Using Non-symmetry and Anti-packing Representation Model for Object Detection

Fei Xiao[1], Jin-Wen Tian[1], Guang-Wei Wang[2], Chang-Qing Chen[3,a]

[1]*School of Automation, Huazhong University of Science and Technology, Wuhan, 430074, China;*
[2]*School of Computer and Information Science, Hubei Engineering University, Xiaogan, 432000, China*
[3]*School of Software, Huazhong University of Science and Technology, Wuhan, 430074, China;*
[a]*ccqclczy@163.com*

Abstract: In this paper, we present a non-symmetry and anti-packing object pattern representation model (NAM) for object detection. A set of distinctive sub-patterns (object parts) is constructed from a set of sample images of the object class; object pattern are then represented using sub-patterns, together with spatial relations observed among the sub-patterns. Many feature descriptors can be used to describe these sub-patterns. he NAM model codes the global geometry of object category, and the local feature descriptor of sub-patterns deal with the local variation of object. By using Edge Direction Histogram (EDH) features to describe local sub-pattern contour shape within an image, we found that richer shape information is helpful in improving recognition performance. Based on this representation, several learning classifiers are used to detect instances of the object class in a new image. The experimental results on a variety of categories demonstrate that our approach provides successful detection of the object within the image.

1 Introduction

In this paper we consider the problem of detecting and localizing object of a generic category, such as horse or car in static images. This is a difficult problem because objects in a category can vary greatly in shape and appearance. Variation arise not only from changes in illumination, occlusion, background clutter and view point, but also due to non-rigid deformations, and intra-class variation in shape and other visual properties among objects in a rich category.

How do we deal with the variation, especial the intra-class and pose variability of object? Most of the current researches have focused on modeling object variability, including several kinds of deformable template models [1, 2], and a variety of part-based, fragment-based models [3,4,5,6,7,8,9]. There are several possibilities to represent object classes. A star shape model [10, 11] can be easily trained and evaluated in contrast to the constellation model [12] or complex graphical models [13]. It allows using as many parts as required, since the complexity scales linearly. Moreover, this model is flexible enough to deal with large variations in object shape and appearance of rigid and articulated structures.

The method of Leibe*etal*.[14] give a highly flexible learned representation for object shape that can combine the information observed on different training examples. Opelt, *et al.*[9] explore a similar geometric representation to that of Leibe*et al*. [14] but use only the boundaries of the object, both internal and external (silhouette).

The pictorial structure models [5, 15] represent an object by a collection of parts arranged in a deformable configuration, where the deformable configuration is represented by spring-like connections between pairs of parts. Crandall *et al.* propose k-fans model [16] to study the extent to which additional spatial constraints among parts are actually helpful in detection and localization. The patchwork of parts model from [6] is similar, but it explicitly considers how the appearance model of overlapping parts interacts to define a dense appearance model for images. It is proved that adding spatial constraints gives better performance.

Another way is to develop a feature set [17, 18, 19, 20] that is robust to variable local shape and the wide range of poses. Lowe [17] proposed a scale invariant feature transform (SIFT), which combines a scale invariant region detector and a descriptor based on the gradient distribution in the detected regions. Geometric histogram [21] and shape context[18] compute a histogram describing the edge distribution in a region. These descriptors were successfully used, for example, for shape recognition of drawings for which edges are reliable features. HOG [22] is similar to that of edge orientation histograms[19, 20], scale-invariant feature transform descriptors [17], and shape contexts [18], but differs in that it on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy.

Our approach has two methods to deal with the variation of object, both global and local. Firstly, we propose a non-symmetry and anti-packing object pattern

representation model (NAM) to represent an object category. The NAM object model consist of several local parts, we call it sub-patterns. The model codes the global geometry of generic visual object categories with spatial relations linking object pattern to sub-patterns.

Secondly, the proposed descriptors of sub-pattern can deal with the local variation of object. Shape based information have been selected as a key component of local features. After reviewing exist edge based descriptors, we show experimentally that grids of Histograms of Edge Direction (G-HED) descriptors significantly outperform existing feature sets for shape class detection. Contour shape have been used in object recognition to a certain extent: Shotton*et al*.[23] and Opelt*et al.* [9] use boundary fragments to represent an object and use boundary matching method to detect object.

The proposed framework can be applied to any object that consists of distinguishable parts arranged in a relatively fixed spatial configuration. Our experiments are performed on images of side views of horses; therefore, this object class will be used as a running example throughout the paper to illustrate the ideas and techniques involved.

The rest of the paper is organized as follow. Section 2 describes the non-symmetry and anti-packing model. Section 3 introduces the sub-pattern descriptor. Section 4 and 5 presentthe framework of our approach. In section 6, experiments on real images show that the model is effective for object detection.

2 Description of Non-symmetry and Anti-packing Pattern Representation Model

The non-symmetry and anti-packing object pattern representation model (NAM) is an anti-packing problem. The idea of the NAM can be described as follows: Given a packed pattern and n predefined sub-patterns $\{p_1, p_2, \cdots, p_n\}$, pick up these sub-patterns from the packed pattern and then represent the packed pattern with the combination of these sub-patterns.

