
Human Detection

— A state-of-the-art survey —

Mohammad Dorgham

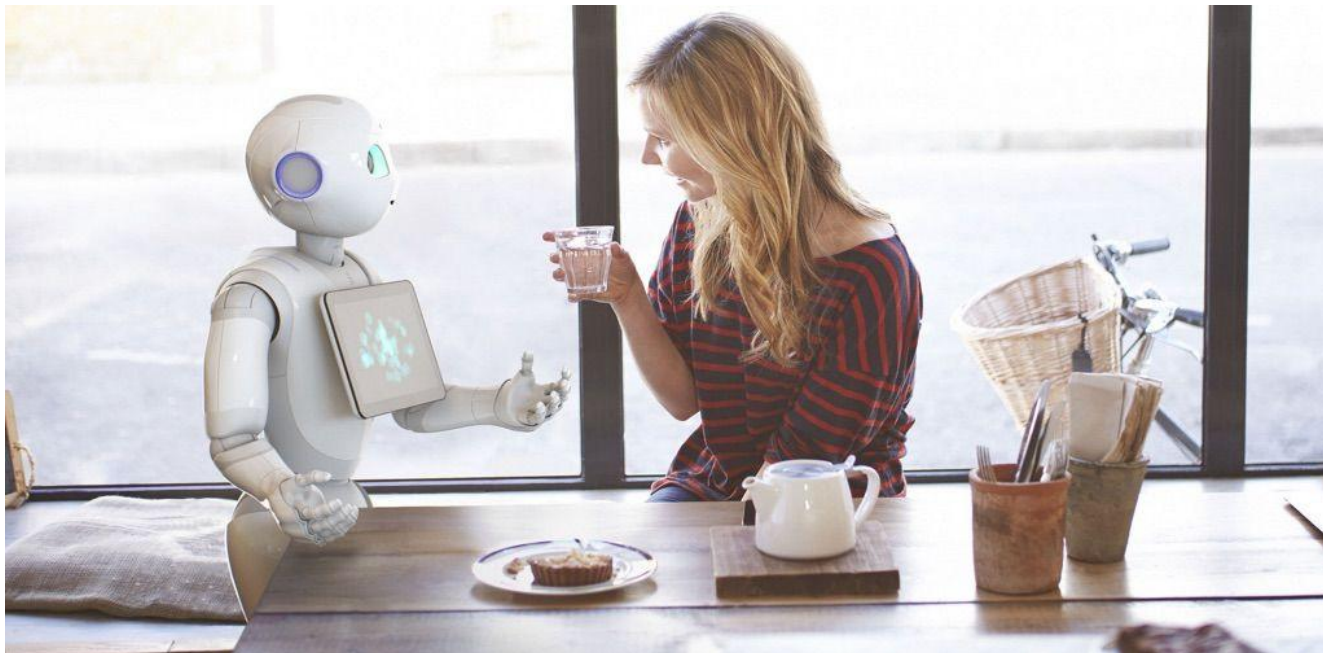
University of Hamburg

Presentation outline

- Motivation
- Applications
- Overview of approaches (categorized)
- Approaches details
- References

Motivation

- Human detection is a key problem in computer vision.
- It leads to applications that can significantly increase quality of life.
- Human detection is essential step for establishing interactions between humans and robots.



<http://robotica.news>

Applications

- Surveillance systems
- Intruder detection
- Driving assistance
- Tracking
- Robotics (human–robot interaction)
- Person re identification
- etc.

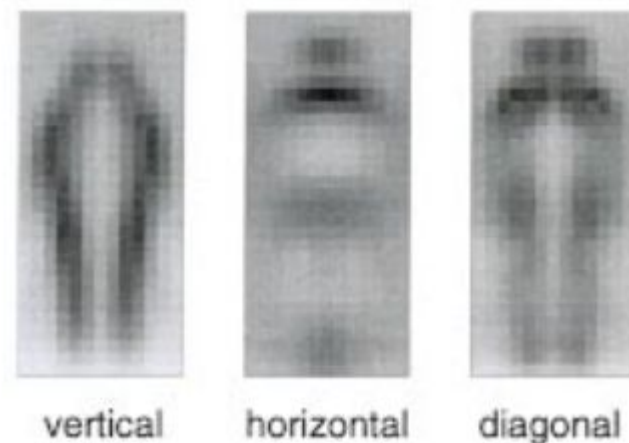
Human Detection Approaches (overview)

