

Planar object recognition based on local feature prediction agreements

GU Chengge, BAO Jianying, SANG Haonan, and MO Jinqiu^a

School of Mechanical Engineering, Shanghai Jiao Tong University, 200240 Shanghai, China

Abstract. Object recognition has drawn great attention in industrial application especially in automated feeding and assembling, for it can greatly improve the line flexibility and save cost. In this paper, a simple but effective method for planar object recognition is presented. This method can deal with objects under complex conditions like occlusion and clutter. The method generates object pose hypothesis from the prediction agreements of different local features in the object. There are two stages contained in our method, offline stage and online stage. At offline stage, the representative parts in the object are chosen as its local features and the recognition template is made. Next at online stage, the matches of different local features are found in the input image. Then the prediction agreements are searched among them in order to generate the final object pose hypothesis. A thin planar object recognition experiment has been conducted under occluded conditions and an improved result is presented compared with the traditional overall matching method.

1 Introduction

Object recognition under complex conditions is a popular topic in machine vision for its promising industrial application potentiality in aspect of automated feeding and assembling. Compared with traditional automation method, machine-vision-based one does not require additional equipment like the feeding lines, thus the system can be more flexible and save cost as well. However, objects are usually randomly placed in a bin, so that the occlusion and clutter inevitably happen, which presents a big challenge in accurate recognition (Figure 1). The main existing problem is that the information derived from the 2D image of occluded objects is incomplete. And the extracted object edges may be missing where the objects overlap due to similar texture and color. Besides, the self-similar clutter background can raise false positive recognition easily.



Figure.1 Object in occluded and clutter condition

In this paper, a simple but effective method for planar object recognition by local feature prediction agreements is presented, which allows to give more accurate prediction under occlusion and clutter conditions. The

method contains offline and online two stages. At offline stage, objects are first split into several representative parts as its local features and for each part-object pair, their local-overall transformation relation is then calculated. At online stage, by finding matches of each local feature in the input image, the local feature matching instance sets are obtained. All the feature instances are then transformed to object pose prediction via transformation relation. Finally, in a Hough-like voting process, different orders of local feature prediction agreements are obtained. And in general, the higher order the agreements are, the more confident the prediction result is.

1.1 Related Work

Recognition for planar object in 2D image is a recurrent topic in machine vision and industry application. There are mainly two technical routes for recognition. One is based on global description of object. In [1], Steger presented a recognition method via overall contour matching, which is robust against occlusion, clutter and nonlinear illumination changes. And this method is widely used in industry application for its good performance. But for highly clutter conditions, this method always fails and may offer some false positive results. The other is focused on local description of object. In [2], Rothwell et al. used projective invariant features of the object shape to do matching and generate hypotheses. Then the hypotheses are merged and verified to obtain final results. In [3], Huang et al. defined a chord angle descriptor based on the angles between two chords for each sample points in contours. By calculating the chord angle descriptor in both model contour and input contour, the distance of the two descriptors is obtain and thus the similarity score is

^a Corresponding author: mojqiu@sjtu.edu.cn

calculated. In [4], Tang et al. used line segment detector (LSD) first to extract straight lines in input image and then connected the line segments and did region division. At last, they completed the object recognition by comparing the shape feature vector to that of the template. In [5], Alberto et.al also used LSD method to detect the contour line segments first. Then they used the line segment and its direction as local descriptor. They found matches between the line segments and the standard CAD model to gain object pose hypotheses. Finally, through a modified Hough-voting scheme, they obtained several best hypotheses as the final results.

1.2 Overview of the paper

Section 2 describes the recognition method based on local feature prediction agreements and its theoretical analysis. In section 3, some experiment results and the corresponding analysis are presented. In section 4, the potential extension of this method to 3D object recognition via 2D image is discussed.

2 Recognition method based on local feature agreements

2.1 Definition and notations

An object or workpiece that is to be recognized in the following is denoted as *Obj*. And all the local features *Obj* contains are $F_{Obj} = \{f_1, f_2, \dots, f_m\}$, where m is the number of its local features. Local features are the representative parts in *Obj*. The given input 2D image is denoted as I , and the matched instance set of F_{Obj} in I is written as $Inst(I, F_{Obj}) = \{f_1(I), f_2(I), \dots, f_m(I)\}$, where $f_i(I), i = 1 \sim m$ is the matched instance set of a single local feature f_i . The instance set contains the pose of every local feature instance, $f_i(I) = \{pose_i^{(1)}, pose_i^{(2)}, \dots, pose_i^{(f_i(I))}\}$, where $|f_i(I)|$ is the instance number of $f_i(I)$. For a 2D image, $pose = \{position = (x, y), angle\}$.

