Volume 6, Issue 9, September 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Object Detection by Feature Matching Method

[1]Gurbinder Singh, [2]Kumar Shanu, [3]Rahul Malhotra, [4]Kulbhushan Rassewat
[1, 3, 4] G.T.B.K.I.E.T., [2] Punjabi University, Patiala
Punjab, India

Abstract— Image matching is a fundamental aspect of computer vision, including object recognition or scene recognition. For the Human recognition system despite of different viewpoints and differences, it is easier to identify the images. This task is still a challenge for computer vision for extracting the scale – invariant and shift – invariant features from images to perform reliable object recognition. In this paper, the object recognition system which can resolve the difficulty of rotations of object, scale changes and illumination are resolved with the help of "Modified SIFT algorithm". It is implemented with different phases such as scale space extreme detection, key point localization and varied threshold value. The Modified SIFT algorithm implemented with MATLAB, which is one of efficient tool to perform image processing. The recognition proceeds by matching individual features to a database of features from known objects using a fast nearest-neighbor algorithm, followed by a Hough transform to identify clusters belonging to single object, and finally performing verification through least-square solution for consistent pose parameters. This approach to recognition can robustly identify objects between clutter and occlusion while achieving near real-time performance.

Keywords— SIFT,

I. INTRODUCTION

Object detection may be a engineering involving laptop vision and image process that deals with police work instances of linguistics objects of a precise category (such as humans, buildings, or cars) in digital pictures and videos. Well-researched domains of object detection embody face detection and pedestrian detection.[1] Object detection has applications in several areas of laptop vision, as well as image retrieval and video police investigation. Object detection is that the method of finding instances of real-world objects akin to faces, bicycles, and buildings in pictures or videos. Object detection algorithms usually use extracted options and learning algorithms to acknowledge instances of an object class.[1] It's unremarkably employed in applications akin to image retrieval, security, police investigation, and automatic vehicle parking systems.

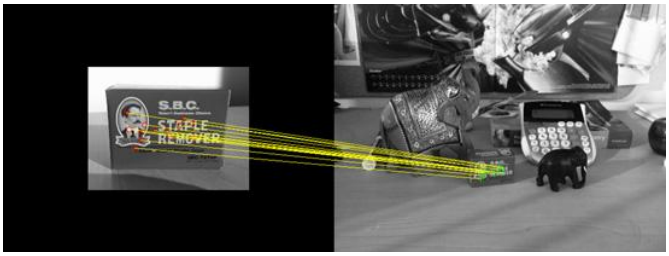


Figure 1: Detecting a reference object (left) in a cluttered scene (right) using feature extraction and matching. RANSAC is used to estimate the location of the object in the test image.

It is task (within pc vision) of finding and distinctive objects in a picture or video sequence. Humans acknowledge a large number of objects in pictures with very little effort, despite the very fact that the image of the objects could vary somewhat in several read points, in many various sizes and scales or maybe once they are translated or turned.[2] Objects will even be recognized once they are part stopped-up from read. This task continues to be a challenge for pc vision systems. several approaches to the task are enforced over multiple decades.[2]

We can acknowledge objects employing a kind of models, including:

- Extracted options and boosted learning algorithms
- Bag-of-words models with options admire SURF and MSER
- Gradient-based and derivative-based matching approaches
- Viola-Jones algorithmic rule
- Template matching
- Image segmentation and blob analysis

Vision

Vision is one amongst the 5 senses of humans, of course the foremost vital one as a result of ninetieth of the knowledge that the human brain receives from the external surroundings is thru vision.[3] To interpret then to move with

totally different objects within the encompassing surroundings is that the main goal of human vision. The existence capabilities of humans e.g. perceiving thousands of objects, distinctive many faces, appreciating beauty, recognizing traffic signs etc. need solely lowest effort[3]. The closing of those tasks could seem easy however it proves the high degree of development of the human vision system.

Computer Vision

Computer Vision could be a field of applied sciences whose main objective is to produce to computers those functions that are gift in human vision. The everyday computer-vision applications embody golem navigation, medical imaging, video streaming, industrial internal control etc. with none doubt, throughout the past decades, spectacular progress has been created during this field, however today the computer-vision applications rely heavily on the human sensory system concerning their lustiness and performance.[4] The human-vision impressed computer-vision systems are promising alternatives as so much because the lustiness and performance of computer-vision solutions are involved. If we expect as researchers in developing computer-vision applications, the question ought to be asked, and answered, what mechanisms are concerned in human vision that build it straightforward for humans however troublesome for computers.[4] Visual Attention is a capability of the vision systems to speedily choose the salient and relevant data/objects in an exceedingly visual scene. The core objective of visual attention is to attain the smallest {amount} doable amount of visual data to be processed to resolve the high level advanced tasks e.g. seeing, which might lead the total vision method to become economical. This visual attention mechanism should be a part of the solution to the question that has been asked on top of.

