PROCEEDINGS OF SCIENCE

Object Detection based on Deformable Parts Model with Global Context

Yuanyuan Wang[1,2]
College of Communication Engineering, Jilin University, Changchun, 130022, China
E-mail: 14487299@qq.com

Xinyu Cui
College of Communication Engineering, Jilin University, Changchun, 130022, China
E-mail: 522022658@qq.com

Mianshu Chen[3]**, Aijun Sang, Xiaoni Li**
College of Communication Engineering, Jilin University, Changchun, 130022, China
*E-mail:*chenms@jlu.edu.cn, sangaj@jlu.edu.cn, 282786447@qq.com

The Deformable part-based model (DPM) is a remarkable algorithm in object detection. In this paper, it is combined with the global information to improve its performance. The gist feature of an image is extracted to capture its global information. After that, the principal component analysis (PCA) is used to reduce the dimensionality of the gist feature. The k nearest neighbor distance (k-NND) is utilized to judge the similarity of an image and an object. To accelerate**our** algorithm, every object is represented as several object models, which can be obtained by affinity propagation (AP). Finally, the score of DPM is merged with one of k-NND to rescore an image. Experimental results show that the introduction of global context is positive for the object detection.

CENet2015
12-13 September 2015
Shanghai, China

[1] Speaker

[2] Corresponding Author

[3] This work was supported by the Natural Science Foundation of Jilin Province(20130101045JC), the Science and Technology Development Plan of International Science of Jilin Province, the Technology Cooperation Projects(20130413053GH), National Natural Science Foundation of China(61171078), the International Science and Technology Cooperation of Science and Technology Development Plan Projects of Jilin Province(20140414013GH) and the Youth Foundation of Science and Technology Development Plan Project of Jilin Province（20130522164JH）.

1. Introduction

The object class detection and recognition is a very active research direction in computer vision, pattern recognition and machine learning fields. Deformable part model was proposed by Pedro Felzenszwalb in 2007 and became the champion in 2007 PASCAL VOC object detection contest [1]. It was attracting wide attention of researchers, and thus a number of improved algorithms were thus proposed for deformable part model, for example: the enhanced model of HOG-LBP [2].

In addition, it has been found that the environmental information of an object played a strong supporting role in the object detection and recognition. Based on the above idea, a data decomposition algorithm for the deformable part model was presented bythe Institute of Automation of Chinese Academy of Sciences in 2011 and won the champion of the PASCAL VOC testing in two consecutive years. The method is mainly based on the variability component model to improve on the bottom of the HOG feature and propose the boosted HOG-LBP feature. Gentle Boost choose part of LBP feature and hog feature are combined and the object detection results have significantly improved. In addition, the model also introduces another important improvement by using a variety of contexts and RBF SVM to carry out contextual learning and make the average accuracy to reach 36.8%; besides, it introduced spatial hybrid modeling and contextual learningin 2011[3-4].

Based on the above reasons, this paper combines the global context information for further improvement of the performance of DPM model. Chapter 2 introduces the deformable parts model and context information, and then Chapter 3 discusses in detail the algorithm structure and the core algorithm. At last, the experiment results demonstrate effectiveness of the algorithm .

2 Deformable Parts Model

Deformable part model (DPM) is a two layers' model, which includes the root layer model and part layer model[1]. The root layer model captures global features of the object and the part layer model captures local features of the object. The object class detection of DPM model is divided into two parts, model training and model testing. At the stage of model training, the templates of object class root and the structure of components are set. According to the image of training object, we extract features of root template and part template; at the same time, we train the elastic constraint of part template according to the differences within the object class. In the course of object class detection, an image is realzied with different scales and its features are extracted. Under certain scale we use root template to match an object while using the part template to match the corresponding parts. The elastic constraint of the position of part template is also taken into account. The overall match score s_{DK} is taken as the score of the object class K in a position.

2.1 Utility of Information of Context

DPM model can be combined with semantic context to re-evaluate the test results and improve the detection effect. The semantic context has significantly increased the average accuracy of its detection.

