ISSN 2229-5518

Review of Various Feature Extraction Technique & Their Application

Shruti Meshram
Department of Computer Science & Engineering
Technocrats Institute of Technology (Excellence)
Bhopal, India
Shrutimeshram1309@gmail.com

Dr. Vivek Sharma
Department of Computer Science & Engineering
Technocrats Institute of Technology
Bhopal, India
Sharma.vivek95@yahoo.com

Abstract— Feature extraction is a process of mapping of the creative higher dimensional signal dimensions into a lower dimensional feature space. One of the mainly difficult jobs in visual object recognition is to shrink the quantity of calculation in feature matching the process removes visual features from images using opt for computer vision methods, and processes the extracted characteristics to generate a condensed record of characteristics annotated with the frame numbers as they appear.

Index Terms— Feature extraction, Scale Invariant Feature Transform (SIFT), feature space.

I. Introduction

Visual perception is the most important and informative way for human beings to obtain information from the surrounding environments. We use visual perceptions to recognize objects and faces and then convert this perceptual information to semantic information for further high-level processing. The interest of using computer to simulate the human perception ability induces extensive study on computer vision, specifically, object and facial recognition.

Computer vision techniques have become relevant with the topic of navigation with the development of different feature extraction methods. Images can be recognized within its features. A feature point is a position which should be unique, invariant to image scaling, revolving, explanation or camera viewpoint, and strong to noise. There are several algorithms to extract features from images, like SIFT [1], SURF [2], BRISK [3], FREAK [4] etc. Several algorithms make available only point locations in an image, and a few algorithms make available a feature descriptor for the each of the illustration points.

A feature descriptor is a vector recitation the exceptional properties of the equivalent feature point. SIFT and SURF has there own descriptors. FREAK creates only points, which can be explained using descriptors like BRIEF [5] or ORB [6]. Image matching is achieved in feature space. If two images have feature points explained with the identical algorithm, and the descriptor vector has N dimensions, then in that N-dimensional space two characteristics from different images neighboring to each other are deliberate a match. In a circumstance like direction-finding, a map can have several, possibly thousands of images. To be on familiar terms with a position, a user requires capturing a frame with a camera, and presenting a feature matching with the images in the diagram. The image with most matches with the captured frame can present the location of the user. There is a challenge of image identification for navigation. As stated before, a map dataset can have several thousands of images. Matching a frame with each one of them will certainly be a interference to real-time retrieval. Image features, which are comparable to each other according to various pre-defined criteria, can be confidential as a visual word. When matching features to get back an image from a database, investigate is presented on these visual words, considerably dropping the search gap. On the other hand, if image retrieval is executed for the purpose of direction-finding in a large map, all the features have to be in use into description to make the illustration word dictionary or codebook, and some features may not be valuable for recognition.

II. Theoretical Concepts

Feature extraction is a mapping of the original higher dimensional signal measurements into a lower dimensional feature space. Their application maintains to develop in a variety of fields gradually. From uncomplicated photogrammetric jobs for instance feature recognition, to the development of difficult 3D modeling software and image's search engine, there are several applications where image matching algorithms play a significant job. Furthermore, this has been a very dynamic part of research in the modern years and as point out by the remarkable quantity of work and documentation issued around this. More than a decade ago, the applications associated with 2D and 3D models and object modernization were essentially for the reason of visual examination and robotics. Nowadays, these purposes now incorporate the use of 2D and 3D models in computer graphics, virtual reality, announcement and others application area. But accomplishing extremely dependable matching consequences from a pair of images is the job that some of the most popular matching techniques are demanding to achieve. But none have been commonly acknowledged. Four feature spaces like time, statistical, spectral, and principal component analysis and combinations thereof have been proposed in the literature for extracting the characteristic elements for identification of bearing faults and classification of military vehicles. The best possible evaluate of the effectiveness of all removed features is Bayes error [7]. With the intention of determine Bayes error, one would need to get hold of the subsequent probabilities, which necessitates the stage a time consuming evaluation of the nonparametric densities. We will find out the effectiveness according to the investigational classification simulations employing the same classification algorithm between the same set of detected occasions using the same data sets.

