SURF 2015

Rotational Invariant Low-Rank Pose Estimation

Joon Sik Kim
Mentor: Yisong Yue
Co-Mentor : Matteo Ronchi

October 17, 2015

Caltech

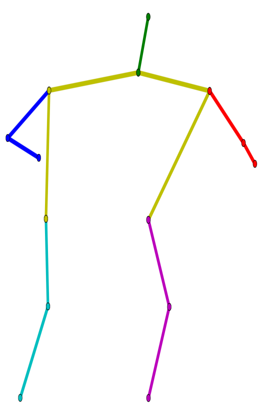
Human poses in images

Leeds Sports Pose Dataset by Sam Johnson and Mark Everingham
[Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation]
In *Proceedings of the 21st British Machine Vision Conference (BMVC2010)*

Questions to ask

How can we train a model to see the similarities among the poses?

How can we compactly express the poses in terms of few "basis" poses that capture all variability of each joints?



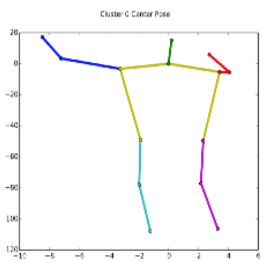
$$= \beta_1 p_1 + \beta_2 p_2 + \ldots + \beta_K p_k = \sum_{i=1}^{K} \beta_i p_i$$

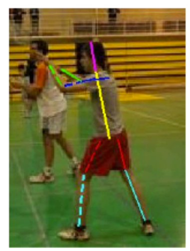
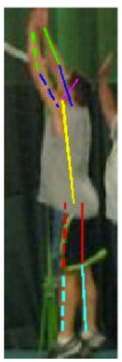
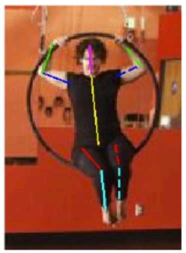
Outline

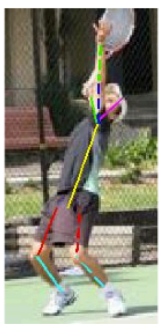
- K-Means Clustering
- Singular Value Decomposition (SVD)
- 2-D Model
  - Problem Formulation
  - Initialization
- 2-D Results
- Conclusion
- Future Work

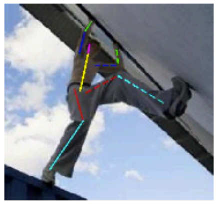
K-Means Clustering

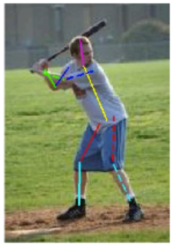
- Clustering based on the distances between the joint coordinates
- Used 8 clusters
- Visualized the mean pose of each cluster with few examples
  - Cluster 0

