Tokyo University of
Agriculture and Technology

Object Shapes Recognition from Haptics

Author:
Marie-Morgane Paumard

Supervisors:
Dr. Ganesh Gowrishankar
Pr. Gentiane Venture

September 10, 2015

Table of Contents

Abstract. To efficiently manipulate an object, robots need to have a notion of its shape. Mostly done from vision, the shape recognition can also be computed from inertia parameters. They can be obtained from a prior manipulation, like blind people do when they face an unknown object. In this paper, we focus on understanding the strategies humans use to get inertia parameters for an unknown object. We propose a method to study these strategies by capturing humans' motions while waving objects, and asking them to guess their shapes. We find out that some people are successful at this task and we demonstrate that the performed movements and strategies depend only on the manipulated object.

1 Introduction

From the writer who tries another fountain pen to the tennis player who wants to purchase a new racket, some people are able to perceive the very little discrepancies when they manipulate slightly different objects than those they use on a regular basis. Based on this consideration, it appears that sightless object manipulation is an untapped way to determine shapes, or at least inertia properties. Endowing machines with such abilitie will certainly be fruitful, especially when visual shape recognition cannot be performed. For instance, blind robots or robots operating in dark or extreme environment can benefit from this capacity. Allegedly, the computing of inertia parameters does not need extra sensors. The data collected from movement suffice. Furthermore, shape recognition from manipulation can improve or corroborate the recognition done from vision.

GVlab has recently proposed an effective method to obtain inertia parameters from robot arm movement [1]. Since this study is human-inspired, a potent way to enhance it is to examine the information people gain from movements such as the weight, the centre of gravity coordinates or the inertia tensor. To improve the shape recognition from movements made by robots, we devise from scratch a method allowing to study and compare how people perform such recognition.

In this paper, we demonstrate that most people have great difficulties to figure out shapes from movement. However, their moves are discriminant enough to determine the type of object they are waving. We thus show that the people who succeed in this task are those who are good at drawing the right conclusions from their movements; they make comparable moves than those who fail.

This report is organized as follows. After giving an overview of the literature, we present the experimental choices we had to make and the protocol we designed to acquire data. Then, we expose the treatments we apply on our dataset and we bring to light some promising facts. Finally, we discuss the results we get in terms of optimization of the performance; we propose several research axis that will benefit from being exploited. We propose further experiments that can be conducted to refute or validate our conclusions.

2 Related work

In the literature, the question of information obtained from movement is barely raised and hardly never examined: there are plenty of papers dealing with weight perception, however the very few that address haptic shape recognition focus on how people draw conclusions from enclosure or contour following. In order to design a proper experiment, we have to examine related topics.

In their tutorial about haptic perception, Lederman and Klatzky [2] establish an overview of the sense of touch. This sense aims to combine tactile, sensorimotor and proprioceptive information to obtain the weight, the shape, the texture, the hardness and the temperature of an object. Bushnell and Boudreau [3] show that haptic sense begins to develop far after the sense of sight: children understand and start playing with shapes at the age of 9.5 months.

If it is possible to get an estimation of the weight without moving, the most precise method to apprehend it is to lift an object. The judgment of weight depends on the manner of lifting and on the size-weight illusion. Jones [4], [5] states that the weight perception is optimal when subjects are adapted to the range of weight they are discriminating. Likewise, children spend more time on unknown objects [3]. In another paper, Jones [6] describes major motor illusions. These illusions will influence our experimental choices. The author explains that real position and its perception are not correlated: the sense of the position derives from cross-calibration of vision and proprioceptive signals. Another illusion is the size-weight illusion: when people see two similar objects of the same size but of different weight, they convince themselves that the biggest one is the heaviest. This impression does not disappear even after lifting each object. Lederman [7] notes that the vision does not play a decisive role in this illusion: if the experimental subject can figure out that one object is bigger, for example by touching it, this knowledge will alter their judgment. For those reasons, we use, in our experiment, a handle to manipulate the objects and do not authorize people to see the objects. Nevertheless, this illusion allows people to lift unknown objects with the good muscular tension. We asked people to begin with small moves so they are not surprised by the weight of the object.

To blindly recognize a shape, people usually follow the contour of the object with their fingers [2]. They can deduce the global shape from inertia parameters. Schedlinski and Link [8] give an overview of the ways to mechanically get these inertia parameters. They proposed static and dynamic methods. The static ones are easier to setup but they can only compute the mass and the centre of gravity coordinates. The inertia tensor should be calculated with dynamic methods. Among the methods described by Schedlinkski and Link, few can be used by humans, notably the pendulum method. This involves using small angular motions to calculate the inertia parameters.