III. Feature Extraction

Identifying features in an image is an essential step in image matching. Each feature is an interest point in the image calculated using fundamental image properties, and a local descriptor which describes the interest point and its neighboring points. In most cases, this descriptor is a high-dimensional vector. The extraction of local image features used in BoF framework involves two steps. The first step is interest point detection and the second step is descriptor computation. In this part, here we give explanation these two operations, and evaluate their significant methods. Figure 2.2 demonstrates keypoint detection and descriptor calculation using the recognized Scale Invariant Feature Transform or SIFT [8] as an example.

2.2.1 Interest Point Detection

Some applications in computer vision, that utilize local image features, may calculate these features on a dense grid. However, for image matching and augmented reality applications, it is useful to search for points of interest in the image. These points should be repeatedly detectable in different images that contain the same object or capture the same scene. These key points should also be invariant to shift, scale and rotation of the object of interest. Key point detection techniques more often than not search for corners or blobs in the image because of their high do again, and the ability to allocate a precise location to these corners or blobs. The accurate localization of key points is a significant footstep in the direction of accomplishing shift invariance, because in the case of shift, the same key point will be detected at a different place in the image. Scale invariance is usually achieved by calculating a scale-space representation of the image [8, 9] and searching for key points at different scales. Finally, rotation invariance is achieved by assigning an orientation to each key point. This orientation depends on the direction of the maximum gradient in the area around the interest point. Some detectors like the DoG detector [8] make available key points with precise localization and high repeatability yet necessitate high computational complication. Others methods are quick and straightforward to calculate yet manufacture key points with low repeatability. For example, Features from Accelerated Segment Test (FAST) detector, excludes non-corner points through simple operations in the pixel domain without gradient computation. However, it operates at a single scale producing non-scale-invariant key points.

2.2.2 Descriptor Computation

After key point detection, the second step in the local feature extraction pipeline is descriptor computation. A descriptor is an efficient description of the canonical patch that can be used for matching similar patches and provides high discrimination against non-similar patches.

Many descriptors like SIFT [8], SURF [9], CHoG, RIFF and GLOH share the common framework that the descriptor consists of Histograms of Gradients (HoG) in the canonical patch located around the detected key point. The canonical patch is first partitioned into spatial bins. SIFT and SURF uses a square grid of spatial bins, while CHoG, RIFF and GLOH use polar spatial binning. A histogram of gradient magnitudes is then estimated for each spatial bin. Such as, SIFT executes an pointed binning of the gradient magnitudes in 8 different directions Gaussian weighting is applied to assign more importance to the center of the canonical patch. An ending normalization pace is executed to provide invariance against brightness changes. Hence, the SIFT descriptor is a 128-dimensional signal that concatenates 16 histograms of gradient magnitudes from the spatial bins.

Similarity of two descriptors is evaluated using a suitable distance metric such as the L_2 norm or symmetric KL divergence. In [10], Mikolajczyk and Schmid evaluate the performance of many feature descriptors in a common framework. Some image retrieval pipelines calculate a global descriptor based on the local feature descriptors in the image. The use of global descriptors provides a compact description of the whole feature descriptor set of an image, and eliminates the need to build an SVT for image retrieval.

International Journal of Scientific & Engineering Research, Volume 7, Issue 4, April-2016
ISSN 2229-5518

IV. Visual Features

Visual feature is a very generic term that may correspond to any distinctive visual characteristic examined in an image. It can be a texture or a colour but also a shape or a corner point. In our circumstances, we search for to determine and illustrate rigid parts of objects by combination of various visual features. Since our system does not offer a method to part visual characteristics, it is essential that each of them is not bring into play to illustrate more than one inflexible neighborhood. Our preference is then maximum value to local visual features i.e. characteristics wrapping a very minute region of the image. A very large number of feature detectors have been developed through the years. At an overview level, they can be divided into the following groups with some overlap:

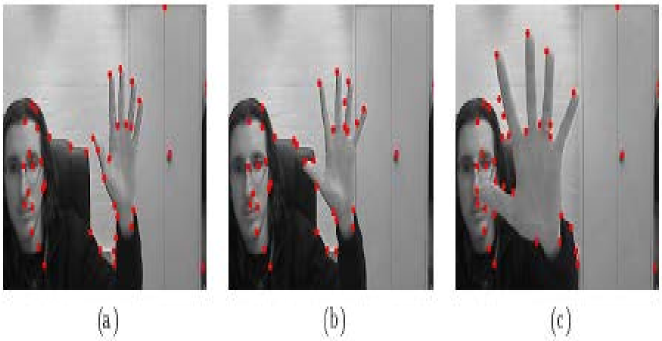

Figure 1: Example of detection of interest points using the Harris detector.

