Appears in CVPR 97, June 17-19, Puerto Rico

Pedestrian Detection Using Wavelet Templates

Michael Oren Constantine Papageorgiou Pawan Sinha
Edgar Osuna Tomaso Poggio

CBCL and AI Lab
MIT
Cambridge, MA 02139

Abstract

This paper presents a trainable object detection architecture that is applied to detecting people in static images of cluttered scenes. This problem poses several challenges. People are highly non-rigid objects with a high degree of variability in size, shape, color, and texture. Unlike previous approaches, this system learns from examples and does not rely on any a priori (hand-crafted) models or on motion.

The detection technique is based on the novel idea of the wavelet template *that defines the shape of an object in terms of a subset of the wavelet coefficients of the image. It is invariant to changes in color and texture and can be used to robustly define a rich and complex class of objects such as people. We show how the invariant properties and computational efficiency of the wavelet template make it an effective tool for object detection.*

1 Introduction

The problem of object detection has seen a high degree of interest over the years. The fundamental problem is how to characterize an object class. In contrast to the case of pattern classification, where we need to decide between a relatively small number of classes, the detection problem requires us to differentiate between the object class and the rest of the world. As a result, the class description for object detection must have large discriminative power to handle the cluttered scenes it will be presented with. Furthermore, in modeling complicated classes of objects (e.g. faces, pedestrians) the intra-class variability itself is significant and difficult to model. Since it is not known how many instances of the class are presented in the scene, if any, the detection problem cannot easily be solved using methods such as maximum-a-posteriori probability (MAP) or maximum likelihood models. Consequently, the classification of each pattern in the image must be done independently; this makes the decision problem susceptible to missed instances of the class and false positives.

There has been a body of work on people detection (Tsukiyama & Shirai, 1985[16], Leung & Yang, 1987[6][5], Rohr, 1993[10], Chen & Shirai, 1994[2]); these approaches are heavily based on motion and hand crafted models. An important aspect of our system is that the model is automatically learned from examples and avoids the use of motion and explicit segmentation.

One of the successful systems in the area of trainable object detection in cluttered scenes is the face detection system of Sung and Poggio [15]. They model face and non-face patterns in a high dimensional space and derive a statistical model for the class of frontal human faces. Similar face detection systems have been developed by others (Vaillant, et al.[17], Rowley, et al.[11], Moghaddam and A. Pentland[8], Osuna et al.[3]).

Frontal human faces, despite their variability, share very similar patterns (shape and the spatial layout of facial features) and their color space is very constrained. This is not the case with pedestrians. Figure 1 shows several typical images of people in our database. These images illustrate the difficulties of pedestrian detection; there is significant variability in the patterns and colors within the boundaries of the body. The detection problem is also complicated by the absence of constraints on the image background. Given these problems, direct analysis of pixel characteristics (e.g. intensity, color and texture) is not adequate. This paper presents a new approach motivated by an earlier piece of work by one of the authors [12] [13] who derived a new invariant called the 'ratio template' and applied it to face detection.

A ratio template encodes the ordinal structure of the brightness distribution on a face. It consists of a set of inequality relationships between the average intensities of a few different face-regions. This design was motivated by the observation that while the absolute intensity values of different regions change dramatically under varying illumination conditions, their mutual ordinal relationships (binarized ratios) remain largely unaffected. Thus, for instance, the forehead is typically brighter than the eye-socket regions for all but the most contrived lighting setups. A small set of such relationships, collectively called a ratio template, provides a powerful constraint for face detection. The emphasis on the use of qualitative relationships also renders the ratio template construct perceptually plausible (the human visual system is poor at judging absolute brightnesses but remarkably adept at making ordinal brightness comparisons). In [13] a scheme for learning such relationships from examples was presented and tested on synthetic images. However, this

[0] Contact authors' email: {oren,cpapa,tp}@ai.mit.edu

Figure 1: Examples of images of people in the training database. The examples vary in color, texture, view point (either frontal or rear) and background.

work left some important issues open. These include a formalization of the template structure in terms of simple primitives, a rigorous learning scheme capable of working with real images, and also the question of applicability to other, possibly more complex, object classes such as pedestrians.

