Signal Processing 142 (2018) 223–230

Contents lists available at ScienceDirect

Signal Processing

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

Domain adaptation with low-rank alignment for weakly supervised hand pose recovery

Chaoqun Hong*, Zhiqiang Zeng, Rongsheng Xie, Weiwei Zhuang, Xiaodong Wang

School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, 361024, China

A R T I C L E I N F O

Article history:
Received 26 February 2017
Revised 23 July 2017
Accepted 25 July 2017
Available online 26 July 2017

Keywords:
Human hand pose recovery
Neural network
Domain adaptation
Low-rank representation

A B S T R A C T

Human hand pose recovery (HPR) in depth images is usually conducted by constructing mappings between 2D depth images and 3D hand poses. It is a challenging task since the feature spaces of 2D images and 3D poses are different. Therefore, a large number of labeled data is required for training, especially for popular frameworks such as deep learning. In this paper, we propose an HPR method with weak supervision. It is based on neural network and domain adaptation is introduced to enhance the trained model. To achieve domain adaptation, we propose low-rank alignment, which aligns the testing samples to the distribution of labeled samples. In this process, autoencoders are used to extract 2D image features and low-rank representation is used to describe this feature space. Therefore, the proposed method is named as Domain Adaptation with Low-Rank Alignment (DALA). In this way, we obtain a robust and non-linear mapping from 2D images to 3D poses. Experiments are conducted on two challenging benchmark datasets MSRA and ICVL. Both the results on a single dataset and across datasets show the outstanding performance of DALA.

1. Introduction

Researchers always try to find some natural Human-Computer Interaction methods. Among them, interaction with hand poses has been proven to be a good candidate. To recognize hand pose, recovery should be achieved first. However, hand pose recovery is a challenging task, especially for marker-free scenarios. For example, Leap Motion uses two monochromatic IR cameras and three infrared LEDs. In this way, the device observes a roughly hemispherical area, to a distance of about 1 m and hand poses can be captured. Leap motion has led a good direction but it still requires an additional device that is not commonly used.

Among the marker-free methods, recovering the poses of human hands from depth images has attracted plenty of attention due to its convenience and low demand on equipments [1–3]. However, its difficulty arises because articulable hands typically have many degrees of freedom (DOF), constrained parameter spaces and self-similar parts [4–7]. All these factors make it difficult to directly fit a model to the depth images [8]. A number of hand pose recovery (HPR) methods have been developed, which can be categorized into two main groups:

(1) Hand parts detection-based HPR. These methods [9,10] use a realistic hand model to generate large synthetic datasets and train a classification model to assign a hand part label to each pixel. The centers of the hand parts are then estimated to form the hand skeleton;

(2) Hand joints detection-based HPR. These methods [11,12] aim to directly predict the location of the joints from depth images. The inverse kinematics (IK) method [13] can then be used to reconstruct the skeleton.

The above-mentioned methods can be all treated as example-based methods or data driven methods. They require a large number of labeled samples to achieve HPR. However, it is not always available. Therefore, we propose a method for weakly supervised hand pose recovery. It trains the mapping model with a partially labeled sample and improves the model by aligning the testing samples to the distribution of training samples. The contributions of this paper are:

- The key contribution is domain adaptation for weakly supervised hand pose recovery. Both the training samples and testing samples are represented in a unified space and aligning parameters are computed in this space. These parameters are used to align the testing samples to the domain of training samples.
- The second contribution is domain adaptation with low-rank representation. Low-rank representation is sparse and the distributions of training samples and testing samples can be

* Corresponding author.
E-mail address: cqhong@xmut.edu.cn (C. Hong).

http://dx.doi.org/10.1016/j.sigpro.2017.07.032
0165-1684/

observed clearly. In this way, the process of alignment can be achieved in low-rank feature space.

- The third contribution is the mapping between 2D depth images and 3D hand poses are computed by a neural network with 2 hidden layers. In this way, their relationship is described on a non-linear manner.

2. Related works

2.1. Domain adaptation

Assume we have two sets of data: a source domain S providing labeled training instances and a target domain T providing instances on which the classifier is meant to be deployed. Data in the source and the target are often distributed differently. S is drawn from a distribution $p(S)$, while T is drawn from a distribution $p(T)$. The learning problem consists in finding a function realizing a good transfer from S to T i.e. it is trained on data drawn from $p(S)$ and generalizes well on data drawn from $p(T)$ [14–17].

Domain adaptation considers the setting in which the training and testing data are sampled from different distributions as mentioned above. Learning setups relating to domain adaptation have been proposed before and published under different names. Daume III and Marcu [18] formalized the problem and proposed an approach based on a mixture model. A general way to address domain adaptation is through instance weighting, in which instance dependent weights are added to the loss function [19]. Another solution to domain adaptation can be to transform the data representations of the source and target domains so that they present the same joint distribution of observations and labels. Ben-David et al. [20] formally analyze the effect of representation change for domain adaptation while Blitzer et al. [21] propose the Structural Correspondence Learning (SCL) algorithm that makes use of the unlabeled data from the target domain to find a low-rank joint representation of the data.

