Human Body-parts Tracking for Fine-grained Behavior Classification

Norimichi Ukita
Nara Institute of Science and Technology
ukita@ieee.org

Atsushi Nakazawa
Kyoto University
nakazawa.atsushi@i.kyoto-u.ac.jp

Abstract

This paper discusses the usefulness of human body-parts tracking for acquiring subtle cues in social interactions. While many kinds of body-parts tracking algorithms have been proposed, we focus on particle filtering-based tracking using prior models, which have several advantages for researches on social interactions. As a first step for extracting subtle cues from videos of social interaction behaviors, the advantages, disadvantages, and prospective properties of the body-parts tracking using prior models are summarized with actual results.

1. Introduction

Video analysis is popular and the most important tool to analyze human behavior in varieties of research field. In autism spectrum disorder (ASD) research, it is reported that autistic individuals shows specific behaviors such as stereotypy, compulsive behavior, sameness, ritualistic behavior, restricted behavior and self-injury[1]. It is also reported that many patients have motor signed such as poor motor planning or toe walking[2]. We believe these anormal behaviors can be automatically found by using computer vision / machine learning frameworks. For example, Wang and their colleagues introduced a machine learning idea to find repetitive structures from a video[3], in particular to find the scene of social games played with parents and infants. In body-parts tracking studies, priors in human body configurations contribute to robust and accurate estimations in human body joint positions/angles[4]. These priors show the probability of human body configurations and obtained through mathematical models such as GPLVM[7] and GPDM[8]. Body-parts tracking using those prior models has advantages for understanding social interactions; 1) priors allow us to track body parts robustly to occlusion caused by the interactions, 2) latent probabilistic models representing the priors show subtle but fundamental differences between different kinds of behaviors, which are crucial for fine-grained behavior classification, and 3) particle filtering in the low-dimensional latent spaces enables

real-time tracking required for online feedbacks in order to control the interactions.

In this work, we tackle the problem to obtain the smooth trajectories of body parts in videos by using prior models. Other promising approaches for body-parts tracking are also mentioned at the end of the paper.

2. Parts Tracking and Behavior Classification

Gaussian Process Dynamical Models (**GPDM**)[8] provide us dimensionality reduction and temporally-smooth transition in the low-dimensional latent space. Figure 1 shows the latent models of two actions, namely walking and jogging. Inherence of the GP allows us to optimize the latent space increasing its generalization and conformity with human body structure and kinematics. Since the latent space represents the fundamental poses and motions, communality among similar actions can be modeled; in Fig. 1, the trajectories of the two actions are almost parallel circles. While the communality is represented, subtle difference among them is also modeled; in Fig. 1, the distance between the two trajectories is larger than temporally-subsequent points in each trajectory.

GPDM with a D-dimensional observation space Y, which is inherently nonlinear, and its d-dimensional latent space X is defined by two mappings; 1) from a point at t to a point at $t+1$ in the latent space, $f_D(\boldsymbol{x})$ where $\boldsymbol{x} \in X$, and 2) from the latent space to the observation space, $f_O(\boldsymbol{x})$. The former mapping gives us the capability of prediction and is useful for human motion tracking.

The extension of GPDM (i.e. GPMGM[4]) was applied to motion prior in human pose tracking. Pose tracking was achieved by image-to-pose regression with particle filtering.

In the learning process, pose data (i.e. joint angles) at each frame is captured with its respective image features (shape contexts[10] in our experiments). for learning an image-to-pose regression function. The motion prior is obtained from the temporal pose data.

In the tracking process, the latent variable of a current pose is estimated by particle filtering with motion prior in the latent space. The current pose is then inferred by pose regression from the estimated latent variable.
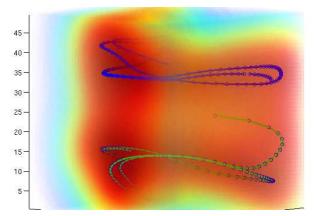
Figure 1. Probabilistic latent models obtained from walking and jogging motions.
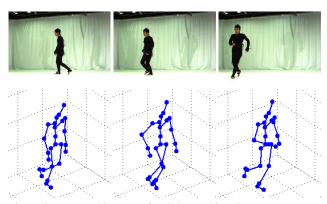

Figure 2. Temporal images of walking and jogging sequences (upper row) and the results of body pose tracking (lower row).