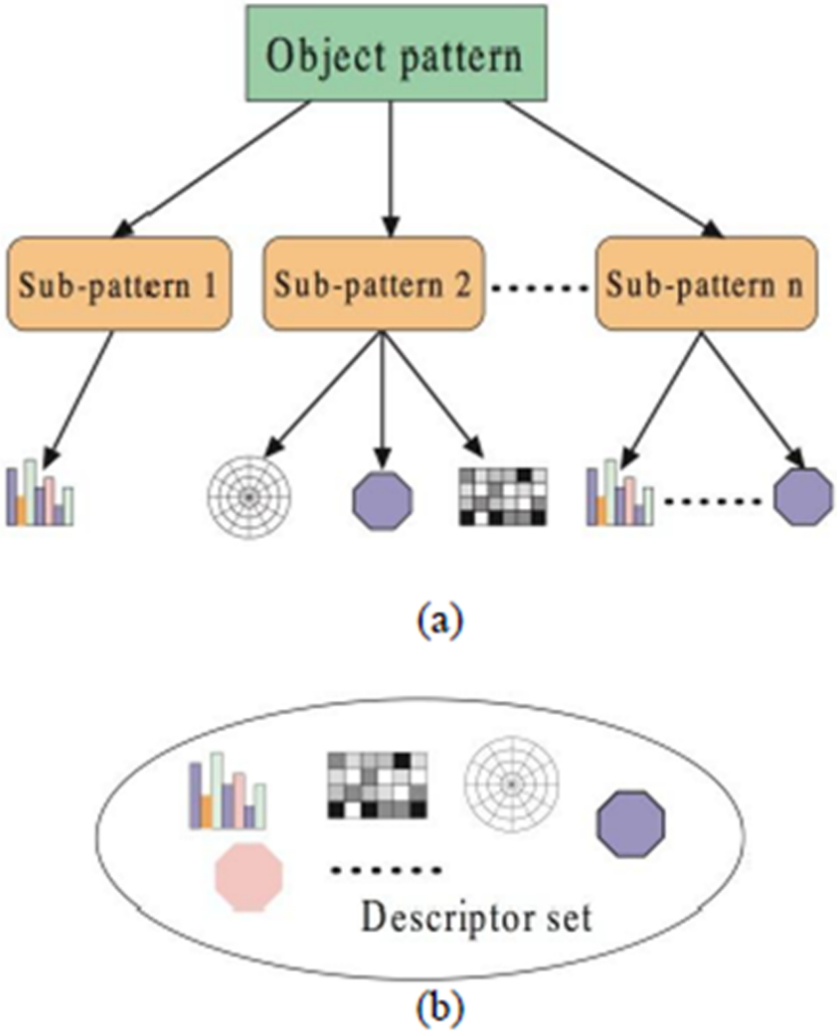

Figure 1..(a)Our hierarchical object model, (b) Descriptor set.

The object pattern representation method is a hierarchical model that codes the global geometry and local appearance of generic visual object categories with spatial relations linking object pattern (top level) to sub-patterns (second level), and local feature cues linking sub-pattern (second level) and local feature class (third

level). See Fig. (1) (a), the object pattern is at the top level, the second and third level is the sub-patterns and local feature descriptors of sub-pattern respectively. Fig. (1) (b) presents the sub-pattern descriptor.