- Human detection based upon video cameras
 - Shape-based detection
 - Motion-based detection
 - Detection based upon multiple cues
- Human detection based upon other sensors
 - IR
 - Radar
 - Laser

Shape based detection

Haar wavelet Features

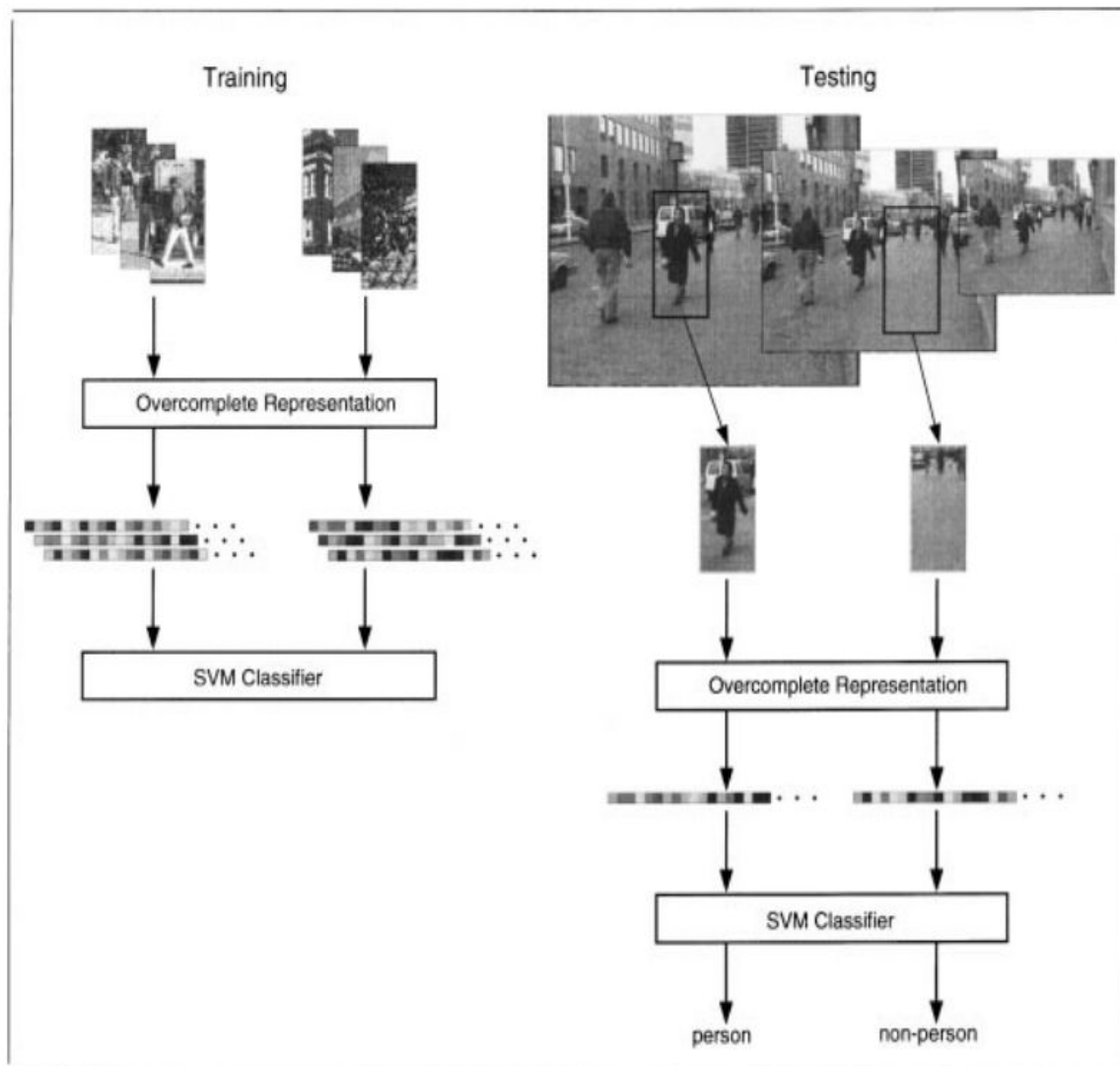
- Wavelet function, identifies local variations in intensity features.
- Haar wavelet transform is applied to the image.
- resulting in a dictionary of features that are then used as training for a classifier.
- strong response from a particular wavelet indicates the presence of an intensity difference at that location in the image, while weak response from a wavelet indicates a uniform area.



