Application of the Nonlinear Wave Metric for Gesture Classification and Gesture Control

András Pálffy, Domonkos Huszár, Lóránt Kovács, András Horváth

†Faculty of Information Tecnhology, Pazmany Peter Catholic University, Budapest, Hungary
1444 Pf. 278, Budapest, Hungary

Abstract—The appearance of intelligent motion sensors and disruptive technologies (like the Wii-controller [1] and Kinect [2]) have revolutionized computer interfaces in gaming in the past decade. However, this revolution restricts itself to the gaming world and these controllers and interfaces are not widely used in other applications. One of the reasons behind this hindrance is the lack of natural and easy to learn gestures and controls. The gestures used by these devices can only be used efficiently if they fit the application framework and can be applied without a long and tedious learning process. Most of the applied learning methods can not be used because comparison of trajectories and spatial-temporal characteristics are very different from image comparison.

In this paper we show how trajectories can be learnt and classified using the non-linear wave metrics and the Leap Motion controller.

1. Introduction

Three dimensional control is extremely important, because our world (at least our perception) is three-dimensional, but computers are usually built with two-dimensional displays. Although the revolutions of virtually three dimensional systems had a huge impact on science as well, most of these applications lack convenient three dimensional control. Real three dimensional control is simulated by control on two-dimensional axes separately, which can be learned but remains more complex and difficult to be applied than an inherently three dimensional control.

The observation and the control of real three-dimensional objects can be extremely important in certain areas where we need to gather real spatial or sometimes spatial-temporal information about the objects. Just to mention a few simple examples: in education (where in engineering and mathematics three dimensional plots of objects and multivariable functions are extremely important for spectacular illustrations and the analyses of such plots can be done easily in detail) or in medical imaging (where a three dimensional display can reveal more information for the doctors).

To achieve our goal we have used the Leap Motion controller [3] which is a small portable USB peripheral device, easily integrable to any environments. The device contains two monochrome infra-red cameras and three light emitting diodes (LED). These LEDs generate a pattern which helps in the estimation of the position and movement of arbitrary objects (e.g. the user's hands, fingers) by range imaging.

We have programmed the controller to be able record four dimensional (three spatial and a temporal) coordinates and compared the efficiency and accuracy of the classification of these trajectories by different metrics.

In section 2 we describe the main properties of gesture control and the devices we have used for our experiments. In section 3 we define the non-linear wave metric and how it can be applied in higher dimensional problem spaces. In section 4 we show and explain our results and the comparison of our method with commonly used metrics. In section 5 we conclude our results.

2. Gesture Recognition with The Leap Motion controller

The Leap Motion Controller observes a roughly hemispherical space, with a detection range up to 60 *centimeters*, which is illuminated by infrared light emitting diodes generating a three-dimensional pattern of dots. The device can provide a datastream with 300 frames per second by synthesizing three-dimensional positions by comparing the stereo-images acquired by the controller.

The smaller observation area and higher resolution of the device differentiates the product from the Kinect and other controllers which are more suitable for whole-body tracking in a space the size of a living room. The Leap Motion Controller may perform better in tasks such as navigating a website, using pinch-to-zoom gestures on maps, high-precision drawing, and manipulating complex three-dimensional data visualizations. Generally, the Leap Motion controller is designed to track fingers and hands precisely, while Kinect can provide a large scale (whole body) three-dimensional point map.

Applications for the controller can be downloaded from the application store[1], however these applications are often experimental, not general and difficult to use. The controller also have an application programming interface that can be used to developed software control to detect arbitrary gestures. These gestures are critical in three dimensional applications, since they may provide a set of inher-

[1] https://airspace.leapmotion.com

ent, easy to use and habitual signals and actions for controlling three-dimensional objects for moving, rotating and zooming.

Although these easy to learn gestures are intuitive and self-explanatory they are not the same for everyone. Different applications can require different gestures for similar or completely same functions. Gestures can also be different for each user based on their experiences, age and background. A more sophisticated user would require many complex gestures which can be hardly distinguished from each other, meanwhile a novice person would use only a few characteristic ones [4]. We can also notice the baby-duck effect in gesture control [5]: the tendency of computer users to prefer the systems that they learn on, and to reject the unfamiliar.

