A Kinect NUI for 3D Medical Visualization

Luigi Gallo, Alessio Pierluigi Placitelli and Giuseppe De Pietro
National Research Council of Italy (ICAR-CNR)
{*luigi.gallo, alessio.placitelli, giuseppe.depietro*}*@na.icar.cnr.it*

Abstract

This proposal introduces a natural user interface (NUI) that allows users to rotate with 3 degrees of freedom, point to and crop 3D reconstructions of anatomical parts by using a Kinect as the only input device. The NUI is built upon a view-independent hand pose recognition module, which allows users to recognize a limited set of pre-defined postures from single, low resolution depth images. We have used principal component analysis to estimate the hand orientation in space, Flusser moment invariants as image features for visual recognition and a multiclass Support Vector Machine to classify the features extracted to a limited set of static hand postures.

1. Introduction

Recent improvements in depth sensing technologies have been giving a new stimulus to the exploration of the potential for touchless interaction in several different application domains. With more researchers exploiting the extremely cheap technology of structured light depth cameras every day, the design of effective natural user interfaces (NUIs) is now rapidly becoming a major challenge in human-computer interaction [3]. Nonetheless, 3D user interfaces able to fully exploit depth sensing technologies are still in their infancy. Interacting with 3D worlds is more complex than with 2D WIMP interfaces, since it requires the user to manipulate the position and orientation of 3D objects involving six degrees of freedom (DOF). Moreover, a truly 3D interaction requires the user not to be constrained to maintain a fixed position in the 3D space. On the contrary, most of the Kinect NUIs proposed so far are able to correctly recognize hand and arm gestures only if the user stands fronto-parallel to the sensor and at a fixed distance from it. At the heart of these difficulties is the need to recognize static and dynamic hand gestures from low resolution depth images of a hand at different distances and differently oriented.

These observations have motivated us to attack these problems by investigating view-independent hand pose recognition from a single depth image. Most of the works in literature approach multi-angle static hand gesture recognition by using multiple sensors. However, most of the approaches require complex calibration steps, so reducing the portability of the system, and force the user to keep her/his hand in a small area of the 3D space. In contrast, our work has been focused on the design of a NUI that takes advantage of the hand posture recognition module to allow users to manipulate 3D objects by using a single Kinect as the input device. The interface we introduce in this proposal has been tailored for 3D medical visualization. The motivation is that, in the medical imaging informatics domain, there is a growing need for uncumbersome, easy to use 3D interaction tools to allow a direct inspection of 3D reconstructions of anatomical parts. Possible applications of the proposed Kinect NUI are in operating rooms, where non-sterilizable devices cannot be used to avoid breaking the asepsis [2], and in medical education, it being a practical and cheap solution to allow physicians and students to share the same interaction space and tools.

2. Hand posture recognition

On each frame, a nearest neighbor filter has been applied to the hand position reported by the Kinect SDK skeleton tracking module in order to extract the point data concerning the right hand, thus producing a smaller 3D point cloud containing hand data only. Then, principal component analysis (PCA) has been employed to estimate the hand orientation, by computing the eigenvectors of the hand scatter matrix. Hand pose compensation is performed by transforming the point cloud data of the hand in such a way that the extracted principal directions are aligned to the frame of the sensor. Then, we have used Flusser moment invariantsto characterize the hand posture image. In our work we have used only the five true invariants, not considering the skew invariant.

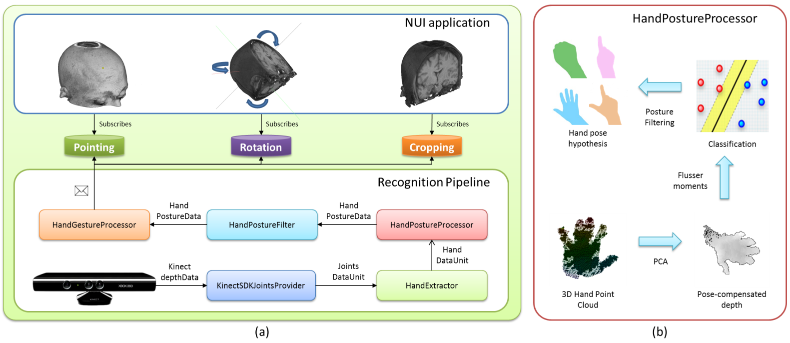

Figure 1. a) Overview of the Kinect NUI for medical visualization and of its hand pose recognition system; b) Steps performed into the HandPostureProcessor module.

In fact, it distinguishes between mirrored images of the same object, which is undesirable in our case since the pose normalization step can produce images reflected across the x- or y-axis. A first hand posture hypothesis is generated by means of a Support Vector Machine classifier, using the Radial Basis Function (RBF) kernel and the one-against-one multiclass strategy. Then, the final hand pose hypothesis has been computed with a threshold-based approach, in which the threshold is a function of the hand orientation and of the hand velocity. Since the faster we move the less precise our movements are, we speculate that the probability of a hand pose transition in a frame is inversely proportional to the hand velocity in that frame. The hand orientation with respect to the sensor, instead, has been considered as an indicator of the degree of certainty of the hand pose hypothesis.

To train and test the recognition system, we have recorded a database of greyscale depth images with a Kinect with four different hand postures: open hand, fist, pointing and index thumb. Six subjects have been required to perform three sequences of hand rotations: two out-of-plane rotations (yaw and pitch) and one in-plane rotation (roll), while standing at about 1 and 2 meters from the sensor. The classification accuracy has been determined using a 10-fold cross-validation. It was found to be 81.18% in the $0.8m - 1.2m$ interval, decreasing to 73.53% in the $1.2m - 2.0m$ interval.

3. NUI implementation details

The Kinect NUI for medical visualization allows the user to point to, crop and rotate with 3 DOF reconstructions of anatomical data[1].We have used the Smoothed Pointing [1] technique to enhance the precision of distal pointing and to smooth rotations. The NUI has been built upon five independent modules, connected as a pipeline by means of a signal/slot paradigm (see fig. 1). The KinectSDKJointsProvider is the first component in the pipeline and relies on the Kinect SDK to segment the depth data and produce a 3D point cloud for each user; it also fills a data packet with the estimated positions of joints and propagates it along the pipeline. Then, the HandExtractor component applies a nearest neighbor filter in order to extract smaller 3D point clouds only containing hand data. As the third step along the pipeline, the HandPostureProcessor employs PCA to further process hand point clouds by aligning the computed principal directions to the frame of the camera through a rigid transformation; Flusser moments are then extracted and the hand posture is recognized using the extracted features in a multi-class SVM classifier based on a RBF kernel. The HandPostureFilter analyzes the received posture data packets and produces the final hand posture hypothesis. Finally, the HandGestureProcessor collects posture packets and joint positions to produce high level event signals to which an application can subscribe.

[1]A video showing the interface in action is available at: www.ihealthlab.icar.cnr.it/index.php/10-news/68-kinectnuimedical.html

References

[1] L. Gallo and A. Minutolo. Design and Comparative Evaluation of Smoothed Pointing: a Velocity-oriented Remote Pointing Enhancement Technique. *Int. J. of Human-Computer Studies*, 70(4):287–300, 2012.

[2] L. Gallo, A. P. Placitelli, and M. Ciampi. Controller-free exploration of medical image data: experiencing the Kinect. In *IEEE CBMS '11*, pages 1–6. IEEE, 2011.

[3] G. Goth. Brave NUI world. *Communications of the ACM*, 54(12):14–16, 2011.