

Design and Evaluation Case Study: Evaluating The Kinect Device In The Task of Natural Interaction In A Visualization System

Rafael M O Libardi

*Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo
São Carlos, 13566-590, Brazil*

mira@usp.br

Agma J M Traina

*Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo
São Carlos, 13566-590, Brazil*

agma@icmc.usp.br

Jose F Rodrigues Jr

*Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo
São Carlos, 13566-590, Brazil*

junio@icmc.usp.br

Abstract

We verify the hypothesis that Microsoft's Kinect device is tailored for defining more efficient interaction compared to the commodity mouse device in the context of information visualization. For this goal, we used Kinect during interaction design and evaluation considering an application on information visualization (over agrometeorological, cars, and flowers datasets). The devices were tested over a visualization technique based on clouds of points (multidimensional projection) that can be manipulated by rotation, scaling, and translation. The design was carried according to technique Participatory Design (ISO 13407) and the evaluation answered to a vast set of Usability Tests. In the tests, the users reported high satisfaction scores (easiness and preference) but, also, they signed out with low efficiency scores (time and precision). In the specific context of a multidimensional-projection visualization, our conclusion is that, in respect to user acceptance, Kinect is a device adequate for natural interaction; but, for desktop-based production, it still cannot compete with the traditional long-term mouse design.

Keywords: Kinect, Natural Interaction, Information Visualization, Participatory Design.

1 INTRODUCTION

Natural interaction, as defined by Valli [1], seeks for techniques to manipulate a given environment in a way that the user shall be able to transition from novice to expert with minimal, or even, no training. Natural interaction is a promising research area that has produced important results; Chang *et al.* [2], for example, use this paradigm to aid in the rehabilitation of patients physically disabled; in another work, Gallo *et al.* [3] employ natural interaction for manipulating medical images through an interface that demands no physical contact.

In this context, in 2010, the Microsoft company introduced the Kinect device; a mechanism whose primary goal was to allow game players to interact without a joystick. Besides its initial motivation, Kinect has given rise to a universe of new possibilities for use and research. As an illustration, a simple search for the term Kinect in the open access Digital Bibliography & Library Project (DBLP) returns over 150 paper titles¹.

¹ <http://www.dblp.org>, accessed in February, 2014

Considering these new possibilities, we have considered Kinect for a visualization research project, carried at the University of Sao Paulo and Embrapa², that could benefit from natural interaction. However, the adequacy of Kinect for this project had to be systematically verified since it represented a different method of operation, not previously tested. In this work, hence, we report on this empirical experimentation according to the following hypothesis:

“The Kinect device is tailored for defining more efficient interaction if compared to the commodity mouse device in the specific context (multidimensional projection) of an information visualization system.”

It is worth noting that we aim specifically at determining the adequacy of Kinect for common tasks regarding information visualization multidimensional projection; that is, identifying clusters and outliers. Despite the specificity of the hypothesis, we discuss design and usability issues with implications to the general usage of Kinect as the interaction device of choice.

Hence, over this methodological hypothesis, we present the steps of the development, including the technologies used to have Kinect integrated and functional; technique Participatory Design (ISO 13407³), used in the design stage of the system; and technique Usability Tests [4][5], used in the evaluation stage. In all the stages, Kinect was an active factor in such a way that we could observe, annotate, and analyze some of its features.

Similar to our work, other authors evaluate the possibilities of Kinect. Andersen *et al.* [21] inspect its adequacy for computer vision; they found that, although with some limitations, Kinect can be useful in many project configurations. LaBelle [22] investigates Kinect's use in a medical setting; she discovered that it can be successfully applied in rehabilitation, but some issues must be satisfied for it to work well. Khoshelham and Elberink [23], in turn, consider Kinect for the task of indoor mapping applications; they conclude that a distance of no more than 3 meters must be respected. Finally, Kefer and Kubinger use Kinect [24] in robotics, they conclude that it has many limitations, but its popular appeal may be of reasonable value. Just like these works, we evaluate Kinect in a specific setting: information visualization. Also similar, our conclusions point the relative adequacy of Kinect, which can be used provided some issues are addressed.

The rest of the paper is organized as follows. Section 2 presents theoretical fundamentals; section 3 describes our methodology; section 4 describes the development; and section 5 presents experiments and analysis. Section 6 concludes the paper.

2 THEORETICAL FUNDAMENTS

In this section, we review concepts related to the present research, providing readers the background necessary to follow the steps of the research.

2.1 Natural Interaction

Computer devices have evolved so that users may interact with computers the same way that they deal with everyday things, that is, with gestures, expressions, movements, voice commands, touch, and vision. This way, users can learn how to use a computer in the same way that they instinctively learn; that is, by observation, manipulation, trial and error [6]. This method of interaction is named natural interaction.

Natural interaction, however, is not suited for every application; Del Bimbo [7] states that industrial automation, production management, stock market, and even word processing software work better with traditional interactive devices. The author also argues that natural interaction is more adequate for applications like object manipulation – the one in the present work, smart environments, personal assistance, and entertainment.

² Empresa Brasileira de Pesquisa Agropecuária

³ <http://zonecours.hec.ca/documents/A2007-1-1395534.NormelSO13407.pdf>

In the last years, great advances occurred in the field of natural interaction; new devices equipped with high resolution cameras, depth sensors, sensitive microphones, and accelerometers brought outstanding possibilities. These devices came mainly from the game industry, whose competitive nature pushes forward the need for innovation. In 2006, game company Nintendo launched the Wiimote control [8], in 2009 Sony announced the PlayStation Move [9] and, in the same year, Microsoft presented the Kinect device – see Figure 1. Each of these devices conveyed solutions for more natural interaction and, due to authoring rights, they all differ in their conceptual design.



FIGURE 1: Interaction Devices from (a) Nintendo, (b) Sony, and (c) Microsoft.

2.2 Microsoft Kinect

Kinect was conceived to be used with game console Microsoft Xbox 360. It has an array of microphones, and a 640x480 pixels RGB camera that captures 30 frames per second combined with depth information obtained from two infrared sensors. Kinect's Software Development Kit [10] efficiently abstracts these features being equipped with voice recognition algorithms to process audio as a sequence of voice commands, or like a dictation stream. It also features video processing to provide regular video streams, or to work as a skeletal tracker, which identifies a set of segments corresponding to 20 human body parts of up to 6 users simultaneously.

