

# A survey of occlusion detection method for visual object<sup>①</sup>

Zhang Shihui (张世辉)<sup>②\*</sup>, He Huan\*, Liu Jianxin\*, Zhang Yucheng\*, Pang Yunchong\*, Sang Yu\*

(\* School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, P. R. China)

(\*\* The Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao 066004, P. R. China)

## Abstract

Occlusion problem is one of the challenging issues in vision field for a long time, and the occlusion phenomenon of visual object will be involved in many vision research fields. Once the occlusion occurs in a visual system, it will affect the effects of object recognition, tracking, observation and operation, so detecting occlusion autonomously should be one of the abilities for an intelligent vision system. The research on occlusion detection method for visual object has increasingly attracted attentions of scholars. First, the definition and classification of the occlusion problem are presented. Then, the characteristics and deficiencies of the occlusion detection methods based on the intensity image and the depth image are analyzed respectively, and the existing occlusion detection methods are compared. Finally, the problems of existing occlusion detection methods and possible research directions are pointed out.

**Key words:** visual object, occlusion detection, intensity image, depth image

## 0 Introduction

Occlusion is not only a geometric phenomenon but also an optical phenomenon. Occlusion can be seen everywhere in the real world. Most vision research fields such as object recognition, 3D reconstruction, object tracking, visual observation, robot grasping, automatic assembly, spacecraft docking etc. involve occlusion. Occlusion makes the visual objects which are actually from different physical surfaces not distinguishable. If the visual system could not take proper measures to deal with occlusion, most computer vision technologies would be invalid even wrong, and the corresponding visual system would lose its value. Thus, the occlusion has become an obstacle factor in studying the related fields further, and it increasingly attracts the attention of scholars. Currently, CMU's professor Martial Hebert, the famous scholar in the field of computer vision and other scholars have already launched a study for the occlusion problem. In addition, in recent years, papers about occlusion problem have also a significant place in the top conferences on computer vision, such as the CVPR, ICCV and ECCV.

The existing literatures on occlusion detection are

analyzed in this paper, the research status of the occlusion detection method for visual object is introduced in detail, and the future research direction in order to provide some help for the researchers in this field is discussed and three sections are included. The related definition and classification for occlusion problem are shown in Section 1. Research status of occlusion detection method is explained and analyzed in Section 2. Section 3 concludes the paper.

## 1 Definition and classification of occlusion

### 1.1 Definition of occlusion

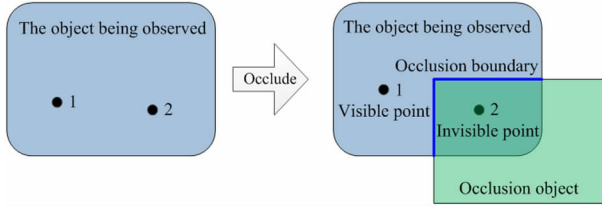
Occlusion mean that all or partial regions of an observed object (visual object) are occluded by other object or some parts of the observed object itself at a certain view, which leads to some regions that can't be observed by the observer (camera), that is to say, the observer can't obtain the information of the occluded regions from current view. So far, there has been no generally accepted definition of occlusion. Fig.1 shows an illustration of occlusion.

As shown in Fig.1, when the occlusion phenomenon occurs, point 1 is visible, point 2 is invisible, and

① Supported by the National Natural Science Foundation of China (No. 61379065) and Natural Science Foundation of Hebei Province (No. F2014203119).

② To whom correspondence should be addressed. E-mail: sshhzz@ysu.edu.cn

Received on June 16, 2015



**Fig. 1** An illustration of occlusion

the dividing line between the visible part and the invisible part is called occlusion boundary. Let  $P = \{p_1, \dots, p_n\}$  denote the set of all points on an object,  $V = \{v_1, \dots, v_n\}$  denotes the set of visibility for all points in  $P$  from current view, and let  $v_i = 0, 1$  or  $2$ . Take any  $p_i \in P$ , then  $v_i \in V$  denotes the visibility of  $p_i$ . Among them, if  $v_i = 0$ , then  $p_i$  is an invisible point; if  $v_i = 1$ , then  $p_i$  is a visible point and an occlusion boundary point; if  $v_i = 2$ , then  $p_i$  is a visible point but not an occlusion boundary point. On this basis, occlusion boundary can be expressed as a set  $E_o = \{p_i \mid p_i \in P, v_i \in V \text{ and } v_i = 1\}$ . Given an object in a scene, if its corresponding set  $V$  has the element of the value of  $0$ , then it shows that the occlusion occurs from current view.

## 1.2 The classification of occlusion problem

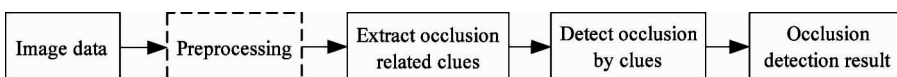
According to the different occurring scope of occlusion, the occlusion problem can be classified into two categories. One is the mutual occlusion between different objects, and the other is the self-occlusion between the multiple parts of a same object. Essentially, they are the same: one object (part) occludes another. Meanwhile, for either mutual occlusion or self-occlusion, pixels in the occlusion region have the same characteristics, so from the perspective of researching occlusion processing method, mutual occlusion and self-occlusion can be dealt together. In addition, according to spatio-temporal attribute difference of occlusion, occlusion problem can be classified into static occlusion and dynamic occlusion. If the occlusion doesn't change with time, it is called static occlusion, otherwise it is called dynamic occlusion. No matter what occlusion it is, mutual or self, and no matter what state it is, static or dynamic, once the occlusion occurs in a visual system, it will affect the operation effects of object recognition, reconstruction, tracking and observation, etc.

