Boosting Speed and Accuracy in Deformable Part Models for Face Image in the Wild

Dinh-Luan Nguyen[1], Vinh-Tiep Nguyen[1], Minh-Triet Tran[1], Atsuo Yoshitaka[2]
[1]Faculty of Information Technology
University of Science, VNU-HCM
Ho Chi Minh City, Vietnam
1212223@student.hcmus.edu.vn, {nvtiep, tmtriet}@fit.hcmus.edu.vn
[2]School of Information Science
Japan Advanced Institute of Science and Technology
Ishikawa, Japan
ayoshi@jaist.ac.jp

***Abstract*— Face detection using part based model becomes a new trend in Computer Vision. Following this trend, we propose an extension of Deformable Part Models to detect faces which increases not only precision but also speed compared with current versions of DPM. First, to reduce computation cost, we create a lookup table instead of repeatedly calculating scores in each processing step by approximating inner product between HOG features and weight vectors. Furthermore, early cascading method is also introduced to boost up speed. Second, we propose new integrated model for face representation and its score of detection. Besides, the intuitive non-maximum suppression is also proposed to get more accuracy in detecting result. We evaluate the merit of our method on the public dataset Face Detection Data Set and Benchmark (FDDB). Experimental results shows that our proposed method can significantly boost 5.5 times in speed of DPM method for face detection while achieve up to 94.64% the accuracy of the state-of-the-art technique. This leads to a promising way to combine DPM with other techniques to solve difficulties of face detection in the wild.**

Keywords-Deformable part models, face detection, non-maximum suppression, image in the wild

I. INTRODUCTION

Face detection is a classic task in computer vision. It is an attractive problem because it has a wide range of applications such as digital cameras, social networks, and surveillance. Although many great strides have been proposed to improve accuracy and speed, this task is still challenging because of the variety of unconstrained images in the wild. Since it is difficult, several ways such as using contour fragments [8], combining low-level features and using probabilistic face structure [2] are created to overcome obstacles. Among these methods, work of Viola and Jones [21] is famous for its high speed and being one of the best frameworks in face detection during early days. However, the performance of this method is far from satisfactory especially when it is applied to images in the wild due to the variety of head pose, occlusion, illumination, etc.

Mapping fixed feature representation to fixed classifier is a common key idea in previous works although they are different in feature representation and learning algorithm. This idea is easily ambiguous in practice since variations of image can exist. For instance, the relative positions of eyes and nose are the same for every face but the pose, expression, and illumination are different. Instead of using fixed features, we use part based model to solve these difficulties. By applying part based model, we can deal with not only deformable objects but also variant lighting ones.

One of efficient methods to create part model is using histogram of oriented gradients (HOG) [1]. Tolerating not only local geometric but also photometric transformation for face in the wild is one of advantages of HOG. However, the original version of HOG has high computation cost since it uses division and inverse trigonometric operations while calculating orientation partition. To take advantage of HOG, we propose pre-calculated lookup table to avoid redundant operations. By applying lookup table, our method shows superior to the original ones in reducing complex calculation.

The characteristic of images in the wild is the variant of object such as pose, illumination, changeable shape, etc. Thus, using method that can deal with variant factors is one of the promising solutions. Not only being based on HOG but also providing root model and part models, Deformable Part Models (DPM) [3] can overcome these challenging characteristics. Root model is used for representing whole general shape while part models are accounted for changeable object's components. To measure the important of parts, DPM has one root model and many part models which are constrained together by score functions. To extract the location of part model, DPM uses sliding window to detect, which is really expensive for computing. Furthermore, DPM creates a pyramid image and train models based on location and scale in each pyramids. Detection score returned by DPM's model is determined by score of appearance minus score of deformation cost. The appearance score is calculated by combining the correlation between sequence of root and parts' filters and HOG feature.

Several works have been proposed to increase accuracy and decrease running time of DPM. Pruning redundant hypothesis early is one of the key ideas to reduce computation cost [4, 16, 17]. Dubout et al. [6] uses Fast Fourier Transform (FFT) to boost correlation. These methods, however, take around 1 second per image for

978-1-4673-8234-2/15 $31.00 © 2015 IEEE
DOI 10.1109/ACOMP.2015.9

face detection in Face Detection Data Set and Benchmark (FDDB).