2.3 Flexible Action and Articulated Skeleton Toolkit (FAAST)

FAAST [11] was conceived as a middleware capable of interpreting the output produced by Kinect simplifying the development of applications. It can also be used as a server that transforms the output of Kinect into simulated keyboard and mouse events, sending these events to client applications. In other words, FAAST translates natural interaction into traditional interaction events, making it possible to have existing applications work with natural interaction with no need to port code. This possibility provides rapid prototyping and testing of interaction schemes, saving on the complexity of adapting existing software.

FAAST is based on two kinds of information: actions and articulated skeleton; actions refer to movements that can be performed by body parts, and the articulated skeleton refers to the scheme defined by such parts; a total of 24 parts, or joints. With a pair (action, joint), it becomes possible to identify natural interaction events performed by a user along with information of time and speed, as for example, (wave, left hand, 5 cm/s), or (to the right, right foot, for 3 seconds). Data like these can be coded by the FAAST server into events like mouse left, mouse click, or character "c", which are understood by existing applications.

We use FAAST to prototype the interaction scheme of our visualization software MetricSPlat – introduced in Section 2.6. We use it during the design and, later, for testing the final scheme.

2.4 Participatory Design

The Participatory Design [12][13] is a methodology to negotiate the traits that a given system will have by consulting its stakeholders – employees, partners, customers, end users. The negotiation works by having the participants to express their interests, preferences, expectations, abilities, and previous knowledge. The idea is to transform the needs and the work routine of the users into design decisions that will lead to higher levels of satisfaction [14][15]. Due to the broad

dissemination of the Participatory Design, it was organized as a norm under authority of the International Organization for Standardization, being named ISO 13407.

The norm ISO 13407 determines five cyclic stages to be carried on: understand and specify the use context; specify the user and the organizational requisites; produce a design solution; and evaluate the design with respect to the requisites.

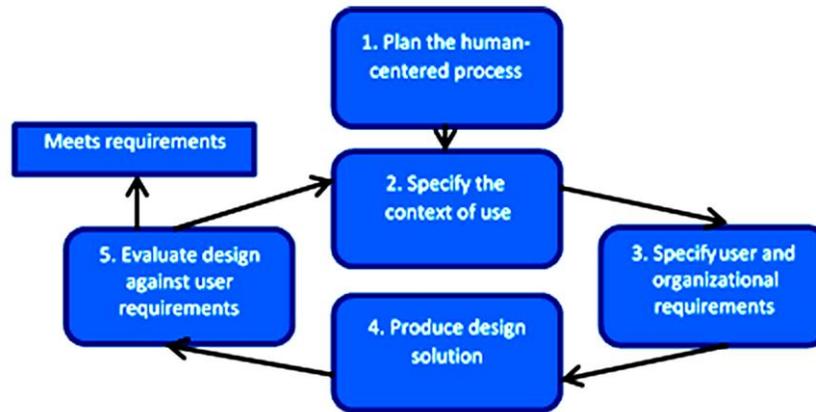


FIGURE 2: Participatory Design Stages.

Figure 2 illustrates the cycle of the Participatory design, according to which five stages are repeated, ideally, until the requirements are met.

2.5 Usability Test

Following the reasoning of user participation, a suitable manner of evaluating a software interface is to use empirical Usability Tests, a user-centered technique also defined in ISO 13407. This kind of evaluation seeks to validate the learning ratio, facility, efficiency in use and in production, user satisfaction, flexibility, usefulness, and security [4][5]. The metrics used to quantify such features include time to accomplish a task, number of mistakes, and ratio of declared user satisfaction. In a complementary work, Nielsen [16] proposes that the detected problems be classified according to three categories: catastrophic, serious or cosmetic.

Despite the fact that there are other techniques for validating software requirements, the usability test is considered an unavoidable process, since it predicts the involvement of the user as the greatest expert. It is also recommended that usability tests be carried on with a small number of users, no more than five; the argument is that once a few users have detected a problem, chances are that the same results will be accomplished with a bigger number of subjects [17][18].

2.6 MetricSPlat

System MetricSPlat [19], which is the target platform in this work, is a system enabled with data retrieval and visualization functionalities. Given a set of multivariate data, MetricSPlat uses content-based data retrieval to answer to similarity-based queries; the answers (subsets of multivariate data), in turn, go through dimensionality reduction algorithm Fastmap [20] to be plotted as a cloud of points in a multidimensional 3D projection. Over the final visualization, one can rotate, translate, scale, and select points – see Figure 3; tasks performed with the mouse, but that could be promptly adapted to Kinect.

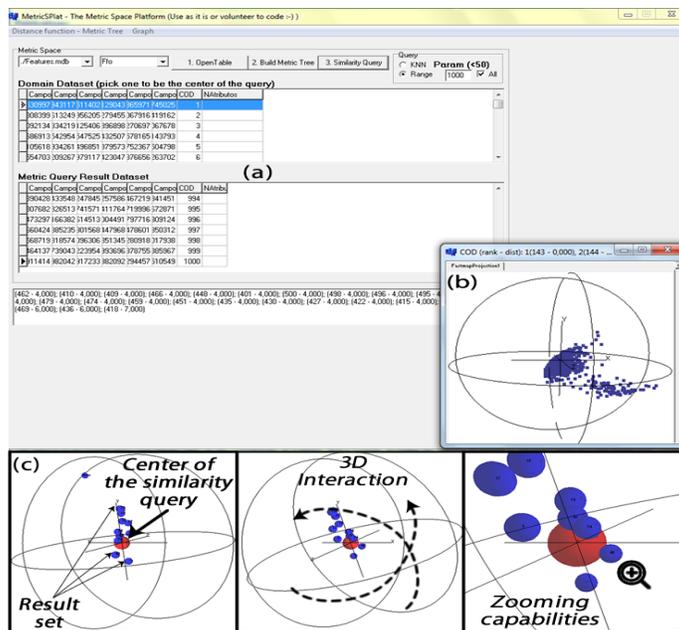


FIGURE 3: Illustration of the MetricSPlat system. (a) The data and the retrieval results. (b) Corresponding cloud of points visualization. (c) Interaction possibilities.

2.7 Embrapa Dataset

We perform experiments over a dataset of agrometeorological data - the Embrapa dataset. The dataset has 9 attributes: precipitation, maximum temperature, minimum temperature, normalized difference vegetation index (NDVI), water requirement satisfaction index (WRSI), average temperature, potential evapotranspiration (ETP), real evapotranspiration (ETR) and measured evapotranspiration (ETM). These data were collected partly with remote sensors (satellite) and partly with in locus samples from sugar cane plantation regions in Brazil. Five Brazilian regions were considered: Araraquara, Araras, Jaboticabal, Jau, and Ribeirao Preto, summing up to 410 records. The goal of the data collection is to visually analyze the relationship between the weather (temperature and humidity) and the agricultural production.