We present an extension of the ratio template, called the *"wavelet template"*, and address some of these issues in the context of pedestrian detection. The wavelet template consists of a set of regular regions of different scales that correspond to the support of a subset of significant wavelet functions. The relationships between different regions are expressed as constraints on the values of the wavelet coefficients. The wavelet template can compactly express the structural commonality of a class of objects and is computationally efficient. We show that it is learnable from a set of examples and provides an effective tool for the challenging problem of detecting pedestrians in cluttered scenes. We believe that the learnable wavelet template represents a framework that is extensible to the detection of complex object classes other than pedestrians.

2 The wavelet template

In this section, we review the Haar wavelet, describe a denser (redundant) transform, and define the wavelet template.

2.1 The Haar dictionary

In this section, for lack of space, we survey the properties of wavelets which are used in the paper; a more detailed treatment can be found in [7] and other standards references on wavelets. As motivated by the work on the template ratio, we were looking for an image representation which captures the relationship between average intensities of neighboring regions. This suggests the use of a family of basis functions, such as the Haar wavelets, which encode such relationships along different orientations. The Haar wavelet representation has also been used for image database retrieval, Jacobs *et al.*[4], where the largest wavelet coefficients were used as a measure of similarity between two images. In our work, the wavelet representation is used to capture the structural similarities between various instances of the class. In Figure 2, we depict the 3 types of 2-dimensional Haar wavelets. These types include basis functions which capture change in intensity along the horizontal direction, the vertical direction and the diagonals (or corners). Since the wavelets that the standard transform generates have irregular support, we use the non-standard 2-dimensional DWT where, at a given scale, the transform is applied to each dimension sequentially before proceeding to the next scale [14]. The results are Haar wavelets with square support at all scales.

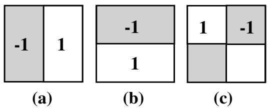

Figure 2: The 3 types of 2-dimensional non-standard Haar wavelets; (a) "vertical", (b) "horizontal", (c) "corner".

The standard Haar basis is not dense enough for our application. For the 1-dimensional transform, the distance between two neighboring wavelets at level n (with support of size 2^n) is 2^n. For better spatial resolution, we need a set of redundant basis functions, or an overcomplete *dictionary*, where the distance between the wavelets at scale n is $\frac{1}{4}2^n$. We call this a *quadruple density* dictionary. As one can easily observe, the straightforward approach of shifting the signal and recomputing the DWT will *not* generate the desired dense sampling. However, one can observe that in the standard wavelet transform, after the scaling and wavelet coefficients are convolved with the corresponding filters there is a step of downsampling. If we do not downsample the wavelet coefficients we generate wavelets with *double density*, where wavelets of level n are centered every $\frac{1}{2}2^n$. To generate the quadruple density dictionary, we compute the scaling coefficients with double density by not downsampling them. The next step is to calculate double density wavelet coefficients on the two sets of scaling coefficients — even and odd — separately. By interleaving the results of the two transforms we get quadruple density wavelet coefficients. For the next scale we keep only the even scaling coefficients of the previous level and repeat the quadruple transform on this set only; the odd scaling coefficients are dropped off. Since only the even coefficients are carried along at all the scales, we avoid an "explosion" in the number of coefficients, yet provide a dense and uniform sampling of the wavelet coefficients at all the scales. As with the regular DWT, the time complexity is $O(n)$ in the number of pixels n. The extension for the 2-dimensional transform is straightforward.

2.2 The wavelet template

The ratio template defines a set of constraints on the appearance of an object by defining a set of regions and a set of relationships on their average intensities. The relationships can require, for example, that the ratio of intensities between two specific regions falls within a certain range. We address the issues of learning these relationships, using the template for detection, and its efficient computation by establishing the ratio template in the natural framework of Haar wavelets. Each wavelet coefficient describes the relationship between the average intensities of two neighboring regions. If we compute the transform on the image intensities, the Haar coefficients specify the intensity differences between the regions; computing the transform on the log of the image intensities produces coefficients that represent the log of the ratio of the intensities. Furthermore, the wavelet template can describe regions with different shapes by using combinations of neighboring wavelets with overlapping support and wavelets of different scales. The wavelet template is also computationally efficient since we compute the transform once for the whole image and look at different sets of coefficients for different spatial locations.

