Enhanced Object Detection using Generalized Hough Transform and RANSAC

Rammah Yousef, Pranaw Kumar, Nabhan Yousef

Abstract: In this paper object detection process has been implemented using the combination of some robust algorithms. The algorithms used are SIFT, GHT and RANSAC. Two models have been proposed which using two of the above mentioned algorithms. In model one features extraction by SIFT has been processed using Generalized Hough Transform (GHT) , where in other model improved RANSAC has been used to eliminate the mismatches in order to achieve better recognition resolution. Overall GHT is found to be superior technique to achieve object detection efficiently.

Keywords : Object detection, SIFT, General hough transform, RANSAC.

I. INTRODUCTION

Being a local feature detecting and descriptive technique , Scale Invariant Feature Transform (SIFT) has been one of the most popular among existing computer vision algorithms .Features that are extracted using SIFT algorithm are invariant to rotation and image scaling. It is also partially robust to the changing view point and the changes in illumination .It is a four stage approach to extract local features from an image. The initial two stages are focused on extracting information of feature location, it also insure repeatable and reliable feature location. However the later two stages have focused on creating feature description. Initially an orientation is assigned to an image which makes the features invariant to rotation and after that the descriptor is created. For calculation of feature descriptor that results a feature vector of 128 elements long, a set of 16 histogram with 8 orientation bins has been used. The extracted features using this algorithm due to its invariance and robustness nature become an excellent candidate for object recognition. SIFT proves to be the most reliable technique to extract invariant features from an image. Adopting this algorithm, a number of object recognition systems have been reported so far.[1-3]. Further some improvement in SIFT have also been proposed in most of the research, especially enhancing its effectiveness [4-6].Obtained local features using SIFT may not always be useful .For characterizing the object for detection purpose. Hence ,it is needed to identify and eliminate features which can negatively affect the performance of the detection task.

Generalized Hough Transform (GHT) proposed by Ballard can be considered as an evidence gathering process. A proper match between a linear pattern and the model description results successful object detection.

II. MATHEMATICAL ANALYSIS OF THE ALGORITHMS USED

Scale Invariant Feature Transform (SIFT)

Scale Invariant Feature Transform is the most used algorithm for features extraction from an image among many proposed algorithms which has been developed by Davis G. Low in 2004[].Basically SIFT algorithm depends on the following steps:

Extrema detection: Gaussian smoothing from the input image and increasing the value of standard deviation (σ) producing difference of Gaussian (DoG) by subtracting successive Gaussian images in pairs. Scale and position of key points are determined by searching for big values in difference of Gaussian (DoG) images.
Enhanced Object Detection using Generalized Hough Transform and RANSAC

(DoG) image is referred as equation (3). \( L(x,y,\sigma) \) represent the blurred version of the original image at the coordinates \((u,v)\)on image with scale \(\sigma\) and this image can be expressed as :

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

(1)

\[
G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)
\]

(2)

\[
D(x,y,\sigma) = L(x,y,\sigma) - L(x,y,\sigma)
\]

(3)

Where \(I(u,v)\) is the original image \((x,y)\) are the distances from the point \((u,v)\). By changing \(k\)-step in equation (3) we get multiple DoG images with various scales. After that maxima and minima of DoG images are detected by comparing the target pixel with its 26 neighbouring pixels among \((3x3)\) scale , this extrema and minima points are regarded as key points.

Key points localization : Some of the detected key points which are sensitive to noise because of poor contrast should be rejected by selecting a suitable threshold value to limit the number of unwanted key points.

Orientation assigning: For each key point the direction of image gradient is determined with its magnitude and orientation that are calculated using equations (16) and (17).

Descriptors computing: A descriptor is made of generating an array from the local neighbourhood of each key point, then gathering descriptors in one array is made with region of \((3x3)\) pixels consisting of 9 pins. So the result is a vector of 128 elements length. These descriptors contain all the information such as: magnitudes and scales and gradient direction.

Some additional procedures should follow the features extracting process for completing the object detection task in good way.

General Hough Transform (GHT)

In order to detect the features of shape lines, circles, arbitrary shapes in a digital image, Hough transform is highly preferred which is regarded as a robust technique to detect multi-dimensional features and also to estimate parameters of the digital image. The motive of this technique is to detect imperfect instances of object within a particular class of shapes. It is achieved by a voting procedure, this voting procedure is furnished out in a parameter space where object candidates are taken as local maxima in accumulator space. It is explicitly constructed by algorithms based on Hough transform. A transformation between Cartesian space and parameter space has been done using Hough transform which results with defining a straight line. Let us consider the case that having straight line shown in (Fig. 1).