II. NEED OF OBJECT RECOGNITION

There is an accord regarding the very fact that it's doable to divide the human vision in 2 main phases: low level vision and high level vision. It's been already established what the most role of every level is, however what's not however clearly outlined is that the frontier between 2 phases. the method starts from the low level part, wherever the visual data is gathered by the retinas at the speed of bits per second. Within the next step, the knowledge gathered is passed to the visual area, wherever this data is processed to extract the knowledge regarding options like color, motion, orientation, depth and form. Then these extracted options ar transmitted to the high level vision wherever it performs its tasks. the most practicality of high level vision is to acknowledge the contents of a scene, that is completed by means that of matching the options of a representative scene to an oversized information that consists of learned and memorized objects.[10] Despite the very fact that human vision has to method immense amounts of visual data which recognition task is of combinatorial nature, there's estimation that giant numbers of objects is recognized by humans in lesser time than two hundred ms. This astonishing performance of the human brain can't be explained by means that of given computation resources, since such associate degree quantity of data can't be processed by roughly neurons of human brain in such a brief time whereas having such a slow rate of response. These facts regarding the high performance of the human brain reveal the high potency of the human vision system.[10]

The high performance of the human sensory system is explained by means that of the existence of a mechanism, within which solely the reduced set of data is taken into account for the high level process, which mechanism is named the visual attention. Hence, the structure of the membrane reinforces the hypothesis of the visual attention mechanism's existence. the very fact that photoreceptors are distributed in heterogeneous fashion over the membrane, ends up in the perception that solely a reduced a part of the sight view is exactly perceived, whereas the remaining a part of the sight view is perceived mistily. Hence, it's necessary to shift the central a part of the membrane i.e. fovea, for the perception of the foremost informative elements of the sight view. Thus, the orientation of the fovea centralis for the choice of informative elements of the visual scene is controlled by the visual attention mechanism.[10]

In the field of pc vision, the complexness of the machine tasks, that ar referred as NP-Complete, like seeing and sensory activity grouping, ar called the fundamental obstacles for the real-world applications. Thus, to beat this complexness issue in computer-vision applications, the connection of the visual-attention paradigm is very accepted.

Indeed, we are able to say that the visual attention could be a preprocessing step that permits the choice of a set of accessible sensory data.[11] Once the salient elements of a visible scene ar detected, then the high level tasks of pc vision will concentrate on these specific scene locations. In seeing ways there are some variable ends up in stereo and mono conditions. The disadvantage of SIFT that slowness of process speed of seeing. These ar computationally expensive and inefficient scanning technique. Among all the on top of object detection and recognition ways we tend to are considering comparison numerous object detection ways. In these ways we'll try and compare the varied blessings and drawbacks of the on top of outlined ways and might check that technique is best in object detection.[12] Within the SIFT technique associate degree object is detected on the premise of properties of the image and also the object i.e. pixel, brightness, frames and also the scale in variances, translation. thus to ascertain the distinction among the results of each the techniques we tend to are comparison the 2 techniques in order that we are able to determine the higher one that uses the image properties expeditiously.

III. RELATED WORK

A) Object Representation Methods:

First step of object trailing is that the illustration of the article of interest. Object may be painted by their form and appearance. In this section, we are going to initial describe the article form representations used for trailing. Then address the joint form and look representations [1].

1. Points: the article is painted by a degree, that is, the centre of mass or it's painted by a collection of points [2]. This representation is appropriate for trailing objects that occupy tiny regions in a picture.

2. Primitive geometric forms: Object shape is painted by a parallelogram, conic [2], etc. Object motion for such representations is typically sculptured by translation, or projective homography transformation. These area unit additional appropriate for representing straightforward rigid objects. they're conjointly used for trailing non rigid objects.

3. Object silhouette and contour: The boundary of an area is outlined by the contour illustration. The region within the contour is termed the silhouette of the article. These area unit appropriate for trailing advanced non rigid shapes [3].

4. Articulated form models: Articulated objects area unit composed of body elements that area unit control along side joints. for instance, the physical body is associate articulated object with body part, legs, hands, head, and feet connected by joints. The relationship between the elements is ruled by kinematic motion models, for instance, joint angle, etc. Articulated objects area unit painted by modeling the constituent elements victimization cylinders or ellipses.

5. Skeletal models. Object skeleton may be extracted by applying medial axis remodel to the article silhouette [4]. This model is employed as a form illustration for recognizing objects.

B) Object Detection Methods:

Tracking mechanism needs associate object detection mechanism once the article initial seems within the video. Once the objects are painted victimization any of the mentioned models next step is to discover the article within the frame. This is often done once the article first seems within the frame or video. The temporal info of the article within the initial frame is extracted to discover it. Some models use quite one frame to extract information; that's done by frame differencing. a number of object detection techniques area unit as follows:

1. Frame differencing: The moving object is decided by shrewd distinction between 2 consecutive pictures. It's robust ability for selection of dynamic environments and it's tough to get complete define of moving object [5].

