Research article

A learning algorithm for model-based object detection

Chen Guodong
Robotics and Microsystems Center, Soochow University, Suzhou, China

Zeyang Xia
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences and the Chinese University of Hong Kong, Shenzhen, China, and

Rongchuan Sun, Zhenhua Wang and Lining Sun
Robotics and Microsystems Center, Soochow University, Suzhou, China

Abstract

Purpose – Detecting objects in images and videos is a difficult task that has challenged the field of computer vision. Most of the algorithms for object detection are sensitive to background clutter and occlusion, and cannot localize the edge of the object. An object's shape is typically the most discriminative cue for its recognition by humans. The purpose of this paper is to introduce a model-based object detection method which uses only shape-fragment features.

Design/methodology/approach – The object shape model is learned from a small set of training images and all object models are composed of shape fragments. The model of the object is in multi-scales.

Findings – The major contributions of this paper are the application of learned shape fragments-based model for object detection in complex environment and a novel two-stage object detection framework.

Originality/value – The results presented in this paper are competitive with other state-of-the-art object detection methods.

Keywords Image processing, Programming and algorithm theory, Computer applications, Object detection, Shape matching, Image segmentation, Shape fragment, Computer vision

Paper type Research paper

Introduction

Detecting and recognizing objects in images is one of the most difficult tasks in computer vision. Scale, rotation, viewpoint changes, occlusion and background clutter (Maged Marghany, 2009; Marghany and Hashim, 2010a, b) are the common challenges. Color, intensity, gradient and depth are always used as the cues for object detection and recognition. Many objects can be accurately represented by their shapes. Shape is a very powerful description of object appearance for detection methods with high precision (Ravishankar *et al.*, 2008; Marghany *et al.*, 2011; Zhu *et al.*, 2008). In the past few years, contour based methods and shape feature based methods have been proposed (Ferrari *et al.*, 2006; Shotton *et al.*, 2005). Most of the methods can be broadly classified as point-based approaches (Belongie *et al.*, 2002), boundary-curve based approaches (Opelt *et al.*, 2006) or shape-region based approaches (Wang *et al.*, 2007). However, there are still some problems in object detection and recognition.

Point-based approaches is always limited to the noises and background clutter in the images. And it is hard to localize the region around the point as the spatial neighborhood of the point. Point-based approaches using examples always involving pure shape matching and handwritten digits. Shape-based methods were not popular due to their sensitivity to segmentation errors. Leibe *et al.* (2008) has proposed novel region based methods for object detection and recognition. Object shape or curve based methods are becoming popular in recent years (Berg *et al.*, 2005; Leordeanu *et al.*, 2007). This is because shape or contour based descriptors are suitable for category-level object detection and recognition. Many categories are characterized by their shape rather than by color and texture. Although several approaches based on contour features have been proposed (Elidan *et al.*, 2006; Bouganis and Shanahan, 2008; ChengEn and Ling, 2011;

The current issue and full text archive of this journal is available at **www.emeraldinsight.com/0260-2288.htm**

Sensor Review
33/1 (2013) 25–39
 [ISSN 0260-2288]
[DOI 10.1108/02602281311294324]

The work related to the paper is partly funded by the Natural Science Fund of Higher Education of Jiangsu Province (11KJB510024) and partly funded by Natural Science Foundation of China, under contract number 61105098. The authors would like to thank the anonymous reviewers and the editors for their many helpful suggestions.

Xu *et al.*, 2009; Stark *et al.*, 2009), there are still some problems in shape modeling and shape fragment feature based object detection.

In this paper, we follow many recent detection and recognition papers in using the contour or shape parts to model the shape. In contrast to many existing methods, we focus on object detection by grouping the contour features in an image according to the model formed by the training image, the contributions of this paper are as follows. First, this paper introduces a novel shape fragment descriptor used for object detection and recognition. Second, a novel two-stage object detection framework is presented in this paper. Third, a novel structure of shape fragment codebook for object detection is used. Forth, the establishment of correspondence between multi-views based on the shape fragment features is build. Figure 1 shows the main flowchart of the proposed approach. Figure 2 is the detailed main framework of the algorithm.

This paper is organized as follows. We first discuss related work on object detection and recognition. Next, we introduce the novel shape fragment descriptor and the shape model. The detail of object detection framework is introduced in the next section. The next section presents the experimental results, which is followed by a section that concludes the paper.

Related works

Many prior works have done on object detection, such as appearance-based methods, model-based methods and so on. Appearance-based methods have become more and more popular, and are used to deal with object recognition tasks. These techniques consider the appearance feature of the object instead of the reconstruction of geometrical properties. The eigen window approach is a typical appearance-based methods. Also there are many object recognition approaches based on local features. These local image patch based descriptors including SIFT, SURF, PCA-SIFT, Harr wavelets rectangular differential feature, parts based binary orientation position histogram, orientation histogram, implicit shape model and histogram of oriented gradients. These descriptors based object recognition methods are popular in recent years. Mikolajczyk proposed to train body parts for human body detection (Mikolajczyk *et al.*, 2004), Ullman uses the combination of the trained parts for object segmentation (Ullman *et al.*, 2001), Leibe proposed to train implicit shape models of object to detect objects (Leibe *et al.*, 2004), Felzenszwalb proposed to use pictorial structures model for object detection (Felzenszwalb, 2005). However, mismatch is the main challenge when using these local descriptors based methods. 3D model-based detection techniques are sensitive to the deformation of the objects. In recent years, multi-view object class detection has received increasing attention. Most approaches present the task by extrapolating known strategies from 2D single-view object class detection, notably by combining classifiers for separate viewpoints. The main challenge in 2D/3D object class modeling by multi-views is the automatic establishment of correspondence between overlapping views. The problem becomes more challenging when the appearance of the training image is different according to the view point change. Some researchers have proposed to include geometric information into the learning process. Most of the approaches employ locally deformable 2D models for discrete viewpoints. This casts the 3D object detection problem into 2D object detection in multi-views.

In order to reduce the mismatches, there are many method combining local and global information to improve the accuracy of recognition. Li proposed to use SIFT descriptor integrating the color and global information for object detection (Li and Ma, 2009). Mortensen proposed use SIFT combining with a global context vector for the correspondence to reduce the mismatches while existing multiple similar local regions (Mortensen *et al.*, 2005). Diplaros uses the color shape context for object detection, which combines the shape and

