Grouping Active Contour Fragments for Object Recognition

Wei Zheng[1,2], Songlin Song[1,2], Hong Chang[1], and Xilin Chen[1]

[1] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, 100190, China
[2] Graduate School of the Chinese Academy of Sciences, Beijing, 100039, China
{wei.zheng,songlin.song,hong.chang,xilin.chen}@vipl.ict.ac.cn

Abstract. In this paper, we try to address the challenging problem of combining local shape features to describe long and continuous shape characteristics. To this end, we firstly propose a novel type of local shape feature, namely Active Contour Fragment (ACF), to encode the shape deformation in a local region. An ACF is automatically learnt from the contours of a specific object class and capable to describe the intra-class shape characteristics based on the point distribution model. Secondly, we combine multiple ACFs into a group, namely Active Contour Group (ACG), to describe the long shape characteristics .We model the ACFs in an ACG using an undirected chain model and estimate the parameters of the chain model in a subspace for accelerating the learning and matching processes of ACGs. Finally, we discriminatively train the classifiers based on ACFs and ACGs in a boosting framework for localizing objects as well as delineating object boundaries. Both qualitative and quantitative evaluations show that our approach is capable of describing long shapes and the proposed recognition algorithm achieves promising performance on the public datasets.

1 Introduction

Shape-based object recognition has been extensively studied [1] [2], since shape is an informative and stable cue to distinguish the target objects from background.

Lots of local shape templates [3] [4] [5] can provide an explicit, effective and efficient representation for shapes, thus they are widely utilized for localizing and delineating objects in real-scene images. Ideal shape features should be both descriptive to fit the object boundaries and discriminative to reject the non-object shapes. Therefore, the local shape templates should be deformable in case they are misaligned with the object boundaries as shown in Fig. 1 (a). However, deformable local shape templates may falsely match to the inner parts or cluttered background since they ignore shape constraints in larger regions. In Fig. 1 (a), we show three local shape templates (red fragments) and their matching results (green fragments). These local shape templates are deformable, and their matching results are in the inner or background edges according to some matching algorithm (to be discussed in Sec. 3.2). The knowledge from perceptual grouping inspires us to group the local shape templates and jointly match them

K.M. Lee et al. (Eds.): ACCV 2012, Part I, LNCS 7724, pp. 289–301, 2013.

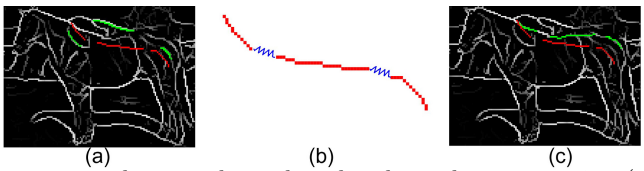


Fig. 1. Comparing matching results with and without chain constraint. (a) Deformable shape templates may fit to inner edges or cluttered background. (b) Bundling shape templates with spring-like connections to describe long shapes. (c) Chained shape templates match to long boundary of horse back.

to images. We believe that the matching results of neighboring shape templates should preserve the continuation property if they match to the same object, i.e., they may simultaneously match to a long and continuous contour instead of scattered ones. Intuitively, we add a spring-connection between the neighboring templates and group the shape templates into a long chain to guarantee the continuation property (see Fig. 1 (b)). The three templates may match to the long boundary of the horse back with the spring-connections as shown in Fig. 1.

Our motivation is to encode the local deformations using deformable shape templates and then group them to describe the long and continuous shape characteristics. We highlight this paper from three folds. Firstly, we propose a novel type of local shape feature, namely, Active Contour Fragment (ACF). An ACF encodes the deformation based on the Point Distribution Model (PDM), which can be learnt from weakly labelled data. Secondly, we combine different ACFs into one group, namely, Active Contour Group (ACG), to describe long and continuous shapes with an undirected chain model. We propose an efficient inference algorithm for the chain model by estimating the parameters in a linear subspace that is learnt from training data. Then, we present an incremental learning algorithm to group the ACFs into ACGs based on an efficient inference algorithm. Finally, we demonstrate the performance of ACGs for object recognition in a boosting framework. The object classifiers can not only localize the objects with bounding boxes but also delineate object boundaries.

We evaluate the proposed approach on Weizmann Horses dataset [5] and ETHZ shape dataset [6]. Experiments show that ACGs can fit to the object boundaries well. The object classifiers based on ACFs and ACGs achieve promising performance on object detection and boundary localization.

2 Related Works

We review the related works from the following two aspects, namely, shape representation and perceptual grouping.

