

Pedestrian Detection by Modeling Local Convex Shape Features

Jungme Park

*Department of Electrical and
Computer Engineering
University of Michigan-Dearborn
MI 48128, USA*

Yun Luo

*Driver Assistance
Group, Robert Bosch,
Plymouth, MI 48170,
USA*

Haoxing Wang, Yi L. Murphey

*Department of Electrical and
Computer Engineering
University of Michigan-Dearborn
MI 48128, USA
yilu@umich.edu*

Abstract

This paper presents a pedestrian model built collectively on a group of strong local convex shape descriptors. The pedestrian model captures the most important features of a pedestrian: head, body contour, arms, legs and crotch, and is robust to variances in appearances and partial occlusions. For an image set of 2571 pedestrians and 4369 car and background images, the pedestrian recognition system, which was built upon the proposed pedestrian model, gave a recognition rate of 98.8% with a false positive rate of 1.56%. Furthermore, the pedestrian recognition requires a very small set of prototypes of pedestrians and non-pedestrians.

1. Introduction

Research in pedestrian detection has been very active in the recent years. It is largely driven by many emerging applications such as intelligent vehicles, perceptual interfaces, image indexing, and smart video surveillance. A variety of global and local features as exemplified by PCA coefficients, Haar wavelets, local receptive fields and contours, have been used in pedestrian detection and classification [1, 2, 3, 4]. Munder & Gavrilla presented an in-depth experimental study on pedestrian classification problem [1] and they found that global features represented by PCA coefficients were inferior to local features. Among the latter, the adaptive features outperformed non-adaptive ones. Furthermore, Munder & Gavrilla found that the best pedestrian recognition system obtained an overall performance of 5% false positives at 90% detection rate, which is still far apart from the performance needed for most real-world applications [1].

In the paper we present a pedestrian model built collectively upon a group of strong local convex shape descriptors. The proposed pedestrian model is robust to imperfect edge images, pedestrians with extraneous

objects, and partially occluded pedestrians in real world images. A pedestrian recognition system is built on this pedestrian model. We tested our system on a data set that contains 2571 pedestrian images and 4369 non-pedestrian images, the performance of the system that uses the proposed pedestrian model is 98.8% recognition rate for a false positive rate of 1.56%. More interestingly, the proposed pedestrian model works well even when the pedestrians are partially occluded. Furthermore the pedestrian recognition system requires a very small number of prototypes, 8 pedestrians and 14 non-pedestrian images.

2. A pedestrian model built upon local convex shapes

Local shapes are considered the most discriminant features shared by different appearances of the pedestrian, man and women with different clothes, short, tall, close to or far away from the camera [5]. Our research focuses on the investigation of local shape features that can sufficiently capture the most important parts of pedestrians such as head, body contour, arms, legs and crotch. We developed a local convex feature representation and detection algorithm that measures the extent to which the detected object contours supporting circle or arc-like local structures at the most promising positions and scales in a given edge image. The algorithm consists of two major components: detecting convex shapes and generating shape context descriptor to capture the distribution of contour points of the detected local convex shape.

2.1 Detecting local convex shapes

We define a local convex shape by the center and the scale of a circle that best approximates the shape contour. A two-stage process has been developed to detect strong convex shapes in an edge image. At the first stage we use the Hough Transformation method

[6] to detect circles. Since Hough Transform is a voting process, noises and discontinuities in edges do not seriously affect the results of the detection. The computation is based on the fact that the gradient lines of the points on the same circle intersect at the center of the circle (see Figure 1). For every gradient direction, a line is drawn and an accumulator matrix is used to count the number of the gradient lines crossing each image grid (x, y) . The grids that have the most counts are considered the centers of circles that represent the strong convex shapes. Figure 1 illustrates the relationship between a local convex shape, edge pixels, the gradient directions, the radius (scale) and the center (C_x, C_y) of the circle. The centers of all strong convex shapes are obtained through a predetermined threshold T such as T being the number of intersected lines, or the top T centers to be selected.

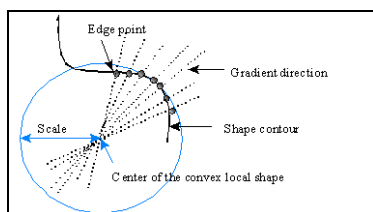


Figure 1. Illustration of a local convex shape.

The second stage involves the calculation of the scale or radius r for each strong convex shape found at the first stage. At each convex shape center (C_x, C_y) , we calculate the distances of all edges to the center. The distance that has the most number of edge points to (C_x, C_y) is the chosen scale.

