Multi-class Obstacle Identification using Shape Descriptors in Video for Autonomous Vehicles

Shumin Liu[1, a], Xujuan Xu[2,b] and Longyun Duan[2,c]

[1] School of Software and Engineering, JiangXi University of Science and Technology, China

[2] Jiangxi Civil Affairs School, Nanchang, China

[a]liushumin_001@163.com, [b]20518713@qq.com, [c]444853203@qq.com

Keywords: Obstacle identification, Shape descriptor, Shape context, Axis of Least Inertia

Abstract: Obstacle identification has been widely studied as part of the broader obstacle detection research area for Autonomous Vehicles (AV). Existing in-vehicle sensing systems are concentrated on obstacle detection for pedestrian or vehicle, and limited work has been conducted on multi-class obstacle classification. In the process of obstacle identification, the selection of classification features is particularly critical. As a set of features to describe a given shape or contour, shape descriptor has attracted much attention in recent years and play an important role in pattern recognition. This paper proposed a shape descriptor based multi-class obstacle identification method where the traffic obstacles (in the front of self-vehicle) be classified into one of four classes: vehicle, lateral moving pedestrian, longitudinal moving pedestrian, and unknown (such as trees, road lamp, barricade etc.). Here a variety of shape descriptors extracted from the contour curve are involved, such as Rectangularity, Compactness, Elongation, Circularity, Shape Context, and Axis of Least Inertia. Finally, the identification results using these descriptors are contrastive analyzed. Though a single shape descriptor does not achieve ideal identification results for traffic obstacle, but this will provide a new idea for multi-class obstacle identification using shape descriptor in video for AV.

I. Introduction

Over the past few years, obstacle identification has been widely studied as part of the broader obstacle detection research area for Autonomous Vehicles (AV). Recently, although great progresses have been made in this field, there still exist many issues to be solved especially for applications in urban traffic scenarios. Compared to highway, urban traffic is much more complex because of cluttered road scenes and diversity of obstacles. However, existing in-vehicle detecting systems are mainly for pedestrians or vehicles, limited work has been conducted on multi-class obstacle classification. Therefore, one of AV challenges for use in urban traffic is developing advanced in-vehicle detecting systems able to reliably detect and identify multi-class obstacles.

Obstacle detection generally consists of two processes [1]: Regions of Interest segmentation (ROIs), target verification and identification. In our previous research works [2-3], we first segmented ROIs from traffic scenario according to their disparity in depth map, then extracted obstacle's contour curve using Snake Model. As a follow-up to the above work, we will extracting a set of shape descriptors from contour curve.

Shape descriptors are some set of features that are computed to describe a given shape or contour curve. It has attracted much attention in recent years and play an important role in pattern recognition, target recognition, product testing, industrial automation, military security and other fields. Reference K. Mahdikhanlou [4] proposed a plant leaf classification method using centroid distance and Axis of Least Inertia Method, [5] structure a signature system based on extended Shape Context Descriptors, and [6] proposed a RGB-D multi-view object detection with object proposals and shape context. However, there are limited work has been conducted on obstacle identification using shape descriptor for AV.

This paper will propose a shape descriptor based multi-class obstacle identification method where the traffic obstacles (in the front of self-vehicle) be classified into one of four classes: Vehicles, Lateral moving pedestrians (LA-P), Longitudinal moving pedestrian (LO-P), and Unknown (such as trees, road lamp, barricade etc.). In the process of obstacle identification, other features are not used

expect shape descriptors. In order to effectively verify the classification effect of different shape descriptors, the test is divided into two stages: 1) pedestrian-vehicle (PV) identification, i.e. the sample sets contain only pedestrians and vehicles; 2) pedestrian-pedestrian (PP) identification, i.e. the sample sets contain only LA-P and LO-P. This paper will provide a new idea for multi-class obstacle identification, though a single shape descriptor does not achieve ideal classification results for traffic obstacle.

In the rest of this paper, section II introduces the shape descriptors. Section III discusses system framework and object identification steps. Section IV shows the experimental results and contrastive analyses the results. Finally, section V concludes this paper.

II. Shape Descriptors

Six main types of shape descriptors are involved in this paper, which are Rectangularity, Compactness, Elongation, Circularity, Shape Context, and Axis of Least Inertia.

