Computing Importance of 2D Contour Parts by Reconstructability

Ge Guo[1,3], Yizhou Wang[2], Tingting Jiang[2], Alan Yuille[4] and Wen Gao[2]
[1] Key Lab of Intelligent Information Processing of CAS,
Institute of Computing Technology, CAS, Beijing 100190, China
[2] Nat'l Eng. Lab for Video Technology, Sch'l of EECS, Peking University, Beijing 100871,China
[3] Graduate University of Chinese Academy of Sciences, Beijing 100039, China
[4] Department of Statistics, University of California, Los Angeles, CA 90095-1554
gguo@jdl.ac.cn, {Yizhou.Wang, ttjiang}@pku.edu.cn, yuille@stat.ucla.edu, wgao@pku.edu.cn

Abstract

Part-based representation has become an essential representation for objects. Different parts of a shape have various importance to shape perception. Concerned by the contributions of local parts to the perception of global shapes, we introduce an importance measure of shape parts from the viewpoint of their "reconstructability". Given one shape part as a local observation, we estimate the global object shapes through a proposed novel shape reconstruction method. The reconstruction quality is determined by part geometry variation and part uniqueness. Then a conditional entropy formulation is introduced to measure the part importance considering the reconstruction quality, part uniqueness and variation.

Extensive experiments demonstrate the benefit of the proposed part importance to object detection applications. Moreover, psychological experiments are conducted to further support the consistency of our importance measure with human perception towards shape parts.

1. Introduction

Many convincing psychological evidences suggest that object parts play a significant role in object perception and recognition (*e.g.* [13, 27]); plenty of successful computer vision practices (*e.g.* [12, 6, 30]) further confirm that the part-based model is an essential representation for objects.

In computer vision literature there are two typical categories of part-based models, (1) *appearance-based models* of the constituent components of objects from image regions (*e.g.* [2, 8]) and (2) *shape-based models* such as 3D volumetric primitives (*e.g.* [1]), medial axes (*e.g.* [31]) and 2D contour fragments.

This paper focuses on the shape-based parts, particularly the *2D contour parts*, for that (1) shape is a robust feature especially for objects with large color / texture variations in its class [21, 14]; and (2) this type of object representation is exploited in a large body of literature and achieves great successes in challenging real applications [9, 26, 20, 7, 24].

To study part-based object representation, one important question to answer is "*Are object parts equally important to a certain visual task?*"

For example, in the context of shape perception and object recognition, usually parts are not equally important according to our experiences. Psychological studies have discovered that, different shape parts provide different retrieval cues for shape perception, and have different abilities for recalling object contours [3]. Furthermore, human vision has powerful abilities of shape reconstruction from parts. Previous studies find that early vision is used to rapidly complete the whole shape when it is partially occluded. Even in cases of severe occlusion, humans can still perceive a complete shape effortlessly [23]. All these motivate us to model and compute part importance from a new viewpoint of shape reconstruction. We do not only take account of local characteristics of a part, but also its contribution to global shape perception. In the literature there exists limited work on the computational modeling of part importance for shape perception. Deriving such a model will provide a mathematical foundation for understanding the computational aspect of shape perception and part representation.

Notice that in this paper we compare the importance of shape parts within one object category, *e.g.* swan neck *vs.* swan back. We do not consider the ability of parts to discriminate between different object categories.

In our methodology, the principle is that the importance of a part comes from its ability to reconstruct the whole object shape, namely the "reconstructability", given the part as a local observation. The reconstructability *i.e.* how well a given local shape part contributes to recovering the global object shapes of that class, is determined by the following two aspects. One is the part's shape property such as

shape geometry and variation, which affects the reconstruction quality. For example, parts of slight variations produce much accurate reconstructions according to shape knowledge of the object category. And a degenerated shape part *e.g.* a flat line greatly damages the reconstruction quality. The other factor is the part uniqueness, *i.e.* how distinct one part is from the others. The more unique a part is, the more important it is. The two aspects are embodied into a conditional entropy formulation, based on which the part importance measure is defined.

We also propose a novel shape completion / reconstruction method from local 2D contour parts. Our idea is to take advantages of partial observation of a shape part, estimate its local deformation and accordingly infer the whole shape's deformation under global optimization.

Moreover, we demonstrate the benefit of the proposed part importance measure to object detection applications. The importance measure is utilized to weight the matching distances of shape parts, in the processes of voting object candidates, inferring matching correspondences and computing total matching scores. In particular, we address severe occlusion cases and show that the use of our importance measure have big advantages, although this scenario breaks most of the existing object detectors. Experimental results demonstrate that it improves the detection rate as well as the localization accuracy by applying part importance to object detection.

In addition, psychological experiments are conducted, in which the task is to implicitly identify objects of the category through parts. The results support that our method are much more consistent with human perception compared with previous work.