## 1.3 The classification of occlusion problem processing method

From the low-level image processing technique to the high-level reasoning task, occlusion problem involves a very wide range of fields. By now, there have been two main solutions to the influence by occlusion. One regards the occlusion as noise, which reduces the interference of occlusion by studying the better robustness method. The other takes the occlusion as processing object, which deals with the occlusion directly by studying the occlusion detection method. For the first way, the robustness of the method can't increase indefinitely, the related method will fail when the occlusion is more serious. But for the second way, the researcher can adopt different strategies according to different occlusion detection results, so it is more general and flexible.

Aiming at occlusion in tracking<sup>[1,2]</sup>, recognition<sup>[3,4]</sup>, segmentation<sup>[5,6]</sup>, 3D reconstruction<sup>[7,8]</sup>, matching<sup>[9,10]</sup>, real-time scene rendering<sup>[11,12]</sup> etc., scholars at home and abroad have proposed corresponding solutions. These methods ensure the robustness and accuracy of related visual task processing algorithm in the condition that occlusion exists in visual object, but they don't provide the clear occlusion detection methods. Therefore, these methods can be considered as the first way to solve the occlusion problem. With the in-depth research of visual information processing technology, occlusion problem becomes an independent research direction, and attracts attentions of scholars at home and abroad. These independent occlusion detection methods can be considered as the second way to solve the occlusion problem. Fig. 2 shows the general flow of occlusion detection by this way.

The image data in Fig. 2 can be a video, multiple images or a single image, and the image can be an intensity image (including gray image and color image) or a depth image. Preprocessing is an optional process, and it may be motion analysis based on optical flow or image segmentation technology. Through the former, the optical flow field or the motion vector field of the visual object can be got, then the motion clues can be obtained, while through the latter, the boundary set which consists of boundaries between various regions by image segmentation can be obtained. Here, the boundary set is generally required so that it can include all



**Fig. 2** General flow of occlusion detection

the occlusion boundaries in the image, on this basis, subsequent processing is processed to determine the occlusion boundary further. Occlusion related clues (features) can be motion clues, appearance clues, depth clues, etc. According to the extracted occlusion related clues, there are two main ideas about designing occlusion detection method. One is determining the occlusion detection results directly by designing clear occlusion discrimination criterion (or occlusion model) and using the threshold, convex optimization and other methods. In this kind of method, the extracted clues are used directly. The other is regarding occlusion detection problem as a classification problem, which designs occlusion detection method based on learning idea. In this kind of method, the implicit occlusion model can be obtained by learning the training data, then determine the final occlusion detection results by the implicit occlusion model. Compared with the former, this method is less sensitive to the actual image data, and has better applicability.

If the image data, the preprocessing, the extracting occlusion related clues and the detecting occlusion in Fig.2 are denoted as  $P = \{p_1, \dots, p_n\}$ ,  $O_{pre} = f_{pre}(P)$ ,  $O_{clue} = \sum_{j=1}^m f_{j\_extract}(O_{pre})$  and  $O_{result} = f_{detect}(O_{clue})$  respectively, then the formal description of occlusion detection can be expressed as

$$O_{result} = f_{detect}\left(\sum_{j=1}^m f_{j\_extract}(f_{pre}(P))\right) \quad (1)$$

where,  $O_{result}$  is the final result of the occlusion detection,  $f_{detect}$  is the occlusion detection method,  $m$  is the number of the kinds of methods for obtaining occlusion related clues,  $f_{j\_extract}$  is the  $j$ th way for obtaining occlusion related clues,  $f_{pre}$  is the preprocessing method.

## 2 Research status analysis of occlusion detection method

Occlusion detection method has got some achievements so far. On the whole, according to the different processing objects of the algorithm, it can be classified into the occlusion detection method based on intensity image and the occlusion detection method based on depth image. The occlusion detection method based on intensity image can be further classified into occlusion detection method based on video or image sequence and occlusion detection method based on a single image. According to the difference of the result, occlusion detection method based on video or image sequence can be further classified into occlusion region detection method and occlusion boundary detection

method. The specific research status analysis for each kind of occlusion detection method is as follows.