2.2.1 Learning the pedestrian template

As shown in Figure 1, it is easy to observe that there are no consistent patterns in the color and texture of pedestrians or their backgrounds in arbitrary cluttered scenes in unconstrained environments. This lack of clearly discernible interior features is circumvented by relying on (1) differences in the intensity between pedestrian bodies and their backgrounds and (2) consistencies within regions inside the body boundaries. We interpret the wavelet coefficients as either indicating an almost uniform area, i.e. "no-change", if their absolute value is relatively small, or as indicating "strong change" if their absolute value is relatively large. The wavelet template we seek to identify will consist solely of wavelet coefficients (either vertical, horizontal or corner) whose types ("change"/"no-change") are both clearly identified and *consistent* along the ensemble of pedestrian images; these comprise the "important" coefficients.

The basic analysis to identify the template consists of two steps: first, we normalize the wavelet coefficients relative to the rest of the coefficients in the patterns; second, we analyze the averages of the normalized coefficients along the ensemble. We have collected a set of 564 color images of people (Figure 1) for use in the template learning. All the images are scaled and clipped to the dimensions 128×64 such that the people are centered and approximately the same size (the distance from the shoulders to feet is about 80 pixels). In our analysis, we restrict ourselves to the wavelets at scales of 32×32 pixels (one array of 15×5 coefficients for each wavelet class) and 16×16 pixels (29×13 for each class). For each color channel (RGB) of every image, we compute the quadruple dense Haar transform and take the coefficient value to be the largest absolute value among the three channels. The normalization step computes the average of each coefficient's class ($\{vertical, horizontal, corner\} \times \{16, 32\}$) over all the pedestrian patterns and divides every coefficient by its corresponding class average. We calculate the averages separately for each class since the power distribution between the different classes may vary.

To begin specifying the template, we calculate the average of each normalized coefficient over the set of pedestrians. A base set of 597 color images of natural scenes of size 128×64 that do not contain people were gathered to compare with the pedestrian patterns and are processed as above. Tables 1(a) and 1(b) show the average coefficient values for the set of vertical Haar coefficients of scale 32×32 for both the non-pedestrian and pedestrian classes. Table 1(a) shows that the process of averaging the coefficients within the pattern and then in the ensemble does not create spurious patterns; the average values of these non-pedestrian coefficients are near 1 since these are random images that do not share any common pattern. The pedestrian averages, on the other hand, show a clear pattern, with strong response (values over 1.5) in the coefficients corresponding to the sides of the body and weak response (values less than 0.5) in the coefficients along the center of the body.

1.18	1.14	1.16	1.09	1.11
1.13	1.06	1.11	1.06	1.07
1.07	1.01	1.05	1.03	1.05
1.07	0.97	1.00	1.00	1.05
1.06	0.99	0.98	0.98	1.04
1.03	0.98	0.95	0.94	1.01
0.98	0.97	0.96	0.91	0.98
0.98	0.96	0.98	0.94	0.99
1.01	0.94	0.98	0.96	1.01
1.01	0.95	0.95	0.96	1.00
0.99	0.95	0.92	0.93	0.98
1.00	0.94	0.91	0.92	0.96
1.00	0.92	0.93	0.92	0.96

(a)

0.62	0.74	0.60	0.75	0.66
0.76	0.92	0.54	0.88	0.81
1.07	1.11	0.52	1.04	1.15
1.38	1.17	0.48	1.08	1.47
1.65	1.27	0.48	1.15	1.71
1.62	1.24	0.48	1.11	1.63
1.44	1.27	0.46	1.20	1.44
1.27	1.38	0.46	1.34	1.27
1.18	1.51	0.46	1.48	1.18
1.09	1.54	0.45	1.52	1.08
0.94	1.38	0.42	1.39	0.93
0.74	1.08	0.36	1.11	0.72
0.52	0.74	0.29	0.77	0.50

(b)

Table 1: Normalized vertical coefficients of scale 32×32 of images with (a) random natural scenes (without people), (b) pedestrians.