Although DPM is usually considered as a time-consuming method, it is one of the state-of-the-art methods for object detection. Therefore, in this paper, we first inherit the original DPM as a baseline for face detection, then we propose a novel method based on DPM to boost not only the speed but also the accuracy to detect faces in unconstrained settings. We combine cascade part running and HOG feature extraction to boost up running time and make it become feasible to be applied in practical application.

Main contribution. There are two main principles in our proposed system. First, in speed acceleration, we create a lookup table with probabilistic codebook quantization so as to approximate the exact inner product and present new early cascading process to eliminate unnecessary computation. Second, in accuracy section, we propose a new integrated face model using DPM for adaptive representation and localization, and new approach for calculating non-maximum suppression.

Experiments on FDDB dataset show that our proposed method not only significantly speeds up the computation but also improves the accuracy for face detection in unconstrained images. Our method can run 5.5 times faster than the original DPM method, and also faster than existing accelerated DPM versions. Besides, our method can achieve up to 94.64% the accuracy of the state-of-the-art method for face detection.

The rest of our paper is organized as follows. Section II reviews related works in face detection and some improvements in both speed and accuracy of DPM. Our primary technical contributions in speed and accuracy are described in Sections III and IV respectively. We describe experimental results using proposed method in Section V. Finally, conclusion is presented in Section VI.

II. Related Works

Some recent works accelerate the speed by calculating magnitude such as coarse-to-fine [4], branch-and-bound [5] and FFT [6]. In DPM's score calculation step, detection score is determined by the sum score of root filter and part filters minus the deformation cost. Each root and part filters score is calculated by multiplying inner product of HOG feature and weight vectors, which takes too much time because of using sliding window and high dimension of feature presentation.

Yan et al. [10] convert star-structure to cascade, which efficiently prune unpromising calculation. Pedersoli et al. [4] proposes a coarse-to-fine approach that convert image into low resolution to cut off computation cost. Work of P. Felzenszwalb [17] prunes redundant hypotheses in part score calculation step. Shotton et al. [8] use contour information to accelerate correlation between root score and part score. In K. Iasonas's work [7], root and part scores are densely calculated by creating correlation between HOG map and root filter and then transform 2D correlation into 1D correlation. Dual Tree Branch-and-Bound introduced in [5] is an efficient bounding based method for detection with DPM. This work substantially accelerates the stages following part computation, which turns complexity from linear to nearly logarithm to make DPM become slightly faster. However, all above works have one disability is that the cost for computing part and root scores is really expensive and this task takes a long time to complete, which is the main bottleneck in speed acceleration for DPM.

In the original version of DPM [3], calculating HOG is one of the most important steps. However, this step takes around 80% of processing time per image on a single thread, which makes DPM slow. Some works such as [18, 19] boost HOG with the computational capacity of GPU but they do not improve algorithm themselves because they just make use of GPU's speed for acceleration. Based on these observations, by using look-up table approach and early cascading method, we significantly improve HOG calculating step described in Section III.

For face detection area, there are two main approaches: single template and part-based. This classification is based on representation of face regardless of classifiers and features. In a single template approach, by using a single detection window, the whole face pattern is captured. Papageorgiou et al. [22] uses Haar wavelet features together with a polynomial Support Vector Machine (SVM). Viola et al. [21] enlarge space-time information to simple Haarlike wavelet features for face detection. In static images, Dalal and Triggs [1] show excellent performance by using dense HOG representation and linear SVM.

On the other hand, in a part-based approach, system captures the pattern of each part and then combines results to make a final decision. In general, part-based approaches can deal with variant appearances of face such as pose, illumination, occlusion, and provide a more complex model for face detection problem. Thus, part-based model shows more favorable performance because the constructed models are rich and flexible. Thanks to its elegant formulation and intuitive interpretation, DPM has established itself as a standard for generic object detection [7, 10, 20]. DPM consists of latent variables for alignment estimation and clustering at training time. To make DPM become robust and able to deal with intra-class variance, multiple components and deformable parts are used. Tree structure trained DPM with supervised part position is applied to face detection [18, 19] and fiducial points estimation [20], showing improved results over baseline DPM.

