3D Interaction environment for free view point TV and games using multiple tablet computers

Rob Dupre, Raul Herrera Acuna, Vasileios Argyriou and Sergio A. Velastin

R.Dupre, R.Herrera, Vasileios.Argyriou, S.Velastin@kingston.ac.uk

Abstract

The use of 3D technologies to represent elements and interact with them is an open and interesting research area. In this article we discuss a novel human computer interaction method that integrates mobile computing and 3D visualization techniques with applications on free viewpoint visualization and 3D rendering for interactive and realistic environments. Especially this approach is focused on augmented reality and home entertainment and it was developed and tested on mobiles and particularly on tablet computers. Finally, an evaluation mechanism on the accuracy of this interaction system is presented.

1. Introduction

3D visualization and representation is used in several areas, especially for entertainment and design, where this technology demonstrated the possibility to improve the understanding and perception of real life concepts and elements.

One of the modern real time applications of 3D representation in Human Computer Interaction is related to the visualization of real time free viewpoint television (FTV). This is a visual media system that enables the user to view a three dimensional scene with the possibility to freely change the point of view without losing the connection by interpolating between the different points of view [1]. Also it allows displaying almost infinite aspects of a scene, providing more information and details about specific environments, actions and interactions. This emergent technology provides the possibility of a revolution on real time broadcasting by improving significantly the interaction with the end users and mainly their engagement, especially in sport related activities, where human actors play a fundamental role [2, 3].

The development of new techniques on 3D rendering and multi view modeling in video games plays a critical role in the improvement of 3D technologies, especially in games based on first person interaction or multiple camera angle displacement. Incorporating videogame engines in the construction and use of realistic environments provides advanced mechanisms for testing and evaluation in several research areas [4]. Also, the influence of video game technologies in the society and the fact that media shape the perception of how the interaction and visualization of environments should be, lead the research on the interaction environments by generating a cross evolution between different knowledge areas [5]. Since augmented reality can fuse 3D virtual objects in real 3D environments in real time, several studies and approaches were introduced to make this technique more effective and efficient. The main application areas are in environments where additional information could be crucial in taking decisions, such as medical visualization, maintenance and repair, military aircraft navigation and targeting [6]. Also, this progress has been integrated in the area of entertainment including the development of virtual environments and focusing on touring and guiding tools for urban locations [7].

The latest innovations in video games have several applications in the film industry, especially in the evolution of the integrated virtual scenarios, augmented reality sequences, character animation and real and artificial graphical elements. These technologies allow more realistic experiences; lower production costs and the possibility to create and integrate novel ideas in the films that would be impossible with the traditional animation or special effect techniques [8]. There are several aspects that have to be considered in the cinematographic integration of virtual and real elements, such as 3D artifact reconstruction, slam, point of view tracking, motion estimation, viewpoint computation and camera planning. Solutions to these research topics have been proposed using filtering, epipolar geometry, segmentation, space correlation and complexity cell reduction [9].

In this paper, we present a novel approach to interact with 3D environments and elements, based on the use of 3D techniques and mobile devices, such as tablet PCs. The proposed methodology allows the user to interact and visualize elements in 3D based on the concept of free viewpoint television. In the following sections we present a brief discussion on previous related works, our proposed methodology and an analysis of the implementation. Finally, we will conclude by presenting its advantages, possible applications in different environments and issues related to our novel human computer interaction approach.

2. Previous work

2.1. Free Viewpoint TV

The interaction mechanisms with 3D scenes have been always a challenging topic for image and video researchers. The difficulty related to these tasks is due to the several factors involved in a full 3D real time multi view point processing, such as representation, capturing, rendering and coding. In the work of Tanimoto [10], an approach capable to deal with all those problems was presented, generating maybe the first efficient 3D free viewpoint TV system. This method is able to generate a complete 3D reconstruction, providing the change to freely view a real scene. This approach also allows image capturing, rendering, texturing and visualization based on a "ray space" method. The "Ray Space" consists of a 3D space representation in four dimensions generated from the reduction of the initial five dimensional ray space by applying dimensionality reduction. The images captured are aligning in parallel and a sine-shaped plane is used to generate concentric mosaics. The "ray-space representation" includes a depth estimation and interpolation unit, which allows us to slice the whole space in 3D scenes.

