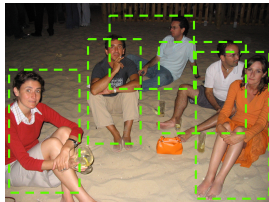**Berkeley Parlab**

Towards Robust, Real-Time Human Detection and Pose Estimation

Michael Anderson, Chao-Yue Lai, Mark Murphy, Lubomir Bourdev, Jitendra Malik, Kurt Ketuzer

Human Detection

- Goal: Draw a bounding box around all humans in an image
  - Applications: surveillence, sociology research, gaming

Poselets

- Novel human detection algorithm created by Lubomir Bourdev at Berkeley
- *Poselet* – Detectable part of a human pose
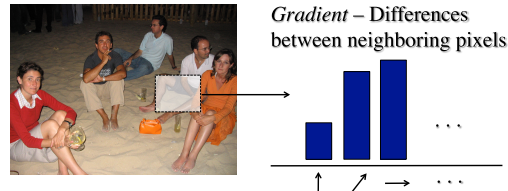- Uses the locations of individual body parts to determine the location of the body

Source: Lubomir Bourdev and Jitendra Malik: *Poselets: Body Part Detectors Trained Using 3D Human Pose Annotations*

Enabling New Applications

- Original system was written in Matlab
  - Several minutes to process a single image

- Many applications demand immediate results

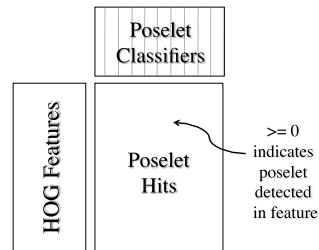- Optimized critical sections in C++ and CUDA to bring runtime to near real-time

Step 1: Feature Extraction

- Histograms of Oriented Gradients

Gradient – Differences between neighboring pixels

- Parallelism: blocks can be independent

Step 2: SVM Evaluation

Poselet Classifiers

HOG Features

Poselet Hits

>= 0 indicates poselet detected in feature

- Each column is an SVM for a specific poselet
- Linear SVMs: Matrix Multiplication

Step 3: Cluster Poselet Votes

- Each poselet votes for center of body
  - Example: Shoulder poselet has a guess for where the rest of the body is located
- Agglomerative clustering of votes

Poselet Prediction

- Fewer poselets considered = reduced runtime
- Learn which poselets are likely to be present in the next frame, based on the current frame.

Runtime with Poselet Selection

Detection Rate with Poselet Selection
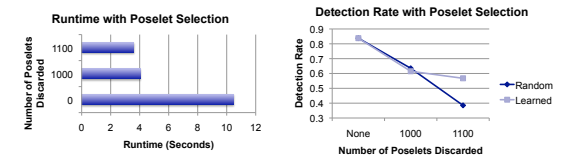
Training videos and test videos from the Carnegie Mellon Motion Capture Database. http://mocap.cs.cmu.edu/

Results

	Original Runtime	Our Runtime	Speedup
HOG Feature Extraction	58 sec.	1.21 sec.	48x
Reshape/SVM/Scan	110 sec.	7.80 sec.	14x
Mean Shift Clustering	11 sec.	0.25 sec.	43x
Agglomerative Clustering	1 sec.	0.12 sec.	10x
Total	**2 min. 12 sec.**	**11.1 sec.**	**17x**

Based on a 1024x768 test image with four people present.

Future Work

- Extend to real-time 3D pose estimation for **video game interfaces**

- Poselets vote for locations of key points in 3D configuration space rather than the torso center