Egovision4Health - Assessing Activities of Daily Living from a Wearable RGB-D Camera for In-Home Health Care Applications

by Grégory Rogez, Deva Ramanan and J. M. M. Montiel

Camera miniaturization and mobile computing now make 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 analysing the gestures and tracking the activities of the wearer. Egovision4Health is a joint research project between the University of Zaragoza, Spain and the University of California, Irvine, USA. The objective of this three-year project, currently in its first year, is 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.

Activities of daily living (ADL) represent the skills required by an individual in order to live independently. Health professionals routinely refer to the ability or inability to perform ADL as a measure of the functional status of a person, particularly in regards to the elderly and people with disabilities. Assessing ADL can help: 1) guide a diagnostic evaluation, 2) determine the assistance a patient may need on a day-to-day basis or 3) evaluate the rehabilitation process. Initial deployment of technologies based on wearable cameras, such as the Microsoft SenseCam (see Figure 1a) have already made an impact on daily life-logging and memory enhancement. We believe that egocentric vision systems will continue to make an impact in healthcare applications as they appear to be a perfect tool to monitor ADL. One unique wearable camera can potentially capture as much information about the subject's activities as would a network of surveillance cameras. Another important benefit is that the activities are always observed from a consistent camera viewing angle, ie in first-person view.

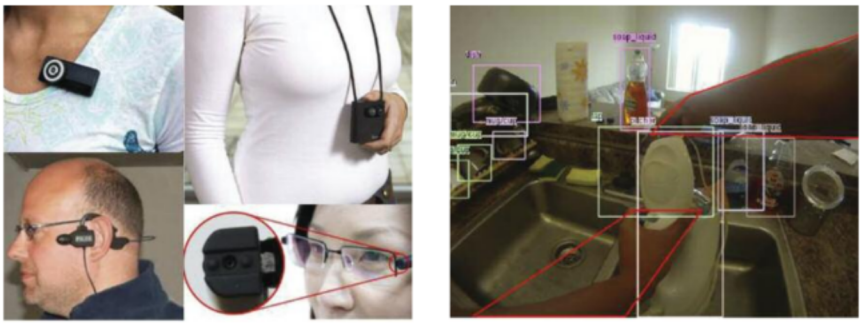

Figure 1 a (left) Examples of wearable cameras (Clockwise from top-left) lapel, neck-worn Microsoft SenseCam, glasses and head-worn camera. Figure 1b (right) Example of a processed image from [1].

Recent work on ADL detection from first-person camera views [1] (Figure 1b) demonstrated an overall performance of 40.6% accuracy was obtained in ADL recognition, and 77% when simulating a perfect object detector. In egocentric vision, objects do not appear in isolated, well positioned photos, but are embedded in a dynamic, everyday environment, interacting constantly with one another and with the wearer. This greatly complicates the task of detection and recognition, especially when an object is being manipulated or occluded by the user's arms and hands.

EgoVision4Health is addressing this problem. Our work is organized along three research objectives: 1) to advance existing knowledge on object detection in first-person views, 2) to achieve advanced scene understanding by building a long-term 3D map of the environment augmented with detected objects, and 3) to analyse object manipulation and evaluate ADL using detailed 3D models.

The analysis of "near-field" object manipulations - and consequently ADL recognition and assessment - could benefit greatly from having all the objects that are likely to be manipulated already located in the 3D environment. For example, if we want to determine whether a person is picking up a mug the wrong way due to an injury it seems important to know where the handle of the mug is, and how it is oriented in 3D. Another advantage of having all the objects already located around the subject is that we can categorize the scene and improve ADL recognition, eg cooking only happens in the kitchen. For a real breakthrough in ADL detection and assessment from a wearable camera, a thorough a priori understanding of the subject's environment is vital.

Since we expect Kinect-like depth sensors to be the next generation of cheap wearable cameras, we use a RGB-D camera as a new wearable device and exploit the 2.5D data to work in the 3D real-world environment. By combining bottom-up SLAM techniques and top-down recognition approaches, we cast the problem as one of "semantic structure from motion" [2] and aim at building a 3D semantic map of the dynamic environment in which the wearer is moving. We plan to model objects and body parts using 3D models and adopt the completely new approach of considering each ADL as the interaction of the 3D hands with 3D objects.