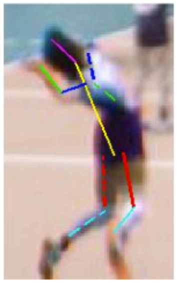




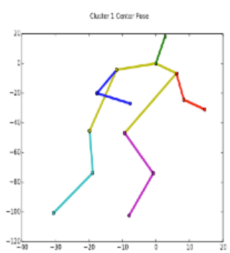


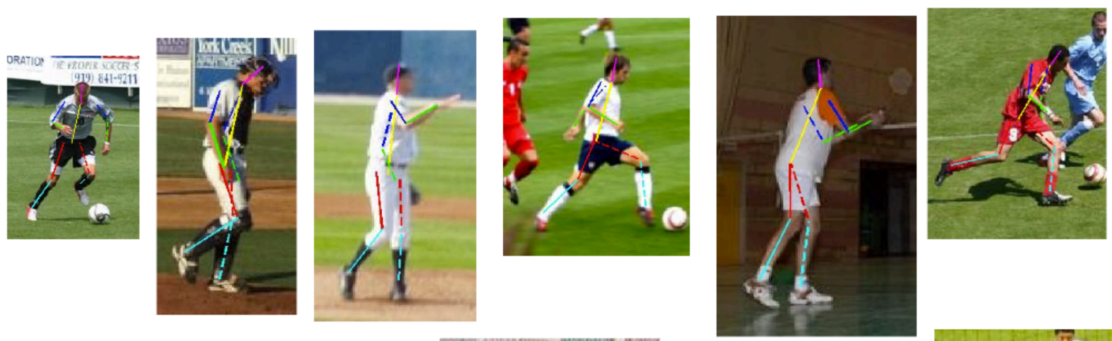


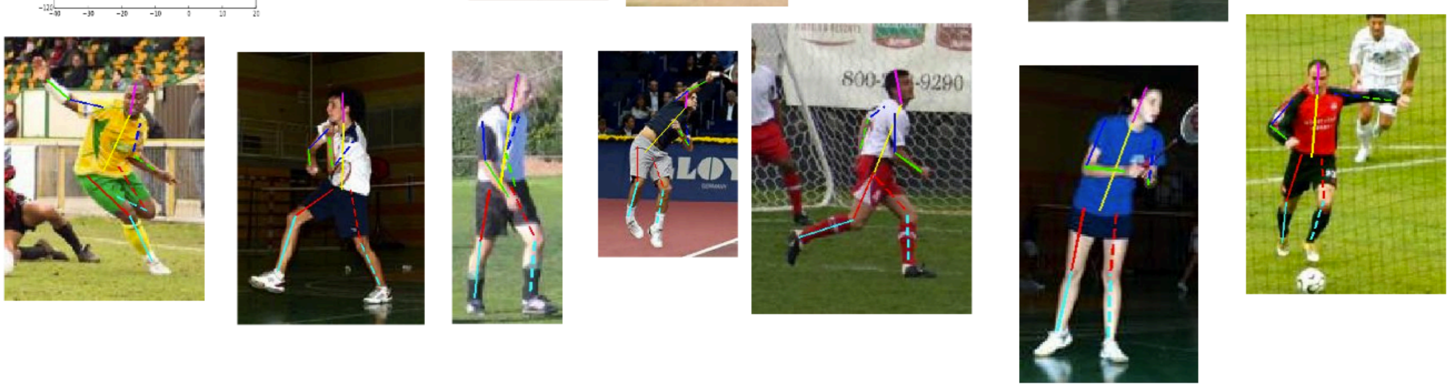
K-Means Clustering - Limitation

- Uncertain about the number of cluster to use for the analysis
- Clusters not fully separated from one another
- Difficult to observe distinctive features from each cluster centers

- Cluster 1







Singular Value Decomposition (SVD)

- Regular SVD solves the optimization problem of

$$\arg\min_{U,V} \sum_i \|X_i - U_i V^T\|^2_{Fro}$$

U : the matrix composed of the basis poses
V : the matrix with coefficients of each basis

- Latent Factor Visualization : plot skeleton figures of the poses

$$U_i' = \mu \pm \beta U_i$$

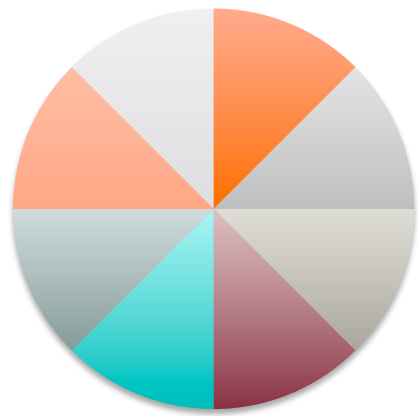
SVD - Visualization

➔ Rotations observed within the basis pose

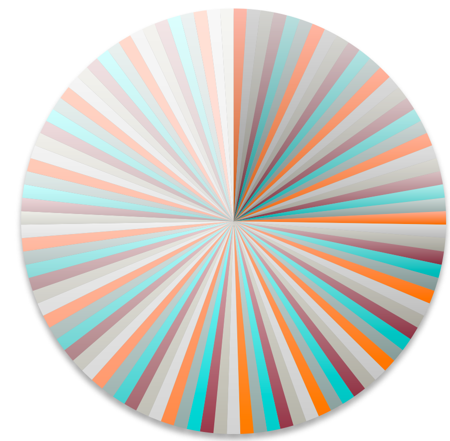
SVD + Viewing Angle ?