Papageorgiou, C, & Poggio T., (2000) A trainable system for object detection, *International Journal of Computer Vision* 38.1, 15-33. Figure 4.

Haar wavelet features (cont.)

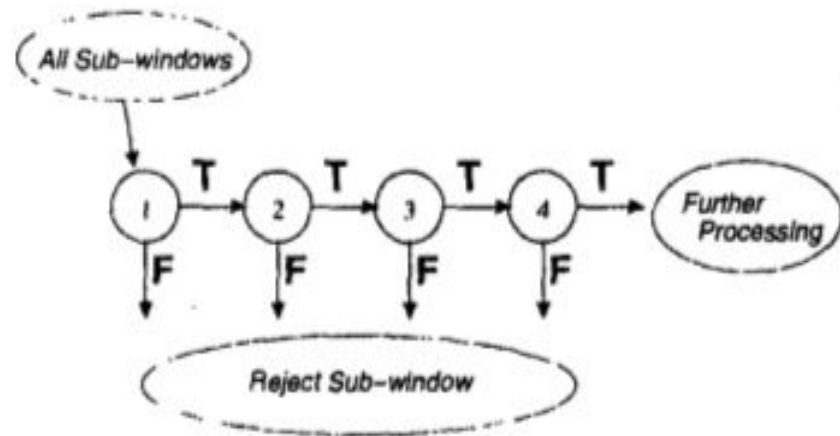
Papageorgiou, Constantine, and Tomaso Poggio.
"A trainable system for object detection."
International Journal of Computer Vision 38.1
(2000): 15-33. Figure 1.



Haar wavelet features (cont.)

- **disadvantages:**
 - large number of Haar features, led to longer processing time.
- **Viola & Jones approach:**
 - ❖ Using cascaded classifiers, each classifier with different feature set.
 - ❖ initial classifier eliminates a large number of negative examples with very little processing.
 - ❖ After several stages of processing the number of sub-windows have been reduced radically.

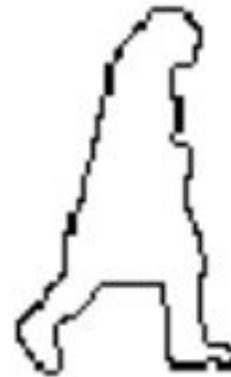
Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-511). IEEE. Figure 5.



Shape based detection (cont.)

Edge Features

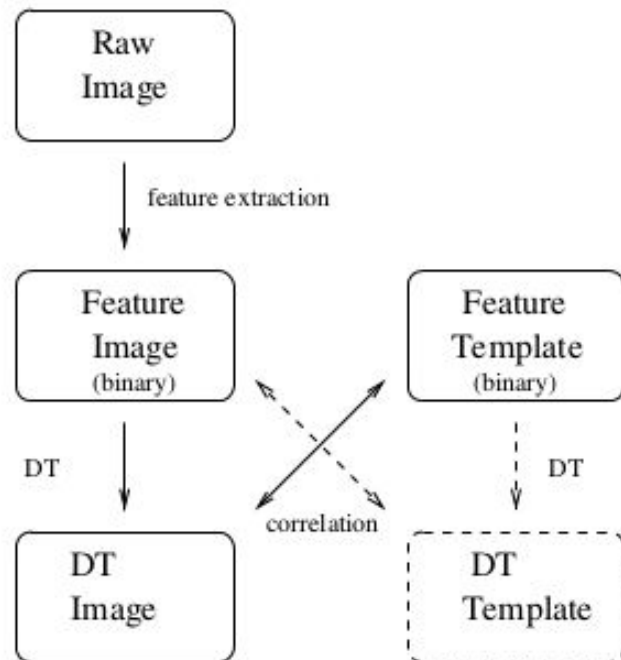
- Edges are points where there is a boundary between two image regions.