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

(4)

Again equation (4) represents a straight line on graph of \((c)\) versus \((m)\). As per the discussion made above, an algorithm for detecting lines in images can be explained as follows:

**Step-1:** All the line points in the image has been formed using a valid line detection scheme.

**Step-2:** The space \((m, c)\) has been quantized in two dimensional matrix \(H\) with proper quantization level.

**Step-3:** The matrix \(H\) has been initialized to zero.

**Step-4:** Each element of the matrix \(H(m, c)\) found corresponding to line point gets incremented by 1. It results a histogram or vote matrix that shows the frequency of line points with respect to particular \((m, c)\) values.

**Step-5:** A threshold histogram \(h\) is defined such that elements with large values are taken. Generalized Hough Transform that works on the principle of template matching is basically the modified version of Hough Transform. This enables the technique to easily detect any arbitrary object residing in its model along with the detection of the object mentioned in analytic equations. Finding the model positions somehow solve the problem of object finding in an image. GHT transforms the problem of finding model position into a problem of finding the transformation parameter which can map the model into respective image. After having the value of the transformation parameter, the model positioning in an image can be easily determined.

GHT implementation uses line information to show mapping from line point orientation to the reference point of the shape. In GHT, \((x, y)\) space is transformed into \((r, \theta)\) space. Hence, a line can generally be represented as:

\[
\rho = x \cos \theta + y \sin \theta
\]

Where \(\rho\) is the perpendicular distance from the origin to the line shown in (Fig. 1).

![Fig. 1 Converting (x, y) space into (ρ, θ) parameter space](image)

Using figure (1) and performing some trigonometric function, it can be shown that the slope of the line segment \(AB\) in \((x, y)\) space is given by:

\[
m = -\cos \theta / \sin \theta
\]

(7)

Similarly the intercept \(y\) from the right angle of triangle \(OPQ\) is given by:

\[
c = \rho / \sin \theta
\]

(8)

Again using equation (7) and equation (8) in general equation of straight line, we can write:

\[
y = -\frac{\cos \theta}{\sin \theta} x + \frac{\rho}{\sin \theta}
\]
Which further can be reduced to equation (6)
In equation (6), x and y are the constant image co-ordinates, where r and θ are considered as variables.

Random Samples Consensus
RANSAC is a method of robust estimation which calculates the parameters of a model on the basis of random voting principles. It can perform even in the conditions when the data space containing most of the outliers. It can also deal with multiple structure data effectively. However RANSAC algorithm represents a technique for separating subset of correct matches from initial matching and help in estimation of parameter of the transformation model. RANSAC algorithm can be explained in the following four steps:

Step-1: A suitable model is selected basing on the transformation model due to the difference found in between the reference image with respect to the sensed image. In order to calculate the various transformation parameter, the required number of the matched points is calculated using equation (10):

\[ q = p / 2 \]  
\[ (10) \]

Here, q represent the number of matched points needed to calculate the transformation parameter. And p shows the number of parameters in each transformation model.

Step-2: In order to select the based model in specific iteration, in each iteration randomly three matching points are selected to calculate transformation parameters for each iteration based on the obtained transformation parameters, the transformation model is calculated as per equation (11):

\[
\begin{bmatrix}
  a & b & c \\
  d & e & f \\
  0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
  x_1 \\
  y_1 \\
  1
\end{bmatrix}
= 
\begin{bmatrix}
  x_i \\
  y_i \\
  1
\end{bmatrix}
\Rightarrow
\begin{align*}
  x_i' &= ax_1 + by_1 + c \\
  y_i' &= dx_1 + ey_1 + f
\end{align*}
\]

\[ (11) \]

Where a, b, c, d, e, f are the transformation parameters. \( x_1,y_1 \) are the coordinates of matching points in a reference image. \( x',y' \) are the coordinates of the transformation model in the sensed image. This transformation model can again be written as: \( HP_e \), such that \( H \) represents transformation parameters and the matching point in the reference image is represented as \( (P_e) \).