A. Rectangularity

The rectangularity Ra of contour represents the ratio of the area of itself to the minimum enclosing rectangle (MER), which can be represented as:

$$R_a = {A_0}/{A_R} \tag{1}$$

where A0 and AR are the area surrounded by the contour and MER respectively, and Ra reflects the area fill level for MER, varies from 0 to 1.

B. Compactness

The compactness C (also called shape factor) is determined by the perimeter C and area factor A of the target region:

$$F = \frac{C^2}{4\pi A} \tag{2}$$

$F=1$ when the target is a circle, otherwise $F>1$. Compactness is a reflection of the compactness of the target region and it has the advantage of insensitivity to scale changes.

C Elongation

Elongation (also known as eccentricity) is also a description of the target's compactness which derived from the region's inertia. Set $A = \sum m_i y_i^2$、 $B = \sum m_i x_i^2$ are the moment of inertia around the X and Y axis of the region respectively, $C = \sum m_i x_i y_i$ is the inertial product, and p, q are two semi-major axis of the inertia ellipse for the region expressed by A, B and C:

$$\begin{cases} p = \sqrt{2/\left[(A + B) - \sqrt{(A - B)^2 + 4C^2}\right]} \\ q = \sqrt{2/\left[(A + B) + \sqrt{(A - B)^2 + 4C^2}\right]} \end{cases} \tag{3}$$

And the elongation E is defined as:

$$E = {p}/{q} \tag{4}$$

where E varies from 0 to 1, and it has maximum value when the target is a circle.

D. Sphericity

In the description of a 2D target, sphericity S is defined as:

$$S = R_i/R_c \tag{5}$$

where R_i and R_c represent the inner and outer circle radius of the target region respectively, and both of the center of the circle are located in the centroid of the region. S is changed between 0 and 1 and it reaches the maximum when the target is a circle region.

E. Shape Context[7]

The basic idea of shape context is to record relative distances from a given point to all other points. SC begins by taking N sample points from an object. Typically those sample points are from shape edges. For point pi of N sample points for an object, there are N - 1 vectors from pi to all other points pj, where i ≠ j. Such vectors strongly indicate the relative positions of all other points pj to pi. The SC of pi is defined with following equation:

$$h_i(k) = \#\{q \neq p_i : (q - p_i) \in bin(k)\} \tag{6}$$

where the bins are normally taken to be uniform in log-polar space and the value of each bin is the number of points that are located into this bin. In general, the log-polar is divided into 12×5 bins which are 12 angles and 5 distances. As a result, there will be N histograms for N sample points of an object and such a set of histograms is the shape context descriptors for the object.

In order to perform matching between two sets of sample points of two objects, denote the shape contexts for points pi and qj as hi(k) and hj(j), respectively. As shape contexts are distributions represented as histograms, it is natural to use Pearson's Chi-square ($\chi2$) test statistics as the "shape context cost" of matching two points:

$$C_{ij} = C(p_i, q_i) = \frac{1}{2} \sum_{k=1}^{K} \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)} \tag{7}$$

where Cij is the cost between points pi and qj, and hi(k) denotes the k-th bin of pi's shape context. Such cost can give a measurement for the quantified similarity between two points. Finally, an n × n cost matrix C can be built. The goal is to find a one-to-one matching that minimizes the total matching cost.

F. Axis of Least Inertia[4]

The axis of least inertia (*ALI*) of a shape is defined as the line for which the integral of the square of the distances to points on the shape boundary is a minimum. Once the *ALI* is calculated, each point on the shape curve is projected on to *ALI*. The two farthest projected points say *E1* and *E2* on *ALI* are chosen as the extreme points. The Euclidean distance between these two extreme points defines the length of *ALI*. Once the axis of least inertia of a shape is obtained, keeping it as a unique reference line, symbolic features are extracted from the shape.

III. Technical Approach

The complete proposed system pipeline is shown in Fig.1. All components are explained in the following sections.
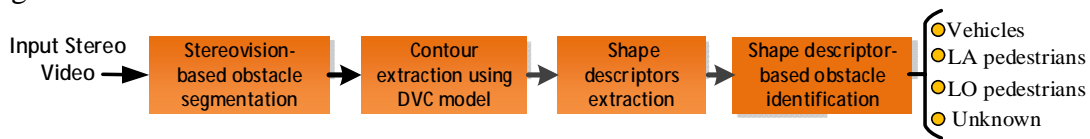

Fig. 1. System framework

A. Stereovision based obstacle segmentation

A key technique of vision-based obstacle detection is how to segment obstacle from background. Object segmentation is a hard task in highly cluttered urban environments using moving cameras. Compared to mono-camera approach like pattern matching and motion analysis, stereovision-based approach can segment obstacles from traffic scenario according to their disparity in depth map. Such an approach can segment multi-class objects no matter what the object shape is and no matter whether the objects are moving or stationary. This is the reason of choosing stereovision in this work.