Gavrila, D. M., & Philomin, V. (1999). Real-time object detection for “smart” vehicles. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* (Vol. 1, pp. 87-93). IEEE. Figure 2.

Edge features (cont.)

Gavrila and Philoman



Edge features (cont.)

- Distance Transform:

labels each pixel of the image with the distance to the nearest obstacle pixel (e.g. boundary or edge).

<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
<u>0</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
<u>0</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
<u>0</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
<u>0</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
<u>0</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>

<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
<u>0</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
<u>0</u>	<u>1</u>	<u>2</u>	<u>2</u>	<u>2</u>	<u>2</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
<u>0</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>2</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
<u>0</u>	<u>1</u>	<u>2</u>	<u>2</u>	<u>2</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
<u>0</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>

Binary Image

Distance transformation

0 is black, 1 is white

Edge features (cont.)

- Sample of the results



Gavrila, D. M., & Philomin, V. (1999). Real-time object detection for "smart" vehicles. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* (Vol. 1, pp. 87-93). IEEE. Figure 7.

Edge features (cont.)

- **Disadvantages:**

high rate of false positives when the camera was near to object.

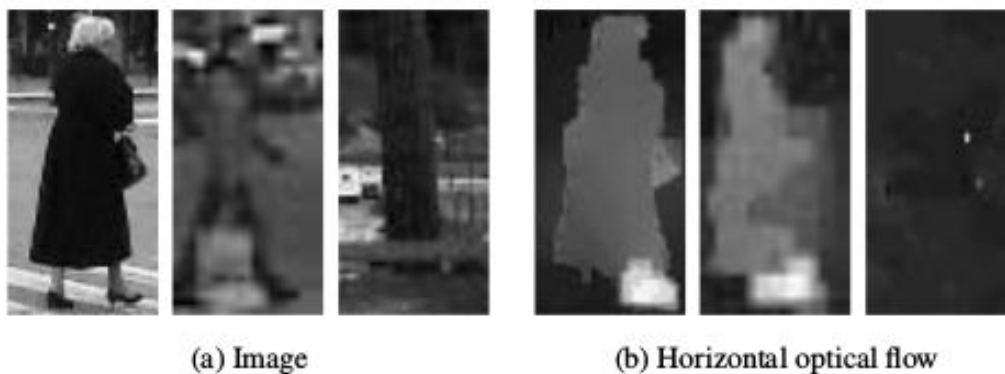


Human Detection Approaches (overview)

- Human detection based upon video cameras
 - Shape-based detection
 - **Motion-based detection**
 - Detection based upon multiple cues
- Human detection based upon other sensors
 - IR
 - Radar
 - Laser

Motion based detection

- Appearance of humans varies hugely due to clothing, identity, weather and amount and direction of light.
- Patterns of human motion, is to a large extent independent of the differences in appearance.



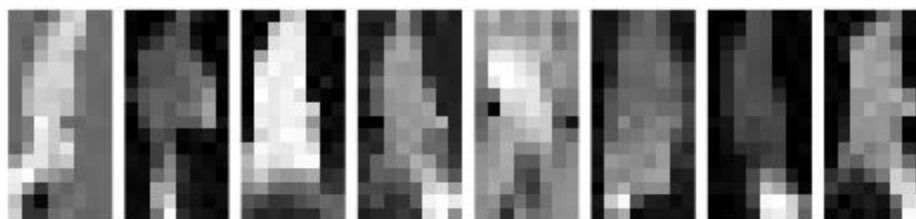
Sidenbladh, H. (2004, August). Detecting human motion with support vector machines. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* (Vol. 2, pp. 188-191). IEEE. Figure 1.

- The tree might be mistaken for a person when viewed at a low resolution if we used shape based detection.

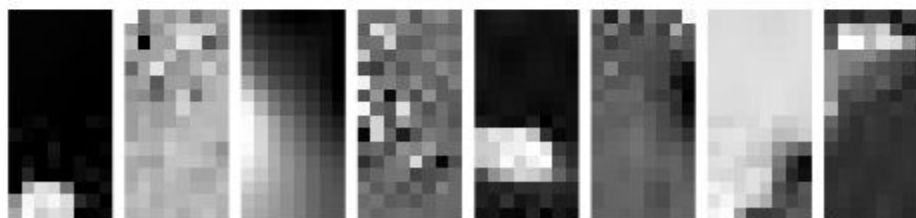
Motion based detection (cont.)