Since our final goal is to improve robot arm or androids, we throw a brief look at existing robotics methods using machine. Several can perform identification of the ten inertia parameters of a rigid body. Some are designed for this single task, such as Niebergall and Hahn's [9] whose inertia parameters identification has an acceptable accuracy. More recently, Katsumata and Venture [1] get satisfying results with an all-purpose robot arm.

3 Data acquisition

To observe if people are efficient in shape identification and how they proceed, we ran several pilot experiments. The collected data allowed us to determine the best parameters and to design the final experiment. It consists in asking 14 people to guess properties and shapes of several objects by waving them. To not bias the results, they had to hold a handle and could not touch nor see the object (Fig. 1).

Fig. 1: Experimental setting: the subject is holding the handle of the object and wears two straps containing the IMUs sensors.

3.1 Shapes choice

We designed and used 7 objects made of wood (Fig. 2a). They all come in a different shape but weight almost the same: $300\,\mathrm{g} \pm 20$. We choose to use this weight according to [4] because people can easily perceive the variations of weight at this range. The shapes are: sphere, cube, lying tube, standing tube, lying rectangle, standing rectangle and plane. The plane is used for demonstration.

The handle is placed so that the weight is evenly distributed and the centres of gravity of the objects are at the same distance of the handle.

(a) Set of shapes used, with handles.

3.2 Sensors choice

Former experiments showed that studying more limb movements than those of the forearm was redundant; we then used two inertial measurement units (IMUs). We attached the first one to the hand and the second one to the forearm, near the elbow. To constrain the movement, experimental subjects were asked to put their elbow on the table and to not move it. We choose to use IMUs because they are accurate enough because we can easily discriminate small moves. We also used a Kinect in order to check the elbow position: we use the method presented by Tuan and al. [10] to fusion Kinect and IMUs data. A camera was used for manual analysis and validation.

3.3 Experimental protocol

Each person had to manipulate nine objects. They were asked to use their favorite hand and to close their eyes. They had to perform some routine movements: those consist in an initialization sequence and a waiting sequence performed between each shape recognition (Fig. 3). The initialization sequence aims to synchronize the signal of both IMUs and correct misalignment issues. The waiting sequence facilitates the segmentation of the data.

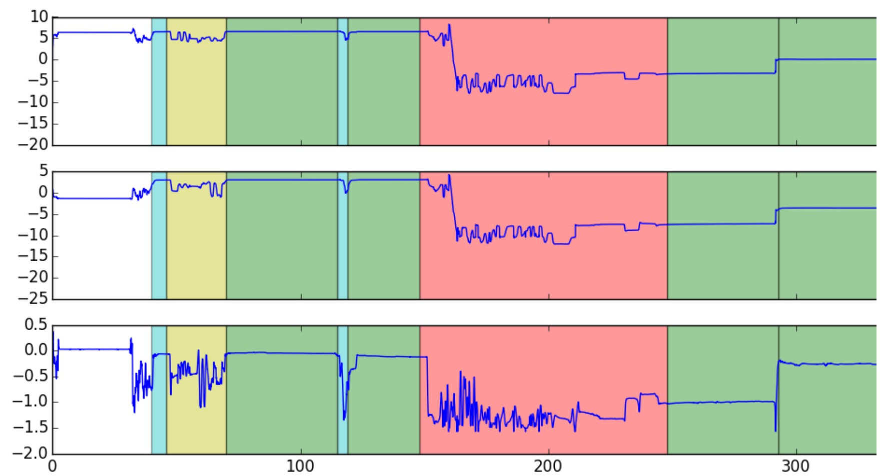


Fig. 3: Manual decomposition of an extract of a hand signal for each axis. The yellow zone corresponds to the initialization sequence, the blue zone is the moves for wearing the eyes mask, the waits are green and the manipulation is red.

In the very first pilot experiments we conducted, people were allowed to see or interact with the objects before blindly waving them. This led to making some people expressing false assumptions due to the size-weight illusion [6]: when an object is bigger, they supposed it was heavier. To thwart such behavior, we use sketches of the shapes. Regarding those that were allowed to manipulate the object before closing their eyes, they got a perfect score. They were able to remember and recognize the shapes they have waved before. It remains to understand the importance of the vision: is there some memory effect when manipulating without sight twice the same object? For the nine recognitions people had to make, we picked three objects and gave them three times each. Experimental subjects were told that they can receive the same object several times. The manipulation was not time-limited because we have seen in previous experiments that it causes stress and the recognition was less effective.

