Boosted Local Structured HOG-LBP for Object Localization

Junge Zhang, Kaiqi Huang, Yinan Yu and Tieniu Tan
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
{jgzhang,kqhuang,ynyu,tnt}@nlpr.ia.ac.cn

Abstract

Object localization is a challenging problem due to variations in object's structure and illumination. Although existing part based models have achieved impressive progress in the past several years, their improvement is still limited by low-level feature representation. Therefore, this paper mainly studies the description of object structure from both feature level and topology level. Following the bottom-up paradigm, we propose a boosted Local Structured HOG-LBP based object detector. Firstly, at feature level, we propose Local Structured Descriptor to capture the object's local structure, and develop the descriptors from shape and texture information, respectively. Secondly, at topology level, we present a boosted feature selection and fusion scheme for part based object detector. All experiments are conducted on the challenging PASCAL VOC2007 datasets. Experimental results show that our method achieves the state-of-the-art performance.

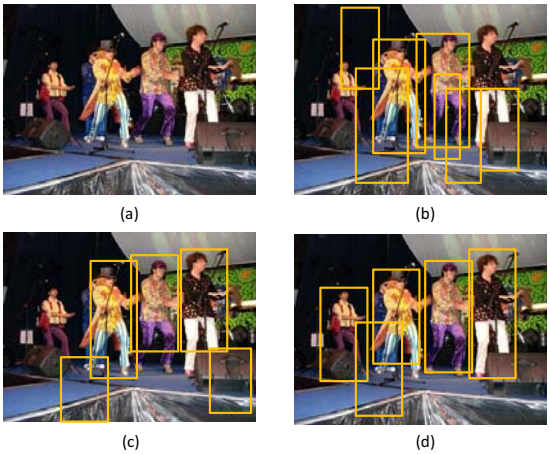

Figure 1. Detection results of different methods. (a) is the original image. (b) is the result of SVM+HOG. (c) is the result from [7] and (d) is the result by the proposed method.

1. Introduction

Object localization is an essential task in computer vision. Impressive performance improvement in object localization has been achieved via the progress in: 1) learning object structure [5, 8, 19, 20, 27] and detector model, and 2) learning low-level feature based appearance model [3, 10, 18, 21, 24, 25].

Detector models mainly include part based models [8, 9, 20, 27] and rigid template models [1, 24, 25]. In part based models, they try to describe the object's structure using several parts and their relationships. Part based models can be considered as top-down structure to tackle the problem of partial occlusion and appearance variations. Part based models [8, 20, 27] have been shown success on many difficult datasets [14]. For these good properties of robustness to deformation, part based model is regarded as a promising method for localizing objects in images. This motivates us to focus on part based model. Rigid template models can not describe the object's structure variations with fixed template. Therefore, they perform well on ideally conditioned database but suffer from those difficult data with deformations. The progress in low-level feature advances the progress of object localization greatly as well. One representative feature is Histogram of Oriented Gradients (HOG) [1]. The others include Pairs of Adjacent Segments (PAS) [10] and Local Binary Pattern (LBP) [17], *etc*.

One important problem in object localization is how to describe object's structure robustly. Part based model as a top-down structure shows its good property of modeling object structure in topology level [20]. But robust low-level feature representation challenges the part based model to obtain better performance. In the field of signal processing, signal is considered structured when the local intensity varies along some preferred orientations [4]. Local structure can be corners, edges or crossings, *etc*. The research in signal processing indicates that there is relation between local energy and local structure. These studies state that using the local energy can represent the local structure [4] well. From

this aspect, previous popular feature HOG and LBP are histogram features. Thus, they can not effectively describe an object's local structure information which is important for object localization.

Motivated by these challenges of robust low-level feature representation for part based model, we address the problem via Local Structured Descriptors based part model. Firstly, we propose Local Structured HOG(LSHOG) in which the Local Structured Descriptor is computed from local energy of shape information, Secondly, similar to LSHOG, we present Local Structured LBP(LSLBP) in which the Local Structured Descriptor is based on texture information. In addition, to tackle the non-linear illumination changes, we clip the large feature value caused by non-linear illumination changes with a truncation item. To reduce the effect of small deformation, we apply spatial weighting which is proved to be robust to aliasing and bin interpolation which can accurately describe histograms in LSLBP. Thirdly, we present a boosted Local Structured HOG-LBP based object detector, and the proposed method achieves the state-of-the-art performance on the challenging PASCAL VOC datasets [14]. Figure 1 gives an example of person detection.

The rest of this paper is organized as follows. Section 2 gives a brief overview of related work. Section 3 introduces the framework of our approach. Section 4 shows and analyzes the experimental results and Section 5 draws conclusions.