B. Contour extraction using DVC Model

Active Contour Models [8], also called Snakes, is a curves defined within an image domain that deform and move to minimize the energy of the fitting error. It is able to extracting contours in a 2-dimensional image, even in presence of gaps or occlusions, together with its dynamic behavior, makes it adequate for delineating rigid objects like vehicles and non-rigid objects like pedestrians. In our previous research work [3], we proposed an improved Snake Model, i.e. DVC Model, in order to

address some limitations of the classic Snake Model. In this work, after the obstacles be segmented from traffic video, we will extract complete contours for them using DVC Model.

C. Shape descriptors extraction for contour curve

Six shape descriptors will be extracted from the target contour curve according to above-mentioned methods, that is Rectangularity, Compactness, Elongation, Circularity, Shape Context, and Axis of Least Inertia, after the contour curve of obstacles are extracted using *DVC Model*.

D. Multi-class obstacle identification using shape descriptor

Shape descriptor play an important role in object identification. After 6 shape descriptors are extracted successfully, they will are used respectively to identify the obstacle. In this stage, the obstacles will be divided into *Vehicles*, *LA-Pedestrian*, *LO-Pedestrian*, and *Unknown*. In order to effectively verify the classification effect of different shape descriptors, the test is divided into two stages: 1) pedestrian-vehicle (PV) identification; 2) pedestrian-pedestrian (PP) identification.

IV. Experimental Results and Discussions

In this paper, the vehicle-mounted stereo camera Bumblebee3 shown in Fig.2, as a means of the obstacle perception, collected a large number of vehicles forward traffic video sequence. After stereovision based obstacle segmentation and DVC Model based contour extraction, there are 400 pedestrian samples with various postures, 400 vehicles samples with different perspective, and 200 other obstacles are gathered from several traffic scenarios in all. The obstacles are located in the front of self-vehicle about 8 to 40 meters, and part of the samples as shown in Fig.2. Moreover, 1000 samples get from the two publicly available datasets (i.e. Daimer pedestrian dataset [9] and Caltech pedestrian dataset [10]) which are gathered with in-vehicle camera. Therefore, the sample set X consist of three subset: the pedestrian subset X1, the vehicle subset X2, and Unknown subset X3, i.e. $\mathbf{X} = \{X_1, X_2, X_3\}$.

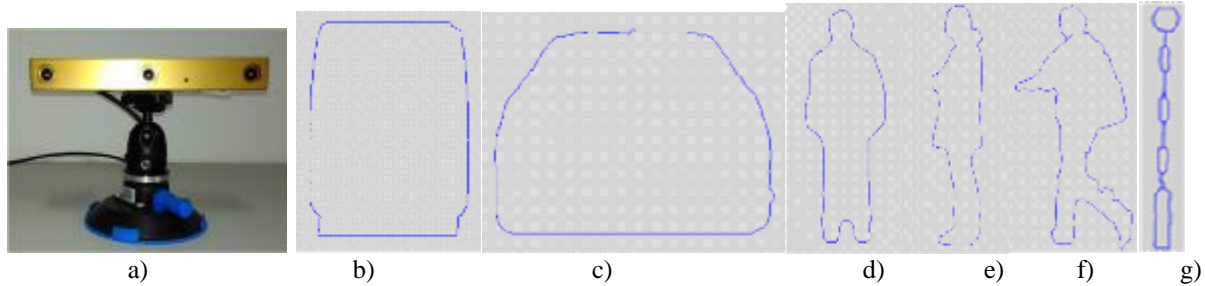| a) | b) | c) | d) | e) | f) | g) |

Fig. 2 Stereo camera Bumblebee3 and some samples of obstacle with contoure curve

1. Experimental Results

These samples will be tested by the 6 shape descriptors, and the experimental results will be shown below respectively.