### 2.1 Occlusion detection method based on intensity image

#### 2.1.1 Occlusion detection method based on video or image sequence

##### (1) Occlusion region detection method

Occlusion region in the video or image sequence is the region consisting of pixels which are visible from current frame while the corresponding pixels from the next frame are invisible. Occlusion region detection method principally uses mismatch of corresponding pixels from temporally adjacent images. Depommier<sup>[13]</sup> is one of the earlier scholars who detected occlusion region based on motion clues between the video frame, and Depommier detected occlusion based on photometric mismatch between consecutive frame in the video by combining motion estimation method. Kolmogorov<sup>[14]</sup> proposed an occlusion detection method based on graph cut. The method is specially designed to detect occlusion in stereo image pairs, so it is not intended for general scenes. Ince<sup>[15]</sup> proposed an occlusion detection method based on geometry estimation, and this method determined occlusion region only by analyzing motion field. Ayvaci<sup>[16]</sup> regarded the occlusion detection problem as a variational optimization problem and used the efficient numerical algorithm to solve global optimum solution to determine the occlusion region. The above mentioned methods are mainly based on motion analysis, and detect occlusion regions by clear occlusion discrimination criterion or occlusion model. With the deepening of research, some scholars regarded the occlusion detection problem as a classification problem, and they had designed the occlusion detection method based on learning idea. Humayun<sup>[17]</sup> extracted a broad spectrum of features and trained a classifier to detect occlusion region. Compared with the occlusion detection method based on motion analysis, the accuracy and adaptability of this occlusion detection method have been improved, but extracting a large number of features requires long computation time. Zhang<sup>[18]</sup> combined learning idea and graph cut to detect occlusion region. The features extracted in this method are more efficient, so compared with existing methods, it has higher accuracy and better real-time performance.

##### (2) Occlusion boundary detection method

Occlusion boundary is a subset of the boundary set which consists of boundaries between various regions by image segmentation. The main basis for detecting occlusion boundary is the discontinuity of optical flow or motion on both sides of the occlusion boundary, and

the discontinuity mainly reflects two aspects: one is the discontinuity of brightness information; the other is the discontinuity of motion vector information. Stein<sup>[19]</sup> proposed an occlusion detection method using a spatio-temporal edge detection operator, which determined the occlusion boundary by analyzing and comparing the motion of the extracted small region in both sides of the edge. Overall, the occlusion boundary detection methods are mainly based on learning idea, and the clues (features) extracted in these methods include motion clues, local appearance clues, depth clues, T-junctions features, etc. Apostoloff<sup>[20]</sup> extracted T-junctions features by using spatio-temporal characteristics in video, and proposed an occlusion detection method based on relevance vector machine (RVM). Sargin<sup>[21]</sup> proposed a probability detection framework based on spatio-temporal lattices to detect occlusion boundary by defining two complementary cost functions. By combining motion clues and local appearance clues, Stein<sup>[22]</sup> proposed an occlusion detection method based on learning. On this basis, He<sup>[23]</sup> introduced the depth clues. He estimated the pseudo-depth information in video, and used pseudo-depth information, motion clues and local appearance clues to train a classifier for detecting occlusion boundary. However, there are two main deficiencies for the method based on learning. One is that the effective sample data is needed to train classifier, and the other is that the method easily leads to the phenomenon of over-fitting. Aiming at the above deficiencies, Jacobson<sup>[24]</sup> proposed an online learning framework for detecting occlusion boundary. Compared with the previous methods, its applicability has been improved in a way, but the accuracy of the detection result needs to be improved.

In view of comprehensive analysis, since most methods use information of optical flow field or motion vector field, any error in the optical flow field or motion vector field may lead to inaccurate occlusion detection result, and the occlusion detection method based on the motion discontinuity implicitly requires that the movement in the video should be rigid transformation (rotation, translation, or their combinations). Moreover, for the occlusion detection method using optical flow estimation, the following two assumptions are required. One is that the brightness of any point in the observed object is constant with time, and the other is that proximal points in the image plane move in a similar way. In addition, the proposed methods in Refs[21-23] require the over-segmentation of the video frame, so the accuracy of the detection result is largely dependent on the segmentation effect.

## 2.1.2 Occlusion detection method based on a single intensity image

Compared with occlusion detection study based on video or image sequence, because the available information in a single intensity image is relatively less, the occlusion detection study based on a single intensity image is more difficult<sup>[19]</sup>, and the achievements are fewer. Ullmann<sup>[25]</sup> is one of the earlier scholars who studied occlusion detection method based on a single intensity image. Assuming the case is in no noise, Ullmann detected the occlusion by determining the mismatched pixels between the template and object image. Gay-Bellile<sup>[26]</sup> presented a specific framework which solved occlusion problem during 2D image non-rigid registration. The limitations of these methods based on matching are that they require the priori knowledge of the visual object, therefore these methods are not suitable for the unknown visual object. Hoiem<sup>[27]</sup> used hierarchical region segmentation method to estimate the depth of the object in a scene and then determined the occlusion boundary by the depth value. The deficiency of this method is that the region segmentation is its foundation, and the depth estimation result also directly affects the accuracy of the occlusion detection result. Chen<sup>[28]</sup> utilized five proposed occlusion clues to understand the occlusion relationship in an image. The main limitation of this method is that the semantic annotation image must be obtained in advance, and therefore the semantic annotation result would affect the occlusion detection result directly. Zhang<sup>[29]</sup> explored an occlusion boundary recovery method from the perspective of depth estimation. Although it is difficult, the result is ideal. The deficiency of this method is that the effect of occlusion detection is largely dependent on the depth estimation accuracy.