2. Related work

This paper focuses on two basic problems: how to accurately describe object structure at feature level and how to fuse multiple Local Structured Descriptors for part based model at topology level.

2.1. Features for object localization

Various visual features such as HOG, LBP, *etc*. have been proposed for object localization. HOG was first proposed for human detection [1]. Ever since then HOG has been proved one of the most successful features in general object localization [14]. During the past few years, many variants of HOG have been presented, such as Co-occurrence Histograms of Oriented Gradients(CoHOG) [26] in which the co-occurrence with various positional offsets is adopted to express complex shapes of object. In [8], contrast-sensitive and contrast-insensitive features are used to formulate more informative gradient feature. LBP was first presented by Ojala *et al*. [17], for the purpose of texture classification. Uniform LBP then was developed to reduce the negative effect caused by noises. In [16], Mu *et al*. stated that the traditional LBP did not perform well in human detection, so they proposed two variants of LBP named by Semantic LBP(S-LBP) and Fourier LBP(F-LBP).Wang *et al*. also proposed a cell-structured LBP [25] dividing the scanning window into non-overlapping cells for human detection.These features(HOG,LBP,CoHOG,S-LBP,*etc*.) are all histogram features which have limitation in describing the object's local structure. In addition, PAS [10] showed attractive performance compared with HOG in recent years. PAS uses the line segments to capture the object's global shape and its structure which is different from HOG and LBP's description schemes. But the boundary detection in PAS is very time consuming which limits its wide applications.

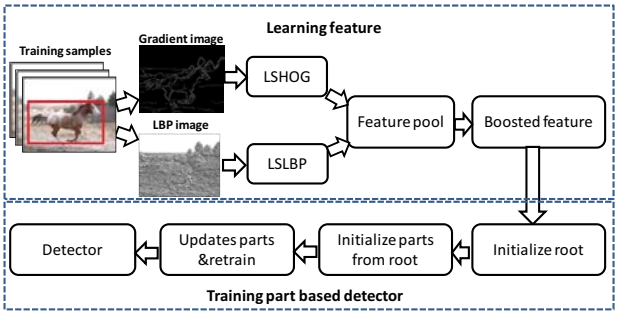

Figure 2. The framework of Local Structured HOG-LBP based part based object detector. This paper mainly focuses on feature construction and multiple features learning for part based model. We perform feature selection in root level. In the training phase, parts models are initialized and updated using the feature learnt from the root. We adopt latent SVM from [8].

2.2. Part based models

Part based models are robust to partial occlusion and small deformation due to their expressive description of object's structure considering the relationships between parts. During the past decade, the most representative part models are the constellation model proposed by Fergus *et al*. [9] and the star-structured part model presented by Felzenszwalb *et al*. [8]. In [9], the parts' locations are determined by the interest points. While in [8], parts' locations are searched through dense feature HOG. Especially, the star-structured part model is discriminatively (For convenience, we refer the method in [8] as DPBM for short) trained and demonstrated state-of-the-art performance in the past several years. In DPBM, an object is represented by a root model and several parts models. The parts' locations are considered as latent information and a latent SVM is proposed to efficiently optimize the model's parameters. DPBM provides a very strong benchmark in the field of object localization. But the performance of DPBM is still limited by the robust low-level feature representation.

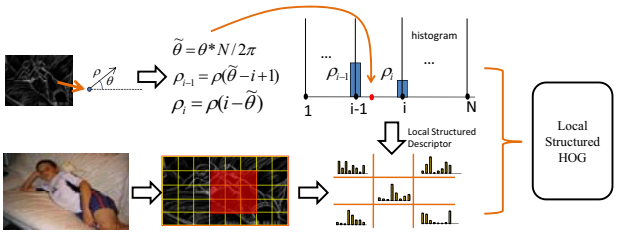

Figure 3. The flowchart of the computation of Local Structured HOG.

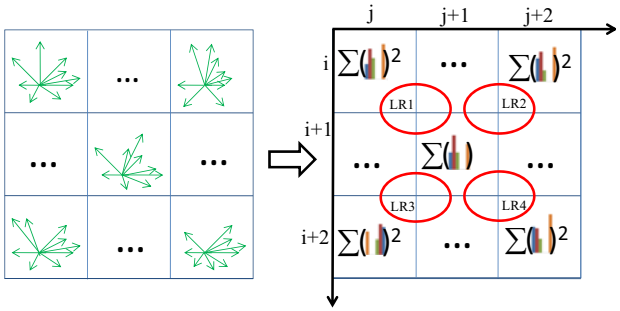

Figure 4. The details of the computation of LSHOG. The left image illustrates the histogram of gradient in each cell. The right image gives the gradient energy via the sum of squares of histogram of gradient in each cell.