Figure 1 The block diagram of the proposed approach

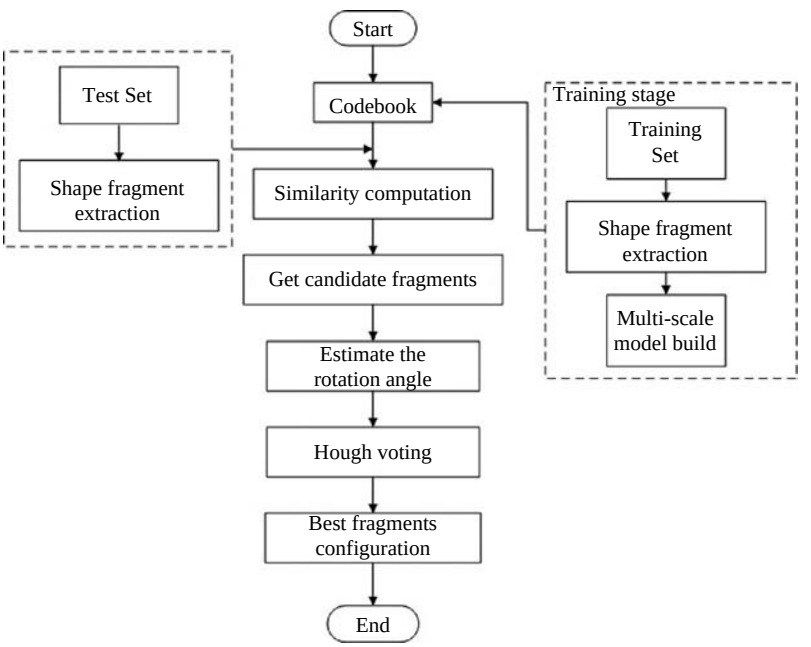
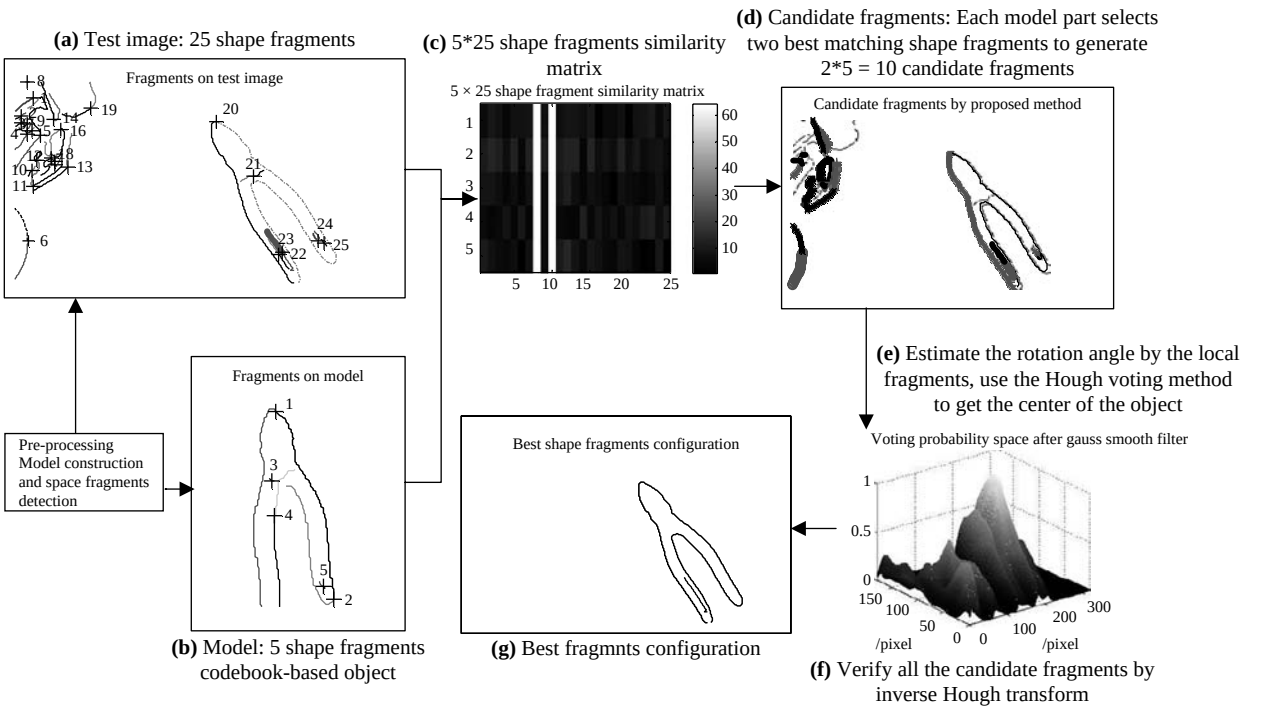


Figure 2 Illustration of the algorithm framework



color information of the image (Diplaros *et al.*, 2003). Thayananthan proposed to combining the shape context and figural continuity to improve the localization and correspondance (Thayananthan *et al.*, 2003). Similarly, Murphy proposed to combine the local and global feature for object detection and localization (Murphy *et al.*, 2006).

In recent years, contour based approaches are becoming popular. This is because many objects are characterized by their shapes rather than by color and texture.

Contour based approaches are mainly divided into two groups: point corresponding approaches (Belongie *et al.*, 2002; Berg *et al.*, 2005; Leordeanu *et al.*, 2007) and boundary segment based approaches (Elidan *et al.*, 2006).

Point corresponding approaches mainly use the property of the point spatial distribution. Belongie proposed the shape context descriptor (Belongie *et al.*, 2002), which use the distribution of all points relative to one point on the edge. Wang presented a family of using shape context which is more robust against deformation and background clutter for the object detection. Marius demonstrated a method of using the pairwise geometric relationship match for object detection and classification (Leordeanu *et al.*, 2007). Elidan used pairwise spatial relations between landmark points (Elidan *et al.*, 2006). Alexandros presented a method using relative point distribution models for object detection (Bouganis and Shanahan, 2008). These point corresponding approaches are always limited to a localized region around the point. In this way, point corresponding approaches are limited by the region size. The boundary segment based method focus on the use of local shape patches, and the local shape patch regions are relative to the length of the segment. Hence, local shape patch based methods are better to handle the scale changes.

Local contour patch based approach can be framed as a correspondence problem between shape patches in a test image and an object model. Ferrari developed an object detection method using chains of connected straight contour segments to form a local descriptor without including nearby clutters. This method can be considered as an extension of shape contexts of points. Lu presented a novel framework for contour based object detection from cluttered environments (ChengEn and Ling, 2011). Xu proposed the contour flexibility, which represents the deformable potential at each point along a contour as the shape descriptor for shape matching (Xu *et al.*, 2009). Stark presented a method using the local shape features for knowledge transfer (Stark *et al.*, 2009). Most of the above mentioned methods used for shape matching by incorporating spatial point or contour segment constraints. In order to match the shape robustly, iterative methods are always involved in these methods.

It is difficult to model the whole model of the object. Shape parts can be directly used to model the object. Most of the methods can be broadly classified as global geometric organization of shape patches or an ensemble of pairwise constraints between point features. Some researchers learned object boundary fragments as the model for object detection. This method analyzes pairs of local shape features by identifying and describing symmetries between the features. Most of the shape based methods learn the model of a specific object class using a set of training shapes. Bai used the contour-based template to detect objects in images (Bai *et al.*, 2009). The active shape model is also in this spirit (Cootes *et al.*, 1995). Thin plate spline method is popular as it is effective for modeling changes in biological forms. In order to enabling object localization in novel image using voting, for each shape patch or point, most of the methods store a pointer to the object center.

The limitation of the above methods is the need of clean training shapes. Some researchers tried to develop new

algorithms not requiring segmented training images. The key idea of this method is to limit the object by using the combinations of object features repeatedly occurring over the training images. Kushal *et al.* tried to learn the object parts model by using the local geometric consistency among the features that make up the part (Kushal *et al.*, 2007). Berg *et al.* (2005) suggested building the model by using pairs of images containing an object instance, by retaining parts matching across several image pairs.

Shape fragment features

Shape fragment detection

After getting the edge map by Canny method, the small edge gaps smaller than 10 pixels are filled by chain code method (Hajjar and Chen, 1999; Bribiesca, 1999; Ghita and Whelan, 2002; Zhijie and Hong, 2008). If the connected points satisfy equation (1), we define the connected edge points group C as a shape fragment C_E:

$$\left| |S-E| - \sum_{i}^{n} C_{vi} \right| < 0.9 \times |S-E| \tag{1}$$

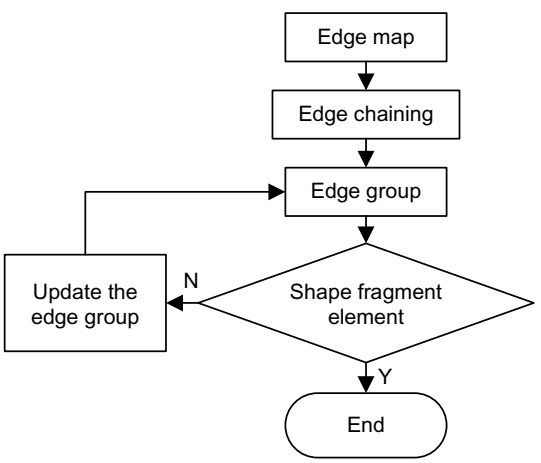
In equation (1), S is the start point of the shape fragment C_E, E is the end point of C_E, and C_{vi} is the edge point, n is the total number of the edge points, $|S - E|$ is the length of the line segment SE.