Former experiments showed that people did not try to grasp the global shape of an object but that they were thinking in terms of its properties. Thus we asked people to describe each object in terms of size and symmetry. Then, we asked them to give each shape a score of 0 to 5 which represents the percentage of chances that the object was the one they held before.

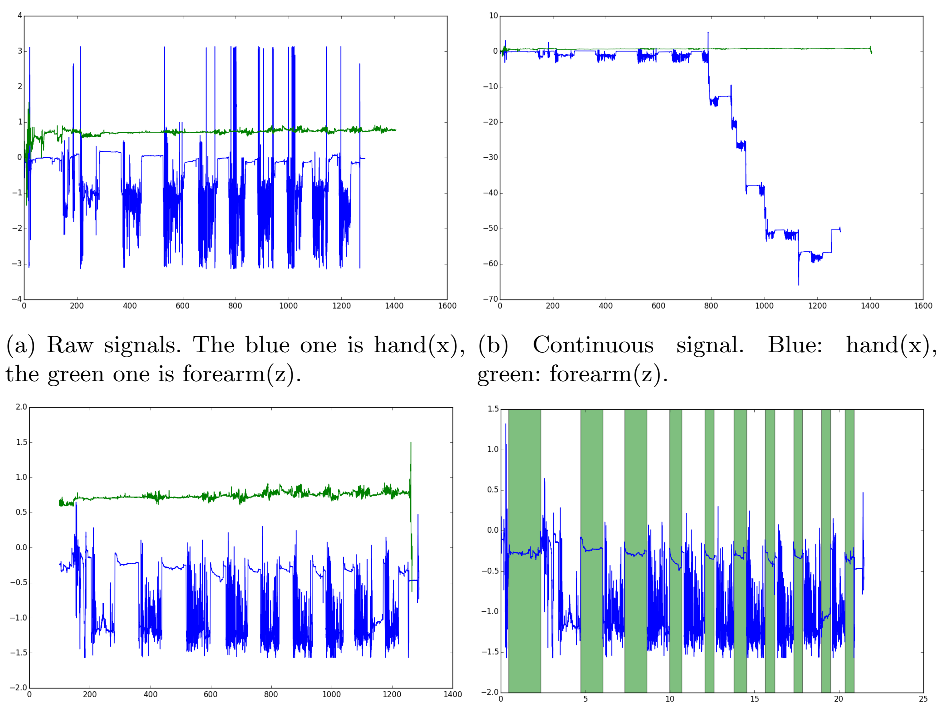
4 Data processing

For each experimental subject, we dispose of the following data: time, 3-d orientations and linear accelerations for both forearm and hand, video of the experiment, 3-d joint positions from Kinect and answers of the participants. To exploit this data, we have to format them and select the ones that will be useful. Before using Support Vector Machine (SVM) or K-Means, we reduced the dimension of the data with the Principal Component Analysis (PCA).

4.1 Preprocessing

Flattening Since the orientation data from the IMUs is varying from $-\pi$ to $+\pi$, we transform the torus signals into continuous signals: those are more easily exploitable (Figure 4b).

Time synchronization The IMUs are not synchronized in time (Figure 4a). We use the initialization sequence to set the time, by aligning peaks (Figure 4c).



(a) Raw signals. The blue one is hand(x), the green one is forearm(z).

(b) Continuous signal. Blue: hand(x), green: forearm(z).

(c) Time synchronization for z-axis.

(d) Segmentation for z-axis. Green zones correspond to the waits.

Fig. 4: IMUs data processing.

Segmentation Each signal of the IMUs contains the data of the nine objects, we thus had to divide them, using the waiting sequences (Figure 4d). In the following, a "move" refers to the complete manipulation of one object.

Submove To compare each move, we divide each into short segments of the same size. We will try to identify similarities, and study the influence of the lenght and the offset. To ease the comparison, the values of each segment are shifted to start at $(0, 0, 0)$.

4.2 PCA

We used the Principal Component Analysis (PCA) [11] in order to find another base to describe the data. This new base will allow us to reduce the data dimension. For instance, from all the submoves made from the orientation data, whose sizes are 600 — 2 joints $\times$ 3 axis $\times$ 100 points per submove, to describe 95% of the variance of the data, a descriptive base of 11 elements is required. The first and second dimension allows us to plot the data (Fig. 5).