The video acquisition is performed using 16 cameras, which corresponds to 16 clients of a server. An upgrade of this technology allows us to obtain the 3D scene captured by 100 cameras, generating a synchronized signal and capturing high resolution videos at 30 fps.

The correction process is performed in real time, applying a geometric alignment under the ray-space, designed for a multi camera setup, using a chess board pattern for calibration and extraction of intrinsic and extrinsic parameters. Then principal component analysis (PCA) is performed to acquire the position of the cameras and, finally, the homographies are determined.

The rendering process applies global optimization for depth estimation using left and right real images as references to generate the virtual view. To prevent errors in depth estimation, the synthetized views contain specific artifacts to guarantee the reliability of the system and in case of inconsistency; the unreliable pixels are subsequently eliminated.

There are two approaches regarding the user interface for FTV. In the first case, the user viewpoint is based on head tracking to generate the corresponding view image. In the second approach the user can freely change the views. Augmented reality and applications

During last decade, augmented reality and its applications started to attract many researchers and was characterized by some fundamental properties [11]:

- Supplements the real world with virtual objects (computer-generated)
- Runs interactively and in real time
- Align real and virtual objects with each other.

This technology lays in the "mixed reality" techniques and is regarded as the preprocessing step of augmented virtuality systems.

Augmented reality potentially can be applied to all senses. There are several advances of tactile augmented reality, especially when it is applied for the treatment of phobias [12]. Also in case of "audible augmented reality", a more realistic experience can be generated by linking visual augmented reality and audio [13].

This technology started in 1960s, but its real growth was during 1990s, when several mixed reality systems were presented. There are several related problems, including resolution, field of view, contrast and lack of brightness. Some of these problems could be overcome by the use for example optical see-through displays. Also, there are other aspects related to the user viewing orientation and position, where even the problem for known environments is almost solved, the case of unknown environments is still a challenging area, especially for outdoors applications. Finally, the calibration problems that were a main issue in the beginning of this technology are reduced using several techniques, such as perspective projection models, redundant sensors and auto compensation of calibration parameters.

2.2. Mutli-touch table tops

Multi touch systems provide the users the ability to control and interact naturally with elements, using dynamic and interactive surfaces capable of identifying multiple contact entry points [14]. Nowadays, multi-touch devices are accessible to everyone, becoming part of several areas of the social interaction and devices such as mobile phones, video game consoles, laptops, etc. [15].

In 1982, N. Mehta, developed the "Flexible Machine Interface" which is considered the first multi touch system [16]. Later, Bell Labs involved in this field, publishing and developing novel interfaces. In 1985 B. Buxton in collaboration with the "Input Research Group" of the University of Toronto developed the "Multi-Touch Tablet" [17]. In the following years, many developments appeared in the area, such as HoloWall [18], DiamondTouch [19], ReacTable [20], TouchLight [21] and Microsoft Surface. Nowadays, there are several development areas for these technologies, such as the integration with other media and devices of the same type (multi-touch) or totally different ones (traditional interfaces), reaching a higher potential of use, functionality and efficiency [22]. A clear example is the development of mobile devices, using integrated features of multi-touch table tops, creating heterogeneous systems, capable of sharing information but keeping specific data only for authorized users [23]. Another related field is the development of software and table top systems able to manage not just the multi-touch interface, but also other ways of interaction, such as natural language [24]. There are several studies on designing systems that can support 3D interaction, resulting to an increase of their capabilities including natural gestures and giving to the users more confidence and comfort when they interact with the devices [25]. Additionally, the interaction devices can improve their performance by combining traditional information representation [26]. A key aspect of the evolution of these devices is related to the integration of collaborative group work. This type of devices will be able to improve interaction between professionals and common people, providing new applications in education [27], in house devices [28] and in architecture design [29].

The conventional interaction devices have been improved during the last decade [30] providing the capability of being “self-contained”, (i.e. they do not have the need of external interaction devices to operate).

Probably the main advantage of the multi-touch table top systems is the nature way of the interaction with them. As a result it is more intuitive, and more similar to the way humans interact with their environment [31]. Studies about the usability of these systems show that they increase the productivity of their users, making the traditional interaction techniques to be obsolete [15]. Also, the speed of learning new configurations increased allowing users to adapt quickly in new environments that could appear incompatible with the usual activities of the user [32].