Otherwise, the connected edge points group C is divided into several shape fragments based on the principle of divide-and-conquer. The flowchart of shape fragment detection is as shown in Figure 3. The detail of the shape fragment detection algorithm is as follows. Define $L_d = |S - E|$ as the length of the line segment SE, $f(x, y)$ is the description equation of line segment SE, the furthest point to line segment SE P(X, Y) on the connected edge points group is computed by equation (2):

$$(\mathrm{X}, \mathrm{Y}) = \arg\max_{(\mathrm{x},\mathrm{y})} \left| \frac{f(x,y)}{L_d} \right| \tag{2}$$

Define point $M = (X, Y)$ as the divide point, the edge points group C is divided into two segments SM and ME. Then judge the segment SM and ME whether satisfy equation (1), if yes, SM and ME are two shape fragments, if not, repeat the above process till all the divided segments satisfy equation (1).

Figure 3 The flowchart of shape fragment detection



The connecting point between two adjacent shape fragments named the key point. Figure 4 shows a simple example of the shape fragment detection algorithm.

Shape fragment descriptor

In order to compare different shape fragments, we need to describe them. Ferrari has pointed out that T and higher orders junctions occur less frequently than simple one-to-one connections. In this paper, we classify all the shape fragment as one kind which contains one key point and two connected line segments. If the local shape patch only has one line segment, we consider this situation as two line segments coincide. If the local shape patch has more than two line segments, we only make use of the two longest line segments.

In order to make the descriptor in a repeatable manner, it is important to define a main direction for the local shape patch. By connecting two points furthest to the key point by a line, we define the slope angle of the line as the main direction. The orientation and the main direction of the shape fragment are as shown in Figure 5. In Figure 5, the original shape fragment is "12", and "1′2′" is the shape fragment rotated the main direction angle around the key point.

One shape fragment can be represented by the following parameters, $C_F = \{C_{E1}, C_{E2}, K, C_S, C_E\}$, here, C_{E1} and C_{E2} represent the edge points of the shape fragment. K is the key point, C_S is the starting point, C_E is the end point. After rotating around the key point, the shape fragment can be represented by $C'_F = \{C'_{E1}, C'_{E2}, K', C'_S, C'_E\}$. θ_1 is the slope angle of the original shape fragment. θ_2 is the slope angle of the rotated shape fragment. The main direction angle of the shape fragment is $|\theta_1 - \theta_2|$. Figure 5 shows an example of the shape fragment slope angle and the main direction.

Center point of the shape fragment can be computed by the key point, start point and the end point of the shape fragment. $G = (K + C_S + C_E)$. The distance vectors of the three points to the center point are:

$$\begin{cases} R_K = \left(R_K^x, R_K^y, |R_K|\right) \\ R_S = \left(R_S^x, R_S^y, |R_S|\right) \\ R_E = \left(R_E^x, R_E^y, |R_E|\right) \end{cases} \tag{3}$$

In which, K represents the key point, S represents the start point, E represents the end point. R_K^x is the distance vector in x-direction, R_K^y is the distance vector in y-direction. $|R_K|$ is the distance of the key point to the center point.

Above all, the shape fragment can be described by equation (4):

$$\left(\frac{R_K}{N}, \frac{R_S}{N}, \frac{R_E}{N}, \theta_1, \theta_2\right) \tag{4}$$

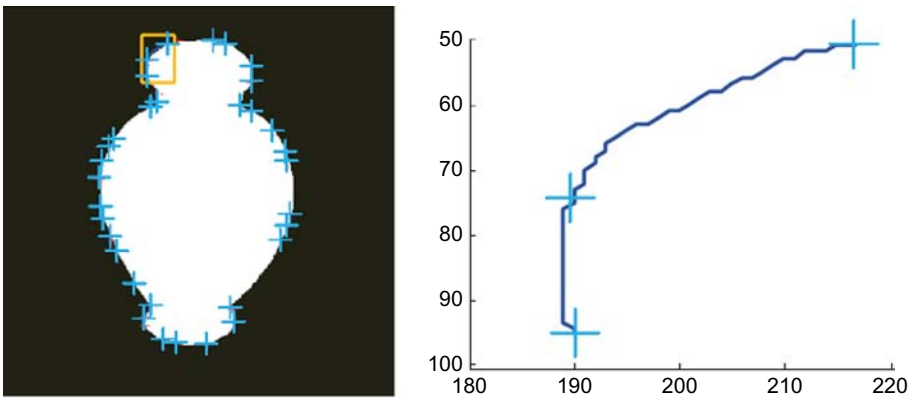
In which, $N = \max(|R_K|, |R_S|, |R_E|)$. R_K, R_S and R_E present the original shape fragments, in order to make the shape fragment descriptor is invariant to rotation, the main direction angle is needed.

The properties of the descriptor

A. Rotation invariance

As we have defined the main direction of the shape fragment, and defined the main direction of the shape fragment as the standard position, the similar shape fragments have the same main direction. Figure 6 shows the example of two similar

Figure 4 A simple example of shape fragment detection



Note: " + " means the keypoint, between two key point is a shape fragment

Figure 5 The orientation and the main direction of the shape fragment

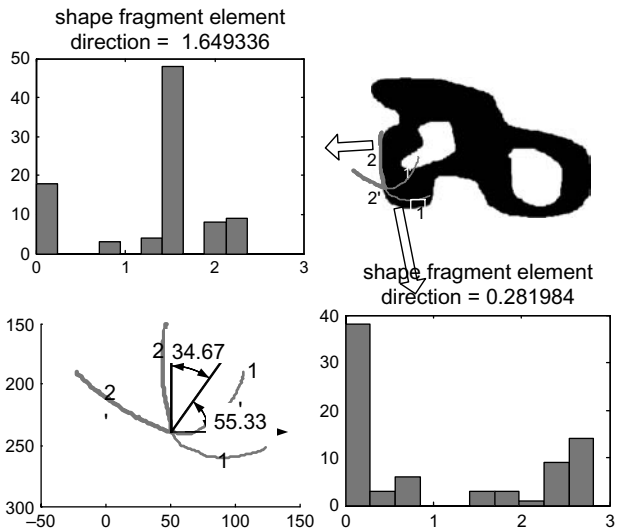


Figure 6 The similar shape fragments have the same main direction after rotation

shape fragments and the corresponding rotated shape fragments.

B. Scale invariance

As shown in equation (4), the descriptor is normalized, the descriptor has the property of scale invariance. In different scale, the similar shape fragments have the same descriptors.

C. Noise robustness

In order to test the influence of the noise to the descriptor, an experiment is done. About 0.2, 0.5, 1, 1.5, 2, 2.5 and 3 percent pepper and salt noises are added to the image, the statistic of the shape fragments main direction is as shown in Figure 7 The main direction errors at different noise levels.

From Figure 7, we can see that the main direction is affected by the noise level. But under 3 percent, the error is smaller the 0.3.

Shape fragment matching

Equation (4) is the descriptor of the shape fragment. The computation equation of the two shape fragments similarity is as shown in equation (5):

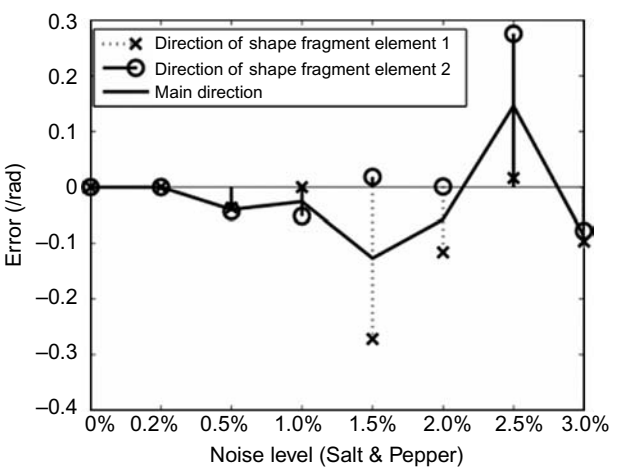
$$T_c(a,b) = D_c(a,b) + \beta\left(\left|\theta_1^a - \theta_1^b\right| + \left|\theta_2^a - \theta_2^b\right|\right) \qquad (5)$$

In which, $D_c(a, b)$ is the distance vector comparison of the two shape fragments. The similarity matrix is shown in Figures 2 and 10. In order to get the most similar shape fragments between the codebook and the object shape fragment, the similarity is sorted according to the similarity of the shape fragments:

$$D_c(a,b) = \left|\frac{R_K^a}{N^a} - \frac{R_K^b}{N^b}\right| + \left|\frac{R_S^a}{N^a} - \frac{R_S^b}{N^b}\right| + \left|\frac{R_E^a}{N^a} - \frac{R_E^b}{N^b}\right|,$$

β is the weight factor, in this paper $\beta = 4$.