A. Rectangularity

Calculate the Rectangularity *Ra* of sample set X according to formula $R_a = A_0/A_R$ respectively, then estimate their means $\hat{\mu}$ and standard deviations $\hat{\sigma}$ by the maximum likelihood estimation method, as shown in Table 1. (Notes: the experimental result for Unknown dose not shown because the value of *Ra* has a wide variety and it usually be identified by exclusive method, the same below.)

Table 1   The statistical property and identification capability of *Ra*

	Pedestrians	Vehicles	LA-Pedestrian	LO-Pedestrian
$\hat{\boldsymbol{\mu}}$	0.5115	0.7887	0.5858	0.4368
$\hat{\boldsymbol{\sigma}}$	0.1052	0.0253	0.0593	0.0862
between-class distance (BCD)	$\hat{\mu}_{PD} - \hat{\mu}_{VH} = \mathbf{2.124}(\hat{\sigma}_{PD} + \hat{\sigma}_{VH})$		$\hat{\mu}_{SPD} - \hat{\mu}_{LRPD} = \mathbf{1.024}(\hat{\sigma}_{SPD} + \hat{\sigma}_{LRPD})$	
identification capability(IC)	96.6%		69.2%	

Fig.3 shows the histogram of *Ra* for different obstacles. Where the dotted curve simulates the distribution of *Ra* according to normal curve, the distance between the peaks of curve represents

between-class distance. It can be found from Fig.3 that the curve overlap-rate for PV is lower, and that for PP is higher. This means that Ra has a stronger identification ability for PV and a weaker ability for PP.

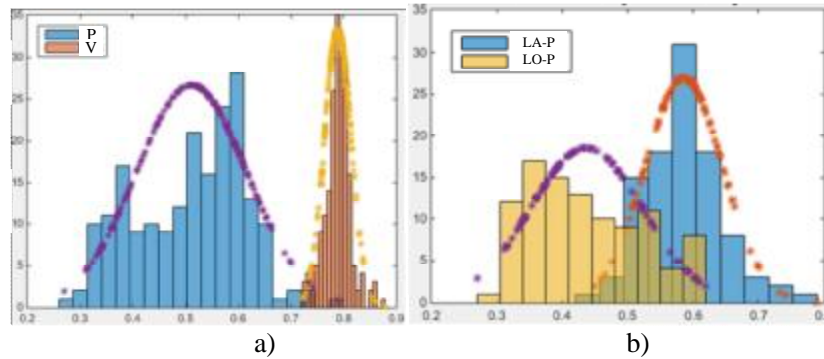

a)   b)

Fig.3 the histogram of Ra and the analog curve with Gaussian normal function, a) for PV, and b) for PP

Obviously, it is difficult to describe the degree of overlap-rate between two categories. In order to qualitatively analyze the identification ability of Ra, means and standard deviations of sample X are involved. Then the identification capability/ probability of Ra can be calculated in terms of the 3σ distribution principle for normal curve, that is: $P(\mu - \sigma < X \leq \mu + \sigma) = 68.3\%$, $P(\mu - 2\sigma < X \leq \mu + 2\sigma) = 95.4\%$, $P(\mu - 3\sigma < X \leq \mu + 3\sigma) = 99.7\%$, and that of other intervals X can be obtained by referring to the normal distribution table. Finally, the between-class distance $d_{i,j}$ between class ω_i and class ω_j can be calculated by $d_{i,j} = |\mu_i - \mu_j|/(\sigma_i + \sigma_j)$, for example, $d_{i,j}$ =2.124 for PV, and the probability for them is 96.6% according to the normal distribution table. The results of $d_{i,j}$ and corresponding identification probability for Ra are shown in Table. 1.

As shown in Table.1, Ra has a 96.6% ability to distinguish between pedestrians and vehicles, and the ability to distinguish LA-Pedestrian and LO-Pedestrian is only 69.2%.

B. Compactness

Compute the compactness F of sample X according to $F = C^2/4\pi A$ respectively, then estimate their μˆand σˆby the maximum likelihood estimation method, as shown in Table 2.

Table 2   The statistical property and classification capability of F

	Pedestrians	Vehicles	LA-Pedestrian	LO-Pedestrian
$\hat{\mu}$	3.8752	1.4040	3.1498	4.5936
$\hat{\sigma}$	1.1636	0.0977	0.7507	1.0457
BCD	$\hat{\mu}_{PD} - \hat{\mu}_{VH} = 1.9592(\hat{\sigma}_{PD} + \hat{\sigma}_{VH})$		$\hat{\mu}_{SPD} - \hat{\mu}_{LRPD} = 0.8037(\hat{\sigma}_{SPD} + \hat{\sigma}_{LRPD})$	
IC	95.0%		57.6%	

C. Elongation

The elongation E of samples are calculated by formula $E=p/q$ and the relevant statistical properties are calculated as shown in Table 3.

