Adaptive training of Hidden Markov Models for stylistic walk synthesis

Joëlle Tilmanne*
University of Mons

Thierry Dutoit
University of Mons

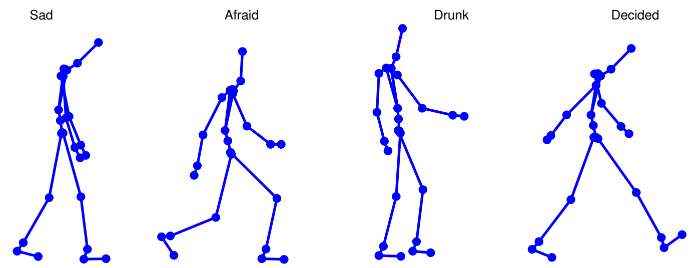

Figure 1: *Postures taken from four different synthesized styles (from left to right: sad, afraid, drunk and decided walks)*

1 Introduction

In this extended abstract, we present the use of Hidden Markov Models (HMMs) in order to synthesize walk sequences with a given style using a small amount of training data from the target style. As a first step, a general model of walk is built. Starting from that model, an adaptive training enables to adapt our model to any particular style using only a small amount of training data. This technique, which was originally developed for speaker adaptation in speech synthesis [Zen et al. 2007], enables to reduce the main problem of machine learning techniques which is the large amount of data needed to train each new model, and to adapt models to the exaggerated style variations of our database that were far from an average walk.

2 HMMs for motion synthesis

Numerous kinds of HMMs have been used to study motion at different complexity levels, whether it was for classification, modeling or synthesis purposes. Some researchers have integrated different forms of "style" variable into their HMMs. Unfortunately there is still no style parameter that can be used to synthesize styles that are very different from the motions on which the system was trained. Increasing the number of styles that can be synthesized means increasing the complexity of the model and re-training the whole model, with the additional problem that enough data must be gathered for each different style.

3 Our Approach

Our approach consists in first building one simple left-to-right HMM for both walk steps (left and right feet) from a large database as normally required. That average model can then be used as a basis for the adaptive training of any style-specific model. A new style-adapted model can thus be build very easily. Compared to existing walk synthesis methods, it only requires a few motion capture steps of the desired walk style and running the adaptive training.

For our database we captured an actor performing walk under eleven different states-of-mind, using an inertial motion capture suit (IGS-190 from Animazoo). Our "style" component consists thus in exaggerated variations that can be far from a plain walk. We chose to model rotations of the 18 captured joints rather than 3D positions in order to insure that the limb length constraints were respected in the synthesized motion. Our angles were converted into the exponential map parameterization which is locally linear and where singularities can be avoided.

For our HMM synthesizer, we adapted the functions implemented for speech within the HMM-based Speech Synthesis System (HTS) to build our procedure [Zen et al. 2007]. We took into account the dynamic aspect of the data, the modeling of the time spent in each state of the HMM (hidden semi-Markov models), and the global variance of the training data to avoid over-smoothed results. The adaptive training is performed based on constrained maximum likelihood linear regression (CMLLR). We adapted thus the latest results of the HMM speech synthesis field to our motion synthesis problem.

Once our adapted model is built, we can synthesize as many walk sequences as we want. The model gives us joint angles and the displacement of the skeleton can be computed using our knowledge of the limb lengths and the step part in which we are.

4 Conclusion

Our method gave very convincing synthesized walk sequences where the styles can easily be recognized (some examples of synthesized motion sequences can be found at `http://tcts.fpms.ac.be/~tilmanne/`). Future work will include an user evaluation to assess the naturalness of the produced motions, and generalization of the use of the style adaptation matrices.

Acknowledgements

This project was partly funded by the Ministry of Région Wallonne under the Numediart research program (grant N0716631) and by the "Fonds pour la formation à la recherche dans l'industrie et l'agriculture (FRIA)".

References

ZEN, H., NOSE, T., YAMAGISHI, J., SAKO, S., BLACK, T. M. A. W., AND TOKUDA, K. 2007. The HMM-based Speech Synthesis System (HTS) Version 2.0. In *Proceedings of the 6th ISCA Workshop on Speech Synthesis, Bonn, Germany*, 294–299.

*e-mail: joelle.tilmanne@umons.ac.be

SIGGRAPH 2011, Vancouver, British Columbia, Canada, August 7 – 11, 2011.
ISBN 978-1-4503-0921-9/11/0008