A Part-Based Probabilistic Model for Object Detection with Occlusion

Chunhui Zhang, Jun Zhang, Heng Zhao, Jimin Liang*

School of Life Science and Technology, Xidian University, Xi'an, Shaanxi, China

Abstract

The part-based method has been a fast rising framework for object detection. It is attracting more and more attention for its detection precision and partial robustness to the occlusion. However, little research has been focused on the problem of occlusion overlapping of the part regions, which can reduce the performance of the system. This paper proposes a part-based probabilistic model and the corresponding inference algorithm for the problem of the part occlusion. The model is based on the Bayesian theory integrally and aims to be robust to the large occlusion. In the stage of the model construction, all of the parts constitute the vertex set of a fully connected graph, and a binary variable is assigned to each part to indicate its occlusion status. In addition, we introduce a penalty term to regularize the argument space of the objective function. Thus, the part detection is formulated as an optimization problem, which is divided into two alternative procedures: the outer inference and the inner inference. A stochastic tentative method is employed in the outer inference to determine the occlusion status for each part. In the inner inference, the gradient descent algorithm is employed to find the optimal positions of the parts, in term of the current occlusion status. Experiments were carried out on the Caltech database. The results demonstrated that the proposed method achieves a strong robustness to the occlusion.

Citation: Zhang C, Zhang J, Zhao H, Liang J (2014) A Part-Based Probabilistic Model for Object Detection with Occlusion. PLoS ONE 9(1): e84624. doi:10.1371/journal.pone.0084624

Editor: José Javier Ramasco, Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), Spain

Received: June 13, 2013; **Accepted:** November 15, 2013; **Published:** January 17, 2014

Copyright: © 2014 Zhang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by the 973 program under Grant No. 2011CB707702 and the National Natural Science Foundation of China under Grant No. 81090272. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jimleung@mail.xidian.edu.cn

Introduction

Object detection [1] is a classical problem in the field of computer vision. Among the numerous methods for object detection, the statistical-based approaches [2–13] have become mainstream. They discriminate the given object from others by learning it, hence achieving a more robust detection. As a typical statistical-based method, the part-based model has attracted increasing attention in the past decade [5–11,14–19]. As the name implies, the part is the local area of the object. The part-based models can capture both the local appearance and spatial structural information of the object simultaneously, which makes these methods robust to the variations of the object pose and appearance to some extent. Similar with other statistical-based methods, the part-based methods also include two fundamental problems: training and detection. The former refers to the modeling of the part appearance and spatial relationship among the parts. The latter refers to the optimization problem for acquiring the information about the parts (such as their position and occlusion status). It is worthy to emphasize that, most part-based methods pay attention to the part areas only. Therefore, they are robust only to the occlusions which do not overlap the part areas (as illustrated in Figure 1.(a)). However, if the occlusions overlap the part areas, as illustrated in Figure 1.(b), not only would the occluded parts be influenced, the non-occluded ones would also be shifted from their right positions due to the spatial relationship among the parts (refer to the experiments below for more details).

Related work and our contributions

In the past years, many part-based methods have emerged, such as the bag model [6,7], constellation model [8,15], pictorial structure model [9,16], star model [10,14], vocabulary based method [17,18] and *k*-fan model [11]. The spatial relationship among their parts is illustrated in Figure 2.

However, The bag model almost consider none of spatial relationship among parts. The other part-based models improve the performance of detection by adding the spatial relationship among the parts. Especially, the pictorial structure model and the *k*-fan model perform fast detection via dynamic programming [20,21], since the appearance of generalized distance transform (GDT) [22] greatly reduces the time complexity of the dynamic programming. However, they does not allow cycles in the spatial relationship [20]. There are many solutions for this problem [23–27], of which the simplest technique is the gradient descent algorithm (GD).

Although the above part-based models belong to the category of sparse representation, they still suffer from the problem of shading the part areas (as demonstrated in Figure 3). In this paper, these shaded parts are named disabled parts. some methods [19] employed a kind of part appearance representation which is robust to the occlusion, but the robustness is limited, especially when the occlusion region is large. In the literature about the occlusion, ignoring the occluded parts from the model is the most intuitive idea to solve the occlusion problem [8,14,28,29]. In general, it can be achieved by estimating a mask for the test image. However, the mask variable is difficult to model in the objective function to

Figure 1. Two kinds of occlusions faced by the part-based model. (a) Occlusion which does not overlap the part areas. (b) Occlusion overlapping the part areas.
doi:10.1371/journal.pone.0084624.g001

achieve a precise inference and cannot work in the case of large occlusion. Papandreou et al. proposed to solve the occlusion problem by using a robust objective function [30], which weakens the role of occlusion parts, but it is difficult to completely eliminate the influence of the occluded areas. Li et al. [31] solved the occlusion problem under a novel RANSAC framework. However, it is difficult to incorporate the spatial relationship to improve the detection of the keypoints. In [32,33], the authors solved the occlusion problem under the sparse framework, which is usually used in the case of batch image processing.