However, one of the main limitations of multi-touch table tops devices is related to their portability. This characteristic of multi-touch table top devices makes them expensive and difficult to use as a simple costumer, limiting its access to organizations that need them for specific tasks.

3. Proposed methodology

In this section we present a novel architecture to overcome the size disadvantage, using software and tablet PCs to create multi touch table tops of variable size and shape focusing on 3D interaction. The general idea is to create multi touch table tops using two or more tablet PCs (or tablet devices) that can “connect” creating a virtual surface and interacting on a delimited workspace area, associated by their relative position. These devices have multi-touch capabilities, “3D position-awareness” and a networking system that can be used to achieve this task.

This idea was based on the appropriate use of data provided by the accelerometers and the gyroscopes of the tablets, which are able to provide the orientation and the displacement of the device.

When the user starts the system, the tablet has to be placed initially on the “centre” of the virtual layer. The centre can be at any place and will correspond to the initial position of the tablet. Once we initialise the system, the centre of the whole “table” will be regarded the same starting position and it will be used as a reference to the related tablet PCs. After that process, the tablet can be “moved” around the virtual layer, showing different parts of the background, independently of its orientation. This is due to its “self-awareness” of the position and movement provided by the accelerometer. A star network topology is used and the first tablet operates as a server for the remaining tablets that joined the table (the tablets can communicate via Bluetooth or by using a wireless network). Since for example the second tablet has the information of the centre location both tablets will be linked as two parts of the whole array creating an “ad-hoc network”, in order to obtain finally a fully functional multi touch table top. This process can be seen in figure 1 (the overall representation and the position of the tablet PCs can be altered).

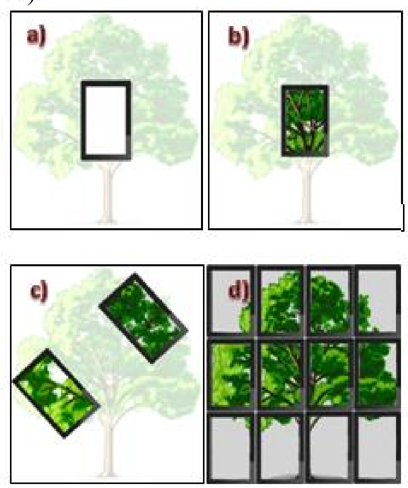

Figure 1: Proposed approach a) Start, b) Initial calibration, c) Multiple tablets, d) Full "table top" layer

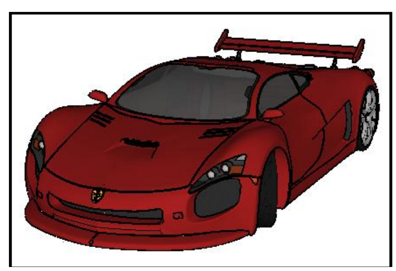

Figure 2: 3D model for calibration

This approach is also extended to support not only 2D interactions but to incorporate also 3D environments, where our virtual layer is now a 3D scene containing models as the one shown in figure 2.

The system starts, again with a random initial view of the scene, which can be the front, side, top-down etc. The initial location and viewing direction will be regarded as the centre of the coordinate system and all the consequent will be estimated relatively to that one. The initial view of the 3D model in our scene will be similar to the 3D image in figure 3 and the user needs to point the tablet on the top of the surface pointing downwards.

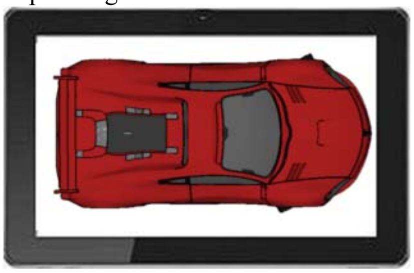

Figure 3: Start position for the system (top-down view)

Once the system is initialised, the user can move freely around the model, getting multiple viewpoints of the model and the 3D scene. This is possible because of the tri dimensional self-awareness of the tablets (based on the measurements from the accelerometers and gyroscopes, which detect every change of the relative position of the tablet from the starting point). The user could choose different views by moving around the initial location and obtain results such as the one in figure 4.