Original SVD

Clustering of Angle



More Clusters



- Disambiguate poses from viewing angle
- Learn a sequence of pose bases that describe how the basis poses vary across viewing angle
- Perform SVD on each viewing angle separately

2-D Model - Optimization

- With the clustered poses, solve the following :

Regularization

$$\arg\min_{U,V,\theta} \sum_{a=1}^{p} \left[\lambda \left(\frac{\|U^{(y)}\|^2}{p} + \|U^{(x)}(a)\|^2 + \|V^T(a)\|^2 + \sum_{a'} \kappa_{a,a'} \|U^{(x)}(a) - U^{(x)}(a')\|^2 \right) + \sum_i \sum_{j \in S^{(a)}} (S_{ij}^{(a)} - U_i(a) V_j^T(a))^2 \right]$$

Spatial regularization

Reconstruction Error term

  - Spatial regularization : ensures smooth transitions of poses across different clusters

➔ Use Stochastic Gradient Descent (SGD)

2-D Model – Initialization (1)

Pose information

Compute Metric

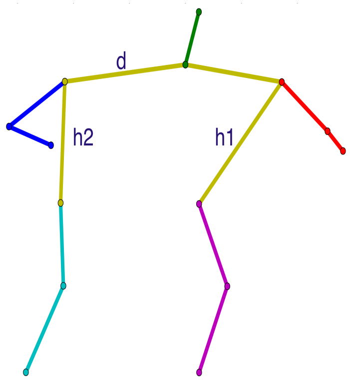
- cluster0
- cluster1
- cluster2
- cluster3
- cluster4
- cluster5
- cluster6
- cluster7

Assigned based on metric

2-D Model – Initialization (2)

Metric for estimating angle view

- Width-height ratio of the body (general magnitude of the rotation)