3 METHODOLOGY

In order to evaluate the hypothesis postulated in the introductory section, we proceeded by defining an experimentation setting over the MetricSPlat system. To this end, we used the FAAST middleware in order to have gesture-based interaction working with the visualization. For the task of determining how MetricSPlat could be better controlled using Kinect, we used the Participatory Design combined to technique Usability Test. This way, the system would have a well-designed interaction scheme and, only after, this scheme would be compared to mouse-based interaction. This was an important step because, otherwise, the Kinect-based interaction could fail simply due to a bad design, leading us to a potentially flawed conclusion. After achieving an acceptable Kinect-based interaction scheme, we proceeded with technique Usability Test to compare devices Kinect and mouse.

3.1 Design Process

The Participatory Design was conducted according to the specifications presented in Table 1, as predicted by standard ISO 13407.

#	Stage	Definitions
1	Planning	<ul style="list-style-type: none"> • six-months period • two users per cycle
2	Specification of user context	<ul style="list-style-type: none"> • reserved room with enough space • a set of tasks to be performed by each user
3	Specify user and organizational requirements	<ul style="list-style-type: none"> • real-time system response • log of usage indicators • experienced users • introductory training • the system shall produce good indicators in accordance to the users' operation
4	Production of a design solution	<ul style="list-style-type: none"> • prototyping using MetricSPlat over FFAST • dynamic design
5	Evaluate design against user requirements	<ul style="list-style-type: none"> • usability Tests
6	Verification of the requisites	<ul style="list-style-type: none"> • documented verification • design decisions endorsed by collected indicators

TABLE 1: Participatory Design Specification.

As will be presented in Section 4.2, system MetricSPlat adapted with natural interaction went through these 6 stages for each prototype version. After each cycle, a new prototype was created in a total of three designs.

3.2 Evaluation Through Usability Tests

For each cycle of the Participatory Design, in stage 5 (Evaluate design against user requirements), the test of the system was conducted with each user going through a sequence of steps: introductory training, dataset description, explanation of the proposed tasks, execution of the tasks, and answering of the satisfaction questionnaire. The questionnaire was scaled 1 (little) through 5 (very much) with the following questions:

- a) How physically bothersome was to take part of this experience?
- b) How psychologically bothersome was to take part of this experience?
- c) How easy was to perform the proposed tasks?
- d) How exhaustive was to perform the proposed tasks?
- e) How convenient was to take part of this experience?

3.3 Hypothesis Evaluation

After the prototype is finished, we comparatively evaluated the Kinect and the mouse – explained in Section 5. We had users performing a set of tasks, the same way we did during the Participatory Design, but with different users, with a broader questionnaire, a log with numerical indicators and a set of observed numerical indicators. The questionnaire was extended with questions:

- f) How appropriate did you consider using Kinect for this experience?
- g) How inclined are you to substitute the mouse with Kinect?

The log of numerical indicators constituted the following set of measures both for the mouse and for the Kinect:

- Number of users;
- Number of users who completed the experience;
- Number of user assistances during the experience;
- Time to complete the experience;
- Number of rotations;
- Number of scales;

4 DEVELOPMENT

Here, we describe the development of the prototype aided with gesture-based interaction.

4.1 FAAST and MetricSPlat

Once the methodology was specified, the first thing was to configure the FAAST server to work with MetricSPlat. We bring together Kinect functionalities to MetricSPlat by means of gesture-based commands rapidly encoded through the FAAST server. Besides this initial configuration, we modified MetricSPlat to log the commands issued by the user; including number of rotations, scales, translations, and session time.

4.2 Design Process

With the system enabled with gesture-based interaction, we worked with the possibilities concerning rotation, scaling, and translation.

First Prototype

The initial interaction scheme was defined following common-sense principles, as illustrated in Figure 4.

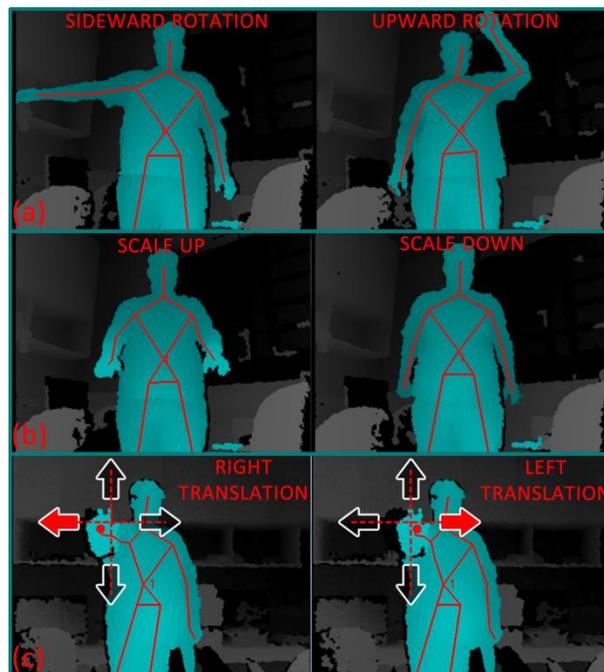


FIGURE 4: Gesture-based interaction design – first prototype: (a) rotation, (b) scale, and (c) translation.

We followed the preparations and the stages predicted in Section 3.1 and 3.2 with two users, who reported results regarding stage 5 of the Participatory design, as presented in Table 2:

Stage	User 1	User 2
How physically bothersome was to take part of this experience?	3	3
How psychologically bothersome was to take part of this experience?	1	1
How easy was to perform the proposed tasks?	4	3
How exhaustive was to perform the proposed tasks?	2	1
How convenient was to take part of this experience?	2	3

TABLE 2: Usability Indicators of the First Prototype.

Besides answering the questionnaire, the users reported two problems that were considered serious:

- Operation “rotation downwards” is misunderstood by the system as operation “scale down”;
- Operation “scale down”, which uses both arms, is bothersome and hard to execute.

These results led us to a new cycle of design and user participation to correct these concerns and, also, to improve the overall satisfaction as reported in the questionnaire.

Second Prototype

In the new version of the system, the scale operation was altered so that the user became able to scale up the system by moving only the head forward, and to scale down by moving only the head backward – as illustrated in Figure 5.