Sidenbladh approach:

1. A set of examples of human and non-human flow patterns is collected manually.
2. SVM classifier is trained with these patterns.
3. At real time detection we convert images to its optical flow representation
4. classification of flow patterns with the trained SVM model to human or non-human.



(a) Human motion patterns

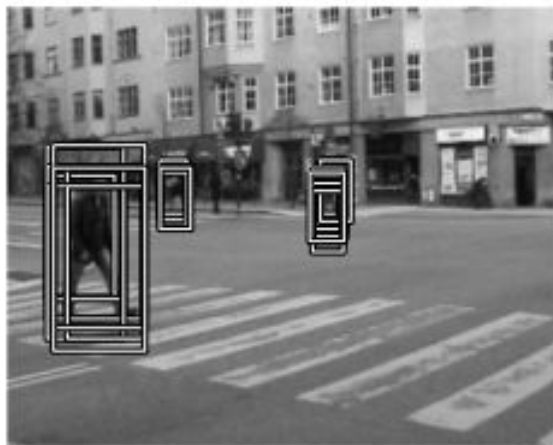


(b) Non-human motion patterns

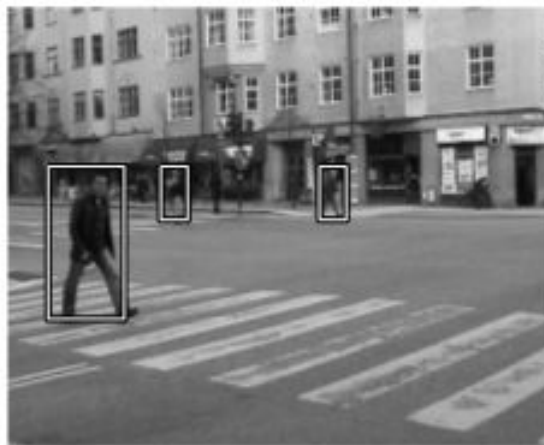
Sidenbladh approach (cont.)

Removing multiple detections

- due to windowing when processing the image, overlapping human pattern candidates corresponding to the same individual are found.
- hits overlapping more than 50% will be replaced by a single window.
- window position and height is the spatial weighted mean of the positions and heights of the overlapping windows.

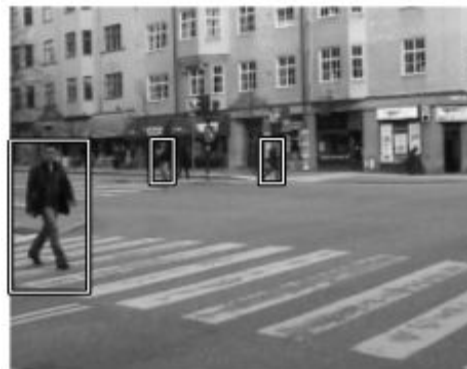


(a) All detections



(b) Multiple candidates removed

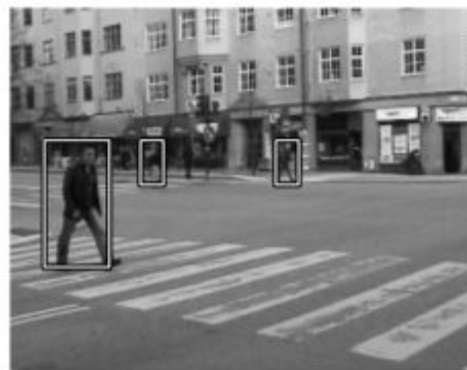
Results sample



(a) Frame 0, image



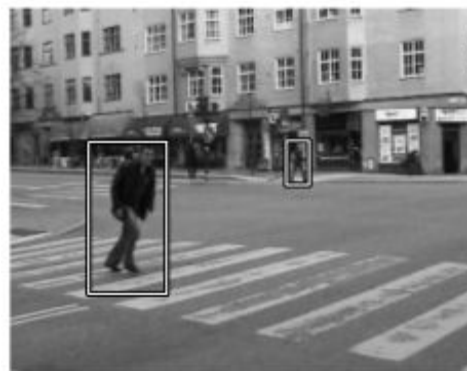
(b) Frame 0, horizontal flow



(c) Frame 10, image



(d) Frame 10, horizontal flow



(e) Frame 20, image



(f) Frame 20, horizontal flow

Sidenbladh, H. (2004, August). Detecting human motion with support vector machines. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* (Vol. 2, pp. 188-191). IEEE.