3. Boosted Local Structured HOG-LBP for part based model

We show our framework of training Local Structured HOG-LBP based part model in Figure 2. The system consists of two parts: learning feature and training deformable part based detector. The first stage is learning feature, including extraction of Local Structured Descriptors based on shape and texture information, and feature selection of LSLBP in a supervised manner. In the stage of training object detector, we firstly train the root model using the learnt feature from the first stage, then initialize parts models from the root model. We use latent SVM [6, 7, 8] to iteratively train the part based detector.

3.1. Local Structured HOG

In this subsection, the details of Local Structured Descriptor based on shape information will be introduced. As shown in Figure 3, the procedures of LSHOG computation include gradient computation, orientation binning, normalization and formulating Local Structured Descriptor. The LSHOG includes both the histogram feature and Local Structured Descriptor. Thus, LSHOG not only describe the shape information through histogram feature, but also capture the relative local structure information through structured descriptor. The former steps are similar to HOG in [1]. Especially, we don't perform gamma/color normalization and Gaussian weighting because we find they have little affect on performance.

The gradient features used in LSHOG include both unsigned gradient and signed gradient [1, 8]. Their orientation range is $0^\circ - 180^\circ$ and $0^\circ - 360^\circ$, respectively. To obtain a cell-structured feature descriptor, the cell size is set to 8×8.

Local Structured Description. As discussed in above section, the original HOG and its variants are still histogram features and can not describe the local structure effectively.

Empirically, the boundary of any object(*e.g.*, person) tends to be continuous and the spatial adjacent regions must have certain structure relation. As mentioned in above section, PAS [10] is used to capture the spatial structure of objects where the length of adjacent segments and their relative angles are encoded in the final descriptor. However, the Berkeley probability boundary detector used in PAS is very time consuming which limits PAS's large scale applications.

In the field of signal processing, local energy based structure representation is widely used for its robustness to noise and aliasing [4]. Inspired by these progresses, we adopt the local gradient energy to capture local structure. We believe the relative local structure between adjacent blocks is more informative. Therefore, we use the relative gradient energy within object's adjacent blocks to capture the local structure.

The computation of LSHOG is illustrated in Figure 4. Let $F_{i,j(i=1,2,\ldots,h;j=1,2,\ldots,w)}$ be the feature map where h, w are the height and width of the feature map, respectively. Let $H_{i,j}$ specify the sum of histogram of gradients at $F_(i,j)$, and let $LR_{i(i=1,2,3,4)}$ be the squared block consisting of four adjacent cells around cell $(i+1, j+1)$. To avoid a large local structure value, for an example, we define LR_1 by

$$LR_1 = \frac{H_{i+1,j+1}}{\sqrt{E_{i,j} + E_{i,j+1} + E_{i+1,j} + E_{i+1,j+1}}} \quad (1)$$

where $E_{i,j(i=1,2,\ldots,h;j=1,2,\ldots,w)}$ is used to denote the gradient energy obtained from the sum of squares of gradient histogram at each cell (i, j) from F. The computation of LR_2, LR_3 and LR_4 is similar to LR_1. Then we can define the Local Structured Descriptor as follows. The Local Horizontal Structure(LHS) is defined as:

$$\begin{aligned} LHS_1 &= \lambda \mid LR_1 - LR_2 \mid \\ LHS_2 &= \lambda \mid LR_3 - LR_4 \mid \end{aligned} \quad (2)$$

The Local Vertical Structure(LVS) is defined as:

$$\begin{aligned} LVS_1 &= \lambda \mid LR_1 - LR_3 \mid \\ LVS_2 &= \lambda \mid LR_2 - LR_4 \mid \end{aligned} \quad (3)$$

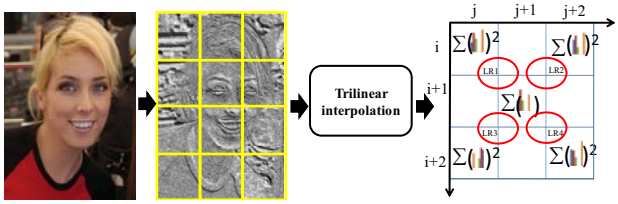

Figure 5. Overview of the computation of LSLBP.