At the end of this process M strong convex shapes are detected and represented by (C_1, \dots, C_M) , where C_i is a vector of two parameters, the center, c_i , and scale σ_i , to represent a convex shape. The convex shapes are ranked in a decreasing order by the following quantitative measure: $f(C_i) = E_i / \rho_i$, where E_i is the number of edge points falling on circle defined by (c_i, σ_i) and ρ_i is the circumference of C_i . The number of M can be determined based on an empirical study.

Figure 2 shows three pedestrian edge images and the top 10 (i.e $M=10$) strong local convex shapes detected by the above process and superimposed on the original images. Note one pedestrian is carrying a brief case and two are carrying backpacks. Collectively the top 10 local convex shapes captured the most important features of each pedestrian: head, body contour, arms, legs and crotch.



Figure 2. Edge images of pedestrian and the top 10 circles detected from the edge images.

2.2. Pedestrian modeling using Local Shape Context Descriptor

The local convex shapes detected using the two-stage process described in section 2.1 are represented by the circularly supported shape context descriptors that are similar to what introduced in [3, 5]. Each local convex shape is represented by a shape context descriptor designed to capture the distribution of its contour points. For each local convex shape, $C=(c, \sigma)$, we generate 4 concentric circles with radii equal to $0.6\sigma, 2\sigma, 4\sigma, 8\sigma$. The circular space is then decomposed into sectors evenly distributed in 32 orientations. This results in 128 bins in the log-polar space. The shape context descriptor is the distribution of the local density of edge pixels over the 128 bins.

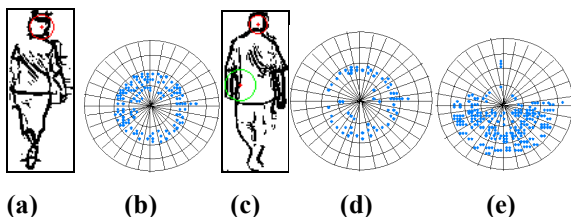


Figure 3. Local convex shape represented in log-polar space. (b), (d) and (e) illustrate the edge distribution in radial-polar bins for local convex shapes represented by the red circle in (a), the red circle and the green circle in (c) respectively.

We use the examples in Figure 3 to illustrate the effectiveness of the shape context descriptor. The three circles shown in Figure 3 (a) and (c) have their shape context descriptors shown (b), (d) and (e) respectively. Figure 3 (b) is the shape context descriptor of the local convex shape outlined by the red circle (head area) in Figure 3 (a). Figure 3 (d) is the shape context descriptor of the local convex shape outlined by the red circle (head area) in Figure 3 (c). The edge distributions in the radial-polar bins in Figure 3 (b) and (d) representing the contour areas of two different pedestrians' heads are similar in comparison to the edge distribution in the radial-polar bins(see Figure 3

(e) that represents the body part marked by the green circle in Figure 3 (c).

A local convex shape context descriptor is formally represented by a histogram of 128 bins with $Bin(h)$ being the number of edge points within the h^{th} bin in the radial-polar space. The similarity between two local shape descriptors is calculated as follows. Let $Bin_p(1..n)$ and $Bin_q(1..n)$, $n=128$ be the radial-polar histograms of two local convex shapes P and Q respectively. We first normalize two histogram functions as: for $k=1, \dots, 128$,

$$D_p(k) = \frac{Bin_p(k)}{\sum_{k=1}^{128} Bin_p(k)}, D_q(k) = \frac{Bin_q(k)}{\sum_{k=1}^{128} Bin_q(k)}.$$

The distance between the two local shape context descriptors is measured by the χ^2 function [7]:

$$\chi^2(D_p, D_q) = \frac{1}{2} \sum_{k=1}^{128} \frac{[D_p(k) - D_q(k)]^2}{D_p(k) + D_q(k)}.$$

The χ^2 function has the range of 0 through 1. A small value of χ^2 implies that two shape descriptors are similar, therefore they are more likely to be in the same class. For example, the similarity of Figure 3 (b) and (d) is $\chi^2 = 0.222$, and the similarity of Figure 3 (b) and (e) is $\chi^2 = 0.473$, and $\chi^2 = 0.472367$ between Figure 3 (d) and (e). Therefore the local convex shape shown by the circle in Figure 3 (a) is more similar to the one shown in the red circle than the green circle in Figure 3 (c).

