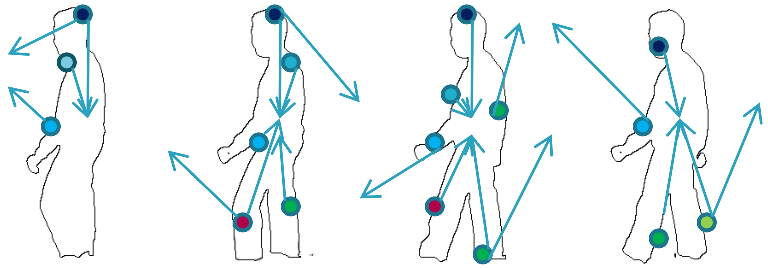
Detecting Humans in Crowded Scenes

ACM Multimedia 2007, Mikel Rodriguez and Mubarak Shah

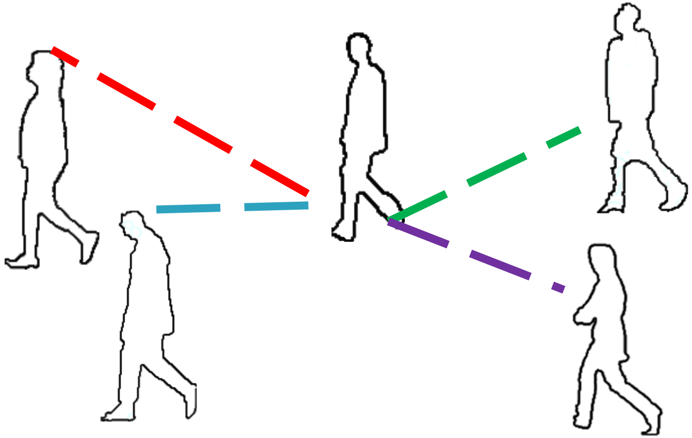

Burgstraße →

Approach overview

- Learn the spatial distribution of local shapes to find human locations.



- Use global shape information to refine initial hypotheses.



Learning local and global shape

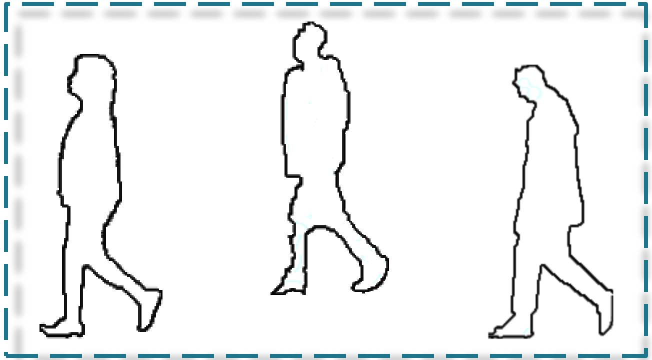
Learn Global posture clusters

Phase I

Learn local shape distribution

Phase II

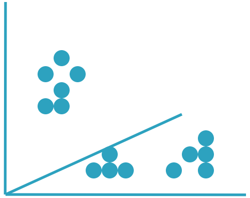
Learning Posture Clusters (global shape)



- Intuition: Use a set of learned posture clusters to initialize segmentations.
  - Initial segmentation contours can them be evolved.

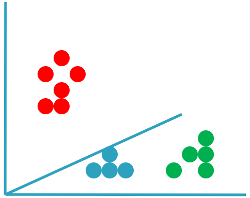
Training Video → Background Subtraction → Foreground blobs → Extract contour → Cluster all contours → Saved cluster centers