This paper considers applying the part-based method in object detection, with special emphasis on occlusion handling, and propose a part-based probabilistic model with an alternative detection scheme. In order to increase the detection accuracy, we constructed a fully connected graph to describe the spatial relationship among parts in the stage of training. Each edge is represented by a 2D Gaussian distribution for the vector difference of the position coordinates. Moreover, we introduced a penalty term to the objective function ensure us to obtain a more accurate detection result. For the occlusion problem, we assigned a binary status variable to each part to indicate whether it is occluded or not, and proposed a method to model the prior probability of the occlusion status variable. Then, according to the Bayesian theory, we constructed a new posterior probability as the objective function. In the stage of detection, We designed two alternative procedures, which are named as the inner inference and outer inference. The former used the GD to determine the positions of the parts given the current occlusion status variable. The outer inference is responsible for determining the occlusion status of the parts according to their current positions. To address the detection efficiently, we adopted a stochastic tentative method in the outer inference. In addition, in the procedure of the detection, we incorporated the validity test mechanism to avoid the invalid inner inference results.

Methods

Consider a model with n parts $V = \{v^1, v^2, \dots, v^n\}$. A detection result of a given image is expressed as $H = \{L, S\}$. The argument $L = \{l_1, l_2, \dots, l_n\}$ is the position variable, where $l_i = \{y_i, x_i\}$ denotes the position of part v^i . The argument $S = \{s_1, s_2, \dots, s_n\}$ represents the occlusion status variable, where s_i is a Boolean variable (if part v^i is a normal part, $s_i = 1$; otherwise, $s_i = 0$ for the disabled part). We define the object function as

$$p(H|I) = (p(I|L, S) \cdot p(L|S) \cdot p(S) \cdot \exp(Q(H))) / p(I), \quad (1)$$

which is the posterior probability of a result H given a test image

I , where $\exp(Q(H))$ is a penalty term and $Q(H)$ is defined in the following subsection.

Construction of the model

In Eq. 1, the posterior probability contains four items. 1. $p(I|H) = p(I|L, S)$, which represents the total matching probability of the normal parts. 2. $p(L|S)$ represents the *priori* probability of a spatial relationship among normal parts. 3. $p(S)$, a *priori* probability of the occlusion. 4. The penalty term $\exp(Q(H))$.

The total matching probability of all of the normal parts is

$$p(I|H) = C_0 \prod_{v^i \in V} g_i(I, l_i)^{s_i}, \quad (2)$$

where C_0 is a constant for a test image I , $g_i(I, l_i)$ is the matching probability of a single part [11].

For a *priori* probability of the spatial relationship $p(L|S)$, we employed the fully connected graph to represent the spatial relationship among the parts. However, as demonstrated in Figure 3(c), the disabled parts will severely affect the detection of the normal parts because of the edges between them. We overcame this problem by discarding the edges connected to the disabled parts. In addition, the positions of the disabled parts were supposed to follow the independent uniform distribution [8]. Therefore,

$$p(L|S) = \left(\prod_{(v^i, v^j) \in E^*} p(l_i | l_j)^{s_i s_j} \right) / M^{n-A}, \quad (3)$$

where A is the number of the normal parts with $s_i = 1$, M is a constant representing the number of possible positions where a part could be placed, E^* is the edge set of a fully connected graph, and $p(l_i | l_j)$ is defined as $p(l_i - l_j)$ which follows the 2D Gaussian distribution.