The Local Diagonal Structure(LDS) is defined as:

$$\begin{aligned} LDS_1 &= \lambda \mid LR_1 - LR_4 \mid \\ LDS_2 &= \lambda \mid LR_2 - LR_3 \mid \end{aligned} \tag{4}$$

And the Local Overall Structure(LOS) is defined by

$$LOS = \lambda^{'} \mid LR_1 + LR_2 + LR_3 + LR_4 \mid \tag{5}$$

The control parameter λ can be taken as a normalization factor for LHS,LVS and LDS. We set

$$\lambda = \frac{\sqrt{\sigma \times 18}}{4} \tag{6}$$

where σ is the maximum possible value for gradient feature. The purpose of Eq.6 is to make Local Structured Descriptor's value be the same order of quantity with histogram feature's value. In LSHOG, we use the truncation value $\sigma = 0.2$, so $\lambda = 0.4743$. For LOS, we find the setting $\lambda^{'} = 0.1$ is enough which has the same purpose as λ. As illustrated above, this coding scheme has several advantages:1) Simple to compute. 2) Robust to small deformation. Because the descriptor is related with the local regions' energy, small deformation would change little in the energy of the corresponding region. 3) Easy to be applied in other pixel based histogram features.

3.2. Local Structured LBP

In this subsection, we will give the details of Local Structured Descriptor based on texture information. As shown in Figure 5, firstly, we compute the uniform binary pattern at each pixel, then the initial cell-structured LBP descriptor is formulated by trilinear interpolation. The final LSLBP consists of both binary patterns histogram and Local Structured Descriptor. The local structure coding scheme is similar with LSHOG.

The LSLBP is computed with the cell size 8×8 to be compatible with LSHOG. Many previous work on LBP did not use the trilinear interpolation which is in fact, very helpful for accurate description of histogram based feature [1].

Similar to LSHOG, we capture the local structure through texture information via LHS, LVS, LDS and LOS. In LSLBP, each cell's energy is computed from the sum of squares of binary patterns histogram. That is,

$$E_{i,j} = \sum_{p=1}^{59} h_p^2 \tag{7}$$

where h_p is the histogram of binary patterns, and p denotes the p_{th} feature in h. In this way, the LSLBP can capture the local structure from the aspect of texture, which is mutual complementary with LSHOG.

According to the coding scheme of LBP, it is invariant to linear illumination changes. In the non-linear case, some LBP values tend to become too large while others' not. In order to reduce the possible negative effect caused by these non-linear changes, we clip the entry of uniform pattern with 0.2. Especially, the entry of non-uniform pattern is often much larger than uniform patterns, so we limit its maximum value to 0.3 empirically. The normalization factors λ and $\lambda^{'}$ are set by the same scheme with LSHOG.

3.3. Learning feature and training detector

In this subsection, we address the problem of combining LSHOG and LSLBP and training part based model with learnt LSHOG-LSLBP. This work is different from [25], in which a rigid template model is trained for human detection using concatenated basic HOG-LBP.

To begin with the details of learning feature, we give the formulation of multiple features combination generally.

Fusion problem. Let's denote the training samples as $\{(x_i, y_i)_{i=1,\ldots,N}\}$ where $x_i \in X$ is the training image and $y_i \in \{+1, -1\}$ is the corresponding class label. We can extract different types of features such as LSHOG, LSLBP, *etc*. which are denoted by $f^l_{i(i=1,\ldots,N,l=1,\ldots,M)} \in F$ where f^l_i denotes the l_{th} feature extracted from sample x_i,N is the number of training samples and M is the total number of feature types. Therefore, the feature combination could be formulated as a learning problem:

$$g : \alpha_1 T_1(f^1) + \ldots \alpha_l T_l(f^l) + \ldots \alpha_M T_M(f^M) \longrightarrow (-1, +1) \tag{8}$$

where T_l is the transformation function of the l_{th} feature and α_l is the corresponding weight. g is the optimization function.

Many popular methods have been proposed to tackle the feature combination problem. They are Multiple Kernel Learning [23, 24], Boosting [11] and subspace learning [12], *etc*. These methods can be roughly divided into two categories: basic feature level and feature subspace. In this paper, we mainly investigate some methods at feature level, including naïve combination, MKL and Boosting methods.

For the above three combination schemes, we take a unified way to learn feature and train the part based object de-

tector using the learnt feature. The whole framework includes two stages: 1)Feature learning stage; 2) Part based model training stage.

Feature learning stage. The goal in this paper is to train a LSHOG-LSLBP based part based detector. Hence, the key problem is how to learn feature for part models. In this work, we use the star-structured part based model [8] and the inference of a detection window for the part based model can be summarized as,

$$score_{subwindow} = sr + \sum_{i=1}^{N} sp_i - \sum_{i=1}^{N} dc_i \qquad (9)$$

where sr is the root score(The rigid template model is analogous to the root model here), sp_i means the score from the i_{th} part filter, dc_i is the deformation cost from the i_{th} part filter and N is the number of parts. In the star-structured part based model, the parts models are initialized from the root model. Therefore, we could perform feature selection on root feature only. In the training part based object detector stage, we use the learnt feature to initialize both root model and parts models. This approach has an important advantage that is the learning procedure does not need to know the parts models' sizes.