Learning Posture Clusters



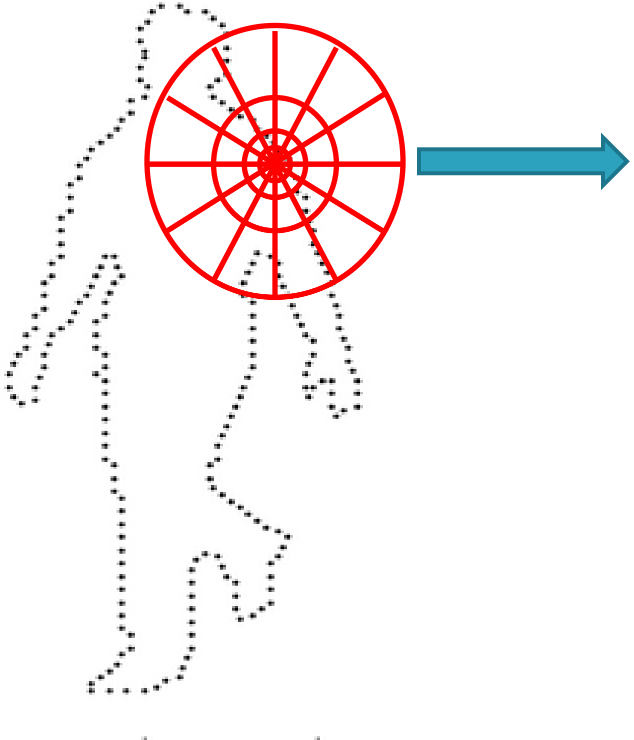
Training Video → Background Subtraction → Foreground blobs → Extract contour → Cluster all contours → Saved cluster centers

Learning Posture Clusters

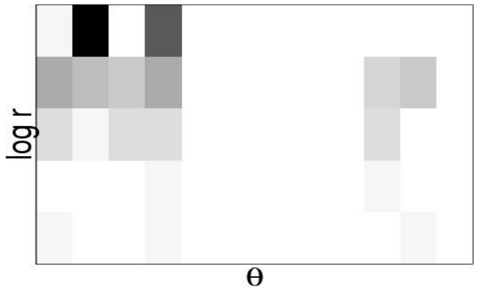


Training Video → Background Subtraction → Foreground blobs → Extract contour → Cluster all contours → Saved cluster centers