Based on part-based idea, Felzenszwalb et al. [2] expresses object as an union of parts, which are used contour segments and HOG to represent their features. By sequential scanning and using mean shift to find the local maximum points in the image scale space then determined the location of the target, Felzenszwalb [3] improves a little in compare with baseline. Shotton [8] proposes a new automatic visual recognition system based only on local contour features, which is capable of localizing objects in space and scale. To capture the main object's structure, Xiang Bai et al. [9] use the skeleton information, which improves non-rigid deformation and modeling articulation. More recently, Yan [10] and Zhu [11] use local parts around landmarks to represent face, and propose a tree

structure deformable model for joint face detection, landmark location and pose estimation with promising performance. However, most of current methods are slow and the accuracy is not really high. Thus, there is a huge need to boost the speed and accuracy of DPM to deal with variance of image occlusion and illumination.

III. Speed Acceleration Method

In this section, we introduce our improvement in computing a part score and propose a method for cascading images.

A. Part score approximation

The whole model for representing an object comprises a root model and several part models. Each model has its own score to represent the position and relation themselves. Thus, computing a model's score is a compulsory step to develop model. In Felzenszwalb's work [3], part score is calculated by using the inner product of HOG features and part weight vectors, which are pre-trained in early stages. Let $S[t]$ denote the score of a part model in position t, Z is the subspace for searching displacement and M is the dimension of HOG, the formula for computing a part score is described in Equation 1.

$$S[t] = \sum_{x \in Z} \sum_{i=1}^{M} W[x,i] H[t+x,i] \quad (1)$$

where $H[x,i]$ and $W[x,i]$ are the i^{th} component of HOG and weight vector in position x respectively. To be specific, supposed $[0,10] \times [0,10]$ is the size of subspace for part filter Z and HOG's dimension M equal to 32, the computation in Equation 1 needs 121x32 summations and multiplications in total. Because of searching every possible position for each part, part score is computed frequently. Based on the observation that these scores in each calculation are the same, we propose to replace the multiplication in Equation 1 by a pre-computed lookup table in which values are approximated to the real calculated one. Equation 1 can be written by:

$$S[x] = \sum_{x} \langle W_x, H_{t+x} \rangle \quad (2)$$

where $W_x = [W[x,1], \dots, W[x,M]]$ is the M-dimension row vector and $H_x = [H[x,1], \dots, H[x,M]]^T$ is the M-dimension column vector.

The lookup table is constructed by quantizing M multiplication to a table that approximates the final output. We create codebook $C = \{C_1, \dots, C_K\}$ for calculating $H[x,i]$ by using K-means (K is hard-chosen) and pre-compute $M.|Z|$ array for representing H_x. This codebook is built based on the possible part location returned by the connection string between a part filter and a root filter. To be specific, from each triple values of root, part filter and connection string, we have one value for part's position. By gathering these scores and applying K-means, we quantize them into K clusters C_i, $i = \overline{1..K}$ representing K highest possible places for part filter.

Furthermore, we create a rank list R for each position x as follows:

$$R[x] = argmin_k(C_k, H_x) \quad (3)$$

Based on the score returned from rank list R and using inverted index structure, we get the approximation as in Equation 4.

$$S[t] \approx \hat{S}[t] = \sum_{x} \langle C_{R[t+x]}, W_x \rangle \quad (4)$$

B. Early cascading method

One disadvantage of DPM baseline and its improvements is using non-cascade model because it takes a long time to process whole root and part models. Since the learned detector comprises one root and some part templates, we propose a new approach to create a cascading method, which significantly boosts our detector to become real time. Our cascading system consists of several stages so that to achieve occlusion invariance in any face region, the score of every part stage is the score of the deformable parts in current stage plus its parent stage in the cascade structure. We also create a threshold in each stage to eliminate unpromising candidates whose scores are not satisfied. With the cascade structure, our detector not only keeps the advantage of model flexibility but also can avoid a lot of redundant computations.

Supposed the cascading method has n stages, the detection score at the i^{th} stage for image I with candidate region α is computed as follows:

$$S_i(I_\alpha) = \beta S_{i-1}(I_\alpha) + R_i(I_\alpha) \quad (5)$$

where $R_i(I_\alpha)$ denotes raw detection score and $0.5 \leq \beta \leq 1$ is the adaptive weight calculated based on $R_i(I_\alpha)$.