Shape representation has been extensively studied recently. Some researchers represent shapes with a group of local shape templates [3] [4] [5] or descriptors [7] [8], while others represent shapes with a global shape model. One merit of local shape features is that they may be robust to the occlusions and T-junctions in the real-scenes. For many local shape templates, there are two additional merits: 1) they describe shapes explicitly with contour segments; 2) they are efficient in computation. Therefore, many researchers utilize the local shape templates

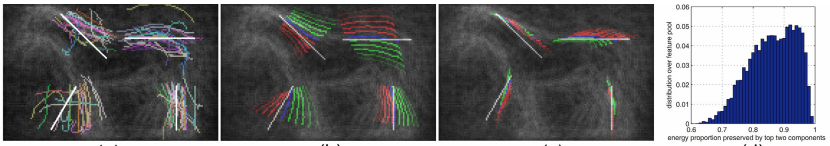


Fig. 2. Learnt ACFs. (a) Shape bases (white line fragments) and best matching contours (color fragments). (b) (c) First and second principal components. (d) Distribution of energy proportion preserved by top two principal components over feature pool.

for object detection, such as edgelet [4], strip [3], contour fragments [5] and boundary fragments [9]. However, the local shape features ignore the global shape constraint and susceptible to background clutter, thus many researchers propose lots of global shape models, such as hand-drawn skeleton or silhouette [10] [11], boundary structure model [12], Shape Boltzmann Machine (SBM) [13] and Point Distribution Model (PDM) [14] [1]. Our approach is different from the above local shape features and global shape models. On one side, the proposed ACGs encode the long and continuous part of shapes instead of the global shape characteristics. On the other side, the proposed ACGs encode more shape constraints than the local shape features.

Perceptual grouping is a hot topic in the area of computer vision. The basic assumption behind these works is that the contour fragments that are related by some perceptually properties should belong to the same object. The perceptual properties include continuation [15], parallelism [16], closure [17] and so on. Such properties may be the driving force for designing feature bundle [18] or learning shape model [8]. Some recent works [1] [6] group k Adjacent Segments (kAS) to represent generic shapes. Different from kAS, the proposed ACGs are class-specific and thus may be more suitable for object recognition of a specific class.

3 Active Contour Fragments

We firstly propose to automatically learn ACFs from the weakly unsegmented training data. Then, we present the matching algorithm of ACFs.

3.1 Learning ACFs from Contours

For a specified object class, we label the training objects with bounding boxes and normalize them into an object window of a specific size (see Table 1). To obtain the contours of objects, we extract the edge maps using the Berkeley edge detector [19] and link the edgels (edge pixels) [20]. The j^{th} object can be represented by a *contour set* as

$$\mathbf{C}_j = \{\mathbf{c}_{j,1}, \mathbf{c}_{j,2}, ..., \mathbf{c}_{j,|\mathbf{C}_j|}\}, \tag{1}$$

where $\mathbf{c}_{j,k}$ is a contour (represented by a list of edgels). As shown in Fig. 2 (a), the horse backs have simple translation variance while the horse legs have both large articulation and non-rigid deformations. Such shape characteristics can be described by the contour fragments. We propose to encode the deformation from

the contour fragments. Our approach consists of two steps, i.e., grouping contour fragments and learning the deformation from the contour fragments.

To group the contour fragments, we utilize a set of shape bases $\{\mathbf{b}_i\}_{i=1,...,N}$. The shape bases are generated by uniformly sampling line segments of 12∼48 pixels in the object window (see Table 1). We match each shape basis to the contour set of training images and find the best matching results according to

$$\mathbf{f}^*_{i,j} = \underset{|\mathbf{f}_{i,j}|=|\mathbf{b}_i|, \mathbf{f}_{i,j} \subseteq \mathbf{c}_{j,k}, \mathbf{c}_{j,k} \in \mathbf{C}_j}{argmin} \Phi(\mathbf{b}_i, \mathbf{f}_{i,j}), \tag{2}$$

where $\mathbf{f}_{i,j}$ is a *contour fragment* that has the same number of edgels with $\mathbf{b}_i$ on $\mathbf{c}_{j,k}$ and $\Phi(\bullet)$ represents the following distance function

$$\Phi(\mathbf{b}_i, \mathbf{f}_{i,j}) = \sum\nolimits_{p=1}^{|\mathbf{b}_i|} \|\mathbf{b}_i(p) - \mathbf{f}_{i,j}(p)\|_D, \tag{3}$$

where $\mathbf{b}_i(p)$ represents the p^{th} edgel of $\mathbf{b}_i$ and we use the same notation in the following. We measure the distance $\| \bullet \|_D$ between two edgels as

$$\|\mathbf{e}_i - \mathbf{e}_j\|_D = \beta((x_i - x_j)^2 + (y_i - y_j)^2)) + (\theta_i - \theta_j)^2, \tag{4}$$

where an edgel $\mathbf{e}_i$ is represented by its coordinates (x_i, y_i) and normal orientation θ_i. The factor β is a constant that balances the importance of position and rotation. We let β equal to $25/A$, where A is the area of the object window. In Fig. 2 (a), we visualize four shape bases and the matched contours on 20 horses.