Because the part based model is based on dense cell-structured feature(LSHOG,LSLBP,*etc*.), learning feature from root still has two strategies: one is learning from features at each cell; The other is from features within the whole detection window. Because our objective is to optimize and classify features from the whole detection window but not from each cell. Therefore, we adopt the latter strategy, *e.g*. learning feature from the detection window. In addition, learning feature procedure is performed for each component to train a part model with multiple components [8] according to aspect ratio.

Part based model training stage. Firstly, we use the learnt feature to initialize the root model. Parts models are then initialized from the root model. Latent SVM [6, 7, 8] is used to train the part models iteratively. The whole algorithm can be found in Algorithm 1.

```
Learnt feature LF_i := ∅;
for component i := 1 to N do
    PF : Extract positive features from i_th root;
    NF : Random sampling from negative samples;
    Learning feature(MKL,Boosting,etc.) from PF, NF;
    Add learnt feature to LF_i;
end
Training part based object detector
for component i := 1 to N do
    Initialize i_th root from LF_i;
    for part j := 1 to N_part do
        Initialize:j_th part from i_th root;
    end
    for Iter k := 1 to K_iter do
        Update models and retrain;
    end
end
```

Algorithm 1: Learning feature and training object detector.

4. Experiments

We evaluate the proposed method on the challenging PASCAL VOC datasets [14] which are widely acknowledged as difficult benchmark datasets for object localization. In PASCAL VOC datasets, there are 20 object classes consisting of person, vehicles(*e.g*.,car, bus), household(*e.g*.,chair, sofa) and animals(*e.g*.,cat,horse) [14]. The criterion adopted in VOC challenge is Average Precision(AP). Our method achieves the state-of-the-art results on PASCAL VOC datasets over other related methods.

Experiments are conducted in three groups:1) Single LSHOG 's experiments designed to validate the effectiveness of Local Structured Descriptor;2) Single LSLBP's experiments developed to validate the effectiveness of trilinear interpolation, truncation and Local Structured Descriptor; 3) Comparison experiments with different combination schemes; 4) The full results of proposed boosted Local Structured HOG-LBP based object detector on PASCAL VOC2007.

Several versions of latent SVM were released at Felzenszwalb's homepage. To avoid confusion, we mention voc-release3.1 [6] as V3 and voc-release4 [7] as V4 shortly. The latent SVM from V4 is only adopted in the full experiments on PASCAL VOC datasets and latent SVM from V3 is used in other experiments. The purpose of using like this is to verify the stability of the proposed method.

4.1. Localization results with LSHOG

To validate the proposed LSHOG, we train a person detector using LSHOG on PASCAL VOC2007 datasets using latent SVM from V3. We achieve 37.4% AP score on person with 1.2% improvement compared with 36.2% from V3. We also do the comparison experiments on aeroplane and dog categories randomly chosen from 20 classes. The results are presented in Figure 6, from which we can see that the improvement is promising.

These results validate that the local structured descriptor can effectively capture more structured information and improve the detection performance. It should be highlighted that the simple coding scheme could be easily extended to other pixel based histogram features.

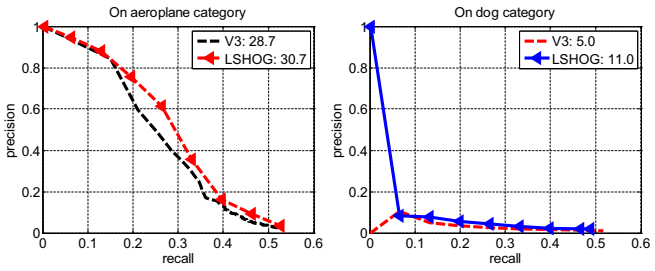

Figure 6. Precision-Recall curve for the categories of aeroplane and dog.

4.2. Localization results with LSLBP

Firstly, we perform the person detection on PASCAL VOC2007 based on traditional LBP without trilinear interpolation, traditional LBP with trilinear interpolation and LSLBP. LSLBP achieves 32.4% best AP score, with an improvement by 1.4% and 2.2% over LBP with trilinear interpolation and traditional LBP, respectively. This result indicates that Local Structured Descriptor and trilinear interpolation are effective.

We also conduct several experiments on person category on VOC2007 to discuss different truncation's strategies. The AP score is 29.2 for LBP without any truncation, 29.4 for LBP with truncation value 0.2 for both uniform pattern and non-uniform pattern and 30.2 for LBP with truncation value 0.2 for uniform pattern and 0.3 for non-uniform pattern. Thus, treating uniform pattern and non-uniform pattern differently and truncating them with 0.2 and 0.3 performs better than others. In addition, the truncation value is set empirically.