Figure 7 The main direction errors at different noise levels



Shape fragment based object detection

Building the shape fragment codebook

For a rigid object, if more than three points can be matched, it is easy to estimate the object pose, then it is easy to match the whole object. But for detecting objects or classifying a certain category of objects, it is more difficult. When view point changes, the appearances of the object are different, besides, the illumination changes and the object deformation occurs, the appearances of the object are different too. As we all know, no matter what extent the differences are, one object has it is own characters. If consider the spatial configuration of objects in different conditions as the representation of the object or the category of the object, it is more flexible to detect the object of known category.

Codebook representation is one of the most popular tools for object detection and classification. There are many prior work have done by using codebook representation (Elidan *et al.*, 2006; Marr and Hildreth, 1980; Bay *et al.*, 2006). In some of the method, they use different methods to group the similar codebook vocabularies. In this paper, we introduce the method we used to build the codebook.

For each object class, we select a few images as training examples. We group all extracted shape fragments to create a codebook of prototypical shape fragments. As described in the previous section, the shape fragments should be extracted from the training set. Let CB presents the learned shape fragment codebook. $CB = \{cb_i\}$, cb_i presents one shape fragment descriptor. cb_i has seven bins, besides the five bins described in equation (4), the distance vector of each point on the shape fragment to the object center should be considered as the other two bins. cb_i is described as equation (6):

$$cb_i = \left(\frac{R_K}{N}, \frac{R_S}{N}, \frac{R_E}{N}, \theta_1, \theta_2, R_{E1}, R_{E2}\right) \tag{6}$$

In which, the first five bins are described in equation (4), R_{E1} and R_{E2} represent the distance vector of each shape fragment point to the object center.

For one point $A(x, y)$, the distance vector of $A(x, y)$ to the object center $O(x_c, y_c)$ is $R(d_x, d_y)$:

$$\begin{pmatrix} d_x \\ d_y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_c \\ y_c \end{pmatrix} \tag{7}$$

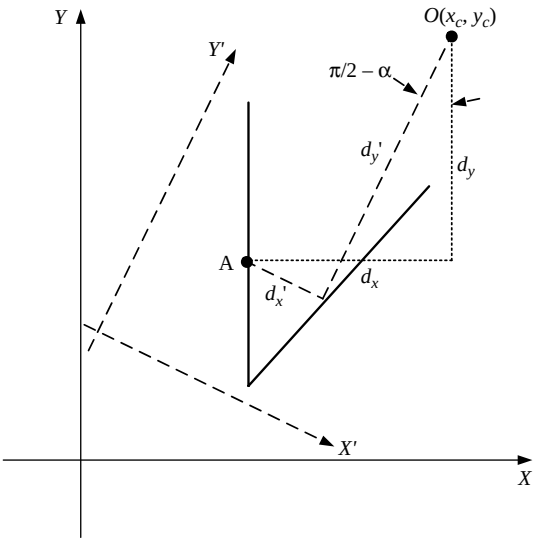
When the shape fragment rotate to the main direction, the rotation angle is $(\pi/2) - \alpha$, α is the main direction angle. When the shape fragment rotate an angle around the key point, the object center point is also rotated. The relationship of the distance vector computation between and after rotate to the main direction is shown in equation (8). Figure 8 shows the relationship illustration of the rotated shape fragments and the original shape fragments:

$$\begin{pmatrix} d'_x \\ d'_y \end{pmatrix} = \begin{pmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{pmatrix} \begin{pmatrix} d_x \\ d_y \end{pmatrix} \tag{8}$$

In which, $\gamma = (\pi/2) - \alpha$.

In order to make the proposed method invariant to the scale, multi-scale images are needed. In this way, the training model of the object is a multi-scale one. Figure 9 are some examples of the shape model composed by shape fragments.

Figure 8 The illustration of the rotated shape fragment



2D object detection

The main frame work of the algorithm is as shown in Figures 1 and 2. In this section, we introduce the detail of the algorithm. The algorithm is a two-stage algorithm. The first stage is the fragment selection stage by using the shape fragment similarity. The flowchart of the first stage is as shown in Figure 10.

The second stage is the Hough based voting stage. The distance of each shape fragment to the object center is fixed. By using the property, we use the Hough method to detect the object.

Let F_i denote the observed feature around a key point location F_i^K. C_i is the image point, $S(O, x)$ denotes the score of object O at a location x. $S(O, x)$ is computed by equation (9):

$$S(O, x) = \sum_i p(O, x, C_i) = \sum_i p(C_i)\, p(O, x|C_i) \tag{9}$$

In the Hough voting stage, the following parameters are needed: $H_v(r, s, \theta)$, where $r = (x_r, y_r)$ is a reference coordinate for the detection object, s is a scale factor, and θ is the rotation angle. The reference coordinate r is always the center of the object.

In the codebook, for each entry C_i, the relative vector to the model center is (x_r, y_r). The relationship between the reference point (x_r, y_r) and the key point $P(p_{ix}, p_{iy})$, of each entry C_i is as shown in equation (10):

$$\begin{pmatrix} x_r \\ y_r \end{pmatrix} = \begin{pmatrix} p_{ix} \\ p_{iy} \end{pmatrix} - \begin{pmatrix} R_{ix} \\ R_{iy} \end{pmatrix} \tag{10}$$

$\begin{pmatrix} R_{ix} \\ R_{iy} \end{pmatrix}$ is the distance vector of the edge point to the reference point. For each point $P_m(p_{mx}, p_{my})$ on the test image, according to the Hough transform principle, 4D accumulator array $H_v(X_r, Y_r, s, \theta)$ is needed. The possible reference point of the edge point $P_m(p_{mx}, p_{my})$ is computed by equation (11):

$$\begin{pmatrix} X_r \\ Y_r \end{pmatrix} = \begin{pmatrix} p_{mx} \\ p_{my} \end{pmatrix} - s \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} R_{ix} \\ R_{iy} \end{pmatrix} \tag{11}$$