2. Optical Flow: In this technique image optical flow field is calculated. And agglomeration process is finished in keeping with the optical flow distribution characteristics of image. This technique gets the entire movement info and detects the moving object. This technique is sensitive to noise, poor anti-noise performance [6].

3. Background subtraction: First step for background subtraction is background modeling. Background Modeling ought to be sensitive therefore on acknowledge moving objects. Background Modeling outputs a reference model. This reference model is employed in background subtraction. In background subtraction every video sequence is compared to the reference model to work out potential Variation. The variations between current video frames to it of the coordinate system in terms of pixels signify existence of moving objects [7]. Currently, mean filter and median filter area unit wide accustomed understand background modeling. The background subtraction technique is to use the distinction technique of this image and background image to discover moving objects. this is often a straightforward algorithmic rule, however terribly sensitive to the changes within the external atmosphere. This technique has poor anti-interference ability. It provides the entire object info within the case background is understood [8]. varied background subtraction models area unit MOG (Mixture of Gaussians), Bayesian call rules, the Codebook-based model, Kernel density estimation [9]. The Codebook algorithmic rule [10] constructs a background model supported a quantization/clustering technique. first for every component a background model is constructed. This model contains one or additional codeword. A codeword could be a system that contains info concerning color, brightness and frequency [10]. Stauffer & Grimson et al. [1] projected a mathematician mixture model supported background model to detect the article. Mixture of Gaussians was accustomed model the component color [1]. The component within the frame is compared with the background model. whereas checking the component is compared with each mathematician within the model until a match is found. If found, the mean and variance of the maths is updated [1]. Non constant quantity Kernel density estimation may be accustomed model the per-pixel background. The component is matched with the component within the background model and with the close pixels [11].

C) Object Tracking Method:

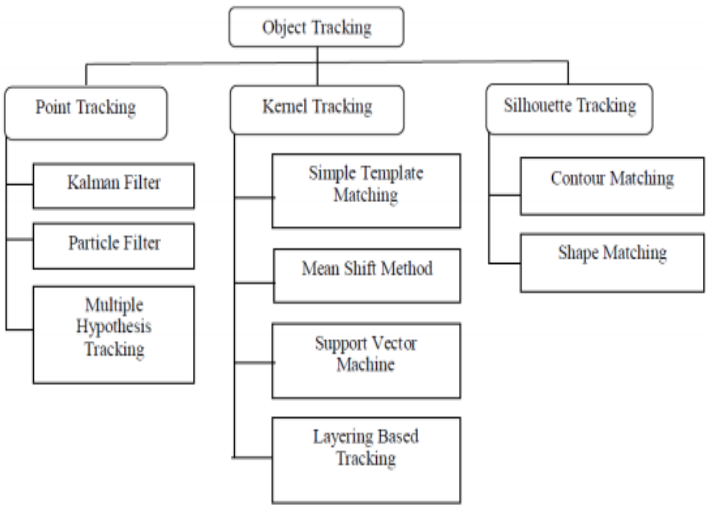


Fig 1. Object Tracking Methods [12]

1. Point Tracking: In image, moving objects area unit diagrammatic by their feature points. Within the incidence of occlusions there's the matter of false detection of object happens. purpose correspondence strategies area unit divided into 2 classes, namely, settled and applied math methods. The settled technique use qualitative motion heuristics to oblige the matter of correspondence. Probabilistic strategies take the thing activity and uncertainties into consideration to determine correspondence [1]. Kalman filters are often used for object following. Kalman filters area unit supported optimum algorithmic processing algorithms. Kalman filter consists of 2 part prediction and correction phases. consecutive state is predic d victimization this set of observation. then this set is updated. The second step provides updates the expected values and offers approximation of consecutive state. Particle filter uses contours, color options or texture mapping for object following. Variable that isn't sampled is chosen. Particle filter samples the variable in step with proposal distribution [12]. MHT (Multiple Hypothesis Tracking) algorithmic rule is AN unvarying algorithmic rule.

Prediction regarding the position of the thing within the frame is created. Then distance live is employed to check the prediction.

2. Kernel Tracking: Kernel following computes the motion of object kind frame to border. The motion of the thing is in variety of constant motion or dense flow field computed in resulting frames. There area unit 2 subcategories density-based look models, and multi read look models. Density-based models area unit straightforward and have relative low machine value. Templates area unit formed victimization image intensity or color feature. There area unit 3 strategies in kernel following approach straightforward example matching, mean shift technique, straightforward vector machine (SVM) and layering based mostly following [12]. Straight forward example matching will track solely single object. The thing of interest is verified with the frame kind the video. It will alter the partial occlusion of the thing. In mean shift technique the thing of interest is outlined victimization rectangular frame. Then the caterpillar-tracked object is separated from background. It uses translation and scaling to trace the thing motion. This technique will alter part occluded objects. The simple vector machine uses coaching set of values. These coaching values area unit positive or negative, positive values contain caterpillar-tracked object and also the negative sample contains the values that aren't caterpillar-tracked. In Layering based mostly following multiple objects are often tracked. conic section is employed to represent the form of the thing and uses layer look supported the intensity. The background motion of the thing is paid then the every pixels chance supported the foreground motion is calculable. This method will alter full occlusion drawback [13].