Overall, the occlusion detection method based on intensity image still needs to be improved further. Meanwhile, the depth information has been preliminarily applied in the occlusion detection method based on the intensity image. For example, Ref. [23] improved the accuracy of occlusion detection result by combining the pseudo-depth information; Ref. [27] used estimated depth information for determining the occlusion relationship between regions in the image more accurately. Although the Ref. [23] and Ref. [27] combine the depth information to detect the occlusion, the depth information is not from the depth image but by imprecise estimating, which makes these methods not having used the depth information enough.

2.2 Occlusion detection method based on depth image

Depth image is a kind of image with special form, and the difference between the depth image and the common intensity image is the content which the pixels represent. For the intensity image, the pixel value represents the intensity or gray level of the point; while for the depth image, the pixel value indicates the distance from the point to the reference plane. As the data of depth image directly reflect the space geometry information of the visual object surface, therefore, the visual information processing technology based on depth image is drawing more and more attention in academia and industry.

At present, in terms of depth image information processing, scholars focus on the study of registration<sup>[30,31]</sup>, segmentation<sup>[32,33]</sup>, estimation<sup>[34,35]</sup>, edge detection<sup>[36]</sup> and application<sup>[37]</sup> etc. In recent years, some scholars abroad have tried to study the approach for specific visual task under occlusion<sup>[38,39]</sup> based on the depth image information of visual object, but they have not provided the clear occlusion detection meth-

ods. Domestic researchers also have done a lot of work for occlusion detection method based on the depth image, and through research, exploration and accumulation in recent years, they have achieved preliminary results<sup>[40-45]</sup>. In Ref. [40] and Ref. [41] in 2010, features based on depth image such as mean curvature, curvature threshold and depth threshold were adopted to detect occlusion. In Ref. [42] in 2011, the authors detected occlusion successfully by using the depth difference and its threshold. The three methods mentioned above use threshold segmentation to detect occlusion, but they can't select threshold automatically, which leads to worse generality, so the accuracy of occlusion detection result in Refs[40-42] still needs to be improved. To further explore other possible occlusion detection method, in Ref. [43] in 2012, the authors studied the occlusion detection problem based on depth image from the perspective of classification by utilizing the maximal depth difference feature and included angle feature. In 2014, three new occlusion related features named eight neighborhood total depth difference feature, maximal area feature and depth dispersion feature are proposed in Refs[44,45], and the

Table 1 Comparison of different occlusion detection methods

Method	Year	Image information				Occlusion type				Experimental image or dataset
		Intensity info	Depth info	Single image	Multiple images	Mutual occlusion	Self-occlusion	Static occlusion	Dynamic occlusion	
Depommier <sup>[13]</sup>	1992	✓			✓	✓	✓		✓	Synthetic image and real scene
Ullmann <sup>[25]</sup>	1992	✓		✓		✓		✓		Program generation
Kolmogorov <sup>[14]</sup>	2001	✓			✓	✓	✓	✓	✓	Tsukuba dataset etc.
Ince <sup>[15]</sup>	2005	✓			✓	✓	✓		✓	Flowergarden etc.
Apostoloff <sup>[20]</sup>	2005	✓			✓	✓	✓		✓	Flowergarden etc.
Stein <sup>[19]</sup>	2009	✓			✓	✓	✓		✓	Optional
Sargin <sup>[21]</sup>	2009	✓			✓	✓	✓		✓	CMU Motion Dataset
Stein <sup>[22]</sup>	2009	✓			✓	✓	✓		✓	CMU data set
He <sup>[23]</sup>	2010	✓	✓		✓	✓	✓		✓	CMU data set
Gay-Bellile <sup>[26]</sup>	2010	✓		✓			✓	✓		Optional
Zhang <sup>[40]</sup>	2010		✓	✓			✓	✓		Stuttgart Range Image Database
Zhang <sup>[41]</sup>	2010		✓	✓			✓	✓		Stuttgart Range Image Database
Humayun <sup>[17]</sup>	2011	✓			✓	✓	✓		✓	Middlebury dataset etc.
Hoiem <sup>[27]</sup>	2011	✓	✓	✓		✓		✓		Geometric Context dataset
Zhang <sup>[42]</sup>	2011		✓	✓			✓	✓		Stuttgart Range Image Database
Ayvaci <sup>[16]</sup>	2012	✓			✓	✓	✓		✓	Middlebury dataset etc.
Jacobson <sup>[24]</sup>	2012	✓			✓	✓	✓		✓	CMU data set
Zhang <sup>[29]</sup>	2012	✓	✓	✓		✓		✓		Geometric Context dataset
Zhang <sup>[43]</sup>	2012		✓	✓			✓	✓		Stuttgart Range Image Database
Chen <sup>[28]</sup>	2013	✓		✓		✓		✓		Geometric Context dataset etc.
Zhang <sup>[44]</sup>	2014		✓	✓		✓	✓	✓		Stuttgart Range Image Database
Zhang <sup>[45]</sup>	2014		✓	✓		✓	✓	✓		Stuttgart Range Image Database
Zhang <sup>[18]</sup>	2015	✓			✓	✓	✓		✓	Middlebury dataset etc.