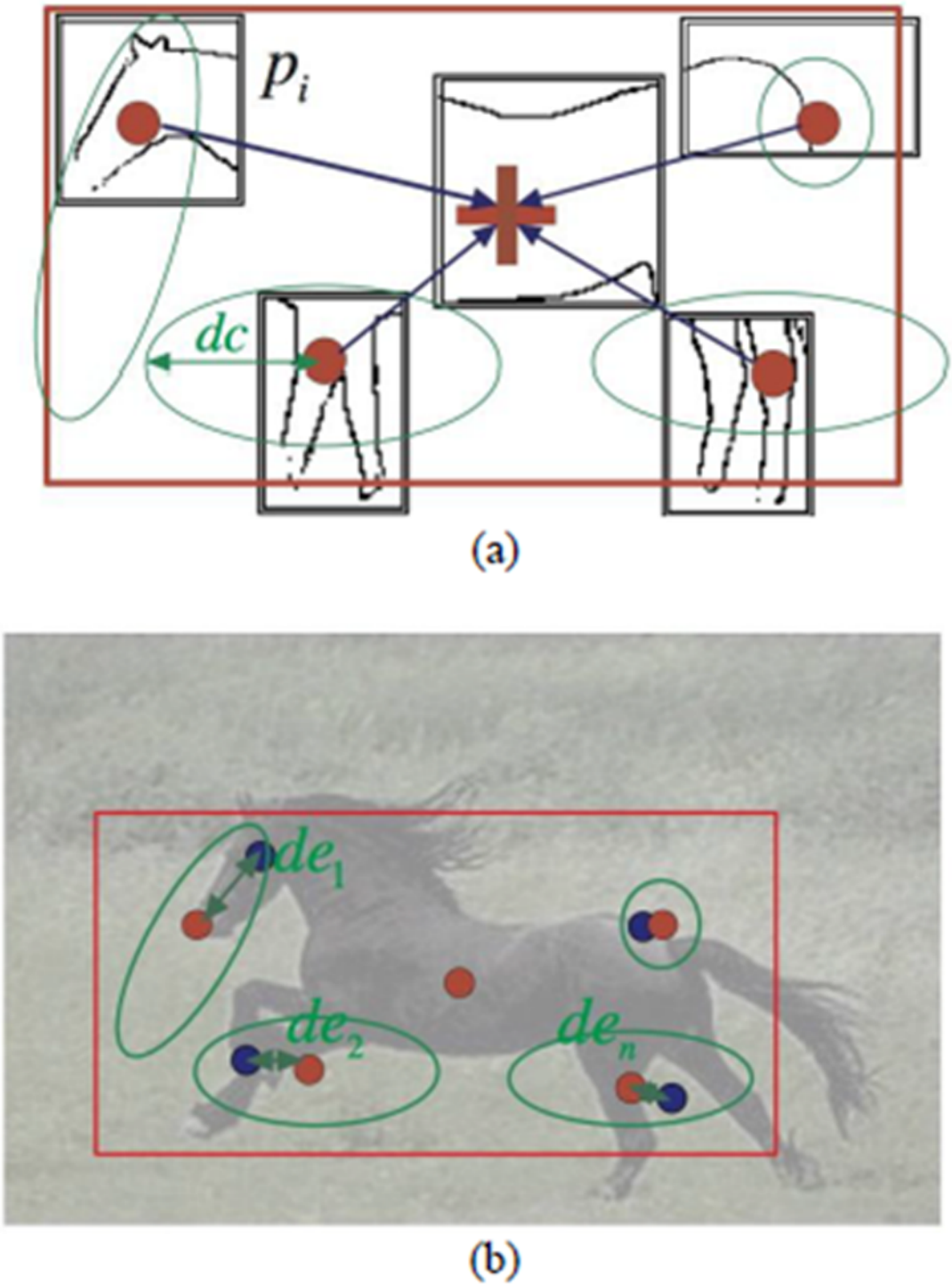

Figure 2. (a)Global spatial structure. Sub-patterns p_1 (black box) are arranged within the bounding box (red). Small red circles show the optimal position and blue ellipses the spatial uncertaintydc. (b)To code the pose variation of a side view horse using global spatial structure. Green arrows show the deviation distancede.

Global Spatial relation: The spatial relations between top level and the second level can be described by global spatial structure. Fig. (2)presents an example of the global spatial structure for horse class.

Local feature encoding: Sub-patterns can be described by a rich set of cues (such as shape, color and texture) inside them. Between second level and third level, we capture different sub-pattern cues from the sub-pattern window, each type of cue is encoded by using an appropriate descriptor, and these encoded information concatenated into a feature vector. In this paper, we use shape information to represent the sub-patterns and edge direction histogram (EDH) to describe the shape information. The detail of encoding sub-pattern feature have presented in Section 3.

Here, we use an object pattern Γto describe an object

category. The object pattern that consists of \$n\$ sub-patterns p_i can be defined by the following expression:

$$\Gamma = \bigcup_{i=1}^{n} p_i(x, r, de, w, \phi(x,r) | \{f_1, f_2, \cdots, f_{m_i}\})$$

Where x is a two-dimensional vector specifying an "anchor" position for sub-pattern p_i relative to the object pattern position; $r$represent the scale of the sub-pattern; de is a deviation vector; w shows the discriminative weight of the sub-pattern. $\phi(x,r))$ denote a feature vector for the i^{th}sub-pattern and $f_j(1 \le j \le m_i)$ is one of the feature descriptors.

The non-symmetry relationship between sub-patterns describes the global structure information of object category and is designed to decouple variations due to affine warps, pose variability and other forms of shape deformations. Anti-packing is a procedure that finding sup-patterns in query images, combining them into an object pattern and classifying.

3 Sub-pattern Description

Sub-pattern can be described by a rich set of cues inside them, such as shape, color and texture. Based on the observation that for a wide variety of common object categories, shape [9, 24, 25] matters more than local appearance [26, 27]. In this paper we use shape information as a key component for object detection.

Since edge points are related to shape information closely, the local shape can often be characterized rather well by the distribution of edge directions. The proposed descriptors are inspired by edge orientation histogram [28, 29, 30, 31, 32], Histograms of Oriented Gradients [22, 33] and shape contexts [18]. They use only edge pixel counts with the direction histogram and they are computed on dense overlapping grids that make the performance so affect.

The implementation of our descriptor combines many of these ideas from early work and more recent contributions.

We first obtain the sub-pattern edge map by using a Canny edge detector and remove spurious edges using a grouping technique inspired by Mahamudet al. [34].

Next, a sub-pattern window is subdivided into aκ× κgrid ("cell"), for each cell accumulating a local 1-D histogram of edge direction over the pixels of the edge within the cell. This step is very similar to region description in [35].

Finally, we accumulate a measure of local histogram over somewhat larger spatial regions ("blocks") and use the results to normalize all of the cells in the block. We refer to the normalized descriptor blocks as grid of Histogram of Edge Direction (G-HED) descriptors.

We now give details of our G-HED implementations and the effects of the various choices on detector performance.

In our experiments reported, we usually use square G-HED's, i.e., $\kappa \times \kappa$ grids block of $\eta \times \eta$ pixel cells each containing direction binsγ, where$\kappa\eta$,γ are parameters. The sub-pattern window is 64×64 pixel and it is divided into 4×4 grid, e.g.,$\kappa = 4$. The various choice of block size ζ is also very important for detector performance. In our experiments, 2×2 grid blocks perform best for sub-pattern detection. The stride (block overlap) is fixed at half of the block size (e.g. a cell).