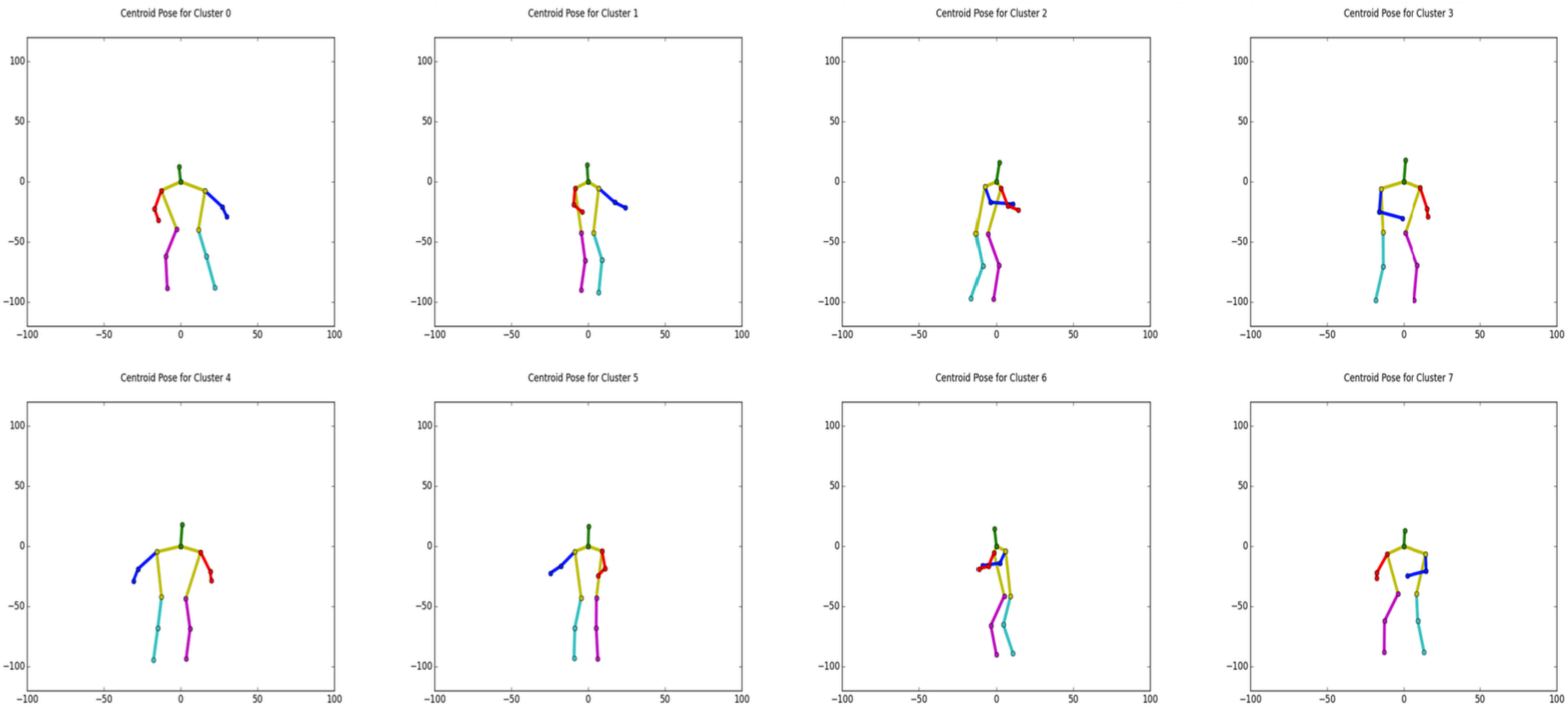


$$\frac{d}{\frac{h_1+h_2}{2}}$$

- Location of right hand w.r.t. right shoulder (the direction pose is facing)

2-D Model – Initialization (3)

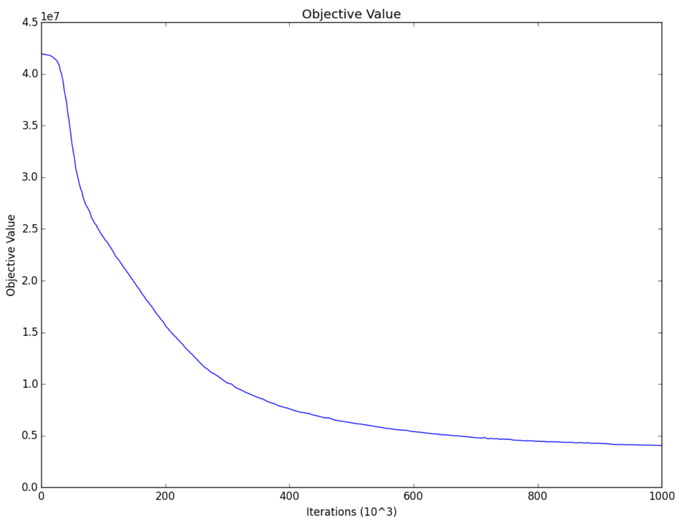
- Clustering of poses based on the angle view



Caltech

2-D Model Results

Minimization of Objective Function



Few examples of pose movements

Conclusion

- Preliminary Findings
  - Some (relatively) rotational invariant basis poses
  - Some rotations exist, primarily due to the small number of clusters
- Improvements
  - Fine-grained clustering at initialization
  - Computational cost of increased number of clusters
  - Other parameter testing

Future Work

- Refine 2-D model
- Ultimate goal: 3-D model

$$\arg\min_{U,V,\theta} \sum_i (S_i - f(UV_i^T, \theta_i))^2$$

  - Reconstructing a 3-D pose information from 2-D images
  - Viewing angle is embedded in to the problem itself

Acknowledgements

- Donor: Samuel P. and Frances Krown
- Mentor: Professor Yisong Yue
- Co-mentor: Matteo Ronchi
- Caltech Computer Vision Lab
- Caltech SFP

Thank You!