Research Modules for Undergraduates in Machine Learning for Automatic Gesture Classification

Sambit Bhattacharya, Bogdan Denny Czejdo and Nick Perez

Department of Mathematics and Computer Science,
Fayetteville State University
Fayetteville, NC 28301
910 672-2466 & 1156
{sbhattac, bczejdo }@uncfsu.edu

Abstract

In this paper we describe ongoing undergraduate research projects that allow us to shift emphasis from teaching to a more active form of student participation. More specifically our projects are on automatic gesture recognition using the Kinect 3D sensor from Microsoft Research and machine learning systems. We have observed the following benefits for our undergraduate students: learning a topic area in AI relatively early; developing proficiency in laboratory practice, specifically, systematic data collection and programming on multiple platforms; learning to use appropriate methodology; applying knowledge to a real situation; learning to analyze data and transform it to various representations; appreciation of scientific experiments and learning what scientific research actually entails.

Introduction

Human robot interaction with automatic gesture recognition is at the frontiers of AI architectures research. It creates many opportunities for detailed experimental research in which undergraduate students can meaningfully participate.

In general, by involving undergraduate students in research we can shift emphasis from teaching to a more active form of student participation (Erkan, Newmark and Ommen, 2009). In this paper we describe ongoing undergraduate research projects on automatic gesture recognition using Kinect 3D sensing technology, machine learning systems and humanoid robots. We are approaching this project building upon the findings described in (Lopatto, 2003) about the benefits of undergraduate research. We have observed the following benefits for our undergraduate students: learning a topic area in AI relatively early; developing proficiency in laboratory practice, specifically, systematic data collection and programming on multiple platforms; learning to use appropriate methodology; applying knowledge to a real situation; learning to analyze data and transform it to various representations; appreciation of scientific experiments and learning what scientific research actually entails. In addition students participating in undergraduate research improve their oral and written communication skills.

We will describe two developed research modules and several in preparation. The first research module involves investigation of the precision of depth sensing and human skeleton tracking with the Kinect sensor. The second module involves experimentation with a machine learning system as applied to gesture recognition using the Kinect. Some examples of modules under development are: dynamic gesture recognition with other advanced machine learning techniques and control and interaction with humanoid robots using gestures.

Precision of Depth Sensing and Human Skeleton Tracking

In this module the students studied how the Kinect can measure depth data (estimates of distance from Kinect to objects in scenery) and can identify and locate human skeleton. This skeleton is identified through positions of skeletal joints computed by Kinect from depth data. They also learned about the parameters and limitations of the device e.g., 20 different joints can be tracked for a single person; full data can be measured for up to two people with partial measurements for more; the Kinect processes and sends the skeletal data along with the depth data at the rate of 30 frames a second.

With such knowledge, our students were challenged to determine the precision of measurements of depth sensing and human skeleton tracking using Kinect. This is a very typical first problem in most scientific research. The precision of measurements need to be determined and its effects for future experiments need to be studied carefully.

After designing and carrying out several experiments our students realized the difficulty and complexity of precision determination. Let us look at Table 1 showing the standard deviation of measurements of Cartesian coordinates of a middle shoulder joint.

Std. Deviation	**x**	**y**	**z**
5 ft.	5.64	0.96	9.50
10 ft.	9.13	3.43	6.41

Table 1. Standard deviation for measurement of Cartesian coordinates of a middle shoulder joint where each coordinate is in millimeters.