The result that the LBP with truncation value 0.2 for uniform pattern and 0.3 for non-uniform pattern performs best indicates that:1) Truncation is helpful for robustness to nonlinear illumination changes; 2)Different truncation for uniform patterns and non-uniform patterns is reasonable.

4.3. Comparisons with different fusion schemes

We compare naïve combination, MKL and Boosting method to find which performs best in combining LSHOG and LSLBP for part based model and also give the analysis.

Naïve combination. Naïve combination directly concatenates different features into a single feature vector, which is the simplest method for feature combination.

Multiple Kernel Learning. MKL has great advantage to handle multiple, heterogeneous data sources and has been widely applied in the problem of feature selection and combination [23, 24]. In this paper, we adopt Generalized MKL(GMKL)[23] for its good generalization property over

category	V3	Naïve	Improvement
person	36.2	37.2	1.0
chair	16.5	15.2	-1.3

Table 1. Detection results of naïve combination on person and chair category.

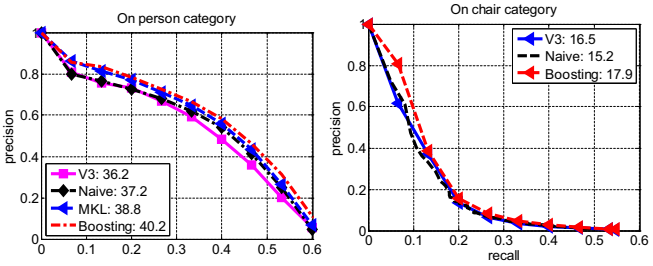

Figure 7. The detection results achieved by the MKL and Boosting based detection results.

kernels combination. And we use linear kernel as base kernel and sum kernel formulation for efficiency.

Boosting. Boosting is one of the most popular methods for feature combination. In our approach, we choose GentleBoost [11, 22] mainly because GentleBoost uses Newton stepping rather than exact optimization at each step, thus it outperforms other boosting methods especially when the training data is noisy and has outliers [13].

Results. In Table 1, we have shown the results of person and chair using naïve combination. To compare with V3 fairly, the HOG from [8] and traditional LBP are adopted in this experiment. As shown in Table 1, the performance is improved by 1% for person, while decreased by 1.3% for chair class. So we find that naïve combination is not always effective for all classes.

Through the experiments discussed in LSHOG and LSLBP subsections, the single LBP based part object detector's performance is worse than that of single HOG. Therefore, it's reasonable to infer that some subsets in LBP features are effective for localization while others not. Inspired by these observations, we could select certain effective subsets in LBP. Thus, we use MKL and boosting to select features. Still the experiments are conducted on person category on PASCAL VOC2007 datasets and a bi-component model using the learnt feature from MKL and GentleBoost is trained, respectively.

In the experiment based on MKL, the penalty C is set to 10 and the maximum iteration is set to 40. The maximum iteration is set to 200 for GentbleBoost.To fairly compare with naïve combination, traditional LBP is adopted in the experiments in which a bi-component model is trained. MKL selects 1850 dimensions out of 3713 in component 1 and 1675 out of 2596 in component 2. While GentleBoost

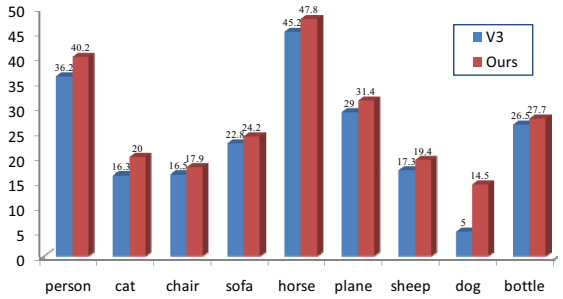

Figure 8. Evaluation of the proposed approach over 9 categories. Note:For a fair comparison, the feature used here is HOG from [8] and traditional LBP.

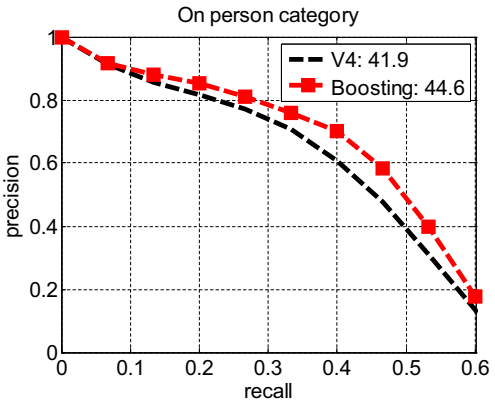

Figure 9. Detection results with the Boosted LSHOG-LSLBP based part object detector using latent SVM from [7].

selects 119 dimensions in component 1 and 131 in component 2.

