Object Detection Based on Minimum Convex Hull and Generalized Hough Transform

Tong Zhang¹, Chenfei Song²*
¹,²Guilin University of Electronic Technology, Guilin, China
*Corresponding author e-mail:vector6@foxmail.com

Abstract. In view of the low efficiency and fallibility of the feature-based matching algorithm in object detection, a new algorithm based on convex hull region and generalized hough transform is proposed. By using the region of interest to narrow the detection region and applying the key feature points to the voting elements of the generalized hough transform, the computation is greatly reduced. Firstly, the fused key feature points are extracted and filtered to generate the convex hull region which is combined with the GBR clustering algorithm to determine the region of interest. Secondly, the key feature points are taken as the voting elements in the generalized hough transform algorithm to obtain the final result. In this paper, dozens of various transformation experiments (scaling, rotation, noise, etc.) based on the proposed algorithm are compared with other similar methods, and it is verified that the method has higher efficiency and accuracy according to the theoretical basis and concrete implementation, which can meet the real-time and high-precision requirements in industrial detection.

1. Introduction
Object detection is an important research topic in the field of machine vision. The purpose is to detect the existence of the object and determine the accurate position of the object according to the object template in an image. Hough Transform (HT), as a global algorithm, detects lines from binary images. Based on global features, the Generalized Hough Transform (GHT) [1] can detect target objects of any shape from images, which has attracted much attention in object detection because of its good anti-interference and noise ability. [2] However, when there is rotation and scaling transformations in the image to be detected, the computational time and memory of the algorithm will increase exponentially. This is the main difficulty encountered by the generalized hough transform in the research and application of object detection.
Therefore, many researches have been made to improve the low efficiency of the GHT algorithm. Ping Fu Fung proposed the Random Generalized hough transform (RGHT)[3], which improved the calculation speed, but its robustness and accuracy decreased significantly when noise pollution or distortion existed in the image. Hu fangming proposed Fuzzy Random GHT (FRGHT)[4], which combines the Fuzzy reasoning system with the Random generalized hough transform. However, the time efficiency of the algorithm will be significantly reduced when the irregular edge proportion of the detected object is too large as well as there is rotation transformation. Thomas and kassim used the displacement vector (DV) feature as the search index and proposed that DV-GHT reduced the dimension of hough space by one dimension. However, the algorithm can not clearly determine the rotation Angle and scaling factor of the target object.

In recent years, the target detection method based on Region of Interest (ROI) has been widely applied. How to select the Region of Interest effectively is the key to its success. Liu chengyun proposed the method of extracting ROI of traffic signs in natural scenes through color enhancement technology, and extracting gradient histogram and multi-channel local binary mode feature to identify traffic signs [5]. The FT (Frenquency-Tuned salient region detection) algorithm proposed by Achanta calculates the difference between the average color of each pixel in the image and the whole image to calculate the significance [6]. Xue feng proposed an image ROI selection method based on SURF feature extraction contribution matrix.[7]Although these method are simple, the significance difference between the significant region and the background is small.

Aiming at above problems, this paper proposes an improved detection algorithm to improve the efficiency of generalized hough transform from two aspects. Firstly, the shi-tomasi and FAST feature points are extracted, and the filtered fusion feature points are used to generate the convex hull region, which is combined with the superpixel clustering to determine the region of interest. Secondly, the key feature points are used as the voting elements of the generalized hough transform to reduce the pixels that need to be voted in the target image. The experimental results show that the proposed algorithm can quickly and accurately locate the region of interest of the target object, also it has a certain inhibitory effect on noise interference in the complex background. In addition, it has a good effect on both the local and global feature under the conditions of scaling and rotation, and has been applied to the actual machine vision detection system.

2. Region of interest detection

2.1. Key feature points detection

Feature detection points are an important part of target recognition[8]. As common local feature, key feature points (interest points and corner points) can be described as follows: The pixel point corresponding to the local maximum of the first derivative; The intersection of two or more edges; Points in the image where the gradient value and gradient direction change at high rates; Where the first derivative is the largest and the second derivative is zero, indicating the discontinuous direction of the edge change of the object.