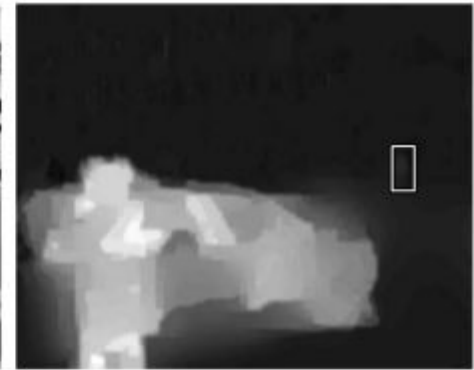
Weakness

- Inability to detect partly occluded humans.

(Figures a,e)



(a) Frame 0, image



(b) Frame 0, horizontal flow



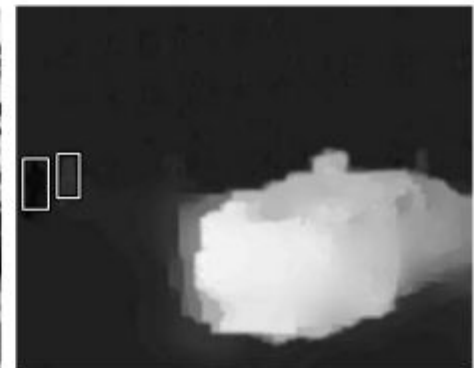
(c) Frame 10, image



(d) Frame 10, horizontal flow



(e) Frame 20, image



(f) Frame 20, horizontal flow

Human Detection Approaches (overview)

- Human detection based upon video cameras
 - Shape-based detection
 - Motion-based detection
 - **Detection based upon multiple cues**
- Human detection based upon other sensors
 - IR
 - Radar
 - Laser

Detection based upon multiple cues

- Individual cues have their limitation under different conditions such as
 - Color cue is greatly affected by change in lighting conditions.
 - Shape cue proved effective for detection of rigid object but vulnerable for detecting nonrigid objects with complex edges.
 - Motion cue could not distinguish detected object from rest of moving object in the scene.
- Combining cues could enhance the performance of detection system.
- Similar to human sensory mechanism which works as multi-sensory system.

Detection based upon multiple cues (cont.)

- Li et al. combined the motion information with shape information for human detection algorithm which was invariant to pose, shadow and occlusion in video.
- Guo et al. proposed Two stage classifier trained through Haar, texture, symmetry, boundary movement and HOG.
- Armanfard et al. proposed a method using texture and edge information for detecting pedestrians.
- Ye et al. combined motion with head and shoulder detection.
- etc.

Human Detection Approaches (overview)

- Human detection based upon video cameras
 - Shape-based detection
 - Motion-based detection
 - Detection based upon multiple cues
- **Human detection based upon other sensors**
 - **IR**
 - Radar
 - Laser

Infrared Sensor

- **The Idea**

- Vision-based sensors are limited under certain conditions such as night.
- Human emit heat energy in the IR domain and this could be categorized effectively from other heat source emissions.
- images captured are less affected by shadow, lighting, texture, color and shadow.

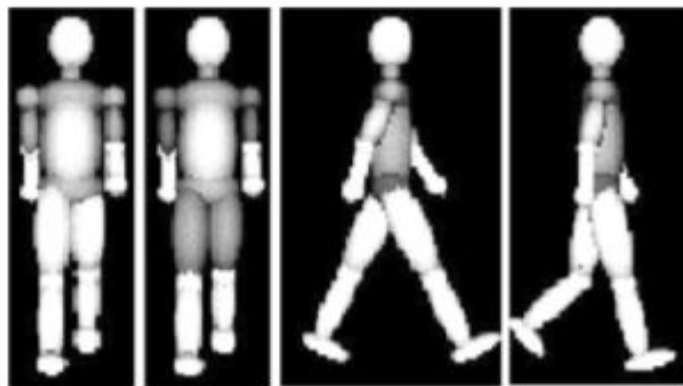
wikipedia.com



Infrared Sensor (cont.)

