www.ietdl.org

Published in IET Image Processing
Received on 14th July 2010
Revised on 22nd November 2010
doi: 10.1049/iet-ipr.2010.0288

ISSN 1751-9659

Database of human segmented images and its application in boundary detection

Q. Wang[1] *S.Y. Li*[2]

[1]*School of Electronics and Control Engineering, Chang'an University, Xi'an 710064, People's Republic of China*
[2]*The Chinese Academy of Sciences, Shanghai Institute of Technical Physics, Shanghai 200083, People's Republic of China*
E-mail: crabwq@mail.ustc.edu.cn

Abstract: This study presents a database that can be used for boundary detection and image segmentation evaluation. Images in our database are based on image pairs instead of single images, which is a remarkable difference from existing databases. This characteristic will greatly facilitate the algorithms combining low-level cues such as brightness, colour and texture with depth information generated from the image pairs to conduct boundary detection and image segmentation tasks. Besides, the authors have put the presented database to use by comparing their proposed boundary detection method with the Berkeley detector and Edison detector. Experimental results show that their method has a better performance.

1 Introduction

Boundary detection and image segmentation are two fundamental problems in computer vision. In the past few decades, many algorithms have been developed, a survey of which can be found in [1–8]. Typically, researchers show their results on a few images and claim that their results look better than others. In fact, from such studies we cannot affirm whether or not their results are better than others indeed, and we also cannot exactly know whether or not the examples they used are typical ones. To properly position the states of the art of existing algorithms, several databases [9–12] containing ground-truth segmentations emerged as benchmarks for assessing the performance of different algorithms. Among these, the Berkeley database [12] presented by Martin *et al.* is the most widely influential one in literature. However, the problems of both boundary detection and image segmentation are hard to solve, and there is still no general purpose solution approaching human level competence according to the evaluation results based on these databases.

In recent years, many sophisticated stereo vision algorithms [13] make it possible to obtain much more accurate depth information than ever before. At the same time, there is evidence [14, 15] showing that depth information, along with other low-level cues such as brightness, colour and texture, can facilitate the process of boundary detection and image segmentation. Then we believe that the techniques combining depth information and other low-level cues to detect boundaries and segment images will greatly improve the results of state-of-the-art algorithms. For this purpose, existing human segmented image databases seem a little impotent to provide as benchmarks, because the images in these databases have no depth information. Based on this deficiency, this paper duly introduces a database consisting of image pairs to fill the gap. Images in our database are selected from Middlebury database [16] and their depth information can be obtained according to stereo vision algorithms. For research based on this database, we can combine the low level features of images with the depth information and then evaluate their results by the ground-truth segmentations. This characteristic is a great improvement compared with existing databases.

The rest of this paper is organised as follows. In Section 2, we describe in detail the construction of our human segmented database. In Section 3, we put the database to use by evaluating the boundary detection results of our proposed algorithm. Finally, conclusion and discussion are made in Section 4.

2 Human segmented image database

In this section, we will give a detailed explanation of the considerations and procedures in the database construction.

2.1 Image selection

The first task in constructing the database is to select a set of images. We choose all the images from Middlebury database [16], which is a publicly available and widely accepted one in stereo vision field. There are 38 groups of colour images in together and each group consists of seven or nine well-calibrated image sequence of a specific actual scene with only horizontal displacement between the adjacent image pair. For each group, only two images are supplied with ground-truth depth maps in the Middlebury database. Based on this platform, researches cannot only test the performance of their stereo matching algorithms according to the ground-truth information but also submit their results to position their algorithms' states of the art on Middlebury website [13]. We select the two images with ground-truth depth maps from each group to be segmented by human. So, there are altogether 38 pairs of images in our database.

2.2 Segmentation tool

In order to easily collect segmentations from a wide range of people, we have developed a toolkit that can be used to divide an image into segments. The software has basic tools to zoom in/out an image. It can also draw smoothly on the image to label different segments with a three-pixel wide pen and check whether the boundary of each segment is closed. After segmentation, the black–white binary boundary map is saved as bmp format.

2.3 Human segmentation procedure

To segment the images, instructions are made for subjects participating in the process: 'Divide each image into closed segments. It is important that all the segments have approximately equal importance. The number of segments in each image is up to you and there is no time limit for your segmentation process.

The initial subject group is 50 graduate students major in computer vision. All of them are first provided with several example segmentations of simple, unambiguous images as a visual description of the task. We make sure that each image has 6–10 segmentations and each participant does not segment one image twice. Fig. 1 shows three most used images (Venus, Teddy, Cones) and six of their ground-truth segmentations in our database.

3 Boundary detection application

Boundary detection is different from what is classically referred to as edge detection. An edge is most often defined as an abrupt change in some low-level image features such as brightness, colour and texture, while a boundary is a contour in the image indicating a change in pixel ownership from one object or surface to another. Much work relating to boundary detection has been done in recent years [17–21] and among these, the Berkeley boundary detector [3] is the most remarkable one. Based on these accomplishments, it is agreed that finding the boundaries of objects and surfaces in a scene is a problem of fundamental importance in computer vision [21] for their applications are extensive. For example, a large body of work on object recognition [22–24] and image segmentation [25–27] relies on boundary detection results. Even in cases where simple low-level features are sufficient for these tasks, it is still desirable to incorporate boundary detection results in order to provide precise object information. However, most of the current work is based on single view image information. One reason we think may be the lack of such a database combing depth maps and low-level information.