We use a gray level coding scheme to visualize the patterns in the different classes of coefficients the values of the coefficients and display them in the proper spatial layout. Coefficients close to 1 are gray, stronger coefficients are darker, and weaker coefficients are lighter. Figures 3(a)-(d) show the color coding for the

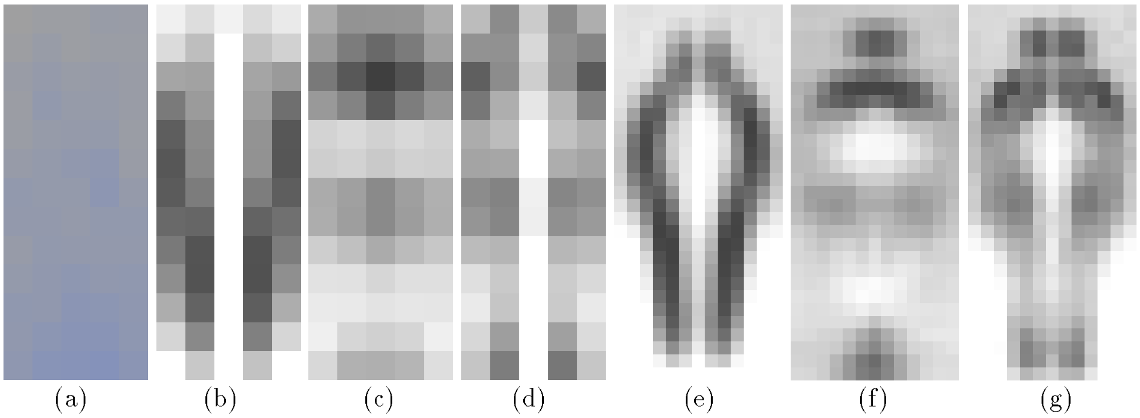

Figure 3: Ensemble average values of the wavelet coefficients coded using gray level. Coefficients whose values are above the template average are darker, those below the average are lighter. (a) vertical coefficients of random scenes. (b)-(d) vertical, horizontal and corner coefficients of scale 32×32 of images of people. (e)-(g) vertical, horizontal and corner coefficients of scale 16×16 of images of people.

arrays of coarse scale coefficients (32×32) and Figures 3(e)-(g) show the arrays of coefficients of the finer scale, (16×16).

Figure 3(a) shows the vertical coefficients of random images; as expected, this figure is uniformly gray. The corresponding images for the horizontal and corner coefficients, not shown here, are similar. In contrast, the coefficients of the people, Figures 3(b)-(d), show clear patterns, with the different classes of wavelet coefficients being tuned to different types of structural information. The vertical wavelets, Figure 3(b), capture the sides of the pedestrians. The horizontal wavelets, Figure 3(c), respond to the line from shoulder to shoulder and to a weaker belt line. The corner wavelets, Figure 3(d), are better tuned to corners, for example, the shoulders, hands and feet. The wavelets of finer scale in Figures 3(e)-(g) provide better spatial resolution of the body's overall shape and smaller scale details such as the head and extremities appear clearer. Two similar statistical analyses using a) the wavelets of the log of the intensities and b) the sigmoid function as a "soft threshold" on the normalized coefficients yields results that are similar to the intensity differencing wavelets. It is intriguing that a basic measure like the ensemble average provides clear identification of the template as shown in Figure 3.

The template derived from learning uses a set of 29 coefficients that are consistent along the ensemble either as indicators of "change" or "no-change". There are 6 vertical and 1 horizontal coefficients at the scale of 32×32 and 14 vertical and 8 horizontal at the scale of 16×16. These coefficients serve as the feature vector for the ensuing classification problem.

3 The detection system

Once we have identified the important basis functions, we can use various classification techniques to learn the relationships between the wavelet coefficients that define the pedestrian class. In this section, we present the overall architecture of the detection system, the classifier we used (the support vector machine), and the training process. We conclude with experimental results of the detection system.