The edge pixel direction is generated by using two 3×3 masks of Sobel operators, e.g., vertical edge mask and horizontal edge mask. For each edge pixel$p_{(i,j)}$, its edge vector can be represented with the vector $D = \{dx_{i,j}, dy_{i,j}\}$, where $dx_{i,j}$ and $dy_{i,j}$ are obtained by using the vertical edge mask and horizontal edge mask on the location $p_{(i,j)}$ of sub-pattern window respectively. Each edge pixel direction is calculated with the equation $\theta_{i,j} = \tan^{-1}\left(\frac{dy_{i,j}}{dx_{i,j}}\right)$. The direction bins are evenly spaced over ("unsigned") or ("signed"). Using 9 bins (e.g. 20° per bin) can obtain better performance.

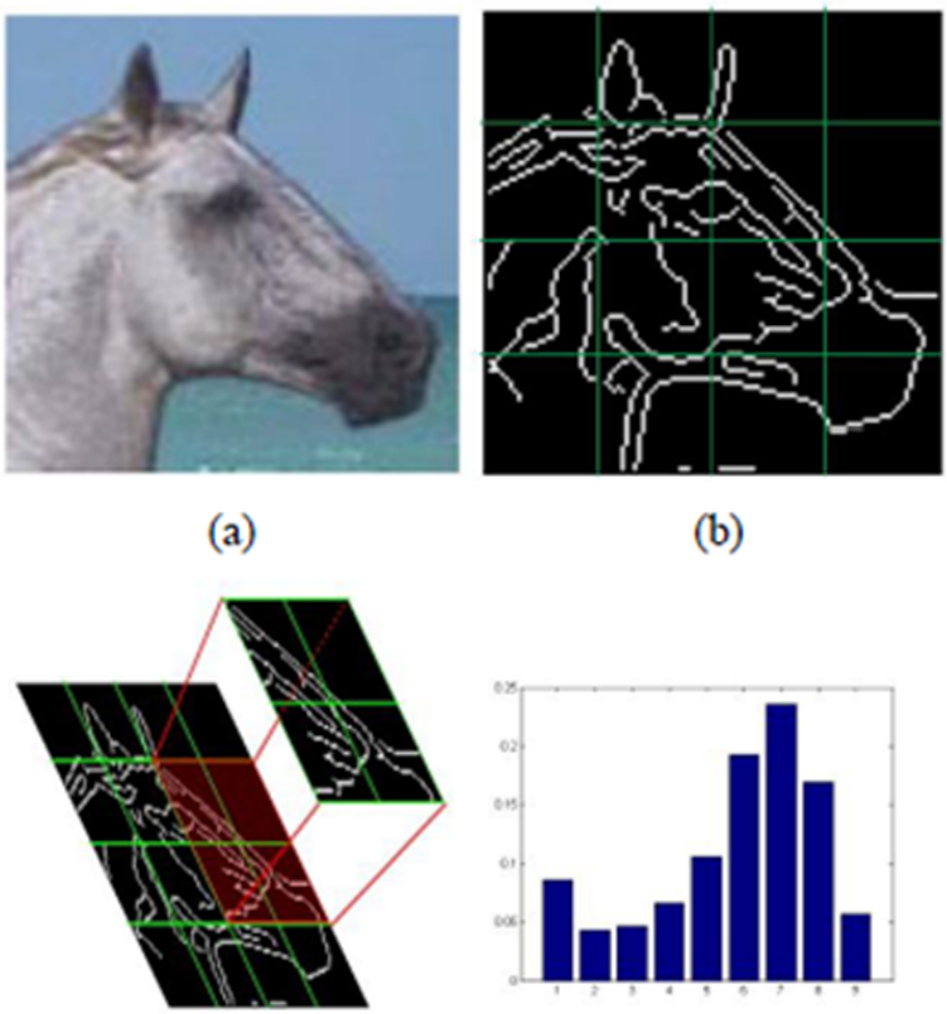

Figure 3. The contour shape descriptor. (a)Original image, (b)Edge map, (c) 4×4 grid on edge map, (d)Edge direction histogram of one block containing 4 grids.

A global direction histogram of a sub-pattern would average too much spatial information to infer pose. Here, we adapt the method of [35] for describing sub-patterns. A sub-pattern is subdivided evenly its bounding box into $n \times n$ grid, and accumulating a local 1-D histogram of edge direction over the edge pixels within the 2×2 grid, as illustrated in Fig.(3). In the experiments reported, we use $n = 4$. The combination of these histograms then represents the descriptor.

4 Detection

The pipeline of our detection framework as follow: first, training a classifier for each sub-pattern. Next, using one classifierto detect hypothesis of object location, i.e., initial detection. After that, a verification scheme is applied to the hypothesis to obtain final detection.

4.1 Learning Classifiers

The task of learning is to establish n classifier $\{Cf_1(\cdot), Cf_2(\cdot), \cdots, Cf_n(\cdot)\}$ for an object pattern with n sub-patterns; each classifier is corresponding to a sub-pattern. Take a classifier for example, given a set of training image windows labeled as positive (object) or negative (nonobjective), each image window is converted into a feature vector as described above. These vectors are then fed as input to a supervised learning algorithm that learns to classify an image window as member or nonmember of the object pattern. In our experiments, two-class LIBSVM had been used for classification (For details of LIBSVM see [36]).

4.2 Detection Hypothesis Using One of the Learned Classifiers

Sliding window classification [25] is a simple, effective technique for object detection. The initial detection problem is to determine whether the query image contains instances of sub-pattern and where it is. Having trained a SVM window classifier, we can detect and localize novel object instances in a test image using a simple sliding-window mechanism [22, 37]. Here, we select the j^{th} sub-pattern $p_j$as an initial detected sup-pattern. The classifier $Cf(\cdot)$ corresponding to sub-pattern p_j is applied to fix-sized windows at various locations in the feature pyramid, each window being represented as a feature vector $\phi(x, r)$, where x specifies the position

of the window in the image, and r specifies the level of the image in pyramid. The following expression represents the classifier $Cf_j(\cdot)$ at one of the sliding windows.