Besides, a hard threshold t_i in each i^{th} stage is also created to remove unqualified candidate. To be specific, if

1:	**For each** region α in image I
2:	**For each** i^{th} stage of n stage
3:	Calculate raw detection score $R_i(I_\alpha)$.
4:	Compute β value based on raw score.
5:	Calculate new detection score $S_i(I_\alpha)$ by equation 5.
6:	Check if new score is satisfied with threshold t_i. If not, move to the next region, otherwise, continue to the next stage.
7:	If the loop reaches the n^{th} stage, push $S_i(I_\alpha)$ into rank list.
8:	**End loop.**
9:	**End loop.**
10:	Re-rank the list to get top M highest detection scores for further process.

Algorithm 1: Proposed algorithm to convert original model to cascade one.

new score returned by Equation 5 is greater than t_i, the candidate is passed to the next stage for further calculation, otherwise it is considered as a negative sample and eliminated immediately. With cascade structure, the detector only pays attention to promising regions and the overwhelming majority of negative samples can be rejected in early stages.

Details for creating the cascading method is described in Algorithm 1. Top M regions returned from this algorithm are used for late processing to find relation between root and part models.

IV. Accuracy Boosting Techniques

Our main contribution in accuracy section is to propose a new formula for face representation and changing the way of calculating bounding box in non-maximum suppression function.

A. Face representation model

In baseline [3] and other improvements [4, 5, 17], all models involve linear filters which are applied to dense feature maps. A feature map is an array whose entries are D-dimensional feature vectors computed from a dense grid of locations in an image. Each feature vector intuitively describes corresponding local image patch. DPM uses HOG to represent features and constrain function to calculate score root and part filters.

We review the standard way of calculating HOG as follows. We create image pyramid with different scales from original input image. In each scaled image, gradient is computed by $[-1\ 0\ 1]$ convolution kernel and its transpose. Based on gradient orientation at each pixel in image, different bins of orientation are added by its corresponding magnitude.

Root and part filter is constructed by HOG's characteristics. The original work of DPM presents a model with 8 part filters which 6×6 pixels per each. However, this default configuration is used for detect wide range of class and it is not efficient for face presentation. In face detection problem, we propose a combined model which is derived from 5-part model and 4-part model. With the frontal face, 5 parts are necessary for 1 forehead, 2 eyes, 1 nose and 1 mouth. Part model for presenting forehead has twice resolution in comparison with others part because of its easy identification. The frontal model score is computed by formula:

$$S_{5-part} = 2^{P_1}(P_2 + P_{2'}) \log(P_3 P_4 + P_3 + P_4) \qquad (6)$$

where $P_{i=1,2,2',3,4}$ are part filters for forehead, 2 eyes, nose and mouth, respectively. This formula can be easily changed by many kinds of function such as polynomial, logarithm, exponential, etc. if those functions satisfy the relation between types of part. To be specific, forehead is the lucid object and easily to be recognized in frontal face so that the whole model is sensitive with any change of this part. Thus, part which represents forehead has an exponential score function. Because two eyes are symmetric, they equally contribute in $P_2 + P_{2'}$ polynomial. Since mouth and nose are asymmetric in whole model, they have slight impact on model score. Thus, we use log function to present this relation.

Similarly, with the occlusion face, 4-part model is constructed with 1 forehead, 1 eye, 1 nose and 1 mouth. Score for calculating 4-part model is described in Equation 7:

$$S_{4-part} = e^{P_3 P_4 + P_3 + P_4} P_2 P_1 \qquad (7)$$

The reason for proposing this score is that when face is occluded or not in frontal view, mouth and nose are easy to be realized. Thus, in this situation, exponential function is used to describe the influence of them. Besides, forehead and eyes are partly occluded and they do not contribute much to face model.

From two mid-model, we create our model by calculate the mean of them. In the feature pyramid $z = (p_0, \ldots, p_n)$, where $p_i = (x_i, y_i, l_i)$ specifies the level and position of the i^{th} filter, model score is calculated by the scores of part filter at their current position minus deformation cost created by the relation between part and root filter.