Finally, domain adaptation can be simply treated as a standard semi-supervised problem by ignoring the domain difference and considering the source instances as labeled data and the target ones as unlabeled data [22]. In that case, the framework is very close to that of self-taught learning [23], in which one learns from labeled examples of some categories as well as unlabeled examples from a larger set of categories. Recent work has investigated several techniques for alleviating the difference: instance reweighting [24,25], sub-sampling from both domains [26] and learning joint target and source feature representations [27,28].

2.2. Low-rank representation

Low Rank Representation(LRR) [29,30] is a recently proposed spectral clustering based method for subspace clustering. It seeks the lowest-rank representation of the data samples. The model of LRR in the noiseless case is:

$$min_z \parallel Z \parallel_*, \ \ s.t. \ X = XZ, \qquad (1)$$

where the nuclear norm $\|.\|_*$ serves as a convex surrogate of the rank function. X is the set of samples to be represented and Z is the set of low-rank coefficients. It is shown by Liu et al. [29] that when the data are noise free and drawn from independent subspaces, the optimal solution to problem (1) is also block diagonal.

2.3. Manifold alignment

The fundamental ideas of manifold alignment are to utilize the relationships of instances within each dataset to strengthen knowledge of the relationships between the datasets and ultimately to map initially disparate datasets to a joint latent space [31]. At the algorithmic level, the approaches described in this paper assume that the disparate datasets being aligned have the same underlying manifold structure. The underlying low-dimensional representation is extracted by modeling the local geometry using a graph Laplacian associated with each dataset. After constructing each of these Laplacians, standard manifold learning algorithms are then invoked on a joint Laplacian matrix constructed by concatenating the various Laplacians to obtain a joint latent representation of the original datasets. Manifold alignment can therefore be viewed as a form of constrained joint dimensionality reduction where the goal is to find a low-dimensional embedding of multiple datasets that preserves any known correspondences across them [32]. Manifold alignment has been widely used in classification or ranking tasks [33–38].

Manifold alignment is closely related to subspace learning or transfer learning. For example, in [39], transfer learning based on decomposition is used for image classification. It is also widely used in cross-modal or multi-modal learning [40,41].

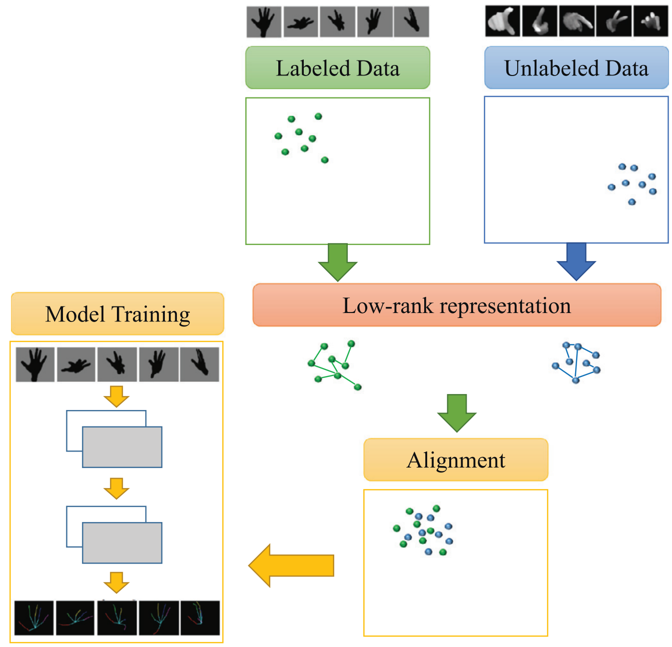

Fig. 1. The flowchart of the proposed method.

3. Domain adaptation with low-rank alignment

3.1. Outline of the proposed method

The proposed method can be summarized in Fig. 1. In the scenario of weakly supervised hand pose recovery, we only have to train the model using examples that are only partially annotated or labeled. Therefore, we need to make use of unlabeled data which have been proven to be useful in the learning task. However, the distributions of labeled data and unlabeled data are usually quite different, which downgrades the performance. Therefore, we make use of domain adaptation to map the unlabeled data to the domain of labeled data. To achieve domain adaptation, we compute the low-rank representations (LRR) of labeled data and unlabeled data. With the low-rank representation, we can find the inherent mapping relationship between labeled data and unlabeled data. In this way, we can map the unlabeled data to the space of labeled data, which is used to improve the performance of hand pose recovery. In this way, the aligned data can be used with the trained model to achieve pose recovery.