Fig. 3. It is very important to place geometric constraints on the location and scale of component detections. Even though a detection may be the strongest in a particular window examined, it might not be at the proper location. In this figure, the shadow of the person's head is detected with a higher score than the head itself. If we did not check for proper configuration and scale, component detections like these would lead to false alarms and/or missed detections of people.

invariance scheme [21]. Haar wavelets of two different scales (16×16 pixels and 8×8 pixels) are used to generate a multiscale representation of the images. The wavelets are applied to the image such that they overlap 75 percent with the neighboring wavelets in the vertical and horizontal directions; this is done to increase the spatial resolution of our system and to yield richer representation. At each scale, three different orientations of Haar wavelets are used, each of which responds to differences in intensities across different axes. In this manner, information about how intensity varies in each color channel (red, green, and blue) in the horizontal, vertical, and diagonal directions is obtained. The information streams from the three color channels are combined and collapsed into one by taking the wavelet coefficient for the color channel that exhibits the greatest variation in intensity at each location and for each orientation. At these scales of wavelets there are 582 features for the 32×32 pixel window for the head and shoulders and 954 features for the 48×32 pixel windows representing the lower body and the left and right arms. This method results in a thorough and compact representation of the components, with high interclass and low intraclass variation.

We use support vector machines (SVM) to classify the data vectors resulting from the Haar wavelet representation of the components. SVMs were proposed by Vapnik [25] and have yielded excellent results in various data classification tasks, including people detection [16], [14] and text classification [9]. Traditional training techniques for classifiers like multilayer perceptrons use empirical risk minimization and lack a solid mathematical justification. The SVM algorithm uses structural risk minimization to find the hyperplane that optimally separates two classes of objects. This is equivalent to minimizing a bound on generalization error. The optimal hyperplane is computed as a decision surface of the form:

$$f(\mathbf{x}) = sgn(g(\mathbf{x})), \tag{1}$$

where

$$g(\mathbf{x}) = \left(\sum_{i=1}^{l^*} y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i^*) + b \right). \tag{2}$$

In (2), K is one of many possible kernel functions, $y_i \in \{-1, 1\}$ is the class label of the data point $\mathbf{x}_i^*$, and $\{\mathbf{x}_i^*\}_{i=1}^{l^*}$ is a subset of the training data set. The $\mathbf{x}_i^*$ are called *support vectors* and are the points from the data set that define the separating hyperplane. Finally, the coefficients α_i and b are determined by solving a large-scale quadratic programming problem. One of the appealing characteristics of SVMs is that there are just two tunable parameters, *Cpos* and *Cneg*, which are penalty terms for positive and negative pattern misclassifications, respectively. The kernel function K that is used in the component classifiers is a quadratic polynomial and is $K(\mathbf{x}, \mathbf{x}_i^*) = (\mathbf{x} \cdot \mathbf{x}_i^* + 1)^2$.

In (1), $f(\mathbf{x}) \in \{-1, 1\}$ is referred to as the *binary class* of the data point x which is being classified by the SVM. As (1) shows, the binary class of a data point is the sign of the *raw output* $g(\mathbf{x})$ of the SVM classifier. The raw output of an SVM classifier is the distance of a data point from the decision hyperplane. In general, the greater the magnitude of the raw output, the more likely a classified data point belongs to the binary class it is grouped into by the SVM classifier.

TABLE 1
Geometric Constraints Placed on Each Component

Component	*Centroid*		*Scale*		*Other Criteria*
	Row	*Column*	*Minimum*	*Maximum*	
Head and Shoulders	23 ± 3	32 ± 2	28×28	42×42	
Lower Body		32 ± 3	42×28	69×46	*Bottom Edge:* Row: 124 ± 4
Right Arm Extended	54 ± 5	46 ± 3	31×25	47×31	
Right Arm Bent		46 ± 3	31×25	47×31	*Top Edge:* Row: 31 ± 3
Left Arm Extended	54 ± 5	17 ± 3	31×25	47×31	
Left Arm Bent		17 ± 3	31×25	47×31	*Top Edge:* Row: 31 ± 3

All coordinates are relative to the upper left-hand corner of a 128×64 *rectangle.*