Although the introduction of semantic context has significant effects, it also has obvious defects. On one hand, the semantic context highly depends on the database and its performance changes sharply on different database; on the other hand, in order to extract context information, we need to run all relevant models of object recognition on every image and obtain the best results of detection, which will result in very high computational cost and high reliability of the object detection algorithm; but the reliability of the object detection algorithm is the problem we have to solve. As a result we try to introduce the global context to improve the detection effect of DPM model[5].

3. Object Class Detection Intergrated with Global Contect

Fig. 1 is the block diagram of object class detection algorithm with the integration of the global information. Firstly, the gist feature of every image in the training set and testing set are a 544-dimensional feature vector. Then the dimensionality reduction of PCA (Principal Component Analysis) is used to reduce the dimension of feature vector to 80; secondly, the k-nearest neighbor distance (k-NND) is utilized to judge the similarity of an image and an object. To accelerate our algorithm, every object is represented as several object models, which can be obtained by affinity propagation (AP). Finally, the score of DPM is merged with the one of k-NND to rescore an image.

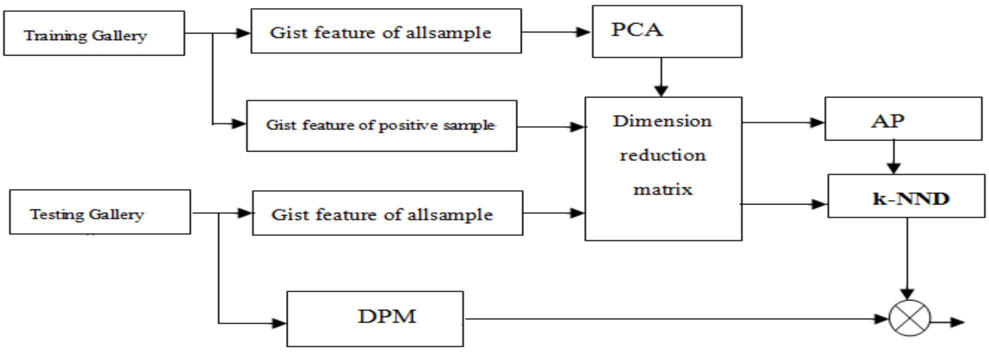

Training Gallery
Gist feature of allsample
PCA
Gist feature of positive sample
Dimension
reduction
matrix
AP
k-NND
Testing Gallery
Gist feature of allsample
DPM

Figure 1: Block Diagram of Object Class Detection Algorithm with Integration of the Global Context

3.1 Extraction of the Global Context based on Gist

In 2007, based on the central filter characteristics of biological, Sigian and Ltti proposed a new gist algorithm to extract information of the scene feature [6].The algorithm combines the gray information, color information and direction information. It has more information compared to the previous model of gist[7]. The process of extraction of gist feature is shown in Fig. 2 [6-7].

The image is divided into three channels of feature, namely, the direction information, the color information and the bright information. The channel of direction characteristics: There are four directions (0,45,90,135) of the gabor filter with filtering on four scales respectively; so the total number of the sub-channels is 16. The channel of color feature contains two filters, the red–green filter and the blue–yellow filter with filters central-surrounded on six scales respectively; thus the total number of the sub-channels is 12. As to the channel of luminance characteristics, it has dark-light filter central-surrounded on six scales; the total number of the sub-channels is 6. Therefore, this model of gist has 34 sub-channels and each channel has 16 sub-dimensional features. The feature of gist extracted from Sigian's algorithm is a feature vector which has 544-dimensional.

The gist algorithm is simple and can achieve the best classification accuracy in some matches of scene classification and in some standard galleries of scenario; thus we chose gist algorithm to capture global information in our paper.

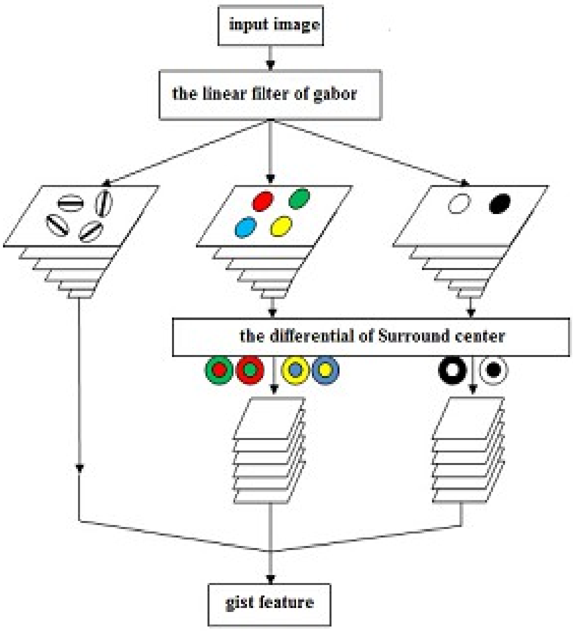

input image
the linear filter of gabor
the differential of Surround center
gist feature