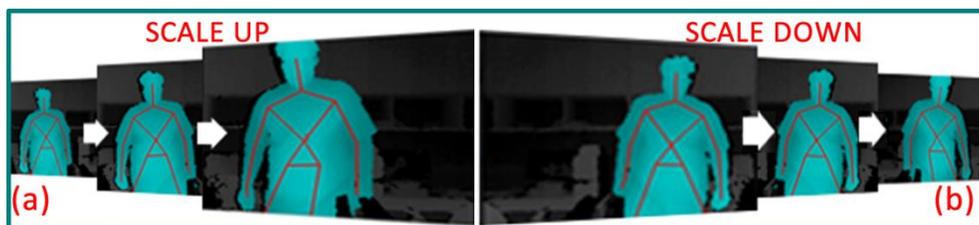


FIGURE 5: Scale operation as defined in the second prototype of the system. (a) Head forward corresponds to scale up; (b) head backward corresponds to scale down.

After altering the first prototype, we had two other users testing the system. The results for the second prototype are reported in Table 3:

Stage	User 1	User 2
How physically bothersome was to take part of this experience?	2	1
How psychologically bothersome was to take part of this experience?	1	1
How easy was to perform the proposed tasks?	5	5
How exhaustive was to perform the proposed tasks?	1	1
How convenient was to take part of this experience?	4	3

TABLE 3: Usability Indicators of the Second Prototype.

One of the users reported one more problem, which was considered a cosmetic problem:

- Operation “rotation” could be more intuitive.

Once more, we verified the possibility of improving the system, what led us to the third prototype.

Third Prototype

In the third prototype, the operation of rotating the system around axis y was defined as the movement of the user’s head and trunk similar to that of peeking left or right. The rotation of the system around axis x was not altered, still being dependent on the movement of the arm, up or down.

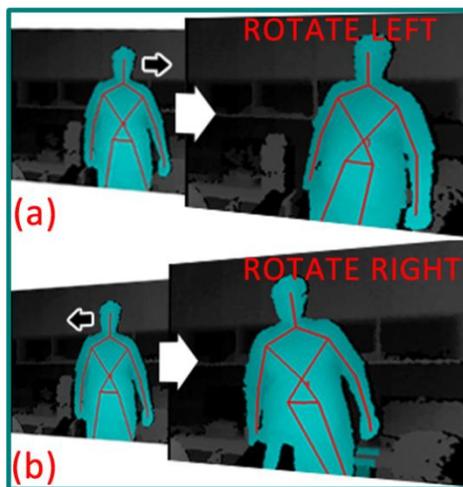


FIGURE 6: Rotation operation as defined in the third prototype. (a) Head and trunk to the left correspond to left rotation; (b) head and trunk to the right correspond to right rotation.

Again, we had two other users testing the system; in this third prototype, the results of the questionnaire were nearly optimum so that we could consider the system sufficiently good.

We observed that the Participatory Design was an interesting approach. Instead of having the system go to production, receiving a demand for alteration from the users, correcting it, and then

resubmitting it for use and appreciation, we had a much shorter development process. The Participatory Design allowed us to quickly determine the problems and to understand what were the users' concerns through a direct dialog.

4.3 Initial Usability Experiments

Over the final prototype, we compare the gesture-based interaction with the traditional point-and-click mouse interaction; in this section we conduct preliminary tests and, in Section 5, we present further experiments to statistically verify our suppositions. For the comparison process, we collected quantitative and qualitative measures, as explained in Section 3.2.

The preliminary experiments – summarized in Section 4 – were performed with five users, each of them using both the interaction mechanisms under investigation.

Measure	Total
Number of users	5
Number of users who completed the task with the mouse	5
Number of users who completed the task with Kinect	5
Number of users who demanded assistance during the experiment with the mouse	3
Number of users who demanded assistance during the experiment with Kinect	3

TABLE 4: User Participation in the Initial Experiments.

4.3.1 Collected Data

As detailed in Sections 3.2 and 3.3, we collected the number of user assistances, time to complete the task, number of rotations and scales.

Number of User Assistances

The number of times the users asked for assistance was not high, neither for Kinect nor for the mouse (Table 5). This observation certifies that with the preceding training that was carried out prior each experimentation, the users became equally prepared, so there was no bias due to lack of proper preparation. Despite the doubts that occurred, both for Kinect and for the mouse, the users were able to accomplish the proposed tasks. Nevertheless, Kinect generated less doubts, possibly due to the intuitive design that gesture-based interaction brings on.

Time to Complete the Task

In Table 6, we can see that Kinect had the highest standard deviation, indicating the users demonstrated a significant variation in relation to their ability in dealing with gesture-based interaction.

Aggregate	Kinect	Mouse
Average	0,6	1,6
Minimum	0	1
Maximum	1	3
Standard deviation	0,54	0,89
Confidence interval of 95%	0,48	0,78

TABLE 5: Summary table for the number of user assistances.

Aggregate	Kinect	Mouse
Average	4,52	2,89
Minimum	3,14	2,14
Maximum	5,95	3,47
Standard deviation	1,12	0,49
Confidence interval of 95%	0,98	0,43

TABLE 6: Summary table for the time to complete the task.

Number of Rotations and Scales

The number of movements in the visualization suggests some extra effort to accomplish interactive actions with Kinect. Both for rotation –Table 7, and for scale –Table 8, the users performed a larger number of interactive moves; many times they had to repeat a sequence of steps due to some imprecision in performing a given task. Together with the higher time to complete the tasks, this was a negative aspect in terms of the Kinect efficiency.

Aggregate	Kinect	Mouse
Average	24,6	16
Minimum	19	11
Maximum	32	22
Standard deviation	4,87	4,52
Confidence interval of 95%	4,27	3,96

TABLE 7: Summary table for the number of rotation movements.

Aggregate	Kinect	Mouse
Average	10,8	3,2
Minimum	7	2
Maximum	13	4
Standard deviation	2,68	0,83
Confidence interval of 95%	2,35	0,73

TABLE 8: Summary table for the number of scale movements.

4.3.2 Questionnaire

In the same manner as in the Participatory Design, in the questionnaire, the users were able to express their perception of the experience from seven different perspectives in a scale ranging from 1 (little) through 5 (very much).

a)/b) How physically/psychologically bothersome was to take part of this experience?

