

# An Approach of Animal Detection Based on Generalized Hough Transform\*

Weimeng Chu<sup>1</sup>, Fang Liu<sup>2</sup>

<sup>1</sup>State Key Laboratory of Robotics and Systems, Harbin Institute of Technology, Harbin, P.R.China

<sup>2</sup>Harbin University of Commerce, Harbin, P.R.China  
chumengmeng.123@163.com

**Abstract** - To detect animal objects under complicated background, a new approach is proposed to detect animal objects using shape feature. At first, a model is built by devising Canny hierarchical structure, then candidate fragments of the back and buttock are extracted using Generalizing Hough Transform, at last the matching task is accomplished based on restricting relationship of relative positions between the back and buttock. The experiment result shows that our approach achieves a promising detection rate under posture variations, scale variations and background clutter.

**Index Terms** - Image Processing; Object Detection; Generalized Hough Transform.

## 1. Introduction

Object detection using shape feature has achieved a great progress in computer vision. Shape feature has high discrimination and recognition accuracy, recently the detection of the simple shape has made great achievements, but it is still difficult to detect animal objects under complicated background. Animal possess highly flexibility, so the extracted fragments can vary also dramatically, as result, it's difficult to detect animals in image.

The commonly used approaches of animal detection are mainly divided into two categories, machine learning methods and template matching methods. Machine learning methods have excellent robustness on affection of many respects, such as background clutter, illumination variation, obscuring, etc. Boosting algorithm is widely used in machine learning methods which update each weak classifier to achieve a strong classifier that has the highest recognition rate [1]. Shotton et al. combined boosting algorithm and Chamfer matching algorithm [2] to detect horses by using a two-stage partially supervised learning approach to train simple star structures that captured from contour fragments feature in a training set [3]. Ferrari et al. accomplished the horses detection by combining the shape fragment descriptor and support vector machine [4]. Machine learning methods require huge amounts of training templates and have lower recognition accuracy, however, template matching methods don't. Belongie et al. proposed a shape context descriptor based on a histogram captured from the relation of distance and angle among points, and have an excellent matching under scaling and translation but subject to background clutter and limited to detect animal [5, 6, 7, 8]. Depending on holistic theories of shape, Toshev et

al. presented a novel shape descriptor to describe object holistic shape feature [9, 10], the above shape model achieve a better matching in INRIA Horses Database but requiring various templates and computationally expensive.

The Hough Transform is an approach of object shape detection by voting, conventional Hough Transform was merely used to detect straight lines, and Ballard extended the theory as Generalizing Hough Transform (GHT) that can detect to arbitrary shape after improving [11]. The GHT construct an R table that associate every edge point of the template with the referenced point has been selected. In the shape detection, calculate the potential referenced points with all edge points of object image according to every vector in the R table. Meanwhile, use accumulator to make a statistic, and the shape detection finish.

We propose a new approach of animal shape fragments, and solve the problem of Canny operator threshold selection by applying Canny hierarchical structure, according to shape deformation of back and buttock of animal isn't obvious, we accurately detect the horses using INRIA Horses Database and the result is desirable and acceptable.

## 2. Algorithm Process

The algorithm process is shown in Fig. 1: At first, we extract edge from object image in preprocessing stage and capture the shape fragments. Extract the shape templates of the back and buttock, and vote for shape fragments obtained from preprocessing stage based on GHT algorithm. The candidate reference points are chosen within the size of the voting value. The real position of corresponding template is obtained by handling the candidate reference points in object image. Finally, the optimal matching program is selected which is based on position constraint of the back and buttock.

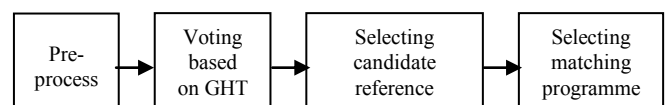


Fig. 1 Algorithm process

## 3. The Preprocessing Stage

Recently, Canny operator is one of the most commonly

\* This work is supported by the theoretical research project (No.SKLR201002D) of State Key Laboratory of Robotics and Systems(Harbin Institute of Technology).

used operators [12], and can select the optimal edge extraction by means of threshold adjustment in the edge extracting process on object images. Different numbers of shape fragments can be obtained by different thresholds. Obtained shape fragments are fewer as the threshold increasing but missing the important shape information readily, and grow with the threshold decreasing but background clutter increases. It is difficult to determine optimal threshold of object image automatically, we build the Canny hierarchical structure to overcome.