Based on the matched contour fragments, we adopt the PDM to model the shape deformation. Suppose the matched contour fragments $\mathbf{f}^*_{i,j}$ distribute in a $2|\mathbf{f}^*_{i,j}|$-dimensional space, the shape model can be obtained by Principal Component Analysis (PCA). The i^{th} ACF ψ_i generated by $\mathbf{b}_i$ is formulated as

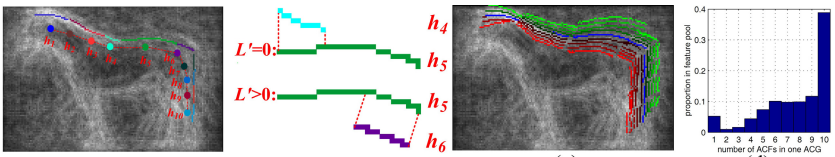
$$\psi_i : \mathbf{t}_i^{\alpha_{i,1},...,\alpha_{i,K}} = \overline{\mathbf{s}}_i + \sum\nolimits_{k=1}^{K} \alpha_{i,k} \mathbf{v}_{i,k}, s.t. |\alpha_{i,k}| \leq 2\sqrt{\lambda_{i,k}}, \tag{5}$$

where $\overline{\mathbf{s}}_i$ is the averaged contour of $\{\mathbf{f}^*_{i,j}\}_{j=1,..,M}$, $\{\mathbf{v}_{i,k}, \lambda_{i,k}\}_{k=1,,K}$ are the top K eigenvectors and eigenvalues returned by PCA and $\mathbf{t}_i^{\alpha_{i,1},...,\alpha_{i,K}}$ is a reasonable shape generated by the shape model. The eigenvectors encode the deformation and the eigenvalues reflect the importance of deformation. Fig. 2 (b) and (c) show the deformation described by the top two components. The white line fragments are shape bases and the blue ones are averaged contour fragments. We vary the first coordinate from $-2\sqrt{\lambda_{i,k}}$ to $2\sqrt{\lambda_{i,k}}$, and the shape deforms from dark red to bright red then from bright green to dark green. Fig. 2 (d) shows that most of ACFs preserve more than 80% energy by the top two principal components.

3.2 Matching ACFs to Images

Hereby, we give the matching algorithm for ACFs. As discussed in Sec. 3.1, most of ACFs can be briefly represented by only two coordinates as in Eqn. 5. Then, the matching energy of the ACF ψ_i is defined as

$$E_{\psi_i}(\mathbf{E}) = \min_{\alpha_{i,1},\alpha_{i,2}} \sum\nolimits_{p=1}^{|\mathbf{t}_i^{\alpha_{i,1},\alpha_{i,2}}|} \min_{\mathbf{e}_p \in \mathbf{E}} \|\mathbf{t}_i^{\alpha_{i,1},\alpha_{i,2}}(p) - \mathbf{e}_p\|_D, \tag{6}$$



(a) (b) (c) (d)

Fig. 3. Illustration of learnt ACGs. (a) Learnt ACG on which numbers are indexes of ACFs. (b) Pairwise energy is measured by connected segments of same length when two neighboring ACFs are of different lengths. (c) First principal component of learnt ACG. (d) Distribution of ACF numbers in each ACG over feature pool.

where $\mathbf{e}_p$ is an edgel of the edge map $\mathbf{E}$. When fixing $\alpha_{i,1}$ and $\alpha_{i,2}$, the energy can be fast calculated via look-up table based on the distance transform as discussed in [5]. To match ACFs to images, we should minimize Eqn. 6 over $\alpha_{i,1}$ and $\alpha_{i,2}$. We quantize $\alpha_{i,1}$ into $N_1 = 11$ values from $-2\sqrt{\lambda_{i,1}}$ to $2\sqrt{\lambda_{i,1}}$ and $\alpha_{i,2}$ into $N_2 = 7$ values from $-2\sqrt{\lambda_{i,2}}$ to $2\sqrt{\lambda_{i,2}}$. Once we find the best configuration $(\alpha^*_{i,1},\alpha^*_{i,2})$, we obtain the matching curve $\mathbf{t}^*$ according to Eqn. 5.

4 Active Contour Groups

We firstly model multiple ACFs with an undirected chain model to describe longer shapes. Then, we give the matching and learning algorithms for ACGs.

4.1 Chain Model for ACGs

We define an *Active Contour Group (ACG)* as a group of ACFs as

$$\Psi_h = \{\psi_{h_1}, \psi_{h_2}, .., \psi_{h_{|\Psi_h|}}\}, \tag{7}$$

where the i^{th} element of the h^{th} ACG corresponds to an ACF. In this paper, we suppose the ACFs in an ACG are organized under an undirected chain model as shown in Fig. 3 (a). Each vertex of the chain corresponds to an ACF and each edge of the chain describes the pairwise relationship of two neighboring ACFs. Formally, we define the neighborhood system of the chain model as

$$\mathbf{N} = \{(\psi_{h_i}, \psi_{h_j}) | \psi_{h_i} \in \Psi_h, \psi_{h_j} \in \Psi_h, |i-j| = 1\}, \tag{8}$$