As we can see, texture less objects like the hand are not well covered by characteristics points. An additional difficulty they can examine is the formation of features on intersections between background and foreground lines (see for example the thumb and pinky on Figure 1.b.

Corners / Interest Points: The term interest points refer to a point in an image which has a well-defined position and can be robustly detected. Similarly, a corner point corresponds to the intersection of two edges and can therefore be distinguished from its undeviating areas. The expression of awareness spot is essentially more common than the term corner point and can correspond to other structures like blobs. Despite this, corners and interest points are sometimes used interchangeably in the literature; which can be a little confusing. So, to make things clear, we focus here on point wise visual features. Wider “interest zones” like blobs will be discussed later.

Interest points are probably the most widely used type of feature in computer vision and many different methods have been proposed to extract them [11]. The success of interest points is due to a number of qualities such as a distinct location, a more often than not prosperous local information substance, steadiness under affine transformations, one thing that is adjacent to them though is their lack of ability to increase uniformly on objects with no or maximum value texture. For example, objects like the hand in Figure 1 do not enclose many forceful corners making these objects difficult to model with a set of interest points. Moreover, features are also created on intersections between background and foreground lines. With the number of valid foreground points quite low, this means there is a high risk to get a significant percentage of outliers in the initialization of the model; increasing the risk for the tracking to fail i.e. at least during the beginning of the model learning phase.

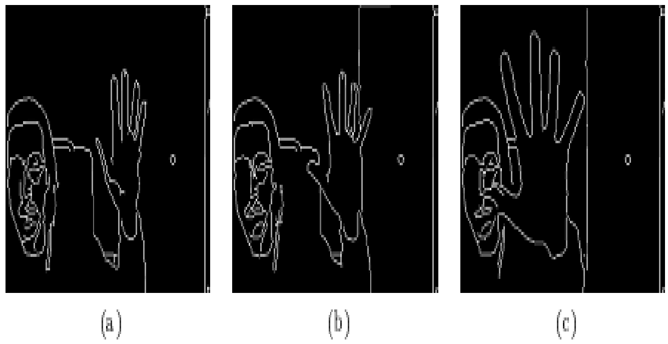

Figure 2: Example of edge detection using the canny edge detector with the same as previous images.

While the result is not ideal, it makes available a extra spontaneous exposure of the hand than the corners/interest points. While the hand would have indeed been difficult to identify using the corners only, it is easily recognized using the edges.

Edges: Edges are locations where there is a boundary between two image regions. In practice, edges are usually defined as sets of points in the image with a strong gradient magnitude in one direction. These points can then be chained together to form a more complete description of an edge. Depending on the length of these chains and their figure, the edges are extra or a smaller amount local and additional or a lesser amount of distinctive from their surrounding means that the longer the chain, the higher the risk of covering more than a single rigid part, and the more complex the shape, the lower the risk to confuse it with its surrounding.

The main drawback of edges is that, locally, they are only defined in one dimension: perpendicular to the gradient. Tracking them consequently in general necessitates more than the local knowledge provided by a single edge. Edges are also very sensible to clutter since they don't have an expressive power comparable with the significance points. On the other hand, edges make available an extreme enhanced exposure of texture-less objects than interest points see the examples in Figure 2. They also are largely invariant to lighting conditions and variations in object colour and texture, making them a better candidate to model object categories. On top of that, they can be matched accurately along the object boundary, while image patches and local descriptor vectors usually used with interest points tend to be more difficult to match when the background is changing.