In this section, some preliminary results of boundary detection are demonstrated. The depth information is fused into our method to generate a boundary map of probabilistic and binary type, and the obtained results are compared with Berkeley detector's [3] and Edison detector's [28], all of which are based on the presented database.

3.1 Boundary detection method

Berkeley boundary detector, assumed to be the most influential and effective one, combines brightness, colour and texture cues to provide a probabilistic boundary map, where for each pixel in the image a probability for being a contour is computed. However, its results are not satisfying in several respects, which will be discussed later. Our method combines Berkeley results with traditional segmentation results and depth information generated by the latest stereo vision algorithm. There are mainly four steps:

1. Unification of Berkeley boundary. The Berkeley boundary suffers from a great shortcoming that the pixels of the same importance do not have the identical probability. Figs. 2*a*–*c* illustrate the problem. In the boundary map, the probability of each pixel being a boundary is scaled from 0–1 to 0–255. The whiter the pixel appears, the more confident it is a boundary. As is evident from the result, boundaries of the same object or surface do not have the same probability. Based on this deficiency, we employ image segmentation results to improve the Berkeley results. One state-of-the-art

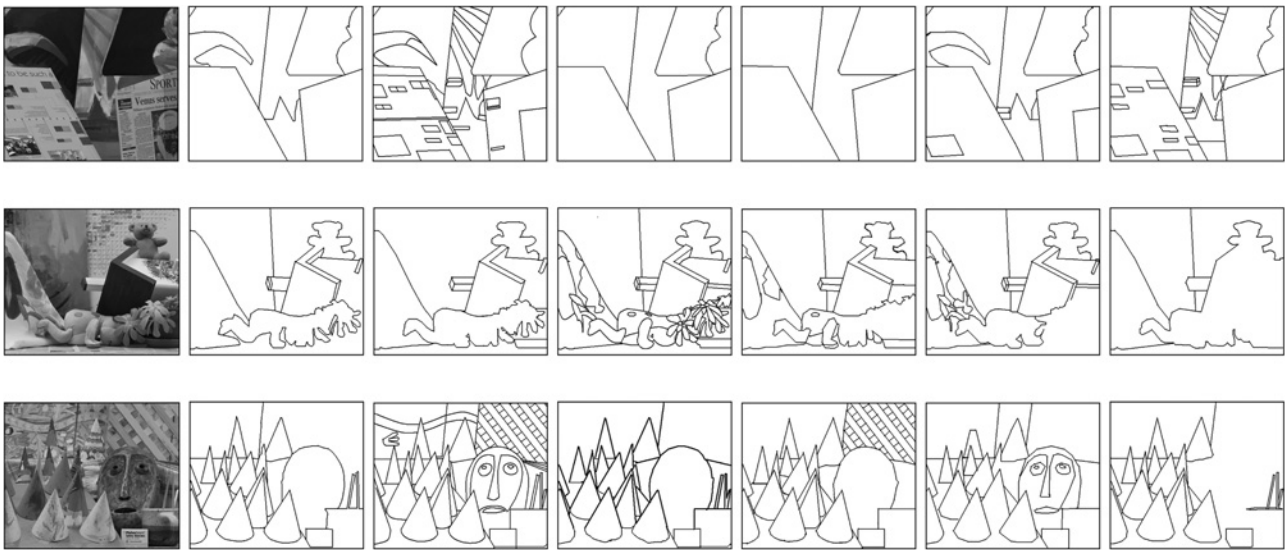

Fig. 1 *Samples of three images (Venus, Teddy and Cones) from our segmentation database*

Left image on each row is the original one and the other ones in the same row are six human segmented boundaries