1.1. Related work

In the literature, the study of the computational aspect of shape part importance to object perception and recognition is quite limited. Previous research has made some progress towards this direction. For example, Sukumar *et al.* [28] adopt the curvature variation measure (CVM) for 3D shape parts. Renninger *et al.* [22] evaluate 2D local parts based on the edgelet orientation distributions, and use it to predict eye fixations in light of the information maximization principle. In both work, the parts with large curvature variation or orientation entropy are considered important. However, in general these methods only consider the local geometric information of shapes. Hoffman and Singh [14] propose three factors as the measurements – the relative size, protrusion degree and boundary strength, in which some global information is considered.

In computer vision, the evaluation of part importance relates to a broader topic – learning weights of features. There are two main streams, the generative and discriminative paradigms. Our method is categorized to the former; and we also discuss the related work of the latter.

First, in the generative paradigm, for example, the minimax entropy learning framework proposes to update feature weights using a learning-by-sampling paradigm [32]. Kersten *et al.* [17] suggest a general principle for determining feature weights under the Bayesian framework, *i.e.* features with more reliable information have higher weights attributed to their corresponding prior constraint. Our observation is that, in addition to Kersten *et al.*'s principle, (which is mainly about reconstruction quality and variation), the feature / part uniqueness factor should also be taken into consideration when evaluating part importance from the viewpoint of shape reconstruction / perception.

Second, in the discriminative paradigm where the main purpose is classification, feature weighting mechanisms are often automatically embodied. For instance, in boosting, weak classifiers can be considered as features, whose weights are updated in an iterative procedure according to classification error rate. In maximum margin based models, kernels are implicit features, and the notion of margin corresponds to the weights of the implicit features (*e.g.* [4]). In conditional random field (CRF) models, feature weights are formulated in the potential functions (*e.g.* [25]). Some other discriminative feature selection methods are based on simple criteria such as feature statistics (*e.g.*[29]) or certain utility functions (*e.g.* [12]).

The rest of the paper is organized as follows. In Section 2 we present the shape part importance model, which is based on a novel shape reconstruction approach as proposed in Section 3. Then the provided shape part importance measure is applied to object detection in Section 4, and verified through psychological experiments in Section 5. Finally Section 6 concludes the paper.

2. Shape Part Importance

In order to learn the importance of different parts for shape perception we provide a computational model based on a conditional entropy formulation.

2.1. Formulation

A certain object category is represented by a set of constituent shape parts $\{x^1, ..., x^N\}$, where N is the number of parts of the category. Assume that we have n instances of each part, $x^a = \{x_i^a\}, a \in \{1, ..., N\}, i \in \{1, ..., n\}$. Here we refer to "shape part" as an object contour segment. In general, we can apply our part importance measure to any contour-based part representation. The parts are assumed to be given, or learned from object shape instances.

We compute the importance of a part by its "reconstructability", specifically by how well the part can be used to reconstruct the global shape. There is a problem however, when a part instance (a contour fragment) is present, but its correspondence to the global shape is unknown. In

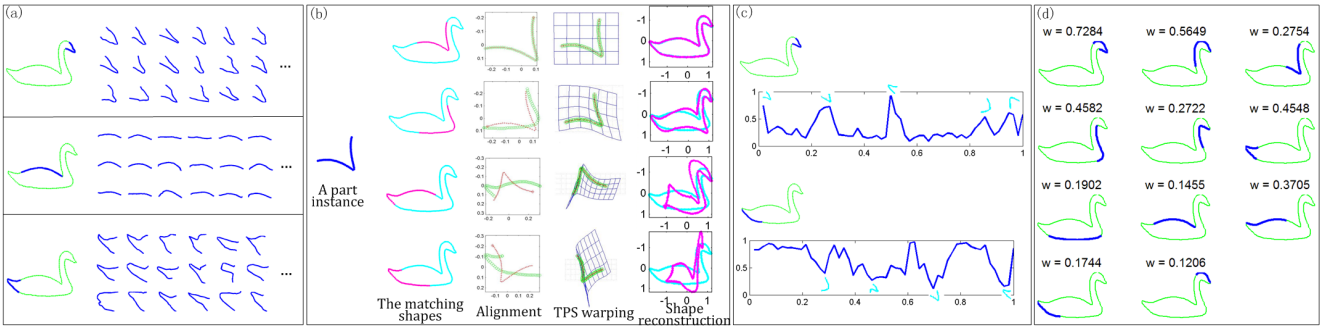

Figure 1. (a) Shape part examples and the part instances. (b) Illustrations of the shape reconstruction process. A part instance is matched to a segment of the object contour, based on which their shape transformation is computed. Then the global transformation is estimated and a complete shape (pink contour) is reconstructed. (c) Distributions of the shape reconstruction probabilities at the normalized locations of the object shape model. The cyan small segments along the curves illustrate some matching segments at different locations of the object contour. (d) The shape parts ranked by the category-specific part importance.

other words, the matching location of the contour fragment to the object shape model is unknown. For example, an almost flat part of the swan class can be considered as either a dorsal part or an abdominal part. Thus, the reconstruction is not unique. The ambiguity of the matching location, which stems from the non-uniqueness of part shapes, induces uncertainty into the shape reconstruction.