For a local feature f_i , denote the local-overall transformation relation by $T_{f_i}^{Obj}$. So, the pose predictions of *Obj* given by the instances of local feature f_i are $H_{Obj}(f_i(I)) = T_{f_i}^{Obj}(f_i(I))$. Given two local features f_i and f_j , if the pose predictions given by their instances reach an agreement, then define the hypotheses based on such consistent predictions as two-order agreement, denoted by $agreement_2 = \{H_{Obj}(f_i(I)) \wedge H_{Obj}(f_j(I)) | i, j = 1 \sim m, i \neq j\}$. We call each element in $agreement_x$ a *Group*. And a *Group* is the combination of pose predictions from different feature instances. For *Obj* with m local features, the hypothesis agreement can reach up to m orders. The hypotheses existing in high order agreement will not be recorded repeatedly in lower one.

With all the agreements obtained, the complete recognition results of *Obj* are denoted as $H(Obj, I) = \{agreement_2 \vee agreement_3 \vee \dots \vee agreement_m\}$.

2.2 The probabilistic analysis of recognition by prediction agreements

Under the same matching technique, applying the method based on local feature prediction agreements can gain higher confidence in object recognition compared with traditional overall matching method.

Given an object *Obj* with m local features, and assume in I , there is a non-occluded *Obj* existing with some $pose_t$. Apply the overall matching method, and suppose the probability of *Obj* existing with $pose_t$ is $P_{ori}(Obj, pose_t) = p_a$. With the same matching technique, there should be m local feature instances detected which belong to m local features respectively. And the m instances should form a *Group* = $\{pose_1, pose_2, \dots, pose_m\}$ in m -order agreement that predicts *Obj* exists with $pose_t$. Suppose the probability of each instance prediction is p_i , then considering their agreements, the probability of *Obj* existing with $pose_t$ is

$$\begin{aligned} P_{loc}(Obj, pose_t) &= 1 - P(\neg(Obj, pose_t) | agreement_m(pose_t)) \\ &= 1 - \prod_{i=1}^m (1 - p_i) \end{aligned} \quad (1)$$

Using the same matching technique, the matching score or probability of local features should be no less than that in overall method, i.e., $p_i \geq p_a$. Compare the two methods,

$$\begin{aligned} P_{loc}(Obj, pose_t) - P_{ori}(Obj, pose_t) &= (1 - p_a) - \prod_{i=1}^m (1 - p_i) \geq (1 - p_1) \left(1 - \prod_{i=2}^m (1 - p_i) \right) > 0 \end{aligned} \quad (2)$$

That is, applying our method, the confidence in objection recognition could be enhanced.

When *Obj* is occluded in I , the overall method may encounter a sharp confidence loss due to the missing of some local information. For our method, the prediction agreement may degenerate to lower order, but the non-occluded local features can still contribute a high confidence to the results.

Next, we would analyze the reliability of our method. That is, check the probability of n ($n \leq m$) predictions given by n local features that agree on *Obj* existing with a certain pose while *Obj* is actually not. The probability is

$$\begin{aligned} P(agreement_n | Obj \text{ not exist with } pose) &= P(agreement_n | \neg(Obj, pose)) \end{aligned} \quad (3)$$

Using Bayes formula,

$$\begin{aligned} P(agreement_n | \neg(Obj, pose)) &= \frac{P(\neg(Obj, pose) | agreement_n) \cdot P(agreement_n)}{P(\neg(Obj, pose))} \end{aligned} \quad (4)$$

Where, according to (1),

$$\begin{aligned} P(\neg(Obj, pos) | agreement_n) &= 1 - P((Obj, pos) | agreement_n) = \prod_{i=1}^n (1 - p_i) \end{aligned} \quad (5)$$