Step-3: On being the transformation parameters are calculated for all the matching points in the reference image, in the sensed image the distance between \( (P_e, HP_e) \) is calculated. If distance is below the threshold value, matching point is accurately merged, otherwise, the matched point both in reference image and sensed image gets deleted. Also here empirically the tolerance error is determined.

Step-4: In each iteration, the number of points that accurately matched is calculated. If this number of accurately matched points exceeds the desired value, or equal to the predetermined maximum number of iterations the algorithm stops. Lastly, based on the transformation model and the matching points having largest number of accurate matching are re-calculated.

III. PROPOSED MODEL

We proposed two methods for the detection goal. Each one of these methods has shown effective results with good accuracy. First we proposed a method depends on both algorithms SIFT and GHT, the flow diagram of this method is shown in (Fig.2).

Fig.2. Flow diagram of object detection method 1

The proposed method follow the steps:

In this work, two images have been captured with a mobile camera having specifications of (1 megapixel) and resolution of (1280x720); this two input images have been referred as the subject for the testing. Out of the two considered pictures, one has been considered as test image and the other one as template image. After both images have been acquired properly, the candidate images are transformed to greyscale images from their RGB nature as preprocessing step. SIFT algorithm has been applied as mentioned earlier, for extracting features goal, and such features became scale and rotation invariant as an advantage of SIFT. As mentioned earlier from both test image and template image faces feature detection and extraction a perfect display of the matched features has been done to be worked on in GHT algorithm for applying line fitting and shape detection. After applying Hough transformation GHT is used to determine the edges of the objects as lines. At last after the proper application of both SIFT and
GHT algorithms a method for bounding the detected object with bounded box is used to show the accurate box which contains the object.

The second work is we worked on the same two images, following the same first two steps in the previous work, and next after the extraction features by SIFT algorithm, it is necessary to match the keypoints from the images (test and template) it is also required to calculate the distance between two matched points and also to consider the keypoints only after thresholding which in turn plays a good role in reducing the calculation complexity. After that a robust regression algorithm RANSAC is applied to estimate the geometric transformation depending on its M-estimator procedure within intention to identify the outliers, it keeps only the inliers that is obtained through estimation process. On successful applying of RANSAC algorithm and the accurate choosing of threshold that separate inliers from outliers and also the determination of iterations number the outliers gets discarded for the best match goal, Further the image show only the inliers.

The other used method that depends on SIFT algorithm and RANSAC algorithm, the flow diagram of this method is shown in (Fig. 3).

First proposed Method:

Generalized Hough Transform (GHT) for Object Recognition:

The general idea behind the Hough Transform is that objects are found through a voting process in a configuration or Hough space. If an evidence for a specific object configuration is found in the image, a vote is added for this configuration. In our case such evidence is a matched feature vector from the template image to the test image. The location, orientation and scale of the feature with respect to the object's coordinate system is known from the training image, so if the feature is also found somewhere in the image, it is possible to calculate the likely location, orientation and scale of the whole object and add a "vote" for that configuration (Fig.4a). After all matched feature vectors have voted, maxima in the configuration space correspond to possible object instances (Fig.4b). In our case, the configuration (or voting) space for the object is four-dimensional, the dimensions are x, y, and the scale and the orientation. The configuration space has to be discretized into bins to be able to accumulate votes. If a configuration should get a vote, the vote is added to the single corresponding bin. The width of the bins (or the "grid resolution") can be chosen independently for each dimension and is crucial for the quality of the final result. If the grid is too fine, votes might spread across too many neighboring bins without a clear maximum. If on the other hand the grid is too coarse, the accuracy of the detection will be bad. For the spatial dimensions, it makes sense to set the bin size with respect to the object's dimensions, for the orientation, a coarse binning of 15°-20° is a good starting point. To find a useful grid size for the scale dimension, you can check the object's dimensions in the object and the test images. Since we want to detect multiple instances of the same object, we cannot simply consider only the maximum value in the voting space! Instead. After the configurations of the objects have been found, they should be indicated by bounding boxes. The extra step is to draw a rectangle in the correct position in the image for better noted result. To display the rectangles at the correct position in the image, the coordinates first have to be transformed to the image coordinate system, according to the detected translation, scale and rotation of the object.
A combination of SIFT algorithm and RANSAC algorithm for extracting and accurately matched points. Firstly, feature points are extracted using SIFT that includes four main steps.