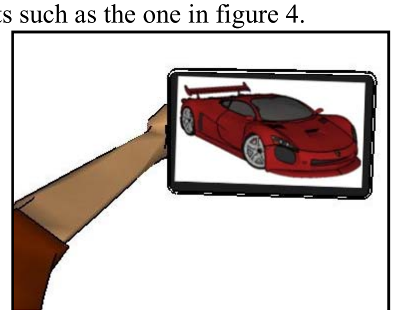

Figure 4: User view of the 3D model in the tablet PC

After the set up of the first tablet and obtaining the relative position of the model, other users can join the "network" in a similar way, transferring information among the tablets regarding their relative positions and directions (see figure 5).

As it can be observed, this 3D approach is more flexible and intuitive than the 2D one, but requires more processing power due to the 3D rendering of the scene, which is possible with the new generation of tablet PCs.

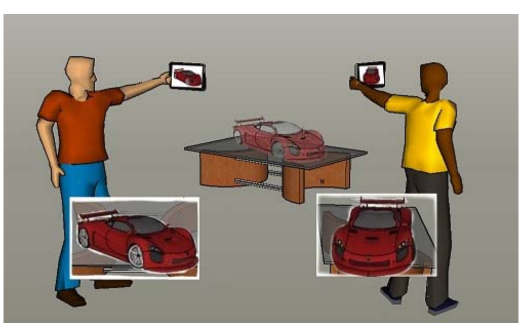

Figure 5: Multiple users with different point of views of the 3D model

4. Implementation and Results

Many features and intelligence have been added to the modern mobile and tablet devices. Free view point TV, tracking and gaming are just a few of the beneficial markets from obtaining positioning information. One option for obtaining this information is through the use of inertial sensors. Since there is no direct conversion between acceleration and position, the signal obtained from these sensors requires further processing. Therefore, a double integral must be applied to the input signal to obtain the current position. This allows velocity information to be obtained using the previous location, while in case of Verlet Integrator the state of the two prior time steps is required. The approach utilized in this work can be applied to any sensing axis, but when positioning is implemented in 3 axes, extra processing is required to consider the earth's gravity effect.

In more details, we are taking the readings from a digital accelerometer, a gyroscope and the compass integrated in a mobile or tablet device, and then combine them to form a rudimentary dead reckoning system. Dead reckoning is the process of estimating the current position by using a pre-determined start point, and then updating the device's position estimate through knowledge of its speed over time, and the direction that is facing. Each new estimate is calculated from old estimates; therefore error accumulation problems may occur.

In most tablet devices a triple axis accelerometer with a digital interface is available and we can sample the acceleration by sending a command over a serial digital protocol which returns a number relating to the acceleration in terms of g-force. This is opposed to an analogue accelerometer which is sampled by reading the voltage of its pins that is proportional to the acceleration.

Initially, the accelerometer was placed in standby mode so it can accept changes to its configuration and it was set to sample at the highest resolution and the fastest data rate. The device now can continuously provide the updated acceleration values at the specified data rate.

Since accelerometers may have offset errors, calibration is required to remove this bias. Therefore, a certain number of calibration readings are taken, before we start our displacement calculations, and averaging them. By subtracting this bias from all the subsequent readings we will have ensured all the readings are properly calibrated to the 0g value.

An update method is considered to sample the accelerometer and perform the calculations for position estimation. Accelerometers are quite noisy, and running averaging is applied to reduce the error introduced by the noise without introducing delays on the frame rate.

Additionally, two thresholds are selected experimentally to indicate the range of the invalid values of the accelerometer and if they fall within these limits we assume them as noise and the acceleration is set to zero.

In order to obtain the position we regard the velocity as the derivative of the position and the acceleration is the derivative of the velocity. Thus integrating the values we have:

$$v = \int(\vec{a})dt \text{ and } s = \int(\vec{v})dt = \int(\int(\vec{a})dt)dt \quad (1)$$

Where $a = \frac{d\vec{v}}{dt} = \frac{d(\overrightarrow{ds})}{dt^2}$ and $v = \frac{d\vec{s}}{dt}$.

We can simplify this approach by introducing a numerical solution and in this case we obtain:

$$v(n) = v(n-1) + a(n)dt \quad (2)$$

$$s(n) = s(n-1) + v(n)dt + \frac{1}{2}a(n)dt^2 \quad (3)$$

By applying this to the acceleration readings, and then again to the velocity calculations, we can obtain an estimate of position.