And $P(agreement_n)$ means the probability that n local feature instance sets could form a n -order

agreement $agreement_n$ for Obj with a certain pose in I , thus

$$P(agreement_n) = \frac{|agreement_n|}{|f_1(I)||f_2(I)| \dots |f_n(I)|} \quad (6)$$

$$= \frac{|agreement_n|}{\prod_{i=1}^n |f_i(I)|}$$

Suppose $\min |f_i| = |f_k|$, then we have $|agreement_n| \leq |f_k|$, thus

$$P(agreement_n) < \frac{1}{\prod_{i=1, i \neq k}^n |f_i(I)|} \quad (7)$$

As for $P(\neg(Obj, pose))$, it can be written as $1 - P(Obj, pose)$. $P(Obj, pose)$ is the probability that Obj exists with a certain $pose$. Given an input image I , Obj exists with limited pose due to its limited number, so for a certain $pose$, the probability that Obj exists is approximately 0, i.e. $P(\neg(Obj, pose)) \approx 1$. Finally, the probability

$$P(agreement_n | \neg(Obj, pose)) < \frac{\prod_{i=1}^n (1 - p_i)}{\prod_{i=1, i \neq k}^n |f_i(I)|} \quad (8)$$

$$< \frac{1}{\prod_{i=1, i \neq k}^n |f_i(I)|}$$

When the number n is set relatively large, the denominator would be a quite large number, and the probability $P(agreement_n | \neg(Obj, pose))$ would be very close to 0, thus the reliability of our method can be guaranteed.

2.3 Obtaining multi-order prediction agreement based on combination numbers

After obtaining the local feature prediction set $H(I, F_{Obj}) = \{H_{Obj}(f_1(I)), H_{Obj}(f_2(I)), \dots, H_{Obj}(f_m(I))\}$, we need to check the prediction agreements between every two of them and form the multi-order prediction agreements. Different with traditional Hough-like voting scheme [6,7], where all the individual hypotheses are put together for pose clustering or voting, our method makes the vote or forms the multi-order prediction agreements among each local feature prediction set. In this case, there would be no agreement containing two pose predictions that belong to the same local feature. Since some local features might be easy to detect, their instance predictions may form a self-clustering in some place, which could lead to the false-positive recognition. With our scheme, such misrecognition can be avoided.

First, define the prediction agreement measurement. Suppose the pose predictions of two local feature f_i and f_j are $H_{Obj}(f_i, I)^p = pose_i^{(p)}$, $H_{Obj}(f_j, I)^q = pose_j^{(q)}$. Refer to [8], give the measurement definition in the form of probability,

$$P\left(pose_i^{(p)} = pose_j^{(q)} \right)$$

$$= \exp \left(- \frac{\left\| position_i^{(p)} - position_j^{(q)} \right\|_2}{n_p \beta_p} \right) \cdot \exp \left(- \frac{\left\| angle_i^p - angle_j^q \right\|_2}{n_a \beta_a} \right) \quad (9)$$

where the n_p and n_a are used for normalization such that the component position and angle are of the same magnitude. β_p and β_a are parameters used to adjust different matching error acceptance.

Set a probability threshold, and using (9), we can obtain the two-order agreement $agreement_2$. Next, we will discuss how to obtain n -order agreement based on $agreement_{n-1}$ and combination numbers. By this method, we can avoid repetitive agreement measurement calculation between instance predictions when finding different agreements.

Code the m local features of Obj with index $1 \sim m$, then for n -order agreement, there would be C_m^n conditions. Denote all the combinations (with ascending order) by $comb(n, \{1, 2, 3, \dots, m\})$, and each one of them indicates which local features form the $agreement_n$. without loss of generality, assume a certain n -order agreement $agreement_n^{(1)}$ has the feature combination $\{1, 2, 3, \dots, n\}$. Consider the $n-1$ order agreement $agreement_{n-1}^{(1)}$ whose combination is $\{1, 2, 3, \dots, l-1, l+1, n\} (1 \leq l \leq n)$, and on the basis of it, we can obtain $agreement_n^{(1)}$ only if we can find one of its agreement Groups in which all its $n-1$ predictions agree with the same feature l prediction in these two-order agreements whose combinations are $\{1, l\}, \{2, l\}, \{3, l\}, \dots, \{l, n\}$. Check the $n-1$ $agreement_2$, and the searching and finding work above in any $n-2$ of them can be substituted for searching in a $n-1$ order agreement which contains the same local feature indices. For example, the searching in former $n-2$ $agreement_2$ can be substituted for searching in $agreement_{n-1}$ with combination $\{1, 2, 3, \dots, l-1, l, l+1, n-1\}$. Given different choices of $n-2$ $agreement_2$, there would be different corresponding $agreement_{n-1}$. But each corresponding $agreement_{n-1}$ just is a combination condition in $comb(n-1, \{1, 2, 3, \dots, n\})$. Therefore, we can infer that to obtain a $agreement_n$ with feature combination $comb(n, \{1, 2, 3, \dots, m\})^{(i)}$ is to find the n predictions groups which appear in all $n-1$ order agreements with combination condition in $comb(n-1, comb(n, \{1, 2, 3, \dots, m\})^{(i)})$.