3.1 System architecture

The system detects people in arbitrary positions in the image and in different scales. To accomplish this task, the system is trained to detect a pedestrian centered in a 128×64 pixel window. Once the training stage is completed, the system is able to detect pedestrians at arbitrary positions by shifting the 128×64 window, thereby scanning all possible locations in the image. This is combined with iteratively resizing the image to achieve multi-scale detection; in our experiments, we scale the image from 0.2 to 1.5 times its original size, at increments of 0.1. At any given scale, instead of recomputing the wavelet coefficients for every window in the image, we compute the transform for the whole image and do the shifting in the coefficient space. A shift of one coefficient in the finer scale corresponds to a shift of 4 pixels in the window and a shift in the coarse scale corresponds to a shift of 8 pixels. Since most of the coefficients in the wavelet template are at the finer scale (the coarse scale coefficients hardly change with a shift of 4 pixels), we achieve an effective spatial resolution of 4 pixels by working in the wavelet coefficient space.

3.2 System training

To train our system, we use a database of frontal and rear images of people from outdoor and indoor scenes. The initial non-people in the training database are patterns from natural scenes not containing people. The combined set of positive and negative examples form the initial training database for the classifier. A key issue with the training of detection systems is that, while the examples of the target class, in this case pedestrians, are well defined, there are no typical examples of non-pedestrians. The main idea in overcoming this problem of defining this extremely large negative class is the use of "bootstrapping" training [15]. After the initial training, we run the system over arbitrary images that do not contain any people. Any detections are clearly identified as false positives and are added to the database of negative examples and the classifier is then retrained with this larger set of data. These iterations of the bootstrapping procedure allows the classifier to construct an incremental refinement of the non-pedestrian class until satisfactory performance is achieved. This bootstrapping technique is illustrated in Figure 4.

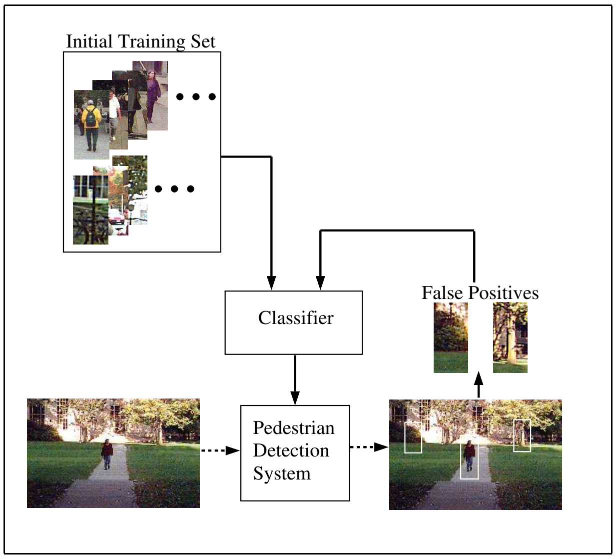

Figure 4: Incremental bootstrapping to improve the system performance.

3.3 Classification schemes

In Section 2.2.1 we described the identification of the significant coefficients that characterize the pedestrian class. These coefficients are used as the feature vector for various classification methods.

3.3.1 Basic template matching

The simplest classification scheme is to use a basic template matching measure. As in Section 2.2.1, the normalized template coefficients are divided into two categories: coefficients above 1 (indicating strong change) and below 1 (weak change). For every novel window, the wavelet coefficients are compared to the pedestrian template. The matching value is the ratio of the coefficients in agreement. A similar approach was used in [12] for face detection with good results. While this basic template matching scheme is very simple — better classification techniques can be applied — it is interesting to see how well it will perform on this more complex task.

3.3.2 Support vector machines

Instead of the simple template matching paradigm we can use a more sophisticated classifier which will learn the relationship between the coefficients from given sets of positive and negative examples. The classifier can learn more refined relationships than the simple template matching scheme and therefore can provide more accurate detection.