Our objective is to use the chain model to encode long and continuous shape characteristics, thus we require the neighboring ACFs to preserve the continuation property. Two neighboring ACFs should satisfy the following two conditions

$$\begin{aligned} &\|\overline{\mathbf{s}}_{h_i}(|\overline{\mathbf{s}}_{h_i}|) - \overline{\mathbf{s}}_{h_{i+1}}(1)\|_2 \leq \Gamma, \\ &\|\overline{\mathbf{s}}_{h_i}(1) - \overline{\mathbf{s}}_{h_{i+1}}(|\overline{\mathbf{s}}_{h_{i+1}}|)\|_2 \geq \max(|\overline{\mathbf{s}}_{h_i}|^2, |\overline{\mathbf{s}}_{h_{i+1}}|^2), \end{aligned} \tag{9}$$

where $\overline{\mathbf{s}}_{h_i}$ is the average contour fragment of ψ_{h_i}. Γ is a constant that is set as 4 in our experiments. Taking the two neighboring ACFs (e.g., ψ_{h_1} and ψ_{h_2}) in Fig. 3 (a) as an example, ψ_{h_1} should be connected with ψ_{h_2} on one end and should not be connected with ψ_{h_2} on the other end. In the learning process (see

Sec.4.2), we require that the neighboring ACFs should satisfy Eqn. 9. Hereby, we define the matching energy of an ACG as

$$E_{\Psi_h}(\mathbf{E}) = \sum\nolimits_{\psi_{h_i}\in\Psi_h} E_{\psi_{h_i}}(\mathbf{E}) + \sum\nolimits_{(\psi_{h_i},\psi_{h_j})\in\mathbf{N}} E_N(\psi_{h_i},\psi_{h_j}), \tag{10}$$

where $E_{\psi_{h_i}}(\bullet)$ is the unary energy of the i^{th} ACF defined in Eqn. 6 and $E_N(\bullet,\bullet)$ is the pairwise energy of two neighboring ACFs, which is defined as

$$\begin{aligned} E_N(\psi_{h_i},\psi_{h_{i+1}}) = \sum\nolimits_{p=1}^{L} \|(\mathbf{t}_{h_i}^{\alpha_{h_i,1},\alpha_{h_i,2}}(L'+p) - \overline{\mathbf{s}}_{h_i}(L'+p))) \\ - (\mathbf{t}_{h_{i+1}}^{\alpha_{h_{i+1},1},\alpha_{h_{i+1},2}}(p) - \overline{\mathbf{s}}_{h_{i+1}}(p))\|_D, \end{aligned} \tag{11}$$

where the symbols are the same with Eqn. 5, $L = \min(|\mathbf{t}_{h_i}|, |\mathbf{t}_{h_{i+1}}|)$ and $L' = \max(|\mathbf{t}_{h_i}| - |\mathbf{t}_{h_{i+1}}|, 0)$. As shown in Fig. 3, we pick up two connected segments of the same length from ψ_{h_i} and $\psi_{h_{i+1}}$ when the neighboring ACFs are of different lengths. Then, we calculate the pairwise energy by accumulating the deformation differences of the two connected segments. For a good matching result, the pairwise energy should be small since the neighboring segments should have similar deformations. As shown in Fig. 3, h_5 and h_6 match to the back of the same horse, thus they should deform to high position or low position simultaneously. Since the neighboring ACFs are connected and have similar deformations in the chain model, the deformable ACG may always preserve the continuation property.

4.2 Learning ACGs from ACFs

To match ACGs to images, we should minimize Eqn. 10 over a $2|\Psi_h|$-dimensional parameter space $\{\alpha_{h_1,1}, \alpha_{h_1,2}, ..., \alpha_{h_{|\Psi_h|},1}, \alpha_{h_{|\Psi_h|},2}\}$ but solving the minimization problem is computational prohibitive. For an ACG that consists of 10 ACFs, the computation complexity of the brute-force search is $O((N_1N_2)^{10})$ if we use the same quantization criterion for $\alpha_{h_i,1}, \alpha_{h_i,2}$ as dicussed in Sec.3.2. Although there are some fast inference approaches (e.g., belief propagation [21]), the computation complexity is still too high. Hereby, we propose an efficient matching algorithm that reduces the computation complexity from $O(N_1N_2)^{10}$ to $O(N_1N_2)$. Then, we present an incremental learning algorithm of ACGs based on the efficient matching algorithm.

The basic idea of the proposed matching algorithm is to constrain the parameter space of Eqn. 10 using the training data. Supposing that Ψ_h is an ACG that is defined in Sec. 4.1, the matching result of Ψ_h on the j^{th} training sample is represented as a $2|\Psi_h|$-dimensional vector

$$\mathbf{\Theta}^*_{h,j} = \{\alpha^*_{h_1,1,j}, \alpha^*_{h_1,2,j}, \alpha^*_{h_2,1,j}, \alpha^*_{h_2,2,j}, \ldots, \alpha^*_{h_{|\Psi_h|},1,j}, \alpha^*_{h_{|\Psi_h|},2,j}\}. \tag{12}$$

$\mathbf{\Theta}^*_{h,j}$ can be obtained according to Eqn. 14 (to be discussed later). Once we have the matching results of training samples, we can derive a linear subspace using PCA. Then, we estimate the matching parameters of Ψ_h according to

$$\Psi_h : \mathbf{\Theta}_h^{\gamma_{h,1},\ldots,\gamma_{h,K}} = \overline{\mathbf{d}}_h + \sum\nolimits_{k=1}^{K} \gamma_{h,k}\mathbf{u}_{h,k}, s.t. |\gamma_{h,k}| \le 2\sqrt{\eta_{h,k}}, \tag{13}$$

where $\overline{\mathbf{d}}_h$ is the averaged vector of the matching results over the training set and $\{\mathbf{u}_{h,k}, \eta_{h,k} | k = 1, ..., K\}$ are the top K eigenvectors and eigenvalues returned by