Learning Posture Clusters

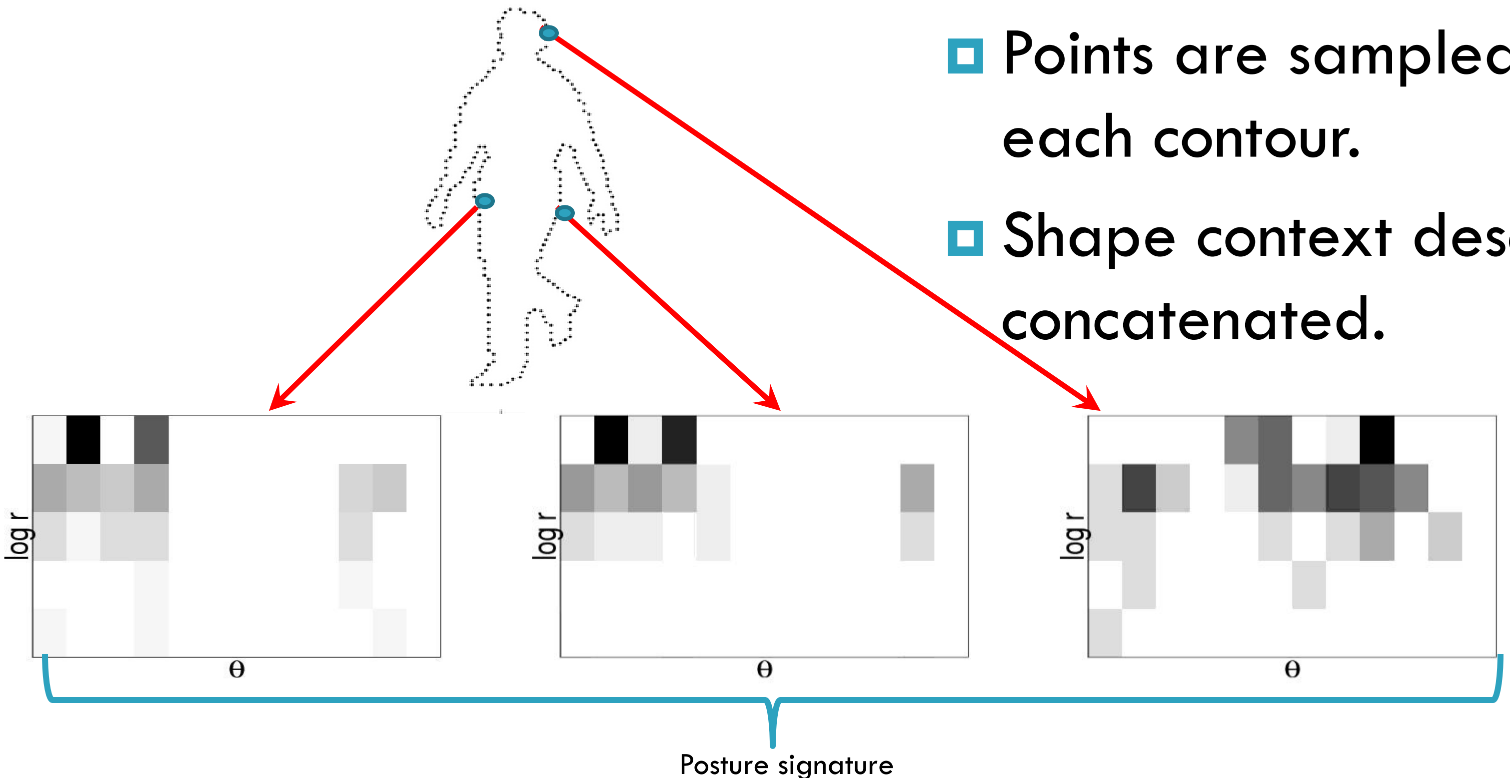


Count the number of points inside each bin, e.g.:



F A compact representation of distribution of points relative to each point (Shape context, Belongie et al)

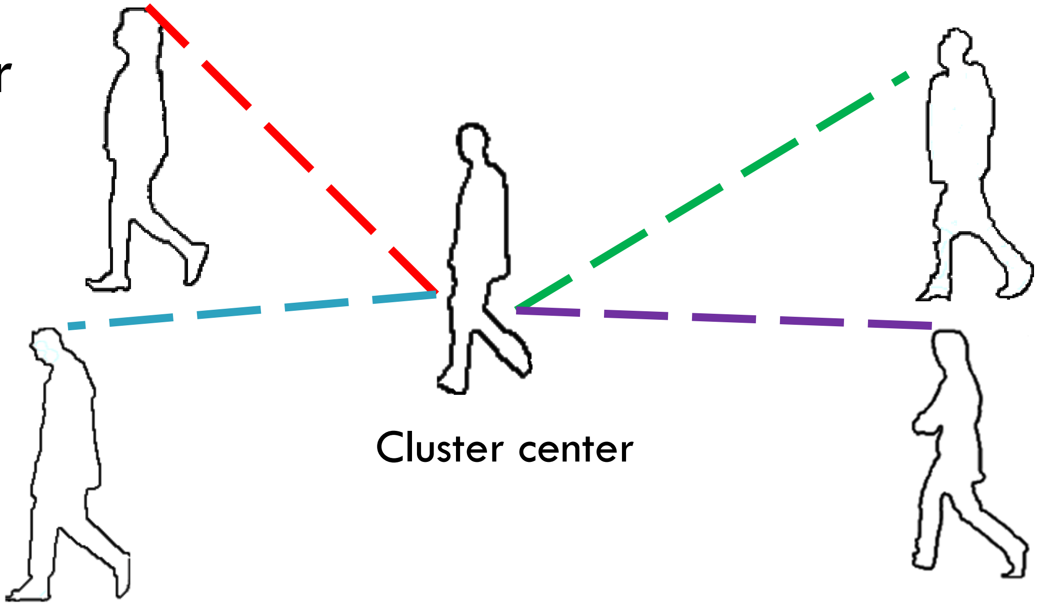
Shape Context



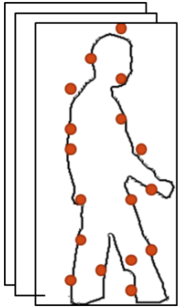
- Points are sampled along each contour.
- Shape context descriptors are concatenated.