The answers – see Table 9 – showed that Kinect significantly required higher physical effort from the users. Meanwhile, the same answers in respect to the psychological demand did not show significant results, as all the users rated this factor at level 1 for Kinect and for the mouse (not shown).

c) How easy was to perform the proposed tasks?

This question indicated that the users found it easy to perform the proposed tasks for both devices – see Table 10, with a significant advantage in favor of the mouse, naturally less demanding than the gesture-based Kinect.

User	Physical bothersome level	
	Kinect	Mouse
1	1	1
2	2	1
3	1	1
4	2	1
5	1	1

TABLE 9: Answers to questionnaire question b) “How physically/psychologically bothersome was to take part of this experience?”

User	Easiness level	
	Kinect	Mouse
1	4	5
2	5	4
3	3	5
4	4	4
5	4	5

TABLE 10: Answers to questionnaire question c) “How easy was to perform the proposed tasks?”

d) How exhaustive was to perform the proposed tasks?

Here, the users declared that the mouse caused no exhaustion at all, while Kinect caused a notable demand of extra energy from the users, as presented in Table 11.

e) How convenient was to take part of this experience?

Table 12 shows that the user satisfaction was higher with Kinect; Table 13 confirms that, on average, the satisfaction ratio was 50% higher when gesture-based interaction was used.

User	Exhaustiveness level	
	Kinect	Mouse
1	2	1
2	1	1
3	3	1
4	1	1
5	1	1

TABLE 11: Answers to questionnaire question d) “How exhaustive was to perform the proposed tasks?”

User	Convenience level	
	Kinect	Mouse
1	4	3
2	5	2
3	3	3
4	4	2
5	5	4

TABLE 12: Answers to questionnaire question e) “How convenient was to take part of this experience?”

Aggregate	Kinect	Mouse
Average	4,2	2,8
Minimum	3	2
Maximum	5	4
Standard deviation	0,83	0,83
Confidence interval of 95%	0,73	0,73

TABLE 13: Summary table for the answers to question e) “How convenient was to take part of this experience?”

f) How appropriate did you consider using Kinect for this experience?

Concerning the appropriateness of the devices, both of them demonstrated good rates for all the users – see Table 14, with a slight advantage in favor of Kinect as detailed in Table 15.

g) How inclined are you to substitute the mouse with Kinect?

This measure, as seen in Table 16, showed that the majority of the users expressed a high concern in respect to changing their interaction paradigm to Kinect. Only one of the users rated his preference below level 3. Clearly, this, and the two former questions, demonstrated a stronger feeling of satisfaction, as expressed by the users, in using the Kinect device.

User	Appropriate level	
	Kinect	Mouse
1	4	4
2	5	3
3	4	4
4	3	3
5	5	3

TABLE 14: Answers to questionnaire question f) “How appropriate did you consider using Kinect for this experience?”

Aggregate	Kinect	Mouse
Average	4,2	2,8
Minimum	3	2
Maximum	5	4
Standard deviation	0,83	0,83
Confidence interval of 95%	0,73	0,73

TABLE 15: Summary table for the answers to questionnaire question f) “How appropriate did you consider using Kinect for this experience?”

User	Inclination Level Mouse » Kinect
1	4
2	2
3	4
4	3
5	4

TABLE 16: Answers to questionnaire question g) “How inclined are you to substitute the mouse with Kinect?”

5 FURTHER EXPERIMENTS

For statistical reasons, we conducted an extended usability experiment to compare the traditional point-and-click mouse interaction against the gesture-based interaction provided by Kinect. Again, we collected quantitative and qualitative measures, but now we had the MetricSPlat

visualization system (section 2.6) used by 20 users, over two different datasets, and considering two analytical tasks.

5.1 Experiment Setting

The 20 users comprised people with mild and advanced experience on computer usage, teenagers and adults, who had never experienced with visualization software. The experiments were conducted with a 52-inch screen at 2 meters distance, in a controlled environment. Before the experiment, the users went through a training phase over the Embrapa dataset (section 2.7) to get familiar with the visualization software, with the devices, and with the tasks they would have to perform.

After training, the users used two datasets from the UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/datasets.html>) for the experimentation phase:

- Cars: a dataset with 8 attributes that describe the characteristics of 398 different car models available at <http://archive.ics.uci.edu/ml/datasets/Auto+MPG>; the dataset has three well-defined clusters, groups of points visually evident; and also a number of outliers, cars that, by far, do not fit in any of the clusters;
- Iris: a dataset with 4 attributes describing the characteristics of 150 flowers of the Iris specie; the dataset has three well-defined clusters (subspecies Setosa, Versicolour, and Virginica); and also some outliers.

With each dataset, the users had two tasks in the following order: 1) identify the clusters; 2) identify the outliers; two of the main tasks usually performed in visual analytics. To do so, it was necessary to interact with the visualizations in a way that the 3D perception would become noticeable – the visualization is not stereoscopic, so the 3D perception comes only when the graphics are in movement. Also, it was necessary to see the data under different perspectives, because, as with any 3D projection, the data is occluded depending on the perspective from which one observes the points. The data and the tasks were defined so that any user could verify these aspects with the minimal training that was provided; we verified that all the users were able to complete the tasks. Although, based on visual analytics, the tasks were chosen and configured not to be over-challenging, as it would deviate from the main purpose of the experiment.

The users were organized in 2 groups of 10 users. The first group performed first using the mouse over the Cars dataset, then using the Kinect over the Iris dataset. The second group performed first using the mouse over the Iris dataset, then using the Kinect over the Iris dataset. We used this configuration for two reasons: first, the users would have totally different datasets to work with each device, so that the tasks would not become easier at the second round with Kinect; otherwise the users would already know the answer. Second, we alternated the datasets for each group/round so that, later, we would be able to verify whether the datasets introduced any bias in the experiment. In summary:

- Users 1 through 10
 - Mouse experimentation: Cars dataset
 - Kinect experimentation: Iris dataset
- Users 11 through 20
 - Mouse experimentation: Iris dataset
 - Kinect experimentation: Cars dataset

Collected Data

For each task, user, and device, we collected quantitative measures: time in seconds, number of rotations, scales, and translations – Section 5.2. We also assessed the perceptions of the users by using a questionnaire with 7 qualitative questions – Section 5.3.

5.2 Quantitative Data

Table 17 presents the quantitative data collected for the task of identifying the clusters; Table 18, presents the quantitative data for the task of identifying the outliers.