doi: 10.1049/iet-ipr.2010.0288

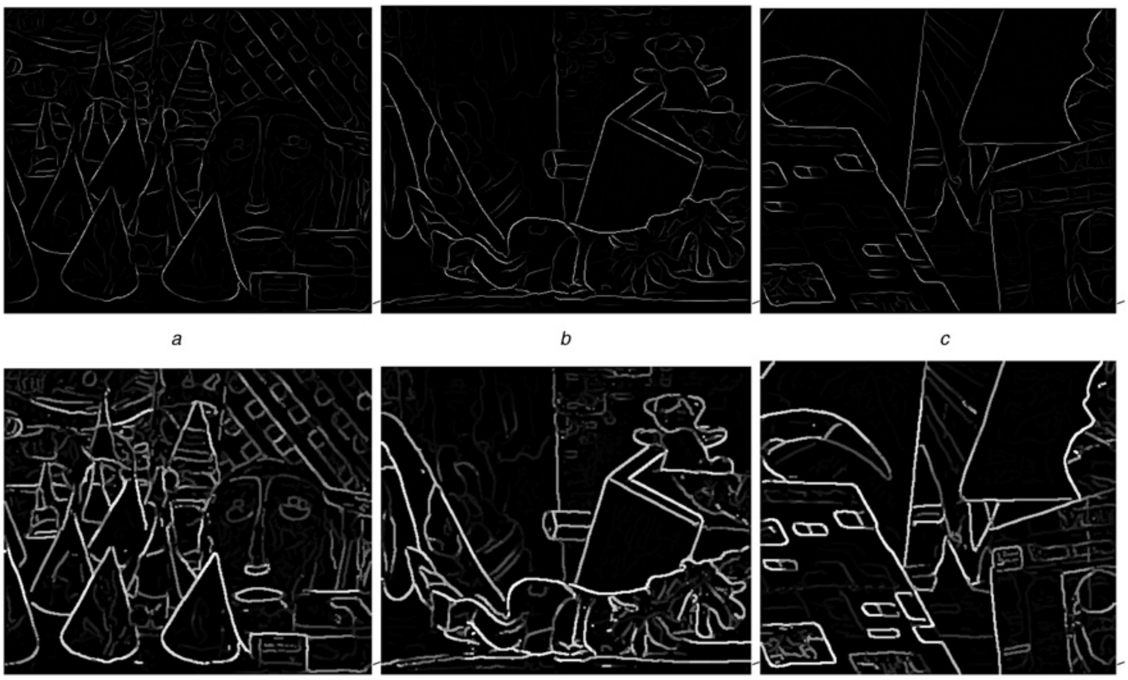

Fig. 2 *Unification illustration of Cones, Teddy and Venus images*

a–c are the original Berkeley boundary
d–f are their corresponding improved results after unification

segmentation algorithm, mean-shift [29], is used to segment the image. Then the segmentation result is taken as a reference to unify the boundary. Our principle is: If the pixels of the Berkeley boundary lie on the same edge of the segment obtained by mean-shift algorithm, their probabilities should be unified by the largest one among them. The segmentation and unification are conducted twice under two sets of parameters, smaller ones and bigger ones. We have tried different parameter combinations to examine its effect on the final results. According to our experimental results, $h_s = 7$, $h_r = 6.5$, $M = 20$ and $h_s = 9$, $h_r = 8.5$, $M = 100$ are the appropriate choice (The meanings of the parameters are explained later.). After that, we expand the one-pixel wide Berkeley boundary to three-pixel in order to evaluate it from our database, whose ground-truth boundary is also three-pixel wide. Figs. 2*d–f* show the corresponding results after unification. We can see clearly that our method improves the Berkeley result a lot. In order to quantitatively justify our conclusion, we threshold the Berkeley boundary and the unified boundary at different levels to compare their corresponding binary map with the ground-truth one. Statistics show that an average of 15% more boundary pixels can be obtained for each image after unification.

2. Probabilistic boundary from segmentation. We will obtain a probabilistic boundary from segmentation results in this step. First, we segment the image by different mean-shift parameter combinations. There are three parameters for the user to specify. The first one h_s, and second h_r, are, respectively, the radius of the spatial dimensions and colour dimensions for gradient estimation. The third one, M (minimum region), controls the number of regions in the segmented image. We determine the plausible meaningful range of each parameter by consulting the original paper and doing a preliminary experiment, through which we can obtain a general idea of the parameters' effects on the algorithm's results. We make sure each parameter samples the entire reasonable parameter space. Our preliminary experiment on dozens of images indicates that the reasonable maximum of the three parameters are, respectively, about 49, 30.5 and 5000. Therefore we specify $7 \times 7 \times 9$ combinations of mean-shift parameters, where $h_s \in \{7, 14, 21, 28, 35, 42, 49\}$, $h_r \in \{6.5, 10.5, 14.5, 18.5, 22.5, 26.5, 30.5\}$ and $M \in \{50, 150, 300, 600, 800, 1000, 2000, 3000, 5000\}$. After segmenting the image under these different parameters, we compute each image pixel's occurrence rate of being a segmented boundary and refer to it as the desired boundary probability. The detected boundary is three-pixel wide and Fig. 3 is an illustration of boundary map from segmentation.

3. Depth boundary. Depth boundary will be integrated into our boundary detection scheme. This is a remarkable difference from other algorithms since we can obtain the depth information by using the image pair in our database and the appropriate stereo matching algorithm ranked on Middlebury website. In this work, one state-of-the-art stereo matching algorithm, which is based on cooperative optimisation [30] and is among the highest ranked algorithm [13] according to the evaluation results based on the known ground-truth depth maps, is employed. The acquired depth map is then input to Canny edge detector to obtain a binary boundary map. There are three main parameters in the Canny algorithm [31], the high threshold, the low threshold and the standard deviation for Gaussian filter. The high threshold is set as the value of the specific pixel's derivative that 70% pixels' derivative in the image are smaller than it. The low threshold is set as the 0.4*high threshold. The standard deviation of Gaussian filter is set as 1. Finally, the boundary is expanded to three-pixel wide. Figs. 4*a–c* show examples of depth map generated by [30] and Figs. 4*d–f* its corresponding boundary results generated by Canny detector.

4. Cue Combination. The above three cues are combined together to generate two kinds of three-pixel wide boundary

doi: 10.1049/iet-ipr.2010.0288

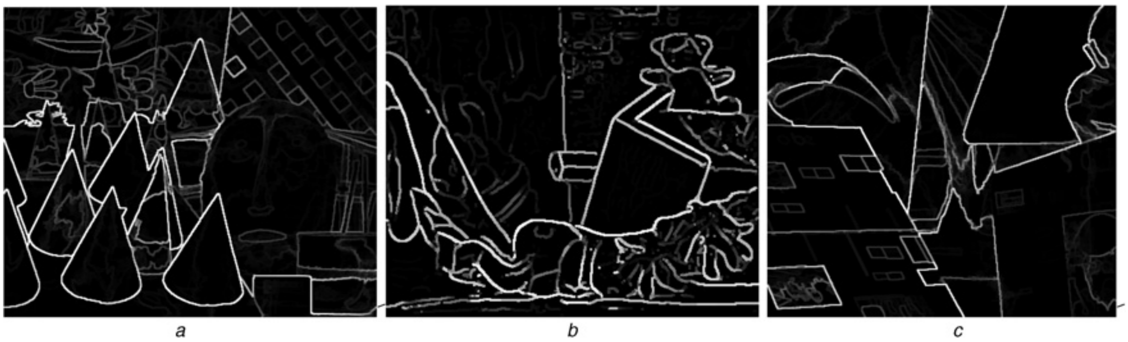

Fig. 3 *Probabilistic boundary obtained from segmentation results*
a–c are, respectively, the results of Cones, Teddy and Venus