In this paper, the algorithm extracts the shi-tomasi corner points and FAST feature points and filter them as key feature points.
2.1.1. Shi-Tomasi corner points detection. The basic idea of the algorithm is to use a window to slide in any direction of the image, and compare the change degree of the pixel gray in the window before and after the slide. When the window moves \([u,v]\), the change of pixel gray in the corresponding window before and after sliding is described as:

\[
E(u,v) = \sum_{x,y} w(x,y)[I(x+u,y+v) - I(x,y)]^2
\]

(1)

Where \((x,y)\) is the corresponding pixel coordinates in the window, and \(w(x,y)\) is the window function. We use binary gaussian function as the initial window function, and use Taylor formula to expand. By \(I(x+u,y+v) = I(x,y) + I_x u + I_y v + o(u^2,v^2)\), it can be inferred:

\[
M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}
\]

(2)

Calculate the two eigenvalues of the autocorrelation function \(M\). When the smaller of the two eigenvalues is greater than the threshold value \(T\), the feature point is retained. In this paper, the feature point detection is based on the shi-tomasi corner point detection operator and color feature are added, which can better eliminate the interference factors in the background and better locate the region of interest. The enhanced corner point detection algorithm obtains the color enhancement matrix \(M\) from the input image, and then the image is color enhanced through the \(M\)-matrix to calculate the enhanced shi-tomasi corner point of the image.

2.1.2. FAST feature point detection. Since the shi-tomasi algorithm may detect too few feature points for some objects with a large proportion of straight lines, the FAST algorithm is further used to detect feature points based on the above initial detection. The algorithm process is as follows:

For a point \(p\) in the image, take \(p\) as the center, the radius of 3 on the round of 16 pixels point, calculate the pixel difference of \(p_1, p_5, p_9, p_{13}\) points in four directions, if there are at least three points exceeds the threshold, \(p\) is selected as a candidate, then calculate the pixel difference between \(p\) and \(p\), if there are at least nine consecutive points exceed the threshold, we consider \(p\) as the expected feature point; After the first step, Non-maximum suppression algorithm is used to remove locally dense feature points. It calculate the response of feature points, the feature points with larger response values are kept and the remaining feature points are deleted.

2.1.3. Feature points fusion. In order to ensure that the final key points can not only describe the shape of the target, but also not be too concentrated, this paper adopts the following methods to extract and filter the fusion feature points.

Firstly, rasterize the detection image, define the upper limit \(T\) of feature points in the grid, extract feature points in each grid, prevent feature points from being too concentrated, and extract the color-boosted shi-tomasi corner points for the image. With the addition of color feature, interference factors in the background can be eliminated better. Calculate the enhanced shi-tomasi corner points of the image. Determine whether to extract FAST feature points based on the number of feature points compared to a given threshold. And define the significant value of two kinds of feature points:

Significant value of shi-tomasi feature point:

\[
S_{shi} = \min(\lambda_1,\lambda_2) \cdot \text{Threshold}
\]

(4)
Significant value of FAST feature point:

$$S_{\text{FAST}} = \sum_{i=1}^{16} (p_i - p)$$  \hspace{1cm} (5)$$

The extracted feature points were sorted according to the significance value to filter the most obvious N feature points.

2.2. Convex hull region generation

Get rid of the feature points within 5 pixels of the image boundary, other residual feature point is used for the minimum convex hull algorithm to get the ROI. The fusion feature can effectively reduce the number of invalid feature points and interference points and make up the shortcoming of the minimum convex hull algorithm that the calculation time is too long due to too many feature points, and gives full play to the advantages of the convex hull region.