3. Silhouette Tracking: This technique generates an object model supported the previous frame. Victimization these object model the thing from every frame is find out. This model are often within the variety of a color bar graph, object edges or the thing contour. Histograms of color and edges are often used because the object models. This technique models the thing look by the sting info obtained within the object silhouette to match silhouettes in consecutive frames [1].

a. Contour following:

Contour following strategies, iteratively progress a primary contour within the previous frame to its new position within the current frame. This contour progress needs that certain quantity of the thing within the current frame overlay with the thing region within the previous frame. Contour following are often performed victimization of 2 completely different approaches. The primary approach uses state area models to model the contour form and motion. The second approach directly evolves the contour by minimizing the contour energy using direct diminution techniques like gradient descent. the foremost important advantage of silhouettes following is their flexibility to handle an oversized kind of object shapes [12].

b. Shape Matching:

These approaches examine for the thing model within the existing frame. Shape matching performance is comparable to the template based mostly following in kernel approach. Another approach to form matching is to search out matching silhouettes detected in 2 successive frames. Silhouette matching, are often thought-about like purpose matching. Detection supported Silhouette is carried out by background subtraction. Models object area unit within the variety of density functions, silhouette boundary, object edges. Capable of handling single object and Occlusion handling are performed in with Hough rework techniques [12].

IV. SIFT METHOD

SIFT (Scale Invariant Feature Transform) algorithm proposed by Lowe in 2004 to solve the image rotation, scaling, and affine deformation, viewpoint change, noise, illumination changes, also has strong robustness. The SIFT algorithm has four main steps: (1) Scale Space Extrema Detection, (2) Key point Localization, (3) Orientation Assignment and (4) Description Generation. The first stage is to identify location and scales of key points using scale space extrema in the DoG (Difference-of-Gaussian) functions with different values of σ, the DoG function is convolved of image in scale space separated by a constant factor k as in the following equation. Where, G is the Gaussian function and I is the image. Now the Gaussian images are subtracted to produce a DoG, after that the Gaussian image subsample by factor 2 and produce DoG for sampled image. A pixel compared of 3×3 neighborhood to detect the local maxima and minima of D(x, y, σ). In the key point localization step, key point candidates are localized and refined by eliminating the key points where they rejected the low contrast points. In the orientation assignment step, the orientation of key point is obtained based on local image gradient. In description generation stage is to compute the local image descriptor for each key point based on image gradient magnitude and orientation at each image sample point in a region centered at key point. These samples building 3D histogram of gradient location and or orientation; with 4×4 array location grid and 8 orientation bins in each sample. That is 128-element dimension of key point descriptor.

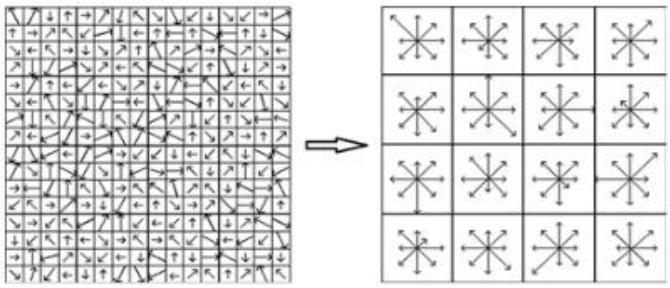


Fig 2: SIFT Descriptor Generation

The key point descriptor is shown on the right side of Fig 1. It allows for significant shift in gradient positions by creating orientation histograms over 4x4 sample regions. The figure shows 8 directions for each orientation histogram, with the length of each arrow corresponding to the magnitude of that histogram entry. A gradient sample on the left can shift up to 4 sample positions while still contributing to the same histogram on the right. So, 4×4 array location grid and 8 orientation bins in each sample. That is 128-element dimension of key point descriptor. SIFT object recognition approach is a view-centered object detection and recognition system with some interesting characteristics for mobile robots, most significant of which is the ability to detect and recognize objects in an un-segmented image. Another interesting feature is the Best-Bin-First algorithm used for approximated fast matching, which reduces the search time by two orders of magnitude for a database of 100,000 key-points for a 5% loss in the number of correct matches. The first stage of the approach consists on matching individually the SIFT descriptors of the features detected in a test image to the ones stored in the object database using the Euclidean distance. As a way to reject false correspondences, only those query descriptors for which the best match is isolated from the second best and the rest of database descriptors are retained. In Figure 2, the matching features between a test and model images can be seen. The presence of some outliers (incorrect pairings of query and database features) can also be observed. Once a set of matches is found, the Generalized Hough Transform is used to cluster each match of every database image depending on its particular transformation (translation, rotation and scale change). Although imprecise, this step generates a number of initial coherent hypotheses and removes a notable portion of the outliers that could potentially confuse more precise but also more sensitive methods. All clusters with at least three matches for a particular training object are accepted, and fed to the next stage: the Least Squares method, used to improve the estimation of the affine transformation between the model and the test images.