PCA. We only consider the top two principal components. To find the best $\gamma_{h,1}$ and $\gamma_{h,2}$, we quantize $\gamma_{h,1}$ into $N_1 = 11$ values from $-2\sqrt{\eta_{h,1}}$ to $2\sqrt{\eta_{h,1}}$ and $\gamma_{h,2}$ into $N_2 = 7$ values from $-2\sqrt{\eta_{h,2}}$ to $2\sqrt{\eta_{h,2}}$ as discussed in Sec. 3.2. Then, we match ACGs to the images by minimizing Eqn. 10 over $\gamma_{h,1}$ and $\gamma_{h,2}$ and the computational complexity is $O(N_1 N_2)$. After obtaining the best parameters $\gamma^*_{h,1}$ and $\gamma^*_{h,2}$, we can derive the matching curve according to Eqn. 13 and Eqn. 5.

Hereby, we give the learning algorithm based on Eqn.13 and explain how to calculate $\mathbf{\Theta}^*_{h,j}$ in Eqn. 12. Every ACG starts from one ACF and grows into a long chain. At first, we select an ACF as the seed of the chain, which is considered as the initial ACG. Then, we find all the candidate ACFs that are neighboring with the head or the tail of the ACG according to Eqn. 9. Supposing that a candidate ACF ψ_c is neighboring with the head of the ACG Ψ_h, we add ψ_c in front of the current ACG's head. For the new ACG, the matching energy on the j^{th} training sample $\mathbf{E}_j$ can be calculated incrementally according to

$$E_{\{\psi_c,\Psi_h\}}(\mathbf{E}_j) = E_{\Psi_h}(\mathbf{E}_j) + E_{\psi_c}(\mathbf{E}_j) + E_N(\psi_c, \psi_{h_1}), \tag{14}$$

We can obtain the new matching result $\{\alpha^*_{c,1,j}, \alpha^*_{c,2,j}, \mathbf{\Theta}^*_{h,j}\}$ by minimizing Eqn. 14. To this end, we search the solution over the solution space $\{\alpha_{c,1,j}, \alpha_{c,2,j}, \mathbf{\Theta}_{h,j}\}$. Since $\mathbf{\Theta}_{h,j}$ can be estimated using the linear subspace according to Eqn. 13, the computation complexity of minimizing Eqn. 14 exhaustively is $O((N_1 N_2)^2)$ if we use the same quantization strategy over the solution space as discussed before. For each candidate ACF, we sum the minimized energy in Eqn. 14 over the training set and select the one with the minimum energy. Finally, we add the selected ACF as the head of the chain and update the subspace in Eqn. 13. When the candidate ACF ψ_c is connected with the tail of the ACG Ψ_h, we can grow the ACG in a similar way. We repeat the growing process until the subspace cannot preserve minimum energy or the number of ACFs in one ACG exceeds the maximum number. We set the minimum energy as 95% of the total energy and the maximum number of ACFs as 10.

Each ACF can grow into an ACG according to the above algorithm. Fig. 3 (a) shows one learnt ACG. In Fig. 3 (c), we show the first principal component of the learnt subspace. In Fig. 3 (d), we can see that most of ACGs have more than 1 ACF. To be mentioned, it is difficult to directly learn long ACFs from images, since the T-junctions and occlusions of the edges may destroy the long shape characteristics. However, the proposed algorithm can group the short shape templates (i.e., ACF) into long shape templates(i.e., ACG).

5 Recognition Algorithm

We utilize ACFs or ACGs as shape features for object recognition in a boosting framework. Similar to [4] [5] [9], we match the ACFs or ACGs to images by minimizing the energy function (Eqn. 6 for ACFs and Eqn. 10 for ACGs) and use the minimum energy as feature scores. Then, we select ACFs or ACGs from feature pools into a cascaded classifier using RealBoost [22]. For acceleration, we use Histogram of Oriented Gradients (HOG) in the first 10 stages (see [23] for detials). We give an evaluation for the HOG filter in Sec. 6.2.

Table 1. Sizes of detection window and feature pools of ACFs and ACGs

	Horses	Applelogos	Bottles	Giraffes	Mugs	Swans
Detection window	150 × 100	80 × 80	50 × 120	150 × 150	90 × 70	120 × 60
Feature pool	10315	6627	4070	14815	6137	7052