Learning Posture Clusters (global shape)

- Posture signatures are clustered using K-means.
- The contour of each cluster center is stored.
  - Represent typical global shapes present in the data.



Local shape codebook



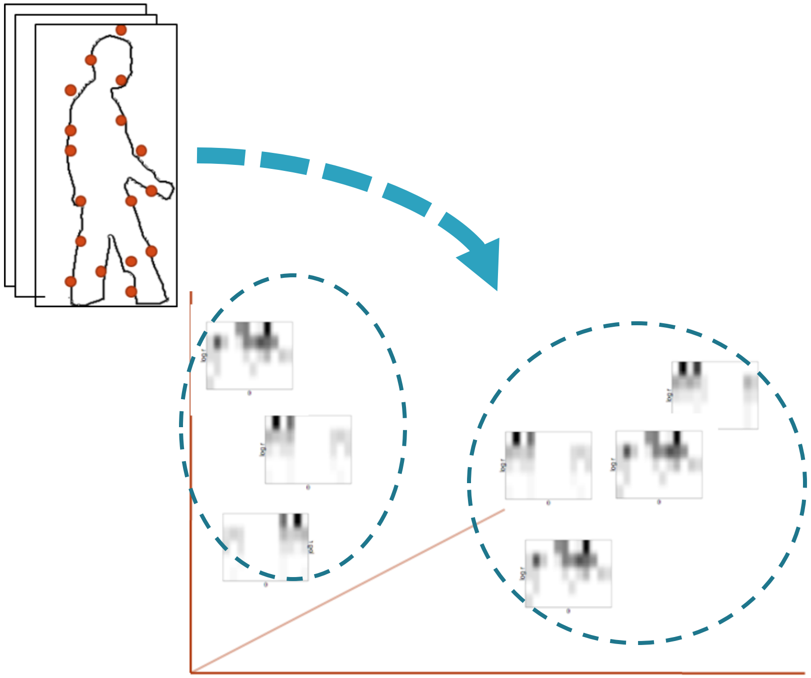
Sample silhouettes using shape context

Cluster SC descriptors into M clusters.

Look for occurrences of codebook

Save centroid vote for each codebook entry (multiple votes may exist per entry)

Local shape codebook



Sample silhouettes using shape context

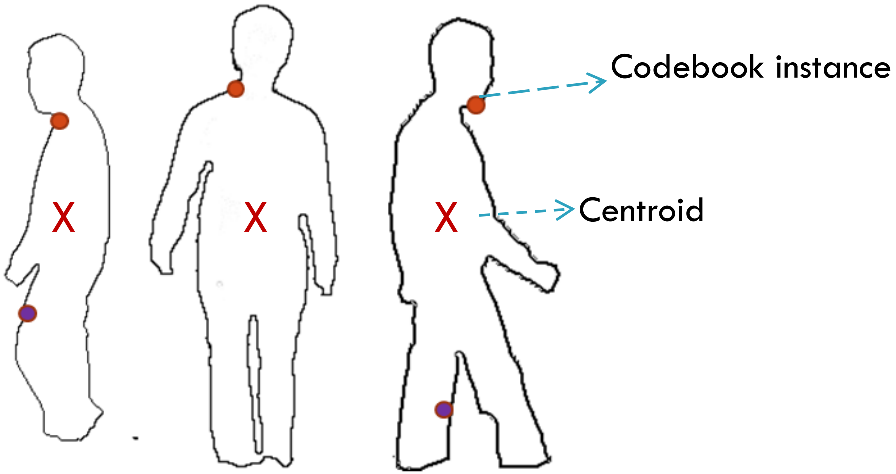
Cluster SC descriptors into M clusters.