The uncertainty of shape reconstruction can be measured by a conditional entropy as Eq. (1).

$$\begin{aligned} h_a &= H(y \mid x^a; \mathcal{T}) \\ &= -\sum_i p(x_i^a; \mathcal{T}) \sum_l p(y_l \mid x_i^a; \mathcal{T}) \log p(y_l \mid x_i^a; \mathcal{T}) \end{aligned} \tag{1}$$

The term $p(y_l \mid x_i^a; \mathcal{T})$ is the probability of the reconstructed shape given each part instance as observation, where y_l is a reconstructed global shape at the matching location l given x_i^a, and with a class-specific object shape $\mathcal{T}$ as reference / expectation for shape reconstruction (we use the mean shape learned from each object class in this paper). The correspondence (matching location) of a part instance to the object shape is treated as a hidden variable; it should be summed over during the reconstruction process. Also the overall contribution of a part to shape reconstruction comes from different part instances; so it is expected to sum over all possible instances of x^a. $p(x_i^a; \mathcal{T})$ is the prior of a part instance, which is considered uniform here.

This conditional entropy formulation is based on shape reconstruction quality at each correspondence, which is reflected by the computation of $p(y_l \mid x_i^a; \mathcal{T})$. When the reconstructed shape y_l is consistent with the shape $\mathcal{T}$, $p(y_l \mid x_i^a; \mathcal{T})$ is high; otherwise it is low. The reconstructed shape is determined by the shape geometry and variation of the part instance, with the assumption that the part instance is pre-aligned (through affine transformations) with the part at location l on the reference shape $\mathcal{T}$. The details of computing $p(y_l \mid x_i^a; \mathcal{T})$ is given in Section 2.2.

Taking account of all the correspondences and part instances, there are two aspects influencing the entropy. First, part uniqueness determines the certainty of the probability distribution $p(y_l \mid x_i^a; \mathcal{T})$. For example in Figure 1 (c) using the swan class as an example, the head part is unique compared to other parts. This uniqueness induces the high probabilities of shape reconstruction when it corresponds to fewer object locations of similar shapes, and low probabilities due to the other correspondences (matching locations). Therefore the distribution of $p(y_l \mid x_i^a; \mathcal{T})$ is peaky in this case and the uncertainty is low. On the other hand, if a part's instance have many good correspondences with the object contour segments at those locations, such as the flat tail fragment in Figure 1 (c), a number of reconstructed shapes with high qualities are induced, thus the uncertainty of the probability distribution $p(y_l \mid x_i^a; \mathcal{T})$ as well as the conditional entropy h_a will increase. Second, part variation, which is reflected by a set of part instances, also influence the uncertainty. Basically the larger the variation of a part instance, the lower the reconstruction quality even at the right correspondence to the object shape, which in turn increases the entropy.

According to the above observations, *Part importance* is defined based on the proposed conditional entropy. The part with lower uncertainty of reconstructing the object shape is considered as higher "reconstructability"; thus the part is more important. We use a logistic function to formulate the importance measure,

$$w_a = \frac{1}{1 + \exp\{c_1 h_a + c_2\}} \tag{2}$$

where c_1 and c_2 are two parameters to enforce the weights fall into a comparable range for different object categories.

2.2. Implementation

In implementation, we uniformly sample a discrete set of normalized object locations $l \in [0, 1]$ as the matching

position, where l is an index of position on the contour $\mathcal{T}$. Each part instance is matched to the segment $\mathcal{T}_l$ located at l[1]. This matching leads to a reconstructed object shape $y_l = \Phi(x_i^a, l; \mathcal{T})$ under some shape transformation constraints, where Φ is a shape reconstruction function. We propose an efficient shape reconstruction method, by *partial shape matching* and *global transformation estimation*, where the former is to compute the transformation between $\mathcal{T}_l$ and x_i^a, and the latter is to infer the optimal global transformation of $\mathcal{T}$ to generate the most possible reconstructed shape (see details in Section 3). Finally, the reconstructed shape y_l is evaluated by the Gaussian distribution in the PCA shape space, $\tilde{p}(y_l) \sim \mathcal{N}(\mathcal{T}, \Sigma)$, where $\mathcal{T}$ and Σ are the mean and covariance learned in the PCA shape space respectively. Figure 1 (c) shows the distributions of $\tilde{p}(y)$ at different matching locations. And the final probability of shape reconstruction $p(y_l \mid x_i^a; \mathcal{T})$ in Eq.(1) is normalized as follows, $p(y_l \mid x_i^a; \mathcal{T}) = \frac{\tilde{p}(y_l)}{\sum_j \tilde{p}(y_j)}, j \in [0, 1]$.

Figure 1 (a) & 4 illustrate the ranked parts according to the importance computed by our model.

In the following section we will present a novel shape reconstruction method, which gives an efficient algorithm for the implementation of $\Phi(x_i^a, l; \mathcal{T})$.