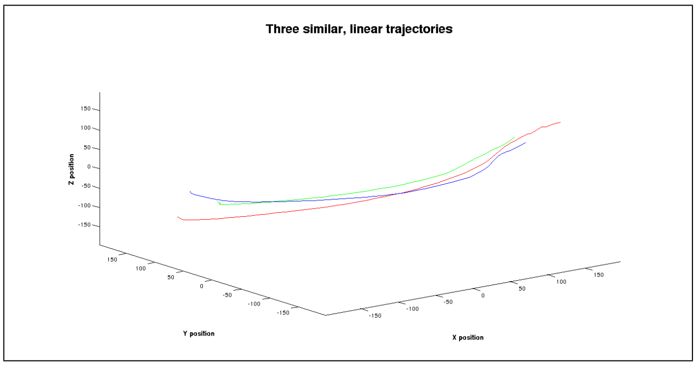
In most gesture recognition softwares gestures are hard-wired and can not be changed or tuned. This results a robust but not user friendly and optimal control. To allow users to have custom and more suitable gestures only one problem has to be solved: the comparison of recorded trajectories. With current devices gestures can be observed and recorded with sufficient spatial and temporal resolution. Recorded gestures can be interpreted as four dimensional inputs (with three spatial dimensions and time being the fourth coordinate). on the other hand the comparison of such data is difficult and computationally expensive. The most serious problem in the comparison is the definition of a proper metric. A proper metric must reflect human intuition and have small distance between similar gestures and larger distance between different ones, since out goal is to make the usage of the device intuitive and convenient.

We have defined such a gesture recognition system based on the non-linear wave metric for the Leap Motion controller implemented in Python, C++ and Java, thus our environment can be easily connected to applications written in these languages. These applications are capable of displaying three-dimensional objects and plots (obviously just two-dimensional images from different angles), but the control of these three dimensional viewpoints are hard to be controlled by regular computer interfaces, which are operating in two-dimensional planes only.

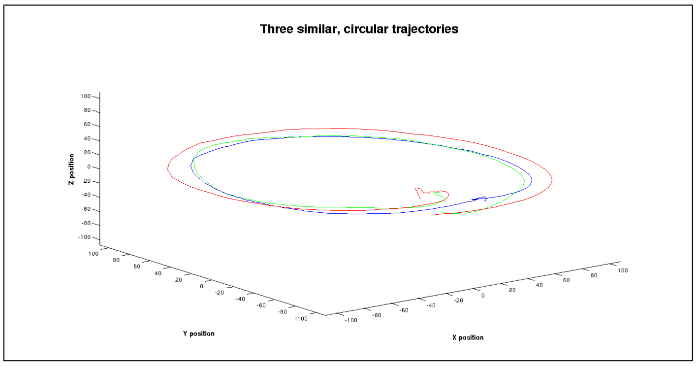
2.1. Gesture classification

The measurement and comparison of gestures is a difficult task because the comparison can not be based on absolute positions. In classification only the relative changes (both spatial and temporal) are determining, the starting point of a trajectory does not matter, since it would be extremely frustrating to always find the exact starting point for a gesture.

This problem is also known in image processing and it is usually solved by moving the object to have a common centroid point. We have applied the very same method here and shifted the trajectories to a common centroid. Naturally this operation can only be done once gestures are finished, but this can be detected by simple movement detection. Some example gestures with similar trajectories, shifted to a common centroid point can be seen on Figure 1.



(a) Line/swipe like trajectories



(b) Circle like trajectories

Figure 1: Three-three examples of two different gestures. The first image shows some line like trajectories in three-dimensions and the second depicts three circle like ones. The fourth, temporal coordinates are neglected on the figure. As it can be seen all four coordinates are necessary for proper classification, since from these data (without the temporal coordinates) the direction of the movement can not be identified, for example it can not be calculated whether a circle is clock-wise or counter clockwise.