Finally, we have incorporated some domain knowledge by defining several heuristic rules on the parameters of the estimated affine transformation to reject those clearly beyond plausibility. Namely:

- Hypotheses with object centers that is too close.
- Hypotheses that have a ratio between the x and y scales below a threshold.

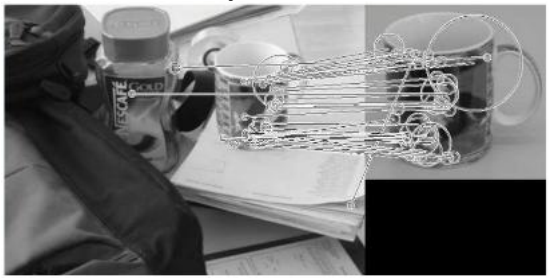


Fig 4: Matching stage in the SIFT object recognition method

V. PROPOSED METHOD

In this section the proposed approach of object detection and the results generated using that proposed approach is defined. For the result generation various parameters are defined like precision and recalling which is described below.

Using the SIFT-operator, the two object images (model and test) are transformed into two SIFT-image feature sets. These two feature sets are divided into subsets according to the octaves in which the feature arise. In order to realize the base procedure mathematically, a quality-integer function $F(x)$ is defined as:

$$F(x)=\begin{cases}\sum_{j=0}^{j=x} Z(M_1^{n-1-x+j}, M_2^{j})\ldots if\ (x<n)\\ \sum_{j=0}^{j=n-1} Z(M_1^{x-n-1+j}, M_2^{j})\ldots if\ (n<x<m) \qquad m\geq n \quad and \quad x\in[0,m+n-2]\\ \sum_{j=0}^{j=m+n-2-x} Z(M_1^{x-n+j}, M_2^{x-m+1+j})\ldots if\ (x\geq m)\end{cases}$$

Among all found matches it can happen that a lot of correct matches exceed Lowe's threshold τ. In order to retrieve these correct matches, the ratio between the Euclidean distance to the nearest and the second nearest feature neighbour must be reduced. This can be done either by reducing the smallest distance $D_1(F_1^i, F_2^{j0})$ or by increasing the next smallest distance $D_2(F_1^i, F_2^{j1})$. An additional result of the research presented in this paper is consideration of the improvement of the original SIFT algorithm with respect to the processing time. As first, it can be shown that the original SIFT procedure and the procedure developed in this paper complete the matching procedure in the same time.[1] Thus,

the complexity of original SIFT-matching procedure is proportional to the product P_1= l.h. Complexity of proposed system is proportional to the following sum of the product.

Recall =*TP* (*TP* +*FN*) ,

1− Pr*ecision* = *FP* (*TP* +*FP*).

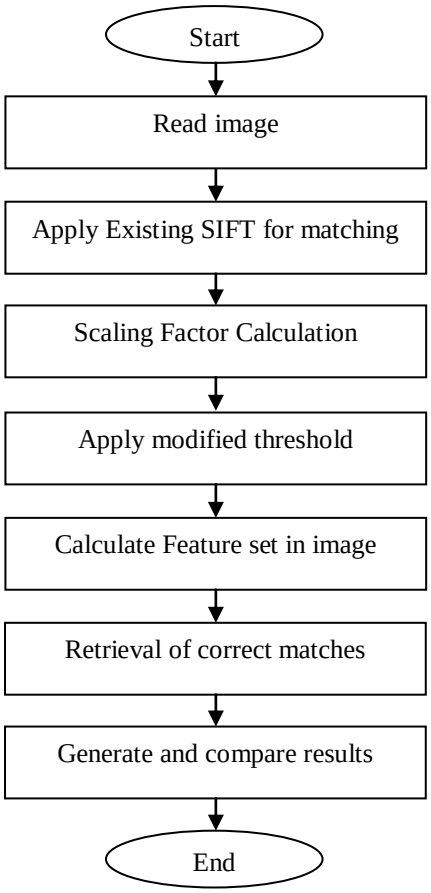
In the very first step an image is to be uploaded for the feature analysis. After the uploading of the image various features of the image will be extracted and saved so that these features can be used for comparison with any other images for detection and recognition purpose.

The SIFT algorithm has four main steps:

(1) Scale Space Extrema Detection
(2) Key point Localization
(3) Orientation Assignment
(4) Description Generation.