The standard deviation was computed based on 50 measurements for each of 2 distances of the human body from the Kinect sensor. These measurements were taken when the middle shoulder joint was stationary but some other joints were moving. This and other similar results were very interesting for the students. Let us summarize the results:

- The 3-standard deviations rule for y value told us that the y coordinate measurement most probably would not vary more than 3 mm for the distance 5 ft. and not more than 10 mm for the distance of 10 ft. This seemed logical since the larger distance to the object could cause larger error of measurement.
- The 3-standard deviations rule for the value z, was completely surprising for students because it told us that z coordinate measurement could vary 30 mm for the distance 5 ft. and only 20 mm for the distance of 10 ft. An explanation of these observations was necessary. Analyzing again the experiments the students realized that they are actually measuring the impact of two variables: precision of Kinect and instability of human body. The middle shoulder joint was actually not stationary but was moving slightly with the movement of other joints e.g. hands. This was due to involuntary small bodily movements when the human subject was trying to hold the pose over time.
- The difference between y and z was that y was relatively stable i.e. no movement up and down while hand was moving. This was due to stability provided by the floor supporting the skeletal frame.

The experiment demonstrated to students the necessity of careful statistical analysis to understand the noisy characteristics of sensory data.

The students also analyzed the measurements for hand joint while moving. Some of the results are shown in Table 2. The results can be summarized as follows:

- The error for x and y value for a given pose told us that y coordinate measurement would not vary much for a comfortable hand position and have a significant error for a very uncomfortable position.

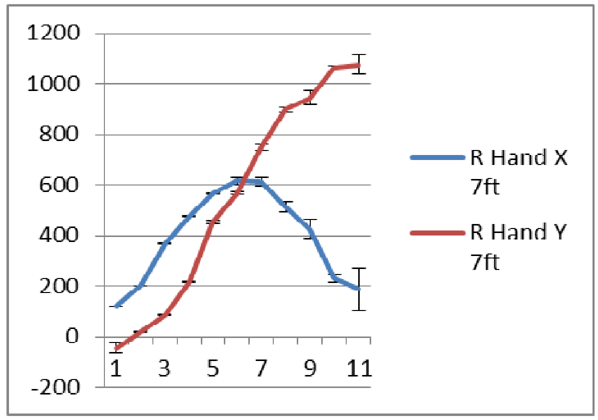

Table 2. The measurements of Cartesian coordinates of a hand joint shown with the computed error. The vertical axis is distance in millimeters and the horizontal axis is pose number where the number ranges from most comfortable to least comfortable type of pose.

Other experiments followed that showed the more detailed picture of Kinect system precision and the human pose precision.

Among the experiments let us mention few:

- placing the human body against the wall so that we can get a better stability of some of the human joints.
- using unmovable manikin
- changing Cartesian coordinates into spherical coordinates

It was very important for students to see that:

- the design of an experiment can affect precision of measurement since many variables can affect the experiment.
- a single solution often was difficult to find and that some type of calibration of results might be the only way to obtain better results.
- changes of data representation (e.g. from Cartesian to spherical) change the description of precision of measurement for better visualization.

- some obvious solution could unintentionally introduce other uncontrolled variables in the experiment e.g. placing the human body against the wall caused problems related with human skeleton tracking through the Kinect system; the students found that the reason for this was the very small difference in depth between human body parts and the wall.

Gesture recognition with machine learning

In this research module the students learned how to apply machine learning to the pattern classification problem. The specific patterns used in our experiments were static gestures or poses of the human body. Students used artificial neural networks to the automatic classification of pose/gesture.

One line of experimental enquiry was to find out what combination of joint data would be optimal for training a neural network to recognize poses. Students also experimented with different number of nodes in the hidden layer to investigate the problem of how well the network is able to generalize based on the same training data. As shown in the table below, given an 8-pose classification problem, where data from 5 joints only were used, the network was able to classify new test cases more accurately when the number of nodes was increased from 10 to 20. However, increasing the number of nodes beyond 20 made the classification less accurate due to over fitting.

In general, students could experiment with various machine learning system parameters. To summarize the students experimented with:

- the number of poses and its impact on neural network training
- the number of joints used in the data set and its impact on the neural network training
- the number of neurons in the hidden layer and its impact on the neural network training
- changing Cartesian coordinates into spherical coordinates in the data set for training neural networks

It was very important for students to see that:

- the training of neural network can produce different results even for the same data set since parts of the data set are randomly selected for verification and testing.
- sufficient variability in training data is required for the neural network to generalize well; specifically, changing the whole body position while holding the same pose was essential for proper automatic recognition of poses.
- changes in data representation (e.g. from Cartesian to spherical) can change the results of neural net training.

