A SEARCH BASED APPROACH TO NON MAXIMUM SUPPRESSION IN FACE DETECTION

E.Zaytseva^{1,2}*, J.Vitrià*^{1,2}

ezaytseva@cvc.uab.es, jordi.vitria@ub.edu
[1] Computer Vision Center, Edifici O, Campus de la UAB, Spain
[2] Departament de Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, Spain

ABSTRACT

Face detectors typically produce a large number of false positives and this leads to the need to have a further non maximum suppression stage to eliminate multiple and spurious responses. This stage is commonly based on considering spatial heuristics: true positive responses are selected by implicitly considering several restrictions on the spatial distribution of detector responses in natural images. In this paper we analyze the limitations of this approach and propose an efficient search method to overcome them. Results show how the application of this new non-maximum suppression approach to a simple face detector boosts its performance to state of the art results.

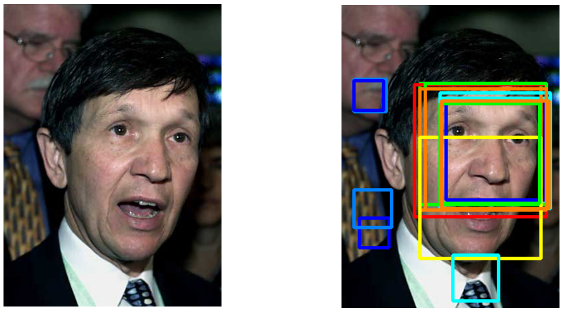
***Index Terms*—** Non maximum suppression, face detection, data driven MCMC.

1. INTRODUCTION

The goal of face detection is to localize faces in a given image and, if present, return their locations and extends. The difficulty in developing a robust face detector arises from the diversity of human faces and changes in environmental conditions as well as due to objects worn by persons that may occlude their face. However, in the computer vision community there still exists a need for improving the results of face detection ([1], [2]), especially when dealing with unconstrained real life images.

State of the art face detectors are based on two different approaches: content-based and context-based. In the first case, the visual content of all image subwindows is considered and processed by a trained classifier [3]. In the second case detectors rely on the information about the environment-object relation to infer the face position and scale in the image [4]. In spite of this basic difference, both approaches suffer from a common problem: the high number of false positives that are generated for an image. To solve this problem, several authors have proposed the use of the spatial response distribution in order to select the true positives. This process is commonly known as *non maximum suppression*. The most basic approach is to process all detection windows with an agglomerative clustering algorithm in order to eliminate spurious responses and to group those that are overlapping. In [5], Atanasoaei et al. present a more advanced model based on a hierarchical model which is built using the detection distribution around a target hypothesis to discriminate between false alarms and true detections. A different approach was proposed in [6] by Takatsuka et al. to study the score distribution in both location and scale spaces.

The main advantage of considering the spatial distribution of responses is its computational simplicity, but has two clear limitations. The first one is that the final face location is restricted to be one of the outputs of the face detector and no other location is analyzed. In figure 1 there is an example of detection distribution which shows a cluster of approximately located (in space and scale) face hypotheses. The second one is that the visual content of the response is not used in the decision but only the spatial distribution of hypotheses. For the case of low resolution detections this fact is admissible but when considering high resolution responses it is not.



(a) Original image

(b) Face detector results

Fig. 1. If final face location is restricted to be one of the outputs of the face detector, the true face hypothesis cannot be found in this image.

The main novelty of this paper is the definition of a unified statistical framework for efficient suppression of false positive responses of a face detector. This framework will allow a two-fold contribution that clearly improves results over classical non-maximum suppression methods: (i) The exact loca-

This work was partially supported by MICINN grants TIN2009-14404-C02-01 and CONSOLIDER-INGENIO 2010 (CSD2007-00018).

tion and scale of the face will not be restricted to the actual outputs of the detector but will be determined by an efficient search strategy on the image; (ii) The visual content of each detection hypothesis will be used for the non-maximum suppression stage. These improvements are possible by the use of an efficient search in the hypothesis space based on a data driven Markov Chain Monte Carlo (DDMCMC) method [7].

The structure of this paper is as follows. The next section describes the statistical model which integrates all available information about faces in an image. Section 3 describes an efficient approach to search all faces in an image. Section 4 shows some results when applying this method to the largest public domain of face databases. Finally, section 5 draws some conclusions and proposes some new research lines.

2. A BAYESIAN MODEL FOR EFFICIENT FACE SEARCHING

To determine the number of faces and their precise positions in an image, we adopt the Bayesian approach that can be divided into three main blocks.

