OBJECT DETECTION USING PICTORIAL STRUCTURE OF GABOR TEMPLATE

Babak Saleh, Mohammad Rastegari
Computer Vision Group, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
saleh@cs.sharif.edu, m.rastegari@ipm.ir

Keywords: Gabor wavelet, Deformable model, Spanning tree, Dynamic programming, Pictorial Structure.

Abstract: Object detection methods are divided into two main branches: In the global approach one extracts low level features and uses machine learning techniques. In the part-based approach one uses deformable templates. We present a Hybrid approach for constructing a deformable template for modeling and detection. Initially one applies Gabor wavelet filters to extract low level features and constructs graphs which resemble shock graphs. A minimum spanning tree (MST) is extracted and is called the pictorial graph. It is used for matching. The pictorial graph is suitable for preserving the visual appearance of the shape of the object and for accommodating shape variances. In this hybrid approach we maintain the generality of the global and the efficiency of part-based approaches. Our algorithm has been applied to a set of test cases and the result shows improved performance as compared to standard object detection methods that do not rely on human intervention.

1 INTRODUCTION

There are two approaches to object detection namely the global and the part-based ones. In the Global approach to object detection features are extracted from raw images and machine learning is used to make models for particular objects (L. Zhang and Dong, ICPR; Y.N. Wu, 2009; Amira and Farrell, 2005). In the part-based approach (Ramanan and Sminchisescu, 2006; Ioffe and Forsyth, 2001; Kohandani et al., 2006; E. Yen, 2005; Lowe, 1991) an object is represented as a set of its parts, and it is represented as a graph with various parts connected by the edges of the graph.

In the global approach gabor wavelets are applied globally to extract a template from a class of objects (e.g., bicycles, cars etc.) For object detection the templates are applied to the given image. A propitious feature of the global approach is that it lends itself to learning algorithms well. A difficulty with this approach is that it is very rigid and cannot accommodate shape variances. The part-based approach improves the performance and alleviates the rigidity problem of the global method but the templates had to be designed separately and rely strongly on human intervention. Computational speed is another issue in the global approach to object detection. The graph models are usually simple trees in the part-based approach and economize significantly on the computational time as compared to the global one. Dynamic programming, belief propagation algorithms can be applied to accelerate the computation in the part-based method (Kohandani et al., 2006).

In our hybrid approach, we initially apply gabor wavelets to extract templates as in the global approach. A key new feature is that we modify the template by first assigning a graph to it and then extracting a minimum spanning tree (MST) from the graph. With the MST we can apply a method similar to that of the part-based approach for the detection of an object in a given image. It should be noted that the tree structure that was extracted from the template allows for the incorporation of shape variances in a given class and ameliorates the corresponding problem which one normally encounters in the global method. Our hybrid approach is computationally efficient since we are essentially dealing with a tree structure as in the part-based method. Furthermore the learning processes in the proposed algorithm do not suffer from the shortcomings of the part-based

method and are easily implementable as in the global approach.

In section 2 the template extraction method that we used is described and is based on (Y.N. Wu, 2009). The extraction of the MST corresponding to a given class is given in section 3. The object identification process is described in section 4 and the empirical results are presented in section 5.

2 TEMPLATE EXTRACTION

We extract a generative template consisting of small components. These small elements are special type of wavelets. Mathematically a gabor function is one of the form:

$$G(x,y) \propto e^{\{-[(\frac{x}{\sigma_x})^2+(\frac{y}{\sigma_y})^2]\}} e^{ix} \qquad (1)$$

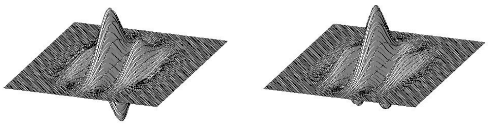
The general form of gabor wavelets are obtained by the translation, rotation and dilation of G; In fact representing a rotation and translation in the form:

$$\begin{aligned} \tilde{x} &= (x'-x)\cos\alpha-(y'-y)\sin\alpha, \\ \tilde{y} &= (x'-x)\sin\alpha-(y'-y)\cos\alpha. \end{aligned}$$

Then the general gabor wavelet with (x,y) as the central position, s the scale parameter and orientation α is given by:

$$B_{(x,y,s,\alpha)} = \frac{G(\tilde{x}/s,\tilde{y}/s)}{s^2}$$

In Fig. 1 an artificial image and its gabor transform are shown.



