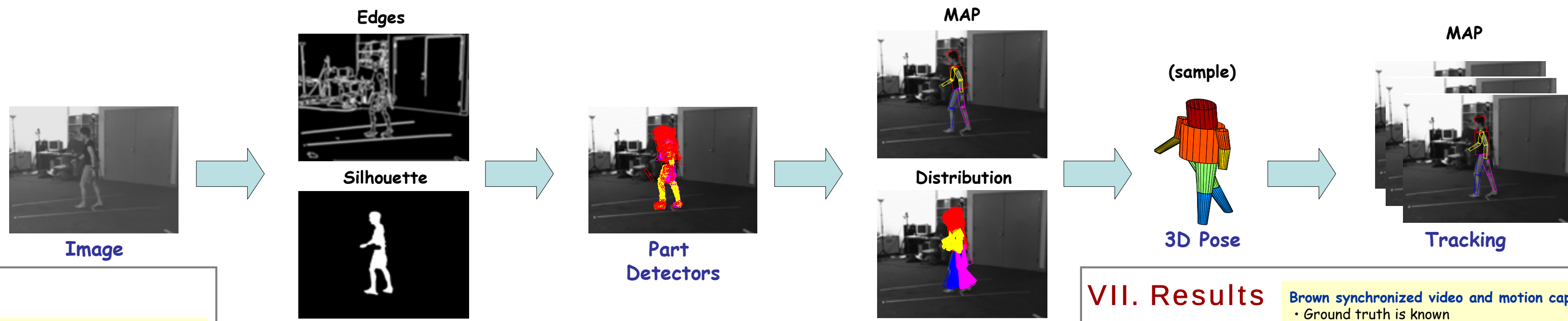
Hierarchical Approach for Articulated 3D Pose-Estimation and Tracking

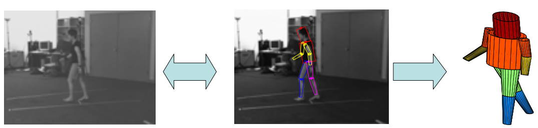
Leonid Sigal and Michael J. Black



I. Introduction

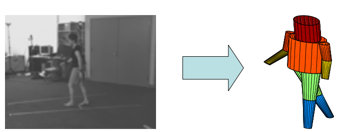
Goals
1) Recover a 3D pose of a person from a single image
2) Assume that image features are imperfect
3) Have a fully probabilistic hierarchical model

Approach: Decompose hard problem into simpler pieces



II. Related Work

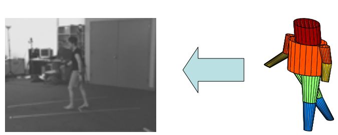
Discriminative approaches
Attempt to learn the mapping from image features (silhouettes) to 3D body pose directly.



Pros: Typically very fast and accurate when test data is like train data.

Cons: Do not generalize well in cases where image evidence is poor.

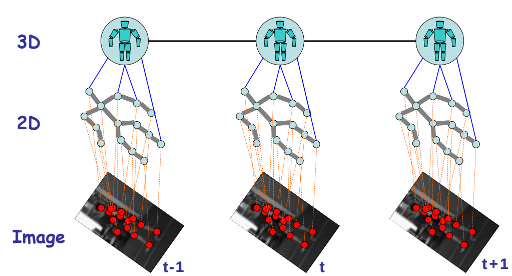
Generative approaches
Postulate hypotheses and test them against image evidence.



Pros: Typically better handle poor image evidence.

Cons: Tend to be very slow.

III. Hierarchical Inference



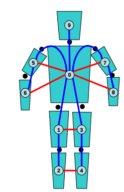
Advantages:
- Allows 3D pose estimation from single monocular image
- Makes use of temporal information when available
- Fully probabilistic framework
- Mediates the complexity of the 3D pose estimation in the absence of perfect silhouette features using an intermediate 2D pose estimation stage

IV. 2D Pose Inference

2D "Loose-Limbed" Body Model: Body is modeled using a graphical model

- Nodes correspond to limbs
- Edges correspond to **Kinematic** and **Occlusion** constraints and are modeled using potentials

Inference is carried out using a variant of Non-Parametric Belief Propagation (PAMPAS) [Isard '03]



V. 2D to 3D Pose Inference

Conditional mapping from 2D to 3D pose is modeled using a Mixture of Experts (MoE) Model [Sminchisescu '05]

$$p(Y \mid X) \propto \sum_{k=1}^{K} p_{e,k}(Y \mid X)\, p_{g,k}(k \mid X)$$

2D Body Pose

3D Body Pose

Sum over all experts*

Expert: probability of the output 3D pose Y according to the k-th expert*

Gaiting Network: probability that input 2D pose X is assigned to the k-th expert*.

*use linear regression as our expert model

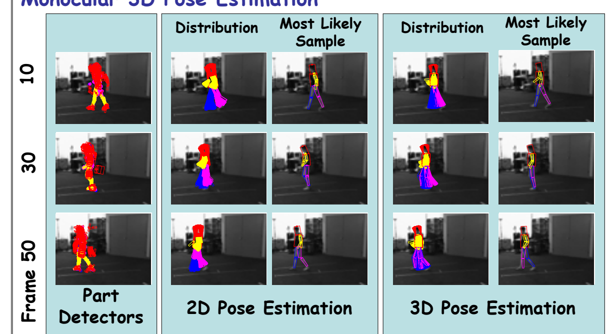
VI. Tracking in 3D

HMM over 3D state:
- Batch inference forward/backwards in time
- Using Non-Parametric Belief Propagation (PAMPAS)

VII. Results

Brown synchronized video and motion capture dataset
- Ground truth is known
- First 50 frames were used for evaluation

Monocular 3D Pose Estimation



3D Tracking

VIII. Conclusions

- New hierarchical framework for 3D pose inference and tracking from monocular video is proposed.
- Complexity of the problem is mediated using an intermediate 2D pose estimation stage
- Results are very encouraging (quantitative evaluation in AMDO '06)

References

- M. Isard. **Pampas: Real-Valued Graphical Models for Computer Vision**, CVPR 2003.
- L. Sigal and M. Black. **Predicting 3D People from 2D Pictures**. IV Conference on Articulated Motion and Deformable Objects, AMDO 2006 (*to appear*).
- L. Sigal and M. Black. **Measure Locally, Reason Globally: Occlusion-sensitive Articulated Pose Estimation**. CVPR 2006.
- C. Sminchisescu, A. Kanaujia, Z. Li and D. Metaxas. **Discriminative Density Propagation for 2D Human Motion Estimation**. CVPR, 2005.

Acknowledgements: This work was partially supported by Intel Corporation and NSF IGERT award #9870676.

Brown University, Computer Science Department