Table 3   The statistical property and classification capability of E

	Pedestrians	Vehicles	LA-Pedestrian	LO-Pedestrian
$\hat{\mu}$	0.9465	0.7064	0.9528	0.9403
$\hat{\sigma}$	0.0227	0.0948	0.0167	0.0261
BCD	$\hat{\mu}_{PD} - \hat{\mu}_{VH} = 2.0391(\hat{\sigma}_{PD} + \hat{\sigma}_{VH})$		$\hat{\mu}_{SPD} - \hat{\mu}_{LRPD} = 0.2921(\hat{\sigma}_{SPD} + \hat{\sigma}_{LRPD})$	
IC	95.8%		22.8%	

D. Sphericity

The Sphericity SP is evaluated by formula $SP = R_i/R_c$ for different sample as shown in Table 4.

Table 4   The statistical property and classification capability of SP

	Pedestrians	Vehicles	LA-Pedestrian	LO-Pedestrian
$\hat{\mu}$	0.1591	0.4746	0.1412	0.1904
$\hat{\sigma}$	0.0391	0.0550	0.0260	0.0327
BCD	$\hat{\mu}_{PD} - \hat{\mu}_{VH} = 3.3528(\hat{\sigma}_{PD} + \hat{\sigma}_{VH})$		$\hat{\mu}_{SPD} - \hat{\mu}_{LRPD} = 0.8382(\hat{\sigma}_{SPD} + \hat{\sigma}_{LRPD})$	

IC	99.9%	60.0%

E. Shape Context(SC)

Follow the description of the shape context above, divided the log-polar into 12×5 bins which are 12 angles and 5 distances, and 60 (i.e. $12 \times 5 =60$) shape features SC_{ij}, i=1...12, j=1…5, are extracted totally after the target contour curve is projected to the corresponding region. Here 100 sampling points are extracted from target contour. An analysis for identification results shows that only SC_{5-3}, SC_{5-4}, SC_{5-9}, SC_{5-10} have an ideal distinguish ability for PV, and none of SC have a ideal ability for PP, the results as shown in Table 5.

Table 5   The statistical property and classification capability of SC_{5-3}, SC_{5-4}, SC_{5-9}, and SC_{5-10}

	SC_{5-3}		SC_{5-4}		SC_{5-9}		SC_{5-10}	
	Pedestrians	Vehicles	Pedestrians	Vehicles	Pedestrians	Vehicles	Pedestrians	Vehicles
$\hat{\mu}$	37.2462	775.3873	36.9447	803.9363	36.3367	755.8725	37.8844	823.4216
$\hat{\sigma}$	50.6259	78.6595	49.8990	70.3161	49.4943	77.7243	51.1295	69.7421
BCD	$5.70(\hat{\sigma}_{PD} + \hat{\sigma}_{VH})$		$6.34(\hat{\sigma}_{PD} + \hat{\sigma}_{VH})$		$5.66(\hat{\sigma}_{PD} + \hat{\sigma}_{VH})$		$6.66(\hat{\sigma}_{PD} + \hat{\sigma}_{VH})$	
IC	100%		100%		100%		100%	

F. Axis of Least Inertia (ALI)

According to the description of Axis of Least Inertia (*ALI*) above to extract the *ALI* of sample *X*. The corresponding statistical properties shown in Table 6.

Table 6   The statistical property and classification capability of *ALI*

	Pedestrians	Vehicles	LA-Pedestrian	LO-Pedestrian
$\hat{\mu}$	9.6294	27.0282	10.6444	8.7844
$\hat{\sigma}$	1.5080	3.5835	1.2057	0.9606
BCD	$\hat{\mu}_{PD} - \hat{\mu}_{VH} = 3.417(\hat{\sigma}_{PD} + \hat{\sigma}_{VH})$		$\hat{\mu}_{SPD} - \hat{\mu}_{LRPD} = 0.8586(\hat{\sigma}_{SPD} + \hat{\sigma}_{LRPD})$	
IC	99.9%		61.0%	

2. Contrast and Analyze

The identification capabilities (IC) of shape descriptors involved above are shown in Table 7. Where the IC for PV be shown in the second column, and that for PP be shown in the third column. The symbol '--' denoted there is no corresponding data for them, i.e. the IC is very low or not being calculated.