Figure 2 shows examples of tracking results. While no quantitative and comparative results are shown due to page limitation, it can be seen that the results are plausible visually. In [4], RSM error of joint positions was around 3cm. We have also confirmed that image-to-pose regression allows us to estimate the configuration of body parts robustly to heavy occlusion in the previous work[9]. Possibility of behavior classification with tracking in the latent spaces has been also demonstrated in [14].

3. Concluding Remarks

This paper discussed the usefulness of human body-parts tracking for acquiring subtle cues in behavior classification. In this paper, we show the possibility to apply the regression techniques of human pose space for the anomaly behavior detection. We are now trying to apply this technique for the MMDB dataset[5, 6] and try to construct the posture prior models of infants, then hope to extend for finding early alert of ASD patient.

We would also explore other approaches for body-parts tracking. For example, even in cluttered environments (i.e. not in organized experimental rooms), body-parts detection by deformable part models[11] work well. The authors have extended the models so that connectivity of parts is evaluated with image features for more robust detection[12]. Detected results of all frames in videos can be connected temporally by a globally-optimal manner[13]. This kind of global optimization might acquire better tracking results rather than online tracking such as the one shown in this paper, though global optimization cannot work in real time.

References

[1] K. S. Lam, M. G. Aman, "The Repetitive Behavior Scale-Revised: independent validation in individuals with autism spectrum disorders," *Journal of Autism and Developmental Disorders*, Vol.37, No.5, pp.855–866, 2007. 1

[2] D. H. Geschwind, "Advances in autism," *Annual Review of Medicine*, Vol.60, pp.367–380, 2009. 1

[3] P. Wang, G. D. Abowd, and J. M. Rehg, "Quasi-Periodic Event Analysis for Social Game Retrieval," *ICCV*, 2009. 1

[4] N. Ukita and T. Kanade, "Gaussian Process Motion Graph Models for Smooth Transitions among Multiple Actions," *CVIU*, Vol.116, No.4, pp.500–509, 2012. 1, 2

[5] J. M. Rehg, G. D. Abowd, A. Rozga, M. Romero, M. A. Clements, S. Sclaroff, I. Essa, O. Y. Ousley, Y. Li, C. Kim, H. Rao, J. C. Kim, L. L. Presti, J. Zhang, D. Lantsman, J. Bidwell, and Z. Ye, "Decoding Children's Social Behavior," *CVPR*, 2013. 2

[6] The Multimodal Dyadic Behavior Dataset, http://cbi.gatech.edu/mmdb/ 2

[7] N. D. Lawrence, "Probabilistic non-linear principal component analysis with Gaussian process latent variable models," *Journal of Machine Learning Research*, Vol.6, pp.1783–1816, 2005. 1

[8] J. M. Wang, D. J. Fleet, A. Hertzmann, "Gaussian Process Dynamical Models for Human Motion," *PAMI*, Vol.30, No.2, pp.283–298, 2008. 1

[9] N. Ukita, M. Hirai, and M. Kidode, "Complex Volume and Pose Tracking with Probabilistic Dynamical Models and Visual Hull Constraints," *ICCV*, 2009. 2

[10] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *PAMI*, Vol.24, No.4, pp.509–522, 2002. 1

[11] P. F. Felzenszwalb, R. B. Girshick, D. A. McAllester, and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models," *PAMI*, Vol.32, No.9, pp.1627–1645, 2010. 2

[12] N. Ukita, "Articulated Pose Estimation with Parts Connectivity using Discriminative Local Oriented Contours," *CVPR*, 2012. 2

[13] H. Pirsiavash, D. Ramanan, and C. Fowlkes, "Globally-Optimal Greedy Algorithms for Tracking a Variable Number of Objects," *CVPR*, 2011. 2

[14] K. Morimoto, Y. Matsuyama, and N. Ukita, "Continuous Action Recognition by Action-specific Motion Models," *MVA Conference*, 2013. 2