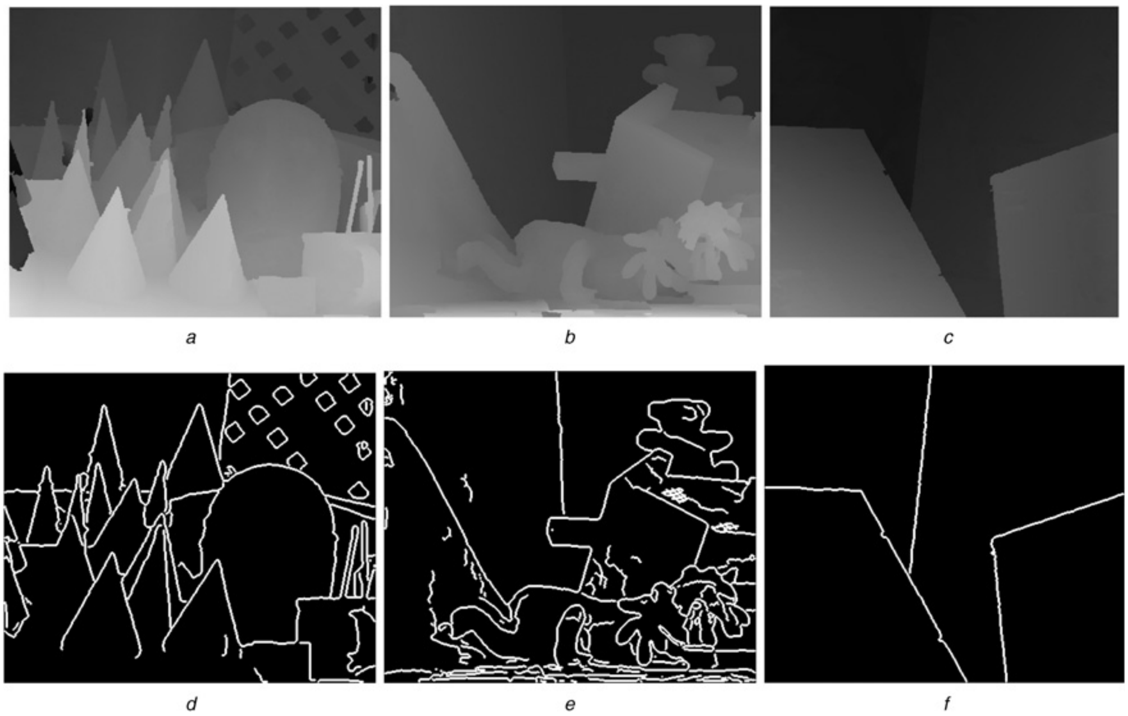

Fig. 4 *Depth map and its corresponding boundary results*
a–c are the depth map of Cones, Teddy and Venus
d–f are their corresponding boundaries obtained by Canny detector

in this step. One is of the same type with Berkeley boundary, which is probabilistic. The other is binary.

For the probabilistic type, the unified Berkeley boundary and segmented probabilistic boundary are averaged first. Then the obtained results are tuned according to the depth boundary. Our process is based on the assumption: Most of the time, depth boundary is the most important one in an image. If the averaged probabilistic boundary coincides with the depth boundary, it should be strengthened; otherwise, it should be weakened. Two regulatory factors are needed to strength or weaken the boundary. According to our experimental results, 1.2 and 0.8 are two appropriate choices. Figs. 5*d–f* show our probabilistic results. Compared with the Berkeley boundary in Figs. 5*a–c*, it is obvious that our detected boundaries are more hierarchical and the boundaries of the same objects or surfaces are more likely the same.

As for the binary type, we detect the boundary by pattern classification approach. Since every pixel in the image is either a boundary pixel or not, this is identically a two-category classification problem. To solve it, we employ our previously proposed piecewise linear classifier [32], which is an improvement for the minimax criterion and an approximation to the theoretical optimum Bayes classifier. The presented classifier consists of three different forms according to the prior intervals. That means the priors are divided into three intervals. When the future estimated prior falls into a particular interval, the specific classifier corresponding to the interval is employed. The decision rule for the proposed classifier is

$$\text{if } \frac{p(\boldsymbol{x}|\omega_1)}{p(\boldsymbol{x}|\omega_2)} > \frac{[1 - P^{\text{o}}(\omega_1)](\lambda_{21} - \lambda_{22})}{P^{\text{o}}(\omega_1)(\lambda_{12} - \lambda_{11})}, \text{ then } \boldsymbol{x} \in \omega_1; \\ \text{otherwise } \boldsymbol{x} \in \omega_2$$

where ω_1 and ω_2 denote the corresponding two categories (boundary and non-boundary), $P^{\text{o}}(\omega_1)$ the prior probability of the specific interval where the estimated prior lies,