Table 7   The identification capability using shape descriptors for Multi-class obstacles

Shape descripors		Identification Capability(%)	
		PV	PP
R_a		96.6%	69.2%
C		95.0%	57.6%
E		95.8%	22.8%
SP		99.9%	60.0%
SC	SC_{5-3}	100%	--
	SC_{5-4}	100%	--
	SC_{5-9}	100%	--
	SC_{5-10}	100%	--
ALI		99.9%	61.0%

It can be seen from the second column of table 5 that all the shape descriptors have an ideal identification capability for PV, and the lowest one is 95.0% and the highest one is 100%. However, it can be found from the third column that most of the IC for PP are weaker, and the highest one is only 69.2% and the lowest almost is 0%.

Through contrastive analysis the IC of 6 shape descriptors, it can be concluded that:

1) All the shape descriptors have a stronger identification capability for PV, and Shape Context factors SC_{5-3}, SC_{5-4}, SC_{5-9}, and SC_{5-10} are the best ones, *SP* is followed, and *C* is the worst.

2) Most of the descriptors have a weaker identification capability for PP. As the best one, the IC of *Ra* is 69.2%, *ALI* is followed, and *SC* is the worst which has hardly any identification capability.

3) Only few descriptors have a better comprehensive identification ability for both PV and PP, where the best ones are *ALI*, *Ra* and *SP*, and the rest are relatively weaker.

V. Conclusion

This paper presents a shape descriptor based multi-class obstacle identification method for AV, where the traffic obstacles (in the front of self-vehicle) be classified into four classes: vehicle, horizontal moving pedestrian, longitudinal moving pedestrian, and unknown (such as trees, road lamp, barricade etc.). As a follow-up work of obstacle segmentation and target contour extraction, a set of shape descriptors, such as Rectangularity, Compactness, Elongation, Circularity, Shape Context, and Axis of Least Inertia are extracted from the contour curve firstly, and then these shape features are used to identify the sample *X* which consist of vehicle, LM-pedestrians, LO-pedestrian, and unknown respectively. Finally, the identification results using different shape descriptor be contrastive analyzed, and it can be concluded that though a single shape descriptor does not achieve ideal classification results for traffic obstacle, but this can provide a new idea for multi-class obstacle identification in video for AV.

Acknowledgements

References

[1] David G, Antonio M L. Survey of Pedestrian Detection for Advanced Driver Assistance Systems[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(7): 1239-1258.

[2] S. Liu, Y. Huang and R. Zhang. Obstacle Recognition for ADAS Using Stereovision and Snake Models[C]. 2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC) October, Qingdao, China, 2014: 99-104.

[3] Y. Huang, S. Liu and R. Zhang. Multi-class obstacle detection and classification using stereovision and improved Active Contour Models [J]. IET INTELLIGENT TRANSPORT SYSTEMS, 2016, 10(3):197-205.

[4] K. Mahdikhanlou and H. Ebrahimnezhad. Plant Leaf Classification Using Centroid Distance and Axis of Least Inertia Method [C]. The 22nd ICEE Shahid Beheshti University, 2014,1690-1694.

[5] M. Adamski1, K. Saeed, etc. Signature system based on extended Shape Context Descriptors[C]. 2013 International Conference on Biometrics and Kansei Engineering, 2013, 267-272.

[6] G. Georgakis and Md. Alimoor Reza. RGB-D Multi-View Object Detection with Object Proposals and Shape Context [C]. IEEE/ RSJ International Conference on Intelligent Robots and Systems (IROS) 2016, 4125-4130.

[7] Z. Wang, Bin Lu etc. Leaf Image Classification with Shape Context and SIFT Descriptors[C]. 2011 International Conference on Digital Image Computing: Techniques and Applications. 2011, 650-654.

[8] M. Kass, A. Witkin, D. Terzopoulos. Snakes: active contour models [J]. International Journal of computer vision, 1987, 1(4): 321-331.

[9] http: //www. science. uva. nl /research /isla /downloads /pedestrians /.

[10] http: //www. vision. caltech. edu /Image_Datasets / CaltechPedestrians /.