Multi Hand Pose Recognition System using Kinect Depth Sensor

O. Lopes[1], M. Pousa[1], S. Escalera[2] and J. Gonzàlez[3]
Computer Vision Center, Campus UAB, Edifici O, 08193, Bellaterra, Spain
[1]{oscar.pino.lopes, miguelpousafreire}@gmail.com, [2]sergio@maia.ub.es, [3]poal@cvc.uab.es

Abstract

Hand pose recognition is a hard problem due to the inherent structural complexity of the hand that can show a great variety of dynamic configurations and self occlusions. This work presents a hand pose recognition pipeline that takes advantage of RGB-Depth data stream, including hand detection and segmentation, hand point cloud description using the novel Spherical Blurred Shape Model (SBSM) descriptor, and hand classification using OvO Support Vector Machines. We have recorded a hand pose dataset of multiple hand poses, and show the high performance and fast computation of the proposed methodology.

1 Introduction

In recent years a great interest and progress in Human-Computer Interaction (HCI) have been performed with the objective of improving the overall user experience [1].

When considering vision-based interfaces and interaction tools, user hand gesture interaction opens a wide range of possibilities for innovative applications. It provides a natural and intuitive language paradigm to interact with computer virtual objects that are inspired in how humans interact with *real-world* objects. However hand pose recognition is an extremely difficult problem due to the highly dynamic range of configurations the hand can show, and to the associated limitation of traditional approaches based solely on RGB data.

More recently with the appearance of affordable depth sensors such as Kinect$^{\text{TM}}$, a new spectrum of approaches were opened. Kinect provided a new source of information that dealt closely with the 3D nature of the objects of interest. This way it is now feasible to extract 3D information from the hand pose using a gloveless input source and, from then on, analyse and estimate the observed pose. This proposal comprises a system for hand pose recognition from a live feed from the Kinect device. This is essentially a classification problem, however the overall results are boosted due to the novel Spherical Blurred Shape Model descriptor created specifically for this task. This descriptor both fullfills temporal complexity requirements and a high discriminative power that can take full advantage from the depth information.

2 Live Hand Pose Recognition Pipeline

The proposed hand pose recognition pipeline (Figure 1) comprises the following main components:

- Hand pose detection and segmentation module, that consists in a two hand scale-invariant blob detector.
- Rotation invariant Spherical Blurred Shape Model descriptor, that performs a point cloud partitioning that harnesses the highly discriminative information provided by the depth data.
- Hand pose classification based on One-versus-One Support Vector Machines (SVM), that were integrated in the pipeline in order to perform online classification of the live data stream.

3 Hand Pose Dataset

The system's overall performance, accuracy wise, depends heavily on its correct training with a representative hand pose dataset. Since currently there was no point cloud hand pose dataset available, it was required to craft one. This was performed using the hand pose detector modification to perform the data

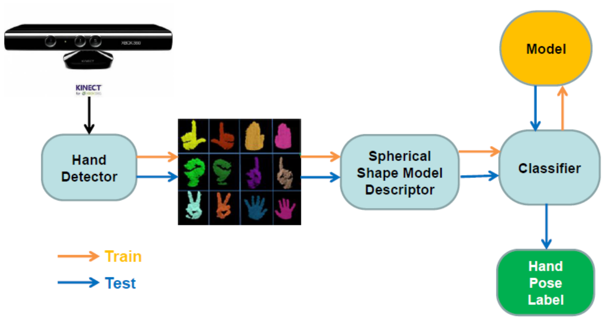

Figure 1: System architecture.

persistence. The dataset comprises 6 classes, 2000 samples each, including both hands.

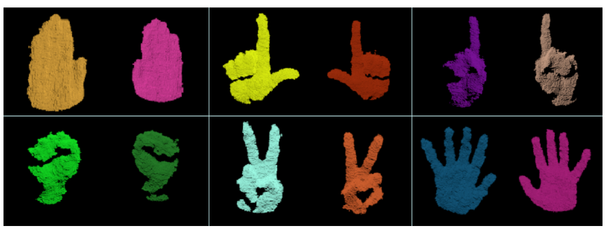

Figure 2: Hand pose dataset.

In order to improve the system robustness, and since the problem is treated as a multi-classification task, a *no pose* class was added, which includes samples captured by the device referring to non-hand elements.

4 Descriptor

The description of the 3D hand is performed using the Spherical Blurred Shape Model (SBSM) descriptor (Figure 3). The SBSM descriptor is based on its 2D counterpart the CSBM descriptor [2]. SBSM decompose a 3D shape with an spherical quantization and perform an assignment based on the distance from the points to the bins centroids. The result is a feature vector that codifies the input shape in a compact manner and has a high discriminative power.

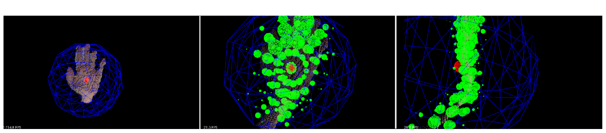

Figure 3: SBSM descriptor.

5 Classification

The classification step is based on SVM. The problem is tackled as a multi-classification task, considering C-SVM using the RBF kernel, combined with a descriptor configured with 257 features, that accordingly to the experimental analysis offers the best balance between performance and final results reliability.

6 Demonstration

The pipeline is demonstrated by visualizing a live feed from the Kinect device, where a single user exposes his hands and the system applies a on-screen label accordingly to the visualized hand pose (see Figure 4). This initial prototype could be included in a more rich HCI environment, using both the hand pose and its spacial location, in order to execute command and control actions, object manipulation, or interface navigation.

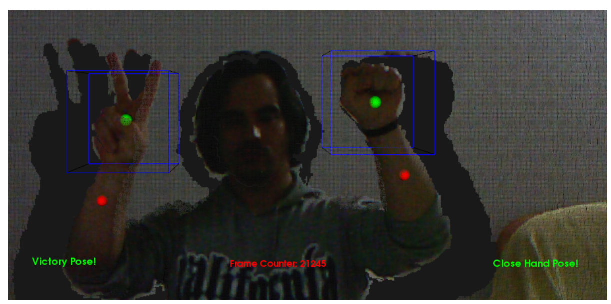

Figure 4: Prototype of the system.

References

[1] F. Periverzov and H. Ilie, "3d imaging for hand gesture recognition: Exploring the software-hardware interaction of current technologies," *3D Research*, vol. 3, pp. 1–15, 2012. [Online]. Available: http://dx.doi.org/10.1007/3DRes.03%282012%291

[2] S. Escalera, A. Fornés ands, O. Pujol, J. Lladó ands, and P. Radeva, "Circular blurred shape model for multiclass symbol recognition," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 41, no. 2, pp. 497 –506, april 2011.