As shown in Figure 7, for the person category, the Boosting based method achieves the best results, by improving 4% over baseline V3, 3% over naïve combination and 2.4% over MKL based method. Specially, it shows that the improvement for chair class is 1.4% while in naïve combination is -1.3%. The result further validates only certain subsets of LBP feature effective for localization. Improvement has been obtained from MKL as well, but inferior to boosting method. MKL selects more features than boosting method, indicating there are still many noisy features not effective for classification. Another disadvantage of MKL is its huge computation cost. Furthermore, we evaluated the approach over other 8 classes randomly chosen from VOC2007 datasets including person, vehicles, households and animals four categories using latent SVM from V3 [6]. The purpose of this experiment is to validate the effectiveness of presented boosted feature fusion scheme. As shown in Figure 8, for the categories with rich texture such as person,horse,dog, the average improvement is about by 4% while for chair, bottle with less texture, the improvement is only by 1% - 2%. Several conclusions can be drawn from these results: 1) Texture complements shape feature for robust feature representation on most categories. 2)Subsets of LBP are effective or better for localization than full LBP. 3) Boosted multiple features fusion scheme for part based model stably improves the localization performance and performs best among these methods. These results also answer the question why we use GentleBoost in this work.

4.4. Full results on PASCAL VOC datasets

Motivated by the above results, and at the same time we intend to validate the stability of the proposed method, we train the boosted Local Structured HOG-LBP based part object detector using the latest latent SVM from V4 [7]. The models in the following experiments are trained with six components.

Firstly, the proposed boosted LSHOG-LSLBP based part detector is compared with V4 on person category. As is shown in Figure 9, an improvement is achieved by 2.7%. The improvement is less than the proposed detector using latent SVM from V3. There possibly exists over-fitting in the training procedure, because the feature dimension in six component model is higher than two component model in previous experiments.

The full results on PASCAL VOC2007 are given in Table 2. These results are all without any context based post-processing.

As shown in Table 2, the proposed method stably outperforms the state-of-the-art part based object detector [7] on all the 20 categories. We outperforms other methods except Oxford-MKL method(Oxford-MKL method adopted four types of multi-level features and achieved very competitive results on VOC2007 datasets) in 16 out of 20 categories. If comparing with Oxford-MKL's method, we obtain the best score in 9 out of 20 and the second best in 8. These methods are all the related representative methods in the past several years. In addition, the mean AP of the proposed method is 34.3% which is the highest among these methods, exceeding Oxford-MKL's method by 2.2%.

5. Conclusions

In this paper, we have presented a boosted Local Structured HOG-LBP based object detector. Firstly, we have proposed two types of local structured features, *i.e.*, Local Structured HOG (LSHOG) and Local Structured LBP (LSLBP). Experimental results have proven the proposed features can describe the object's local structure effectively and improve the detection performance. Then, we have presented a boosted multiple features fusion scheme to tackle

	plane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	meanAP
V4[7]	28.9	59.5	10.0	15.2	25.5	49.6	57.9	19.3	22.4	25.2	23.3	11.1	56.8	48.7	41.9	12.2	17.8	33.6	45.1	41.6	32.3
best2007[15]	26.2	40.9	9.8	9.4	21.4	39.3	43.2	24.0	12.8	14.0	9.8	16.2	33.5	37.5	22.1	12.0	17.5	14.7	33.4	28.9	23.3
UCI[2]	28.8	56.2	3.2	14.2	**29.4**	38.7	48.7	12.4	16.0	17.7	24.0	11.7	45.0	39.4	35.5	**15.2**	16.1	20.1	34.2	35.4	27.1
LEO[27]	29.4	55.8	9.4	14.3	28.6	44.0	51.3	21.3	20.0	19.3	**25.2**	12.5	50.4	38.4	36.6	15.1	19.7	25.1	36.8	39.3	29.6
Oxford-MKL[24]	**37.6**	47.8	**15.3**	15.3	21.9	**50.7**	50.6	**30.0**	17.3	**33.0**	22.5	**21.5**	51.2	45.5	23.3	12.4	**23.9**	28.5	45.3	**48.5**	32.1
Proposed	36.7	**59.8**	11.8	**17.5**	26.3	49.8	**58.2**	24.0	**22.9**	27.0	24.3	15.2	**58.2**	**49.2**	**44.6**	13.5	21.4	**34.9**	**47.5**	42.3	**34.3**

Table 2. Full results on PASCAL VOC 2007 challenge datasets. best2007 was the best results submitted to the VOC2007 challenge [15]. The V4 is from [7] without context based post-processing. The UCI [2] method adopts multi-object layout to do object detection. The LEO method [27]used a latent hierarchical model to represent an object. Oxford-MKL method [24] used four types of multi-level feature and achieved very competitive results on VOC2007. Our method has no context rescoring.

the problem of multiple features combination for part based model. And the proposed method achieves very competitive results on PASCAL VOC datasets.