The minimum convex hull generation algorithm is as follows:

a) Pick the point with the smallest x-coordinate in figure 1, if there are more than one point, pick the point with the smallest y-coordinate in the points.

b) Scan the other points from bottom to top, use the exterior product of the vector to Represents the position relationship between vectors.

$$|c| = |a| \times |b| \times \sin \langle a, b \rangle$$  \hspace{1cm} (6)$$

If $|c|>0$, Vector a is in the clockwise direction of vector b, so vector a is just below vector b.

If $|c|=0$, as two vectors are collinear, the position relation is determined by the order of the distance from the origin.

c) The graham scan is performed in the order of the vertexes obtained in step b. When scanning to left turn($p_0, p_4, p_5$), keep the point($p_4$); When scanning to right turn($p_8, p_6, p_3$), delete the point($p_6$). The result is the set of vertexes of the convex hull.

2.3. Superpixel Segmentation

Superpixels can represent the pixels of an image as a whole. A series of pixels with similar distance and feature such as color and brightness are merged into one superpixel. The superpixel can not only preserve the boundary of the image, but also make the subsequent operation of the image easier. In this paper, Simple linear iterative clustering (SLIC)[9] algorithm is used to segment the image into superpixels. The SLIC algorithm is simple in structure and requires few parameters, which can
effectively divide the image into image blocks with different sizes and shapes for specific purposes. Figure 2 is the effect of image segmentation using SLIC algorithm. It can be seen that the image is divided into a number of irregular representative regions.

The 3d normal distribution of color fitting of all pixel points was used as the representation of superpixel feature, and the feature of each superpixel were expressed as:

\[ S_i = (N(m, n)nc(x, y)) \] (7)

\( N(m, n) \) is the three-dimensional normal distribution fitted by all pixel points in the superpixel, \( m \) is the vector composed of the average value of each dimension in the normal distribution, \( n \) is the covariance matrix, \( c(x,y) \) is the spatial center of the superpixel, and the distance between different superpixels is calculated by using Wasserstein distance on L2.

\[ \text{tr} \left( \frac{1}{2} \ln \left( \frac{m_i}{m_j} \right) \right) = W_{ij} \] (8)

Where \( \text{tr} \) represents the trace of a matrix.

![Figure 2. Superpixel based on SLIC](image)

2.4. Super pixel clustering and ROI determination

Based on the extraction of superpixels, this paper adopts GBR Clustering algorithm to classify superpixels[10]. In the algorithm, Semi-positive definite relaxation is used to solve the normalized cuts.

\[ \min y^\top Ly \quad \text{s.t.} \quad Ay = \varphi, \quad \varphi \neq 0 \] (9)

Where \( e^\top = (1, 1, \ldots, 1) \), \( A = e^\top e \), \( L \) is the Laplace matrix, the calculation method is as follows:

\[ D(i, i) = \sum_j W(i, j) \] (10)

\[ W(i, j) = \exp \left( -\frac{\delta(v_i - v_j)^2}{2\sigma^2} \right) \quad \sigma > 0 \] (11)

\[ L = D - W \] (12)

Where \( W \) is the similarity matrix of each vertex in \( G=(V,E) \); \( D \) is the diagonal matrix.

Through Lagrange multiplier \( \lambda \), take the partial derivatives of \( A \) and \( B \) respectively to get the clustering result:

\[ y^* = L^{-1}A^\top(AL^{-1}A^\top)^{-1}\varphi \] (13)

Combining the results of super pixel classification and the convex hull region, the final probability of the region of interest is:

\[ p_m = \frac{|n \cap h|}{|n|} \] (14)
In the above formula: $|n|$ represents the number of pixels within a class in a cluster, $|n \cap h|$ represents the number of pixels in both the convex hull region and the class.

Figure 3. Final result of region of interest

3. **Object detection algorithm based on generalized hough transform**

Based on the improved algorithm proposed by Thomas and Kassim [11], this paper takes the displacement vector feature as the index of the look-up table. The DV-GHT algorithm and the algorithm in this paper are integrated, and the image multi-resolution analysis is combined. As mentioned above, the approximate location of the target object is first determined to determine the ROI region, and then the fusion feature is adopted as the voting element to further reduce the amount of voting computation.