Let $w^{(m,n)}$ denote the conditional probability of part v^n being shaded under the condition that v^m is disabled. If the mean distance $d_{(m,n)}$ between v^m and v^n satisfies that $d_{(m,n)} < 2c_o$ (c_o is a constant), we have proved that (please refer to the *Appendix S1* for the deduction.)

$$\begin{aligned} w^{(m,n)} &= 2 \left(4c_o^2 - d_{(m,n)}^2 \right)^{1.5} / (3c_o^2 \pi \times d_{(m,n)}) \\ &\quad - 2 \sqrt{4c_o^2 - d_{(m,n)}^2} / (\pi \times d_{(m,n)}) + (d_{(m,n)} / \pi c_o) \\ &\quad \times \log \left((2c_o + \sqrt{4c_o^2 - d_{(m,n)}^2}) / d_{(m,n)} \right). \end{aligned} \quad (4)$$

Then a *priori* probability $p(S)$ is calculated as

$$p(S) = \left(\prod_{(v^i, v^j) \in E^*} R(s_i, s_j) \right) / U, \quad (5)$$

where

$$R(s_i, s_j) = \begin{cases} 1 - (1 - p_t)(2 - w^{(i,j)}), & s_i = 1 \text{ and } s_j = 1, \\ (1 - p_t)w^{(i,j)}, & s_i = 0 \text{ and } s_j = 0, \\ (1 - p_t)(1 - w^{(i,j)}), & \text{otherwise,} \end{cases} \quad (6)$$

represents the joint probability of the occlusion status about the part pair (v^i, v^j) . Where p_t is a constant standing for the probability of a part being present, and U is a normalization constant.

Figure 2. Spatial relationship among parts. (a) Bag model. (b) Constellation model. (c) Pictorial model. (d) k -fan model ($k=1,2,3$ from left to right).

doi:10.1371/journal.pone.0084624.g002

The penalty term $Q(H)$ aims at improving the detection results by emphasizing the weak parts, and also for regularizing the argument space of the objective function. It is defined as

$$Q(L, S) = \min_{V_S} (W_i(L, S))$$

$$= \min_{V_S} \left(\log(g_i(I, l_i)) + \left(\sum_{(v^i, v^k) \in E^*} (s_k \cdot \log(p(l_i | l_k))) \right) / \sum_{(v^i, v^k) \in E^*} s_k \right), \quad (7)$$

where V_S is the part set in which the occlusion status of each element is 1.

Substitute Eq. 2, 3, 5, 7 into Eq. 1, the posterior probability can be rewritten as

$$p(H|I) = \frac{C \times \prod_{(v^i, v^j) \in E^*} R(s_i, s_j) \times \prod_{v^i \in V} g_i(I, l_i)^{s_i} \times \prod_{(v^i, v^j) \in E^*} p(l_i | l_j)^{s_i s_j} \times \exp(Q(H))}{M^{n-A}}, \quad (8)$$

where $C = C_0 / [U \cdot p(I)]$ is a constant for a test image I . Applying the logarithm and minus operations to both sides of Eq. 8, we have

$$E(H) = (n - A) \log(M) - \sum_{(v^i, v^j) \in E^*} \log(R(s_i, s_j)) - Q(H)$$

$$- \sum_{v^i \in V} s_i \cdot \log(g_i(I, l_i)) - \sum_{(v^i, v^j) \in E^*} s_i s_j \cdot \log(p(l_i | l_j)). \quad (9)$$

The above expression is the right objective function for detection.

Detection

In the step of the detection, we look for an optimal detection result $H^* = \{L^*, S^*\}$ with minimum energy, which is

$$H^* = \arg \min_H (E(H)). \quad (10)$$

In this paper, we adopted the strategy of alternative optimization to solve the above problem. The basic idea is to let S search in the space of S (called the outer inference), and after each movement of S , the inner inference searches the current optimal part positions L^{*S} . These two procedures are performed alternately until the terminal conditions are satisfied.

Figure 3. Influence of the disabled parts on the detection of the normal ones. (a) is the manual label of the face. In (b), the eyes have been occluded (the occlusion is represented by the dotted box). (c) shows the detection results that the disabled parts degrade the detection of the normal ones. (d) illustrates the detection results after the occluded parts are discarded from the model.

doi:10.1371/journal.pone.0084624.g003

Figure 4. Flow chart of the outer inference.
doi:10.1371/journal.pone.0084624.g004

Inner inference and Outer inference. In the inner inference, given the current status vector $S = \{s_1, \dots, s_n\}$, Eq. 9 can be expressed as

$$E_S(L) = - \sum_{v^i \in V} s_i \cdot \log(g_i(I, l_i)) - \sum_{(v^i, v^j) \in E^*} s_i s_j \cdot \log(p(l_i | l_j)) - Q(L, S) + C', \quad (11)$$

where C' is a constant. Eq 11 is the right objective function in inner inference. we used the gradient descent algorithm (GD) to search the current optimal position variable for each part $L^{*S} = \{l_1^{*S}, \dots, l_n^{*S}\}$.