AdaBoost classifier and the Random Forest classifier are used to detect the occlusion boundary in depth image respectively. Although these methods have some generality, and in a way improve the detection accuracy, the adaptability and classification models are still insufficient.

Overall, the current research at home and abroad on the occlusion detection method based on depth image is not deep enough, and it is conjectured that the depth information may contain many valuable features to be further explored.

### 2.3 Comparison of different occlusion detection methods

Table 1 shows the comparison of different occlusion detection methods.

To facilitate further research on occlusion problem, the quantitative comparison results among different occlusion detection methods for all kinds of images are provided in the following tables. When the experiment results are analyzed,  $\{x_{gt}\}$  denotes the set of oc-

clusion pixels in the Ground Truth and its size of  $n_{gt}$ .  $\{x_d\}$  denotes the set of occlusion pixels detected by occlusion detection methods and its size of  $n_d$ .  $\{x_{gt}\} \cap \{x_d\}$  denotes the set of correct occlusion pixels detected by occlusion detection methods and its size is  $n_m$ . Then, precision  $R_p = n_m/n_d$ , recall  $R_r = n_m/n_{gt}$ , and  $F-score = 2R_pR_r/(R_p + R_r)$ . Average precision  $R_{ap}$  denotes the average of all the precisions for the test data, average recall  $R_{ar}$  denotes the average of all the recalls for the test data, and average  $F-score$  denotes the average of all the  $F-scores$  for the test data.

Table 2 shows the average precision  $R_{ap}$ , average recall  $R_{ar}$ , average  $F-score$  and average time-consumption for three representative occlusion region detection methods based on image sequence. Because Middlebury dataset is widely adopted in occlusion region detection and AlgoSuit dataset includes more comprehensive occlusion cases, the image sequences in Middlebury dataset and AlgoSuit dataset are used for testing these methods in the experiment.

Table 2  $R_{ap}$ ,  $R_{ar}$ , average  $F-score$  and average time-consumption for different occlusion region detection methods

Method	$R_{ap}(\%)$	$R_{ar}(\%)$	Average $F-score$	Average time-consumption (s)
Ayvaci <sup>[16]</sup>	74.30	64.68	0.6724	1015.5
Humayun <sup>[17]</sup>	73.63	85.29	0.7684	6066.5
Zhang <sup>[18]</sup>	74.43	86.22	0.7795	48.9

In Table 2, Ayvaci<sup>[16]</sup> is based on motion analysis, while Humayun<sup>[17]</sup> and Zhang<sup>[18]</sup> are based on learning idea. As shown in Table 2, the occlusion region detection results are better than the result by Ayvaci<sup>[16]</sup>. Because of requiring long computation time for extracting a large number of features, the average time-consumption in Humayun<sup>[17]</sup> is much longer than the others. Overall, the method in Zhang<sup>[18]</sup> outperforms the method in Humayun<sup>[17]</sup>. It should be noted that the sample data is needed to train classifier in Humayun<sup>[17]</sup> and Zhang<sup>[18]</sup>, while Ayvaci<sup>[16]</sup> doesn't require any training. So the method in Zhang<sup>[18]</sup> is better than others when the sample data is provided, while the method in Ayvaci<sup>[16]</sup> is more suitable for the case without sample data.

Table 3 shows the average precision  $R_{ap}$  and  $F-score$  for three representative occlusion boundary detection methods based on video. The CMU dataset, which is widely adopted in occlusion boundary detection field, is used for testing these methods in the experiment.

In Table 3, Stein<sup>[22]</sup> and He<sup>[23]</sup> depend on different clues and combine the classifier to detect occlusion boundary. Jacobson<sup>[24]</sup> detects occlusion boundary by

Table 3  $R_{ap}$  and average  $F-score$  for different occlusion boundary detection methods

Method	$R_{ap}(\%)$	$F-score$
Stein <sup>[22]</sup>	39.60	0.4800
He <sup>[23]</sup>	34.50	0.4700
Jacobson <sup>[24]</sup>	39.90	0.4386

updating a set of weights for the online learning. As shown in Table 3, the occlusion boundary detection results by Stein<sup>[22]</sup> and He<sup>[23]</sup> are better than the result by Jacobson<sup>[24]</sup> according to the comprehensive index  $F-score$ . But the sample data is needed to train classifier in Stein<sup>[22]</sup> and He<sup>[23]</sup>, while Jacobson<sup>[24]</sup> uses self-learning idea and doesn't require training. So the methods in Stein<sup>[22]</sup> and He<sup>[23]</sup> are better than the method in Jacobson<sup>[24]</sup> when the sample data can be obtained, while the method in Jacobson<sup>[24]</sup> is more suitable for the case of obtaining the sample data not easily.

Table 4 shows the quantitative comparison results for representative occlusion detection methods based on a single intensity image. Min depth estimation and max depth estimation are two methods used in Hoiem<sup>[27]</sup>. The Geometry Context dataset consisting of a wide variety of scenes is used for testing these methods in the

experiment. Firstly, standing object region is classified into two cases with visible ground-contact points and without visible ground-contact points. Secondly, because there is no Ground Truth for occlusion detection based on a single intensity image, a person is asked to observe results for these methods to decide which method's performance is most accurate. Finally, these methods are compared quantitatively though the percentage of the corresponding method performs best in all the methods for each kind of region over all the testing images.