3. Shape Reconstruction

There is much previous work on shape completion, such as amodal completion [16] which makes use of heuristics *e.g.* local continuities, proximity and global regularities; and curve completion *e.g.*[18] under the rules such as smoothness or curvature-based constraint. However these methods have limited abilities to deal with severe occlusion. Whereas in our problem, the observed part may be a small portion of the object shape. In this section a novel shape reconstruction method is proposed to handle the problem.

The process of shape reconstruction consists of two steps, called *partial shape matching* and *global transformation estimation* (Figure 2).

In the first step, each instance of the part x^a is matched to the segments at uniformly sampled locations of the reference shape (mean shape) as described above. The transformations between the reference segment and the part instance are obtained from the shape matching algorithm TPS-RPM [5]. Specifically, denote the point set of the part instance x_i^a as X, and that of reference shape $\mathcal{T}$ and the reference segment $\mathcal{T}_l$ as TandT_P respectively. The affine transformation A_P and the warping deformation W_P from T_P to X are estimated. The transformed T_P is,

$$X_P = T_P \cdot A_P + K_P \cdot W_P \tag{3}$$

where K_P is the TPS kernel [5] generated from T_P.

[1]Notice that the parts and the mean shape are learned from a set of training object instances normalized to the same scale, as in Section 4.1. Thus the length of $\mathcal{T}_l$ is taken the same as the length of the part instance.

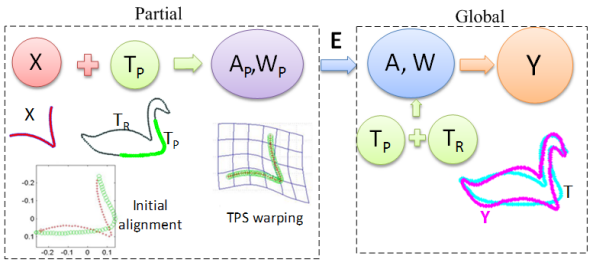

Figure 2. The proposed method of shape reconstruction from partial matching to global shape transformation estimation. The observed part X is matching with the segment T_P on the mean shape. And the transformation (A_P, W_P) between the two is computed. Based on this, the optimized global transformation (A, W) is estimated, and the complete shape Y is reconstructed by transforming the mean shape $T = T_P \cup T_R$ using (A, W).

In the second step, we use the local transformations to estimate the optimal global affine transformation A and deformation W, and further reconstruct a complete shape Y by transforming the shape T.

$$Y = T \cdot A + K \cdot W, \tag{4}$$

where K is the TPS kernel generated from T. By assuming the consistence of the affine transformations $A = A_P$, Eq.(4) can be rewritten as a decomposed form

$$\begin{pmatrix} Y_P \\ Y_R \end{pmatrix} = \begin{pmatrix} T_P \\ T_R \end{pmatrix} \cdot A + \begin{pmatrix} K_P & B \\ B^T & K_R \end{pmatrix} \cdot \begin{pmatrix} W'_P \\ W_R \end{pmatrix} \tag{5}$$

where $T_R = T \setminus T_P$ is the remaining point set on the mean shape other than the matching segment; K_P, K_R are the TPS kernels from T_P, T_R respectively; W consists of W'_P and W_R, both to be estimated, and the new deformation W'_P may be slightly different from the original W_P, which is computed only from the local matching of T_P and X.

W is obtained by solving an optimization problem under two constraints: (1) the reconstructed whole shape should entertain the part instance as much as possible, and (2) the amount of global deformation is as small as possible. The objective is to minimize the following energy function,

$$\begin{aligned} E &= E_D + \lambda_1 E_P + \lambda_2 E_W \\ \text{where } E_D &= ||Y_P - X_P||^2, \\ E_P &= ||W'_P - W_P||^2 \\ E_W &= \text{trace}(W^T K W). \end{aligned} \tag{6}$$

E_D, E_P, E_W account for the distance error, the change of the partial wrapping deformations, and the amount of the global deformation. The parameter λ_1 controls the partial deformation constraint, and λ_2 is set to be the same as the parameter related to the warping deformation in the TPS-RPM shape matching between T_P and X [5].

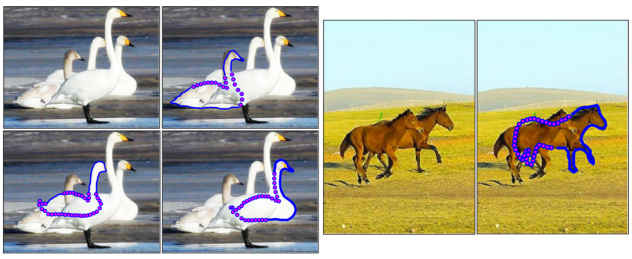

Figure 3. Shape completion examples. The blue contour segments are the observed object boundaries; and the pink dotted lines denote the reconstructed contours.

The optimization problem is solved by methods such as least square and gradient-descent. In consequence the reconstructed shape can be obtained from Eq.(5), which provides the implementation of Φ in Section 2.2. Illustrations are shown in Figure 1 (b).

Notice that the part X is pre-aligned to T_P before the TPS matching. Hence our shape reconstruction is orientation-invariant, and so is the part importance model proposed in this paper.