First block is concentrated on computing a prior distribution of the faces on the image. To get this distribution, all hypotheses of where faces could be found are collected. Generally any face detector, such as the Viola and Jones face detector [3], can be used for this purpose.

Second block is all about the way to measure goodness of fit of a proposed face configuration to the real image. This measurement is done in terms of template matching techniques with a large set of high quality templates. We have learned 44 face templates, which includes frontal, lateral and intermediate positions of faces. They also include faces of people with glasses, beards, etc. Templates are coded as histograms of oriented gradiens (HoG) and are computed from a high resolution data set.

Third block defines a Data-Driven Markov Chain Monte Carlo method which iterates until a stable face distribution is found in the image. Due to the data-driven techniques to guide the Markov chain search we are able to achieve an important speed-up in comparison to classical MCMC algorithms. In our case the data term takes into account the initial distribution of the faces and the template matching results.

2.1. Computing a priori distribution of faces.

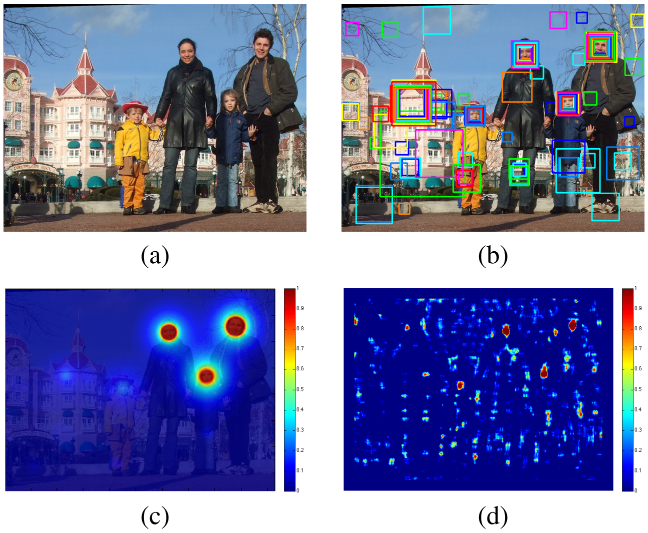
To compute the prior face distribution in an image we can use any face detector that can work at a regime level that provides a high true positive rate with a moderate false positive rate. This characteristic can be obtained by almost any commonly available face detector, such as the Viola and Jones face detector. In this case, the wanted regime can be obtained by inhibiting the implemented non maximum suppression stage, which is based on an agglomerative clustering method. Thus, by taking into consideration all image subwindows that are accepted by the detector, the face configuration prior distribution $\pi(\cdot)$ of the image can be estimated. It indicates where faces are likely to be seen, and more importantly, where not to look for them. We represent this information as prior density function $\pi(f)$ where f represents a set of faces on an image.

Given an image and the face hypotheses generated by a face detector (each one indexed by their image coordinates and scale), we can use a kernel density method to estimate the prior function. It provides a smooth probability distribution function of the spatial distribution of faces at each possible scale, as can be seen in figure 2(c). The only parameter to set up is the bandwidth of the smoothing function. This parameter is defined as a function of observation's scale and number of observations of the same scale.

Given a set of N face hypotheses in an image and K different scales, we group the hypotheses in K different scale subsets $\mathbf{X}_j \ : \quad \mathbf{X} = \cup_{j=1..K} \mathbf{X}_j$. For each group we can estimate the smoothing bandwidth as in [8]:

$$\alpha_j = 1.06 * \min(\sigma, \frac{q_{0.75} - q_{0.25}}{1.34}) * n^{-1/6}, \tag{1}$$

where σ is the sample standard deviation of the location of a member of $\mathbf{X}_j$, $q_{0.25}$, $q_{0.75}$ are their 25% and 75% sample quartiles, and n is the total amount of the samples in $\mathbf{X}_j$. As a kernel function, the Gaussian function is adopted. In figure 2(c) it is shown an example of the estimated prior face distribution for an image.



(a) (b)

(c) (d)

Fig. 2. (a) Original image, (b) Face detector results, (c) Estimated prior distribution, (d) Likelihood based on HoG detector.

2.2. Face likelihood estimation

To compute the goodness of fit a proposed configuration of faces to the image, a common way is to match a set of calibrated face templates at each proposed location of the image and then compute a probability estimate of the proposal.