Step-1: Extreme detection of scale space where the algorithm choose extreme points of signal space as candidate feature points. The image $I(x, y)$ scale space can be defined as mentioned in equation (1), again, the extreme in the result of difference of Gaussian is detected and image convolution is performed as

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$  \hspace{1cm} (12)

where the symbol $*$ represent the 2-D convolution operation.

Step-2: Key point localization is done by Taylor expansion in signal space of difference of Gaussian function construction:

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X$$  \hspace{1cm} (13)

By solving this formula we get:

$$\hat{X} = -\frac{\partial D^T}{\partial X} \left( \frac{\partial^2 D}{\partial X^2} \right)^{-1}$$  \hspace{1cm} (14)

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X} X$$  \hspace{1cm} (15)

Step-3: Assignment of orientation :for each feature point a main direction is assigned which contains gradient magnitude $m(x,y)$ and gradient direction $\theta(x,y)$ Where:

$$m(x, y) = [(L(x,y+1) - L(x,y-1))^2 + (L(x+1,y) - L(x-1,y))^2]^{1/2}$$  \hspace{1cm} (16)

$$\theta(x, y) = \arctan \frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}$$  \hspace{1cm} (17)

Step-4: Finding the key points descriptors : The image region surrounding the critical point is divided into blocks and the gradient histogram in each block is calculated in the end a unique 128 dimensional vector is generated. After that the main role for RANSAC algorithm is needed to remove the mismatches knowing that RANSAC algorithm follow the steps:

Step-1: The proportion of correct points that have true matches has been increased.

Step-2: A transformation function has been determined according to the projection transformation between the two images.

Step-3: Finalizing the final model regarding the RANSAC steps that mentioned earlier.

IV. RESULT AND DISCUSSION

MATLAB version R2017a has been used as software tool. One laptop with core i5 processor (2.7 GHz) and (8GB) memory RAM and Intel (R) HD graphics (4160MB total memory), the images are captured using mobile camera (1 Megapixel) and resolution of (1280x720).

Fig.5 and Fig.6 are the candidate images that our work is done on and we referred to them as test image and template image respectively. Fig.7.a and Fig.7.b are the results of applying SIFT algorithm on both of the images which give all keypoints regarding to this algorithm.
Enhanced Object Detection using Generalized Hough Transform and RANSAC

Fig. 7.a and Fig. 7.b are the results of applying SIFT algorithm on both of the images which give all keypoints regarding to this algorithm. It is realized that from Fig. 10 that RANSAC algorithm can’t dismiss all the mismatches and few outliers remains ,but in fact this algorithm is voting procedure dependent and its result depend only on the number of inliers detected ,the final projection homography result is shown in Fig.11. When the edge threshold value is big we get more number of matches and less residual squared distance error(RMS). Table 1 show experimental results of RANSAC.

**TABLE 1. RANSAC parameters with different value of thresholding**

| Threshold | Number of features in test image | Number of features in template image | Number of matched keypoints | Best number of inliers | Number of iterations needed | RMS error | Correct matching rate |
|-----------|---------------------------------|-------------------------------------|-----------------------------|------------------------|---------------------------|-----------|---------------------|
| 0.5       | 1136                            | 111                                 | 111                         | 72                     | 25                         | 1.8482    | 0.64               |
| 1.5       | 1136                            | 111                                 | 110                         | 70                     | 5                          | 1.8482    | 0.63               |
| 1.9       | 1136                            | 111                                 | 74                          | 67                     | 5                          | 1.0684    | 0.90               |

The peak threshold and the edge threshold values from SIFT algorithm has been chosen to limit the number of features detected to make the calculation complexity less and was taken as : peak_threshold=2, edge_threshold=15. From this table it is realized that when we increase the matched keypoints threshold we get less number of iterations and more correct inliers regarding to the total number of matches. It is clear also that correct matching ratio is also better where we got 90% ratio when we increased the threshold to the value 1.9.

One important term to show the quality of performance of RANSAC algorithm is known as root mean-squared error (RMS) which is calculated from this formula:

$$\text{RMS} = \sqrt{\frac{\sum_{k}^{N} (X_k - X'_k)^2}{N}}$$

(18)

Where N represents number of inliers , the numerator represent the residual squared distance errors for all the points.