The first stage is to identify location and scales of key points using scale space extrema in the DoG (Difference-of-Gaussian) functions with different values of σ, the DoG function is convolved of image in scale space separated by a constant factor k as in the following equation.[1]

Where, G is the Gaussian function and I is the image. Now the Gaussian images are subtracted to produce a DoG, after that the Gaussian image subsample by factor 2 and produce DoG for sampled image. A pixel compared of 3×3 neighborhood to detect the local maxima and minima of D(x, y, σ). In the key point localization step, key point candidates are localized and refined by eliminating the key points where they rejected the low contrast points[1]. In the orientation assignment step, the orientation of key point is obtained based on local image gradient. In description generation stage is to compute the local image descriptor for each key point based on image gradient magnitude and orientation at each image sample point in a region centered at key point. these samples building 3D histogram of gradient location and or orientation; with 4×4 array location grid and 8 orientation bins in each sample. That is 128-element dimension of key point descriptor.



Flow Chart

VI. RESULT AND DISCUSSION

Proposed ISIFT Scheme: is an approach for extracting distinctive invariant features from images. It has been successfully applied to a variety of computer vision problems based on feature matching including object recognition, pose estimation, image retrieval and many others. The main idea is to divide the features extracted from both the test and the model object image into several sub-collections before they are matched. The features are divided into several sub-collections considering the features arising from different octaves that are from different frequency domains. In the existing scheme threshold of 0.8 provides 95% of correct matches as positive and 90% of false matches as negative. The

total amount of the correct positive matches must be large enough to provide reliable object recognition. In the proposed scheme the number of correct matches is to be increased using the feature set and threshold based upon the Euclidian distance.

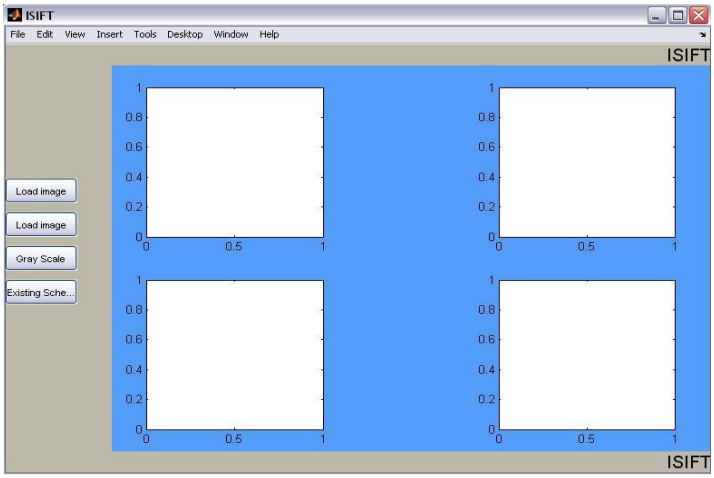


Figure 5: GUI

In the figure 5 the GUI for the current research is displayed. In this GUI we can upload images and convert it to gray scale which is a requirement for feature extraction and then can apply existing SIFT feature extraction technique to generate key-points and match among 2 images or multiple images.

Figure 6: Image Upload

Figure 6 is presenting screen in which one image is uploaded in it. Using this uploaded image a SIFT based existing scheme is applied and generate key-points and match them to generate results.

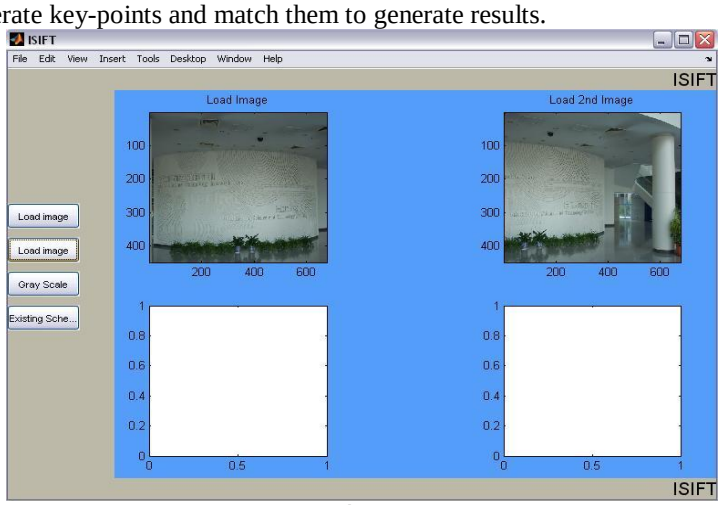


Figure 7: 2nd Image upload

Figure 7 is presenting screen in which 2nd image is uploaded in it. Using this uploaded image a SIFT based existing scheme is applied and generate key-points and match them to generate results.