Figure 9 Shape models composed by shape fragments

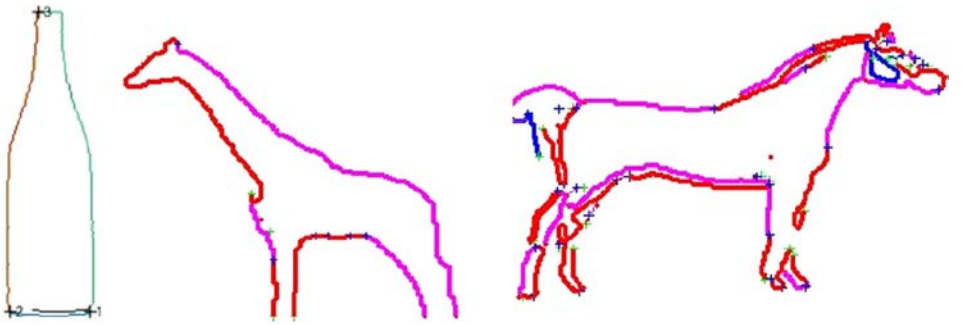
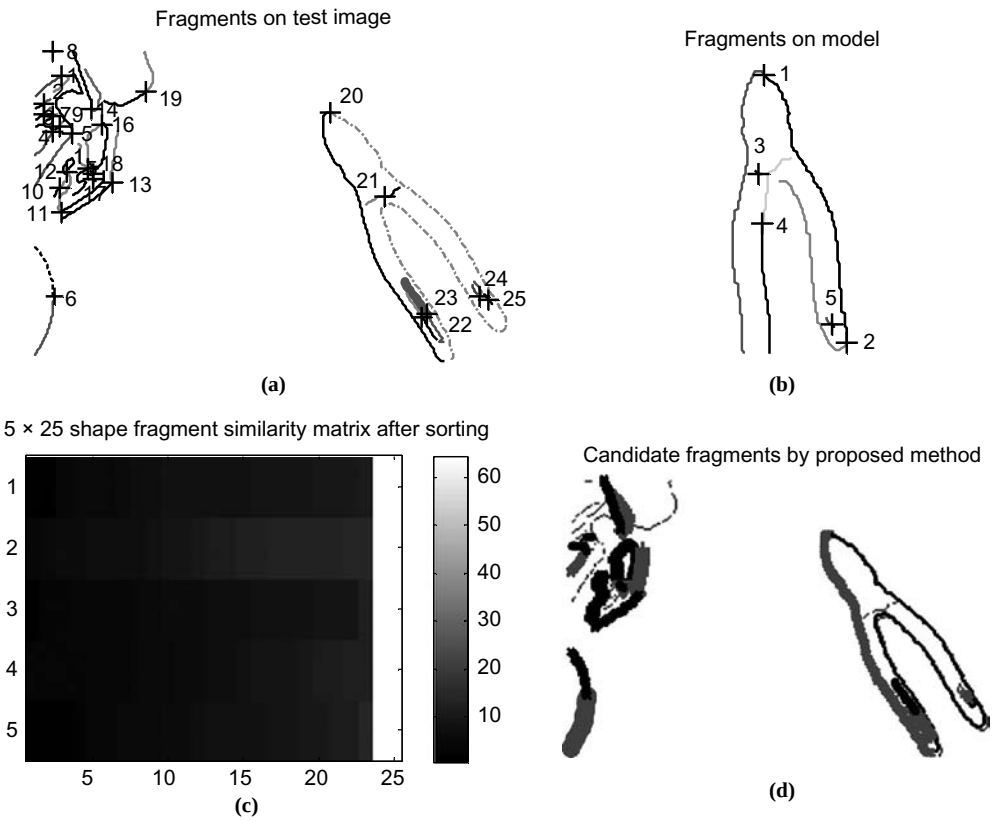


Figure 10 Get the candidate fragments based on the shape fragment similarity



By matching all the shape fragments of the test image with the codebook entry, the location of deemed points of reference can be found by voting method. Prepare a 4D accumulator array $H_v(X_r, Y_r, s, \theta)$ and initialize the value to 0, for each possible matched entry C_i, compute the candidate reference points using the following equation:

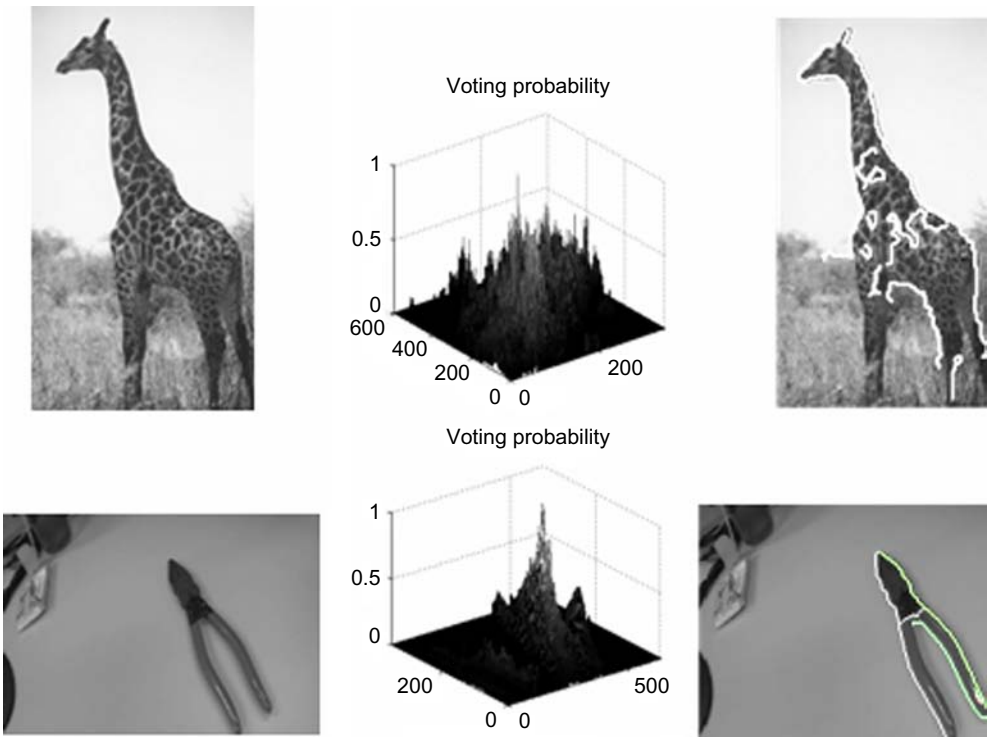
$$H_{v+1}(X_r, Y_r, s, \theta) = H_v(X_r, Y_r, s, \theta) + 1 \qquad (12)$$

All elements in H satisfying $H_v(X_r, Y_r, s, \theta) > threshold$ represent the reference points of the test image. By using the inverse Hough transform, the shape fragments belonging to the object are detected. Figure 11 shows the Hough voting space distribution and the detection results.

According to the Hough space distribution of the reference point, the center of the object can be determined. However, a further step is still needed to verify whether the center obtained by Hough voting method is the right one. A two step method is used to verify:

- *Step 1: inverse Hough transform.* Each point on the shape fragment is tested by equation (11), if more than half of the amount satisfy equation (11). The whole shape fragment is selected as the candidate shape fragment. If the center point got by Hough transform is the correct one, a serial of shape fragments satisfying equation (11) can be got. When there exists background clutter or more than one object exists, more than one peaks will got from the Hough voting space. In this case, all the possible reference points should be taken into equation (11) to verify the shape fragments.
- *Step 2: verify the shape fragments by the propriety of the object.* If all the shape fragments belonging to one object, the shape

Figure 11 Hough voting space and part of the detection results



Notes: The first column is the original image, the second column is the Hough voting space, the third column is the detection results by the proposed method

fragments should have the relationship of *AND*/*OR*. All the candidate shape fragments belonging to one object should connected with others, if the candidate shape fragment is an isolated one, this shape fragment is not the correct one.

3D object detection

A. Shape modeling using multi-view constraints

The 2D object shape model is a graph of interconnected parts. The constraints of the interconnected parts include two parts. One is the multi-view geometrical constraint of the same object part, and the other is the constraint of the object parts in one view. In this section, we introduce the work of the learning of 2D/3D object model. For 2D object modeling, we only need to set up the constraint between all views for the same object part. But for 3D object modeling, our goal is to build a model that is a multi-view representation of a 3D object class. We choose a shape patch based model for describing an object. For each viewpoint, an object is an ensemble of geometrically arranged shape parts. For each part of the 3D object, the corresponding 2D projection part will appear in more than one view.

We start with several viewpoints around the object. A number of viewpoints are captured by a camera. The 3D object generative model is constructed by using viewpoints and the object shape patches that are linked across different views.

B. Multi-view constraints

A simple 3D projection model is shown as Figure 12. The relationship of two views can be represented by:

- the fundamental matrix;
- the planar projective transformation; and
- the quadratic transformation.

In these two views, if local shape patch C_i and C_j can be matched, we represent the relationship of the two shape patch by the planar projective transformation. A planar projective transformation is a linear transformation on homogeneous three-vectors represented by a non-singular 3 by 3 matrix:

$$k\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

or more briefly, $kp' = Hp$.

We use the method described in last section. We use the key point represent the local shape fragment. The match of two images is shown in Figure 13, and the initial matching error and the reprojection error estimated by RANSAC method is shown in Figure 13 right. By using the multi-view relation constraints, it is easy to remove the shape patches with no relevance in all views.