At beginning, edges are extracted with threshold from 0.9 to 0.2 by interval of -0.1. According to experience, usually the contours of the back and buttock appear entirely while the threshold value is 0.2. Subtracting the previous edge image from the next one orderly, we can achieve increased shape fragments at each reducing threshold. Set A is made up of the edge image of the threshold value 0.9 and other images obtained by decreasing order. It can be seen in Fig. 2 that, the background clutters distribute in different layers of set A by above processing. It's easier to match back and buttock against the background clusters obtained by a specific threshold. However, the whole edge curves of back and buttock have different intensities, and the hierarchical structure results in false matching due to dispersing the whole edge curves. After testing, we discover that the edge curves of back and buttock separated in different layers can form a connected domain. The incomplete shape fragments are integrated with connected domain approach in the different layers of set A, and defined as connected algorithm by connected domain approach.

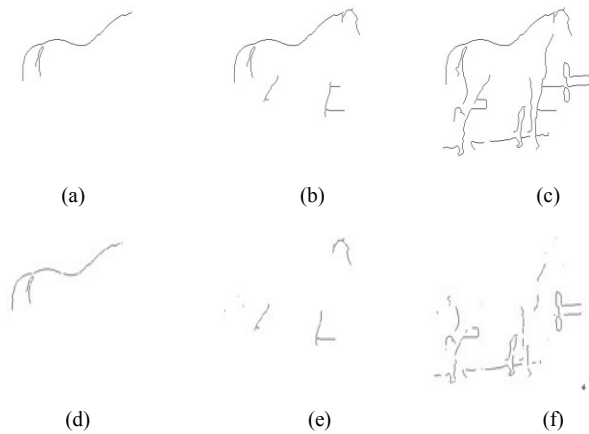


Fig. 2 Canny hierarchical image.(a)-(c) show that edge image of threshold is 0.9, 0.8, and 0.7 respectively; (d)-(e) are the part of set A, and show that edge image of threshold is 0.9, between 0.8 and 0.9, and between 0.7 and 0.8 respectively.

The connected algorithm consists of two cycles, starting from the first layer image of set A defined as “seed image”, in which obtains the curve segment endpoints as “seed points”. Obtained the connected curve segments from the second layer image merge with the “seed image”, which are subtracted from the second layer image. Later extracting the “seed points” from the new “seed image” once more, and obtained the connected

curve segments from the third layer image. The process above is repeated until the last layer of set A, as the intrinsic cycle. Then the “seed image” is changed to the second layer of the set A as the extrinsic cycle. The connected algorithm finished while the “seed image” is the last layer image of set A.

Effectiveness of the connected algorithm is shown in Fig. 3, the curve segment is more complete and the background clusters do no increase significantly.

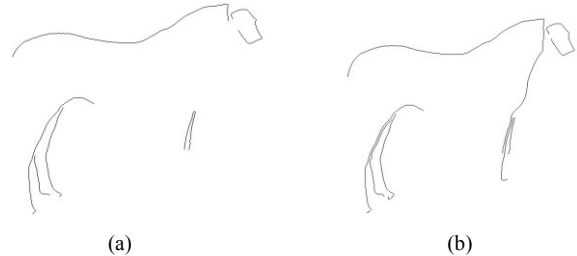


Fig. 3 Connected algorithm result.(a) is original image; (b) is processed image.

#### 4. Extracting the Candidate Reference Points

Due to the background clusters, it is difficult to extract the shape fragments accurately. Here, the hierarchical object images obtained from the pre-processing stage are matched by GHT, and extract the candidate reference points by restricting the overlap ratio between the templates and matching fragments.

##### A. Extracting Template and Building Model

Extracting templates process is shown in Fig. 4. Appropriate original image is selected to extract templates, and defining the fragment between the tail and neck as the back template, the fragment between tail and leg as the buttock template.

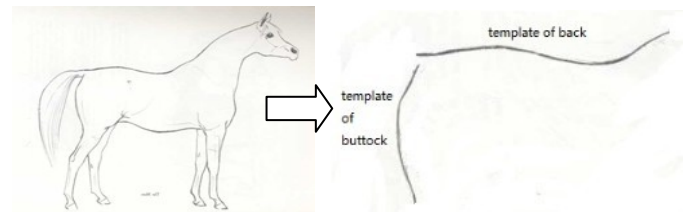


Fig. 4 The process of extracting templates

After that, we build the R table with two templates separately using GHT by selecting common reference point.