Figure 3 shows some shape reconstruction / completion examples of occluded objects in natural images using the proposed method. We can see that although the observed parts have various deformations from the shape model, the object contours are still well recovered. This demonstrates the strong reconstruction power of the proposed method under shape deformation and severe occlusion.

4. Object Detection with Part Importance

In this section we demonstrate that, by exploiting the proposed part importance in object detection, the performances of object detectors are improved; and in particular, we show the distinct advantages of detecting severe occluded objects.

4.1. The parts

The shape parts of each object category is learned using the object outlines from the ETHZ Shape Classes [11] and INRIA-horses datasets [15]. First, the object instances within one class are aligned in location and orientation, and normalized in scale. Then, kAS (k-Adjacent Segments) detector [9] is adopted to generate/learn a set of contour segments (k = 2,3 in this paper) as shape parts (see Figure 1 (d) & 4). We use this detector to learn parts since kAS can capture the low- and mid-level shape configurations and is robust to clutters in the images. Figure 1 (a) shows some learned part instances. Notice that some parts are overlapped. Similar parts (those near to each other and with large overlaps) are grouped into one.

For testing images, the object detection and boundary localization experiments are conducted on the extracted edge maps [19] in the presence of background clutter and occluding objects.

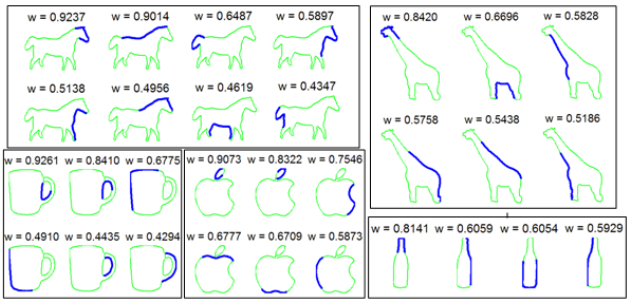

Figure 4. Ranked important parts of the ETHZ [11] and Inria-horses datasets[15]. The importance values are shown on top of each highlighted part. For some classes, only several top ranked important parts are displayed.

4.2. Part importance in object detection

We apply the part importance measure to one of the state-of-the-art object detectors using the shape-based representations [10]. The general detection process contains two steps, (1) a Hough-style voting scheme is utilized to obtain rough estimates for location and scale of object candidates; (2) with the estimates as initialization, a non-rigid shape matcher is used to localize exact object boundaries.

The proposed part importance is exploited in the detection process as follows. First, in the Hough voting, each kAS feature in the test image is matched with the shape parts to vote the object center, and the voting score is not only determined by the similarity of the matching pair, the strength of the feature, but also weighted by the part importance. This importance weight will relatively enhance the votes of important parts. Second, in the shape matching process from the object shape model to an object candidate, part importance is used to infer correspondences in the shape matching and also to compute the total matching score. The intuition is that, important parts should find more matches with the object candidate and lower matching cost.

Notice that the shape matching is based on a point-based representation, and the importance of each contour point on the object model is inherited from the part it belongs to. For overlapping parts, the importance value of a common point takes the maximum of the importance values of all the parts that contain the point.

The shape matching method is adapted from the TPS-RPM algorithm [5]. An deterministic annealing process is adopted to jointly estimate the correspondence and transformation between two shapes. The target is to minimize the

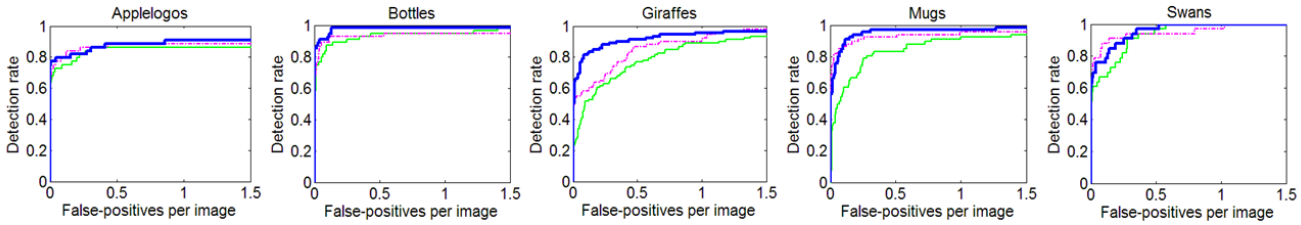

Figure 5. Performance comparisons of the object detectors with (blue curve) / without (pink curve) part importance, and that of the method in [10] using one hand-drawing object for each class as the shape model (green curve).

