2018-00663 - Deep Neural Networks for End-to-end Human 3D Shape Retrieval in Single Images [PhD Campaign, Campagne Doctorants Centre recherche Inria Rhône-Alpes]

**Type de contrat :** CDD de la fonction publique  
**Niveau de diplôme exigé :** Bac + 5 ou équivalent  
**Fonction :** Doctorant

A propos du centre ou de la direction fonctionnelle

Inria the French national institute for research in computer science and control, is dedicated to fundamental and applied research in information and communication science and technology (ICST). Inria has a workforce of 3,800 people working throughout its eight research centers established in seven regions of France.

Grenoble is the capital city of the French Alps. Combining the urban life-style of southern France with a unique mountain setting, it is ideally situated for outdoor activities. The Grenoble area is today an important centre of industry and science (second largest in France). Dedicated to an ambitious policy in the arts, the city is host to numerous cultural institutions. With 60,000 students (including 6,000 foreign students), Grenoble is the third largest student area in France.

Contexte et atouts du poste

Hosting teams: THOTH (https://thoth.inrialpes.fr) and MORPHEO (http://morpheo.inrialpes.fr) are joint research teams of the Inria Grenoble Rhône-Alpes research center and of the Laboratoire Jean Kuntzmann, a joint research unit of CNRS, Inria, Grenoble Alpes university and Grenoble Institute of Technology.

- The main objectives of the THOTH team are: (i) designing and learning structured models capable of representing large-scale visual information; (ii) learning visual models from minimal supervision or unstructured meta-data; and (iii) large-scale learning and optimization. An additional focus of Thoth is on collection of appropriate datasets and design of accompanying evaluation protocols.

- MORPHEO’s main objective is the ability to perceive and to interpret moving shapes using multiple camera systems for the analysis of animal motion, animation synthesis and immersive and interactive environments. Multiple camera systems allow dense information on both shapes and their motion to be recovered from visual cues. Such ability to perceive shapes in motion brings a rich domain for research investigations on how to model, understand and animate real dynamic shapes.

Supervisors:  
Advisor: Cordelia SCHMID  
Co-advisors: DR Grégory ROGEZ & Dr Jean Sébastien FRANCO

**Mission confiée**

While the success of deep learning methods in recent years is undeniable, as applied to a wide array of classical computer vision problems ranging from image feature extraction to semantic segmentation, its applicability to higher level 3D problems is still an open challenge. Fueled by the renewed interest in Virtual Reality fields, and several highly innovative developments in the industry, a growing trend in the community is to examine applicability of such inference tools to estimation of 3D shape and pose characteristics [1]. The ability to retrieve such information from videos is relevant to a broad number of applications, from self-driving cars, where spatial understanding of surrounding obstacles and pedestrians plays a key role, to augmented reality applications such as virtual change rooms that can offer the E-commerce industry a virtual fitting solution for clothing or bodywear.

In the research community, recent works have shown the success of deep network architectures for the problem of retrieving 3D features such as kinematic joints [2,3]...
or surface characterizations [4], with extremely encouraging results. Such successes, sometimes achieved with simple, standard network architectures such as AlexNet [5] or VGG [6], naturally raise the question of applicability of these methodologies for the more challenging problem of end-to-end full 3D shape retrieval. Is it possible to design an architecture that produces full 3D shapes corresponding to humans or objects observed in an input image or a sequence of input images? This is the problem we propose to tackle in this PhD thesis proposal.

Naturally, however simple its formulation, this objective raises several key challenges. First, there is an unsolved representational issue. While the comfort zone of CNNs is in dealing with regular 2D input and output grids, in this case study, the gap must be bridged somewhere in the envisioned architectures between the still 2D nature of inputs, and a 3D shape parameterization yet to be defined. Second, the dimensionality of the problem is considerably higher than what existing 3D networks have been shown to handle, because the parameterization sought is no longer restricted to a subset of the variability, e.g. kinematic pose of humans, but to an intrinsically finer description, which should also accurately account for shape surface details. Third, the training sets to be designed for this problem are yet to be designed and produced, requiring a particularly demanding definition and acquisition effort. The large data variability of 3D problems has motivated some initial efforts to produce fully synthetic training sets [7], where such variability can be scripted. Yet recent successful methods underscore the necessity for as realistic as possible training data, for both the general applicability of the estimation, and to keep the underlying network architecture simple, as devoid as possible of any domain specific adaptations.

Principales activités

This 3-year PhD effort is grounded in a solid environment bringing together the expertise of internationally recognized researchers in the visual recognition field on one hand (THOTH), and 3D tracking and capture on the other hand (MORPHEO). The PhD candidate will benefit from state of the art 3D capture equipment through the Equipep Kinovis platform, which we will leverage to produce highly detailed 3D capture data for training sets in a highly controlled environment, which can in turn help the methods produce accurate results in general acquisition situations. The experience in data acquisition and deep architecture training that the candidate will gain during his MSc project ensure he will be in ideal conditions for starting this PhD work. We expect this thesis to yield breakthrough methods with broad impact on 3D vision, with publication at top level conferences in computer vision, computer graphics, and 3D capture communities.

References:


Key-words: Computer vision, CNN, Deep learning, Human 3D shape, Human 3D pose

Compétences

Master in Computer Science or Applied Mathematics.
Creative and highly motivated
Solid programming skills, Python, C++ and/or Matlab
Solid mathematics knowledge in linear algebra, geometry, and statistics.
 Fluent English or French spoken.
Prior courses or knowledge in the areas of computer vision, computational geometry, mesh processing, computer graphics, signal processing, machine learning is a plus

Avantages sociaux
• Subsidised catering service
• Partially-reimbursed public transport
• Social security
• Paid leave
• Flexible working hours
• Sports facilities

Rémunération


Monthly salary after taxes: around 1596,05€ for 1st and 2nd year. 1678,99€ for 3rd year. (medical insurance included).