Let F_i is a filter for the i^{th} part, d_i is a 4-dimensional vector representing coefficients of a quadratic function, and v_i is a 2-dimensional vector representing parts' position, the deformation cost is calculated as Equation 8:

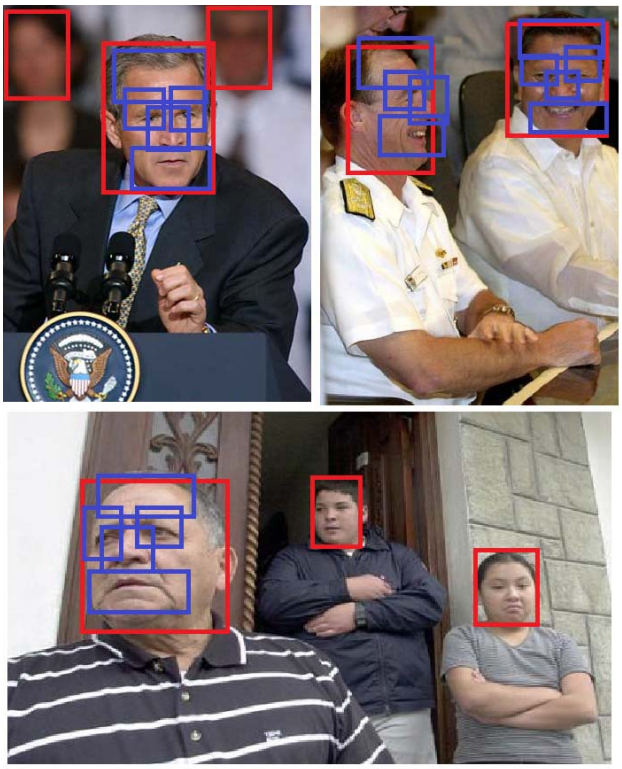

Figure 1. Detection with fusion model in FDDB dataset. 5-part model is used to deal with frontal face while 4-part model is suitable for detecting occluded ones. Root model (red bounding box) detects the general face shape and part models (blue bounding box) detect face's components.

$$S_{(p_0,\ldots,p_n)} = \sum_{i=0}^{n} F_i' . \delta(H, p_i) - \sum_{i=1}^{n} d_i . \theta(dx_i, dy_i) \quad (8)$$

where $\delta(dx, dy) = (\mathrm{dx}, \mathrm{dy}, dx^2, dy^2)$ are the part features of deformation and

$$\theta(dx_i, dy_i) = (x_i, y_i) - (2(x_0, y_0) + v_i) \quad (9)$$

denotes the displacement of the i^{th} part.

Figure 1 demonstrates our proposed model. Decision for choosing 4-part model, 5-part model, or only root model depends on size of face. This new model has advantage over the old one is that it can present face in many view such as left, right, up and down, which is suitable for face detection in the wild.

B. *Intuitive non-maximum suppression*

In Felzenszwalb's works [2, 3, 17] and other works using DPM as a core model [4, 5, 9, 19], non-maximum suppression (NMS) is only computed by calculating the coordinate of each bounding box and then compare each other to conclude if they overlap 50% area of the smaller bounding box or not. If the value is greater than fifty percent, it discards the bigger. Otherwise, the current box is not change. Formulation for choosing bounding box is described in Equation 10. This way to calculate suppression is not sufficient since it discards the uncommon area between the big and small one. Besides, if the list of candidate bounding box contains a few correct ones, this way of calculating leads to wrong returned result.

$$\frac{S(big\ box) \cap S(small\ box)}{S(small\ box)} \geq 50\% \quad (10)$$

Based on meticulous examination the advantages and disadvantages of the old method, we propose new method as described in algorithm 2 to make use of existing advantages and overcome defects in the old one.

1:	Set an zero matrix having the same size of input image.
2:	**For each** bounding box in the candidate list
3:	Cumulate the matrix area by overlapping bounding box score value.
4:	Save the size of the smallest bounding box.
5:	**End loop.**
6:	Calculate area position having the maximum score value.
7:	At the center point of max score area, expand the size of new detection window equal to the smallest bounding box's size.

Algorithm 2. Non-maximum suppression score based computation

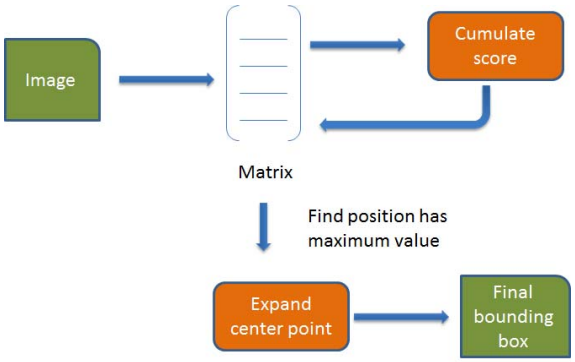

Figure 2. Proposed method for calculating bounding box for intuitive non-maximum suppression technique.