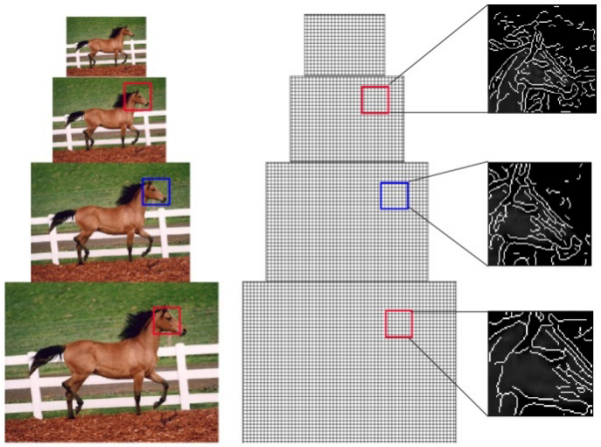

Figure 4. Image pyramid (Left), feature pyramid (Middle) and exampleof feature windows (Right).

A threshold α is introduced to determine whether the window is positive or contain an instance. If $s_{p_j} > \alpha$, then,the window is positive, $h_j = (x, r)$ is a hypothesis and we put the h_j into the sub-pattern hypothesis set $H = \{h_{j,1}, h_{j,1}, \cdots, h_{j,k}\}$. Lowering the threshold increases the correct detections but also increases the false positives; raising the threshold has the opposite effect. In our experiment, we use$\alpha = 0.5$.

The feature pyramid illustrated in Fig.(4), which is similar to [33], specifies a feature map for a finite number of scales in a fixed range. In practice we compute feature pyramid by computing a standard image pyramid via repeated soothing and subsampling, and then computing a feature map from each level of the image pyramid. A test image is scaled to sizes ranging from 0.48 to 1.2 times the original size, each scale differing from the next by a factor of 1.2.

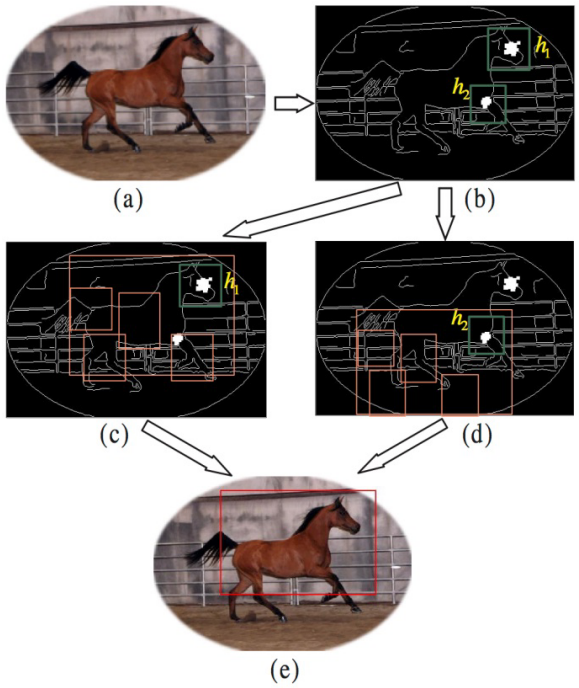

Figure 5. Initial detection and verification. (a)Query image, (b) Hypothesis set of initial detection $\{h_1, h_2\}$, (c)Process of verification for h_1,(d)Process of verification forh_2 and (e)Final result.

5 Verification

These hypothesis are then refined through a verification scheme to obtain final detection result. Fig. (5) introduces the process. The first step is to generate a hypothesis h_Γ of the object pattern Γ by applying a transformation $T(\cdot)$ to $h_j$andΓ. Fig.(5) (b) illustrate the transformation procedure.$h_j$is one of hypothesis in set H. The transformation $T(\cdot)$ exploits the rough localization provided by the spatial relationship between the sub-patterns and the object pattern. Then the transformation for the hypothesis h_j and object patternΓ is characterized by:

$$Sp_i = Cf_i(\emptyset(L_i))$$

Where $L_i = (x_i, r_i, de_i)(1 \le i \le n)$ is the expected location of sub-patterns beside h_j. This transformation provides not only position x_i, scale estimation r_i but also deviation de_i of sub-patterns in the object pattern.

Next, classifier$Cf_i(\cdot)$ is applied to the corresponding window at location L_i.

Where s_{p_j} is to determine that whether location L_i contains sub-pattern p_j. The overall verification score$S_{ver}(h_\Gamma)$for object patternΓis a combination of the sub-patterns detection result s_{p_i}:

$$S_{ver}(h_\Gamma) = \sum_{i=1}^{n} (w_i \cdot s_{p_i} - de_i)$$

wherew_i is the discriminative weight of sub-pattern p_j, de_i is the deviation of sub-pattern from the optimal position. The verification of object pattern was illustrated in Fig.(5)(c) and (d). When the value of the S_{ver} is above a threshold β, the hypotheses position $h_\Gamma$contains an instance of the object pattern.

6 Experiments

We present extensive experimental evaluation, involving several existing data sets covering eight diverse shape-based object classes for a total of more than 900 test images.