Apart from this problems one needs good and properly defined metrics to compare two trajectories. We have to keep in mind that in case of gestures temporal changes are not that determining as relative position changes. A little bit slower or faster gesture is usually considered the same, meanwhile a rotated gesture, pointing into an other direction is usually identified as a different gesture (which often triggers a some-how opposite action) by a human observer.

With different parameters the user must have the ability to tune the spatial and temporal properties of the metric. These properties can not be handled by commonly used metrics (we have no not that they can be modified artificially to become problem dependent), but can be solved by the Wave metric.

3. The non-linear wave metric

The grayscale version of the wave metric can be formulated in a PDE model where an image - either binary or gray-scale - is defined as a real valued function $I(x, y)$: $[0, N]^2 \rightarrow [0, 1]$. Zero values stand for background pixels and ones code pixels with maximum intensity. The two objects to be compared are two images I_in and I_ref with identical dimensions to I. The dynamical equation defining the gray-scale wave metric comparing two images is the following nonlinear partial differentiation equation (PDE) system:

$$\frac{\partial I_1(x_1, x_2 \ldots, x_n, t)}{\partial t} = \sum_{i=1}^{N} D_i \frac{\partial^2 I_1}{\partial x_i^2} + \gamma(I_{max} - I_1) \quad (1)$$

$$\frac{\partial I_2(x_1, x_2 \ldots, x_n, t)}{\partial t} = \omega(I_{max} - I_1) \quad (2)$$

where $I_{max}(x, y) = \max(I_{in}(x_1, \ldots, x_n, 0), I_{ref}(x_1, \ldots, x_n, 0))$ contains the pixel-wise maximum of I_1 and I_2 , $I_{min}(x_1, \ldots, x_n) = \min(I_{in}(x_1, \ldots, x_n, 0), I_{ref}(x_1, \ldots, x_n, 0))$, $I_1(x_1, \ldots, x_n, 0) = I_{min}$, $I_2(x_1, \ldots, x_n, 0) = 0$, $\gamma > 0$, $\omega > 0$ and D_i are constants defining the wave propagation Δ is the Laplace operator. The final wave is the map of the steady state solution of I_2. Here the change of magnitude of the wave at a pixel is also depends on the difference between the two pixel values.

Thee wave propagation in each direction can be controlled independently in each dimension by the D_i parameters. If the propagation coefficient is zero in a given direction then the differences along that dimension will be ignored and the metric should be calculated for only those dimensions which coefficients are non-zero.

It was shown that the gray-scale wave metric combines all advantages of the previously mentioned two methods [6] and can be applied for image comparison and medical imaging[7]. Both the Hausdorff and the Hamming metrics can be derived from the gray-scale wave metric. This itself would motivate the usage of this method, considering fast computation on CNNs[8].

Although these metrics can be used on two-dimensional inputs, their extension of higher dimensions, especially in case of changing objects (spatio-temporal comparison) has not yet been investigated.

In higher dimensions, like in case of gesture comparison it can easily happen that the objects are non-overlapping according to one or more dimensions, or they only have an extremely small common set of elements and unfortunately all of the commonly applied metrics will require overlapping objects for comparison.

To overcome this problem we need to introduce a fade-out in the objects and blur the them in the selected dimensions. This way we can extend the objects and the overlapping regions, creating non-empty intersections. The wave propagation can start properly from such initial condition and the blur will fade out according to the distance from the centroid, from the position of the original object. The further the object is, the lower the intensity will be and this will cause lower magnitude change in the wave -this way the magnitude change of the wave will be proportional to the original distance between the elements. This will ensure the properties and axioms of the metric and will give an applicable result.

4. Experiments and Results

We have recorded 7 different sets of trajectories, with four different and independent measurements in all of them. The seven sets were the following: fast lines, slow lines (both approximately rectilinear movements), large and small circles with both clockwise and counterclockwise motion and standby sets (without any movements or gestures).