Another issue that needs to be addressed is the estimation of the end of the movement. If we move our accelerometer from a stationary point for a certain distance and then bring it to rest again, we should get an equal and opposite amount of acceleration in both directions. If we don't read an equal amount, our velocity will remain constant at a certain number. A heuristic approach is utilized and the number of zero readings of acceleration in a row is calculated. If that number is above a predefined threshold the velocity is reduced gradually to zero.

Regarding the direction of the movement in 2D space the readings from the digital compass are used and a number in the range of 0 to 359.9 degrees is obtained.

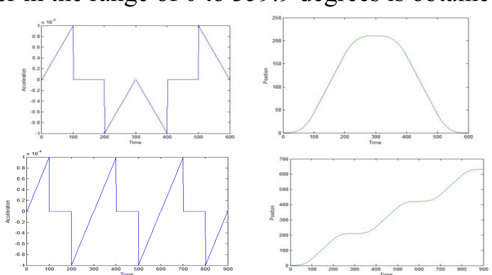

Figure 6: Artificial examples with the acceleration on the left column and the estimated position versus time on the right column.

From the experiments performed we can see in figure 6 two cases of artificial acceleration and the estimated position. In the first case the object returns to its initial location since the same but inverted acceleration is applied on the tablet. Regarding the experiments with real data few examples are shown in figure 7 and it can be observed that in some cases the estimation of the end of the movement is essential due to the significant amount of noise. Finally, some images of the developed system on a mobile device are shown in figure 8, indicating the features of the implemented positioning mechanism.

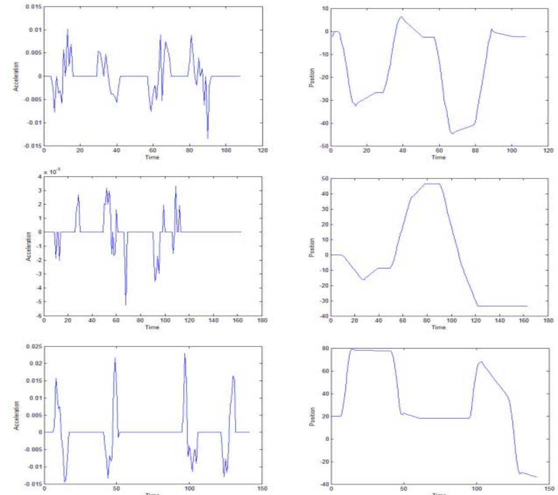

Figure 7: Real examples with the acceleration on the left column and the estimated position versus time on the right column.

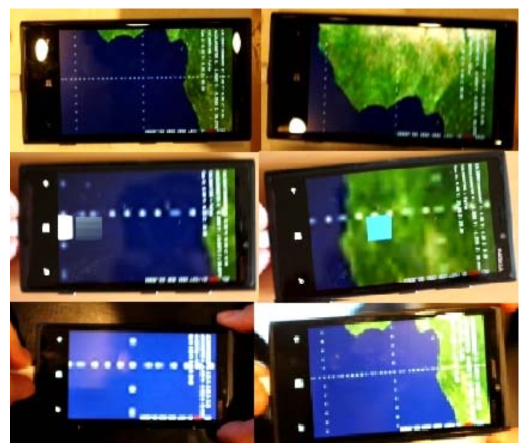

Figure 8: Positioning mechanisms implemented on a mobile device on each axis (first row – up and down, second row – left and right; and third row zoom in and out).

5. Conclusions

In this paper a novel human computer interaction mechanism with applications in augmented reality, gaming, tabletops and mainly on free view point TV was presented. It provides the ability to interact both in two and three dimensions allowing the users to work and interact with the same environment in groups. Further applications of this approach could be on personal and car navigation, back-up GPS, anti-theft devices, map tracking, 3-D gaming, PC mouse, plus many others. Also, experiments were performed to evaluate the accuracy of the positioning algorithm showing the potential of this system by improving further the human computer interaction mechanisms.

6. References

[1] M. Tanimoto. Overview of free viewpoint television. Signal Process Image Communication, 21(6):454-461, 2006.

