Applied Soft Computing 17 (2014) 153–166

Contents lists available at ScienceDirect

Applied Soft Computing

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

Comparing evolutionary algorithms and particle filters for Markerless Human Motion Capture

Enrique Yeguas-Bolivar*, Rafael Muñoz-Salinas, Rafael Medina-Carnicer, Angel Carmona-Poyato

Department of Computing and Numerical Analysis, Maimonides Institute for Biomedical Research (IMIBIC), University of Córdoba, Córdoba, Spain

A R T I C L E I N F O

Article history:
Received 16 November 2011
Received in revised form 7 November 2013
Accepted 10 January 2014
Available online 22 January 2014

Keywords:
Markerless Motion Capture
Tracking
Particle filters
Evolutionary algorithms

A B S T R A C T

Markerless Human Motion Capture is the problem of determining the joints' angles of a three-dimensional articulated body model that best matches current and past observations acquired by video cameras. The problem of Markerless Human Motion Capture is high-dimensional and requires the use of models with a considerable number of degrees of freedom to appropriately adapt to the human anatomy.

Particle filters have become the most popular approach for Markerless Human Motion Capture, despite their difficulty to cope with high-dimensional problems. Although several solutions have been proposed to improve their performance, they still suffer from the *curse of dimensionality*. As a consequence, it is normally required to impose mobility limitations in the body models employed, or to exploit the hierarchical nature of the human skeleton by partitioning the problem into smaller ones.

Evolutionary algorithms, though, are powerful methods for solving continuous optimization problems, specially the high-dimensional ones. Yet, few works have tackled Markerless Human Motion Capture using them. This paper evaluates the performance of three of the most competitive algorithms in continuous optimization – Covariance Matrix Adaptation Evolutionary Strategy, Differential Evolution and Particle Swarm Optimization – with two of the most relevant particle filters proposed in the literature, namely the Annealed Particle Filter and the Partitioned Sampling Annealed Particle Filter.

The algorithms have been experimentally compared in the public dataset HumanEva-I by employing two body models with different complexities. Our work also analyzes the performance of the algorithms in hierarchical and holistic approaches, i.e., with and without partitioning the search space. Non-parametric tests run on the results have shown that: (i) the evolutionary algorithms employed outperform their particle filter counterparts in all the cases tested; (ii) they can deal with high-dimensional models thus leading to better accuracy; and (iii) the hierarchical strategy surpasses the holistic one.

1. Introduction

Markerless Motion Capture (MMOCAP) is an emerging field with potential applications in areas like the animation industry [1], medical rehabilitation [2], and video surveillance [3], among others. The goal is to automatically determine the pose of a subject from the images of a video sequence. For that purpose, a human body model is rendered on the images so as to determine the model's configuration that best matches the input images. The sequential nature of the problem allows us to use the solutions obtained in one frame as the starting point for the search in the next one. In order to cope with three-dimensional ambiguities and occlusions, the most effective setup consists of several cameras simultaneously observing the scene from different points of view. Ultimately, the problem is formulated as a continuous optimization problem that seeks the best angles for the joints of the human body model given the information available in the previous and the current images.

MMOCAP is considered in the computer vision community as a difficult problem because of the many challenges that it presents. First, it is a high-dimensional problem. In order to obtain accurate results, the body model employed needs to have a suitable number of joints so as to adapt properly to all human poses. Second, the computing power is a limiting factor. The operations needed to evaluate solutions are computationally very expensive. Therefore, it is of paramount importance to obtain good solutions in as few iterations as possible. Finally, an appropriate balance is required between local and global search. While in most of the cases local

* Corresponding author. Tel.: +34 957212289.
E-mail addresses: eyeguas@uco.es (E. Yeguas-Bolivar), rmsalinas@uco.es (R. Muñoz-Salinas), rmedina@uco.es (R. Medina-Carnicer), ma1capoa@uco.es (A. Carmona-Poyato).

1568-4946/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2014.01.007

search achieves good solutions, occlusion and ambiguities in the camera configuration might require the use of global optimization to discover the correct solution once the conflicting situation has finished.

Most of the solutions emerging from the computer vision community are particle filtering variants [4–7], which try to cope with the high dimensionality of the model. Some solutions partition the search space taking advantage of the hierarchical nature of the problem [5,6]. Other solutions tackle the problem by dividing the search into multiple stages [4,7]. In spite of the advances achieved, most of the proposed solutions suffer from the curse of dimensionality and they need to rely on simple human models (which lead to suboptimal tracking results), or require a high number of evaluations to achieve accurate results.

Over the past few years, the field of global optimization has been very active, producing different kinds of deterministic and stochastic algorithms for optimization in the continuous domain [8]. Among the stochastic approaches, evolutionary algorithms [9] offer a number of exclusive advantages that make them an attractive choice, e.g., robust and reliable performance, global and local search capability, little or no information requirement, etc. Despite of their advantages, there are few works in the literature devoted to test the performance of evolutionary algorithms on the MMOCAP problem [10–12].