6. ACKNOWLEDGEMENT

This work is supported by National Natural Science Foundation of China (Grant No.60875021,60723005), NLPR 2008NLPRZY-2, National Hi-Tech Research and Development Program of China (2009AA01Z318), Key Project of Tsinghua National Laboratory for Information Science and Technology.

References

[1] N. Dalal and B. Triggers. Histograms of Oriented Gradients for Human Detection. *In CVPR*, 2005.

[2] C. Desai, D. Ramanan, C. Fowlkes, and U. C. Irvine. Discriminative Models for Multi-class Object Layout. *In ICCV*, 2009.

[3] T. Deselaers and V. Ferrari. Global and Efficient Self-Similarity for Object Classification and Detection. *In CVPR*, 2010.

[4] R. S. J. Estepar. *Local Structure Tensor for Multidimensional Signal Processing. Applications to Medical Image Analysis*. PhD thesis, University of Valladolid, Spain, 2005.

[5] Felzenszwalb, P. F., Huttenlocher, and D. P. Pictorial Structures for Object Recognition. *In IJCV*, 2005.

[6] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester. Discriminatively Trained Deformable Part Models, Release 3.

[7] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester. Discriminatively Trained Deformable Part Models, Release 4.

[8] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object Detection with Discriminatively Trained Part Based Models. *In TPMAI*, 2010.

[9] R. Fergus, P. Perona, and A. Zisserman. Object Class Recognition by Unsupervised Scale-Invariant Learning. *In CVPR*, 2003.

[10] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid. Groups of Adjacent Contour Segments for Object Detection. *TPMAI*, 2008.

[11] J. Friedman, T. Hastie, and R. Tibshirani. Additive Logistic Regression: a Statistical View of Boosting. *Annals of Statistics*, 1998.

[12] S. Hussain and B. Triggs. Feature Sets and Dimensionality Reduction for Visual Object Detection. *In BMVC*, 2010.

[13] R. Lienhart, E. Kuranov, and V. Pisarevsky. Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection. *In PR*, 2003.

[14] E. Mark, L. Gool, C. K. Williams, J. Winn, and A. Zisserman. The Pascal Visual Object Classes (VOC) Challenge. *In IJCV*, 2010.

[15] E. Mark, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007(VOC2007) Results.

[16] Y. D. Mu, S. C. Yan, Y. Liu, T. Huang, and B. F. Zhou. Discriminative Local Binary Patterns for Human Detection in Personal Album. *In CVPR*, 2008.

[17] T. Ojala, M. Pietikainen, and D. Harwood. A Comparative Study of Texture Measures with Classification based on Featured Distributions. *In PR*, 1996.

[18] F. Porikli, P. Meer, O. Tuzel, and O. Tuzel. Human Detection via Classification on Riemannian Manifolds. *In CVPR*, 2007.

[19] P.Schnitzspan, M.Fritz, S.Roth, and B.Schiele. Discriminative Structure Learning of Hierarchical Representations for Object Detection. *In CVPR*, 2009.

[20] P. Schnitzspan, S. Roth, and B. Schiele. Automatic Discovery of Meaningful Object Parts with Latent CRFs. *In CVPR*, 2010.

[21] W. Schwartz and L. Davis. Learning Discriminative Appearance-Based Models Using Partial Least Squares. *In Proceedings of the XXII Brazilian Symposium on Computer Graphics and Image Processing*, 2009.

[22] A. Torralba, K. P. Murphy, and W. T. Freeman. Sharing Features: Efficient Boosting Procedures for Multiclass Object Detection. *In CVPR*, 2004.

[23] M. Varma and B. R. Babu. More Generality in Efficient Multiple Kernel Learning. *In ICML*, 2009.

[24] A. Vedaldi, V. Gulshan, M. Varma, and A. Zisserman. Multiple Kernels for Object Detection. *In ICCV*, 2009.

[25] X. Wang, T. X. Han, and S. Yan. An HOG-LBP Human Detector with Partial Occlusion Handling. *In ICCV*, 2009.

[26] T. Watanabe, S. Ito, and K. Yokoi. Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection. *Advances in Image and Video Technology*, 2009.

[27] L. Zhu, Y. Chen, A. L. Yuille, and W. T. Freeman. Latent Hierarchical Structural Learning for Object Detection. *In CVPR*, 2010.