Image and Vision Computing 28 (2010) 836–849

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Discriminative human action recognition in the learned hierarchical manifold space

Lei Han *, Xinxiao Wu, Wei Liang, Guangming Hou, Yunde Jia

Beijing Laboratory of Intelligent Information Technology, Beijing Institute of Technology, Beijing 100081, PR China

ARTICLE INFO

Article history:
Received 28 February 2009
Received in revised form 30 June 2009
Accepted 2 August 2009

Keywords:
Human action recognition
Discriminative model
Hierarchical manifold learning
Mutual invariant
Motion pattern

ABSTRACT

In this paper, we propose a hierarchical discriminative approach for human action recognition. It consists of feature extraction with mutual motion pattern analysis and discriminative action modeling in the hierarchical manifold space. Hierarchical Gaussian Process Latent Variable Model (HGPLVM) is employed to learn the hierarchical manifold space in which motion patterns are extracted. A cascade CRF is also presented to estimate the motion patterns in the corresponding manifold subspace, and the trained SVM classifier predicts the action label for the current observation. Using motion capture data, we test our method and evaluate how body parts make effect on human action recognition. The results on our test set of synthetic images are also presented to demonstrate the robustness.

1. Introduction

Human action recognition has been an active research area over two decades [1–3] and recently has attracted increasing interest from computer vision and pattern recognition community. In particular, it has a wide range of applications, e.g. intelligent interface, digital entertainment, surveillance and security system [4,5].

Most researches on human action recognition mainly aimed at daily human actions such as walking and jumping [6], sitting and lying [7]. For either the entire human body or most body parts, these actions are obviously different in both 2D appearances and 3D joint positions. However, many human actions are only distinct in one or few body parts, and they are difficult to distinguish from each other only using the whole human body. In this paper, we will concentrate on the recognition of these subtle actions, such as walking, marching, dribbling a basketball with right arm, walking with stiff arms, walking with wild legs. The recognition of subtle actions has many promising applications, such as sport motion analysis, digital entertainment and virtual reality, where the reliable poses of human bodies are available.

Generally, there are two crucial problems involved in human action recognition. One is extracting useful and compact motion features, and the other is modeling reference motion. Feature extraction involves two process levels: low-level feature extraction and high-level feature representation. We do not pay much attention to the low-level feature extraction and present results on publicly available motion capture data as well as on the joint position sequences estimated from datasets of synthetic images. Even if 3D positions are available, action recognition is also difficult due to the high dimensionality of pose space which not only increases computational complexity but may also hide key features of actions. Fortunately, the subset of allowable configurations in original pose space is usually restricted by human biomechanics, and human motion can be expected to lie on a low-dimensional manifold. In this paper, we use hierarchical latent variable space analysis to discover a hierarchical manifold space of human motion. Moreover, the mutual invariant features (i.e. motion patterns) are extracted from each manifold subspace.

Human actions evolve dynamically over time. They can happen at various timescales and may exhibit long-range dependencies. Furthermore, the transition between simple actions naturally has ambiguity of temporal segments and overlapping. One of the most common approaches for human action recognition is to use Hidden Markov Model (HMM) or its variants [7,8]. However, a strong assumption that observations are conditionally independent is usually made in such generative models in order to ensure computational tractability. The restriction makes it difficult or impossible to accommodate long-range dependencies among observations or multiple overlapping features of observations at multiple time steps. Conditional random fields (CRFs) use an exponential distribution to model the entire sequence given the observation sequence [6,9–12], which avoids the independence assumption between observations and is able to incorporate both overlapping features and long-range dependencies into the model. CRF and most of its variants can estimate a label for each observation, so

* Corresponding author. Tel.: +86 10 68940955.
E-mail addresses: hanlei@bit.edu.cn (L. Han), wuxinxiao@bit.edu.cn (X. Wu), liangwei@bit.edu.cn (W. Liang), jiayunde@bit.edu.cn, gmhou@bit.edu.cn (Y. Jia).

0262-8856/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2009.08.003