(a) (b)

Figure 1: An artificial image and its gabor transform. (a)An artificial input image. (b)Output field, resulted by applying gabor function.

For template extraction we first fix a category consisting of one class of objects, e.g. passenger cars. The samples have the same parameters such as pose, size, orientation, etc. Then we apply gabor filters in every possible combinations of location and orientation with a fixed scale on the training image set. The output is a set of coefficients which are computed in the familiar way as an inner product. A template is selected by a voting algorithm, more precisely known as the Shred Sketch Algorithm, applied to the outputs of gabor filters as in (Y.N. Wu, 2009). A selected output is referred to as a component of a generative template.

3 EXTRACTING MST STRUCTURE

The template graph is a weighted complete graph with nodes (or vertices) corresponding to the wavelet coefficients. The weight assigned to an edge is the "Euclidean distance" between the centers of the wavelets whose coefficients correspond to the vertices of the edge. This is a very complex graph we extract a minimum spanning tree (MST) by which we mean a spanning tree such that the sum of the weights is a minimum. In general specifying an MST has $O(n^2)$ complexity. To reduce the amount of calculation we consider the Delaunay triangulation of the plane with the given nodes. This is no longer a complete graph and we retain the weights for the surviving edges. By a standard method (Kruskal, 1956) we extract an MST from this graph as shown Fig. 2. The complexity of this operation is $O(n\log n)$.

Having modified the active template structure to that of an MST we tested the algorithm on several shape categories to ensure that it provides a faithful representation of the class of objects in the category. Five examples are shown in Fig. 2. Notice that the MST's are "good" representations of the objects.

4 MATCHING ALGORITHM

In this section we are going to explain how we can find the best matching of the extracted template (or MST) to an object in a given image. In order to find this best matching we will modify a well-known method (P.F. Felzenszwalb, 2000) for the efficient matching of pictorial structures. A pictorial structure (Fischler and Elschlager, 1986) is a template given by a number of parts with connections between them (Fig. 3(a)). This connection is elastic like a spring and shrinks or dilates according to the movements of the parts (Fig. 3(b)). This form of connection gives the parts the capability to change their location with respect to each other in a deformation in a smooth manner.

To quantify the "goodness" of a match we introduce several quantities. Let l_i be the 4-vector describing the location, scaling and orientation of a wavelet coefficient. Given an image I we want to define a cost function (Global Matching Function) for a given pictorial structure which is a sum of both the cost for fitting this pictorial structure to an object in the image and a deformation cost with respect to parts' positions, their distances, rotations, and scales (Felzenszwalb and Huttenlocher, 2005). In other words the cost function measures both how well each element

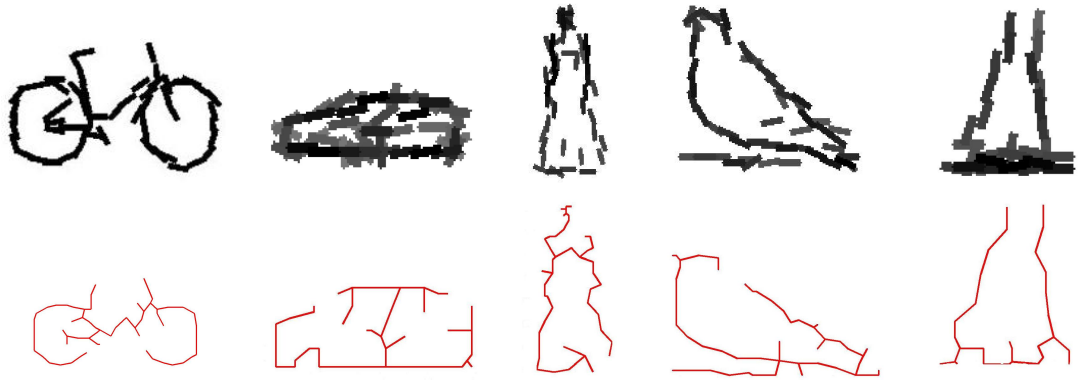
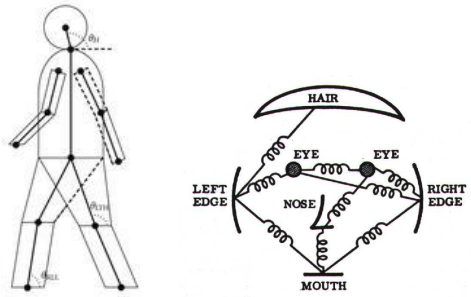