Blobs/Regions of Interest: In computer vision, the expression region of interest more often than not submits to a region in the image that is either brighter or darker than its neighboring i.e. in the case of colour images; this definition must be true for as a minimum one way through channel. These regions can be distinguished with for example, convolutions with kernels such as a Laplacian of Gaussian or a difference of Gaussians. This approach is for example used by the SIFT detector [12] which then tends to overlap with the techniques used for the interest points detectors. Another popular detector is the Maximally Stable Extremal Region (MSER) detector originally defined in [13].

Skeleton and Ridges: Skeleton points or medial axis points make available a perceptive, compact representation of a shape, making them demanding for many applications in computer vision. Various algorithms have been extended for skeleton extraction using for example space transform, topological thinning, Voronoi diagrams or gradient vector flow [14]. Apart from their high sensitivity to noise in the object's boundary, the skeleton points cannot be computed directly from the raw image and require a contour or a silhouette with the intention of be taken out. It can consequently be used in arrangement with contour points [15] but is difficult to use as a feature on its own. Moreover, skeleton extraction requires a high computational cost in order to deal with all the edges detected inside the object and therefore not relevant for the skeleton extraction.

Ridges are defined as the local maxima of the image intensity. They are useful to detect the medial axis of elongated objects or parts. These features are not frequent in computer vision since they are highly scale sensitive and are only effective if the objects are significantly lighter or darker in the inverted image than they surrounding.

.

V. Literature Survey

Visual feature extraction with scale invariant feature transform (SIFT) is commonly exploited for object recognition. On the other hand, its instantaneous accomplishment goes through from extensive latency, intense totaling, and high memory storage for the reason that of its frame level calculation with iterated Gaussian blurs process. Consequently, this paper author recommends a new technique using real time layer parallel SIFT (LPSIFT) with fundamental image, and its parallel hardware intend with an on-the fly based feature extraction engine run so that only incomplete transitional consequences have to be accumulated for that is proficient to calculate 2000 feature points for HD1080p30 at 100 MHz. The proposed plan [16] implements the layer parallel restructured box kernel to substitute iterated Gaussian blur operations for easy and parallel computation. This also diminishes the latency to a small number of image lines as an alternative of several frames.

Here they compared [16] with the novel SIFT algorithm, the proposed technique come within reach of decreases the computational quantity by 90% and memory usage by 95%. After the concluding accomplishment utilize on 580-K gate count with 90-nm CMOS technology, and recommends 6000 feature points/frame for VGA images at 30 frames/s and ~2000 feature points/frame for 1920×1080 images at 30 frames/s at the clock rate of 100 MHz. With these performances, the existing design straightforwardly accomplishes the real time require with considerably lower cost, which saves 56% gate count and 90.4% memory cost when evaluated to the earlier plan.

Feature extraction and matching is at the support of many computer vision difficulties, for example objects recognition or arrangement from movement. Existing techniques for evaluating the presentation of well-liked image matching algorithms are offered and rely on expensive descriptors for detection and matching images. In particular, the technique evaluates the different type of images under which each of the algorithms evaluated here present [17] to its highest or achieving maximum efficiency. The effectiveness is calculated in expressions of the numeral of equivalents founds by the algorithm and the number of type I and type II errors come a crossed when the algorithm is hardened in opposition to a definite match up of different images. Existing relative learning asses the concert of the algorithms based on the consequences acquired in unusual criteria such as momentum, understanding, occlusion, and some others criteria are used. This learning addresses the restrictions of the obtainable relative instruments and distributes a take a broad view standard to find out earlier the level of competence anticipated from a matching algorithm given the type of images estimated. The algorithms and the particular images used inside this work are divided into- two groups: feature-based and texture based.

