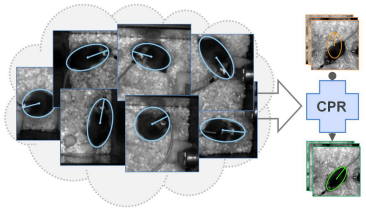
Cascaded Pose Regression

Piotr Dollár
Dept. of Electrical Engineering
California Institute of Technology
pdollar@caltech.edu

Peter Welinder
Computation and Neural Systems
California Institute of Technology
welinder@caltech.edu

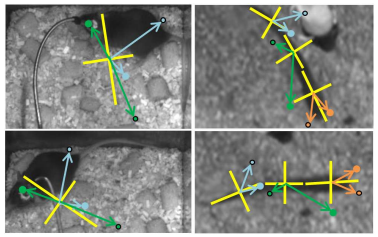
Pietro Perona
Dept. of Electrical Engineering
California Institute of Technology
perona@caltech.edu

Overview



Cascaded Pose Regression (CPR)
- Principled approach to pose estimation
- Built using pose-indexed features
- Learning from training examples
- Natural coarse to fine estimation

Pose-Indexed Features

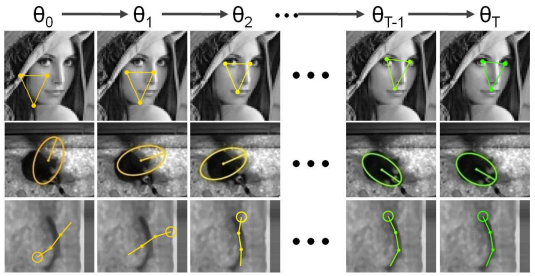


Key property: weak invariance

Training Algorithm

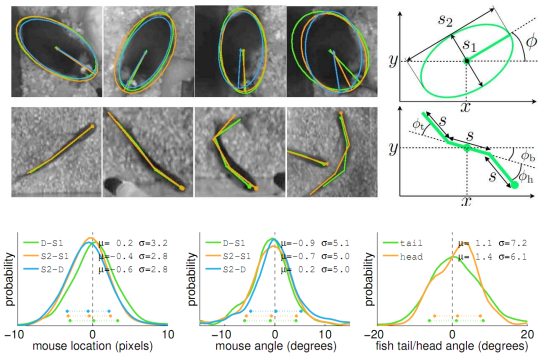
Input: Data (I_i, θ_i) for $i = 1 \dots N$
1: $\theta^0 = \arg\min_\theta \sum_i d(\theta, \theta_i)$
2: $\theta_i^0 = \theta^0$ for $i = 1 \dots N$
3: **for** $t = 1$ to T **do**
4: $\quad x_i = h^t(\theta^{t-1}, I_i)$
5: $\quad \widetilde{\theta}_i = \overline{\theta}_i^{t-1} \circ \theta_i$
6: $\quad R^t = \arg\min_R \sum_i d(R(x_i), \widetilde{\theta}_i)$
7: $\quad \theta_i^t = \theta_i^{t-1} \circ R^t(x_i)$
8: $\quad \epsilon_t = \sum_i d(\theta_i^t, \theta_i) / \sum_i d(\theta_i^{t-1}, \theta_i)$
9: $\quad$ If $\epsilon_t \geq 1$ stop
10: **end for**
11: Output $R = (R^1, \dots, R^T)$

Cascaded Pose Evaluation

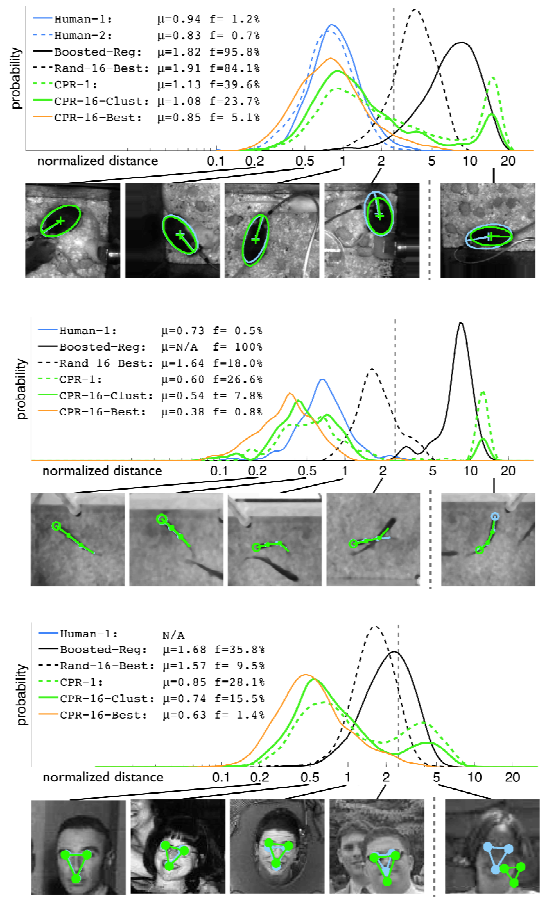


Input: Image I, initial pose θ^0
1: **for** $t = 1$ to T **do**
2: $\quad x = h^t(\theta^{t-1}, I)$ // compute features
3: $\quad \theta_\delta = R^t(x)$ // evaluate regressor
4: $\quad \theta^t = \theta^{t-1} \circ \theta_\delta$ // update θ^t
5: **end for**
6: Output θ^T

Human Consistency Analysis

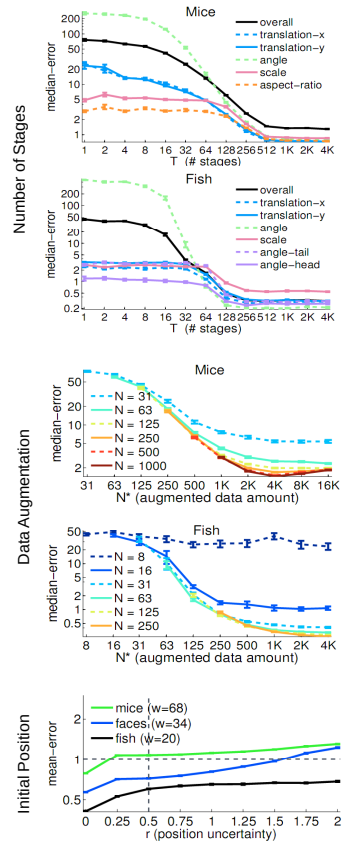


Results



Human: human versus human performance.
Boosted-Reg: boosted regression [17] using same features as CPR.
Rand-16-Best: oracle selects the best of 16 random poses.
CPR-1: CPR with a single ($K = 1$) starting pose.
CPR-16-Clust: CPR with 16 starting poses followed by clustering.
CPR-16-Best: CPR with 16 starting poses, oracle selects best.

Performance Breakdown



Conclusion

Advantages:
- Principled learning driven approach
- General, robust and accurate
- Data augmentation
- Fast (2-3ms per image)

Disadvantages:
- Pose-indexed feature design
- Supervision (~250 labels)

Future Work:
- Integrate with detection
- Integrate with tracking

CPR