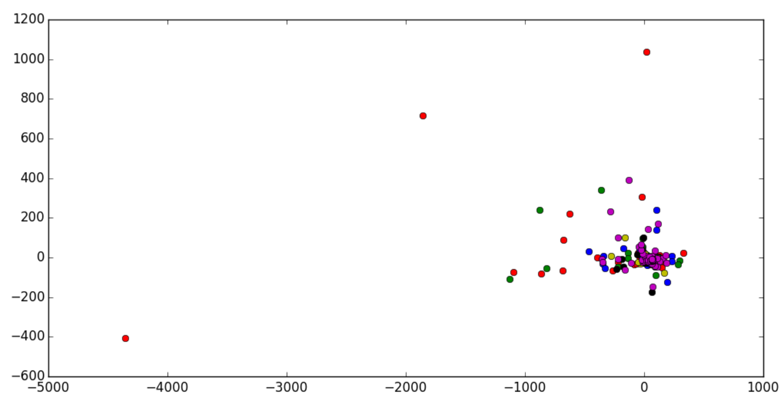


Fig. 5: Orientation data reduced with PCA in two dimensions. Each color represents a different shape. The submove size is 1000.

4.3 Machine learning

We have studied the result of both unsupervised and supervised machine learning algorithm. Unsupervised learning allows us to see if there are similarities in the submoves. With supervised learning, we can compare submoves with the interpretations of the experimental subjects or the objects.

Unsupervised learning We use K-Means [12] because it is general-purpose algorithm. Knowing the number of clusters, it allows us to find similar submovements.

Supervised learning Support Vector Machines (SVM) [13] draws our attention because it is effective in high dimensional spaces, even if the number of sample is lower than the number of the dimensions. This algorithm computes a classification based on training sample and their labels, then we try to predict the class of the test sample. When the training data is a part of the data, we apply our classifier on the rest of the data, the test sample. We use cross-validation to get better estimation of the success rate. When the training data and the test data are the same, the percentage of well-classified data indicates how distinct our data is.

5 Results

In this section, we examine the relation between movement data, the interpretation made by the experimental subjects and the shape of the objects. Thus, if we can deduce shapes from the interpretations, it implies that people are sucessful in recognition. If there is no link to be found between interpretation and shapes, it means that people failed. Nevertheless, the study of the relation between moves and shapes will reveal how useful the moves are; the subjects may have done efficient movements but may be unskilled to draw the right conclusions. Besides, we will discuss the influence of the shapes and study the moves of the successful subjects.

5.1 Overview of the data

First, we made some statistics on the interpretations made by all the experimental subjects. They first described the objects in terms of size and symmetry. Most of them are reasonably well characterized: we get 76% of correct evaluation. Fig. 6 shows the results of parameters identification for the best-recognized shapes: the cube. The accuracy of the estimation of the parameters depends on various factors. The weight and the size is estimated by comparison with the previous objects. The sizes for each axis is usualy evaluated by balancing moves in the desired axis. Some people make move in the other axis to compare and understand if the object is taller in one direction. The symmetry is the feeling that the shape will be the same if Y and Z-axis were inverted. People usually make rotative motions. They are less efficient in symmetry identification than size recognition. The curve is the feeling that the object is curve like a sphere: people do not perform specific moves but try to guess.

Each possible shape received a note from 0 to 5. If we consider that a score of 3 or more means that the object is recognized, we got 36% of object correctly identified. If we use the note to weight each group, we got 33% of recognition.

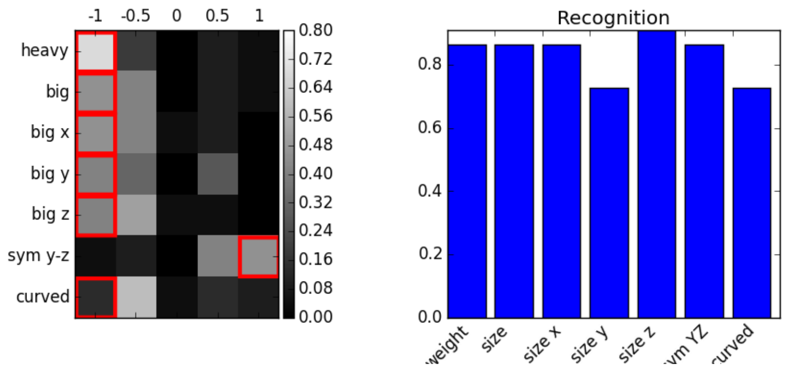


Fig. 6: Answers and success rate for the cube. The left figure shows the answers of the experimental subjects. The red squares indicate the right answers. If the answer is 1, it means True. The right figure is the success rate: the manipulation is successful if the answers are ± 0.5 close of the red squares.