Figure 2: Extracted Templates and the corresponding MST. The first row shows input templates consisting of gabor wavelets. The second row is the result of applying MST on these templates



(a) (b)

Figure 3: (a) Pictorial structure of Human body, different parts and their connection are illustrated this picture designed by (Borgefors, 1984) (b) Spring connection in a Pictorial Structure. It shows that in pictorial structure of a face different parts can slightly change their positions. This picture designed by (Felzenszwalb and Huttenlocher, 2005)

fits the image data and how well the relative positions of the elements agree with the model. Mathematically the global matching function or cost on an image I is expressed as

$$\mathfrak{L}(I) = \min\Big(\sum_{(v_i,v_j)\in E} d_{ij}(l_i,l_j) + \sum_{v_i\in V} m_i(I,l_i)\Big), \quad (2)$$

where E is the edge set of the MST, $d_{ij}(l_i,l_j)$ is the Euclidean distance between l_i's, $m_i(I,l_i)$ represents the distance between the wavelet l_i and the wavelet transform of the corresponding point in the image I, and the min is taken over all possible matching.

The naive exhaustive search computation of $\mathfrak{L}$ is very inefficient and its complexity is of the order $O(h^4)$, where h is the number of possible location. In (P.F. Felzenszwalb, 2000) the authors showed that if the given pictorial structure is a tree, then one can apply a dynamic programming algorithm which reduces the complexity to as low as $O(h)$. The tree structure helps represent the deformation cost of each element as a function of its children's costs. This method results in a recursive algorithm for the minimization problem as follow: For every leaf node of this tree it is sufficient to find the best location which has the minimum matching cost. For other nodes one finds the best place constrained by the children's locations which have been determined in the preceding iteration. This recursive algorithm is summarized by the equation 3

$$B_j(l_i) = \min_{l_j}\Big(d_{ij}(l_i,l_j) + m_j(I,l_j) + \sum_{v_c\in C_j} B_c(l_j)\Big) \quad (3)$$

where C_j is the set of children of j, $B_c(l_j)$ is the best location for the children of j (represented by the v_j), which $B_j(l_i)$ is the best location of the j^{th} part given its parent location l_i. This recursive algorithm reduces the complexity to $O(h^2)$. To achieve a lower complexity we make use of the generalized distance transform (GDT) (P.F. Felzenszwalb, 2004) on the tree. GDT is a weighted version of usual distance transform (Borgefors, 1984; Borgefors, 1986). We used expanded GDT in four dimensions: (x,y) -components of the center of the gabor, s-scale parameter and r-orientation. Using the naive approach finding the best configuration has complexity $O(h^n)$, and the dynamic programming approach (A. Amini, 1990) reduces it to $\boldsymbol{O}(h^2 n)$, where h is the number of possible locations for each part and n is the number of parts. In our application h is large (and n is relatively small) which makes the computation infeasible. But state of the art generalized distance transform reduces