The classification technique we use is the support vector machine (SVM) developed by Vapnik et al.[1][18]. This recently developed technique has several features that make it particularly attractive. Traditional training techniques for classifiers, such as multilayer perceptrons (MLP), use empirical risk minimization and only guarantee minimum error over the training set. In contrast, the SVM machinery uses structural risk minimization which minimizes a bound on the generalization error and therefore should perform better on novel data. Another interesting aspect of the SVM is that its decision surface depends only on the inner product of the feature vectors. This leads to an important extension since we can replace the Euclidean inner product by any symmetric positive-definite kernel $K(x, y)$ [9]. This use of a kernel is equivalent to mapping the feature vectors to a high-dimensional space, thereby significantly increasing the discriminative power of the classifier. For our classification problem, we find that using a polynomial of degree two as the kernel provides good results.

It should be observed, that from the view point of the classification task, we could use the whole set of coefficients as a feature vector. However, using all the wavelet functions that describe a window of 128×64 pixels, over a few thousands, would yield vectors of very high dimensionality, as we mentioned earlier. The training of a classifier with such a high dimensionality would in turn require too large an example set. The template learning stage of Section 2.2.1 serves to select the basis functions relevant for this task and to reduce their number considerably (to a very reasonable 29).

4 The experimental results

To evaluate the system performance, we start with a database of 564 positive examples and 597 negative examples. The system then undergoes the bootstrapping cycle detailed in Section 3.2. For this paper, the support vector system goes through three bootstrapping steps, ending up with a total of 4597 negative examples. For the template matching version a threshold of 0.7 (70% matching) was empirically found to yield good results.

Out-of-sample performance is evaluated over a test set consisting of 72 images for both the template matching scheme and the support vector classifier. The test images contain a total of 165 pedestrians in frontal or near-frontal poses; 24 of these pedestrians are only partially observable (e.g. with body regions that are indistinguishable from the background). Since the system was not trained with partially observable pedestrians, we would not even expect a perfectly trained system (with the current template) to detect these instances. To give a fair account of the system, we present statistics for both the total set and the set of 141 “high quality” pedestrian images. Results of the tests are presented in Table 2 for representative systems using template matching and support vectors.

The template matching system has a pedestrian detection rate of 52.7%, with a false positive rate of 1 for every 5,000 windows examined. The success of such a straightforward template matching measure suggests that the template learning scheme extracts non-trivial structural regularity within the pedestrian class.

