

Publishable Summary

Project Acronym: EgoVision4Health

Project Code: PIOF-GA-2012-328288

Project Title: Assessing Activities of Daily Living from a Wearable RGB-D Camera for In-Home Health Care Applications

Scientific Coordinator: Prof. J.M.M. Montiel / Dr Cordelia Schmid

Researcher: Dr Grégory Rogez

Period Covered: July 2013 - June 2016

Project Web page: <http://www.gregrogez.net/research/egovision4health/>

We summarize the research activity performed by Dr. Rogez during the outgoing phase at the University of California, Irvine with Prof. D. Ramanan and during the return phase at Inria Grenoble with Dr C. Schmid.

1. Summary description of the project objectives

Camera miniaturization and mobile computing now makes it feasible to capture and process videos from body-worn cameras such as the Google Glass headset. This egocentric perspective is particularly well-suited to recognizing objects being handled or observed by the wearer, as well as analyzing the gestures and tracking the activities of the wearer. The main goal of the three-year project EgoVision4Health, was to investigate new egocentric computer vision techniques to automatically provide health professionals with an assessment of their patients' ability to manipulate objects and perform daily activities. The main research objectives were: 1) to introduce the use of wearable RGB-D cameras and advance existing knowledge on object detection in first-person views, and 2) to analyze object manipulation and daily activities using detailed 3D models of the human body (hands, upper-body, full-body).

2. Description of the work performed

In the first months of the project, Dr Rogez created a prototype of wearable RGB-D camera by chest-mounting an Intel Creative camera using a GoPro harness. He then collected and annotated (full 3D hand poses) the first RGB-D benchmark dataset of real egocentric object manipulation scenes. He developed a semi-automatic labelling tool to accurately annotate partially occluded hands and fingers in 3D. In a second phase, Dr Rogez developed his own rendering engine which synthesizes photorealistic RGB-D images of egocentric object manipulation scenes. This led to the creation of a large scale training dataset of synthetic egocentric RGBD images that Dr Rogez used to train several new computer vision algorithms for detection and recognition of hands during everyday object manipulations. Then, Dr Rogez analyzed functional object manipulations, this time focusing on fine-grained hand-object interactions. He made use of a recently developed fine-grained taxonomy covering everyday interactions and created a large dataset of 12000 RGB-D images covering 71 everyday grasps in natural interactions. This Grasp UNderstanding dataset (GUN-71) is publicly available. In the last period of the fellowship, Dr Rogez addressed the more general problem of full-body 3D pose estimation in third-person images. This is relevant in case of interactions of the camera wearer with other persons observed from the wearable camera. He developed a new data synthesis technique to generate large-scale (2 millions images) training data that were later used to train Deep Convolutional Neural Networks. This dataset is also publicly available.

3. Description of the achieved results

With the research activity performed during the outgoing phase, Dr Rogez reached a large part of the objectives described in the proposal. He introduced the use of wearable RGB-D cameras and advanced existing knowledge on hand and object detection in first-person views. In particular, he defined and developed the new concept of "*Egocentric Workspace*" and the associated spherical encoding of depth features. This concept allowed developing a new computer vision based method that estimates the 3D pose of an individual's upper limbs (arms+hands) from a chest mounted depth-camera reaching state-of-the-art results in real-time. Dr Rogez then analyzed functional object manipulations during daily activities and explored the problem of contact and force prediction (crucial concepts in functional grasp analysis) from perceptual cues. His analysis reveals the importance of depth for segmentation and detection, and the

effectiveness of state-of-the-art deep RGB features for detailed grasp understanding. Finally, Dr Rogez tackled the more complex problem of full-body 3D pose estimation in RGB images and achieved outstanding results. He artificially augmented a dataset of real images with new synthetic images and showed that Convolutional Neural Networks (CNN) can be trained on artificial images and generalize well to real images. His end-to-end CNN classifier for 3D pose estimation outperforms state-of-the-art results in terms of 3D pose estimation in controlled environments and shows promising results in the wild.