Based on the analysis above, the procedure to obtain multi-order prediction agreement is given:

Procedure: obtain multi-order prediction agreements

Input: local feature prediction set $H(F_{Obj}, I)$

Steps:

- 1: calculate all the $agreement_2$
- 2: **For** $n = 3, 4, \dots, m$ **do**
- 3: **For** i-th combination $comb_n$ in $comb(n, \{1, 2, 3, \dots, m\})$ **do**
- 4: Find $agreement_{n-1}$ with feature

combination $comb(n-1, comb_n)$ and assign them to ag_1, ag_2, \dots, ag_n ;

5: Take ag_1 and ag_2 , compare the first $n-2$ feature prediction components in their *Groups*, keep *Groups* with mutual components and delete the rest;

6: **if** ag_1 or ag_2 is empty

7: $agreement_n^{(i)} = \emptyset \rightarrow \text{continue}$;

8: Take the last feature prediction of ag_1 and ag_2 to form a two-order agreement $ag' = \{ag_1^{(n-1)}, ag_2^{(n-1)}\}$;

9: Compare ag' with the last two feature prediction components in *Groups* of ag_3, ag_4, \dots, ag_n , keep *Groups* with mutual components and delete the rest;

10: **if** ag' is empty

11: $agreement_n^{(i)} = \emptyset \rightarrow \text{continue}$;

12: Take the first $n-3$ feature prediction components in *Groups* of ag_3, ag_4, \dots, ag_n , compare them with the corresponding feature components in *Groups* of ag_1/ag_2 , keep *Groups* with mutual components in ag_1/ag_2 , and delete the rest;

13: **if** ag' is empty

14: $agreement_n^{(i)} = \emptyset \rightarrow \text{continue}$;

15: $agreement_n^{(i)} = \{ag_1, ag_2^{(n-1)}\}$;

16: **endfor**

17: **endfor**

Output: $agreement_2, \dots, agreement_m$

2.4 Object pose derivation from multi-order agreement

Assume there is a *Group* in a certain n -order prediction agreement, $gp = \{pose_1, pose_2, \dots, pose_n\}$, in which the probability of each pose prediction is p_i . We will talk about the final object pose hypothesis derivation from two different aspects.

From the aspect of their prediction possibility, naturally, the pose prediction with higher probability should gain a larger weight in final pose synthesis. Therefore, give the pose hypothesis based on probability,

$$pose_{prob} = \frac{\sum_{i=1}^n p_i \cdot pose_i}{\sum_{i=1}^n p_i} \quad (10)$$

Our method is established on the prediction agreements between different local features, so intuitively, the prediction pose which is more agreed with others should gain a larger synthesis weight. The target "more agreed" can be described by their space aggregation. Here, we define a position weight to describe their space aggregation. First, give the average pose of the *Group*

$$pose_{avg} = \frac{1}{n} \sum_{i=1}^n pose_i \quad (11)$$

Then the position weight of every pose in *Group* can be defined as

$$w_i = \frac{\exp(b_i)}{\sum_{i=1}^n \exp(b_i)} \quad (12)$$

Where b_i is the parameter that describes the space aggregation of $pose_i$ and $pose_{avg}$. According to (9), we

can use $b_i = P(pose_i = pose_{avg})$.