##### B. Extracting the Object Based on GHT

Typically, voting for object of arbitrary shape is executed in generalizing Hough space  $H(r,s,\theta)$ , which involves of three parameters,  $r = (x,y)$  as the coordinate of the reference point,  $s$  as the scale factor,  $\theta$  as the angle of rotation. According to slight rotation, we vote in parameter space  $H(r,s)$ .

The factor  $s$  consists of two parts, one is the image scale factor  $s_1$ , and the other is the object scale factor.  $s_1$  is the ratio of the size object image divided by template image,  $s_1$  is fixed value because the size of object image and template image is constant.  $s_2$  is the ratio of the length of object image and template image which is variable, and can limit range of  $s_2$ . After multiple testing,  $s_2$  ranges from 0.6 to 1 with interval of 0.1. Finally, the value of  $s$  is

$$S = S_1 \cdot S_2 . \quad (1)$$

The relative positions in the R table are recalculated on different scale of parameters space while the R table and parameters space are determined. Then we obtain different R tables on the range of  $s$ , vote for all layers of the set A by traversing all edge points  $E(E_{ix}, E_{iy})$  of each layer of the set A with R tables recalculated, and store the vote values in relevant scale parameters space. Calculating all the coordinates of reference points:

$$r = E - s \cdot R . \quad (2)$$

According to general voting mode, adding one per reference point in parameters space:

$$H(r, s) = H(r, s) + 1 . \quad (3)$$

However, due to effects of deformation, occlusion, background clusters of shape fragments; general voting mode has a low detection accurate rate for extracting the location of reference points. Therefore to choose the weighted specific points method for voting. Basis for GHT, an arbitrary edge point calculate with R table can generate a group of voting points which is similar to shape of template, more vote values are weighted on the red marking line segment which is shown in Fig. 5, as in

$$H(r, s) = H(r, s) + k \quad (k > 1) . \quad (4)$$

Above method is proved to extract fragments that similar to template more effectively.



Fig. 5 Location of weighted voting

### C. Selection of the Candidate Reference Points

Some fragments similar to template are obtained from a certain layer of set A by connected algorithm and weighted method, but the background clusters also have some fragments similar to template, and extract optimal location of reference point cannot be extracted from voting maximum of each layer

merely. Therefore, candidate reference points are chosen from location of front  $n$  sizes of local domain vote value. And  $n$  are 20 by experience.

To reduce computational cost, some candidate reference points are deleted by constraining overlap ratio between template and matching fragment. The actual distribution of template in object image can be obtained by combining reference points with information of R table. That edge image of threshold of Canny operator is 0.3 is selected as object image. With a view to possible deformation, object shape fragments which are obtained in actual distribute points of certain range are fetched as compensation. Defining reference point as centre, taking rays to each point of template as direction, and computing overlap ratio at the directions of rays. Existing points of object fragments at the directions of rays express overlap, and deleting unqualified fragment of overlap ratio less than 0.4.

### 5. Detection Method

The back and buttock of animal have relatively fixed position relation, and position variance is quite small, therefore, object images are detected by restricting the relative position of back and buttock.

After analysis on whole shape of horse, according to properties of GHT, the left endpoint  $E_{back}(x_{back}, y_{back})$  of back and the upper endpoint  $E_{hip}(x_{hip}, y_{hip})$  of buttock are restricted to limit whole position of back and buttock. All candidate reference points are combination of two, by which obtain the locations of two endpoints. To avoid occlusion, two endpoints are selected from the actual location of template in object image. As shown in Fig. 6, the probability  $p_{dis}$  of Gaussian density is computed with position relation between two endpoints by selecting suitable mean  $\mu$  and variance  $\sigma$ .

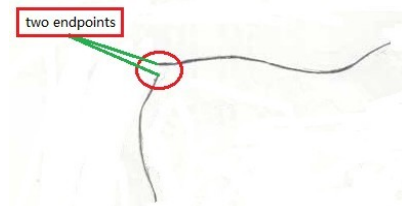


Fig. 6 The left endpoint of back and the upper endpoint of buttock

Higher  $p_{dis}$  indicate that positional relationship between the template and the back and buttocks is more similar, as in

$$p_{dis} = \phi(dis | E_{back} - E_{hip}) . \quad (5)$$

Where

$$\phi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (6)$$

The scaling proportion coefficient of back and buttock of horse is quite close in real images, so  $p_{dis}$  multiplied by scaling coefficient  $S_{dis}$  makes the probability larger while scaling proportion is close, and vice versa.