To learn a 3D object representation, we first use a set of training images of a single object to initialize the model. Not like the 2D image, all views contains all the whole object. We build up a codebook from all the training image. When we detect the 2D object from the test image, what we need to do is to match the features extracted from the test image with the codebook. For 3D object, from different view point, the object has different appearance. If we want to model the whole model of the 3D object, it is not easy.

C. 3D object model

We build the 3D object model from a set of training images. We assume that each training image contains one instance of the

Figure 12 Projection model

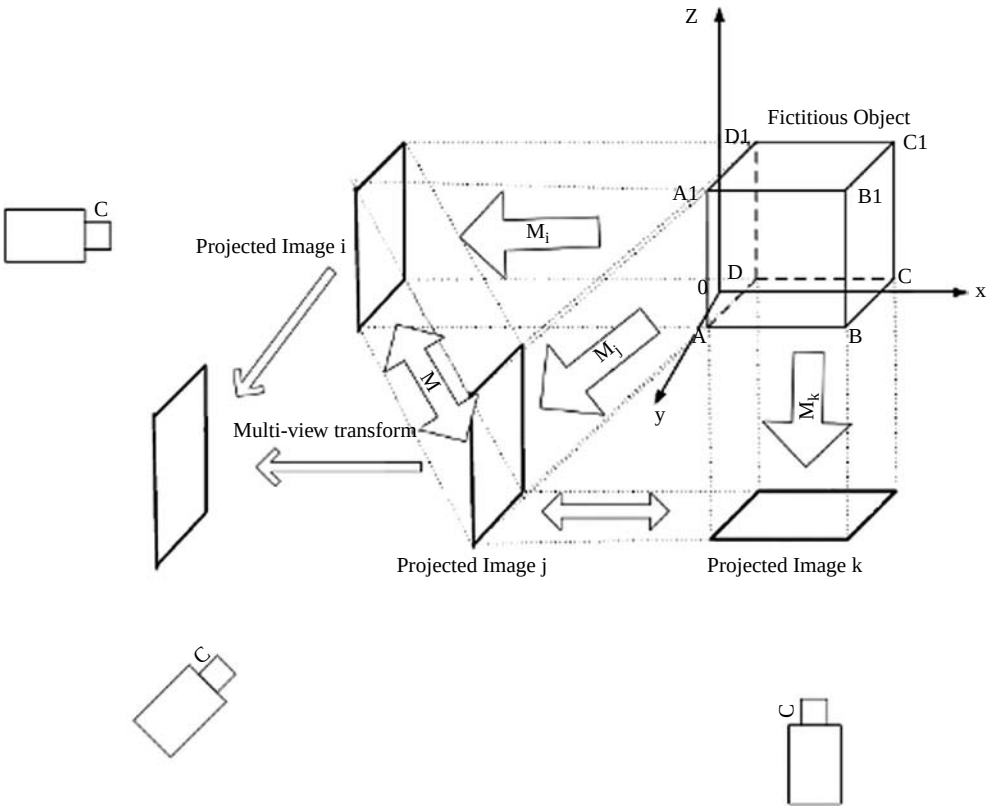
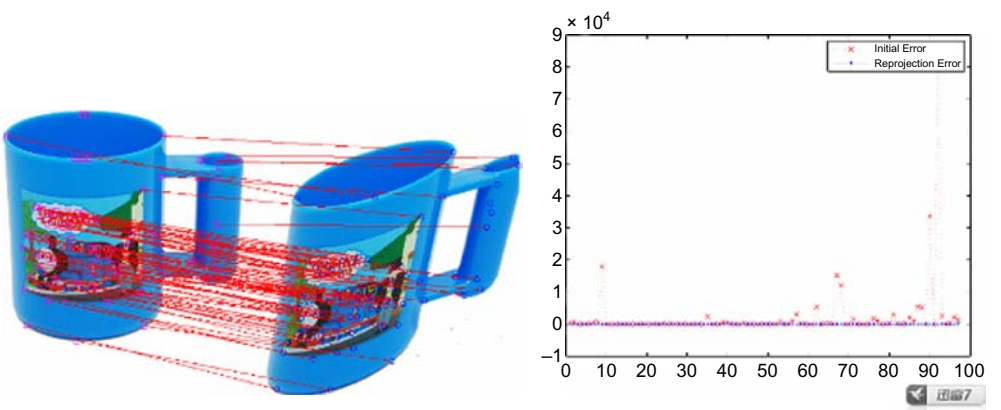


Figure 13 Left: an matching example using the method proposed and right: the initial matching error and the reprojection error estimated by RANSAC method



target object. The task of learning is to start with this set of images, extract local shape fragment features. According to the introduction in last section, a multi-view constraint is used to delete non-object features. We link together all the features form a full 3D model. Below is a brief sketch of model forming algorithm:

1 Obtain a set of M candidate local shape patch features based on shape similarity measured by $T_c(a,b)$ across two training images.
2 Run RANSAC algorithm on M to obtain a new (and smaller) set of matches $MF \in M$ based on $xiFxj = 0$, where F denotes the fundamental matrix.
3 Further refine the matches using RANSAC to obtain a set of MH matches such that $xi.HFxj = 0$, where $MH \in MF \in M$.

Experimental results and evaluation

Results and comparison

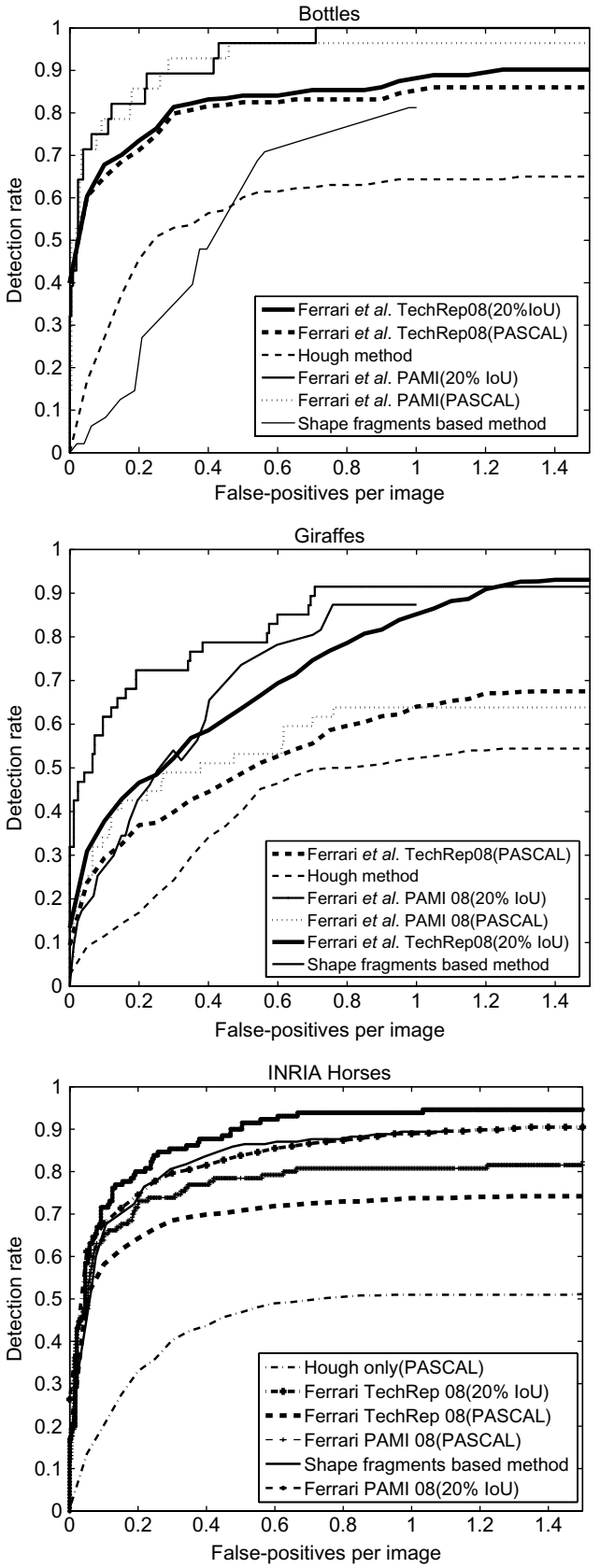
We first report our results on the ETHZ shape classes. ETHZ shape classes has five different object categories with 225 images in total. The objects are surrounded by extensive background clutter and have interior contours.