We have defined face templates (models) as a variation of the Dalal-Triggs detector[9], where single filters based on histogram of oriented gradient (HOG) features are used to represent a face. These models are trained using a discriminative procedure that only requires the bounding boxes of faces in a set of images.

To measure the goodness of fit of a face to a model, it is enough to compute the dot product between a set of weights that represent a face model M and the HOG features within a window W that correspond to the current face hypothesis:

$$Score = M(x_0, y_0) \cdot W(x_0, y_0) \tag{2}$$

By using the discriminative method presented in [10] we have built 22 different face models, which includes frontal, lateral and intermediate positions of faces, and also people with glasses, beards, etc. Each model can be transformed to its symmetric version, so the total number of model filters is 44. In figure 3 several of these filters are shown.

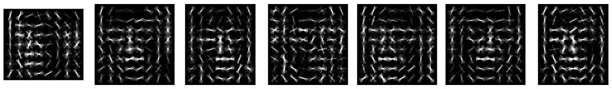


Fig. 3. Example of different models of a face. Observe that model dimensions can vary.

The final score for a face hypothesis is defined as the maximum value of the scores of different face models: $s = \max_{i=1..44} Score_i$. The higher value of the score indicates higher probability of detection of the face.

In order to be able to compare different scores in different image locations it is necessary to transform scores to probabilities. To this end we use a variation of of the Platts method proposed in [11], which approximates the posterior by a sigmoid function:

$$Pr(W = 1|M) \approx P_{A,B}(s) = \frac{1}{1 + exp(As + B)} \tag{3}$$

Based on the obtained values of A and B any score value s can be transformed to a likelihood probability. An example of the distribution of face likelihood values in an image is shown in figure 2(d).

3. A DD-MCMC METHOD FOR FACE SEARCHING

The Markov chain Monte Carlo (MCMC) method is a tool to sample high dimensional distributions that can be used for optimization. If the objective function is multimodal, MCMC presents global convergence to the maximum value because it ensures the Markov chain to reach it with a high probability. The convergence speed of MCMC can be improved by using data driven MCMC (DD-MCMC). The traditional MCMC method randomly walks through the parameter space while DD-MCMC employs some heuristics from data to guide the walks.

DD-MCMC proceeds with an iterative sampling procedure that proposes a local update to a current configuration and then decides stochastically whether or not to accept the new configuration based on the value of an acceptance ratio:

$$a(f, f') = min(1, \frac{p(f')q(f', f)}{p(f)q(f, f')}) \tag{4}$$

where f and f' is current and proposed configurations, $p(\cdot)$ is the a posteriory distribution evaluated for a given configuration and $q(f, f')$ is the probability of proposing a transition from f to f'. In our implementation we allow two types of transitions: an update of a hypothesis scale and an update of a hypothesis position. We define each iteration of the DD-MCMC as the update of only one face hypothesis.

The a posteriory distribution at each iteration is approximated by:

$$p(f) = \prod_{i=1}^{n} \pi(f_i) L(f_i) \tag{5}$$

where $\pi(\cdot)$ is the prior distribution of the proposal, $L(\cdot)$ is its likelihood value and n is the number of elements. In this way, during the MCMC iteration we must only evaluate the likelihood value of the changed element because the prior value is already estimated.

Update moves are restricted with respect to the initial parameters. As an example, an update of the scale of an hypothesis can not be changed by more than 25% of its original value.

4. EXPERIMENTS

The proposed algorithm was tested on the Face Detection Data Set [12]. Due to algorithms simplicity the proposed method works at 10 frames per second (using a non optimized program in Matlab and C).

Initial faces hypoteses for DD-MCMC are generated by a non-homogeneous Poisson point process with a prior distribution as intensity function. The length of the MCMC search is experimentally limited to 100 iterations with respect to each face hypotheses of the image. At the end of MCMC iterations, an agglomerative clustering of ”very similar” [1] face hypotheses is performed in order to generate the final hypotheses.

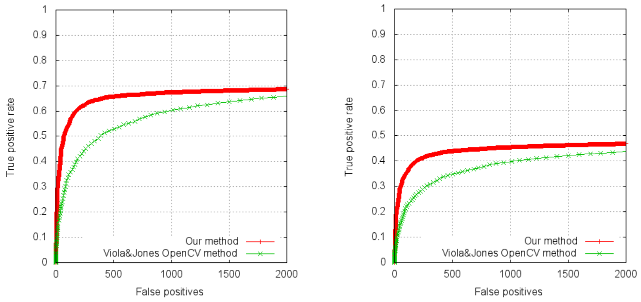
The obtained results, using the area under the ROC curve metric, for the Viola and Jones face detector can be seen in figure 4. It represents two ways of scoring the detections in an image: discrete score, and continuous score. In the first case, if the ratio of the intersection of a detected region with an annotated face region is greater than 0.5, a score of 1 is assigned to the detected region, and 0 otherwise. In the second

[1] The intersection of two face hypotheses must be greater than 95%.

case, the value ratio of the intersection of a detected region with an annotated face region is directly used as score value. Further details for the evaluation procedure can be found in the FDDB technical report [12].