This paper [17] has estimated five feature detection techniques for image deformation. SIFT is measured very slowly and not well at scaling alterations, at the same time as it is invariant to regular change, enlighten transforms and affine transformations. And from this wide categorization only three of the most extensively used algorithms are evaluated: color histogram, FAST (Features from Accelerated Segment Test), SIFT (Scale Invariant Feature Transform), PCA-SIFT (Principal Component Analysis-SIFT), F-SIFT (fast-SIFT) and SURF (speeded up robust features). The presentation of the Fast-SIFT (F-SIFT) feature detection techniques are evaluated for scale changes, rotation, blur, illumination changes and affine transformations. Fast SIFT is faster than usual SIFT and become visible good in unusual characteristics but SIFT is enhanced concert than fast SIFT. SURF is quick and has excellent performance as the equivalent as SIFT, but it is not constant to rotation and enlightenment transforms. It was bringing to a closed that F-SIFT have the most excellent in general presentation beyond SIFT and SURF but it experiences from become aware of very a small number of characteristics and consequently matches. It is suggested that the assets of this algorithm to be get better by generating a new accomplishment make available with a matching constituent. It is essential to get better F-SIFT by ever-increasing the quantity of characteristics it can identify. But unique concern should be taken to safeguard the strength of the algorithm and keep away from the detection of ineffective characteristics. All the researches use duplicate skill quantity and the number of accurate matches for the appraisal dimensions. SIFT here find its strength in most circumstances even though it's deliberate. F-SIFT is the best ever one with good concert as the equivalent as SURF, SIFT, PCA-SIFT show its improvements in rotation and clarification transforms.

Here Bag-of-words representation is put into operation and make an effort on the 10-class visual concept detection problem, Here author [18] discover the new concept move toward of "DURF+ERT+SVM" accomplishes much enhanced detection presentation than "SIFT+ERT+SVM", which give you an idea about that the number of illustration image patches is a considerable aspect. And by using intense example and SURF-like descriptor, DURF is around 10 times faster than SIFT in feature extraction. Joining the DURF and SIFT gives the most excellent consequence in the researches, which shows that even enhanced consequence can be anticipated if other example approaches and feature extraction algorithms are collective. The detection concert on the perceptions that have enormous intraclass dissimilarity, for example *person* and *bird*, is not adequate. Accumulation spatial information into the bag-of-words representation strength enhances the discovery concert on these difficult classes, which is the future work.

The investigational effects on this paper [18] give you an idea about that "DURF+ERT+SVM" do better than "SIFT+ERT+SVM" both in detection presentation and calculation competence. As well, uniting DURF and SIFT consequences in still enhanced detection show. Concurrent object detection using SIFTS and RANSAC is also attempting on easy objects can detected, e.g. drink can, and excellent effect is accomplished.

In the modern earlier period, the recognition and localization of objects based on local point characteristics has grow to be a extensively acknowledged and make the most of technique. Along with the most well-liked characteristics are at this time the SIFT features, the additional current SURF characteristics, and region-based features such as the MSER. For time-critical purpose of object recognition and localization systems working on such characteristics, the SIFT features are excessively deliberate i.e. 500–600 ms for images of size 640×480 on a 3GHz CPU. The quicker SURF accomplishes a calculation time of 150–240 ms, which is still excessively deliberate for dynamic tracking of objects or visual serving applications. In this paper [19], author try to present a arrangement of the Harris corner detector and the SIFT descriptor, which calculates characteristics with a high replicate and very good matching material goods within approx. 20 ms. While just calculating the SIFT descriptors for calculated Harris interest points would lead to an advance that is not extent invariant, here author has also try demonstrate how scale-invariance can be accomplished without a time-consuming scale space investigation. Additionally, they try to present results from experiments on simulated image data over and above on real image data from the humanoid robot ARMAR-III working in a kitchen background demonstrated the realistic applicability and presentation of the recommended features of doing well application of the proposed features within our method for recognition and localization of textured objects. A wide-ranging investigational assessment demonstrates the convenient applicability of our idea.

In this paper [20], author will talk about the modern computer vision literature on 3D object recognition. Here author will try to commence a general idea of the existing approaches of various significant difficulties in visual recognition, to investigate their strong points and limitations. In convectional application domain in computer vision object recognition is a basic application area in computer vision. For many decades, it is assume about as an area of widespread examine particularly in 3D objects. But 3D object recognition can be characterized as the undertaking of pronouncement and recognizing objects in the real world from an image or a video progression. It is still a latest study area in computer vision for the reason that it has many disputes for instance perspective discrepancies, scaling, illumination transforms, partial occlusion, and environment clutter. Many approaches and algorithms are proposed and implemented to overcome these challenges.