All categories have significant intra-class variations, scale changes and illumination changes. We use four shape based classes (bottles, giraffes) to verify the proposed method. The advantages of our method are not sensitive to rotation, background clutter and scale changes. For training, we use half the positive examples. All the remaining examples are used for testing. In the training stage, we use the training images to model the shape and set up the object codebook. Figure 14 shows part of the test results using proposed method.

We compare our results with (Ferrari *et al.*, 2008) for comparison, which is also based on the shape fragments. Figure 15 compares our method with the method of Ferrari by plotting recall against false positives per image (fppi). It turns out that our method outperforms the Hough method (Ferrari *et al.*, 2008) and Ferrari's test result under the PASCAL criterion: a detection is counted as correct if the area of intersection between its bounding-box and the ground-truth bounding-box exceeds 50 percent of their union (PASCAL Challenge, 2006). This is mainly because our method use local shape patches around the key point rather than the common contour groups without any constraints. And we use local shape patches, it is invariant to background clutter and

Figure 14 Part of the test results

Figure 15 Detection performance under the PASCAL criterion and comparison



point noises. In this way, this method is better than point to point correspondence methods.

The influence of the edge detection parameter

In this manuscript, Canny edge detection method is used. In order to test the influence of the edge detection parameter to the results, a serial of different edge detection parameter is given. We get the object detection results when the Canny edge detection parameters are 0.1, 0.2, 0.3, 0.4 and 0.5. According to the "PASCAL" criterion, when the detection area is more than 50 percent of the ground truth area, the object is considered to be detected. The test results of the bottle under different Canny edge detection parameters are as shown in Figure 16.

The edge detection result is also influenced by the noise. Figure 17 shows the detection result in different "salt and pepper" noise levels.

Test results analyze when existing many objects and occlusion

As the shape fragments are local features, the proposed method is invariant to occlusion. The object detection problem is converting to the problem of finding the extreme value in the Hough space. When many objects exist in the images, it is the problem of finding many extreme values in the Hough space. An example is shown in Figure 18.

Test results analyze on the scale invariance

Although the descriptor of the shape fragment is invariant to scale, the different shape fragments are extracted in different scales. And the Hough voting method is sensitive to the scale. Hence, the proposed method has its limitation to the scale changes. In order to make the proposed method is insensitive to the scale changes, a multi-scale model is built in the training stage. When the training set and the test set are in a certain scale range, the proposed method is invariant to scale. As the scale of the multi-scale model is discrete, and the test image is not exactly the same size with the model, the proposed method must robust to the scale in a certain range. That means, even if the model is a single scale model, the proposed method can detect the object from the test image in a certain scale range. Figure 19 shows examples of the voting space when the test image is 1/2 of the model and twice the model image.

Conclusions and discussion

In this paper, a shape fragment based object detection method is proposed. In the training stage, the object class model can be represented by the codebook of shape fragments in multi-scale. By comparing the shape fragments of the detect image to the codebook, a distribution of object center image is successfully achieved. Experimental comparisons with other methods were carried out to evaluate the proposed method, and test results shows the method are more distinctive and robust to image transformation and background clutter.

Figure 16 Comparison of the detection results under different Canny parameters

Note: From left to right, the Canny detector parameters are from 0.1 to 0.5

Figure 17 Detection results in different noise levels

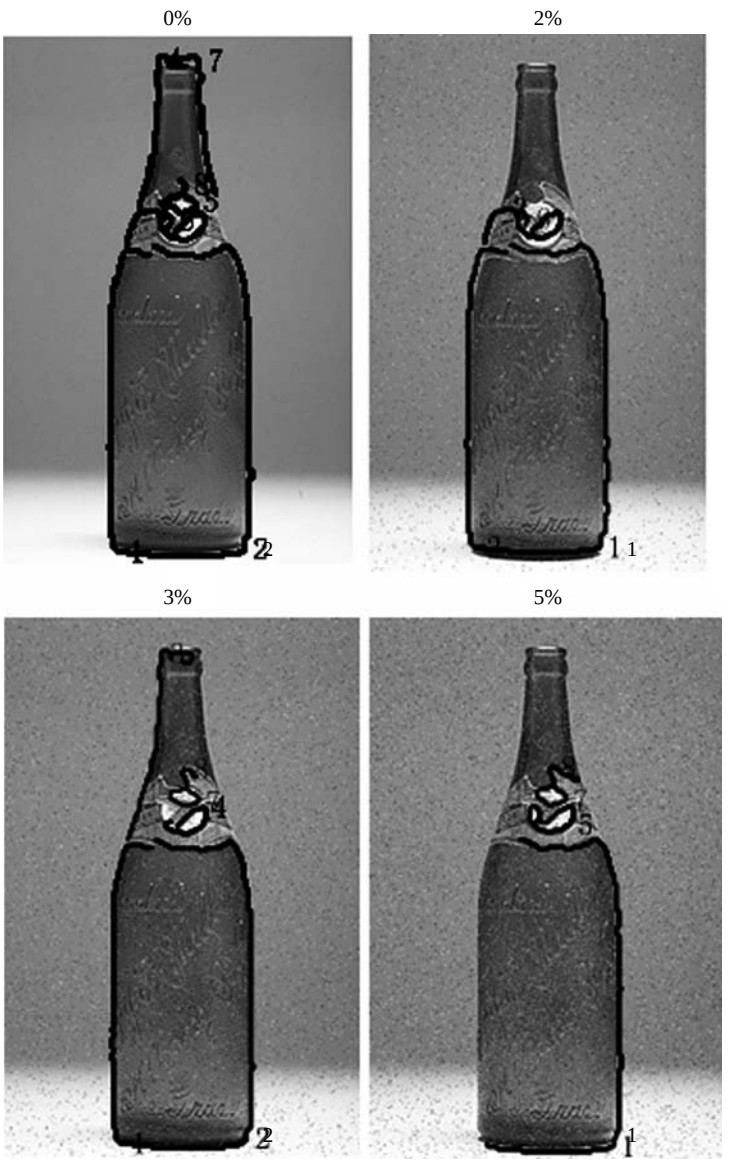
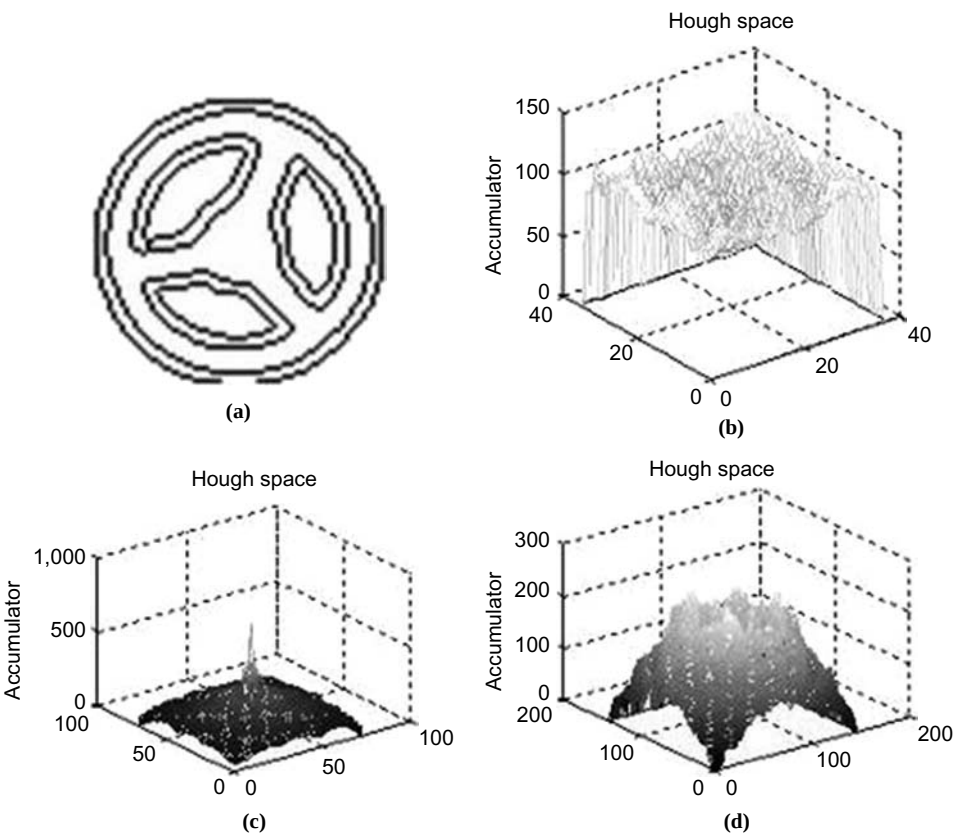