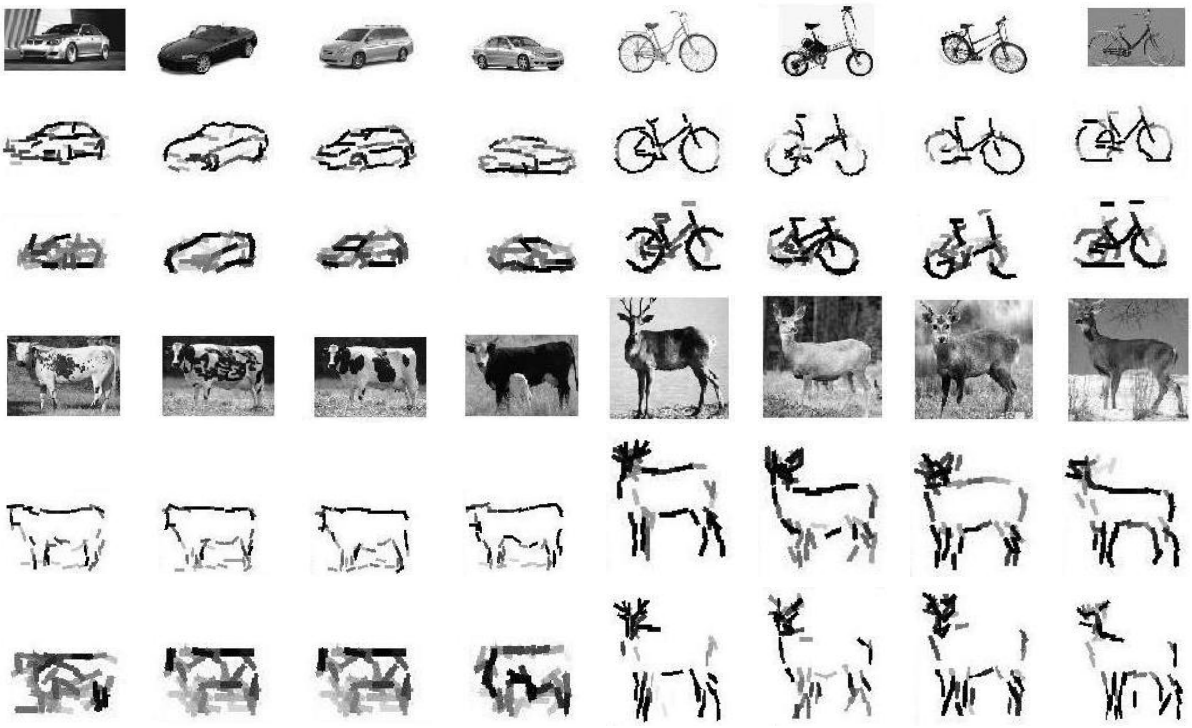


Figure 4: Qualitative results. The first and fourth rows show input test images for different categories, taken from (Y.N. Wu, 2009), the second and fifth rows are the results of the proposed method and the third and sixth rows show the results attained by (Y.N. Wu, 2009)

the complexity to $O(hn)$ or $O(h)$ since n is a small number independent of h.

5 RESULTS

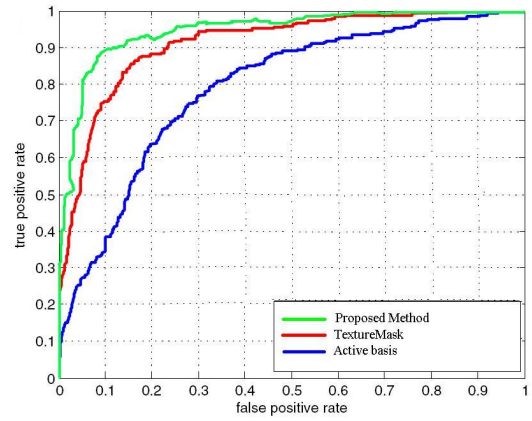
In this section we present some experimental results of our approach and a comparison is made with the Active Basis methods (Y.N. Wu, 2009). We implemented our method with MATLAB on a PC with 1.8 GHz CPU and 512 MB RAM and selected 60 gabor wavelets for the template. We used gabor bases at 8 orientations and 10 scales. Our experiments were carried out on the database provided by (Y.N. Wu, 2009). Fig. 4 shows a qualitative comparison where the results of four test cases obtained by the proposed and the Active Basis (Y.N. Wu, 2009) methods are shown. As one can see our method (the second and fifth rows) has qualitatively superior performance relative to the Active Basis one (the third and sixth rows). Fig. 5 demonstrates a quantitative comparison. The RoC curves of the two test cases for three methods are shown. The RoC curve based on the proposed method showed significant improvement relative to Active Basis and the Texture Mask methods (Y. N. Wu and Zhu, 2008). We did not compare with the part-based method because the latter approach requires human intervention.