The easiest object to recognize is the standing rectangle (Fig. 7); the cube get better parameters recognition but is often confused with the sphere. The lying tube and the lying rectangle are also confused. It is almost not the case for the standing tube and the standing rectangle, since the tube is the same in Y and Z axis. However, the standing tube is often confused with the cube and the sphere because of they are symmetrics and their centre of gravity is at the same distance of the handle.

5.2 Results of machine learning

We choose to focus our research on two main ideas: first, we compare the moves with K-Means and try to find if people uses different strategies. Then, we try to understand if there are links between movements and shapes with SVM.

K-Means application We apply K-Means on our orientation dataset, for different number of classes (from 3 to 2000). We found out that most objects are mainly described by some small moves around the initial position. Regardless the number of classes used, this cluster contains more than 70% of the moves done by the people. The remaining moves are well dissociated. When comparing results of the experimental subjects, we ascertain that the people were doing the same kind of moves: the clusters are not correlated to the people who do the moves.

SVM We apply PCA on our orientation data. We reduced the dimension of our data so that we can explain 95% of the variance of the data. Then, we use SVM, with the same learning and testing set. Having the submovement size varies, we find that we can correctly associate 71.0% to 95% of the data with the correct

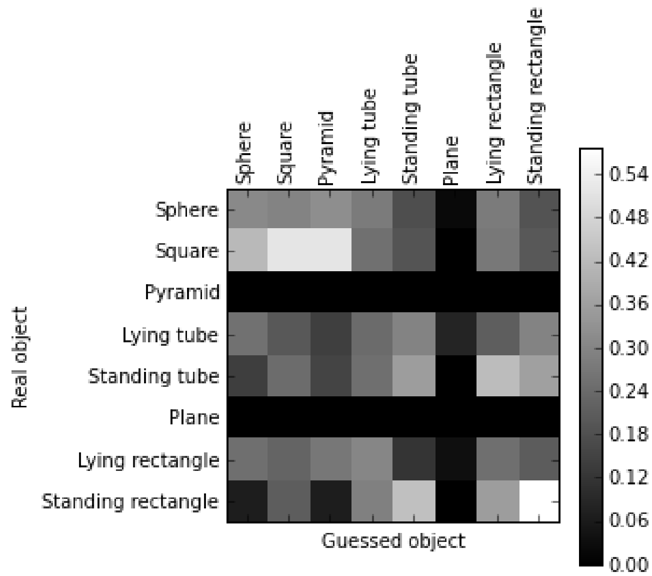


Fig. 7: Confusion matrix. The lighter a result is, the more answered it is. The plane is black because it was not given. The pyramid object does not exists, but it is listed in the answer list.

label. This result means that the moves contain some information that can be discriminatory. If we study the results for each person individually, we get 100% of correctly associate data with the label.

Then, we apply SVM on training datasets whose sizes are 10% of the orientation dataset. We use cross-validation and we obtain a recognition rate varying between 41% and 62% for different values of the size of the submovement. The results are 10% higher when studying people individually. This difference is mostly due to IMUs (see 6.1, second paragraph). Applying SVM on the movements give better results than asking people what shape they have in hands.

6 Discussion

6.1 Observations on the experiment

Significant improvements can easily be done to get better results. For example, the handle was not optimal: its edges guided the choices of moves performed during the shape recognition. A better handle would have been tubular and more light. Incidentally, the weight of the objects should be studied in detail and be exactly the same for each. It will also be interesting to know the influence of the position of the centre of gravity.

The motion capture system has some flaws: first, we had issues with synchronization. More importantly, when comparing two experimental subjects, problems related to IMUs arise. The orientation of the sensors cannot be exactly the same and the distance of the arm differs from one individual to another. We can use the initialization sequence to produce an automatic calibration method. Another solution to solve this issue is to use a single model of arm: the movements should be resized and their orientation corrected.