Because shape of buttock is relatively simple, as shown in Fig. 7 that multiple positioning problem appears in matching process, i.e. reference points is not same but cannot distinguish matching locations.

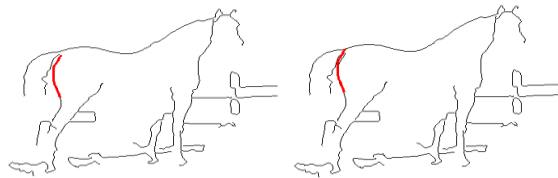


Fig. 7 Multiple positioning of buttock

We insert the buttock overlap ratio  $CR_{hip}$  obtain from section B to detected equation, and solve above problem using the value of overlap ratio. According to (7), final result is the maximum of pairs of reference points:

$$Match\_back\_hip = Max(p \cdot S_{dis} \cdot CP_{hip}) \quad (7)$$

## 6. Experiments

The algorithm is tested on the INRIA Horses Database. The INRIA Horses Database consists of 170 images of horses which include different environments, proportions, postures and the number of horses. To test our method, we detect the shape fragments combining both the back and the buttock of the object. Due to the influence of light, environment and posture, the fragments of the back and buttock of some images in the database are incomplete or complete missing, and 70 of them which meet the following conditions are chosen to match: all of the selected images have the complete shape fragments of back and buttock, same direction and only one horse in each image after detecting by Canny operator.

To compare with other methods, the PASCAL standard is used for evaluation, i.e. if the detection bounding box and the real bounding box overlap by more than 50%; it illustrates that the object is detected. Compared with the Ferrari method, the Table 1 lists the detection rates of the two methods, and illustrates that the algorithm has a better detection rate.

TABLE I Comparison of Detection Rate

Algorithm	Detection rate
Ferrari method	80.8%
Our method	88.6%

A part of detected images is shown in Fig. 8. The red area is detected result, and the green frame is the bounding box of positive image. The examples show that the algorithm is not only adaptive and robust for deformation, but also more accurate in matching and locating.



Fig. 8 Detection results

## 7. Conclusion

We propose a GHT-based animal detection method in this paper. Using horses as the detection target, a single template is employed to achieve accurate detection of horses of different postures in complex background. It solves the problem of Canny operator threshold selection by the Canny hierarchical algorithm and connected method, and it has a better result in scaling and deformation. Finally, the validity of algorithm is verified by experiment.

## 8. References

- [1] A. Torralba, K. P. Murphy, W.T. Freeman, "Sharing features: Efficient boosting procedures for multiclass object detection." CVPR04, pp.762-769, 2004.
- [2] D.M. Gavrila, "Multi-feature hierarchical template matching using distance transforms." ICPR98, pp.439-444, 1998.
- [3] J. Shotton, A. Blake, R. Cipolla, "Contour-based Learning for Object Detection[C]."Tenth IEEE International Conference on Computer Vision, vol.1, pp.503-510, 2005.
- [4] Ferrari V., Fevrier L., Jurie F., Schmid C., "Group of adjacent contour segment for object detection." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.30, no.1, pp.36-51, 2008.
- [5] S. Belongie, J. Malik, J. Puzicha, "Shape Context: A New Descriptor for Shape Matching and Object Recognition[C]." Neural Information Processing Systems Conference, 2000.
- [6] S. Belongie, J. Malik, J. Puzicha, "Shape Matching and Object Recognition Using Shape Context [J]." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, no.4, pp.509-522, 2002.
- [7] A. Thayananthan, B. Stenger, P. Torr, et al, "Shape Context and Chamfer Matching in Cluttered Scenes[J]." 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp.127-133, 2003.
- [8] G. Mori, S. Belongie, J. Malik, "Efficient Shape Matching Using Shape Context [J]." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no.11, pp.1832-1837, 2005.
- [9] Toshev A., Taskar B., Daniilidis K., "Shape-Based Object Detection via Boundary Structure Segmentation." International Journal of Compute Vision, vol.99, no.1, pp.123-146, 2012.
- [10] Toshev A., Taskar B., Daniilidis K., "Object detection via boundary structure segmentation." CVPR, pp.950-957, 2010.
- [11] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shape." Pattern Recognition, vol.13, no.2, pp.111-122, 1981.
- [12] J. Canny, "A computational approach to edge detection." PAMI, vol.8, no.6, pp.679-698, 1986.