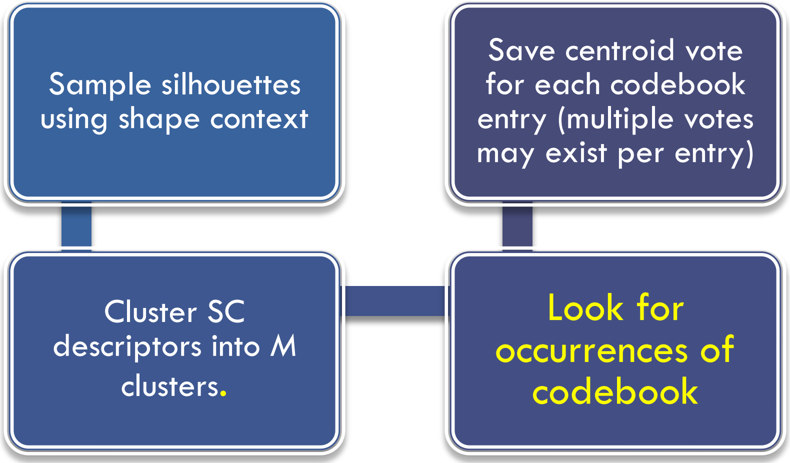
Look for occurrences of codebook

Save centroid vote for each codebook entry (multiple votes may exist per entry)

Local shape distribution

- Learn the spatial distribution of local shapes.





Local shape distribution

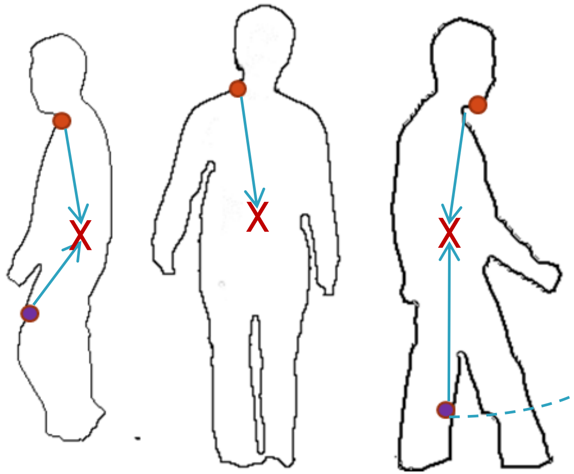
- Learn the spatial distribution of local shapes.

Sample silhouettes using shape context

Save centroid vote for each codebook entry (multiple votes may exist per entry)

Cluster SC descriptors into *M* clusters.

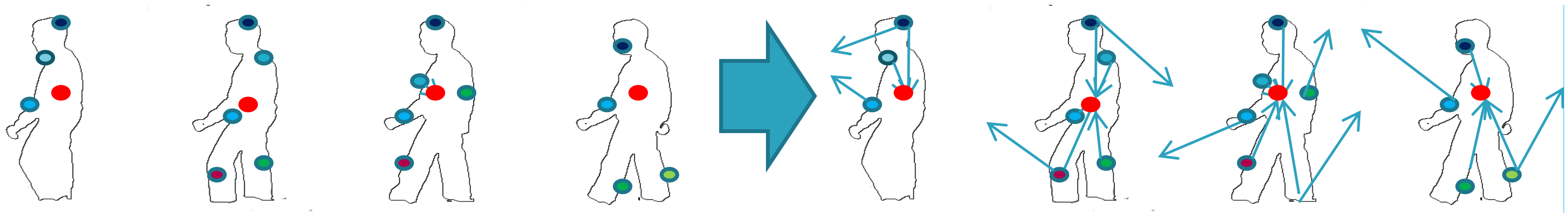
Look for occurrences of codebook



Shape Context Descriptor	**Centroid Vote**	**Global posture cluster**
…	…	…
…	…	…