The key idea for proposing this method is from observation that if one area was detected by many bounding boxes, it also has the high possibility of real face. So by following Algorithm 2, we get the bounding box which covers highest possibility area of face. Figure 2 shows the method for calculating non-maximum suppression described in Algorithm 2.

V. Experimental Results

Dataset. FDDB [12] is famous for being a benchmark for face detection. It contains 2845 images including 5171 faces collected from the news photograph and displays large variation in background, illumination, appearance, and pose. In this dataset, the ground-truth provides the coordinator of ellipse which cover face region. Some state-of-the-art techniques use FDDB as a standard dataset and their results are published on FDDB website. Figure 3 shows some images in dataset with their ground-truth in FDDB dataset.

Evaluation. To be fair when comparing our result with other works, we use FDDB's evaluation protocol provided in the dataset.

A. *Accuracy experiments and results.*

First, we compare our result with the default configuration in DPM in order to prove the efficiency of our method. Table 1 illustrates the result of using model for face representation with many configurations. The default model with 1 root filter and 8 part filters is just suitable for generic object while new model shows the efficiency for face detection.

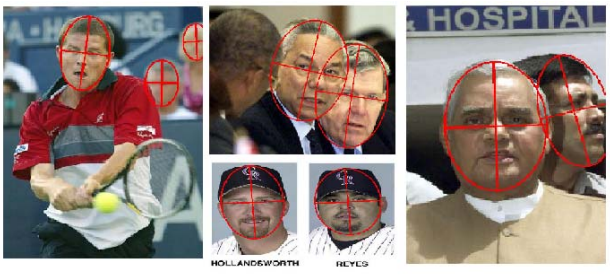

Figure 3. Some example images and annotations in FDDB dataset.

TABLE 1. COMPARISION BETWEEN DIFFERENT CONFIGURATION IN MAP

Configuration	*mAP*
DPM with default model	65.7%
Default DPM with intuitive NMS	68.2%
4-part model	69.2%
5-part model	70.1%
Combination of 4-5 part model	76.7%
Proposed method	**79.5%**

As we can see from Table 1, 4-part model or 5-part model itself slightly improves the mAP in overall. Since the images in DPM dataset cover wide range of conditions, the combination between these models significantly boosts the result (from 69.2% and 70.1%, the fusion model get up to 76.7% in mAP).

Besides, by using intuitive non-maximum suppression, we avoid a lot of wrong bounding boxes returned from rank list. By making use of the advantage of intuitive NMS, the eliminating unpromising region candidate improves 2.5% in default DPM and 2.8% in our system.

Second, in figure 4, we show the average 10 subfolders discrete ROC in FDDB dataset and compare our result with some state-of-the-art techniques include top 4 academic methods, and top 2 commercial systems. These techniques comprise Olaworks[1] - the best of commercial system, IlluxTech[2], the best academic result Yan [15], Jain [13], Mikolajczyk [14], and ViolaJones [21]. These results are published in FDDB dataset's website so that it is fair to compare with them. Result of our method shows that we nearly get the state-of-the-art in accuracy while our speed is superior, which is analyzed in the speed experiment results.

Figure 5. Some of difficult situations in FDDB dataset are correctly detected by our method, which includes low illumination, occlusion, different face pose, and blur condition

The ROC curve results are in Figure 4. Our proposed method can achieve 79.5% true positive rate in 1000 false positives images. The best commercial system Olaworks face detector gets 82.2% true positive rate and the best academic system Yan et al.[15] gets 82.4% true positive at the same number of false positive images. Furthermore, our result is also better than famous ViolaJones detector, Mikolajczyk, Jain, and commercial Illuxtech system.

Our method can deal with not only easy situation face conditions but also difficult ones. Figure 5 shows some difficult examples detected by our system. Although faces in these images have different illumination, occlusion or blur, our system can detect them correctly.