Figure2 : Schematic of Feature Extraction of Gist

3.2 PCA Dimensionality Reduction

PCA (Principal Component Analysis) is a statistical analysis method used to find out the principal contradiction of things. It can parse out the main factors from diverse things to reveal the essence of things and simplify complex problems; as a result, PCA is mainly used for reduction of data dimensionality. This is to say that solving a projection matrix can drop the multi-dimensional vector of the original data to a low dimension from high-dimension and reach certain contribution rate. In this paper, while faced with the high-dimensional of gist feature, PCA is used to reduce the dimensionality of gist feature and the computational complexity.

3.3 AP Clustering

Affinity Propagation (AP) Clustering is a clustering algorithm proposed in the journal Science. It clusters according to similarity (such as Euclidean distance) of N data points [8]. These similarities may be symmetrical or asymmetrical. AP algorithm does not have to pre-set the number of clusters; on the contrary, it will make all the data points as a potential clustering center, calling exemplar.

Whether the k-point can become a cluster center or is not determined by a evaluation criteria of similarity value $s(k,k)$, $s(k,k)$ is the similarity when jserves as the cluster center of i) of the diagonal matrix S. The bigger the value of $s(k,k)$ is, the greater likelihood of the point will become the cluster center. The value is also known as the reference level p

(preference). AP algorithm delivers two types of messages: responsibility $r(i,k)$ and availability $a(i,k)$, in which, $r(i,k)$ represents the numerical information that the point of i sends to the k of candidate cluster center, indicating that whether k point is suitable as cluster center of i point. $a(i,k)$ represents numerical information that the k point of candidate cluster center is sent to i point to reflect whether i point chooses k point as cluster center of i point.). The bigger the value of $r(i,k)$and $a(i,k)$ is, the greater likelihood of the point of k will become the cluster center and the point of i belongs to the cluster which the center is k point.

Here are the formulas of $r(i,k)$and $a(i,k)$

$$r(i,k)=s(i,k)-max\{a(i,j)+s(i,j)\}\, j\in\{1,2,\cdots,N,j\neq k\} \tag{3.1}$$

$$a(i,k)=min\{0,r(k,k)+\sum_{j}\{max(0,r(j,k))\}\}\, j\in\{1,2,\cdots,N,j\neq k,j\neq i\} \tag{3.2}$$

$$r(k,k)=p(k)-max\{a(k,j)+s(k,j)\}\, j\in\{1,2,\cdots,N,j\neq k\} \tag{3.3}$$

As can be seen from the above formulas that $r(k,k)$ $a(i,k)$ andincreases with the increase of $p(k)$, which increases the likelihood of k point to become the final cluster center. Thus, the increase or decrease of p can increase or decrease the number of clusters of AP output.

In addition, AP algorithm continuously updates the attractiveness and attribution values of each point, which is a feature vector of gist of an image in the article through an iterative process until the number of quality exemplar is m; at the same time, the rest of data points will be assigned to the corresponding clustering. In the paper, AP is used to find the clustering of feature vectors and accelerate k-nearest neighbor to reduce the computational complexity.

3.4 k Nearest Neighbor Classifier

The k-nearest neighbor algorithm was firstly presented by Hart and Cover in 1968. As this algorithm is simple and the classification works well, it has attracted wide concern and research, and achieved considerable development in the latest half century. At present, the algorithm has been widely used in the machine learning based on statistics.

According to the basic idea of k-nearest neighbor algorithm, in order to adapt them to the image detector, the concept of k-nearest neighbors' scores is introduced. Specific algorithm is shown as follows: to identify the most similar k samples, the similarity between X sample which has to be detected with all AP cluster centers is calculated. Finally, the score of k-nearest neighbor of X sample is calculated by Equation 4.

$$S_X=\frac{1}{k}\sum_{K=1}^{k}S_{XK} \tag{3.4}$$

$$S_{XK}=e^{-\frac{1}{d_{xK}}} \tag{3.5}$$

The score of S_{XK} is score between x with sample of K, and$K=1,2,\cdots,k$.

