Learning a Non-Parametric Articulated Pedestrian Representation

Chris Stauffer, Lily Lee & Eric Grimson

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: Modeling highly articulated objects such as people is an extremely difficult problem in computer vision. This work investigates an alternative to the two extremes of using explicit models and using "exemplars." Rather than fitting an articulated model to a silhouette or treating the silhouettes holistically, this work explains how to derive a sparse basis to represent the observed silhouettes. The coefficients of the basis vectors indicate the presence of a body part at a particular location.

Motivation: The advantage of this representation is the ability to exploit structure in the coefficient values. We can factor the basis into "functional groups" representing particular limbs in different locations. We are investigating how a Boltzmann Machine can be used to constrain coefficient vectors to the manifold of "valid" silhouettes. This factorable, sparse representation holds promise for generative modeling and activity recognition.

Previous Work: Articulated models have become vastly more useful in the recent past. Unfortunately, model specification, model initialization, and tracking stability have limited their use for general applications. Constraining the models to the manifold of "valid" positions is difficult, although it has been done for specific types of actions [2].

Image-based approaches to articulated tracking have been severely limited. Using correlation between silhouettes, some gait work has been accomplished[3]. Other gait recognition work have used descriptions of motion in regions of a walking silhouette[6], but these regions are decided in arbitrary manners and do not strongly correspond to our intuitive understanding of functional groups of a human body. Using exemplars has been effective for certain classes of activities [1], but is not factorable and not compact.

Approach: Our approach involves first deriving a sparse basis to represent the motion silhouettes. To prove the general concept we created a dataset of silhouettes of an articulated figure with each of its five random independently moving limbs. Figure 1 shows an example subsequence.

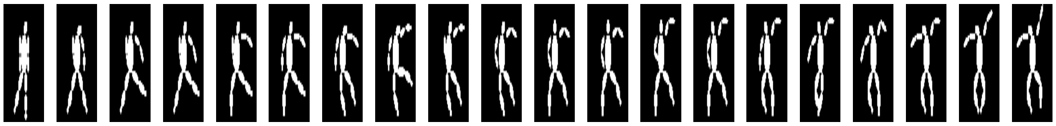


Figure 1: This figure shows an example sequence of the synthetically generated silhouettes.

After normalizing the silhouettes so that each pixel's variance was uniform, we use NMF[5][2] to determine 100 basis vectors representing different body parts in different locations. After converging, we have the basis shown in Figure 2a. This is a sparse, localized basis for representing the silhouettes.

As was stated earlier, one of the advantages of this representation is that the statistics of the coefficients tell us a lot about the silhouettes. One such advantage that is particularly true of our dataset is the different limbs are independent. To exploit this fact, we have done a binary, hierarchical decomposition of the basis to find groups of basis vectors which have the minimum correlation. This decomposition performed as in [7]. These groups of basis vectors correspond to sets basis vectors representing particular limbs in different locations. Figure 2b shows the corresponding (negative) correlation matrix. The four groups of basis vectors correspond to the two arms and the

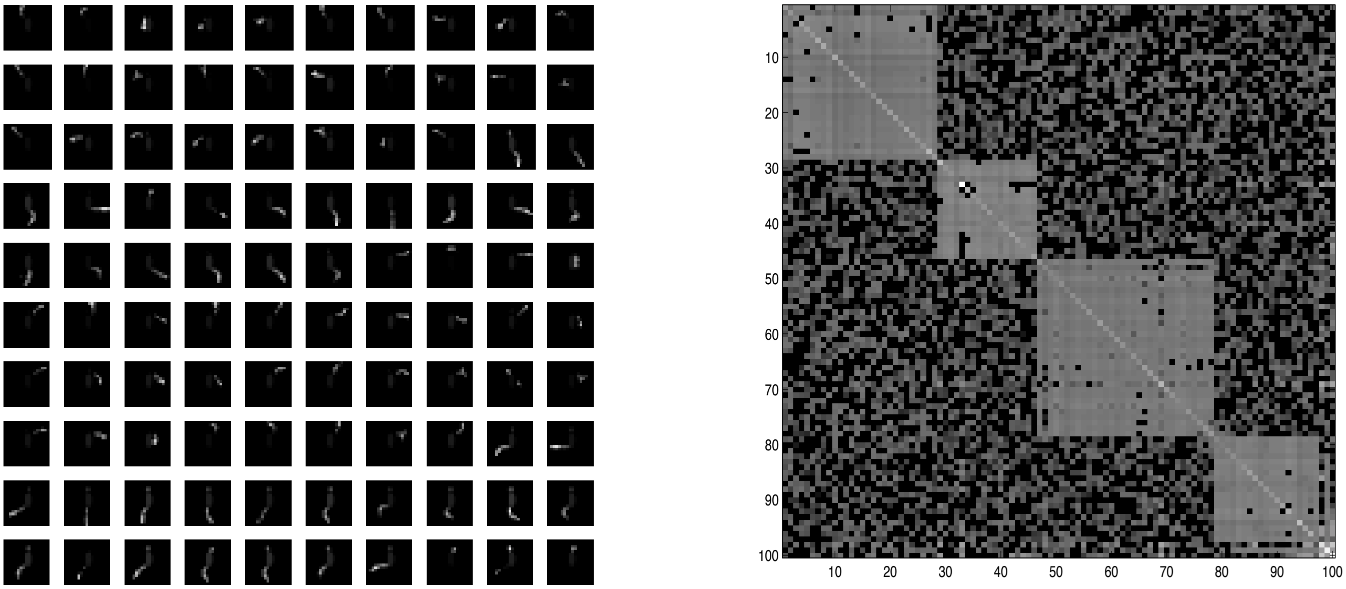
[2] This is closely related to Thomas Hofmann's "aspect model."[4]

two legs that were segmented at the second level of the decomposition hierarchy.

Impact: This work allows a descriptive basis for articulated motions to be automatically derived from simple tracking data. Further, this basis holds promise for both generative modeling of human activities and recognition of classes of human activities.

Future Work: We seek to further develop this work to determine what types of bases are derived for different types of human and non-human activities (e.g., tennis, karate, construction vehicles). Also, we will investigate generative models and articulated activity recognition.

Research Support: This work has been funded by DARPA under contract number N00014-00-1-0907 administered by the Office of Naval Research.



(a) (b) .

Figure 2: (a) is the basis of 100 features derived from the data. (b) is the corresponding (anti)correlation matrix. The elements are sorted based on the hierarchical decomposition and correspond to sets of features representing the four independent body components.

References:

[1] A. Bobick and J. Davis. The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3), March 2001.

[2] Matthew Brand. Shadow puppetry. In *Proceedings of the International Conference on Computer Vision*, Kerkyra, Greece, 1999.

[3] Ross Cutler and Larry Davis. Robust real-time periodic motion detection, analysis, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):781 –796, 2000.

[4] Thomas Hofmann. Probabilistic latent semantic analysis. *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI'99)*, 1999.

[5] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[6] L Lee. Gait dynamics for recognition and classification. Technical Report AIM-2001-019, MIT AI Lab Memo, Sept. 2001.

[7] Chris Stauffer and Eric Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, Aug 2000.