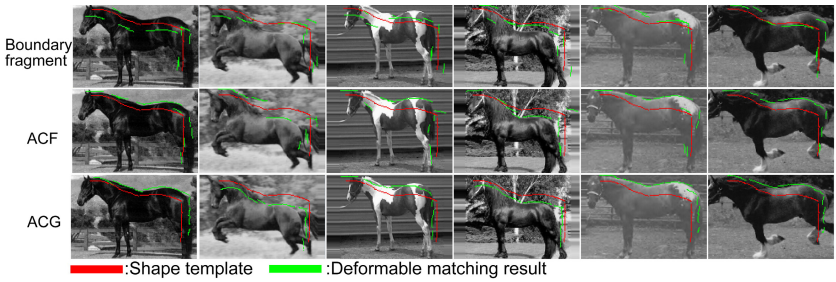


Fig. 4. Comparing matching curves of boundary fragments,ACFs and ACGs

The boundary localization is only conducted on the object windows. An object is identified when a detection window is classified as positive at a false positive rate of 10^{-5}. The final boundary localization results can be obtained by averaging the results of multiple sliding windows in all the scales. To this end, we give the voting algorithm in one detected window. For each ACF (or ACG), we can obtain a matching curve after minimizing Eqn. 6 (or Eqn. 10). Each point $\mathbf{m}$ on the matching curve casts a vote for the detection window according to

$$P(b_{\mathbf{x}}=1)=\sum_{b_{\mathbf{m}}=0,1}P(b_{\mathbf{x}}=1|b_{\mathbf{m}})P(b_{\mathbf{m}}) \quad =\mathtt{I}(\mathbf{x}==\mathbf{m})\frac{W^{+}_{f_{\mathbf{m}}}}{W^{+}_{f_{\mathbf{m}}}+W^{-}_{f_{\mathbf{m}}}}, \qquad (15)$$

where $\mathbf{x}$ represents any point in the detection window, $b_{\mathbf{x}}$ identifies whether the point $\mathbf{x}$ is a boundary ($b_{\mathbf{x}}=1$) or not ($b_{\mathbf{x}}=0$), $P(b_{\mathbf{x}}=1|b_{\mathbf{m}}=0)$ is supposed to be 0, $\mathtt{I}(\mathbf{x}==\mathbf{m})$ equals to 1 if $\mathbf{x}==\mathbf{m}$ or else equals to 0, $f_{\mathbf{m}}$ is the matching score corresponding to the matching point $\mathbf{m}$, $W^{+}_{f_{\mathbf{m}}}$ (or $W^{-}_{f_{\mathbf{m}}}$) is the proportion of positive (or negative) samples when the feature response equals to $f_{\mathbf{m}}$. $W^{+}_{f_{\mathbf{m}}}$ and $W^{-}_{f_{\mathbf{m}}}$ can be efficiently derived using look-up table since $f_{\mathbf{m}}$ is calculated in the detection process. Fig. 8 shows some boundary localization results.

6 Experiments

We show the effectiveness and efficiency of the proposed matching approach in this section. In Table 1, we list the sizes of detection windows and feature pools of ACFs and ACGs for each object class used in our experiments.

6.1 Comparing Matching Approaches on Weizmann Horses Dataset

We adopt Weizmann Horses dataset to evaluate the proposed matching approach. We use the training-testing split (100 horses for training and 456 images

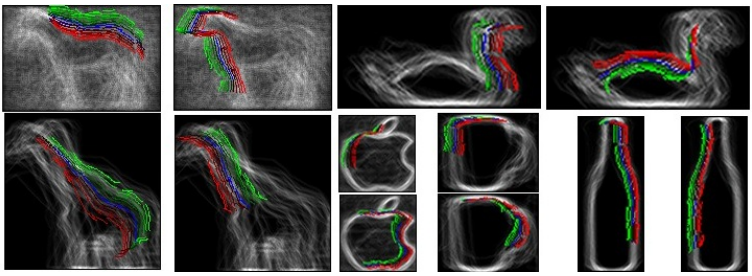


Fig. 5. Top two selected ACGs capture long and salient shape characteristics

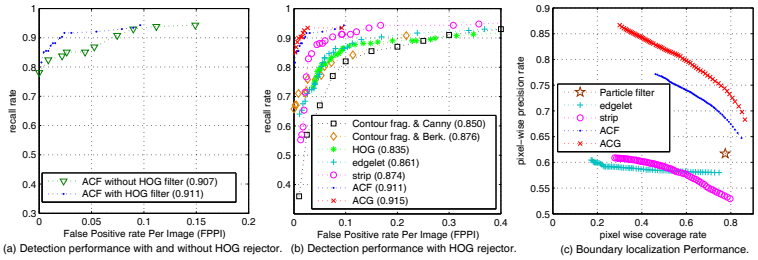


Fig. 6. Detection and boundary localization performance on Weizmann Horses

including 228 horse images for testing) as in [5] [24].We learn ACFs and ACGs from the 100 horses and match them to the testing horses. We compare the matching results of ACFs, ACGs and boundary fragments [9]. The boundary fragments use the averaged contour fragments of ACFs as templates and match to images by translating the templates in a local region of 10×10 pixels. We omit the quantitative comparison due to space limit, but present some qualitative comparison in Fig.4. It can be seen that both ACFs and ACGs can fit to the object boundaries more accurately than the boundary fragment. Furthermore, ACGs do not only fit the object boundaries well, but also preserve the continuation of the shape. We also show the top two features that are selected by the boosting algorithm in Fig. 5. Apparently, these features capture the long and salient shapes and delineate the major shapes for each object class.