The hypothesis at the basis of our proposal is that the EU's well-established technology in mapping for robotics and the latest computer vision techniques can be cross-fertilized for boosting egocentric vision, particularly ADL recognition. Our goals are motivated by the recent advances in object detection, human-object interactions and ADL

detection [1] obtained by UC Irvine's group, as well as by the expertise of the University of Zaragoza in robust and real-time Simultaneous Localization and Mapping (SLAM) systems [3].

The tools currently available in each of these domains are not powerful enough alone to account for the diversity and complexity of content typical of real everyday life egocentric videos. Current maps, composed of meaningless geometric entities, are quite poor for performing high-level tasks such as object manipulation. Focusing on functional human activities (that often involve interactions with objects in the near-field), and consequently on dynamic scenes, adds to the challenging and interesting nature of this problem, even from a traditional SLAM perspective.

On the other hand, ADL detectors perform poorly in the case of small objects occluded by other surrounding objects or by the user's body parts. Research breakthroughs are thus required, not only in vision-based ADL recognition and SLAM, but also in exploiting the synergy of the combination.

EgoVision4Health is financed by the European Commission under FP7-PEOPLE-2012-IOF through grant PIOF-GA-2012-328288.

Links:

http://www.gregrogez.net/research/egovision4health/

http://cordis.europa.eu/projects/328288

References:

[1] H. Pirsiavash, D. Ramanan: "Detecting activities of daily living in first-person camera views", in proc. of IEEE CVPR, 2012, pp. 2847-2854

[2] N. Fioraio, L. Di Stefano, "Joint Detection, Tracking and Mapping by Semantic Bundle Adjustment", in proc. of IEEE CVPR, 2013, pp. 1538-1545

[3] J. Civera, A. J. Davison, J. M. M. Montiel: "Structure from Motion using the Extended Kalman Filter", Springer Tracts in Advanced Robotics 75, Springer 2012, pp. 1-125.

Please contact:

Grégory Rogez
Aragon Institute for Engineering Research (i3A),
Universidad de Zaragoza, Spain
E-mail: grogez@unizar.es

Applying Random Matrix Theory Filters on SenseCam Images

by Na Li, Martin Crane, Cathal Gurrin and Heather J. Ruskin

Even though Microsoft's SeneseCam can be effective as a memory-aid device, there exists a substantial challenge in effectively managing the vast amount of images that are maintained by this device. Deconstructing a sizeable collection of images into meaningful events for users represents a significant task. Such events may be identified by applying Random Matrix Theory (RMT) to a cross-correlation matrix C that has been constructed using SenseCam lifelog data streams. Overall, the RMT technique proves promising for major event detection in SenseCam images.

Microsoft's SenseCam is a lifelogging camera with a fisheye lens that is worn, suspended around the neck, to capture images and other sensor reading in an automatic record of the wearer's every moment. SenseCam can thus collect a large amount of data, even over a short period of time, with a picture typically taken every 30 seconds, and an average of 4,000 images captured in a typical day. Even though experience shows that the SenseCam can be an effective memory-aid device, serving to improve recollection of an experience, users seldom wish to refresh their memory by browsing large image collections. Hence, tools are required to assist in the management, organization and analysis of these large data sets, eg, to automatically highlight key episodes and, ideally, to classify these in terms of importance to the life logger. Our previous research has shown that SenseCam time series exhibit a strong long-range correlation, indicating that the series do not constitute a random walk, but are cyclical, with continuous low levels of background information picked up constantly by the device [1]. Further, we have shown that a cross-correlation matrix can be analysed to highlight key episodes, thus identifying boundaries between daily events [2].

However, due to the finite length of time series available to estimate cross correlations, the matrix contains "random" contributions. As a consequence, a percentage of noise or routine event inclusion in processing is inevitable. (This phenomenon can also be observed in other domains such as the analysis of financial data and wireless communications [3]). A well-known method for addressing this type of problem is to apply Random Matrix Theory (RMT).

The aim is to compare the properties of the cross-correlation matrix C with those of the random correlation matrix R, separating the content of C into two groups: (a) the part reflecting properties of R ("noise") and (b) the part that deviates from R (and contains information on major events). Figure 1 compares the probability distribution of a typical cross-correlation matrix with that for the random correlation matrix. We note the presence in the former of a well-defined "bulk" of eigenvalues, which fall within the bounds for the latter. We also note deviations for a number of the largest and smallest eigenvalues. This suggests that the cross-correlation matrix captures many major events from the data stream, but also contains substantive noise.

The deviations of the probability distribution of the cross-correlation matrix from RMT suggest that such deviations should also be observed in the corresponding eigenvector components. In order to interpret the meaning of the