報告番号	※甲　　　第　　　号

主　論　文　の　要　旨

論文題目　Example-Based Methods for Estimating 3D Human Pose from Silhouette Image using Approximate Chamfer Distance and Kernel Subspace
(近似Chamfer距離とカーネル部分空間を用いた例に基づくシルエット画像からの人物姿勢推定法)

氏　　名　曹　暉

論　文　内　容　の　要　旨

In this dissertation, we have introduced an effective and efficient example-based approach to recovering 3D human body pose from a single silhouette. Our approach mainly involves two stages: retrieving candidate poses from a large database and re-ranking the candidate poses using kernel subspace.

Research Background

Recovering 3D human pose from a single image remains one of the fundamental problems in computer vision, with potential applications in surveillance, video editing / annotation, human computer interface, and entertainment. The depth ambiguities of 3D-2D projection, part occlusion, clothes variation, and high degrees of freedom of the body pose make this 3D reconstruction problem particularly hard to cope with.

Existing approaches to this problem can be categorized into three groups: Model-based approaches fit a 3D human model to an image by minimizing a cost function; Learning-based approaches directly infer poses from image features depending on a learned parametric model; Example-based approaches store a set of training examples whose 3D poses are known and search examples resembling the input image. However, these approaches become incapable when it is necessary to handle a wide range of 3D human poses and high efficiency is also required. The model-based approaches are time-consuming and sensitive to the initialization of pose. Learning-based approaches are fast but it currently can only deal with a limited set of typical human poses. Example-based approaches may be a good choice for dealing with wide range of 3D human poses, but the issue of high time and memory complexities must be addressed.

Problem Definition

This thesis focuses on estimating 3D body pose from silhouette. We do not make use of rich visual information such as clothes pattern, skin color or face pattern, to segment and label body parts such as the head, torso, thighs, calves and arms. The background clutter problem is avoided, while other possible problems — including pose and appearance variances, self-occlusion and depth ambiguity — will be handled.

As a summary, our goal is to estimate human pose from silhouette in a generic setting of arbitrary pose and arbitrary viewpoint, with assumptions as:

1. The torso of the target is approximately parallel to the imaging plane;

2. There is no serious external occlusion;

3. The whole body is visible.

Contributions

Pose from silhouette is essentially an ill-pose problem due to the high-dimensional state space combined with the unknown factors like parts occlusion, appearance variation and varying view position. In most cases, it is hard (or impossible) to provide a single optimal solution to this problem. The ultimate goal of this dissertation is to provide a few plausible poses estimates which are, hopefully, exactly or close to the ground-truth.

We achieve this goal by two-stage processes. At the first stage, we obtain a certain number of candidate poses by searching over a large database of examples. At the second stage, these candidate poses are re-ranked by kernel subspace projection. Finally, a few top candidates are selected as the pose estimates. These processes are briefly explained below.

Constructing a large database If the input silhouette well matches some examples in the database which is composed of a large number of human silhouettes annotated with the corresponding 3D body poses, the poses are probably close to the input's pose. Intuitively, this idea is capable of dealing with any complex poses if a large variety of examples being preserved in the database. To create a large database, we first created a medium-size source dataset by means of 3D human character rendering software and various human motion capture data. We then created a larger database containing two millions of examples by means of valid half-body combinations.

Approximate Chamfer Matching It is usually infeasible for most of good image matching algorithms to deal with a large database. The chamfer distance has proved to be an efficient and effective tool for shape comparison in many computer vision works. However, applying the chamfer distance to the large-scale matching task is still costly.

To address this computational issue, we propose two approximate methods to accelerate computing chamfer distance.

The first method, which we refer to as eigen-chamfer matching, utilizes eigenspace approximation to distance transform in computing chamfer distance so that the computation of chamfer distance shifts to a low-dimensional subspace. This new method is able to efficiently complete a linear scan to a two million examples database, while the achieved estimation performance is yet competitive to the exact chamfer distance.

The second method, which is referred to as joint-chamfer matching, utilizes a part-to-whole strategy for searching pose candidates. The half-body candidate poses are first retrieved by means of partial chamfer matching, from which valid half-body combinations are picked out subject to the pre-computed combination constraints. The further evaluation on selected combinations is also efficient as it involves a small number of simple arithmetic (addition and minimization) operations on known half-body distances. This method is computationally extremely efficient and the current implementation can work near in real-time.

Kernel Subspace Re-Ranking The image similarity is not optimal and probably inconsistent to the desired pose similarity. In other word, sometimes when query is different to database images due to body size, clothes, *etc*, irrelevant poses may be overestimated while relevant poses may be underestimated. Thus, it is necessary to re-rank the candidate poses in combination with other complemental knowledge.

We propose two kernel subspace ranking methods: KPCA-ranking and KCCA-ranking. Kernel Principal Component Analysis (KPCA) and Kernel Canonical Correlation Analysis (KCCA) are used to learn nonlinear subspaces characterizing the underlying structure of image-pose pairs. Depending on the kernel subspace, candidate poses are ranked based on ranking criteria: subspace projection loss for KPCA-ranking and correlation score for KCCA-ranking. The kernel subspace ranking is complemental to image similarly ranking so that the combination of them perform better than either does alone.

Organization of the Thesis

The outline of this thesis is as follows.

Chapter 1 provides an overview of the work presented in the thesis.

In Chapter 2, we briefly review previous works on human body pose estimation from a single image. Previous works are classified into three groups: model-based, learning-based and examples-based approaches. The model-based approaches are further divided into top-down and bottom-up according to the strategy of exploring the human pose space.

Chapter 3 describes the human pose database and its construction process. we first create a medium-size source dataset of $14,964$ pose examples from the collected human motion capture sequences. Then, we exploit the half-body combination strategy to enlarge the source dataset into a 130 times larger database consisting of 2 million pose examples. The combination constraint between any two half-body poses is calculated based on their body orientation and pose proximities, thus ensuring the allowable combinations valid poses. In addition to explicitly preserving the large database, we also create a compact half-body database comprised of only half-body examples and the combination constraints.

Chapter 4 describes the first stage of our pose estimation approach. For a given input image, we obtain a set of candidate poses from the database. We propose two approximate chamfer matching algorithms to do this work. The eigen-chamfer matching method uses the eigenspace approximation to speedup the chamfer distance. The joint-chamfer matching method first searches half-body candidates by partial shape matching. Further evaluation are carried out on the valid half-body combinations subject to the combination constraints. Additionally, a refined matching step via adding silhouette cue and image normalization method are introduced. Finally, we present and discuss the estimation results for both synthesis and real image dataset.

Chapter 5 describes the second stage of our pose estimation approach. We use the Kernel PCA and Kernel CCA to learn kernel subspaces which can well characterize the nonlinear manifold of image and pose. The ranks by the kernel subspace are combined with the original (image similarly) ranks so that the more likely candidate poses for a given input image can be assigned higher ranks. We present and discuss synthesis experiment and compared the experimental results before and after re-ranking.

Chapter 6 summarizes this work and provides some speculation on future research directions.