In the outer inference, given L^{*S} , the object function Eq. 9 becomes the function depending on S only:

$$E(S) = (n - A) \log(M) - \sum_{(v^i, v^j) \in E^*} \log(R(s_i, s_j)) - Q(L^{*S}, S) - \sum_{v^i \in V} s_i \cdot \log(g_i(I, l_i^{*S})) - \sum_{(v^i, v^j) \in E^*} s_i s_j \cdot \log(p(l_i^{*S} | l_j^{*S})). \quad (12)$$

As aforementioned, the outer inference is responsible for determining the occlusion status variable. At the beginning of

Figure 5. Two-part sub-model.
doi:10.1371/journal.pone.0084624.g005

the outer inference, the occlusion status variable is assumed to be $S = \{1, 1, \dots, 1\}$, implying that no occlusion happens to any part. The aim of the outer inference is to find the next probable S to reduce the value of Eq. 12.

Due to the discreteness of the S space, we adopted a stochastic tentative method to address the outer inference. In each iteration, we calculated the gradient vector $h = G(S) = \{\partial E / \partial s_1, \dots, \partial E / \partial s_n\}$ for Eq. 12. If $(s_j = 0 \& h(j) < 0) \vee (s_j = 1 \& h(j) \geq 0)$ holds, we consider the s_j as a feasible descending bit, and consider the s_j which has the minimal value of $|h(j)|$ as the most irresolute bit. The procedure of the outer inference is illustrated in Figure 4 and detailed in Table 1.

The validity test is used to validate whether the inner inference has obtained a feasible result. For a two-part model, if the likelihood

$$\alpha(l_1, l_2) = g_1(I, l_1) g_2(I, l_2) p(l_2 | l_1) \quad (13)$$

is larger than some threshold, this two-part model is defined to pass the validity test. A full-part model is defined to pass the validity test if there is at least one two-part sub-model passing the validity test. Step 3g to Step 3h is the procedure of inner inference, and can avoid the solutions from deviating from the right occlusion status variable. Finally, after we have obtained the output, we could estimate the position of the disabled parts just by the spatial relationship among all of the parts, i.e. minimizing the following expression:

$$K(L) = - \sum_{(v^i, v^j) \in E^*} \log(p(l_i | l_j)), \quad (14)$$

where $\{l_i | v^i \in V_{S^*}\}$ is the known variable, which has been obtained in Algorithm 1 (Table 1).

Results and Discussion

In this section, we tested the performance of our method on the Faces dataset in the Caltech database [8,34], and the performance of l-fan [11] is compared. For this dataset, as done in [11], six parts were selected: the left eye, the right eye, nose, the left corner of the mouth, the right corner of the mouth and the chin (defined as the part 1,2,...,6 respectively). In our experiment, the distance error e is defined as the mean distance of n parts from the

Table 1. Algorithm 1: Outer inference.

Input: The initial occlusion status variable $S = \{1, 1, \dots, 1\}$, the initial energy value $O_i = \inf$;

Step 1. Use the simulated annealing algorithm to obtain the initial position, then obtain current optimal position L^{*S} via the inner inference, if it passes the validity test (explained below), go to Step 3, otherwise go to Step 2;

Step 2. Use all of the two-part sub-model (illustrated in Figure 5) to make an inner inference (determine the positions of these two parts), until a two-part sub-model T whose results can pass the validity test appear. Then, estimate the approximate positions of other parts except for the two parts in T , and obtain their position by GD further. If none of the two-part models can pass the validity test, quit the outer inference in failure;

Step 3. **While** the result S^* is not altered and the maximum iteration number m is not reached,

(a) Calculate gradient vector h ;

(b) If $E(S) < O_i$, S^* is updated as S , and O_i is updated as $E(S)$;

(c) In S , search the feasible descending bits;

(d) If there is no feasible descending bit, invert the most irresolute bit in S , and go to Step 3g, otherwise go to Step 3e;

(e) If there is only one feasible descending bit, invert it, and go to Step 3g, else go to Step 3f;

(f) If there are at least two feasible descent bits, therein invert the corresponding bit with a probability proportional to its gradient absolute value;

(g) Carry out the GD algorithm for $E_S(L)$, if the results cannot pass the validity test, go to Step 3h;

(h) Choose a different bit to invert again randomly, go to Step 3g;

End While;

Output: The status result S^* and $\{l_i^* | i \in V_S\}$.

doi:10.1371/journal.pone.0084624.t001

detection results to its corresponding ground truth. The smaller e is, the better the given model performs on this specific test image.