Table 4 The quantitative comparison results for occlusion detection methods based on a single intensity image

Method	Percentage for ground-contact region (%)	Percentage for ground-uncontact region (%)
Min depth estimation <sup>[27]</sup>	7.7	22.5
Max depth estimation <sup>[27]</sup>	9.3	13.8
Zhang <sup>[29]</sup>	83.0	63.7

In Table 4, both Hoiem<sup>[27]</sup> and Zhang<sup>[29]</sup> use hierarchical region segmentation method to estimate the depth of the object in a scene and then determine the occlusion boundary by the depth value. Compared with the method in Ref. [29], the methods in Ref. [27] don't consider the depth change in the region, so they are easy to go wrong while marking the occlusion relation between adjacent regions. Moreover, for the region without visible ground-contact points, the methods in Ref. [27] take the depth of a certain adjacent region as its depth, so some occlusion boundaries will be lost. As shown in Table 4, the methods in Ref. [27] are not as good as the method in Ref. [29].

Table 5 shows the average precision  $R_{ap}$ , average recall  $R_{ar}$  and  $F$ -score for all the occlusion boundary detection methods based on depth image. The Stuttgart Range Image Database (SRID) is utilized for testing these methods in the experiment.

Table 5  $R_{ap}$ ,  $R_{ar}$  and average  $F$ -score for different occlusion detection methods based on depth image

Method	$R_{ap}$ (%)	$R_{ar}$ (%)	$F$ -score
Method in Ref. [40]	27.39	26.38	0.2688
Method in Ref. [41]	51.91	52.57	0.5224
Method in Ref. [42]	52.28	55.98	0.5407
Method in Ref. [43]	95.48	96.95	0.9621
Method in Ref. [44]	96.64	97.41	0.9702
Method in Ref. [45]	97.40	97.26	0.9733

For the methods in Table 5, threshold segmentation is used to detect occlusion in Refs[40-42], and the methods in Refs[43-45] are based on learning. As the threshold segmentation methods in Refs[40-42] can't select threshold automatically, which leads to worse generality, so the occlusion detection results of Refs[40-42] are obviously worse than the results of Refs[43-45]. Ref. [43] uses the method by training classifier to detect occlusion in depth image for the first time, which makes the result better than the result of threshold segmentation in essence. Ref. [44] adds the process of feature selection and meanwhile optimizes the above process, which improves the detection result. Ref. [45] uses AdaBoost to detect occlusion, which improves the detection result further.

2.4 Research status summary of occlusion detection method

By the above analysis, Table 6 shows the research status of current visual object occlusion detection field.

Specifically, there exists following problems to be solved.

(1) Deficiencies still exist in the existing occlusion detection methods

Existing occlusion detection methods based on intensity image are mainly achieved by video or image sequences, so generally deficiencies exist like large spatio-temporal cost, low detection accuracy, limited applicability and so on. A few occlusion detection methods based on single intensity image have the problems with application scope and accuracy, because they take the priori knowledge, region segmentation and semantic annotation as the premise. Therefore, occlusion detection method based on intensity images is still worthy of further study, and a feasible idea is to study the occlusion detection method based on online learning. In addition, currently scholars have worked on specific visual task in the condition of occlusion based on depth image, but they don't provide clear method on occlusion detection. Compared with the method based on certain number of intensity images, occlusion detection method based on a single depth image has the advantages of spatio-temporal cost and real-time. Although some researchers have proposed several occlusion detection methods based on a single depth image, the proposed methods are still insufficient in adaptability and detection accuracy, etc. So it is necessary to study the occlusion detection method based on the depth image further.

(2) The excavation of occlusion related feature in depth information is still not deep enough

Table 6 Research status of occlusion detection

Research status of occlusion detection	Major performance	Solution approach
The accuracy of occlusion detection method needs to be improved	The number of detected occlusion point is small; The number of incorrectly detected occlusion point is big; Precision-Recall curve is not ideal.	Starting from the method itself to improve the accuracy; Starting from detection result itself to optimize the result.
The applicability of occlusion detection method needs to be improved	Most methods are non-adaptive	Studying the occlusion detection method based on online learning
The quantitative evaluation criteria of occlusion detection method needs to be improved	The Ground Truth dataset for single intensity image or single depth image doesn't exist yet	Make related Ground Truth dataset

In addition to depth value jump feature, existing available features which can be used to detect occlusion in the depth information are rarely known by people. Existing research results indicate that the occlusion detection effect which simply uses depth value jump feature still needs to be improved, and many valuable features hidden in the depth information should be further explored. Once these features are found, the research based on the depth information will contribute to obtain occlusion detection method with higher accuracy, stronger applicability and better real-time performance, in order to meet the performance requirement of the active vision system better.