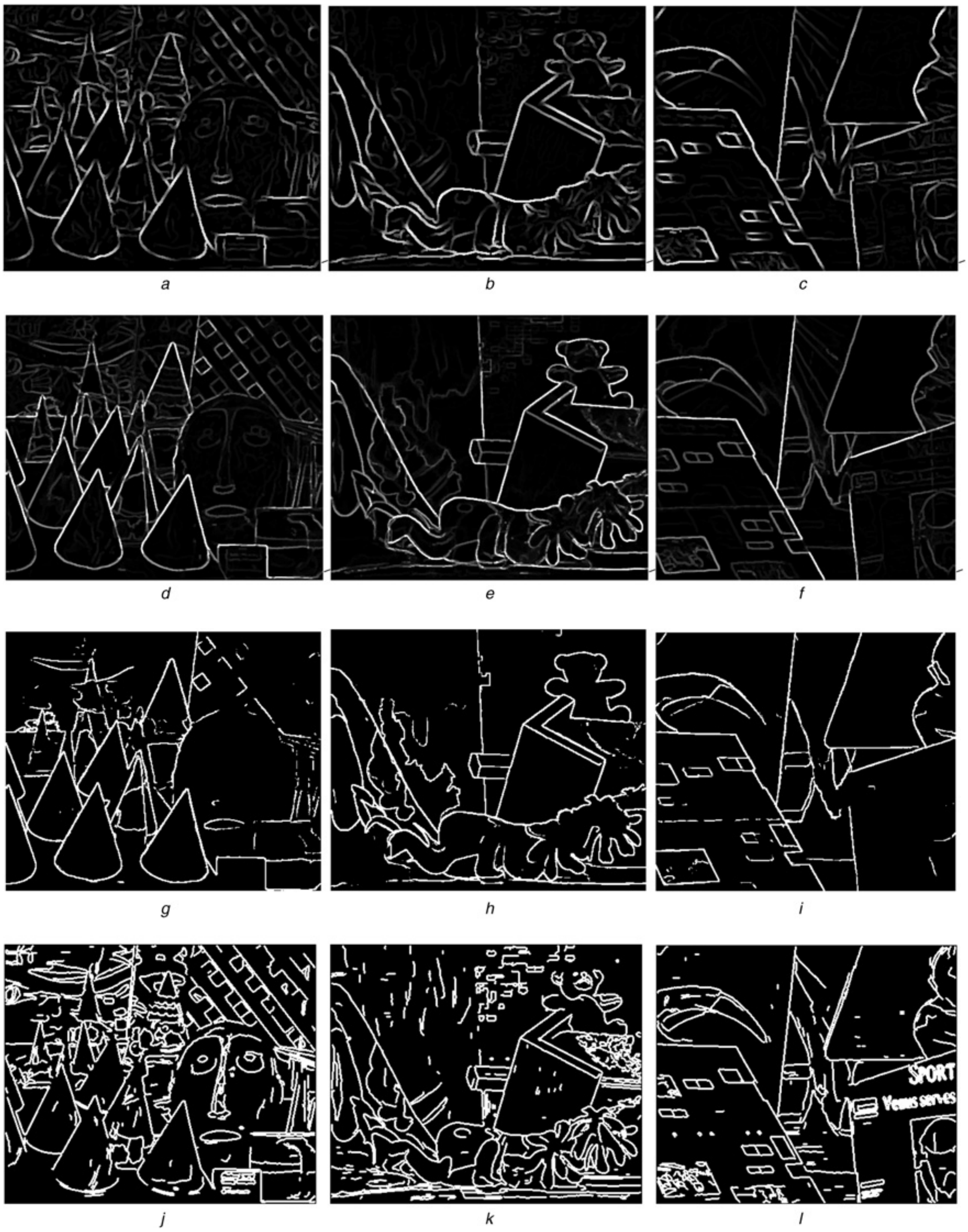

Fig. 5 *Cue combination illustration and comparison with Berkeley and Edison results*

a–c are the original Berkeley boundary of Cones, Teddy and Venus
d–f are the results of our probabilistic type
g–i are the results of our binary type
j–l are the results of Edison detector

$p(\boldsymbol{x}|\omega_i)$ the class-conditional probability density function for $\boldsymbol{x}$ conditioned on ω_i and λ_{ij} be the loss incurred for deciding ω_j when the true state of nature is $\omega_i(i, j = 1, 2)$.

In the decision process, the pixel's probability value obtained from the cue combination is treated as feature vector $\boldsymbol{x}$ and the estimated prior. For the class-conditional probability density function and loss function, the mostly used Gaussian density and zero-one loss function are employed. The Gaussian parameters are determined by maximum likelihood estimation [32]. The prior interval is divided into [0, 1/3],(1/3, 2/3) and [2/3, 1]. Figs. 5*g–i* are the final results.

3.2 Comparison of experimental results

We compare our method with existing methods in this section. Since the detected boundary is of probabilistic and binary type, we present the comparison results, respectively.