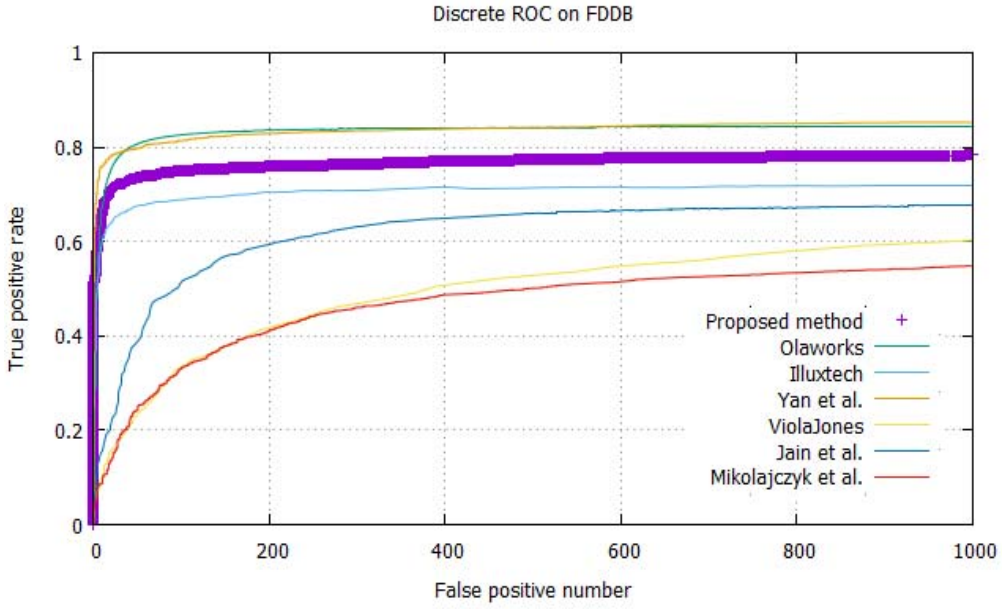

Figure 4. Result comparison with state-of-the-art on FDDB dataset.

[1] *http://eng.olaworks.com/olaworks/main/*
[2] *http://illuxtech.com*

TABLE 2. COMPARISON AVERAGE TIME (MEASURED BY SECOND) WITH SOME STATE-OF-THE-ARTS ON FDDB DATASET.

	Feature extraction	*Detection*	*Total*
Baseline DPM	0.46	4.36	4.82
Branch-Bound (DPM)	0.43	2.75	3.18
Cascade (DPM)	0.45	0.99	1.44
FFT (DPM)	0.48	0.98	1.46
Coarse-to-fine (DPM)	0.67	0.95	1.62
Using part score approximation	0.38	0.86	1.24
Using cascading method	0.24	0.77	1.01
Proposed Method	**0.19**	**0.68**	**0.87**

B. Speed experiments and results.

The proposed method is implemented based on DPM release 5. Besides the implementation of DPM release 5, we compare accelerated DPM versions, including cascade [10], branch-bound, coarse-to-fine, and FFT. All these methods except coarse-to-fine use the default configuration in DPM release 5, where the number of levels in an octave is 10, HOG bin size is 8, part number for each component is 8 and component number is 6. For coarse-to-fine DPM, the setting advised by the paper [4] is used, where component number is 4. We use these works' code to apply in FDDB dataset. The average feature extraction time, detection time and total time are reported in Table 2. Detection time equals to sum of the root and parts computation time. For fair comparison, all the codes run on the same PC with 2.53GHz Intel Core i5 Ram 8GB. The configuration of cascading method for formula in Equation 5 is $K = 256$ for K-mean clustering and top M result in Algorithm 1 is equal to 25.

Results from Table 2 shows that using part score approximation and early cascading method significantly boost running time. We accelerate the performance up to 5.5 times comparing to baseline DPM and 1.4 times comparing to some works of DPM improvement. The proposed method reduces unnecessary computation operator and redundant functions in early stages.

VI. CONCLUSION

In this paper, four novel techniques are proposed to solve the bottleneck of deformable part models. Part score approximation and early cascading method is proposed to boost up running time of DPM. Besides, combination between 4-part model and 5-part model together with new way of calculating non-maximum suppression help significantly increase accuracy. Although our method is nearly get state-of-the-art in accuracy but the speed is really huge improved in comparison to other techniques having the same accuracy. Techniques discussed in this paper also can be used for generic object detection and accelerate related models, which use part model. Additionally, our experimental results demonstrated that our method is more robust and superior than some DPM improvements and has the potential for practical use in detecting not only faces but other objects in the future.