(3) The quantitative evaluation criteria of occlusion detection method needs to be established

This is mainly in the lack of evaluation criteria to assess the occlusion detection method. Currently, only CMU, Middlebury and AlgoSuit have the hand-labeled Ground Truth based on intensity image video. But there is no related Ground Truth in terms of occlusion detection method based on a single intensity image or a single depth image, and researchers primarily evaluate the occlusion detection method by visual observation (qualitative evaluation). Therefore, in order to evaluate related method more objectively, evaluation criteria of occlusion detection method needs to be established or formulated, so that relevant method can be easily evaluated in the quantitative way. It is considered that the quantitative evaluation of occlusion detection method at least contains two aspects: real-time and accuracy. The real-time evaluation criteria mainly denote the run-time of the method. Combining with the general flow of occlusion detection method presented in this paper, time consumption of preprocessing, occlusion clues extraction, occlusion detection and the whole process can be calculated respectively, and then the

real-time of method can be evaluated based on the above information. The accuracy evaluation criteria mainly denote precision  $R_p$ , recall  $R_r$  and  $F-score$  of the occlusion detection method. Comprehensive and credible evaluation results can provide important reference for the improvement of occlusion detection method, so establishing quantitative evaluation criteria of occlusion detection method is significant.

3 Conclusions

Occlusion problem is one of the challenging issues in vision field for a long time. Perceiving and detecting occlusion should be one of the abilities for an intelligent vision system. For a visual object with occlusion phenomenon occuring, if visual system doesn't have the ability to detect occlusion, it will lead to the fact that the visual system can't accomplish the tasks such as recognition, reconstruction, tracking and other operations. Therefore, it's crucial to study the visual object detection method and technology. This paper introduces the research status of current occlusion detection technology, describes the characteristics and deficiencies of existing methods, and discusses the issues that researchers should focus on. It is hoped it will be helpful for researchers in related fields. With the development of machine learning theory and the in-depth research of occlusion related feature in the depth information, the advantage of occlusion detection method combined with the depth information based on online machine learning idea will be prominent gradually in terms of real-time, accuracy and applicability. It is believed that, in the near future, researchers will be able to achieve outstanding achievement in this promising research direction, in order to meet the technology requirement in the field of computer vision better.



## References

- [ 1 ] Zhao Q, Yang Z, Tao H. Differential earth mover's distance with its applications to visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(2): 274-287
- [ 2 ] Ricco S, Tomasi C. Dense Lagrangian motion estimation with occlusions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, USA, 2012. 1800-1807
- [ 3 ] Hsiao E, Hebert M. Occlusion reasoning for object detection under arbitrary viewpoint. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, USA, 2012. 3146-3153
- [ 4 ] Yang M, Zhang L, Shiu S, et al. Gabor feature based robust representation and classification for face recognition with Gabor occlusion dictionary. *Pattern Recognition*, 2013, 46(7): 1865-1878
- [ 5 ] Feldman D, Weinshall D. Motion segmentation and depth ordering using an occlusion detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(7): 1171-1185
- [ 6 ] Zhang Q, Ngan K. Segmentation and tracking multiple objects under occlusion from multiview video. *IEEE Transactions on Image Processing*, 2011, 20(11): 3308-3313
- [ 7 ] Lee S J, Park K R, Kim J. A SfM-based 3D face reconstruction method robust to self-occlusion by using a shape conversion matrix. *Pattern Recognition*, 2011, 44(7): 1470-1486
- [ 8 ] Beeler T, Bradley D, Zimmer H, et al. Improved reconstruction of deforming surfaces by cancelling ambient occlusion. In: Proceedings of the European Conference on Computer Vision, Florence, Italy, 2012. 30-43
- [ 9 ] McAuley J J, Caetano T S. Fast matching of large point sets under occlusions. *Pattern Recognition*, 2012, 45(1): 563-569
- [ 10 ] Pizarro D, Bartoli A. Feature-based deformable surface detection with self-occlusion reasoning. *International Journal of Computer Vision*, 2012, 97(1): 54-70
- [ 11 ] Liang X, Ren W, Yu Z, et al. Improved coherent hierarchical culling algorithm based on probability computing model. *Journal of Software*, 2009, 20(6): 1685-1693 (In Chinese)
- [ 12 ] Szirmay-Kalos L, Umenhoffer T, Tóth B, et al. Volumetric ambient occlusion for real-time rendering and games. *IEEE Computer Graphics and Applications*, 2010, 30(1): 70-79
- [ 13 ] Depommier R, Dubois E. Motion estimation with detection of occlusion areas. In: Proceedings of the IEEE International Conference on Acoust, Speech and Signal Processing, San Francisco, USA, 1992. 269-272
- [ 14 ] Kolmogorov V, Zabih R. Computing visual correspondence with occlusions using graph cuts. In: Proceedings of the IEEE International Conference on Computer Vision, Vancouver, USA, 2001. 508-515
- [ 15 ] Ince S, Konrad J. Geometry-based estimation of occlusions from video frame pairs. In: Proceedings of the IEEE International Conference on Acoust, Speech and Signal Processing, Philadelphia, USA, 2005. 933-936
- [ 16 ] Ayvaci A, Raptis M, Soatto S. Sparse occlusion detection with optical flow. *International Journal of Computer Vision*, 2012, 97(3): 322-338
- [ 17 ] Humayun A, Aodha O M, Brostow G J. Learning to find occlusion regions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, USA, 2011. 2161-2168
- [ 18 ] Zhang S, He H, Kong L. Fusing multi-feature for video occlusion region detection based on graph cut. *Acta Optica Sinica*, 2015, 35(4): 0415001:1-10 (In Chinese)
- [ 19 ] Stein A, Hebert M. Local detection of occlusion boundaries in video. *Image and Vision Computing*, 2009, 27(5): 514-522
- [ 20 ] Apostoloff N, Fitzgibbon A. Learning spatiotemporal T-junctions for occlusion detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005. 553-559
- [ 21 ] Sargin M E, Bertelli L, Manjunath B S, et al. Probabilistic occlusion boundary detection on spatio-temporal lattices. In: Proceedings of the IEEE International Conference on Computer Vision, Kyoto, Japan, 2009. 560-567
- [ 22 ] Stein A, Hebert M. Occlusion boundaries from motion: Low-level detection and mid-level reasoning. *International Journal of Computer Vision*, 2009, 82(3): 325-357
- [ 23 ] He X, Yuille A. Occlusion boundary detection using pseudo-depth. In: Proceedings of the European Conference on Computer Vision, Heraklion, Crete, Greece, 2010. 539-552
- [ 24 ] Jacobson N, Freund Y, Nguyen T Q. An online learning approach to occlusion boundary detection. *IEEE Transactions on Image Processing*, 2012, 21(1): 252-261
- [ 25 ] Ullmann J R. Analysis of 2-D occlusion by subtracting out. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, 14(4): 485-489
- [ 26 ] Gay-Bellile V, Bartoli A, Sayd P. Direct estimation of nonrigid registrations with image-based self-occlusion reasoning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(1): 87-104
- [ 27 ] Hoiem D, Efros A, Hebert M. Recovering occlusion boundaries from an image. *International Journal of Computer Vision*, 2011, 91(3): 328-346
- [ 28 ] Chen X, Li Q, Zhao D, et al. Occlusion cues for image scene layering. *Computer Vision and Image Understanding*, 2013, 117(1): 42-55
- [ 29 ] Zhang S, Yan S. Depth estimation and occlusion boundary recovery from a single outdoor image. *Optical Engineering*, 2012, 51(8): 087003:1-11
- [ 30 ] Sharp G C, Lee S W, Wehe D K. Maximum-likelihood registration of range images with missing data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(1): 120-130
- [ 31 ] Wei H, Zhang Y, Liu S, et al. An algorithm on registration of multi-view range images based on SIFT feature matching. *Journal of Computer Aided Design & Computer Graphics*, 2010, 22(4): 654-661 (In Chinese)
- [ 32 ] Li S, Zhao D. Gradient-based polyhedral segmentation for range images. *Pattern Recognition Letters*, 2003, 24