3 Pedestrian recognition using local convex shape descriptors

Pedestrian recognition is carried out by matching the local convex shape descriptors of an input image with the pedestrian prototypes and non-pedestrian prototypes in our pedestrian recognition system. The matching function is non-linear, it takes into account the relative location, orientation and context descriptor of each local convex shape. Let us denote L_{Aj} as the j^{th} Local Convex Shape (LCS) of an image A and $L_{Aj} = (c_{Aj}, r_{Aj}, D_{Aj})$, where c_{Aj} is the location of the LCS, r_{Aj} is the respective scale, D_{Aj} is the local convex shape context descriptors explained in section 2.2.

Assume an input image X is matched with a prototype object image Γ , where Γ and X are represented by their respective sequence of LCS, $\Gamma_{LCS} = \{L_{\Gamma 1}, \dots, L_{\Gamma k}\}$ and $X_{LCS} = \{L_{X1}, \dots, L_{Xm}\}$. The matching procedure is described as follows.

For every LCS, $L_{Tj} = (c_{Tj}, r_{Tj}, D_{Tj})$ in prototype image Γ_{LCS} , $j = 1, \dots, k$, we search for a possible matching

LCS in input image X_{LCS} that its location is close to c_{Tj} . A $\omega \times \omega$ search window, $\omega(c_{Tj})$ is placed in the input image space centered at c_{Tj} . If a local convex shape of X , $L_{Xi} = (c_{Xi}, r_{Xi}, D_{Xi})$ has its center c_{Xi} within the search window, $\omega(c_{Tj})$ and the orientation of the vector $(c_{Xi} - O_X)$ is similar to the orientation of the vector $(c_{Tj} - O_\Gamma)$, then $L_{Xi} = (c_{Xi}, r_{Xi}, D_{Xi})$ is a candidate for matching with $L_{Tj} = (c_{Tj}, r_{Tj}, D_{Tj})$, where O_X is the center of input image X , and O_Γ is the center of prototype image Γ . Let us denote such a candidate set as $\xi_j = \{L_{Xj1}, \dots, L_{Xjh}\} \subseteq X_{LCS}$ with respect to the local convex shape, $L_{Tj} = (c_{Tj}, r_{Tj}, D_{Tj})$ in prototype Γ_{LCS} . The similarity between $L_{Xjl} = (c_{Xjl}, r_{Xjl}, D_{Xjl}) \in \xi_j$ and $L_{Tj} = (c_{Tj}, r_{Tj}, D_{Tj})$ is measured by the χ^2 distance. If $\{\chi^2(D_{Xjl}, D_{Tj}) < th \mid L_{Xjl} = (c_{Xjl}, r_{Xjl}, D_{Xjl}) \in \xi_j\} \neq \emptyset$, then there is a match from X_{LCS} to L_{Tj} , where th is a pre-set threshold. The matching score between X and Γ is measured by the number of local convex shapes in Γ_{LCS} that have matched local convex shapes in X_{LCS} . The input image X is classified as a pedestrian if y is positive, $y = \text{sign}(\sum_{i=1}^p F(X_{LCS}, \Gamma_{LCS}^i))$, where Γ_{LCS}^i is the top i^{th} matched prototype based on the matching score, p is a odd number and $F(X_{LCS}, \Gamma_{LCS}^i) = 1$ if Γ^i is a pedestrian prototype otherwise $F(X_{LCS}, \Gamma_{LCS}^i) = -1$.

4. Experiments and Conclusion

We have implemented a pedestrian recognition system based on the algorithms presented in section 2 and 3. We also generated a prototype group of 8 pedestrian images and 14 non-pedestrian images. 4 prototype pedestrian images were selected manually from a set of 197 pedestrian images (UMD_images) we acquired in urban areas using a camera system mounted on a moving vehicle, 2 prototype pedestrian images from the MIT-Label-Me [8], 2 prototype pedestrian and 5 prototype background images from INRIA [9], and 9 prototype car images from MIT-CBCL car data site [10]. The 8 prototype pedestrian images are used as the positive class, and the prototypes of cars and background images are used as the negative class.