ACKNOWLEDGMENTS

This work is supported by Japan Student Services Organization and Honors Program of University of Science, VNU-HCM.

REFERENCES

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human Detection" in Computer Vision and Pattern Recognition (CVPR), vol. 1, 2005, pp. 886–893.

[2] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model" in Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.

[3] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models" ,Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on, vol. 32, no. 9, 2010, pp. 1627–1645.

[4] M. Pedersoli, A. Vedaldi, and J. Gonzalez. "A coarse-to-fine approach for fast deformable object detection" in Computer Vision and Pattern Recognition (CVPR). IEEE, 2011, pp. 1353-1360.

[5] I. Kokkinos. "Rapid deformable object detection using dual-tree branch-and-bound" in Advances in Neural Information Processing Systems (NIPS), 2011.

[6] C. Dubout and F. Fleuret. "Exact acceleration of linear object detectors" in European Conference on Computer Vision (ECCV), Springer, 2012, pp. 301-311.

[7] K. Iasonas., "Rapid Deformable Object Detection using Bounding-based Techniques", Technical Report 7940, INRIA, 2012.

[8] J. Shotton, A. Blake, R. Cipolla. "Multiscale Categorical Object Recognition Using Contour Fragments",IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2008, pp. 1270-1218.

[9] X. Bai,et al. "Active skeleton for non-rigid object detection", International Conference on Computer Vision (ICCV), 2009 IEEE, pp. 575-582.

[10] J. Yan, X. Zhang, Z. Lei, D. Yi, and S. Li. "Structural models for face detection", in Automatic Face and Gesture Recognition (FG). IEEE, 2013, pp. 1-6.

[11] X. Zhu and D. Ramanan. "Face detection, pose estimation, and land-mark localization in the wild", in Computer Vision and Pattern Recognition (CVPR). IEEE, 2012, pp. 2879-2886.

[12] V. Jain and E. Learned-Miller. "FDDB: A benchmark for face detection in unconstrained settings", Technical report, University of Mas-sachusetts, Amherst, 2010.

[13] V. Jain and E. Learned-Miller. “Online domain adaptation of a pretrained cascade of classifiers”, in Computer Vision and Pattern Recognition (CVPR). IEEE, 2011, pp. 577-584.

[14] K. Mikolajczyk, C. Schmid, and A. Zisserman. “Human detection based on a probabilistic assembly of robust part detectors”, in European Conference on Computer Vision (ECCV), 2004, pp. 69-82.

[15] J. Yan, Z. Lei, L. Wen and S. Z. Li. “The Fastest Deformable Part Model for Object Detection”, in Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2497-2504.

[16] A. Vedaldi, A. Zisserman. “Sparse Kernel Approximations for Efficient Classification and Detection”. In Computer Vision and Pattern Recognition (CVPR), 2012, pp. 2320-2327.

[17] P. Felzenszwalb, R. Girshick, and D. McAllester. “Cascade object detection with deformable part models”. In Computer Vision and Pattern Recognition (CVPR). IEEE, 2010, pp. 2241-2248.

[18] H. Pirsiavash, D. Ramanan, and C. Fowlkes. “Bilinear classifiers for visual recognition”. In Neural Information Processing Systems (NIPS), 2009, pp. 1482-1490.

[19] H. Pirsiavash and D. Ramanan. “Steerable part models”. In Computer Vision and Pattern Recognition (CVPR), IEEE, 2012, pp. 3226–3233.

[20] Zhang, Ning, et al. “Deformable part descriptors for fine-grained recognition and attribute prediction”. In International Conference on Computer Vision (ICCV), 2013, pp. 729-736.

[21] P. Viola, M. J. Jones, and D. Snow. “Detecting pedestrians usingpatterns of motion and appearance”. International Journal of Computer Vision , 2005, 63(2):153–161.

[22] C. P. Papageorgiou, M. Oren, and T. Poggio. “A general framework for object detection”. In International Conference on Computer Vision (ICCV), 1988, pp. 555-562.