6.1 Evaluation Criteria

In this section, we investigate the performance of our system under the PASCAL criterion. For a detection to be marked as correct, its inferred bounding box b_{inf} must agree with the ground truth bounding boxb_{gt} based on an overlap criterion as $\frac{b_{inf} \cap b_{gt}}{b_{inf} \cup b_{gt}} \geq 0.5$. Each b_{gt} can match to only one b_{inf} , and so spurious detections of the same object count as false positives

Table 1. Performance of our detection system on INRIA horse data set, containing 170 positive images and 170 negative images.

β	No. of correctdetections TP	No. of falsedetections FP	RecallRTP/170	PrecisionPTP/(TP+FP)
20	152	110	0.8941	0.5802
30	138	107	0.8118	0.5633
40	130	98	0.7647	0.5702
50	126	86	0.7412	0.5942
60	121	74	0.7118	0.6205
80	110	53	0.6471	0.6748
100	103	45	0.6059	0.6959
120	87	39	0.5118	0.6905
150	67	26	0.3941	0.7204
180	41	14	0.2412	0.7455
210	23	7	0.1353	0.7667

Table 2. Performance of our detection system on WEIZMANN horse data set.

β	No. of correctdetections TP	No. of falsedetections FP	RecallRTP/328	PrecisionPTP/(TP+FP)
20	299	36	0.9144	0.8925
30	294	34	0.8991	0.8963
40	289	31	0.8838	0.9031
50	283	28	0.8654	0.9099
60	266	24	0.8135	0.9172
80	229	17	0.7003	0.9309
100	206	11	0.6299	0.9493
120	185	8	0.5657	0.9585
150	149	6	0.4557	0.9612
180	106	4	0.3242	0.9636
210	62	2	0.1896	0.9688

When a detection system is put into practice, we are interested in knowing how many of the objects it detects, and how often the detections it makes are false. This trade-off is captured more accurately by a variation of the recall-precision curve, where

$$Recall = \frac{TP}{nP}$$

$$Precision = \frac{TP}{TP + FP}$$

whereTP is the number of true positives, FP is the number of false positives and nP the total number of positives in data set. The first quantity of interest, namely, the proportion of objects that are detected, is given by the recall. The second quantity of interest, namely, the number of false detections relative to the total number of detections made by the system, is given by

$$1 - Precision = \frac{FP}{TP + FP}$$

Plotting recall versus (1-precision), therefor, expresses the trade-off.

Performance is also evaluated by plotting detection rate (DR) versus the incidence of false positives (false positives per image (FPPI)) while varying the detection threshold, where

$$Detection\ rate = \frac{TP}{nP}$$

$$Fals\ positives\ per\ image = \frac{FP}{nN}$$

Where nN is the total number of images in data set. Comparison between different methods is mainly based on two points on the DR/FPPI plot, at 0.3 and 0.4 FPPI.

6.2 INRIA Horse and Weizmann-Shotton Horse

INRIA horse [38]. This challenging data set consists of 170 images containing one or more horses, seen from the side, and 170 images without horses. Horses appear at several scales and against cluttered backgrounds.

Weizmann-Shotton horse [23]. Shotton*et al.*propose another horse detection data set, composed of 327 positive images containing exactly one horse each and 328 negative images.The INRIA and Weizmann are very challenging data sets of horse images, containing different breeds, colors, and textures, with varied articulations, lighting conditions, and scales.

We present our results in Table 1 and 2. The different detection results are obtained by varying the threshold parameter β as described in Section 4.3.

We compare our method with Dalal*etal.*[22] and Ferrari *etal*. [25] on the INRIA and Weizmann horses data sets. We randomly select 100 training images per category in Caltech101 and Google images to train classifiers. Dalal's method is currently the state of the art in human detection and has proven very competitive on other classes as well. The object detection method by Ferrari *etal*. achieved considerable gains on many object categories. Like ours, their object detectors are based on sliding a window subdivided into tile but uses different feature descriptors.

The results are displayed in Fig.(6). Our detector achieves a substantially higher performance than HOG. Compared to [25] our method achieved same fine performance. However, their criterion (a detected bounding box is considered correct if it overlaps≥ 20%) is rather loose, and it might consider as correct also rather inaccurate detections.

6.3 ETHZ Shape Classes

The ETHZ shape database(collected by V. Ferrari *et al.*[39]) consists of five distinctive shape categories (apple logos, bottles, giraffes, mugs and swans) in a total of 255 images. All categories have significant intra-class variations, scale changes, and illumination changes. Moreover, many objects are surrounded by extensive background clutter and have interior contours.