6.2 Recognition Results on Weizmann Horses Dataset

We evaluate our method on object detection using PASCAL IoU 50% criterion [25]. In Fig. 6 (a) (b), we plot the recall rate against False Positives Per Image (FPPI) and give the Average Precision (AP) in the legend for each approach. We give a quantitative evaluation for the 10-stage HOG filter proposed in Sec.5. We implement two cascaded classifiers using ACFs: one uses 10-stage HOG filter and the other uses only ACFs. We can see that the one with HOG filter is slightly better than the other, and the HOG filter can also accelerate the search process. Thus, we use the 10-stage HOG filer in the following experiments. We compare our approach with other shape-based approaches, namely, *contour fragment* [5],

Table 2. Comparing APs of different approaches on ETHZ shape dataset

Shape classes	Applelogos	Bottles	Giraffes	Mugs	Swans
Fan shape model [2]	0.866	**0**.975	**0**.832	0.843	0.828
Many-to-one [8]	0.845	0.916	0.787	**0**.888	**0**.922
Grouping with PF [28]	0.844	0.641	0.617	0.643	0.798
OB+GB [29]	0.675	0.781	0.585	0.559	0.661
Dominant Set [18]	0.705	0.761	0.687	0.625	0.773
kAS [6]	0.351	0.733	0.391	0.476	0.273
TPS-RPM [1]	0.689	0.643	0.333	0.585	0.390
Ours (ACFs)	0.910	0.847	0.791	0.853	0.674
Ours (ACGs)	**0**.920	0.864	0.782	0.856	0.653

edgelet [4] and *strip* [3]. We use 10-stage HOG filer in the cascaded classifiers with ACFs and ACGs. For contour fragment, we report two results, which are based on Canny edge (referred from [5]) and Berkeley edge (implemented by ourselves). The proposed approach uses deformable shape templates while contour fragment uses fixed shape template, which is the only difference between these two approaches. We can see the deformation algorithm improves the detection performance by 3% $\sim$ 4%. Our approach outperforms the other shape features since ACFs and ACGs are capable of learning the deformation of horses before the matching process. Recently, some approaches [24] [26] achieve even better performance than our approach. These approaches combine shape features with texture and color features, while our approach only uses shape features.

To quantitatively evaluate our method on boundary localization, we show the coverage-against-precision curves [1] averaged over correct detections when the false positive rate equals 10^{-5} and the curves are obtained by varying the threshold of boundary probability. In Fig. 6 (c), We compare our approach with *particle filter* [27], *edgelet* [4] and *strip* [3]. For particle filter, we refer to the reported result. For other approaches, we vote for boundary probability according to Eqn. (15). Apparently, our approach outperforms all the other approaches. Compared with ACFs, ACGs do not obviously improve the detection performance but they improve the boundary localization performance. The possible reason is that ACGs can capture long and salient shape characteristics and suppress the false matches on the background clutter.

6.3 Recognition Results on ETHZ Shape Dataset

We evaluate our approach in detection and boundary localization on the ETHZ shape dataset. This dataset includes 5 classes, namely, *applelogo*, *bottle*,*giraffe*, *mug* and *swan*. We follow the same training and testing splits suggested by [6].

For object detection, we report the results of our approach and some related methods in Table 2. For *kAS* and *TPSRPM*, we estimate the APs from [1]. Obviously, both ACFs and ACGs achieve competitive performance. For evaluating boundary localization, we compare the results of our approach with those of *TPSRPM* and *Bounding Boxes (B.B.)* reported in [1] in Fig. 7. Apparently, both

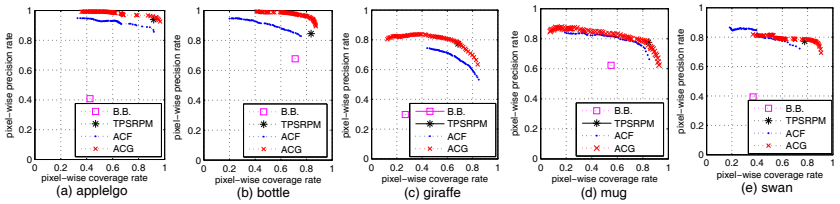


Fig. 7. Quantitative evaluation of boundary localization using coverage-against-precision curves on ETHZ shape dataset

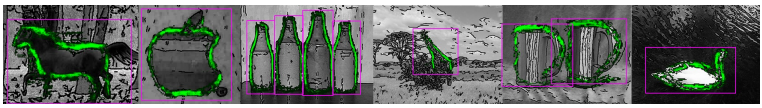


Fig. 8. Recognition examples. Black points are edgels while green points indicate boundary probability. Bright green means higher confidence and vice versa.

ACFs and ACGs substantially outperform the *B.B.* especially for the less rectangular objects, i.e., swans and giraffes. ACGs achieve similar or better performance comparing with *TPSRPM*. Comparing with ACFs, ACGs achieve better precision in most cases, as ACGs can suppress many false matches on the inner or background edges. Fig. 8 shows some recognition examples.

