Volume 4, Issue 3, March 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Hand Gesture Recognition based on SIFT Algorithm

Robinson Paul[1] Asst. Professor BVM, India

Ami Parikh[2] Research Student BVM, India

Shivani Patel[3] Research Student BVM, India

Himani Zala[4] Research Student BVM, India

Shikha Maheshwari[5] Research Student BVM, India

Abstract— This paper presents recognition of gestures using SIFT (Scale Invariant Feature Transform) algorithm. In gesture based technology, devices can get visual input and recognize person's gesture without touching keys or screens and use them to convey information or for controlling application. Hand gesture recognition allows various complicated operations to be performed on machine only by using fingers or hand movements reducing the need of physical contact. Thus, gesture recognition is considered to be highly adaptive interface between human and machines.

Keywords—Hand Gesture Recognition—SIFT Algorithm

I. INTRODUCTION

The present era of emerging technology in various areas, such as communication, medical, military, home automation and many others, gesture provide a natural and intuitive way to convey information. Gestures are powerful tools of communication among humans for non verbal communication. Hand gesture recognition can be used to enhance human–computer interaction by replacing traditional input devices such as keyboard and mouse. Therefore, reduces the complexity of interaction between humans and computer and developing a user friendly application. In future, almost all technology/ application would use natural human gestures to interact and control the devices without actually touching them.[6] Hand gestures can be classified in two categories: static and dynamic. A static gesture is a particular hand configuration and pose, represented by a single image. A dynamic gesture is a moving gesture, represented by a sequence of images. The stop sign is an example of static gesture and waving hand means goodbye is an example of dynamic gesture. Two common approaches for hand gesture recognition are Device-based and Vision-based approaches. Device - based method uses glove, stylus or position tracker, whose movements send signals that the system uses to identify gesture. Glove method uses special glove-based sensor device to extract information of position, orientation of fingers but they are quite expensive and also have much complexity therefore inefficient to use. Vision-based method is based on how we perceive information about our surroundings. Vision-based method includes 3D hand/arm modeling and Appearance modeling. We have used bare hand gesture recognition for the matching of images. [7]

II. LITERATURE REVIEW

Many methods for hand gesture recognition using visual analysis have been proposed for hand gesture recognition. Xia Liu and Kikuo Fujimura have proposed the hand gesture recognition using depth data [1]. Attila Licsar and Tamas Sziranyi have developed a hand gesture recognition system based on the shape analysis of the static gesture [2]. Another method is proposed by E. Stergiopoulou and N. Papamarkos [3] is based on color segmentation. Byung-Woo Min, Ho-Sub Yoon, Jung Soh, Yun-Mo Yangc and Toskiaki Ejima have suggested the method of hand gesture recognition using Hidden Markov models [4]. Rotation Invariant method is widely used for texture classification and recognition.

Image matching is an important computer vision problem and has application in various domains such as robot localization, content-based medical image retrieval, and image registration. Comparing images remains a challenging task because of issues such as variation in illumination conditions, differences in image orientation etc. Researchers have recently focused to local features in an image, which are invariant to common image transformations and variations. Mainly two broad steps are found in any local feature-based image-matching scheme. The first step involves detecting features (also referred to as key points or interest points) in an image in a repeatable way. Repeatability is important in this step because robust matching cannot be performed if the detected locations of key points on an object vary from image to image. The second step involves computing descriptors for each detected interest point. These descriptors are useful to distinguish between two key points.

METHODS:

• **Subtraction Method:** This method involves a simple subtraction between two images, pixel per pixel to compare them.
• **Gradient Method:** This method involves detecting edges in an image and counting the bright pixels that comprise them for each row or column and then making a comparison.
• **Principal Component Analysis:** The aim is to compute the Eigenvectors of the various pictures and then to express each image with its principal components (Eigenvectors).

• **Rotation Invariant:** This method involves two steps basically. First, achieving gray scale and secondly, achieving rotation invariance of an image. [8]

III. SIFT ALGORITHM

Scale Invariant Feature Transform (SIFT) algorithm was proposed by David Lowe [5]. SIFT is an algorithm to detect and extract local features in images. The SIFT algorithm is highly robust to scale invariance, illumination invariance and rotation invariance. A image matching method first detects points of interest, then selects a region around each point, and finally associates with each region a descriptor. Correspondences between two images can then be established by matching the descriptors of both images. The algorithm packages keypoints in each pixel location as [row, column, scale]. The acos of key vectors are sorted and the first 128 values are used as the feature vector for an image. The input image is compared with all its keypoints with the database image vectors where the nearest neighbour has angle less than the distance ratio, the keypoints are taken as matched. The maximum keypoints matched image is retrieved as that character.