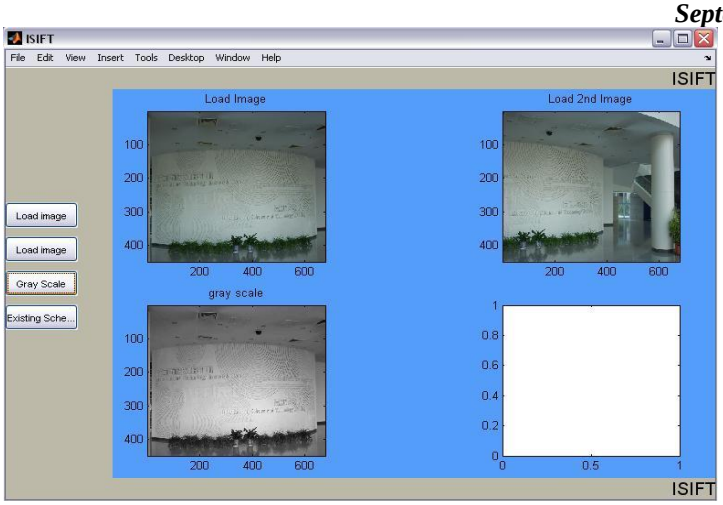


Figure 8: RGB to Gray scale image

Figure 8 is presenting screen in which images are to be converted into gray scale image which is a prime requirement for feature extraction. Using this uploaded gray scale image a SIFT based existing scheme is applied and generate key-points and match them to generate results and then it will be compare with the proposed scheme.

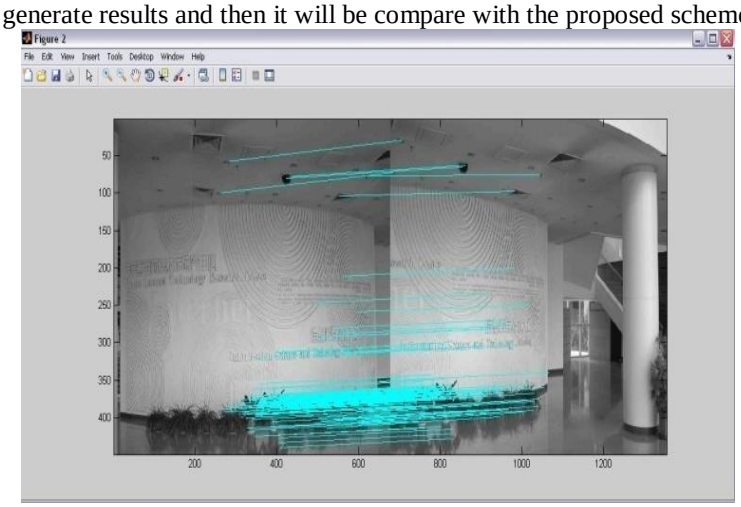


Figure 9: Existing Scheme

Figure 9 is presenting resultant screen in GUI. This figure describing the output of the research in which the two images are compared to generate key-points and to extract and match it. In this figure the uploaded images are compared and find out the various key-points in them and then these key-points are compared to generate results.

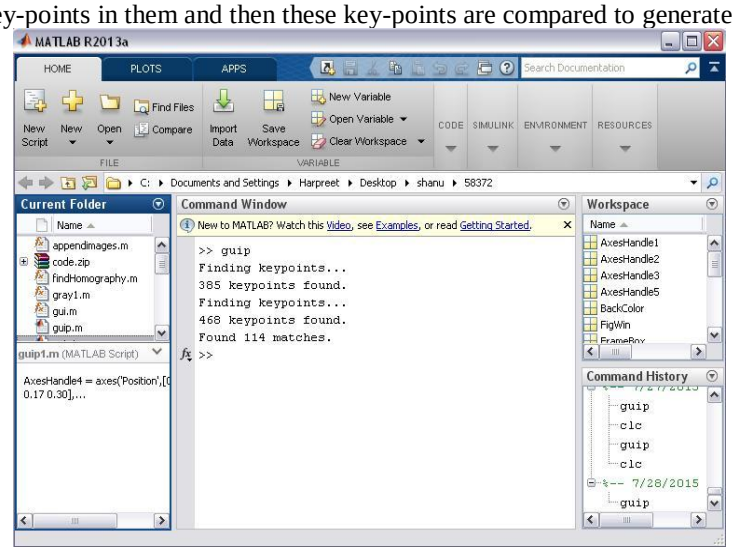


Figure 10: Results for existing scheme

Figure 10 is presenting results for the images uploaded in it. Using this uploaded image a SIFT based existing scheme is applied and generate key-points and match them to generate results. In this figure it is very clear that there are 385 key-points generated but matched only 114.

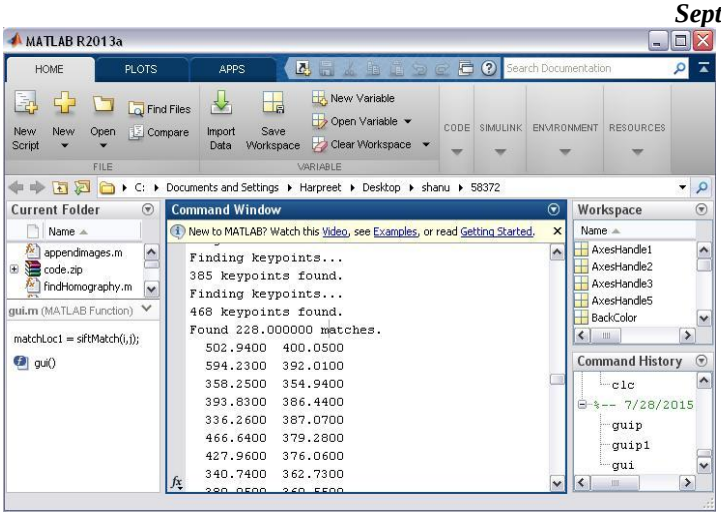


Figure 11: Results for proposed scheme

Figure 11 is presenting results for the images uploaded in it. Using this uploaded image a SIFT based existing scheme is applied and generate key-points and match them to generate results. In this figure it is very clear that there are 385 key-points generated but 228 that is more than that of existing scheme.