Two experiments have been conducted. First experiment was conducted on the image data set that contains 2571 pedestrian images and 4369 non-pedestrian images collected from the following sources: 924 pedestrian images from the MIT-CBCL

pedestrian data [11], 308 pedestrian and 293 car images from the MIT-Label-Me [8], 1142 pedestrian and 3560 background images from INRIA [9], 197 pedestrian images from UMD_images, and 516 car images from the MIT-CBCL car data [10]. The images are in a broad range of background, resolution and contrast. The pedestrians in these images are mingled with other objects such as vehicles, buildings, trees and roads; some pedestrians are carrying briefcases, backpacks, purses or shopping bags. The images are first being processed by the Histogram Equalization to reduce the sensitivity to low contrast and changes in illumination conditions followed by the Sobel edge detection. The local convex descriptors are generated by the algorithm described in section 2.1 and 2.2. The local convex descriptor of each image is matched with every prototype image. The top 5 matched prototypes are used in the voting process to generate the ROC curve: if the number of votes for pedestrian among the top 5 matched prototype is equal to or greater than i , $i = 1\sim 5$, the input image is a pedestrian. The system performances on the entire set of 6940 images for $i = 1\sim 5$ are used to generate the ROC curve shown in Figure 4.

For the false positive rate of 0, the system detected 74.5% of the pedestrian images, and for a false positive rate of 1.56% the recognition rate on the entire pedestrian set is 98.8%.

In order to perform a quantitative study on the robustness of our pedestrian, we applied 10%, 20%, 30%, 40% and 50% occlusion to all 308 pedestrian images from the MIT-Label-Me set in four directions, top, left, right and bottom. We applied the pedestrian detection system to these occluded images and the results are shown in Figure 5. The recognition rates are stable for occlusions up to 20% in all four directions. The system recognition rate is still above 90% with 30% occlusions at the top, left and right. This shows that the pedestrian recognition system is robust to partial occlusions.

We have presented a pedestrian model built upon strong local convex shapes. A pedestrian recognition system has been developed based on the pedestrian model. We conducted experiments on a wide range of pedestrian and non-pedestrian images. The results show that the pedestrian recognition system gives high recognition rates with low false positive rates, and is robust to partial occlusions.

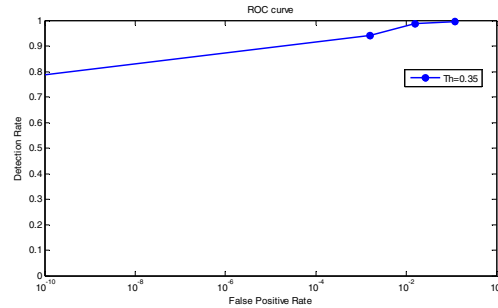


Figure 4. The ROC curve performance.

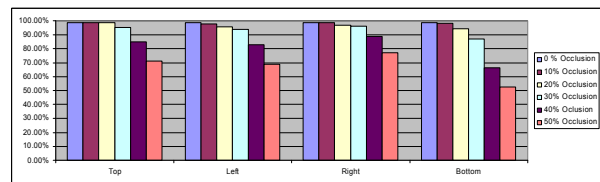


Figure 5. Performance analysis of occluded pedestrian images.

Acknowledgment

This research is supported in part by a grant from the TRW Education Foundation.

References

- [1] S. Munder and D. M. Gavrila, "An Experimental Study on Pedestrian Classification," IEEE PAMI, vol.28, pp. 1863-1868, 2006.
- [2] P. Viola, M. Jones, and D. Snow, "Detecting Pedestrians Using Patterns of Motion and Appearance," Proc. Int'l Conf. Computer Vision, pp. 734-741, 2003.
- [3] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," IEEE PAMI 24, pp. 509-522, 2002.
- [4] Ying Wu and Ting Yu, "A Field Model for Human Detection and Tracking," IEEE PAMI, vol. 28, May 2006.
- [5] F. Jurie and C. Schmid, "Scale-invariant shape features for recognition of object categories," In Proc. IEEE Conf. Comp. Vision Pattern Recognition, 2004.
- [6] E. R. Davies, "A modified Hough scheme for general circle location," Pattern Recognition Letter, Vol. 7, pp.37-43, 1988.
- [7] H. Zhang and J. Malik, "Learning a discriminative classifier using shape context distances", CVPR 2003, vol11, pp 242-247.
- [8] MIT-Label-Me (<http://people.csail.mit.edu/torralba/research/LabelMe/js/LabelMeQueryObjectFast.cgi>).
- [9] INRIA Dataset (<http://pascal.inrialpes.fr/data>).
- [10] MIT-CBCL Car Data (<http://cbcl.mit.edu/software-datasets/CarData.html>).
- [11] MIT-CBCL Pedestrian Data (<http://cbcl.mit.edu/software-datasets/PedestrianData.html>).