6.2 Validity of our results

Most people usually identify an object with two or three shapes and give each a note of 2 or 3. However, two people show high confidence: they give mostly extreme notes and usually pick one single shape. Since their success rate is the same as other people, their behaviors affect our data distribution. Scheduling another experiment with more people will allow us to decrease the influence of the self-confidence.

Even if they were told that they could receive several times the same object, most people act as if they had the opportunity to manipulate every shape. They try to associate each shape they moved with each possible shape; when they have an object in hand for the second time, they got worse score. These behaviors imply that it is not possible yet to make conclusions on the influence of memory of previous objects.

6.3 Additional ways to study our data

So far, we did not exploit all the data we had: Kinect data and angular acceleration data were not studied. They could be used to enhance the accuracy of our orientation measures and to solve the misalignment issues. We used all-purpose machine learning algorithms. We should optimize them better and compare them with other machine learning algorithms. We can also compare our results with those obtained without PCA. Since we use labels, we can replace PCA by LDA.

The study of the successful individuals will elucidate whether they were lucky or if they are good at interpreting their movements.

6.4 Further steps

So far, we have obtained inertia parameters by asking for some specific moves. Being capable to compute them during normal manipulation of objects will certainly be of great interest. Still, before doing so, we must sharpen our understanding of what is an useful moves. Ideally, we would be able to get rid of the handle — which is currently essential to not bias the results.

Another interesting experiment would be to extend our objects set to objects that contain liquid or are made with deformable material. We can also imagine the case of an unknown object placed in a bigger box: the manipulation of the box will make the object move and we can get profitable information from the variation of the inertia parameters of the box.

7 Conclusions

In this paper, we proposed a method to examine the recognition of object shapes from movements. Since the shape and the inertia parameters were very rarely studied, we were inspired by the various works on weight and forces perception. We ran pilot experiments to find what will make our experiment effective. We applied various analysis methods to our data in order to get an overview of the study paths we have and to decide what methods deserve further optimization. The few results we obtained are encouraging: with good parameterization, it is conceivable to obtain a grasp of the shape. In our experiment, we find out that the moves made slightly differ according to the shape of the object; people usually adopt the same kind of moves to identify parameters such as the size or the symmetry. The recognition of object shapes from movement is a promising area.

References

1. T. Katsumata, G. Venture: ロボットアームによる手持ち質量の 力学 同定 と形状. DVD Proceedings ROBOMEC, 2015.
2. S. J. Lederman, R. L. Klatzky: Haptic perception: A tutorial. Attention, Perception & Psychophysics, 2009, Vol.71, No.7, 1439-1459.
3. E. W. Bushnell, J. P. Boudreau: The Delevopment of Haptic Perception During Infancy. The psychology of Touch, 1991, edited by M. A. Heller, W. Schiff, Chapter 6.
4. L. A. Jones: Perception of Force and Weight: Theory and Research. Psychological Bulletin, 1986, Vol.100, No.1, 29-42.
5. Lynette A. Jones: Peripheral Mechanisms of Touch and Proprioception. Canadian Journal of Physiology and Pharmacology, 1994, Vol. 72, No. 5, 484-487
6. L. A. Jones, Motor illusions: What do they reveal about proprioception? Psychological Bulletin, Vol. 103, No. 1, 1988, 72-86
7. R. R. Ellis, S. J. Lederman: The role of haptic versus visual volume cues in the size-weight illusion. Perception & Psychophysics, 1993, Vol. 53, No. 3, 315-324.
8. C. Schedlinski, M. Link: A Survey of Current Inertia Methods Parameter Identification Methods. Mechanical Systems and Signal Processing, 2001, Vol.15, No.7, 189-211.
9. M. Niebergall, H. Hahn: Identification of the Ten Inertia Parameters of a Rigid Body. Nonlinear Dynamics, August 1997, Vol. 13, No. 4, 361-372
10. Y. Tuan, X. Meng, D. Tao, D. Liu, C. Feng: Upper limb motion tracking with the integration of IMU and Kinect Neurocomputing, 2015, Vol.159, No. 1, 207-218
11. H. Hotelling, Analysis of a Complex of Statistical Variables with Principal Components: Journal of Educational Psychology, 1933.
12. J. A. Hartigan, M. A. Wong: Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society, 1979, Series C, Vol.28, No.1, 100-108.
13. C. Cortes, V. Vapnik: Support-vector networks. Machine Learning, 1995, Vol.20, No.3, 273.
14. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay: Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 2011, Vol. 12, No. 1, 2825-2830.