On the basis of position weight, the pose hypothesis is

$$pose_{posi} = \frac{\sum_{i=1}^n w_i \cdot pose_i}{\sum_{i=1}^n w_i} \quad (13)$$

Combine these two aspects, the final hypothesis is given

$$pose_{pred} = \alpha_s \cdot pose_{sw} + \alpha_p \cdot pose_{pw}, \alpha_s + \alpha_p = 1 \quad (14)$$

In general, $\alpha_s = \alpha_p = 0.5$.

To obtain the confidence of final hypothesis via the analysis in section 2.2, we have to transfer the confidence of each pose prediction to that at $pose_{pred}$. The probability transfer can be implemented via (9). Define the probability transfer coefficient $c_i = P(pose_i = pose_{pred})$. Then according to (1), the confidence of the final hypothesis for gp is

$$P(Obj, pose_{pred}) = 1 - \prod_{i=1}^n (1 - c_i p_i) \quad (15)$$

3 Experiment

To evaluate our method, we conduct an experiment on platform Halcon-12.0 and Matlab(R2016b) to identify the planar object which is randomly piled up in Figure 1. Also, the results of our method and the traditional one by overall contour matching are compared. In our experiment, the scale changes are not considered, because the planar object is very thin and the distance change from objects in different stack layers to our 2D camera can be neglected.

3.1 Offline Stage

At offline stage, we make the object template for recognition (Figure 2). The object parts which are pointed out with rectangles are the chosen local features. The arrow in Figure 2 is the reference set to represent the whole part. And the relative position between each local feature and the reference arrow gives the local-overall transformations.

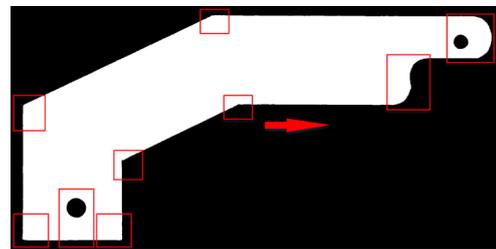


Figure 2. Object template for recognition

3.2 Online Stage

At online stage, given an input image, edge detection algorithm is first performed to extract object edges. However, due to the similar surface texture and color, the gradient where objects overlap has a very low magnitude and the standard edge detection methods like Canny operator [9] tend to fail, giving rise to the edge missing.

To overcome this problem, we first apply LSD method [10] to extract the edge segments (Figure 3) and then superimpose them on the original image to enhance object edges.

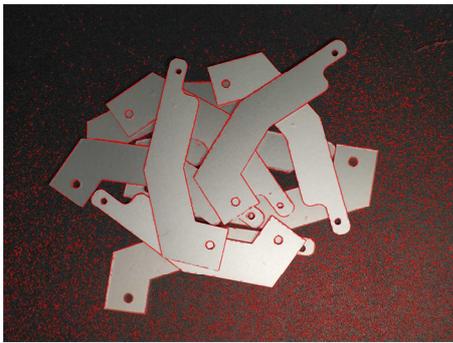


Figure 3. Edge segments extracted by LSD

We use “Matching Assistant” in Halcon-12.0 to find local feature instances in the input image via its contour matching function. The matching results are given in Figure 4. As we can see, many false positive instances are presented and in the meanwhile, several true feature instances are missing. The false positive instances can be easily rejected because no strong agreements are made with them. As for the missing true instances, so long as not all the local features of an object instance are inaccessible, we still can identify it through the predictions of its rest local features via our agreement scheme.

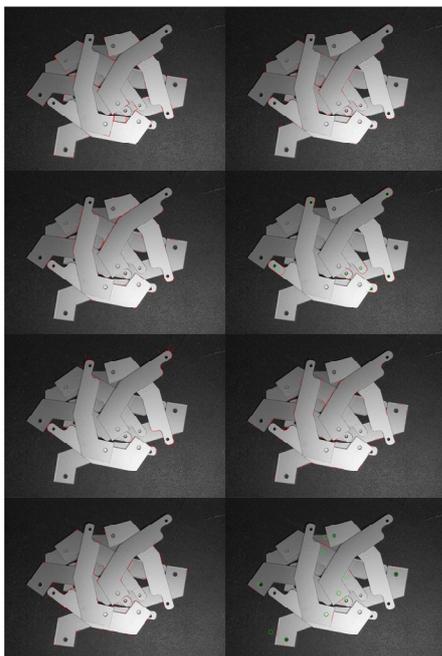


Figure 4. Local feature matching results

To visualize our prediction agreements scheme, we draw all the object pose predictions as arrows in Figure 5 and use different colors to distinguish between predictions from different local features. The arrow tail indicates the predicted position, while its direction indicates the predicted angle. And the length of it represents the confidence that the prediction holds. We can see that all the arrows belong to false positive instances seem to be randomly dispatched. Comparatively, the ones of true

positive instances are gathered together with highly consistent directions. According to the algorithm in section 2.3, we can obtain eight prediction agreement *Groups*, and they are marked with black circles with dashed lines in Figure 5.