7 Conclusion and Future Works

This paper proposes a novel type of local shape features (i.e., ACFs) and an algorithm to group the local shape features together (i.e., ACGs) for object recognition. The proposed features encode the local deformation based on the PDM and the grouping algorithm combines multiple local shape features to describe long and continuous shapes. We assemble ACFs and ACGs under the boosting framework for localizing objects as well as delineating object boundaries. Qualitative and quantitative experiments show that the proposed approach is effective and efficient for object detection and boundary localization. Furthermore, the proposed approach does not require segmented data in the training process.

The ACGs are still not global shape features although they can describe long and continuous shape characteristics. We will consider some other perceptual properties (e.g., closure and symmetry) in the grouping process to make ACGs more global and further improve the performance of object recognition. Furthermore, ACGs might be promising for foreground segmentation based on the shape information. We will pursue this possibility in the future.

Acknowledgement. This work is partially supported by Natural Science Foundation of China (NSFC) under contract Nos. 61025010, 60832004, and 61001193; and Beijing Natural Science Foundation (New Technologies and Methods in Intelligent Video Surveillance for Public Security) under contract No. 4111003.

References

1. Ferrari, V., Jurie, F., Schmid, C.: From images to shape models for object detection. International Journal of Computer Vision 87, 284–303 (2010)
2. Wang, X., Bai, X., Ma, T., Liu, W., Latecki, L.: Fan shape model for object detection. In: CVPR (2012)
3. Zheng, W., Liang, L.: Fast car detection using image strip features. In: CVPR (2009)
4. Wu, B., Nevatia, R.: Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. International Journal of Computer Vision 75, 247–266 (2007)
5. Shotton, J., Blake, A., Cipolla, R.: Multiscale categorical object recognition using contour fragments. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 1270–1281 (2008)
6. Ferrari, V., Fevrier, L., Jurie, F., Schmid, C.: Groups of adjacent contour segments for object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 36–51 (2008)
7. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 509–522 (2002)
8. Srinivasan, P., Zhu, Q., Shi, J.: Many-to-one contour matching for describing and discriminating object shape. In: CVPR (2010)
9. Opelt, A., Pinz, A., Zisserman, A.: A Boundary-Fragment-Model for Object Detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part II. LNCS, vol. 3952, pp. 575–588. Springer, Heidelberg (2006)
10. Bai, X., Wang, X., Latecki, L., Liu, W., Tu, Z.: Active skeleton for non-rigid object detection. In: ICCV (2009)
11. Ferrari, V., Tuytelaars, T., Van Gool, L.: Object Detection by Contour Segment Networks. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part III. LNCS, vol. 3953, pp. 14–28. Springer, Heidelberg (2006)
12. Toshev, A., Taskar, B., Daniilidis, K.: Object detection via boundary structure segmentation. In: CVPR (2010)
13. Eslami, S., Heess, N., Unit, G., Winn, J.: The shape boltzmann machine: a strong model of object shape. In: CVPR (2012)
14. Cootes, T., Taylor, C., Cooper, D., Graham, J., et al.: Active shape models-their training and application. Computer Vision and Image Understanding 61, 38–59 (1995)
15. Rothwell, C., Zisserman, A., Forsyth, D., Mundy, J.: Planar object recognition using projective shape representation. International Journal of Computer Vision 16, 57–99 (1995)
16. Lowe, D.: Three-dimensional object recognition from single two-dimensional images. Artificial Intelligence 31, 355–395 (1987)
17. Ming, Y., Li, H., He, X.: Connected contours: a new contour completion model that respects the closure effect. In: CVPR (2012)
18. Yang, X., Liu, H., Latecki, L.J.: Contour-based object detection as dominant set computation. In: ACCV (2010)
19. Martin, D., Fowlkes, C., Malik, J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Transactions on Pattern Analysis and Machine Intelligence 26, 530–549 (2004)

20. Suzuki, S., et al.: Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics and Image Processing 30, 32–46 (1985)
21. Pearl, J.: Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann (1988)
22. Schapire, R., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. Machine Learning 37, 297–336 (1999)
23. Zhu, Q., Yeh, M., Cheng, K., Avidan, S.: Fast human detection using a cascade of histograms of oriented gradients. In: CVPR (2006)
24. Gall, J., Yao, A., Razavi, N., Van Gool, L., Lempitsky, V.: Hough forests for object detection, tracking, and action recognition. IEEE Trans. Pattern Anal. Mach. Intell. 33, 2188–2202 (2011)
25. Everingham, M., Zisserman, A., Williams, C., Van Gool, L.: The pascal visual object classes challenge 2006 (voc 2006) results (2006)
26. Shotton, J., Blake, A., Cipolla, R.: Efficiently combining contour and texture cues for object recognition. In: BMVC (2008)
27. Yang, X., Latecki, L.J.: Weakly Supervised Shape Based Object Detection with Particle Filter. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 757–770. Springer, Heidelberg (2010)
28. Lu, C., Latecki, L., Adluru, N., Yang, X., Ling, H.: Shape guided contour grouping with particle filters. In: ICCV (2009)
29. Schlecht, J., Ommer, B.: Contour-based object detection. In: BMVC (2010)