For the probabilistic type, boundaries are depicted with a probability of 0–1. To evaluate its hierarchy, thresholds

doi: 10.1049/iet-ipr.2010.0288

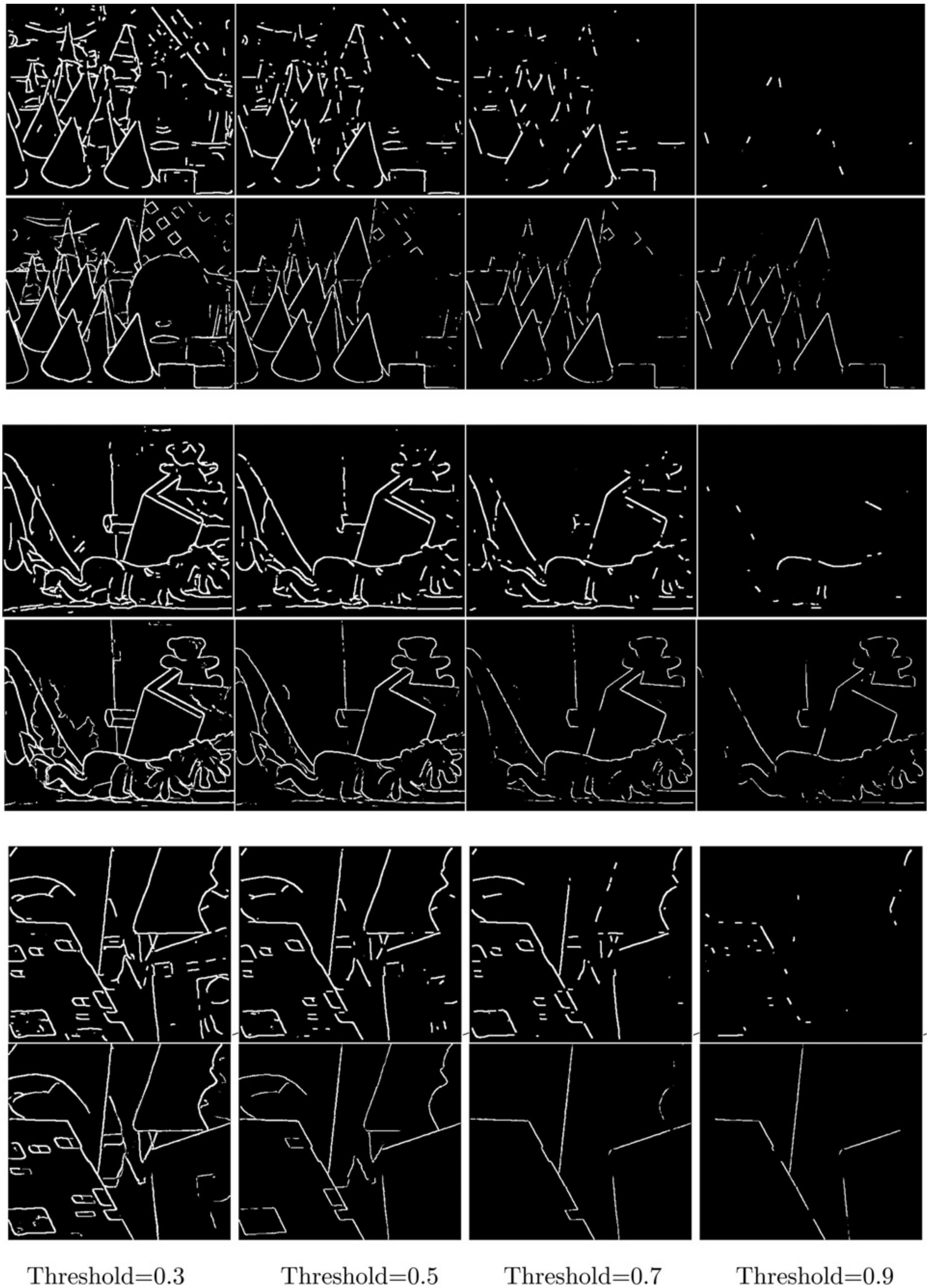

Fig. 6 *First two rows list Cones binary maps from the probabilistic boundary, where the Berkeley results lie on top and our results bottom*

For each row from left to right, the thresholds are, respectively, set to 0.3, 0.5, 0.7 and 0.9. The second two rows and the third two are, respectively, Teddy and Venus images

from low to high are set to generate different binary boundaries at different levels. The higher the threshold is, the more evident boundaries emerge. Unfortunately, the Berkeley boundaries are not satisfying in a hierarchical way. If we set a higher threshold, some desirable boundaries may disappear. Otherwise, if we set a lower