User	Time to complete (s)		# Rotations		# Scales		# Translations	
	Mouse	Kinect	Mouse	Kinect	Mouse	Kinect	Mouse	Kinect
1	<i>96</i>	197	<i>8</i>	14	<i>3</i>	7	2	0
2	<i>124</i>	143	<i>10</i>	18	<i>3</i>	6	1	1
3	<i>158</i>	162	<i>11</i>	9	<i>3</i>	4	0	0
4	<i>148</i>	199	<i>9</i>	15	<i>4</i>	9	1	0
5	<i>92</i>	252	<i>5</i>	12	<i>2</i>	6	0	0
6	<i>88</i>	121	<i>4</i>	6	<i>2</i>	3	0	0
7	<i>74</i>	214	<i>4</i>	12	<i>3</i>	5	0	1
8	<i>82</i>	189	<i>4</i>	8	<i>3</i>	5	0	0
9	<i>120</i>	143	<i>7</i>	10	<i>5</i>	6	0	0
10	<i>74</i>	178	<i>4</i>	9	<i>3</i>	8	0	0
11	<i>111</i>	186	<i>4</i>	7	<i>3</i>	7	0	0
12	<i>74</i>	172	<i>3</i>	5	<i>2</i>	7	0	0
13	<i>87</i>	163	<i>4</i>	8	<i>3</i>	5	0	0
14	<i>112</i>	174	<i>4</i>	7	<i>4</i>	9	0	0
15	<i>95</i>	192	<i>5</i>	7	<i>5</i>	8	0	0
16	<i>83</i>	179	<i>4</i>	8	<i>4</i>	7	0	0
17	<i>127</i>	144	<i>5</i>	7	<i>5</i>	7	0	0
18	<i>131</i>	143	<i>4</i>	6	<i>3</i>	7	0	0
19	<i>147</i>	127	<i>6</i>	8	<i>4</i>	6	1	0
20	<i>83</i>	196	<i>3</i>	8	<i>4</i>	8	0	0

TABLE 17: Quantitative data for task “identify clusters”; Cars dataset in italic, Iris dataset in regular type.

User	Time to complete (s)		# Rotations		# Scales		# Translations	
	Mouse	Kinect	Mouse	Kinect	Mouse	Kinect	Mouse	Kinect
1	<i>87</i>	275	<i>12</i>	16	<i>6</i>	13	0	0
2	<i>156</i>	294	<i>15</i>	20	<i>2</i>	10	2	0
3	<i>253</i>	221	<i>10</i>	12	<i>8</i>	5	3	1
4	<i>132</i>	287	<i>9</i>	12	<i>4</i>	7	2	0
5	<i>145</i>	276	<i>7</i>	10	<i>4</i>	8	0	0
6	<i>147</i>	253	<i>6</i>	10	<i>4</i>	7	0	0
7	<i>96</i>	277	<i>5</i>	13	<i>4</i>	6	0	0
8	<i>173</i>	245	<i>9</i>	13	<i>4</i>	7	1	1
9	<i>159</i>	232	<i>8</i>	11	<i>4</i>	7	0	1
10	<i>146</i>	314	<i>7</i>	15	<i>4</i>	9	0	0
11	<i>154</i>	159	<i>6</i>	5	<i>3</i>	2	1	0
12	<i>134</i>	289	<i>5</i>	9	<i>4</i>	6	0	1
13	<i>148</i>	185	<i>6</i>	9	<i>3</i>	8	1	0
14	<i>159</i>	161	<i>3</i>	5	<i>3</i>	2	0	0
15	<i>144</i>	217	<i>7</i>	9	<i>3</i>	10	0	1
16	<i>148</i>	276	<i>9</i>	13	<i>4</i>	9	2	2
17	<i>174</i>	162	<i>6</i>	8	<i>6</i>	9	1	0
18	<i>92</i>	185	<i>3</i>	9	<i>2</i>	8	0	0
19	<i>188</i>	211	<i>8</i>	9	<i>5</i>	7	0	0
20	<i>130</i>	114	<i>5</i>	4	<i>3</i>	4	1	0

TABLE 18: Quantitative data for task “identify outliers”; Cars dataset in italic, Iris dataset in regular type.

Analysis of the quantitative data for the datasets

Here, we present the statistical raisings used to evaluate the influence of the datasets. In this evaluation, and in the ones that follow, we use the following steps: simple summing for overall

analysis; verification of normality so that we could decide whether t-test could be used; checking of variances equality to decide which t-test to use; one-tail t-test to verify significance of influence; and two-tail t-test to verify oriented influence.

In Table 19, for task “identify clusters”, Cars dataset vs Iris dataset, the one-tail t-test and the two-tail t-test indicate that there is no significance influence in the user experience due to the different datasets. The summations (row Total (sum)) indicate some variation from one dataset to the other, but they are not significant to contradict the t-test over the time measure.

Dataset	Time to complete (s)		# Rotations		# Scales		# Translations	
	Cars	Iris	Cars	Iris	Cars	Iris	Cars	Iris
Total (sum)	2,732	2,848	137	155	102	96	4	3
Kolmogorov-Smirnov Normality test (normal if > p-value 0.05)	0.41	0.72	0.09	0.03	0.01	0.12	< 0.0001	< 0.0001
Mean ± stdv	136.6 ± 41	142.4 ± 49	non-normal		non-normal		non-normal	
F-Test (variances are equal if p-value > 0.05)	0.2 -> equal variances							
One-tail t-test equal variances (significant if p-value < 0.05)	0.34 -> not significant							
Two-tail t-test equal variances (significant if p-value < 0.05)	0.69 -> not significant							

TABLE 19: Quantitative statistical analysis for task “identify clusters”, Cars dataset vs Iris dataset.

In Table 20, for task “identify outliers”, Cars dataset vs Iris dataset, the one-tail t-test shows that the datasets slightly influence the experimentation by a very small factor (area of confidence) for p-value 0.05; the two-tail t-test, on the other hand, states that there is no influence. Besides that, the summations do not indicate any strong evidence of influence. Actually, if we look close, the Iris dataset is indeed a little trickier for outlier detection.

Dataset	Time to complete (s)		# Rotations		# Scales		# Translations	
	Cars	Iris	Cars	Iris	Cars	Iris	Cars	Iris
Total (sum)	3,453	4,145	168	190	109	115	12	9
Kolmogorov-Smirnov Normality test (normal if > p-value 0.05)	0.09	0.17	0.01	0.51	0.00	0.44	0.00	0.00
Mean ± stdv	0.17 (equal)		non-normal		non-normal		non-normal	
F-Test (variances are equal if p-value > 0.05)	0.04 (significant)							
One-tail t-test equal variances (significant if p-value < 0.05)	0.08 (not significant)							
Two-tail t-test equal variances (significant if p-value < 0.05)	3,453							

TABLE 20: Quantitative statistical analysis for task “identify outliers”, Cars dataset vs Iris dataset.