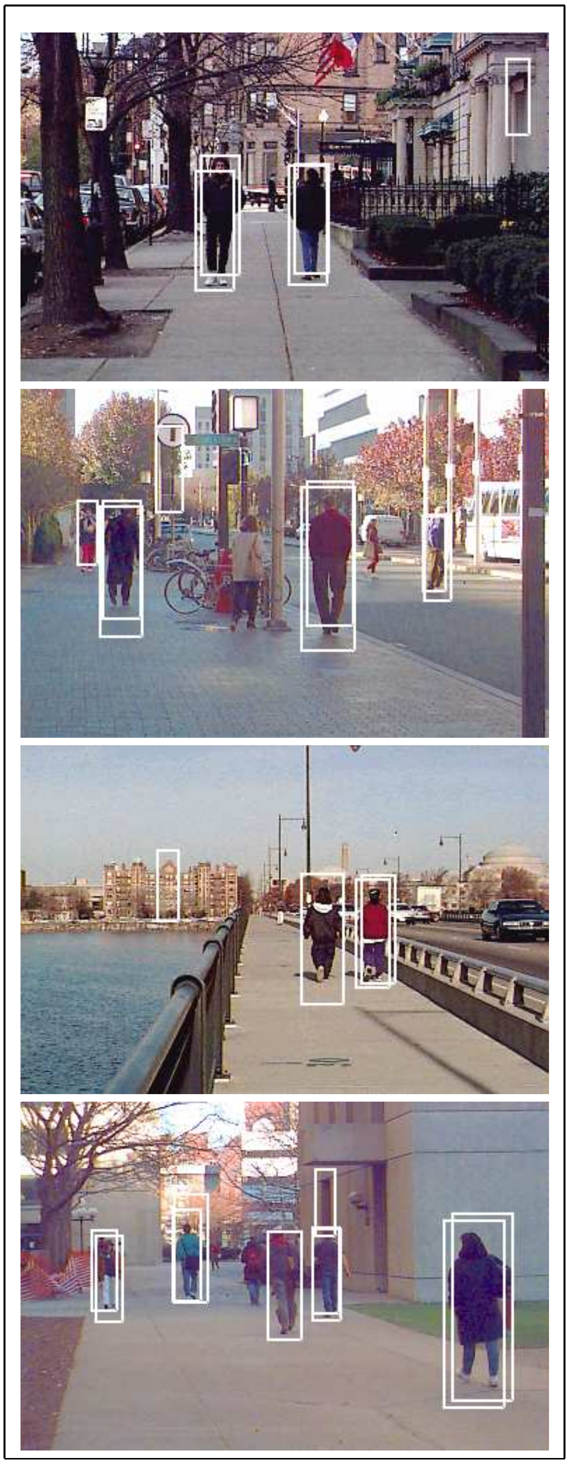

Figure 5: Results from the pedestrian detection system. These are typical images of relatively complex scenes that are used to test the system.

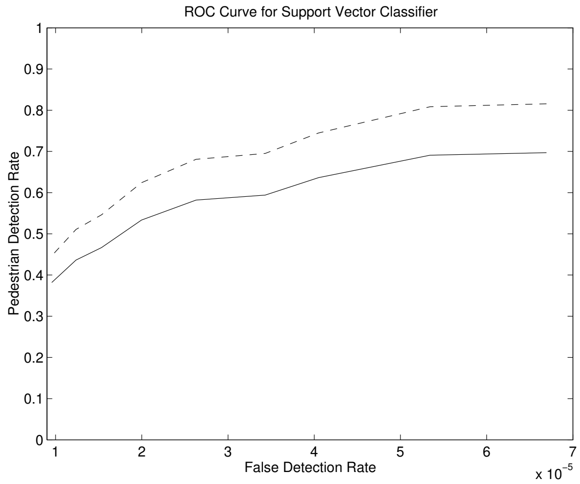

Figure 6: ROC curves for the support vector detection system; the bottom curve is over the entire test set, the top curve is over the "high quality" set.

	Detection Rate	*False Positive Rate (per window)*
Template Matching	52.7% (61.7%)	1:5,000
SVM	69.7% (81.6%)	1:15,000

Table 2: Performance of the pedestrian detection system; values in parentheses are for the set of "high quality" pedestrian images.

For the more sophisticated system with the support vector classifier, we perform a more thorough analysis. In general, the performance of any detection system exhibits a tradeoff between the rate of detection and the rate of false positives. Performance drops as we impose more stringent restrictions on the rate of false positives. To capture this tradeoff, we vary the sensitivity of the system by thresholding the output and evaluate the ROC curve, given in Figure 6. From the curve we can see, for example, that if we have a tolerance of one false positive for every 15,000 windows examined, we can achieve a detection rate of 69.6%, and as high as 81.6% on the "high quality" set. As we expect, the support vector classifier with the bootstrapping training performs better than the "naive" template matching scheme.

In Figure 5 we show typical images that are used to test the system. These are very cluttered scenes crowded with complex patterns. Considering the complexity of these scenes and the difficulties of pedestrian detection in natural outdoor scenes, we consider the above detection rate to be high. It is interesting to observe that most of the false positives are patterns with overall proportions similar to the human body. We believe that additional training and refinement of the current system will reduce the false detection rate further.

The system is currently trained only on frontal and rear views of pedestrians. Training the classifier to handle side views can be accomplished in an identical manner and is our next extension to the system.

5 Conclusion

In this paper, we introduce the idea of a wavelet template and demonstrate how it can be learned and used for pedestrian detection in a cluttered scene. The wavelet template defines an object as a set of regions and relationships among them. A key idea is to use a wavelet basis to represent the template, yielding not only a computationally efficient algorithm but also an effective learning scheme.