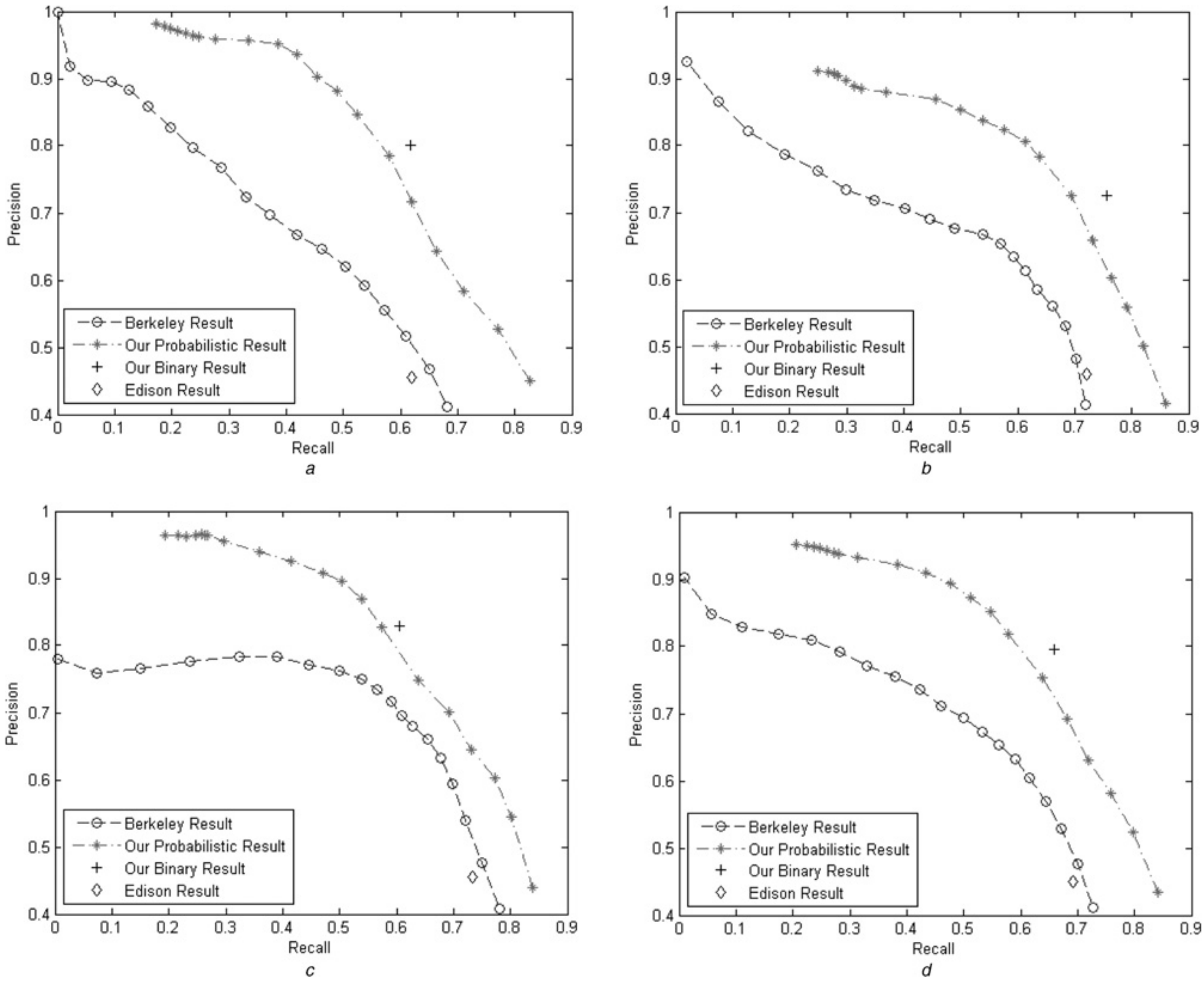

Fig. 7 *Performance of P–R curves for the proposed method*

a–c are, respectively, precision-recall curves of Cones, Teddy and Venus images
d is the curves averaged on all the images of our database

one, many unwanted boundaries may emerge. Fig. 6 shows an example of three images. The first two rows list Cones binary boundary maps, where the Berkeley results lie on top and our results bottom. For each row from left to right, the thresholds are, respectively, set to 0.3, 0.5, 0.7 and 0.9. The second two rows and the third are, respectively, Teddy and Venus images. We can see from Fig. 6 that our results outperform the Berkeley results in a more hierarchy manner.

For the binary type, we compare the results with Edison detector. Figs. 5*g–i* are our binary results and Figs. 5*j–l* are the Edison results. We can see clearly that Edison detector generate more details than our results. However, these details are not needed in our boundary detection process.

In order to quantitatively evaluate our method, we employ the precision–recall curve, which is a standard evaluation technique in the information retrieval community and has been used for evaluating edge detectors [33, 34]. Precision is the fraction of detections that are true positives rather than false positives, while recall is the fraction of true positives that are detected rather than missed. Based on our ground-truth database, we compute the two indexes in a way similar to [3]. We process all the images in our database to compare our results with the Berkeley's and Edison's. Figs. 7*a–c* show the results of three images Cones, Teddy and Venus. Fig. 7*d* is the results averaged on all the images in our database. Since our binary results and Edison results are definitely binary, their *P–R* curves decrease to a point. Fig. 6 tells us that our method can detect much more accurate boundaries than the Berkeley detector and Edison detector.

However, the processing speed of the presented method is much slower than the other two detectors. On the computer with Intel Pentium processor 1.6 GHz, 512MB Memory, the Berkeley detector will cost an average of 3 s, the Edison 0.3 s but the present method 5 s.

4 Conclusion

In this paper, we present a database of images segmented by human subjects along with an application for boundary detection. The images in our database are based on image pairs. This will greatly facilitate the algorithms combining depth information and other low-level cues to detect boundaries or segment images. Besides, we put the database to use by proposing a method for detecting boundaries and compare our results with the Berkeley and Edison detectors. Experiments show our method outperforms the two detectors but for the limited paper length, only a few figural results are demonstrated. Other results that do not appear in this paper conform to the conclusion. Our future work is to make the database available on the website and to employ it to develop a segmentation algorithm.