We first demonstrated the influence of the disabled parts on the normal ones. We chose 200 images from the Faces dataset to train a 1-fan model, and chose 100 test images to construct a test dataset by shading the right eye and the right corner of the mouth in each test image. Then, we compared the discarded 1-fan model (the disabled parts had been discarded from the 1-fan model) (shown in Figure 6 (b)) with the original 1-fan model (shown in Figure 6 (a)). We used the distance error e as the evaluation index. The distance error e of all of the test images was normalized to $[0, 1]$. The comparison result is illustrated in Figure 7, where $N(e)$ is the distribution function of the distance error. The horizontal axis represents normalized e , the vertical axis represents the percentage

of test images whose distance error is smaller than e . It is obvious that the higher the curve is, the better the model performs. It should be emphasized that only the normal parts were gathered to calculate the distance error e . Figure 7 shows that the detection results of the discarded 1-fan were much better than that of the original 1-fan model, due to discarding the disabled parts. In other words, the disabled parts will severely affect the detection of the normal parts if they are not handled properly. Once the occlusion happens, the matching degree of the disabled parts is very likely to be low at the right positions, so they must search other positions to minimize the objective function, which would increase the deformation of the edge connecting them. As a result, the adjacent normal parts would tune their positions to reduce the edge cost (as illustrated in Figure 3.(c)). For this reason, we discarded the

Figure 6. 1-fan model and discarded 1-fan model.

doi:10.1371/journal.pone.0084624.g006

Figure 7. Influence of disabled parts on the Faces dataset.
doi:10.1371/journal.pone.0084624.g007

disabled parts in our method. The performance will be demonstrated in the following experiments.

In the second experiment (partially occluded experiment), we compared the proposed model with the 1-fan model and demonstrated the detection accuracy of our method when one or two parts were partially occluded. Both models were trained by 200 images selected randomly from the Faces dataset. We randomly selected another 100 images to construct the two test datasets. The first dataset, termed as the one-part-shaded test dataset, was constructed by shading part(1), part(2), ..., part(6) respectively with different occlusion degrees for each test image. The occlusion degrees are defined as the ratio of the occlusion area to part area varied from about 44% (the size of the occlusion region was 40×40) to 100% (the size of the occlusion region was 60×60), 11 values. The number of test images in the one-part-shaded dataset was $100 \times 6 \times 11 = 6600$. The second dataset, termed as the two-parts-shaded test dataset, was constructed by shading 7 kinds of adjacent part pairs (i.e., part(1,2), part(1,3),

part(1,4), part(4,6), part(2,3), part(2,5), part(5,6)) respectively with different occlusion degrees for each test image. The number of images in the second dataset is $100 \times 7 \times 11 = 7700$. The test images in the two-parts-shaded test dataset are illustrated in Figure 8. The distance error e was also used as the evaluation index for detection accuracy. The average distance errors for all of the test images are plotted in Figure 9. As a typical instance, the results on the part(1,2)-shaded test are listed in Table 2.

Figure 9 shows that the average distance error for our model is almost constant and much smaller than that of the 1-fan model when the occlusion degree changes from 44% to 100%. These results are due to the fact that the disabled parts were discarded from our model, and could not affect the detection of the normal parts. For the 1-fan model, the average distance error on the one-part-shaded dataset was smaller than that on the two-part-shaded dataset. What is more, the average distance error for the 1-fan model increase with the increase in the occlusion degree. Specifically from Table 2, once the occlusion degree exceeded

Figure 8. Sample images in the two-part-shaded test dataset. (a) is an image with part(1,2) being shaded (the occlusion degree is 81%). (b) is an image with part(4,6) being shaded (the occlusion degree is 64%). The purple solid boxes represent the part regions and the black dotted boxes represent the occlusion regions.
doi:10.1371/journal.pone.0084624.g008

Figure 9. Average distance error of the partially occluded experiment.
doi:10.1371/journal.pone.0084624.g009

70%, the average distance error increased sharply. That is because the information of the face held by part(1,2) (see Figure 8 (a)) was more than the other parts, thus the occlusion of part(1,2) greatly misguide the 1-fan model.