Overall, considering tables 19 and 20, we can verify that the extra complexity of the Iris dataset is of nearly 11% for time; small enough to assume that the datasets were fairly similar in the experiments – we did not expect to have identical datasets, as it is not feasible.

Analysis of the quantitative data for the devices

In Table 21, for task “identify clusters”, mouse vs Kinect, the one-tail t-test indicates that, yes, there are significant differences in using the different devices. This was already expected as the

mouse and the Kinect provide very different interaction mechanisms. In the same way, the two-tail t-test indicates that using the Kinect demands more time than using the mouse for the same task. This fact is clearly visible in the summations, except for the number of translations, small for both devices.

Device	Time to complete (s)		# Rotations		# Scales		# Translations	
	Mouse	Kinect	Mouse	Kinect	Mouse	Kinect	Mouse	Kinect
Total (sum)	2,106 s	3,474 s	108	184	68	130	5	2
Kolmogorov-Smirnov Normality test (normal if > p-value 0.05)	0,06 -> Normal	0,56 -> Normal	0,0003	0,0030	0,0007	0,1100	< 0.0001	< 0.0001
Mean ± stdv	105,3±26,7	173,7±31,7	non-normal		non-normal		non-normal	
F-Test (variances are equal if p-value > 0.05)	0,23 -> equal variances							
One-tail t-test equal variances (significant if p-value < 0.05)	0,0000000039 -> significant							
Two-tail t-test equal variances (significant if p-value < 0.05)	0,0000000078 -> significant							

TABLE 21: Quantitative statistical analysis for task “identify clusters”, mouse vs Kinect.

In Table 22, for task “identify outliers”, mouse vs Kinect, the one-tail t-test again indicates that there are significant differences in using the different devices. Also again, the two-tail t-test indicates that the Kinect demands more time for completing the task. The summations for this task – “identify outliers”, are even more revealing than for task “identify clusters” (Table 21).

Device	Time to complete(s)		# Rotations		# Scales		# Translations	
	Mouse	Kinect	Mouse	Kinect	Mouse	Kinect	Mouse	Kinect
Total (sum)	2,965	4,633	146	212	80	144	14	7
Kolmogorov-Smirnov Normality test (normal if p-value > 0.05)	0.08 -> Normal	0.09 -> Normal	0.37 -> Normal	0.39 -> Normal	0.14 -> Normal	0.14 -> Normal	< 0.00001	< 0.00001
Mean ± stdv	148.2 ± 36	231.6 ± 56	7.3 ± 2.9	10.6 ± 3.9	4 ± 1.4	7.2 ± 2.6	non-normal	
F-Test (variances are equal if p-value > 0.05)	0.03 -> unequal		0.1 -> equal		0.004 -> unequal			
One-tail t-test equal/unequal variances (significant if p-value < 0.05)	1.7 E-06 -> significant		0.002 -> significant		0.00002 -> significant			
Two-tail t-test equal variances (significant if p-value < 0.05)	3.4 E-06 -> significant		0.004 -> significant		0.00004 -> significant			

TABLE 22: Quantitative statistical analysis for task “identify outliers”.

If we consider the data of both tables, we can verify that the difference between the mouse and the Kinect was of nearly 38% in seconds, 36% in number of rotations, and 46% in number of scales. The number of translations was too small for both devices. The quantitative experimentation, hence, demonstrated that the Kinect device is less efficient in terms of time and number of operations in the context of information visualization; more specifically, in interacting with a multidimensional projection for the tasks of cluster and outliers identification.

5.3 Qualitative Data

Table 23 presents the qualitative data collected after the user experience. The 7 questions of the questionnaire were meant to measure how the users perceived the usability of each device considering general adequacy, acceptance, easiness, physical and psychological adequacy, besides an explicit declaration of preference.

- A) How did you appreciate using the specified device for this experience?
- B) How appropriate did you consider the specified device for this experience?
- C) How easy did you consider the commands (rotation, scaling, and translation) necessary to use the specified device?
- D) How physically comfortable was to perform this experience with the specified device?
- E) How psychologically comfortable was to perform this experience with the specified device?
- F) How physically facile (non-exhausting) was to perform this experiment with the specified device?
- G) How inclined are you to substitute the mouse with Kinect?

User	A)		B)		C)		D)		E)		F)		G)
	Mouse	Kinect											
1	5	5	5	5	4	3	4	3	4	4	4	4	3
2	2	5	4	5	5	4	4	4	4	4	4	3	4
3	4	5	5	5	3	4	4	3	4	4	4	4	5
4	4	5	4	5	5	5	4	2	4	4	4	3	5
5	3	5	4	4	4	4	4	3	4	4	4	3	5
6	4	4	4	4	5	4	4	3	4	4	4	3	4
7	3	4	5	5	5	5	4	3	4	4	4	4	3
8	3	5	3	5	5	4	4	4	4	4	4	3	5
9	3	4	5	4	4	3	4	2	4	4	4	4	4
10	4	5	5	5	5	4	4	2	4	4	4	4	3
11	5	4	4	5	5	5	4	3	4	4	4	4	4
12	4	4	4	4	5	5	4	4	4	4	4	3	4
13	4	4	5	5	4	5	4	4	4	4	4	3	5
14	5	5	4	4	4	4	4	4	4	4	4	3	5
15	4	5	4	4	3	4	4	3	4	4	4	2	3
16	5	5	5	3	3	4	4	4	4	4	4	4	5
17	5	5	5	5	4	5	4	4	4	4	4	2	5
18	5	4	4	5	3	5	4	4	4	4	4	3	5
19	5	5	5	4	5	5	4	4	4	4	4	3	4
20	4	5	5	5	5	4	4	3	4	4	4	3	4
Accumulated score	81	93	89	91	86	86	80	66	80	80	80	65	85
	%	%	%	%	%	%	%	%	%	%	%	%	%

TABLE 23: Quantitative data of the experiment – 7 questions regarding perception of usability.

The qualitative data demonstrated that the users aproved using the Kinect device, more than the mouse, as seen in questions A, B, and G. According to questions C and E, the devices were

equally easy to use and psychologically comfortable. Questions D and F, in turn, showed that, physically, the mouse is more adequate than the Kinect.