Based on the problems cited above, this paper uses the region of interest to narrow the search space, and uses the fusion feature points to reduce the voting elements. Combining above methods, the improved generalized hough transform algorithm based on key feature points is proposed. For the image to be detected, the region of interest of the image is obtained by clustering algorithm and convex hull region, and the fused key feature points are detected by filtering Shi-tomasi corner points and FAST feature points.

In the object accurate detection stage, the feature points in the region of interest are taken as the voting elements of the target object. Since the determination of the region of interest greatly reduces the influence of the image edge region and complex background, and reduces the edge points $n_t$ of the image to be detected, the time complexity of the detection algorithm is reduced theoretically. The specific algorithm flow is as follows:

1. The image to be detected is segmented into super pixel regions by SLIC algorithm.
2. Extract the enhanced Shi-tomasi corner points[12,13] and FAST feature of the image. Considering that the feature points extracted based on FAST algorithm do not have scale invariance, the image was down-sampled to generate pyramid images and feature points were extracted from multi-layer pyramid images. The image to be detected was divided into raster based on image size, the upper limit $K$ of feature points was defined (obtained through multiple experiments, $K$ was related to image size and feature point extraction accuracy, generally $(0.005-0.02) *$ image pixel number $n_m$), the two feature points were sorted according to significance, and the $N$ optimal feature points were filtered.
3. After removing the feature points that are too close to the edge of the image, the minimum convex hull of the remaining feature points is calculated according to the minimum convex hull generation algorithm. The region of interest is obtained by combining the superpixel classification region with the convex hull region.
(4) Detection of target object position based on GHT algorithm. For template figure, after downsampling by 0.5, the R-table is created, for the image to be detected, take the filtered feature points as vote elements of generalized hough transform, because the image data sets is not big, there is two layers of image pyramid generated, the detection start from the top, then get the zoom factor and the rotation Angle estimation which is the input of the next layer, as a result the hough space is reduced further, and time complexity is cut.

**Figure 4.** Object detection flowchart based on convex hull and generalized hough transform

### 4. Experimental results and analysis

In order to verify the correctness and validity of this algorithm, on the computer (CPU 2.3 Ghz i5), multiple sets of experiment is conducted which contains a variety of image transformation or interference like rotation, scaling, multiobject interference. In the experiments, the key feature points threshold \( K = (0.005-0.015) \times n_m \), scale \( s = 0.5-2 \), rotation \( \theta = 0 \sim 180^\circ \), a variety of position as shown in figure 5 for some of experiment results. Design the method for calculating the matching error:

\[
E = \sqrt{(x' - x_c)^2 + (y' - y_c)^2}
\]  

(15)

Where \((x', y')\) is the ideal coordinate of the searched target center in the image to be detected.

Compared with GHT, RGHT, FRGHT and DV-GHT, since there are many experimental variables and a large number of experiments, this paper only lists the mean computing time and mean matching error of each group of experiments as shown in table 1.
Table 1. The experimental results compared with RGHT, FRGHT and DV-GHT

| Scaling S | Rotation θ | Method        | Matching error E | Computing time T/s |
|-----------|------------|---------------|------------------|-------------------|
| 1.0       | 30         | RGHT          | 6.2136           | 3.0365            |
|           |            | FRGHT         | 2.0361           | 3.5880            |
|           |            | DV-GHT        | 2.8760           | 2.2842            |
|           |            | Proposed algorithm | 2.6164       | 0.8496            |
| 0.5       | -45        | RGHT          | 6.9687           | 3.8130            |
|           |            | FRGHT         | 2.6218           | 3.8617            |
|           |            | DV-GHT        | 3.1926           | 2.3849            |
|           |            | Proposed algorithm | 2.9760       | 0.8762            |
| 1.25      | 120        | RGHT          | 7.2135           | 3.8671            |
|           |            | FRGHT         | 2.6165           | 3.9648            |
|           |            | DV-GHT        | 2.9413           | 2.674             |
|           |            | Proposed algorithm | 2.9075       | 0.8749            |