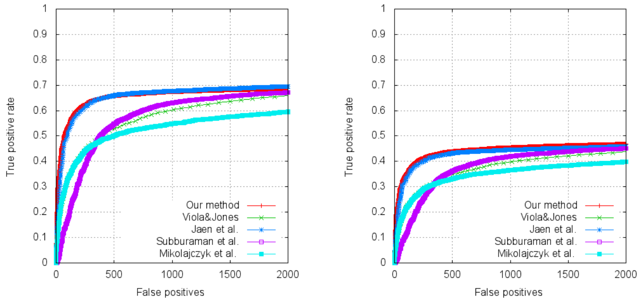
In the first exepriment, the area under the discrete ROC curve increased from 56.56% up to 65.63% while the area under the continuous ROC curve increased from 37.36% up to 44.21%.

In the figure 5 there is a comparison between the Viola and Jones algorithm that uses our non-maximum suppression method and several state of the art methods. The experiment shows that by just changing the standard non maximum suppression stage the Viola and Jones face detector can get a state of the art performance.



(a) Discrete score (b) Continuous score

Fig. 4. Performance of our method for the Face Detection Data Set. The Viola and Jones face detector, as implemented in OpenCV 2.1, is used as base detector.



(a) Discrete score (b) Continuous score

Fig. 5. Performance comparison with the state of the art methods. Reference to the methods can be find in http://vis-www.cs.umass.edu/fddb/results.html

5. CONCLUSIONS

We propose an efficient search method as an alternative to the non-maximum suppression stage in face detection. Experiments we performed confirm the efficiency of the method and its better performance when evaluated in the largest available face data set. This kind of methods could be applied not only to face detection tasks, but can easily be generalized to object detection problems, taking into account that we can define a good high resolution likelihood model. As a future work, the data-driven MCMC will be extended to a trans-dimensional version of the MCMC search procedure. This will allow dimensionality changes in face configurations and thus, the automatic estimation of the number of faces in an image.

6. REFERENCES

[1] A. Frome, G. Cheung, and et. al, “Large-scale privacy protection in google street view,” in *ICCV*, 2009.

[2] C. Zhang and Z. Zhang, “A survey of recent advances in face detection,” in *Microsoft Res. Tech. Rep., MSR-TR-2010-66*, 2010.

[3] P.A. Viola and M.J. Jones, “Rapid object detection using a boosted cascade of simple features,” in *CVPR*, 2001, pp. 511–518.

[4] S. Segui, M. Drozdzal, P. Radeva, and J. Vitrià, “An integrated approach to contextual face detection,” in *ICPRAM*, 2012.

[5] C. Atanasoaei, C. McCool, and S. Marcel, “A principled approach to remove false alarms by modelling the context of a face detector,” in *Proc. BMVC*, 2010.

[6] H. Takatsuka, M. Tanaka, and M. Okutomi, “Distribution-based face detection using calibrated boosted cascade classifier,” in *Proc. ICIAP*, 2007.

[7] Zhuowen Tu, Song-Chun Zhu, and Heung-Yeung Shum, “Image segmentation by data driven markov chain monte carlo,” in *ICCV 2001*, 2001.

[8] H. Liu, M. Xu, H. Gu, A. Gupta, J. Lafferty, and L. Wasserman, “Forest density estimation,” *J. Mach. Learn. Res.*, 2011.

[9] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *CVPR*, 2005.

[10] Pedro F. Felzenszwalb, Ross B. Girshick, and et. al, “Object detection with discriminatively trained part-based models,” *IEEE TPAMI*, 2010.

[11] Hsuan-Tien Lin, Chih-Jen Lin, and Ruby Weng, “A note on Platt’s probabilistic outputs for support vector machines,” *Machine Learning*, 2007.

[12] Vidit Jain and Erik Learned-Miller, “Fddb: A benchmark for face detection in unconstrained settings,” Tech. Rep., University of Massachusetts, Amherst, 2010.