To further evaluate the performance of our method, we constructed four test datasets by complete shading one, two, three, or four parts (named the completely shading experiment). We carried out our algorithm on these test datasets. To evaluate the occlusion status variable S , we used two evaluation indices: the occlusion false alarm probability p_f and the occlusion false dismissal probability p_d for all of the bits in the occlusion status variable S . We defined the occlusion false alarm probability p_f as the probability that the bit in S was wrongly estimated as 0, but it

was actually 1. We defined the occlusion false dismissal probability p_d as the probability that the bit in S was wrongly estimated as 1, it was actually 0. To evaluate the position variable L , we also used the distance error e as the evaluation index. The results of complete shading experiment are shown in Table 3. We did not compare our method with the 1-fan model in this experiment because the 1-fan model almost cannot work in the case where three or more parts are shaded.

From Table 3, we can see that both p_f and p_d increase as the number of disabled parts increases. That is because when more parts are occluded, it will be more difficult to obtain valid results in Step 1 of Algorithm 1 (Table 1), and the number of the valid two-part sub-models will be also reduced in Step 2 of Algorithm 1 (Table 1). Table 3 also shows that average distance error increases as the number of disabled parts increases. The reasons, except for those illustrated above, also lie in that the disabled parts are estimated only by the spatial relationship with the normal ones. The experimental results in Table 3 demonstrate that our method is competent for object detection even though most parts are occluded.

Table 2. Average e of the 1-fan model and our method when part(1,2) was shaded.

Occlusion degree of each part	Average e of 1-fan (pixel)	Average e of Our model (pixel)
44.00%	6.6096	2.6243
49.00%	6.9762	2.6026
54.00%	7.3334	2.5952
59.00%	7.7257	2.6025
64.00%	8.1099	2.5939
69.44%	8.7296	2.5946
75.11%	11.6345	2.6098
81.00%	11.9330	2.6090
87.11%	14.0736	2.6091
93.44%	16.0132	2.6021
100.00%	18.3148	2.5845

doi:10.1371/journal.pone.0084624.t002

Table 3. The results of complete shading experiment on the Faces dataset.

Test set	p_f	p_d	Average e (pixel)
One-part-shaded test datasets	0.13%	0.67%	2.4057
Two-part-shaded test datasets	0.42%	0.50%	2.6009
Three-part-shaded test datasets	2.78%	1.11%	3.8830
Four-part-shaded test datasets	5.33%	1.83%	5.8602

doi:10.1371/journal.pone.0084624.t003

Supporting Information

Appendix S1 Deduction of $w^{(i,j)}$. (PDF)

Figure S1 Diagram of calculating $w^{(i,j)}$. (TIF)

Acknowledgments

This research was supported in part by the Program of the National Basic Research and Development Program of China (973) under Grant

No. 2011CB707702, and the National Natural Science Foundation of China under Grant Nos. 81090272 and 41031064.

Author Contributions

Conceived and designed the experiments: CHZ JZ. Performed the experiments: CHZ. Analyzed the data: CHZ JZ. Contributed reagents/materials/analysis tools: CHZ HZ. Wrote the paper: CHZ JZ HZ JML.