(a)template image           (b)target image               (c) result

Figure 5. Some experiment results of the algorithms

As mentioned above, the time complexity of the classical generalized hough transform is \(n_s/R_s)n_s\theta_r\). The proposed algorithm determines the region of interest quickly to narrow the vote space, greatly reduce the influence of image edge region and the complex background. Besides, the importance sampling matching algorithm based template weaken the multiobject interference, taking the key feature points as the vote elements instead of the object contour edge reduce the vote computation \(n_s\) further. Due to the Combination with multi-resolution analysis, the time complexity of detection algorithm is reduced theoretically, and the influence of background and noise is greatly reduced. By the experimental results in table 1, figure 5 and described above, The proposed improved generalized hough transform algorithm based on feature points speed up computation and keep the accuracy and the stability compared with the former detection algorithm (RGHT, FGHT, DV-GHT), In addition, the proposed algorithm keep the high efficiency in rotation, scaling, multiobjects or complex background interference.
5. Conclusion

In this paper, a new object detection algorithm combining region of interest and generalized hough transform base on key feature points is proposed. The theoretical basis and realization of the method are discussed in detail. Through the above experiments, the algorithm has the advantages of small computation, high accuracy, and good efficiency under the influence of rotation, scaling, multi-object, etc. It has been proved that it can better meet the requirements of high precision and real-time in industrial detection.

6. Acknowledgments

This work was supported by The National Natural Science Foundation of China (61661017, 61967005, U1501252), the National Natural Science Foundation of Guangxi Province, China (No.2017GXNSFBA198212), the China Postdoctoral Science Foundation funded project (No.2016M602923XB), the Key Laboratory of Cognitive Radio and Information Processing Ministry of Education (CRKL190107, CRKL160104).

References

[1] BALLARD D H, Generalized Hough Transform to detect arbitrary shapes. Pattern Recognition, 1981, 13(2):111-122.
[2] Wu peng, wang yue, cui ya-ping. Insulator identification method based on generalized hough transform and complex network community. the 37th China control conference.2018.
[3] FUNG P-F, LEE W-S, KNG I. Randomized generalized Hough transform for 2-D grayscale object detection. ICPR96:Proceeding of the 13th International Conference on Pattern Recognition. 1996, 2:511-515.
[4] Hu fangming, peng guohua. Target matching based on fuzzy random generalized hough transform. Computer application, 2010, 9081(2010)11-035.
[5] Liu chengyun, chang liangliang. Detection of traffic signs based on the region of interest and hog-mblbp characteristics. Journal of electronics and information technology, 2016, 38(5):1092-1098.
[6] ACHANTA R, HEMAMI S, ESTRADA F. Frequency-tuned salient region detection. CVPR, 2009:1597-1604.
[7] Xue feng, gu jing, cui guoying et al. Image ROI selection and retrieval method based on SURF feature contribution matrix. Journal of computer-aided design and graphics, 2015, 27(7):1271-1277.
[8] Tran A , Manzanera A . Mixing Hough and Color Histogram Models for Accurate Real-Time Object Tracking. International Conference on Computer Analysis of Images and Patterns. 2017.
[9] Achanta R, Shaji A, Smith K, et al.SLIC superpixel compared to state-of-the-art superpixel methods. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2012, 34(11):2274-2282.
[10] Lee C H , Zaiane O R , Park H H , et al. Clustering high dimensional data: A graph-based relaxed optimization approach. Information Sciences, 2008, 178(23):4501-4511.
[11] Thomas A D H. Compressing the parameter Space of the Generalized Hough Transform. Pattern Recognition Letters, 1993, 13(2):107-112.
[12] Zhao qian, cao jialin, hu yueli. Significance region detection combined with gaussian multi-scale transformation and color complexity calculation. Journal of instrument and instrumentation, 2012, 33(2):405-412.
[13] Van de Weijer J, Gevers T. Boosting color saliency in image feature detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2006, 28(1):150-156.