Following are the major stages of computation used to generate the set of image features [5]:

Scale-space extrema detection: 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.

Keypoint localization: At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measures of their stability.

Orientation assignment: One or more orientations are assigned to each keypoint location based on local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations.

Keypoint descriptor: The local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination.

Algorithm Steps:

- Constructing the scale space: Scale space local extrema detection stage of filtering is used to identify locations and scales that are identifiable from different views of same object.
- Finding Gaussian difference: DoG can be computed by taking difference between two images, one with scale k times another. DoG technique is used to detect stable keypoints location in scale space.

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \qquad (1)$$

Where, $G(x,y,\sigma)$ is variable scale Gaussian
$I(x,y)$= input image

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) \qquad (2)$$

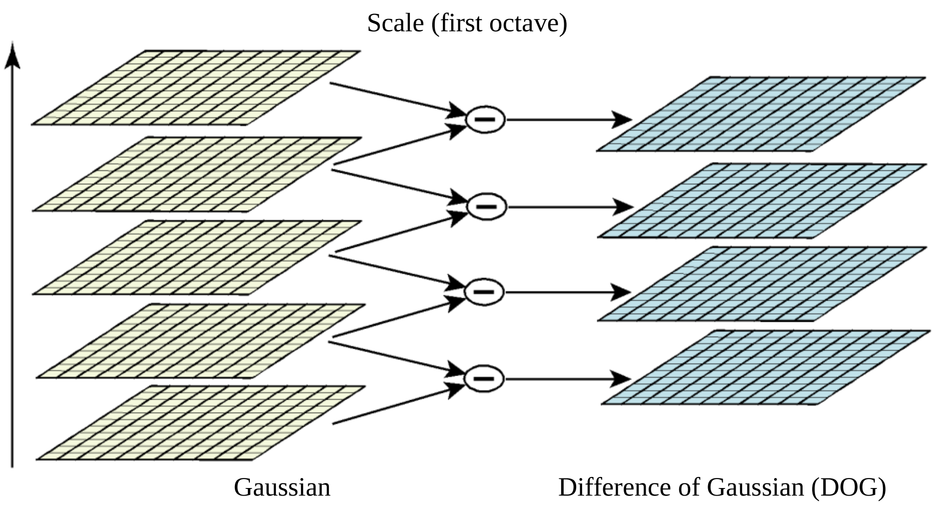

Fig.1 Difference of Gaussian (Source: Reference 5)

- Identify min and max from its neighbouring points: To determine local maxima and minima of $D(x,y,\sigma)$ each point is compared with its 8 neighbours at the same scale and its 9 neighbour up and down one scale.
- Identify potential feature points.
- Filter edge and low contrast responses using sobel edge detector
- Assign key point orientation which is used to create histogram for all values >0.9
- Build key point descriptor

FLOWCHART:

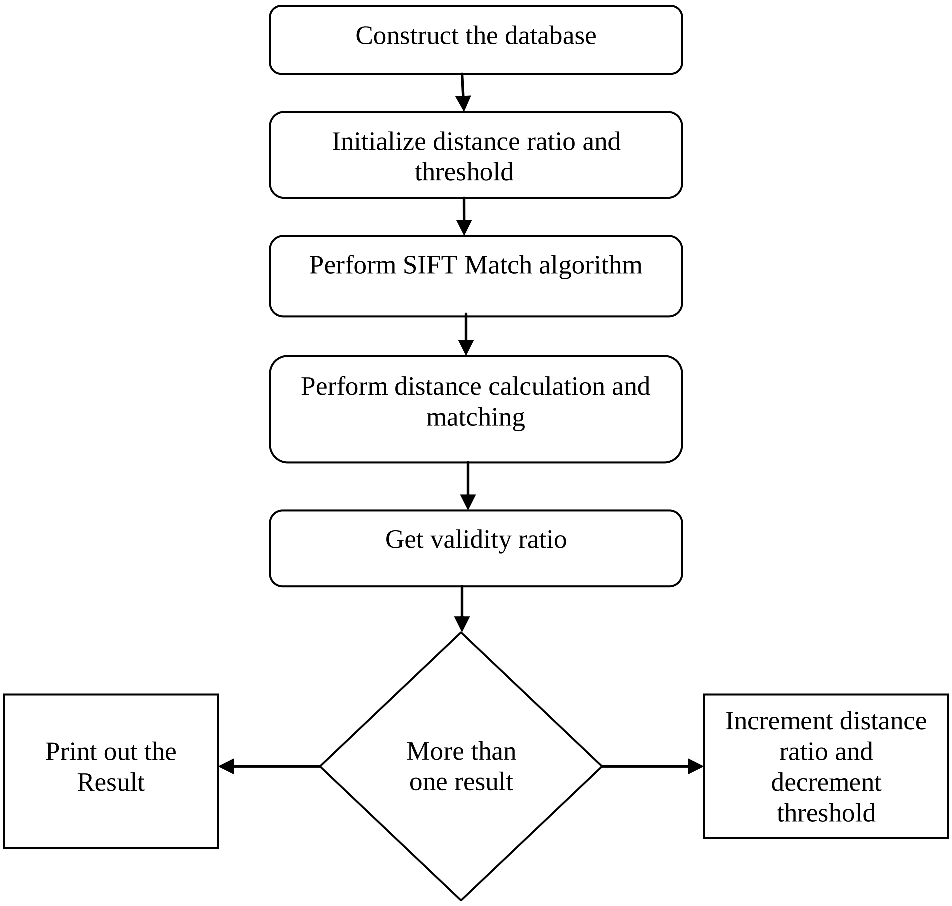

Fig.2 Flowchart

IV. EXPERIMENTAL RESULTS

In the training stage, we captured few training images with size of 50×50, 180x180 and 210x210 for every new hand posture and saved in folder named database. We have basically shown here 6 gestures named as start, stop, forward, backward, left and right.

Fig.3 Database set of the hand gesture (210x210 pixels)

In the testing stage, we captured 50 images with size 50×50, 180x180 and 210x210 for hand gesture similar to database with different angle variation. Of them, the analysis of few images are shown in the table1

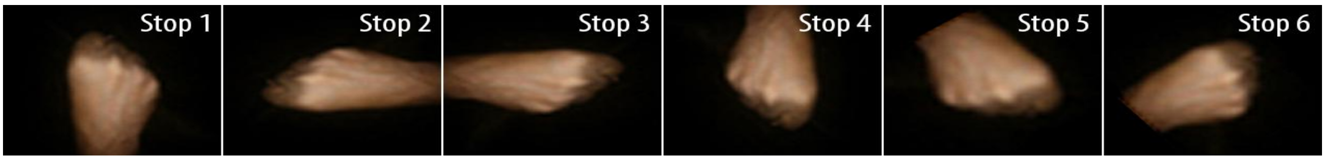

Fig4. (a) Stop gesture images with variation in angles (210x210 pixels)

Fig4. (b) Left gesture images with variation in angles (210x210 pixels)

Fig4. (c) Right gesture images with variation in angles (210x210 pixels)

The test images (210x210 pixels) were taken by camera and each of them rotated about 45 degrees from the original positions, yet the training image is easily recognized. We also varied the angle by 30 degree, 90 degree, 135 degree, 180 degree and many more from the original positions, yet almost every time the training images shows the correct result.

Table1. Analysis of gestures with different angle variation and threshold values (210x210 pixels)

Gesture Name	Number of images	Incorrect	Threshold value	Correct	Accuracy (%)	Recognition time (Second)
		0	0.035	10	100%	2.590562
		0	0.025	10	100%	2.403256
Start	10	0	0.020	10	100%	2.356359
		0	0.015	10	100%	2.291780
		0	0.010	10	100%	2.356432
		1	0.035	9	90%	2.522720
		1	0.025	9	90%	2.325567
Stop	10	1	0.020	9	90%	2.348552
		1	0.015	9	90%	2.092920
		1	0.010	9	90%	2.297399
		1	0.035	9	90%	2.361985
		1	0.025	9	90%	2.386890
Forward	10	1	0.020	9	90%	2.449288
		1	0.015	9	90%	2.267449
		1	0.010	9	90%	2.291959
		1	0.035	9	90%	2.448171
		1	0.025	9	90%	2.367726
Backward	10	1	0.020	9	90%	2.344366
		1	0.015	9	90%	2.092920
		1	0.010	9	90%	2.269714
		0	0.035	10	100%	2.540166
		0	0.025	10	100%	2.526730
Left	10	0	0.020	10	100%	2.461143
		0	0.015	10	100%	2.421002
		0	0.010	10	100%	2.406417
		0	0.035	10	100%	2.326266
		0	0.025	10	100%	2.387476
Right	10	0	0.020	10	100%	2.386695
		0	0.015	10	100%	2.291461
		0	0.010	10	100%	2.313750

Initially the threshold value was kept 0.035, but by decreasing the threshold value, all the test images were recognized correctly, which shows the robustness of the images. The total time taken to recognize the set of particular images is nearly 2 sec. The result analysis showed that we were able to match the correct images irrespective of the changes in rotation angle and maintained good accuracy. The result table shows the analysis of 10 different images with variation of gesture, but we have observed it for 50 images of different variation, still it is working properly and giving correct and accurate recognition. We have also observed it by resizing the images to 50x50 pixels and 140x140 pixels, still the robustness is observed.