Table 1. Performances of object detection & localization WITHOUT / WITH part importance at 0.4 FPPI.

	applelogos	bottles	giraffes	mugs	swans	horses
AUC	0.843 / 0.911	0.928 / 0.991	0.798 / 0.936	0.946 / 0.979	0.935 / 0.963	0.701 / 0.860
Boundary accuracy	0.937 / 0.960	0.895 / 0.913	0.790 / 0.853	0.852 / 0.880	0.839 / 0.866	0.733 / 0.789

following objective function,

$$E_{\text{TPS-RPM}} = \sum_i \sum_j w_i m_{ij} ||z_j - \psi(v_i))||^2 - \sum_i \sum_j (1 - w_i) m_{ij} + t \sum_i \sum_j m_{ij} \log m_{ij} + \lambda_2 \text{trace}(W' K W) + \lambda_3 |A - I|. \quad (7)$$

where v_i and z_j denote a point of the learned parts and the candidate object contour in the test images respectively; m_{ij} represents their correspondence (a soft assignment); ψ is the shape transformation; and w_i is the importance of v_i, which is integrated as a weight into Eq. (7) ($w_i \in [0, 1]$ in this paper). The first term is a weighted matching distance to penalize large matching cost of important parts. The second term encourages more important parts to be matched. The third term is an entropy barrier function with the temperature parameter t as defined in [5]. The fourth and the last terms penalize the total amount of deformation and affine variation, respectively. I is the identity matrix.

During the iteration process, the correspondence is updated by

$$m_{ij} = \frac{1}{t} \exp\{-\frac{w_i (z_j - \psi(v_i))' (z_j - \psi(v_i))}{2t}\}. \quad (8)$$

Finally, to compute the total matching energy for scoring the potential object, we use

$$E_s = \beta_1 \sum_i \sum_j w_i m_{ij} ||z_j - \psi(v_i))||^2 + \beta_2 \frac{\sum_i w_i \cdot \mathbf{1}(m_i > \theta)}{\sum_i w_i} + \beta_3 \text{trace}(W' K W) + \beta_4 |A - I|. \quad (9)$$

The terms are similarly defined as in Eq.(7) except its third term. Here m_i is the maximum value of $m_{ij}, j \in \{1, ..., k\}$ where k is the number of the candidate points. $\mathbf{1}(\cdot)$ is an indicator function, where θ is a matching threshold. $(\beta_1, \beta_2, \beta_3, \beta_4) = (8, 0.2, 0.2, 0.5)$ for all the object classes in the experiments.

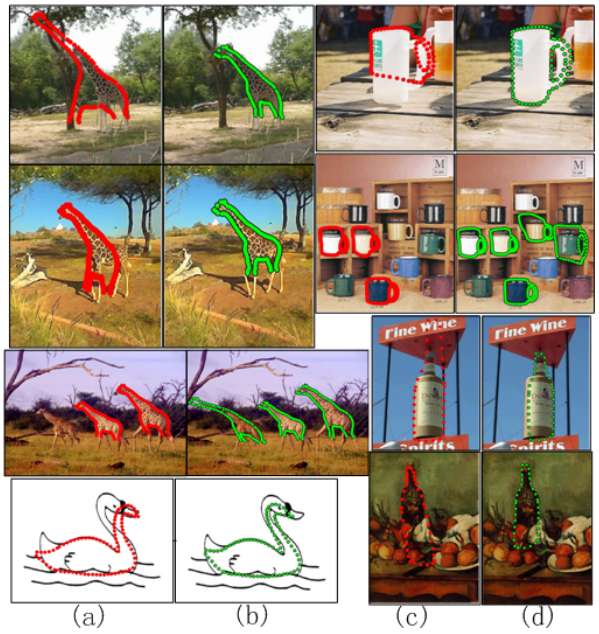

Figure 6. Comparison results of object boundary localization by the method (a)(c) with part importance and (b)(d) without part importance.

4.3. Experimental results and discussions

Figure 6 shows some comparisons of the localized object boundaries by the proposed method using part importance (in green) and those without part importance (in red). It shows that, the method by exploiting part importance locates object boundaries much more accurately. We compute the boundary accuracy similarly to that of [10]. It is the mean value of (1) the percentage of the true positive object boundary points, and (2) the percentage of the recalled groundtruth outline points. The results are shown in Table 1.

It is also found that the detection rate is improved by considering part importance, as shown by the AUC (Area

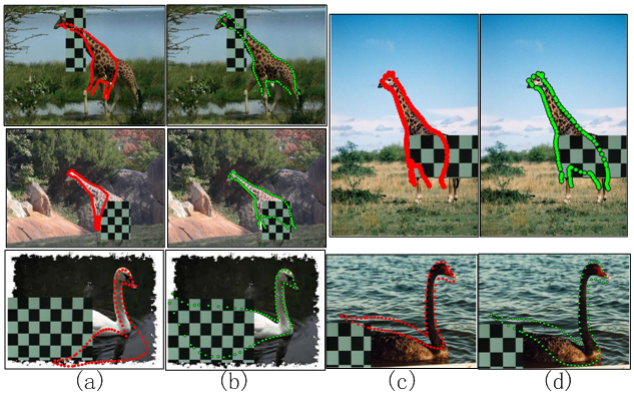

Figure 7. Comparison results of object boundary localization under occlusion by the method (a)(c) with part importance and (b)(d) without part importance.