6 CONCLUSIONS

We reviewed concepts related to natural interaction, to the Kinect device, and to techniques Participatory Design and Usability Test according to ISO 13407. The union of these concepts was used to draw experiments so that we would be able to evaluate the following methodological hypothesis:

“The Kinect device is tailored for defining more efficient interaction if compared to the commodity mouse device in the specific context (multidimensional projection) of an information visualization system.”

Our results demonstrated that quantitatively – time, and number of operations; and qualitatively – physical effort and easiness, the Kinect device was not as efficient as the mouse. We can say that Kinect demands extra effort and attention from the users who, standing with their entire body, are dueled to make the right moves at the right moments in order to achieve their intents; in contrast to the mouse interaction, with users sitting and moving only one hand.

Despite of that, the qualitative subjective measures pointed out a higher user satisfaction in respect to the convenience, and to the adequacy of Kinect; at last, the users declared a reasonable higher desire to substitute the mouse with Kinect. This fact reflects the interactive appeal, and the game-oriented design of Kinect, which intentionally leads to interesting feelings of immersion inside the computational system.

From these remarks we conclude that the methodological hypothesis was not full-filled, since it was stated in terms of efficiency, and not in terms of satisfaction. From our observations, we infer that, for production settings – like those observed in desktop visualization systems, the extra effort that we quantitatively measured for Kinect may be cumulative for long-term activities. In such circumstances, the lack-of-efficiency aspect of Kinect tends to overcome the short-term satisfaction that this device promotes. Accordingly, we have enough arguments to refute the proposed hypothesis.

Acknowledgements

This work received support from the following funding agencies: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq 560104/2010-3), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2011/13724-1) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes).

7 REFERENCES

- [1] Valli, A. “The Design of Natural Interaction.” *Multimedia Tools Applied*, vol. 38 issue 3, pp. 295-305, 2008.
- [2] Chang, Y. J., Chen, S. F., and Huang, J. D. “A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities,” *Research In Developmental Disabilities*, vol. 32 issue 6, pp. 112-128, 2011.
- [3] Gallo, L., Placitell, A. P., and Ciampi, M. “Controller-free exploration of medical image data: experiencing the Kinect,” in *Proceedings of the International Symposium on Computer-Based Medical Systems*, pp. 1-6, 2011.
- [4] Nielsen, J. *Usability Engineering*. Cambridge: Academic Press, 1993.
- [5] Salvendy, G. “Handbook of human factors and ergonomics,” New York: Wiley, 2012.

- [6] Calle, J., Martínez, P., Del Valle, David, and Cuadra, D. "Towards the Achievement of Natural Interaction," in *Engineering the User Interface, 1st ed.* Redondo, M., Ed. London: Springer, pp. 1-19, 2009.
- [7] Del Bimbo, A. "Special issue on natural interaction," *Multimedia Tools and Applications*, vol. 38 issue 3, pp. 293-294, 2008.
- [8] Nintendo "Wii - Official Website at Nintendo," Internet: <http://www.nintendo.com/wii>, 2012 [May 9, 2012].
- [9] Sony "PlayStation®Move - PlayStation®3" Internet: <http://br.playstation.com/ps3/playstation-move>, 2012 [May 10, 2012].
- [10] Microsoft "Kinect - Xbox.com." Internet: <http://www.xbox.com/pt-BR/Kinect>, 2012 [Apr. 26, 2012].
- [11] Suma, E., Lange, B., Rizzo, A., Krum, D., and Bolas, M. "FAAST: The Flexible Action and Articulated Skeleton Toolkit," in *Proceedings of the IEEE Virtual Reality*, pp. 247-248, 2011.
- [12] Löwgren, J., and Stolterman, E. (2004) *Thoughtful Interaction Design: A Design Perspective On Information Technology*. Chicago: Mit Press, 2004.
- [13] Kaptelinin, V., & Bannon, L. J. "Interaction Design Beyond the Product: Creating Technology-Enhanced Activity Spaces," *Human-Computer Interaction*, vol. 27 issue 3, pp. 277-309, 2012.
- [14] Asaro, P. M. "Transforming society by transforming technology: the science and politics of participatory design." *Accounting, Management and Information Technologies* vol. 10 issue 3, pp. 257-290, 2000.
- [15] Berg, A., Gulden, T. "Participatory design for well-being," in *Design education for future wellbeing: proceedings of the 14th International Conference on Engineering and Product Design Education*, pp. 531-536, 2012.
- [16] Nielsen, J. "Cost of User Testing a Website." Internet: <http://www.useit.com/alertbox/980503.html>, 1998 [May 20, 2012].
- [17] Virzi, R.A., "Refining the Test Phase of Usability Evaluation: How Many Subjects is Enough?" *Human Factors* vol. 34 issue 4, pp. 457-468, 1992.
- [18] Brock, D., Kim, S., Palmer, O., Gallagher, T., Holmboe, E. "Usability Testing for the Rest of Us: The Application of Discount Usability Principles in the Development of an Online Communications Assessment Application," *Teaching and learning in medicine*, vol 25 issue 1, pp. 89-96, 2013.
- [19] Jose Rodrigues, Luciana A. S. Romani, Agma J. M. Traina, Caetano Traina Jr., *Combining Visual Analytics and Content Based Data Retrieval Technology for Efficient Data Analysis*. International Conference on Information Visualisation, pp. 61-67 IEEE Press.
- [20] Faloutsos, C., and Lin, K. I. *FastMap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets*. Proceedings of the SIGMOD international conference on Management of data, ACM Press, pp. 163-174, 1995.
- [21] M.R. Andersen, T. Jensen, P. Lisouski, A.K. Mortensen, M.K. Hansen, T. Gregersen and P. Ahrendt, *Kinect Depth Sensor Evaluation for Computer Vision Applications*, Technical Report ECE-TR-6, 39 pages, AARHUS University, Dept. of Engineering, 2012.

- [22] Kathryn LaBelle, Evaluation of Kinect Joint Tracking for Clinical and In-Home Stroke Rehabilitation Tools, Thesis, 66 pages, Notre Dame University, Indiana, 2011.
- [23] Kourosh Khoshelham, Sander Oude Elberink, Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications, Journal Sensors, volume 12, p. 1437-1454, 2012.
- [24] Martin Kefer, Wilfried Kubinger, Evaluation of Kinect depth sensor for use in mobile robotics, Proceedings of the International DAAAM Symposium, volume 22, issue 1, p. 147-148, 2011.