V. CONCLUSION

The user can interact with the virtual environment using hand gestures. Relatively simple and fast, which can run in real-time on a workstation. The Algorithm is based mainly on using SIFT features to match the image to respective sign by hand gesture. Some modifications were made to increase the simplicity of the SIFT algorithm. Applying the algorithm on the training set, we found that it was able to identify the correct sign by hand gesture or to declare 'No Match' in case of no match condition. The algorithm was robust to scale difference, rotation by any angle from the test image. Maintain accuracy, even when different hands are used.

ACKNOWLEDGMENT

We would like to express our gratitude and sincere thanks to a great many people who helped and supported us during the conceptualization and implementation of the project as well as while writing the report. Our deepest thanks to Professor Robinson Paul, the Guide of the project for guiding and correcting various documents of ours pertaining to the project, with attention and care. He has taken pain to go through the project and made necessary correction and improvements as and when needed. We express our thanks to the Mr. Hiren Patel and staff of Birla Vishwakarma Mahavidyalaya Engineering College, specially our Head of Department Mr. Bhargav Goradiya for providing us with the infrastructure and facilities to accomplish the project. Lastly, we would like to thank the non-teaching staff of the ET department for their support. Without the support of the above mentioned, this project would be a distant reality.

FUTURE SCOPE

This Gesture recognition system can be further used in many application like home automation, banking, Robotics like autonomous robot to test its usability in the context of a realistic service task, smart watches. We can control application like Air Conditioner, TV, Mixer, Fan, Lights, etc. through gesture without use of switch. Even by using special image compression method [9], we can reduce the recognition time and design the specific chip for the same.

REFERENCES

[1] Xia Liu and Kikuo Fujimura, " Hand Gesture Recognition using Depth Data", Proc. of the Sixth IEE International conference on automatic Face and Gesture Recognition, pp. 529-534, 2004.

[2] Attila Licsar and Tamas Sziranyi, "Supervised training based hand gesture recognition system", Proc. of the 16th International Conference on Pattern Recognition, Vol. 3, pp 30999 – 31003, 2002.

[3] E.Stergiopoulou and N.Papamarkos: "A New Technique on Hand Gesture Recognition", Proc of the IEEE International Conference on Image Processing, 2657-2660, 2006

[4] Byung-Woo Min, Ho-Sub Yoon, Jung Soh, Yun-Mo Yangc and Toshiaki Ejima : "Hand Gesture Recognition Using Hidden Markov Models", Proc. of the IEEE International conference on Systems, Man and Cybernetics, vol 5, pp. 4232 -4235, 1997.

[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, Nov. 2004.

[6] Fakhreddine Karray, Milad Alemzadeh, Jamil Abou Saleh and Mo Nours Arab , "Human-Computer Interaction: Overview on State of the Art", International Journal on Smart Sensing and Intelligent systems, vol. 1, no. 1, march 2008

[7] Nicu Sebe, Michael Lew, Thomas S. Huang, Computer Vision in Human-Computer Interaction , ECCV 2004 Workshop on HCI, Prague, Czech Republic, May 16, 2004. Proceedings, ISBN: 978-3-540-22012-1 (Print) 978-3-540-24837-8 (Online), Volume 3058, 2004

[8] Prateem Chakraborty, Prashant Sarawgi, Ankit Mehrotra, Gaurav Agarwal, Ratika Pradhan , "Hand Gesture Recognition: A Comparative Study, Proceedings of the International MultiConference of Engineers and Computer Scientists 2008 Vol IIMECS 2008, 19-21 March, 2008, Hong Kong

[9] Robinson P. Paul, Chintan K. Modi and Rahul K. Kher , "A simple and novel algorithm for medical image compression", Int. J. Biomedical Engineering and Technology, Vol. 10, No. 4, 2012, Inderscience *Enterprises Ltd.*

.