Summary and future directions

In this paper we described two developed research modules for undergraduate research. The first research module involves investigation of the precision of depth sensing and human skeleton tracking with the Kinect sensor. The second module involves experimentation with a machine learning system as applied to gesture recognition using the Kinect.

# of neurons	Training Error for 4 Networks	Posture 1	Posture 2	Posture 3	Posture 4	Posture 5	Posture 6	Posture 7	Posture 8
10	0.012273	1.0000	0.0001	0.0005	0.0006	0.0001	0.0000	0.0000	0.0000
	0.006415	0.0011	0.9989	0.0000	0.0000	0.0001	0.0001	0.0002	0.0012
	0.23589	0.0034	0.0000	0.4856	0.0252	0.0000	0.0000	0.0000	0.0000
	0.006819	0.0000	0.0011	0.6991	0.9052	0.0599	0.0001	0.0014	0.0001
	Avg Err:	0.0011	0.0013	0.0000	0.0000	0.9566	0.0001	0.0000	0.0000
	0.065349	0.0000	0.0172	0.0000	0.0000	0.0002	0.9854	0.0000	0.0083
		0.0006	0.0000	0.0006	0.0006	0.0000	0.0000	0.9934	0.0048
		0.0000	0.0000	0.0001	0.0000	0.0000	0.0001	0.0135	0.9862
20	0.00014574	0.9993	0.0000	0.0000	0.0000	0.0063	0.0000	0.0010	0.0000
	7.71E-06	0.0002	0.9981	0.0000	0.0001	0.0007	0.0006	0.0001	0.0003
	0.00036695	0.0000	0.0000	0.9998	0.0000	0.0000	0.0000	0.0000	0.0002
	0.00070239	0.0000	0.0000	0.0000	0.9997	0.0000	0.0017	0.0000	0.0001
	Avg Err:	0.0002	0.0009	0.0007	0.0000	0.9805	0.0000	0.0000	0.0000
	0.0003057	0.0000	0.0009	0.0000	0.0002	0.0000	0.9986	0.0000	0.0003
		0.0003	0.0000	0.0005	0.0000	0.0000	0.0000	0.9999	0.0000
		0.0000	0.0000	0.0000	0.0004	0.0000	0.0000	0.0000	0.9998

Table 3. The results for 8-pose classification problem.

We are currently developing more research modules which explore new experimental directions based on and related to our existing work. Some examples of modules under development are: dynamic gesture recognition with other advanced machine learning techniques and control and interaction with humanoid robots using gestures. Dynamic gesture recognition requires more advanced techniques such as time-compressed templates and time-delayed neural networks (Mitra and Acharya, 2007). We are planning to use the approach of research through replication of research (Erkan, Newmark and Ommen, 2009). In this approach undergraduate students will be verifying results from other pioneering research through experimentation with large volumes of data.

In another future direction we will combine the machine learning techniques with robot control to achieve human robot interaction. In this project students will find a novel application of the previously done research. In addition this will open the door to applications in social and rehabilitation robotics from which students will be able to choose their niche for development.

Acknowledgments

We acknowledge the National Science Foundation (award id: 0959958), internal grants from the Graduate School of Fayetteville State University, and the Belk Foundation for supporting this research.

References

Lopatto, D. 2003. The essential features of undergraduate research. Council on Undergraduate Research Quarterly, 23, 139–142.

Mitra, S.; Acharya, T. 2007. Gesture Recognition: A Survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol.37, no.3, pp.311-324.

Erkan, A.; Newmark, S.; and Ommen, N. 2009. Exposure to research through replication of research: a case in complex networks. In Proceedings of the 14th annual ACM SIGCSE conference on Innovation and technology in computer science education (ITiCSE '09). ACM, New York, NY, USA, 114-118.