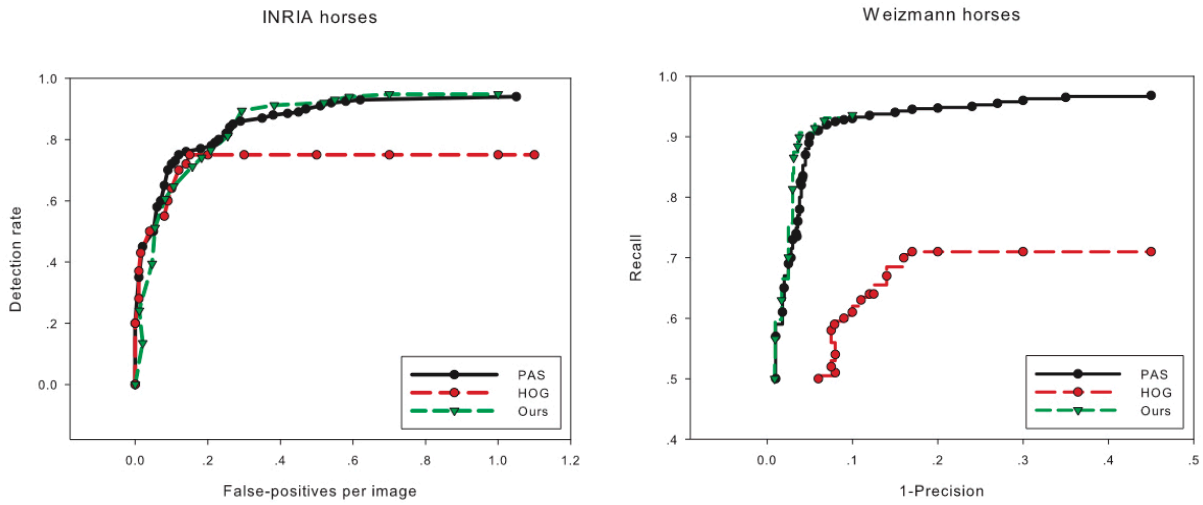

Figure 6..Comparison between our detector, Ferrari *et al*: [25] PAS-based one and Dalal*et al*: [22] HOG-based one.

Table 3. Comparison of detection performance with [Ferrari et al., 2007] on the ETHZ shapedata set at 0.4 FPPI.

	Apple	Bottle	Giraffe	Mug	Swan	average
Our	88.6	83.4	83.9	83.5	87.5	85.4
Fritz	89.9	76.8	90.5	82.7	84.0	84.8
Ferrari	83.2	83.2	58.5	83.6	75.4	76.8

We compare to [39] and [40]} on the ETHZ shape database. Experiments are conducted in 5-fold cross-validation. We split the entire set into half training and half test for each category, and average performance from 5 random splits is reported. This is consistent with the implementation in [39, 40] which reported the state-of-the art detection performance on this data set. Table 3 shows our comparison to [39] on each of the categories. Average over all categories we improve the performance of [39] by 8.6% to 85.4%. On apple logos, giraffes and swans, we improve the performance by 5.4, 25.3 and 12.1% respectively. On bottles and mugs, our approach performs comparable. We account the performance on the bottles and mugs to the shape which is less discriminant with respect to the background. As the data set was designed to test shape-based approaches, the improvements obtained by our approach underlines the versatility and adaptively of the hierarchical representation.

Figure.7.. Detections for INRIA and Weizmann horses data sets when the value of threshold is 50. In the middle row, the rightmost imageshows a missed detection.

7 Conclusion

We have proposed a non-symmetry and anti-packing object pattern representation model (NAM) to represent a object category. This model can effectively codes global structure of object. The object pattern model consists of several part sub-patterns. Each sub-pattern is represented by a rich set of image cues(shape, color and texture). One sub-pattern descriptor can incorporate several feature descriptors. We selected appropriate feature descriptor for the sub-pattern to deal with the local variation. In our work, the edge direction histogram descriptor introduced to describe the shape information of sub-patterns. Based on this representation, several learning classifiers were trained to detect sub-pattern instances. The proposed framework can be applied to any object category that consists of distinguishable parts arranged in a relatively fixed spatial configuration.

8 Conflict of Interest

This article content has no conflict of interest

Acknowledgement

The authors wish to acknowledge the support of the National Natural Science Funds of China under Grant No. 61370092, Natural Science Funds of Hubei Province under Grant No. 2013CFC005, and Hubei Engineering University Teaching Research Project (Grant No. 2015A38).

References

1. T.F. Cootes, G.J. Edwards, and C.J. Taylor.Active appearancemodels.IEEE Trans. Pattern Analysis and Machine Intelligence,pages 681–685, 6 2001.
2. J. Matthews and S. Baker. Active appearance models revisited.International Journal of Computer Vision, 63(6):135–164,2004.
3. R. Fergus, A. Perona, and A. Zisserman.A sparse object categorymodel for efficient learning and exhaustive recognition.In IEEE conf. on Computer Vision and Pattern Recognition,2005.
4. S. Agarwal, A. Awan, and D. Roth.Learning to detect objectsin images via a sparse, part-based representation.IEEE Trans.Pattern Analysis and Machine Intelligence, pages 1475–1490,2004.
5. P.F. Felzenszwalb and D.P. Huttenlocher.Pictorial structuresfor object recognition. International Journal of ComputerVision, pages 55–79, 2005.