Although this set is far from complete it is enough make experiments about trajectory comparison with the non-linear wave metric. It contains three different and generally used trajectories, which are differing in both spatial (line and circle like trajectories) and temporal (clockwise and counterclockwise circular trajectories) properties. It also reveals an important requirement of gesture recognition systems: the classifier must handle inputs where there are no gestures applied. We found two methods to determinate whether there was a gesture performed. One can either use different time windows on the recorded trajectories to identify when a gestures was started and to cut a part that contains the gesture or one can also start gesture identification when the movement is larger than a previously given threshold and stop the classification when the movement settles, the average change is lower than a threshold. We have selected and used the latter method, because it requires less computational power.

We have compared our recorded trajectories based on three different metrics: the Hamming-, Hausdorff-distances and the extended non-linear wave metric. Metric comparison was based on the relative difference between the metrics using the difference between two relatively similar, linear trajectories from the same class as reference. We calculated the relative difference from this comparison for each metric in percentage and used this for our comparison, because in case of classification a gesture is classified by its distance from the references. It its smaller than a given threshold from a reference trajectory the gesture will be classified. Some example comparisons can be seen in Table 1.

Our aim was to show that in all examined cases the distance for similar trajectories was the lowest and it was the

Table 1: . Because the number of recorded trajectories were high we were not able to list all the possible pairs here,we restricted ourselves to some relatively similar and some completely different trajectories. Rows are containing comparisons for different trajectory sets, Circle Cw and Circle CCw present clockwise and counterclockwise circular trajectories. The second column contains the distance measured by the Hamming metric, the third column shows the Hausdorff distance and the last column contains the same data for the non-linear wave metric. It can be seen that the new method resulted in the smallest distances between the similar gestures and the biggest between the different ones.

Traj1/Traj2	Ham.	Haus.	Ext. Wave
Circle/Line	35.2%	36.5%	17.0%
Circle/Circle	20.3%	16.2%	5.0%
Line/Line	42.6%	18.4%	2.4%
Circle CW/Circle Ccw	19.7%	17.9%	45.0%

highest for distinct trajectories. As it can be seen from the results the non-linear wave metric was the best representation of trajectory comparison amongst these methods. All of the gestures were classified correctly by our approach and non of the two other metrics could solve this task.

5. Conclusion

We have examined how the extension of the non-linear wave metric can be used in the comparison of four-dimensional trajectories. We have compared our method with commonly used metrics, like the Hamming- or Hausdorff-distance. from our experiments it can clearly be seen that amongst these metrics our method is the most suitable and applicable for gesture classification or for other type of classification in high-dimensional problem spaces.

We have developed a gesture recognizing environment based on the Leap Motion controller. Our system uses the non-linear wave metric for gesture classification and it can identify and selected previously learnt and stored gestures. This way the user can define a gesture set that is the most suitable and user-friendly for a given, specific task.

Acknowledgement

The support of the grants TÁMOP-4.2.1.B-11/2/KMR-2011-0002 and TÁMOP-4.2.2/B-10/1-2010-0014 are gratefully acknowledged.

References

[1] Torben Schou , Henry J. Gardner: "A Wii remote, a game engine, five sensor bars and a virtual reality theatre" *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*, 2007

[2] Iason Oikonomidis, Nikolaos Kyriazis, Antonis Argyros: "Efficient model-based 3D tracking of hand articulations using Kinect" *Proceedings of the British Machine Vision Conference*, September 2011

[3] Daniel Terdiman: "Leap Motion: 3D hands-free motion control, unbound" *Cutting Edge. c—net*, 2013

[4] Micahel B DuCharme: "The Operating System Handbook" *maben.homeip.net*, 1994

[5] Allen Kent; James G. Williams; Rosalind Kent: "Baby duck syndrome: Imprinting on your first system makes change a very hard thing" *CRC Press*, 2005

[6] István Szatmári, Object comparison using PDE-based wave metric on cellular neural networks *International Journal of Circuit Theory and Applications, pp 359-283, 2006. 8.7*

[7] Andras Horvath, Daniel Hillier: Topographic Metrics for Image Segmentation *12th IEEE international workshop on cellular nanoscale networks and applications, Berkeley, California, pp 1-4, 2010*

[8] István Szatmári, A nonlinear wave metric and its cnn implementation for object classification *Journal of VLSI Signal Processing, pp 437-447, 1999*