References

1. Amit Y (2002) 2D Object Detection and Recognition: Models, Algorithms, and Networks. Cambridge, MA: MIT Press.
2. Moghaddam B, Pentland A (1997) Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19: 696–710.
3. Rowley H, Baluja S, Kanade T (1998) Rotation invariant neural network-based face detection. *Computer Vision and Pattern Recognition*: 38–44.
4. Schneiderman H, Kanade T (1998) Probabilistic modeling of local appearance and spatial relationships for object recognition: 45–51.
5. Fischler M, Elschlager R (1973) The representation and matching of pictorial structures. *IEEE Transactions on Computers* 100: 67–92.
6. Csurka G, Dance C, Fan L, Willamowski J, Bray C (2004) Visual categorization with bags of keypoints. *Workshop on statistical learning in computer vision, ECCV 1*: 22.
7. Lazebnik S, Schmid C, Ponce J (2005) A maximum entropy framework for part-based texture and object recognition. *IEEE International Conference on Computer Vision* 1: 832–838.
8. Fergus R, Perona P, Zisserman A (2003) Object class recognition by unsupervised scale-invariant learning. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2: II-264–II-271.
9. Felzenszwalb P, Huttenlocher D (2005) Pictorial structures for object recognition. *International Journal of Computer Vision* 61: 55–79.
10. Felzenszwalb P, McAllester D, Ramanan D (2008) A discriminatively trained, multiscale, deformable part model. *IEEE Conference on Computer Vision and Pattern Recognition*: 1–8.
11. Crandall D, Felzenszwalb P, Huttenlocher D (2005) Spatial priors for part-based recognition using statistical models. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1: 10–17.
12. Zhao X, Satoh Y, Takauji H, Kaneko S, Iwata K, et al. (2011) Object detection based 241 on a robust and accurate statistical multi-point-pair model. *Pattern Recognition* 44: 1296–1311.
13. Yang X, Liu H, Jan Latecki L (2011) Contour-based object detection as dominant set computation. *Pattern Recognition* 45: 1927–1936.
14. Fergus R, Perona P, Zisserman A (2005) A sparse object category model for efficient learning and exhaustive recognition. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1: 380–387.
15. Fei-Fei L, Fergus R, Perona P (2003) A bayesian approach to unsupervised one-shot learning of object categories. *IEEE Ninth International Conference on Computer Vision*: 1134–1141.
16. Li S, Lu H, Zhang L (2012) Arbitrary body segmentation in static images. *Pattern Recognition* 45: 3402–3413.
17. Wen M, Wang L, Wang L, Zhuo Q, Wang W (2007) Object class recognition using snow with a part vocabulary. In: Slezak D, Szczuka M, Duentzsch I, Yao Y, editors. *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*. Berlin: Springer. pp. 526–533.
18. Agarwal S, Awan A, Roth D (2004) Learning to detect objects in images via a sparse, part-based representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 26: 1475–1490.
19. Mikolajczyk K, Schmid C, Zisserman A (2004) Human detection based on a probabilistic assembly of robust part detectors. In: Pajdla T, Matas J, editors. *Computer Vision-ECCV 2004*. Berlin: Springer. pp. 69–82.
20. Szeliski R (2010) *Computer vision: Algorithms and applications*. Berlin: Springer.
21. Jiang X, Große A, Rothaus K (2011) Interactive segmentation of non-star-shaped contours by dynamic programming. *Pattern Recognition* 44(9): 2008–2016.
22. Felzenszwalb P, Huttenlocher D (2012) Distance transforms of sampled functions. *Theory of computing* 8: 415–428.
23. Chou P, Brown C (1990) The theory and practice of Bayesian image labeling. *International Journal of Computer Vision* 4: 185–210.
24. Kirkpatrick S, Gelatt C Jr, Vecchi M (1983) Optimization by simulated annealing. *Science* 220: 671–680.
25. Frey B, MacKay D (1998) A revolution: Belief propagation in graphs with cycles. *Advances in neural information processing systems* 10. Cambridge, MA: MIT Press. pp. 479–485.
26. Boykov Y, Veksler O, Zabih R (2001) Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23: 1222–1239.
27. Kolmogorov V, Zabih R (2004) What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26: 147–159.
28. Zhou Z, Wagner A, Mobahi H, Wright J, Ma Y (2009) Face recognition with contiguous occlusion using markov random fields. *IEEE 12th International Conference on Computer Vision*: 1050–1057.
29. Gross R, Matthews I, Baker S (2004) Constructing and fitting active appearance models with occlusion. *IEEE Conference on Computer Vision and Pattern Recognition*: 72.
30. Papandreou G, Maragos P (2008) Adaptive and constrained algorithms for inverse compositional active appearance model fitting. *IEEE Conference on Computer Vision and Pattern Recognition*: 1–8.
31. Li Y, Gu L, Kanade T (2009) A robust shape model for multi-view car alignment. *IEEE Conference on Computer Vision and Pattern Recognition*: 2466–2473.
32. Wagner A, Wright J, Ganesh A, Zhou Z, Ma Y (2009) Towards a practical face recognition system: Robust registration and illumination by sparse representation. *IEEE Conference on Computer Vision and Pattern Recognition*: 597–604.
33. Peng Y, Ganesh A, Wright J, Xu W, Ma Y (2010) Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Conference on Computer Vision and Pattern Recognition*: 763–770.
34. Fei-Fei L, Fergus R, Perona P (2004) Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. *Computer Vision and Pattern Recognition Workshop, 2004 CVPRW '04 Conference on* 12: 178.