6. Y. Amit and A. Trouve. Pop: Patchwork of parts models forobject recognition. International Journal of Computer Vision,pages 267–282, 2007.
7. L. Kyoung-Mi. Component-based face detection and verification.Pattern Recognition Letters, pages 200–214, 2008.
8. H. Schneiderman and T. Kanade. Object detection using thestatistics of parts. International Journal of Computer Vision,pages 151–177, 2004.
9. A. Opelt, A. Pinz, and A. Zisserman.A boundary-fragmentmodelfor object detection. In European conf. on Computer
10. Vision, 2006.
11. J. Shotton, A. Blake, and R. Cipolla.Contour-based learningfor object detection.In International conf. on Computer Vision,2005.
12. B. Leibe, E. Seeman, and B.Schiele.Pedestrian detection incrowded scenes. In IEEE conf. on Computer Vision and PatternRecognition, pages 878–885, 2005.
13. R. Fergus, P. Perona, and A. Zisserman. Object class recognitionby unsupervised scale-invariant learning. In IEEE conf.on Computer Vision and Pattern Recognition, pages 264–271,2003.
14. E. Sudderth, A. Torralba, W. T. Freeman, and A. Wilsky.Learning hierarchical models of scenes, objects, and parts.In International conf. on Computer Vision, pages 1331–1338,2005.
15. B. Leibe, A. Leonardis, and B. Schiele.Robust object detectionwith interleaved categorization and segmentation. InternationalJournal of Computer Vision, pages 259–289, 2008.
16. P. Felzenszwalb, D. McAllester, and D. Ramanan.A discriminativelytrained, multiscale, deformable part model. In IEEEconf. on Computer Vision and Pattern Recognition, pages 151–168, 2008.
17. D. Crandall, P. Felzenszwalb, and D. Huttenlocher. Spatialpriors for part-based recognition using statistical models.InIEEE conf. on Computer Vision and Pattern Recognition, 2005.
18. D.G. Lowe. Distinctive image features from scale-invariantkeypoints. International Journal of Computer Vision, 60(2):91–110, 2004.
19. S. Belongie, J.Malik, and J. Puzicha.Matching shapes. InInternational Conference on Computer Vision, pages 454-461,2001.
20. W. T. Freeman, K. Tanaka, J. Ohta, and K.kyuma. Computervision for computer games. In 2nd International Conferenceon Automatic Face and Gesture Recognition, pages 100–105,1996.
21. W. T. Freeman and M. Roth. Orientation histograms for handgestrue recognition.In Intel. Workshop on Automatic FaceandGesture-Recognition, pages 296–391, 1995.
22. A. Ashbrook, N. Thacker, P. Rockett, and C. Brown.Robustrecognition of scaled shapes using pairwise geometrichistograms. In Proc. Sixth British Machine Vision Conf., pages503–512, 1995.
23. N. Dalal and B. Triggs. Histgrams of oriented gradients forhuman detection.In IEEE conf. on Computer Vision andPattern Recognition, 2005.
24. J. Shotton, A. Blake, and R. Cipolla. Multi-scale categoricalobject recognition using contour fragments. IEEE Trans.Pattern Analysis and Machine Intelligence, 2008.
25. J. Sullivan, O. Danielsson, and S. Carlsson. Exploiting partbasedmodels and edge boundaries for object detection.InDigital Image computing, 2008.
26. V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid. Groups of adjacent contour segments for object detection. IEEETransaction on MAMI, 2008.
27. G. Bouchard and B. Trigs.Hierarchical par-bbased visualobject categorization.In IEEE conf. on Computer Vision andPattern Recognition, 2005.
28. L. Zhu, Y. Chen, A. Yuille, andW. Freeman. Latent hierarchicalstructural learning for object detection.In IEEE conf. onComputer Vision and Pattern Recognition, 2010.
29. F. Mahmoudi and J. Shanbehzadeh. Image retrieval based onshape similarity by edge orientation autocorrelogram. PatternRecognition, pages 1725–1736, 2003.
30. D. Tao, X. Tang, X. Li, and X. Wu. Asymmetric bagging andrandom subspace for support vector machines-based relevancefeedback in image retrieval. IEEE Trans. Pattern Analysis andMachine Intelligence, pages 1088–1099, 2006.
31. T. Dacheng, T. Xiaoou, and L. Xuelong. Which components areimportant for interactive image searching? IEEE Transactionson Circuits and Systems for Video Technology, 2008.
32. G. Xinbo, X. Bing, T.Dacheng, and L. Xuelong. Imagecategorization: Graph edit distance + edge direction histogram.Pattern Recognition, pages 3179–3191, 2008.
33. K. Young-Woo and O. Il-Seok.Watermarking text documentimages using edge direction histograms. Pattern RecognitionLetters, pages 1243–1251, 2004.
34. Pedro F. Felzenszwalb, Ross B. Girshick, D. McAllester, andD.Ramanan. Object detection with discriminatively trained partbased models. IEEE Trans. Pattern Analysis and MachineIntelligence, 2010.
35. S. Mahamud, L. R. Williams, K. K. Thornber, and K. Xu.Segmentation of multiple salient closed contours from realimages.IEEE Transaction on MAMI, 2003.
36. ChunhuiGu, J. J. Lim, P. Arbelaez, and J. alik. Recognitionusing regions.In IEEE conf. on Computer Vision and PatternRecognition, pages 1030–1037, 2009.
37. Chih-Chung Chang and Chih-Jen Lin. Libsvm: a libraryfor support vector machines. Software available athttp://www.csie.ntu.edu.tw/_cjlin/libsvm, 2001.
38. P. Viola and M.jones.Rapid object detection using a boostedcascade of simple features. In IEEE conf. on Computer Visionand Pattern Recognition, pages 1–8, 2001.

39. A. Opelt, A. Pinz, M. Fussenegger, and P. Auer.Generic objectrecognition with boosting.IEEE Trans. Pattern Analysis andMachine Intelligence, pages 516–431, 2006.
40. V. Ferrari, L. Fevrier, and C. Schmid. Accurate object detectionwith deformable shape models learnt from images. In IEEEconf. on Computer Vision and Pattern Recognition, pages 564–571, 2007.
41. M. Fritz and B. Schiele. Decomposition, dscovery and detectionof visual categories using topic models. In IEEE conf. onComputer Vision and Pattern Recognition, pages Part I:511–518, 2008