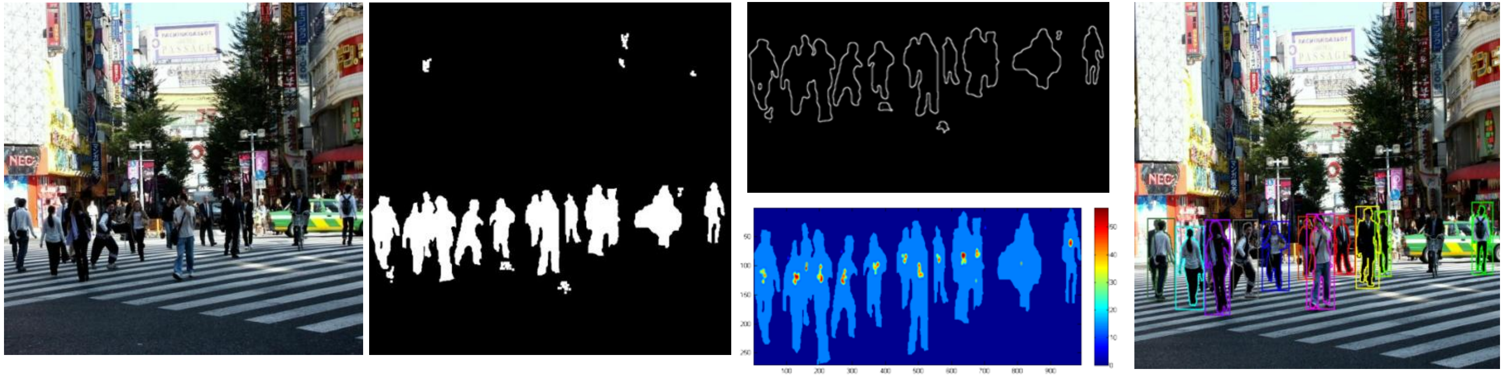
Local shape

- Instances of local shape codebook vote for locations and postures of humans in the scene.
- We search for consisted hypothesis by finding maximums in a voting space.



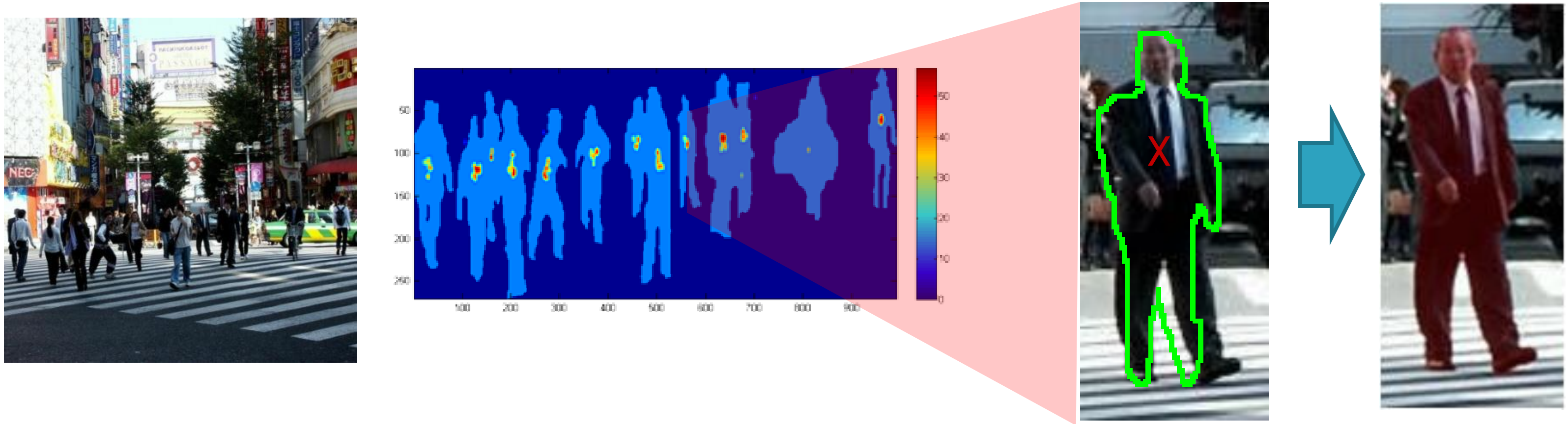
Detection

- Contours of foreground blobs are sampled.
- Centroid votes are cast
- Maximums in voting space correspond to human locations



Segmentation

- Given the centroid locations of humans in the scene:
  - We initialize segmentations using the learned posture clusters.