- **Bertozzi et al. approach:**

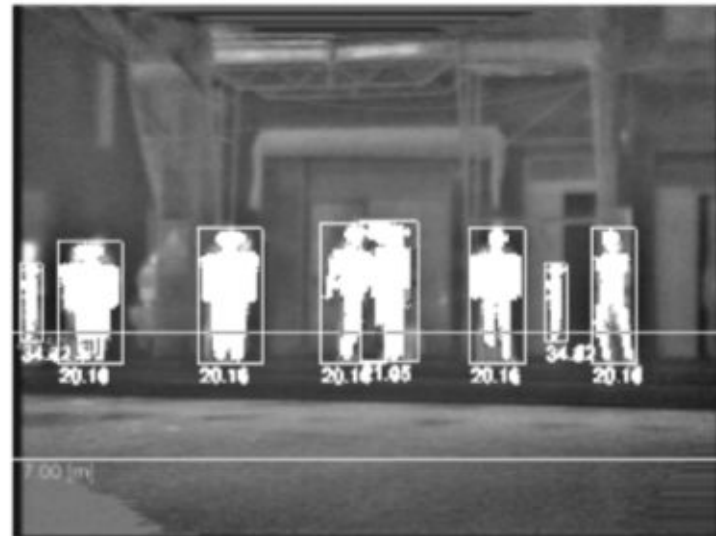
1. Localization of warm symmetrical objects with specific aspect ratio and size.
2. candidates filtering to remove errors, based on non-pedestrian characteristics.
3. candidates validation on the basis of a match with a model of a pedestrian.



Models representing different clothings, postures, and points of view.

Bertozzi et al. (cont.)

Result sample



Radar Sensor

- Radar is an object-detection system that uses radio waves to determine the range, angle, and velocity of objects.
 - The radar signals are less affected by noises (i.e. fog, rain, dust, smoke, fire, etc.).
 - it can be used for detection through walls.
-
- ❖ Li et al. exploited life characteristics of human such as breathing, motion, etc. for detection of through wall human detection.
 - ❖ The parameters from periodic motion of human were extracted using Fast Fourier Transform and S transform and used for locating the trapped human.

Laser Sensor

- **The idea**

- laser range finder transmits beams into different directions which will hit the object placed at different distance.
- The receiver captures the reflected beam and measure the time between transmitted and reflected beam.
- From this time difference the distance of object is measured.
- They are used for human detection due to their ability to capture geometry of object, measure distance information accurately.
- *Mucientes and Bugari* introduced pattern classifier for learning the spatial temporal pattern of human.
- **Disadvantages:** detection using only laser sensor is restrictive because no color information which is necessary for distinguishing humans standing close to each other.

Multiple Sensors

- More than one sensor outputs are combined for robust human detection under challenging environmental conditions.
- ❖ *Choi and Park* used thermal and video camera for human detection.
- ❖ *Bellotto and Hu* combined information from laser and video sensor.
- ❖ *Susperregi et al.* proposed a multimodal approach for human detection using mobile robot. Color image, depth and thermal images were obtained.
- ❖ etc.

Human Detection Approaches (overview)

- **Human detection based upon video cameras**
 - Shape-based detection
 - Motion-based detection
 - Detection based upon multiple cues
- **Human detection based upon other sensors**
 - IR
 - Radar
 - Laser

References

- ❖ Walia, G. S., & Kapoor, R. (2014). HUMAN DETECTION IN VIDEO AND IMAGES—A STATE-OF-THE-ART SURVEY. *International Journal of Pattern Recognition and Artificial Intelligence*, 28(03), 1455004.
- ❖ Papageorgiou, C., & Poggio, T. (2000). A trainable system for object detection. *International Journal of Computer Vision*, 38(1), 15-33.
- ❖ Gavrila, D. M., & Philomin, V. (1999). Real-time object detection for “smart” vehicles. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* (Vol. 1, pp. 87-93). IEEE.
- ❖ Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-511). IEEE.
- ❖ Sidenbladh, H. (2004, August). Detecting human motion with support vector machines. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* (Vol. 2, pp. 188-191). IEEE.
- ❖ Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7), 674-693.