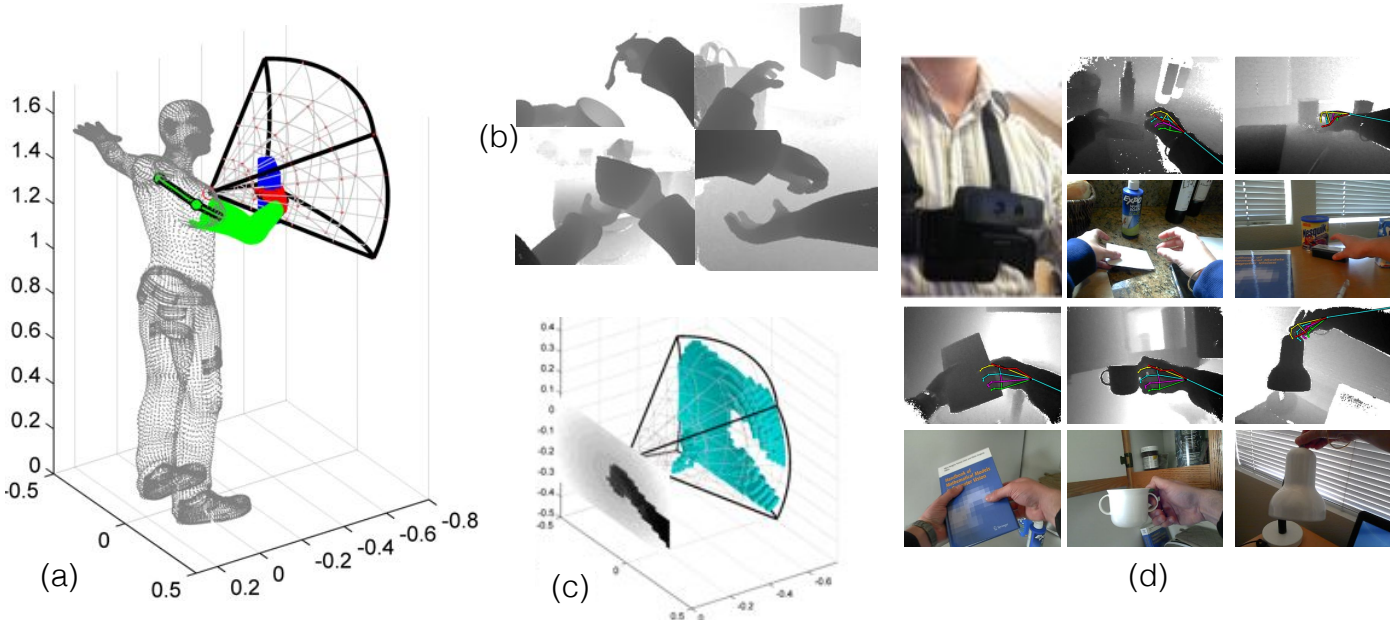


Figure 1. Egocentric pose estimation. (a). Synthetic egocentric camera mounted on a virtual avatar and egocentric workspace. (b) Examples of synthetic training depth images. (c) Depth features computed on the whole egocentric workspace for classification. (d) Our prototype (upper-left) of wearable RGB-D camera and pose estimation results in real egocentric RGB-D images.

4. Results and their potential impact and use (including the socio-economic impact and the wider societal implications of the project so far).

The main motivation of EgoVision4Health is assistive technology. Clinicians watch and evaluate patients performing everyday hand-object interactions for diagnosis and evaluation. A system with patient-wearable camera would allow for long-term monitoring and have an important socio-economic impact. A demonstrator recognizing manipulation of object during basic activities of daily living has been developed.

Egocentric/wearable cameras seem to be emerging as a significant topic in Computer Vision, and Dr Rogez’s depth-based take on things will prove to be impactful in the field. His publications are being cited, his results and the different datasets that he constructed are being used by other research groups. In particular, his new grasp dataset GUN-71 is different from past work (usually addressed from a robotics perspective) in terms of its scale, diversity, and combination of RGB and depth data. Anonymous reviewers said that this work “makes a nice contribution to the community”, representing “a great effort” that introduces a “valuable” dataset that “is more comprehensive than any other comparable dataset available, and is likely to be used in future research”. Dr Rogez introduced the traditionally-robotics problem of visual grasp prediction, including contact+force prediction, to a computer vision audience. Currently, such work on the robotics-vision boundary is targeted exclusively for robotics conferences. This exclusion limits progress for computer vision. The first feedbacks suggest that Dr Rogez’s work will help remedy this.

Contact Details:

Dr Grégory Rogez

Inria Grenoble - Rhône-Alpes, 655 avenue de l’Europe, 38330 Montbonnot-Saint-Martin, FRANCE

gregory.rogez@inria.fr