The k-nearest neighbor algorithm is an algorithm based on statistics, which greatly depends on the distribution of the training data samples. Faced with the feature of k-nearest neighbor algorithm, the article improves the k-nearest neighbor algorithm by adjusting the sample number of the sample library, which maintains the training library to meet the needs of k nearest neighbor algorithm.

As to the k-nearest neighbor algorithms, we have a question: suppose all training samples be $\left(c_1, c_2, \cdots c_k\right)$, and the samples were tested when k=1 and the correct rate can reach 100%; we used the same test sample of library for training at k = 3, how much the correct rate can reach or at k=5? Here we select the data of 2000*100 (pos330, neg1670) as the training library of sample. Test results are shown as Table 1.

In Table 1, we can see that with the increase of k, the ratio of NT (negative true) and pos becomes bigger and bigger, the rate of false detection is higher and higher, even more than 50%. This is simply a disaster for the target recognition; therefore, to improve the classification capacity, our idea is to adjust the training data to improve test results. Assume that all samples of training sample be$(c_1, c_2, \ldots c_k)$, and set the value of k and test training library by sample, and then rejoin NT (negative true) samples and FP (false positive) samples of detection errors to the training library, again and again, until jump out the cycle when the overall detection precision reaches a threshold.

	NT		FP		accuracy
k=3	100	30.3%	89	5.3%	90.55%(1811/2000)
k=5	135	40.9%	90	5.4%	88.75% (1775/2000)
k=7	170	51.5%	83	5.0%	87.35% (1747/2000)

Table 1 : Ratio of NT, FP and Accuracy of the Increase of k

According to the idea, we designed a corresponding program. The results are shown in Table 2. We can see that with the increase of iteration, NT and FP dropped sharply and the overall detection accuracy increased; as the a result, the iterative algorithm can effectively improve the classification ability of k nearest neighbor algorithm.

In the paper, the score of an image is merged with the score of DPM and k-nearest neighbor (as shown in Formula 6). The score is used to calculate the evaluation criteria of Recall Rate and Precision Rate, which has positive effect on experimental results than the score of DPM.

$$S_{X(totle)} = S_X \times S_{XD} \tag{3.6}$$

	NT		FP		accuracy
iteration=1	100	30.3%	89	5.3%	90.55%(1811/2000)
iteration=2	28	8.5%	83	5.0%	94.45%(1889/2000)
iteration=3	3	0.9%	12	0.7%	99.25%(1985/2000)

Table 2 : Ratio of NT, FP and Accuracy of the Increase of Iteration

4. Experiment Results and Analysis

The experiments use a standard image library (the number of standard images of training library is 2111 and the number of image of verify gallery is 2221) which used the target recognition in PASCAL2010 as experimental data. The experiments used Recall Rate, Precision Rate and Average Precision (AP) as the evaluation criteria, and used the average precision whether improved or not as a standard to determine the effectiveness of the algorithm. In the end, the corresponding experimental data were obtained through MATLAB to verify the accuracy and the superiority of the algorithm.

These three indicators are borrowed from information retrieval, index classification, recognition, translation and other areas. In order to make the index suitable to the field of image detection, we redefine them as