Experiments

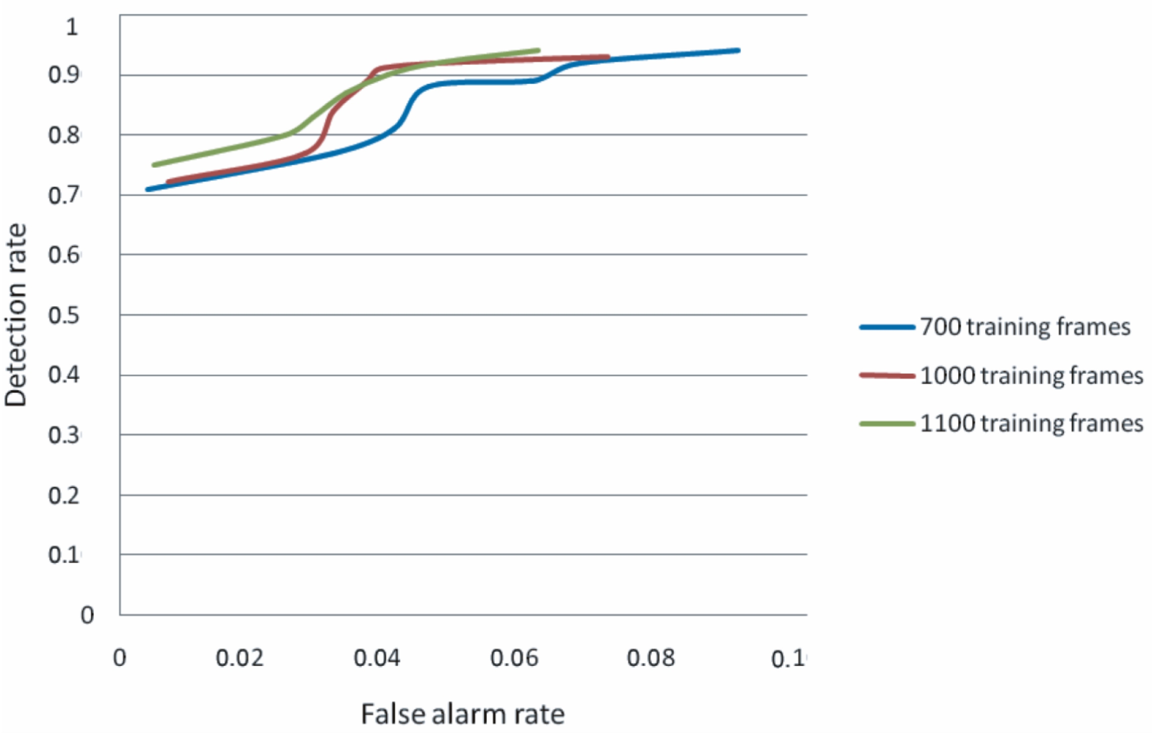
- Our testing database consisted of a wide range of scenes, totaling 34,100 frames in size and contained a total of 312 humans for which the torso is visible.
- The size of the humans across the video sequences averaged 42x65 pixels.
- Training set sizes:
  - Run1: 700 frames
  - Run 2: 1,000 frames
  - Run 3: 1,100 frames

Experiments

Experiments

Experiments and Results

Recognition performance based on a range of training set sizes.



Conclusion

- Local shape distribution represents a powerful cue which can be integrated into existing lines of research.
- Given a set of initial hypothesis, global shape clusters aid the segmentation process.
- Encouraging results given the difficulty of the data.

Thank you