CONCLUSION

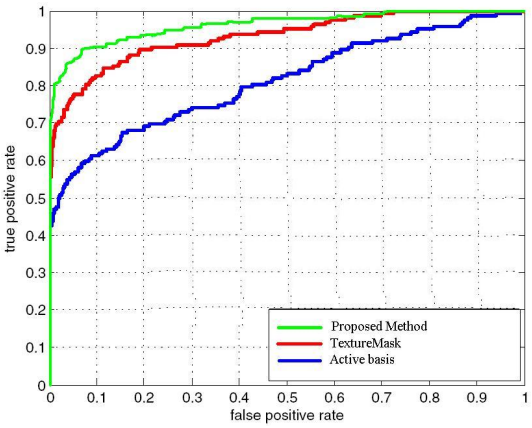
A new method for object detection was presented. The proposed algorithm is a hybrid method that combines a generative template similar to the global approach to object detection and a fast algorithm for fitting this template into images in order to achieve object detection. This matching algorithm is based on dynamic programming algorithm.

The novelty and advantages of the proposed method are

1. It has the flexibility of accommodating visual variances as in the part-based approach.
2. It lends itself easily to learning algorithms as in the global approach.
3. It does not require human intervention as in the part-based approach.



(a)



(b)

Figure 5: (a) RoC curves for Car samples, in the front viewpoint,(b) RoC curves correspondent to Motorbike samples

4. Its computational complexity is low as in the part-based approach.

ACKNOWLEDGEMENT

The authors would like to thank Prof. Mehrdad Mirshams Shahshahani for his great support, valuable discussions and supervision.

REFERENCES

A. Amini, T. Weymouth, R. J. (1990). Using dynamic programming for solving variational problems in vision. In *Vol. 12, No. 9, pp. 855-867*. PAMI.

Amira, A. and Farrell, P. (2005). An automatic face recognition system based on wavelet transforms. In *ISCAS (6)*, pages 6252–6255.

Borgefors, G. (1984). Distance transformations in arbitrary dimensions. In *Vol. 27, No. 3, pp. 321-345*. CVGIP.

Borgefors, G. (1986). Distance transformations in digital images. In *Vol. 34, No. 3, pp. 344-371*. CVGIP.

E. Yen, A. S. (2005). Image recognition via deformable template. statistical methodology. In *pp. 213-225*. Statistical Methodology.

Felzenszwalb, P. F. and Huttenlocher, D. P. (2005). Pictorial structures for object recognition. *International Journal of Computer Vision*, 61(1):55–79.

Fischler, M. and Elschlager, R. (1986). The representation and matching of pictorial structures. In *Vol. 22, No. 1, pp. 67-92*. IEEE Trans. On Computers.

Ioffe, S. and Forsyth, D. A. (2001). Mixtures of trees for object recognition. In *CVPR (2)*, pages 180–185.

Kohandani, A., Basir, O. A., and Kamel, M. S. (2006). A fast algorithm for template matching. In *ICIAR (2)*, pages 398–409.

Kruskal, J. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. The American Mathematical Society.

L. Zhang, H. Gong, T. W. and Dong, J. (ICPR). Deformable template combining alignable and non-alignable sketches. 2008.

Lowe, D. G. (1991). Fitting parameterized three-dimensional models to images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(5):441–450.

P.F. Felzenszwalb, D. H. (2000). Efficient matching of pictorial structures. IEEE Conference on Computer Vision and Pattern Recognition.

P.F. Felzenszwalb, D. H. (2004). Distance transforms of sampled functions. Cornell Computing and Information Science Technical Report TR2004-1963.

Ramanan, D. and Sminchisescu, C. (2006). In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 206–213.

Y. N. Wu, Z. Si, H. G. and Zhu, S.-C. (2008). Active basis model, shared sketch algorithm, and sum-max maps.

Y.N. Wu, Z. Si, H. G. S. Z. (2009). Learning active basis model for object detection and recognition. IJCV.