5 Acknowledgment

This project was supported by the Special Fund for Basic Scientific Research of Central Colleges, Chang'an University, CHD2011JC124

6 References

1 Health, M.D., Sarkar, S., Sanocki, T., Bowyer, K.W.: 'A robust visual method for assessing the relative performance of edge-detection algorithms', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (12), pp. 1338–1359
2 Basu, M.: 'Gaussian-based edge-detection methods-a survey', *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.*, 2002, **32**, (3), pp. 252–260
3 Martin, D., Fowlkes, C.C., Malik, J.: 'Learning to detect natural image boundaries using local brightness, color, and texture cues', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2004, **26**, (5), pp. 530–549
4 Ando, S.: 'Image field categorization and edge/corner detection from gradient covariance', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (2), pp. 179–190
5 Ge, F., Wang, S., Liu, T.: 'Image-segmentation evaluation from the perspective of salient object extraction'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2006, pp. 1146–1153
6 Zhang, H., Fritts, J.E., Goldman, S.A.: 'Image segmentation evaluation: a survey of unsupervised methods', *Comput. Vis. Image Underst.*, 2008, **110**, pp. 260–280
7 Unnikrishnan, R., Pantofaru, C., Hebert, M.: 'Toward objective evaluation of image segmentation algorithms', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2007, **29**, (6), pp. 929–944
8 Shaffrey, C.W., Jermyn, I.H., Kingsbury, N.G.: 'Phychovisual evaluation of image segmentation algorithms'. Proc. Advanced Concepts for Intelligent Vision System, Ghent, Belgium, 2002
9 http://homepages.inf.ed.ac.uk/amos/afreightdata.html
10 http://www.isi.uu.nl/Research/Databases/SCR/
11 http://archive.ics.uci.edu/ml/datasets/Statlog+(Image+Segmentation)
12 Martin, D., Fowlkes, C., Tal, D., Malik, J.: 'A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics'. Proc. Eighth IEEE Int. Conf. Computer Vision, 2001, pp. 416–423
13 http://vision.middlebury.edu/stereo/eval/
14 Benlamri, R., Al-Marzooqi, Y.: 'Free-form object segmentation and representation from registered range and color images', *Image Vis. Comput.*, 2004, **22**, (9), pp. 703–717
15 Ntalianis, K.S., Doulamis, N.D., Doulamis, A.D., Kollias, S.D.: 'A feature point based scheme for unsupervised video object segmentation in stereoscopic video sequences'. Proc. IEEE Int. Conf. Multimedia and Expo, 2000, pp. 1543–1546
16 http://vision.middlebury.edu/stereo/data/
17 Stein, S., Hoiem, D., Hebert, M.: 'Learning to find object boundaries using motion cues'. Proc. IEEE Int. Conf. on Computer Vision, 2007, pp. 1–8
18 Elder, J.H., Zucker, S.W.: 'Computing contour closure'. Proc. fourth European Conf. on Computer Vision, 1996, pp. 1611–3349
19 Felzenszwalb, P., McAllester, D.: 'A min-cover approach for finding salient curves'. Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition Workshop, Washington, DC, USA, 2006, pp. 185–185
20 Jermyn, I.H., Ishikawa, H.: 'Globally optimal regions and boundaries as minimum ratio weight cycles', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, pp. 1075–1088
21 Ren, X.F., Fowlkes, C.C., Malik, J.: 'Learning probabilistic models for contour completion in natural images', *Int. J. Comput. Vis.*, 2008, **77**, pp. 47–63
22 Berg, A.C., Berg, T.L., Malik, J.: 'Shape matching and object recognition using low distortion correspondences'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2005, pp. 26–33
23 Ferrari, V., Fevrier, L., Jurie, F., Schmid, C.: 'Groups of adjacent contour segments for object detection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2008, **30**, (1), pp. 36–51
24 Leordeanu, M., Hebert, M., Sukthankar, R.: 'Beyond local appearance: category recognition from pairwise interactions of simple features'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2007, pp. 1–8
25 Rotem, O., Greenspan, H., Goldberger, J.: 'Combining region and edge cues for image segmentation in a probabilistic gaussian mixture framework'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2007, pp. 1–8
26 Micusık, B., Hanbury, A.: 'Automatic image segmentation by positioning a seed'. Proc. European Conf. on Computer Vision, 2006, vol. 3952, pp. 1611–3349
27 He, X.M., Zemel, R.S., Ray, D.: 'Learning and incorporating top-down cues in image segmentation'. Proc. European Conf. on Computer Vision, 2006, pp. 1611–3349
28 Meer, P., Georgescu, B.: 'Edge detection with embedded confidence', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, pp. 1351–1365
29 Comaniciu, D., Meer, P.: 'Mean shift: a robust approach toward feature space analysis', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (5), pp. 603–619
30 Wang, Z.F., Zheng, Z.G.: 'A region based stereo matching algorithm using cooperative optimization'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2008, pp. 1–8
31 Canny, J.: 'A computational approach to edge detection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1986, **8**, (6), pp. 679–698
32 Wang, Q., Wang, Z.F.: 'Classifier design method based on piecewise linearization', *Pattern Recognit. Artif. Intell.*, 2009, **22**, (2), pp. 214–222
33 Abdou, I., Pratt, W.: 'Quantitative design and evaluation of enhancement/thresholding edge detectors', *Proc. IEEE*, 1979, **67**, pp. 753–763
34 Bowyer, K., Kranenburg, C., Dougherty, S.: 'Edge detector evaluation using empirical roc curves'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1999, vol. 1, pp. 354–359