This work performs an experimental comparison of three relevant evolutionary algorithms namely Covariance Matrix Adaptation Evolutionary Strategy (CMAES) [13], Differential Evolution (DE) [14], and Particle Swarm Optimization (PSO) [15], with two popular variants of particle filters proposed for the MMOCAP problem namely Annealed Particle Filter (APF) [4] and Partitioned Sampling Annealed Particle Filter (PSAPF) [6]. The selection of these evolutionary algorithms is based on the results reported in the IEEE International Conferences on Evolutionary Computation (CEC) competitions on real parameter optimization [16], for which these three algorithms have continued to secure top ranking. With regards to the particle filters selected, the first one (APF) can be considered the base line algorithm taking into account the high number of cites obtained. The PSAPF has been selected for being considered the hierarchical version of the APF.

This work analyzes the behavior of the evolutionary algorithms using both holistic and hierarchical strategies so as to decide which strategy is preferable. Additionally, two different articulated models have been employed. The first one, comprised of 28 degrees of freedom (DOF), is similar to those used in most of the previous approaches in the literature. The second one, comprised of 39 DOF, aims at evaluating the capability of the tested algorithms to deal with more realistic (and thus complex) body models.

The experiments have been conducted on nine sequences of the HumanEva-I dataset [17], and analyzed by means of a non-parametric statistical analysis [18]. The results obtained show that: (i) the evolutionary algorithms tested provide significantly better results than particle filters; (ii) the 39-DOF body model employed in this work obtains better results than the other one; and (iii) the hierarchical strategy surpasses the holistic one in all the evolutionary algorithms applied. Additionally, our experiments determine the minimum number of evaluations required for an optimal trade-off between precision and speed for the different algorithms in the experiments.

The remainder of this paper is structured as follows. Section 2 revises the related work. Section 3 formulates the problem of pose estimation, describes the body models, cost function and evaluation strategies. Section 4 introduces the basis of the particle filters and the evolutionary algorithms employed. Finally, Section 5 explains the experimentation carried out, while Section 6 draws some conclusions.

2. Related works

Most of the solutions proposed for the MMOCAP problem fall into two main categories: holistic and hierarchical approaches. The former employs a global optimization approach to fit the model's parameters. As a consequence, the computational requirements are high and the search method may suffer from premature convergence because of the high dimensionality. On the other hand, hierarchical strategies exploit the underlying hierarchical structure of the articulated model assuming that body parts can be localized independently from each other. Then, the problem is divided into smaller problems that can be more easily solved. However, hierarchical approaches have several drawbacks. First, the optimal partitioning may not be obvious (e.g. arms first or legs first?) and it may change over time (e.g. when the legs cross during walking). Second, while partitioning might not make a difference for noise-free simulation data, in practice an incorrectly estimated early partition due to noisy data may irrevocably mislead the outcome [6].

2.1. Computer vision approaches

The first solutions emerging from the computer vision community consist in the use of particle filters. In particular, the Condensation algorithm is the most prevalent of such algorithms and has been widely employed for the tracking task [19]. However, when applied to the MMOCAP problem, it has been repeatedly shown that it suffers from the *curse of dimensionality*. Therefore, Deutscher and Reid proposed the APF [4], which combines the ideas of the Condensation and the Annealed Search so as to improve the tracking results. With the aim of reducing the complexity of the search in many dimensions, the particles are evaluated in layers. In the first layers, the objective function is smoothed so as to allow escaping from local minima. In the last layers, the objective function is more peaked thus concentrating on exploiting the best solutions. Corazza et al. propose also a custom version of adapted fast simulated annealing [20] for body tracking using as input data a visual hull reconstruction and an a priori model of the subject.

Another popular approach for tracking articulated objects is the use of Partitioned Sampling (PS) [5]. The technique was initially employed for tracking several objects using particle filters, but then successfully applied to hand tracking. Unlike the APF, PS imposes a strong partition of the search space. As previously indicated, the main problem consists in determining the optimal partitioning. Bandouch et al. proposed the PSAPF [6] as an attempt to combine the strengths of PS and the Annealed Search. To do so, they incorporate the APF within a PS framework by applying an appropriate weighted resampling in each sub-space. As they report, they are able to cope with high-dimensional models, but at the cost of employing a very high number of evaluations per frame (around 8000). Recently Gall et al. [7] have proposed a multi-layer framework that combines stochastic optimization, filtering, and local optimization search. Their approach runs on several stages. First, a simulated annealing is employed. And then, the solutions obtained are refined by filtering and local optimization. Although they report good tracking results, the main drawback of their approach is the limitation imposed to the body model tested (only 28 DOF).

Chang and Lin [21] proposed a progressive particle filter. The main idea was to combine a standard particle filter with the mean shift strategy [22] and a hierarchical search. However, their solution was only tested on video sequences with a single camera in a non-public dataset. It is not clear if their approach would work on sequences with multiple views.

In spite of the advances achieved over the last years, mobility limitations often are imposed to the body models employed so as to obtain reasonable performance in manageable computing times.