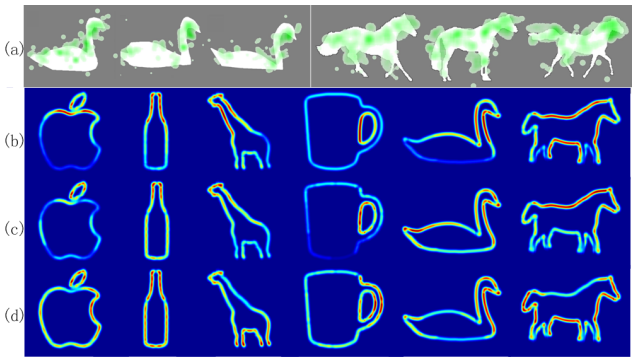

Figure 8. The psychological experiments. (a) Some examples of stimuli (shape silhouettes in white) and the fixation densities (the more green the higher densities). (b)-(d) The heat maps of part importance from the human fixation, our model and the edge orientation distribution respectively (the more red the more important).

Under ROC Curve) statistics in Table 1 and the DR-FPPI (detection rate vs. false positives per image) curves in Figure 5. It worth notice that our experimental results without importance is not the same as that of [10], we learn the parts from object outlines; the results of [10] is based on a hand-drawing object for each class as the shape model.

Part importance for occlusion. Occlusion is a frequently-happening phenomenon in natural images. It breaks most of the state-of-the-art object detection and recognition methods. Here we show that by integrating part importance into object detectors, it greatly improves the detection performance even under severe occlusion.

Due to the lack of dataset for occlusion cases in the literature (to our best knowledge), we synthesize a dataset in which the occluded objects are generated by randomly placing a checkboard (with a random size) in the original images of the ETHZ-dataset (5 random trials for each original test image, so totally hundreds of occlusion images). We can see from the examples in Figure 7 that the proposed importance-integrated method is able to find object boundaries more reliably and accurately than the method without considering part importance, especially when large portions of occlusion happens. On average, the AUC and boundary location accuracy are improved from 0.597 to 0.740 and 0.821 to 0.877 on the occlusion dataset respectively.

The occlusion happens randomly on the object parts according to our synthesized data. We find in our experiments that, in the case of large portions of important parts are occluded, there is little improvement for the detection task by using part importance; it is also difficult for the method without importance to detect the targets.

5. Psychological experiments

In order to verify the proposed importance model, we conduct psychological experiments, in which subjects are asked to implicitly identify objects of the category by observing the object silhouettes. Each object category has 30 instances. The stimulus, an object silhouette, is displayed in the center of the screen for 2 seconds for everyone of 30 subjects. During observation, the subjects' eye fixations are recorded.

Table 2. Performance of importance evaluation

KL dist.	Uniform	Renninger'07	Ours
applelogos	5.9211	3.1491	1.7703
bottles	4.4891	2.6078	0.88269
giraffes	5.3887	2.6752	1.7362
mugs	5.4796	2.1745	1.0639
swans	5.7895	2.3014	1.6353
horses	4.1358	3.0118	2.3233

From the eye fixations we compute human fixation-scores for different parts as follows. First, the density map of the fixations are calculated using the kernel-based method, in which a 2D Gaussian kernel is employed on each fixation (Figure 8 (a)). Then, we collect the fixations with respect to each part based on the aligned and matched shape contours of the category. The fixation-score of a part is measured by the mean value of the accumulated densities in the local band region around the corresponding contour part of the stimuli.

We compare different part importance measures to their fixation-scores (denoted by a vector I_0). Let I_1 be the proposed importance measure, and I_2 be the measure according to the local edge orientation histograms by Renninger *et al.*[22]. In addition, if the parts are considered of equal importance, it corresponds to a uniform distribution $I_3 = 1/N$ (N is the number of the parts). The differences between the three measures to human fixation data are shown in Table 2, which is evaluated by the KL divergences between I_0 and I_1, I_2, I_3 respectively. Figure 8 (b-d) shows the heat maps based on different measures. The results demonstrate that

our model is more coherent to human perception of shape parts.

6. Conclusion and Discussion

A novel method of shape part importance measurement is proposed by the "reconstructability" of a part. It is successfully applied to object detection applications, especially in severe occlusion scenarios. The psychological experiments also verify the consistency of our model with shape perception of humans.

Our current implementation of the proposed method still has some limitations. For example, it is not very robust to articulation, since large articulation induces large part pose variations w.r.t a global shape, rather than deformation of shape parts. In the future, we shall augment our model by introducing pose variables to discount the articulation problem in shape reconstruction based on local parts.

