rectified image by applying SURF Feature point detection and matching. Here in images of dataset there is difference in disparity in point of two images taken which is caused by geometry displacement of scene hence conversion of RGB image to gray scale is necessary. In a two horizontally displaced plane camera difference in relative planes of image is observed in baseline of the two cameras. A grayscale image helps in providing pixel information and helps in applying algorithm during convolution of image and this conversion also helps in disparity mapping of two images. Therefore, images are converted to grayscale for applying algorithm as in Figure 4. When two RGB Images of left and right camera are combined there is formation of composite Image that is used to find difference of feature points where left right Images are overlapped. In order to know the difference found in blobs of images Left camera Image is represented by Red color and Right camera Image is represented by Cyan color as shown Figure 5.

Next is detection of SURF points detection in Composite Image which is defined as interesting point of Image. In examination of region of Image there were found 30 interest points that is to be matched between two Images left and right by putative matching as in images. The putative point is extracted from valid blobs in Composite image that are defined as Interest points based on correspondence between images. These are smoothened by higher level pyramids in scale spaces where edges and blobs could be found for matching. The putatively matching points are used in finding average difference of points to obtain exact point of descriptor of blobs. Later Outliers are removed and matching is done for inliers points of both images. Finally, rectified Image is obtained by applying squared sum of difference between pixels as in Figure 7. This rectified image is image rectification of composite image that was represented by Red and Cyan colors used for two images identification. If optical axes of images are parallel then accordingly epipolar lines converge that simplifies in providing matching of images. Aligning of cameras to form image alignment based on configuration of features forms rectified images. And distance between corresponding matches in shift of stereo pairs forms disparity mapping of images used in finding in pixel wise matching. The disparity mapping of two images is done for analyzing pixels in the Composite image over scale for finding depth scale estimation of combined image as in Figure 8.
IV. CONCLUSION

In this paper, we present an efficient feature-based method for Image Fusion in stereo vision that uses extraction and matching of image feature points by SURF algorithm. For fusing of two images, implementation is done using MATLAB software as it is user-friendly and provides an interactive environment for various computations. By using SURF Algorithm, there is a new way to match two images feature points in a feasible way with reduced mismatch points. One further direction in improvable work could be applied for Multiview video frames in stereo vision matching.

REFERENCES