Key Points in Image		Improved SIFT matching		Modified SIFT Matching	
Left	Right	Matching Time(sec)	Number of inliers	Matching Time(sec)	Number of inliers
217	229	0.025	133	0.020	139
777	640	0.230	325	0.212	340
3014	2233	4.950	683	3.879	749
6871	6376	47.790	856	43.462	976

VII. CONCLUSION

In this paper varied phases of object tracking system viz. object detection, object classification and object chase has been studied. Obtainable strategies for these phases are explained in details and variety of defect and limitations were highlighted in every technique. Entirely different strategies for object detection area unit frame distinction, optical flow and background subtraction. As from the results and discussion section the proposed accuracy and efficiency is 95.5 and 98.5 as compared to existing which is 91 and 96.5 respectively[16]. Same in case of Recall vs precision, Number of inliers and matching time that are 0.8, 1000 and 43 sec in case of proposed scenario as compared to existing which possess values as 0.72, 820 and approximately 49 sec in existing respectively. From the current research it may be clarified that the proposed technique is much better than that of existing SIFT as it is 20% more efficient and accurate. Advance study is also dispensed to incorporate notice economical algorithmic program to cut back process value and reduce the time needed for chase the item for kind of videos containing distributed characteristics.

REFERENCES

[1] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. Acm Computing Surveys (CSUR), 38(4):13, 2006.

[2] COMANICIU, D., RAMESH, V., AND MEER, P. 2003. Kernel-based object tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence 25, 5, 564–577.

[3] YU, S. X. AND SHI, J. 2004. Segmentation given partial grouping constraints. IEEE Trans. Patt. Analy. Mach. Intell. 26, 2, 173–183.

[4] BALLARD, D. AND BROWN, C. 1982. Computer Vision. Prentice-Hall.

[5] Rupali S.Rakibe, Bharati D.Patil, "Background Subtraction Algorithm Based Human Motion Detection", International Journal of Scientific and Research Publications, May 2013

[6] Abhishek Kumar Chauhan, Prashant Krishan, "Moving Object Tracking Using Gaussian Mixture Model And Optical Flow", International journal of Advanced Research in Computer Science and Software Engineering, April 2013

[7] J.Joshan Athanesious, P.Suresh, "Systematic Survey on Object Tracking Methods in Video", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) October 2012, 242-247.

[8] Himani S. Parekh1, Darshak G. Thakore 2, Udesang K. Jaliya 3,"A Survey on Object Detection and Tracking Methods". International Journal of Innovative Research in Computer and Communication Engineering. Vol. 2, Issue 2, February 2014

[9] Enrique J. Fernandez-Sanchez, Javier Diaz and Eduardo Ros, "Background Subtraction Based on Color and Depth Using Active Sensors". Sensors 2013, July 13

[10] P.Siricharoen T.H.Cchalidabhongse S.Siddihichai, "Robust Outdoor Human Segmentation based on color-bassed Statistical Approach and edge combination" Proceedings of IEEE, vol. 22, August 2010.

[11] Marius Leordeanu, Rahul Sukthankar and C Sminchristia,"Generalized Boundaries from Multiple Image Interpretations" Proceedings of IEEE, vol.36,july 2014

[12] Qiang chen Zheng song, Rogerio Feris" Efficient Maximum Appeareance Search for Large-scale object detection" Proceedings of IEEE ,vol. 6.Dec 2013.

[13] Arnau Ramisa, David Aldavert, Shrihari Vasudevan, Ricardo Toledo, Ramon Lopez de Mantaras, "Evaluation of the SIFT Object Recognition Method For a Mobile Robotic Platform",vol. 32, April 2013.

[14] Karel Zimmermann,David Hurych,Tomas Svoboda"NON-Rigid object detection with local interleaved sequential Alignment" proceedings of IEEEvol. 36 April 2014.

[15] Khushboo Khurana, Reetu Awasthi, "Techniques for Object Recognition in Images and Multi-Object Detection", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), ISSN: 2278 – 1323, Vol. 2, Issue 4, April 2013

[16] Faraj Alhwarin, Chao Wang, Danijela Risti -Durrant, Axel Gräser, "Improved SIFT-Features Matching for Object Recognition",institute of automation,university of Bremen,vol. 5, Nov 2008