The obtained homography matrix is:

$$\text{Homography \ matrix} = \begin{bmatrix} 0.2914 & -1.2203 & 642.5182 \\ 1.5123 & -0.0013 & -870.7510 \\ 0.0006 & 0.0000 & 0.6231 \end{bmatrix}$$

It is clear that GHT algorithm has discarded all matches that don’t belong to the same line alignment , thus the object is being determined accurately with bounding box , the matches are shown in Fig.8. The detected object is surrounded with a rectangular box depending on GHT algorithm and Hough transform feature Fig.9.

Fig. 5 and Fig. 6 are the candidate images that our work is done on and we referred to them as test image and template image respectively.
V. CONCLUSION

Thus an object detection has been done accurately using the two proposed models. In first proposed model, object detection technique SIFT is followed by GHT algorithm and in the second model the GHT is replaced by RANSAC algorithm. RANSAC parameters with different values of thresholding has been obtained. Among the two proposed models, model one supposed to be superior technique for detecting objects purpose.

REFERENCES

1. Suga, K. Fukuda, T. Takiguchi, and Y. Arikii, “Object recognition and segmentation using sift and graph cuts,” in 19th International Conference on Pattern Recognition, Dec. 2008, pp. 1–4.
2. Q Fan, K. Barnard, A. Anur, A. Efrat, and M. Lin, Matching slides to presentation videos using sift and scene background matching,” in MIR ’06: Proceedings of the 8th ACM international workshop on Multimedia information retrieval. New York, NY, USA: ACM, 2006, pp. 239–248.
3. J. Chen and Y.-S. Moon, “Using sift features in palmprint authentication,” 19th International Conference on Pattern Recognition. ICPR 2008, Dec. 2008, pp. 1–4.
4. F. Liu and M. Gleicher, “Region enhanced scale-invariant saliency detection,” IEEE International Conference on Multimedia and Expo, vol. 0, pp. 1477–1480, 2006.
5. Y. Ke and R. Sukthankar, “Pca-sift: a more distinctive representation for local image descriptors,” in 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2004, pp. 06–11.
6. Y. Cui, N. Hasler, T. Thomählen, and H.-P. Seidel, “Scale invariant feature transform with irregular orientation histogram binning,” in ICIAR 09: Proceedings of the International Conference on Image Analysis and Recognition. Springer, 2009.
7. D.H. Ballard “Generalizing the Hough Transform to detect arbitrary shapes”, Pattern Recognition, Vol. 13, No. 2, pp. 111–122, 1981.
8. J. Illingworth and J. Kittler “A Suvey of the Hough Transform” CVGIP 44, pp 87–116, 1988.
9. P. Hough. Method and means for recognizing complex patterns. U.S Patent 3,069,654, 1962.
10. Kimura and T. Watanabe. An extension of the generalized hough transform to realize affine-invariant two-dimensional (2d) shape detection. In 16th International Conference on Pattern Recognition ICPR, volume 1, page 10065, 2002.
11. N. Kryzi, Y. Eldar, and A. Bruckstein. A probabilistic hough transform. Pattern Recognition, 24(4):303–316, 1991.
12. P. Kultaken, L. Xu, and E. Oja. A new curve detection method: randomized hough transform. Pattern Recognition Lett, 11:331–338, 2002.
13. H. Li, M. A. Lavin, and R. J. L. Master. Fast hough transform: A hierarchical approach. Computer Vision, Graphics, and Image Processing, 36(2-3):139–161, 1986.
14. M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” Comm. Assoc. Comp. Mach, vol. 24, no. 6, pp. 381–395, 1981.
15. O. Chum and J. Matas, “Matching with PROSAC - progressive sample consensus,” in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, San Diego, CA, volume 1, June 2005, pp. 220–226.
16. D. Nister, “Preemptive RANSAC for live structure and motion estimation,” in Proc. 9th Intl. Conf. on Computer Vision, Nice, France, volume 1, October 2003, pp. 199–206.
17. P. H. S. Torr and D. W. Murray, “The development and comparison of robust methods for estimating the fundamental matrix,” International J. of Computer Vision, vol. 24, no. 3, pp. 271–300, 1997.
18. P. H. S. Torr and A. Zisserman, “Robust parameterization and computation of the trifocal tensor,” Image and Vision Computing, vol. 15, pp. 591–605, August 1997.
19. P. H. S. Torr and A. Zisserman, “MLESAC: A new robust estimator with application to estimating image geometry,” Computer Vision and Image Understanding, vol. 78, pp. 138–156, 2000.