Finally, they will present [20] exacting confronts in 3D object recognition move towards that have been used in recent times. In addition to, probable ways for upcoming investigate will be presented in this area of applications. Through this investigation here they become aware of that an enormous arrangement of the research meeting pointed on reactive acknowledgment, to some amount, on the feature selection period of the recognition difficulty exclusive of taking into thoughtfulness the consequences of various cost limitations chat about in the investigation.

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression, "One of us (R. B. G.) thanks . . ." Instead, try "R. B. G. thanks". Put applicable sponsor acknowledgments here; DO NOT place them on the first page of your paper or as a footnote.

REFERENCES

[1] David G. Lowe. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision, 60(2):91-110, November 2004. URL: http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94.

[2] Herbert Bay, Tinne Tuytelaars, and Luc Gool. SURF: Speeded up robust features. In Ale Leonardis, Horst Bischof, and Axel Pinz, editors, Computer Vision-ECCV 2006, volume 3951 of Lecture Notes in Computer Science, pages 404-417. Springer Berlin Heidelberg, 2006. URL: http://dx.doi.org/10.1007/11744023_32.

[3] Stefan Leutenegger, Margarita Chli, and Roland Yves Siegwart. BRISK: Binary robust invariant scalable keypoints. In Computer Vision (ICCV), 2011 IEEE International Conference on, pages 2548-2555, Nov 2011. doi:10.1109/ICCV.2011.6126542.

[4] Alexandre Alahi, Raphael Ortiz, and Pierre Vandergheynst. FREAK: Fast retina keypoint. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 510-517, June 2012. doi:10.1109/CVPR.2012.6247715.

[5] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. BRIEF: Binary robust independent elementary features. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios, editors, Computer Vision - ECCV 2010, volume 6314 of Lecture Notes in Computer Science, pages 778-792. Springer Berlin Heidelberg, 2010. URL: http://dx.doi.org/10.1007/978-3-642-15561-1_56.

[6] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. ORB: An efficient alternative to SIFT or SURF. In Computer Vision (ICCV), 2011 IEEE International Conference on, pages 2564-2571, Nov 2011. doi:10.1109/ICCV.2011.6126544.

[7] M.C. Wellman, N. Srour and D.B. Hillis, "Acoustic feature extraction for a neural network classifier", Army Research Laboratory - Technical Report, ARL-TR-1166, January 1997.

[8] D. Lowe. Distinctive Image Features from Scale-invariant Keypoints. International Journal of Computer Vision, 60(2):91-110, November 2004.

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

[10] K. Mikolajczyk and C. Schmid. Performance Evaluation of Local Descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(10):1615-1630, October 2005.

[11] Herbert Bay, Tinne Tuytelaars, Van Gool, and L. SURF: Speeded Up Robust Features. In 9th European Conference on Computer Vision, 2006.

[12] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vision, 60(2):91-110, 2004.

[13] J. Matas, O. Chum, M. Urban and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In British Machine Vision Conference, volume 1, pages 384-393, 2002.

[14] M. Sabry Hassouna and Aly A. Farag. On the Extraction of Curve Skeletons using Gradient Vector Flow. Computer Vision, IEEE International Conference on, pages 1-8, 2007.

[15] Nagesh Adluru and Longin Jan Latecki. Contour Grouping Based on Contour-Skeleton Duality. Int. J. Comput. Vision, 83:12-29, 2009.

[16] Liang-Chi Chiu, Tian-Sheuan Chang, "Jiun-Yen Chen, and Nelson Yen-Chung Chang, "Fast SIFT Design for Real-Time Visual Feature Extraction" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 22, NO. 8, AUGUST 2013.

[17] M.M. El-gayar, H. Soliman, N. meky, "A comparative study of image low level feature extraction algorithms" Egyptian Informatics Journal (2013) 14, 175–181.

[18] Ran Tao, "Visual Concept Detection and Real Time Object Detection" 2008.

[19] Pedram Azad, Tamim Asfour, Rudiger Dillmann, "Combining Harris Interest Points and the SIFT Descriptor for Fast Scale-Invariant Object Recognition" 2009.

[20] Khaled Alhamzi, Mohammed Elmogy 3D Object Recognition Based on Image Features: A Survey" International Journal of Computer and Information Technology (ISSN: 2279 – 0764) Volume 03 – Issue 03, May 2014.