Figure 18 Detection results when there are multi-objects and occlusion

Figure 19 Analyze of the scale problem



Notes: (a) Template image; (b) voting space distribution when the test image size is half of the template image size; (c) voting space distribution when the test image size is the same as the template; (d) voting space distribution when the test image is twice the template image size

References

Bai, X., Li, Q., Latecki, L., Liu, W. and Tu, Z. (2009), "Shape band: a deformable object detection approach", *CVPR*, pp. 418-25.

Bay, H., Tuytelaars, T. and Van Gool, L. (2006), *Surf: Speeded Up Robust Features*, Lecture Notes in Computer Science, Vol. 3951, pp. 404-17.

Belongie, S., Malik, J. and Puzicha, J. (2002), "Shape matching and object recognition using shape contexts", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24 No. 4, pp. 509-22.

Berg, A., Berg, T. and Malik, J. (2005), "Shape matching and object recognition using low distortion correspondences", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 26-31.

Bouganis, A. and Shanahan, M. (2008), "Flexible object recognition in cluttered scenes using relative point distribution models", *19th International Conference on Pattern Recognition*, pp. 1-5.

Bribiesca, E. (1999), "A new chain code", *Pattern Recognition*, Vol. 32 No. 2, pp. 235-51.

ChengEn, L. and Ling, H. (2011), "Contour based object detection using part bundles", *Computer Vision and Image Understanding*, Vol. 114 No. 7, pp. 827-34.

Cootes, T., Taylor, C., Cooper, D. and Graham, J. (1995), "Active shape models their training and application", *Computer Vision and Image Understanding*, Vol. 61 No. 1, pp. 38-59.

Diplaros, A., Gevers, T. and Patras, I. (2003), "Color-shape context for object recognition", *IEEE Workshop on Color and Photometric Methods in Computer Vision*, pp. 167-73.

Elidan, G., Heitz, G. and Koller, D. (2006), "Learning object shape: from drawings to images", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (*CVPR*), Vol. 2, pp. 2064-71.

Felzenszwalb, P. (2005), "Representation and detection of deformable shapes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27 No. 2, pp. 208-20.

Ferrari, V., Tuytelaars, T. and Van Gool, L. (2006), *Object Detection by Contour Segment Networks*, Lecture Notes in Computer Science, Vol. 3 14 pp. 14-28.

Ferrari, V., Fevrier, L., Jurie, F. and Schmid, C. (2008), "Groups of adjacent contour segments for object detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30 No. 1, pp. 36-51.

Ghita, O. and Whelan, P.F. (2002), "Computational approach for edge linking", *Journal of Electronic Imaging*, Vol. 11 No. 4, pp. 479-85.

Hajjar, A. and Chen, T. (1999), "A VLSI architecture for real-time edge linking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21 No. 1, pp. 89-94.

Kushal, A., Schmid, C. and Ponce, J. (2007), "Flexible object models for category level 3D object recognition", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 21-8.

Leibe, B., Leonardis, A. and Schiele, B. (2004), "Combined object categorization and segmentation with an implicit shape model", *ECCV 2004 Workshop on Statistical Learning in Computer Vision*, pp. 17-32.

Leibe, B., Schindler, K., Cornelis, N. and Van Gool, L. (2008), "Coupled object detection and tracking from static cameras and moving vehicles", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30 No. 10, pp. 1683-98.

Leordeanu, M., Hebert, M. and Sukthankar, R. (2007), "Beyond local appearance: category recognition from pairwise interactions of simple features", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8.

Li, C. and Ma, L. (2009), "A new framework for feature descriptor based on SIFT", *Pattern Recognition Letters*, Vol. 30 No. 5, pp. 544-57.

Marghany, M. and Hashim, M. (2010a), "Developing adaptive algorithm for automatic detection of geological linear features using RADARSAT-1 SAR data", *International Journal of the Physical Sciences*, Vol. 5 No. 14, pp. 2223-9.

Marghany, M. and Hashim, M. (2010b), "Texture entropy algorithm for automatic detection of oil spill from RADARSAT-1 SAR data", *International Journal of the Physical Sciences*, Vol. 5 No. 9, pp. 1475-80.

Marghany, M., Hashim, M. and Moradi, F. (2011), "Object recognitions in RADARSAT-1 SAR data using fuzzy classification", *International Journal of the Physical Sciences*, Vol. 6 No. 16, pp. 3933-8.

Marr, D. and Hildreth, E. (1980), "Theory of edge detection", *Proceedings of the Royal Society of London. Biological Sciences, Series B*, pp. 187-217.

Mortensen, E., Deng, H. and Shapiro, L. (2005), "A SIFT descriptor with global context", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 184-90.

Murphy, K., Torralba, A., Eaton, D. and Freeman, W. (2006), *Object Detection and Localization Using Local and Global Features*, Lecture Notes in Computer Science, Vol. 4170, pp. 382-400.

Opelt, A., Pinz, A. and Zisserman, A. (2006), *A Boundary-fragment-model for Object Detection*, Lecture Notes in Computer Science, Vol. 3952, pp. 575-88.

PASCAL Challenge (2006), "2006 visual object classes", available at: www.pascal-network.org/challenges/VOC/voc2006

Ravishankar, S., Jain, A. and Mittal, A. (2008), "Multi-stage contour based detection of deformable objects", *ECCV*, pp. 483-96.

Shotton, J., Blake, A. and Cipolla, R. (2005), "Contour-based learning for object detection", *Proc. ICCV*, Vol. 1, pp. 503-10.

Stark, M., Goesele, M. and Schiele, B. (2009), "A shape-based object class model for knowledge transfer", *Twelfth IEEE International Conference on Computer Vision*, pp. 373-80.

Thayananthan, A., Stenger, B., Torr, P. and Cipolla, R. (2003), "Shape context and chamfer matching in cluttered scenes", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 122-7.

Ullman, S., Sali, E. and Vidal-Naquet, M. (2001), *A Fragment-based Approach to Object Representation and Classification*, Lecture Notes in Computer Science, Vol. 2059, pp. 85-102.

Wang, L., Shi, J., Song, G. and Shen, I. (2007), *Object Detection Combining Recognition and Segmentation*, Lecture Notes in Computer Science, Vol. 4843, pp. 189-99.

Xu, C., Liu, J. and Tang, X. (2009), "2d shape matching by contour flexibility", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31 No. 1, pp. 180-6.
Zhijie, W. and Hong, Z. (2008), "Edge linking using geodesic distance and neighborhood information", *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pp. 151-5.
Zhu, Q., Wang, L.M., Wu, Y. and Shi, J.B. (2008), "Contour context selection for object detection: a set-to-set contour matching approach", *ECCV2008*, pp. 774-87.

Further reading

Ferrari, V., Jurie, F. and Schmid, C. (2010), "From images to shape models for object detection", *International Journal of Computer Vision*, Vol. 87 No. 3, pp. 284-303.
Viola, P. and Jones, M. (2001), "Rapid object detection using a boosted cascade of simple", *Proc. IEEE CVPR*, pp. 1-10.

Corresponding author

Chen Guodong can be contacted at: guodongxyz@gmail.com