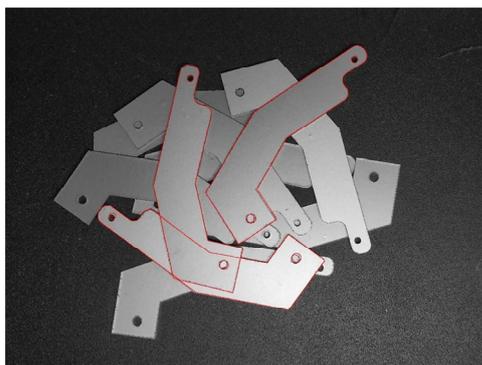
Figure 5. Visualization of local feature instance predictions

Refer to section 2.4, eight object pose hypotheses can be derived from those eight prediction agreement *Groups*. The results are showed in Figure 6(a) We also calculate the probabilities that these hypotheses hold. In experiment, we set the similarity probability 0.9 for two positions with 10-pixel difference in both dimensions or two angles with 2-degree difference. Based on that assumption, we can determine the parameters in (9). Then, according to (15), we get the result that all probabilities are greater than 0.99.

We also use “Matching Assistant” in Halcon-12.0 to implement the traditional overall matching experiment. The same edge enhanced image is fed as its input and the results are shown in Figure 6(b) The matching parameters are set as recommended in Halcon and the lowest matching score is set to be 0.5. Only three matches are presented with probabilities 0.53, 0.54 and 0.57 respectively which is far less than our method. Comparatively speaking, our method can give more matches with higher confidence under the same complex working conditions.



(a)



(b)
Figure 6. (a) Matching results of our method
(b) Matching results of traditional method

4 Conclusion and discussion

We have presented a simple and reliable method for planar object recognition. This method focuses on the local information of an object, so that it can handle the conditions with incomplete object information like occlusion. The matching hypothesis is given through the consistent predictions of different local features in an object, thus it can obtain high confident results even in clutter situations.

The thought obtaining hypothesis through the prediction agreements of local features can also be generalized in 3D objection recognition. Choose different feature surfaces of an object and find their matches in 3D point data. Then with our method, the 3D object pose hypothesis can also be obtained.

Reference

1. S. Carsten. Similarity measures for occlusion, clutter, and illumination invariant object recognition. *joint pattern recognition symposium*, 148-154 (2001).
2. R. Charlie, et al. Planar object recognition using projective shape representation. *International Journal of Computer Vision* **16.1**, 57-99(1995).
3. Huang W.G., Yang J.Y., Zhu Z.K., Hu D.M., Chord angle feature description for occlusion of shape matching. *Optics and Precision Engineering*, **23.6**, 1758-1767(2015)
4. Tang Y., Wu Q.X., Zhu F., Methods research of plane parts recognition and localization based on regional division. *Machinery Design & Manufacture*, **10**, 172-175(2015).
5. P. Alberto, S. Tonello, and E. Menegatti. Flexible 3D localization of planar objects for industrial bin-picking with monocular vision system. *conference on automation science and engineering*, 168-175(2013).
6. T. Federico, and Luigi D. Stefano. Object Recognition in 3D Scenes with Occlusions and Clutter by Hough Voting. *pacific-rim symposium on image and video technology*, 349-355(2010).
7. C. Choi, et al. Voting-based pose estimation for

robotic assembly using a 3D sensor. *international conference on robotics and automation*, 1724-1731(2012).

8. C. Schmid. A structured probabilistic model for recognition. *computer vision and pattern recognition*, 485-490(1999).
9. J. F. Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8.6**, 679-698(1986).
10. R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, LSD: A fast line segment detector with a false detection control. *IEEE Trans. Pattern Anal. Mach. Intell.*, **32.4**, 722-732(2010).