The success of the wavelet template for pedestrian detection comes from its ability to capture high-level knowledge about the object class (structural information expressed as a set of constraints on the wavelet coefficients) and incorporate it into the low-level process of interpreting image intensities. Attempts to directly apply low-level techniques such as edge detection and region segmentation are likely to fail in the type of images we analyze since these methods are not robust, are sensitive to spurious details, and give ambiguous results. Using the wavelet template, only significant information that characterizes the object class — as obtained in the learning phase — is evaluated and used.

In summary, in our approach a pedestrian template is learned from examples and then used for classification, ideally in a template matching scheme. It is important to realize that this is not the only interpretation of our approach, though it is the one originally suggested by [12] and is the one emphasized throughout the paper. An alternative, and more general, point of view considers the step of learning the template as a dimensionality reduction stage. Using all the wavelet functions that describe a window of 128×64 pixels would yield vectors of very high dimensionality, as we mentioned earlier. The training of a classifier with such a high dimensionality would in turn require too large an example set. The template learning stage of Section 2.2.1 serves to select the basis functions relevant for this task and to reduce their number considerably (to a very reasonable 29). A classifier – such as the support vector machine – can then be trained on a small example set. From this point of view, learning the pedestrian detection task consists of two learning steps: 1) dimensionality reduction, that is, task-dependent basis selection and 2) training the classifier. In this interpretation, a template in the strict sense of the word is neither learned nor used.

In any case, it seems that the approach described in this paper, combined with the related strategy used previously to learn face detection, may well generalize to several other object detection tasks.

Acknowledgements

This research was supported by DARPA, ONR and Daimler-Benz.

References

[1] B. Boser, I. Guyon, and V. Vapnik. A training algorithm for optim margin classifier. In *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, pages 144–52. ACM, 1992.

[2] H.-J. Chen and Y. Shirai. Detecting multiple image motions by exploiting temporal coherence of apparent motion. *Computer Vision and Pattern Recognition*, pages 899–902, 1994.

[3] R. F. Edgar Osuna and F. Girosi. Support vector machines: Training and applications. *MIT CBCL-Memo*, May 1996. In preparation.

[4] C. Jacobs, A. Finkelstein, and D. Salesin. Fast multiresolution image querying. *SIGGRAPH95*, August 1995. University of Washington, TR-95-01-06.

[5] M. Leung and Y.-H. Yang. Human body motion segmentation in a complex scene. *Pattern Recognition*, 20(1):55–64, 1987.

[6] M. Leung and Y.-H. Yang. A region based approach for human body analysis. *Pattern Recognition*, 20(3):321–39, 1987.

[7] S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–93, July 1989.

[8] B. Moghaddam and A. Pentland. Probabilistic visual learning for object detection. Technical Report 326, Media Laboratory, Massachusetts Institute of Technology, 1995.

[9] F. Riesz and B. Sz.-Nagy. *Functional Analysis*. Ungar, New York, 1955.

[10] K. Rohr. Incremental recognition of pedestrians from image sequences. *Computer Vision and Pattern Recognition*, pages 8–13, 1993.

[11] H. Rowley, S. Baluja, and T. Kanade. Human face detection in visual scenes. Technical Report CMU-CS-95-158, School of Computer Science, Carnegie Mellon University, July/November 1995.

[12] P. Sinha. Object Recognition via Image Invariants: A Case Study. In *Investigative Ophthalmology and Visual Science*, volume 35, pages 1735–1740. Sarasota, Florida, May 1994.

[13] P. Sinha. Qualitative image-based representations for object recognition. *MIT AI Lab-Memo*, No. 1505, 1994.

[14] E. Stollnitz, T. DeRose, and D. Salesin. Wavelets for computer graphics: A primer. *University of Washington, TR-94-09-11*, September 1994.

[15] K.-K. Sung and T. Poggio. Example-based learning for view-based human face detection. A.I. Memo 1521, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, December 1994.

[16] T. Tsukiyama and Y. Shirai. Detection of the movements of persons from a sparse sequence of tv images. *Pattern Recognition*, 18(3/4):207–13, 1985.

[17] R. Vaillant, C. Monrocq, and Y. L. Cun. Original approach for the localisation of objects in images. *IEE Proc.-Vis. Image Signal Processing*, 141(4), August 1994.

[18] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, 1995.