- (12): 2069-2077
- [33] Pamplona S, Silva L, Bellon O R P. Automatic face segmentation and facial landmark detection in range images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2010, 40(5): 1319-1330
  - [34] Zhang G, Jia J, Wong T, et al. Consistent depth maps recovery from a video sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(6): 974-988
  - [35] Yang W, Zhang G, Bao H, et al. Consistent depth maps recovery from a trinocular video sequence. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Providence, USA, 2012. 1466-1473
  - [36] Ikeda K, Matsunuma C, Masuda H. Robust edge detection and GPU-based smoothing for extracting surface primitives from range images. *Computer-Aided Design and Applications*, 2011, 8(4): 603-616
  - [37] Shotton J, Sharp T, Fitzgibbon A, et al. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 2013, 56(1): 116-124
  - [38] Jang I Y, Cho J H, Seo M K, et al. Depth image based 3D human modeling resolving self-occlusion. In: *Proceedings of the ACM SIGGRAPH Posters*, Los Angeles, USA, 2008. 105
  - [39] Liu Y. Automatic range image registration in the Markov chain. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(1): 12-29
  - [40] Zhang S, Zhang Y, Kong L. Self-occlusion detection approach based on depth image. *Journal of Chinese Computer Systems*, 2010, 31(5): 964-968 (In Chinese)
  - [41] Zhang S, Zhang Y, Kong L. Self-occlusion detection algorithm combining depth image and optimal segmentation threshold iteration. *Chinese High Technology Letters*, 2010, 20(7): 754-757 (In Chinese)
  - [42] Zhang S, Gao F, Kong L. A self-occlusion detection approach based on range image of vision object. *ICIC Express Letters*, 2011, 5(6): 2041-2046
  - [43] Zhang S, Liu J. A self-occlusion detection approach based on depth image using SVM. *International Journal of Advanced Robotic Systems*, 2012, 9(12): 230-1-8
  - [44] Zhang S, Liu J, Kong L. Using random forest for occlusion detection based on depth image. *Acta Optica Sinica*, 2014, 34(9): 0915003:1-12 (In Chinese)
  - [45] Zhang S, Pang Y. Occlusion boundary detection method for depth image based on ensemble learning. *Acta Metrologica Sinica*, 2014, 35(6): 569-573 (In Chinese)

**Zhang Shihui**, born in 1973. He received his Ph.D degree from Yanshan University in 2005. Now he is a professor and a Ph. D supervisor at the School of Information Science and Engineering, Yanshan University. His research interests include visual information processing and pattern recognition.