References

[1] I. BiedermanRecognition-by-components: a theory of human image understanding. *Psychol Rev.*, 92:115–147, 1987. 1

[2] G. Bouchard and B. Triggs. Hierarchical part-based visual object categorization. In *IEEE Conf. Computer Vis. and Patt. Recognition*, 2005. 1

[3] G. H. Bower and A. L. Glass. Structural units and the redintegrative power of picture fragments. *J. of Experimental Psychology*, 2:456–466, 1976. 1

[4] H. Cai, F. Yan, and K. Mikolajczyk. Learning weights for codebook in image classification and retrieval. In *IEEE Conf. Computer Vis. and Patt. Recognition*, 2010. 2

[5] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2-3):114–141, 2003. 4, 5, 6

[6] D. J. Crandall and D. Huttenlocher. Weakly supervised learning of part-based spatial models for visual object recognition. In *Euro Conf. Computer Vis.*, 2006. 1

[7] A. Dubinskiy and S. C. Zhu. A multi-scale generative model for animate shapes and parts. In *Proc. of Int'l Conf. on Computer Vision*, 2003. 1

[8] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial structures for object recognition. *Int'l J. of Computer Vision (IJCV)*, 2005. 1

[9] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid. Groups of adjacent contour segments for object detection. *IEEE Trans. on Patt. Analysis and Machine Intelligence (PAMI)*, 2008. Dataset: www.vision.ee.ethz.ch/∼calvin/datasets.html. 1, 5

[10] V. Ferrari, F. Jurie, and C. Schmid. From images to shape models for object detection. *Int'l J. of Computer Vision (IJCV)*, 2009. 5, 6, 7

[11] V. Ferrari, T. Tuytelaars, and L. V. Gool. Object detection by contour segment networks. *Euro Conf. Computer Vis.*, 2006. 5

[12] O. Freifeld, A. Weiss, S. Zuffi, and M. J. Black. Contour people: A parameterized model of 2d articulated human shape. *IEEE Conf. Computer Vis. and Patt. Recognition*, 2010. 1, 2

[13] D. D. Hoffman and W. Richards. Parts of recognition. *Cognition*, 18, 1984. 1

[14] D. D. Hoffman and M. Singh. Salience of visual parts. *Cognition*, 63:29–78, 1997. 1, 2

[15] F. Jurie and C. Schmid. Scale-invariant shape features for recognition of object categories. In *CVPR*, 2004. Dataset: lear.inrialpes.fr/data. 5

[16] G. Kanizsa and W. Gerbino. Amodal completion: Seeing or thinking? *J. Beck (Ed.), Organization and representation in perception*, pages 167–190, 1982. Hillsdale, New Jersey: Lawrence Erlbaum Associates. 4

[17] D. Kersten, P. Mamassian, and A. Yuille. Object perception as bayesian inference. *Annu. Rev Psychol*, 2004. 2

[18] B. Kimia, I. Frankel, and A. Popescu. Euler spiral for shape completion. *Int'l J. of Computer Vision*, 54(1/2):157–180, 2003. 4

[19] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. on Patt. Analysis and Machine Intelligence (PAMI)*, 26(5):530–549, 2004. 5

[20] A. Opelt, A. Pinz, and A. Zisserman. Learning an alphabet of shape and appearance for multi-class object detection. *Int'l J. of Computer Vision (IJCV)*, 2008. 1

[21] Z. Pizlo. A theory of shape constancy based on perspective invariants. *Vision Research*, 34:1637–1658, 1994. 1

[22] L. K. Renninger, P. Verghese, and J. Coughlan. Where to look next? eye movements reduce local uncertainty. *J. of Vision*, 7(3):1–17, 2007. 2, 7

[23] R. A. Rensink and J. T. Enns. Early completion of occluded objects. *Vision Research*, 38:2489–2505, 1998. 1

[24] P. Sala and S. Dickinson. Contour grouping and abstraction using simple part models. *ECCV*, 2010. 1

[25] P. Schnitzspan, S. Roth, and B. Schiele. Automatic discovery of meaningful object parts with latent CRFs. In *IEEE Conf. Computer Vis. and Patt. Recognition*, 2010. 2

[26] J. Shotton, A. Blake, and R. Cipolla. Multi-scale categorical object recognition using contour fragments. *IEEE Trans. on Patt. Analysis and Machine Intelligence (PAMI)*, 2008. 1

[27] K. Siddiqi, B. Kimia, and K. Tresness. Parts of visual form: Psychophysical aspects. *Perception*, 25:399–424, 1996. 1

[28] S. R. Sukumar, D. L. Page, A. F. Koschan, A. V. Gribok, and M. A. Abidi. Shape measure for identifying perceptually informative parts of 3d objects. In *Proc. IEEE 3rd Inter'l Symp. on 3D Data Processing, Visualization and Transmission (3DPVT)*, 2006. 2

[29] S. Ullman. Object recognition and segmentation by a fragment based hierarchy. *Trends Cogn. Sci.*, 11:58–64, 2007. 2

[30] L. Zhu, Y. Chen, A. Yuille, and W. Freeman. Latent hierarchical structural learning for object detection. In *IEEE Conf. Computer Vis. and Patt. Recognition*, 2010. 1

[31] S. C. Zhu. Stochastic jump-diffusion process for computing medial axes in markov random fields. *IEEE Trans. on Patt. Analysis and Machine Intelligence (PAMI)*, 1999. 1

[32] S. C. Zhu, Y. N. Wu, and D. B. Mumford. Filters, random field and maximum entropy(frame): Towards a unified theory for texture modeling. *IEEE Trans. on Patt. Analysis and Machine Intelligence (PAMI)*, 1998. 2