$$Recall = RD / T \tag{4.1}$$

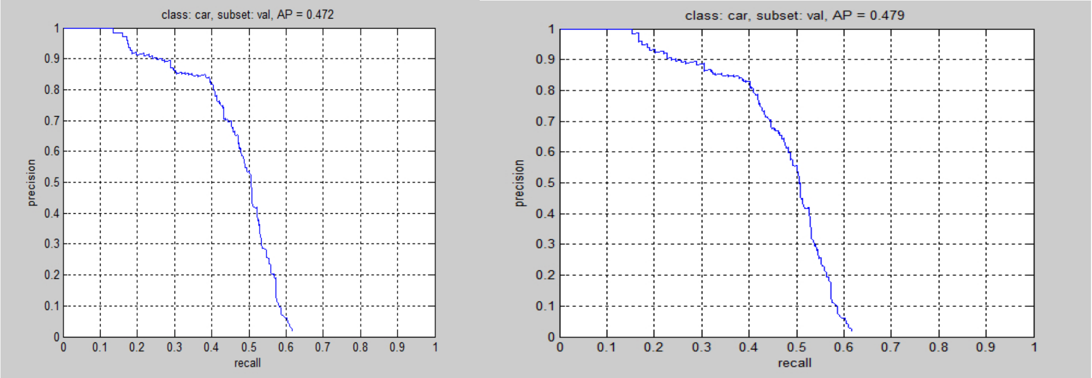
$$Precision = RD / TD \tag{4.2}$$

RD is the number of objects to be detected correctly; T is the total number of objects in the gallery; TD is the total number of objects to be detected.

$$AP = \frac{1}{r}\sum_{i=1}^{r} i / the\ position\ of\ image\ be\ detected\ at\ i\ th \tag{4.3}$$

	Orignal	The article
car	0.472	0.479
person	0.536	0.534
aeroplane	0.541	0.541

Table 3 : Average Precision of Original Article Algorithm



(a)DPM testing AP=0.472 (b) After the review AP=0.479

Figure 3: Contrast Figure of Test Results of Car Object Class

The paper carried out experiment as described in Chapter III. By Fig. 6, we can find the original AP = 0.472 and get AP = 0.479 after a similarity calculation of gist features. Furthermore, the image libraries of car, person, and aeroplane as experimental subjects were selected to verify the effectiveness of the algorithm. The experimental results are shown in Table3, from which we can find the Average Precision （AP） of the part of Gallery has improved through the algorithm. Scene aeroplane-class of objects appearance is single. Upon fusing of the two contextual information, the testing effect is similar. The scene of car-class object appearance is more complex. It may be roads, streets, or parking garage. The detection results are improved more obviously. The scene of person-class objects almost has no rule to follow and the relative area occupied by such objects in the whole image is also the smallest; therefore, the correlation of the scene is the weakest and the detection result is the worst. In summary, the proposed method is more suitable for the detection scene which is slightly more complicated, and the detection target has rules to follow.

5. Conclusion

To sum up, this paper describes an effective method for object detection and Average Precision（AP）of the part of Gallery has been improved through the algorithm. At the same time can be seen, the algorithm also has a lot of room for improvement. For example, this paper mainly uses the global context of the information, no use of local context information. In fact, we can try to use gist algorithm to extract the global context, and integration of two kinds of context information, so as to further improve the object class detection performance.

References

[1] P. Felzenszwalb, D. McAllester, D. Ramanan. A . *Discriminatively Trained, Multiscale, Deformable Part Model.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Anchorage, Alaska, USA, 1-8(2008).

[2] Y. N. Yu, J.G Zhang, Y.Z Huang, S. Zheng, W.Q. Ren, C. Wang, K.Q. Huang, T.N. Tan, *Object Detection by Context and Boosted HOG-LBP*[C], in PASCAL Visual Object Challenge Workshop, European Conference on Computer Vision (ECCV), pp，252010

[3] Y.Z. Huang, K.Q. Huang, Y. N. Yu and T.N. Tan. *Salient coding for image classification.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado, USA, 1753–1760(2011).

[4] J.G Zhang, Y.Z Huang, K.Q. Huang, Z.F. Wu, and T.N. Tan. *Data decomposition and spatial mixture modeling for part based model*[C]. Proceedings of the Asian Conference on Computer Vision (ACCV), Daejeon, Korea, 123–137(2012).

[5] B. C. Russell, A. Torralba, C. Liu, R. Fergus, W.T. Freeman, *Object recognition by scene alignment*[C], Neural Information Processing Systems Foundation（NIPS）,1-8 2007

[6] C. Siagian and L.Itti. *Rapid Biologically-inspired scene classification using features shared with visual attention*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 29(2), 300–312(2007)

[7] K. Murphy, A. Torralba, W. Freeman, *Using the forest to see the tree:a graphical model relating features*, objects and the scenes, NIPS,2003.

[8] B. J. Frey,Ducck D.*Clustering by passing messages between data points.Science*, 315(5814):972-976(2007).