[2] J. Carranza, C. Theobalt, M.A. Magnor and H. Seidel. Free-viewpoint video of human actors. ACM Transactions on Graphics (TOG), 22(3):569-577, 2003.
[3] L. Yang, T. Yendo, M.P. Tehrani, T. Fujii and M. Tanimoto. Artifact reduction using reliability reasoning for image generation of FTV. Journal of Visual Communication and Image Representation, 21(5): 542-560, 2010.
[4] R. Andreoli, R. De Chiara, U. Erra and V. Scarano. Interactive 3d environments by using videogame engines. Proceedings of the Ninth International Conference on Information Visualisation :515-520, 2005.
[5] S. Poole. Trigger happy: Videogames and the entertainment revolution. London:4th State, 2004.
[6] R.T. Azuma. A survey of augmented reality. Presence-Teleoperators and Virtual Environments, 6(4):355-385, 1997.
[7] S. Feiner, B. MacIntyre, T. Höllerer and A. Webster. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. Personal and Ubiquitous Computing, 1(4):208-217, 1997.
[8] K. Kennedy and R.E. Mercer. Planning animation cinematography and shot structure to communicate theme and mood. Proceedings of the 2nd international symposium on Smart graphics:1-8, 2002.
[9] C. Lino, M. Christie, F. Lamarche, G. Schofield and P. Olivier. A real-time cinematography system for interactive 3D environments. Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation: 139-148, 2010.
[10] M. Tanimoto, M.P. Tehrani, T. Fujii and T. Yendo. Free-viewpoint TV. Signal Processing Magazine IEEE, 28(1):67-76, 2011.
[11] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier and B. MacIntyre. Recent advances in augmented reality. Computer Graphics and Applications IEEE, 21(6):34-47, 2001.
[12] A.S. Carlin, H.G. Hoffman and S. Weghorst. Virtual reality and tactile augmentation in the treatment of spider phobia: a case report. Behaviour research and therapy, 35(2):153-158, 1997.
[13] J.M. Rozier. Hear&There: An augmented reality system of linked audio. MIT, Master Thesis, 2000.
[14] K. Kin, M. Agrawala and T. DeRose. Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation.Proceedings of Graphics interface 2009: 119-124, 2009.
[15] S. Malik, A. Ranjan and R. Balakrishnan. Interacting with large displays from a distance with vision-tracked multi-finger gestural input. Proceedings of the 18th annual ACM symposium on User interface software and technology:43-52, 2005.
[16] D. Saffer. Designing Gestural Interfaces: Touchscreens and Interactive Devices. O'Reilly, 2008.
[17] A. Mazalek. Media Tables: An extensible method for developing multi-user media. MIT Thesis (Ph. D.), 2005.
[18] N. Matsushita and J. Rekimoto. HoloWall: designing a finger, hand, body, and object sensitive wall. Proceedings of the 10th annual ACM symposium on User interface software and technology:210, 1997.
[19] P. Dietz and D. Leigh. DiamondTouch: a multi-user touch technology. Proceedings of the 14th annual ACM symposium on User interface software and technology:219-226, 2001.
[20] S. Jorda, M. Kaltenbrunner, G. Geiger and R. Bencina. The reactable. Proceedings of the international computer music conference (ICMC 2005):579–582, 2005.
[21] A.D. Wilson. TouchLight: an imaging touch screen and display for gesture-based interaction. Proceedings of the 6th international conference on Multimodal interfaces:69-76, 2004.
[22] W. Buxton, R. Hill and P. Rowley. Issues and techniques in touch-sensitive tablet input. Proceedings of the 12th annual conference on Computer graphics and interactive techniques 19(3):215-224, 1985.
[23] T. Döring, A. Sahami Shirazi and A. Schmidt. Exploring gesture-based interaction techniques in multi-display environments with mobile phones and multi-touch table. Proceedings of the International Conference on Advanced Visual Interfaces:419, 2010.
[24] E. Tse, C. Shen, S. Greenberg and C. Forlines. Enabling interaction with single user applications through speech and gestures on a multi-user tabletop.Proceedings of the working conference on Advanced visual interfaces:336-343, 2006.
[25] Y. Takeoka, T. Miyaki and J. Rekimoto. Z-touch: an infrastructure for 3d gesture interaction in the proximity of tabletop surfaces. ACM International Conference on Interactive Tabletops and Surfaces:91-94, 2010.
[26] N. Villar, S. Izadi, D. Rosenfeld, H. Benko, J. Helmes, J. Westhues, S. Hodges, E. Ofek, A. Butler and X. Cao. Mouse 2.0: multi-touch meets the mouse. Proceedings of the 22nd annual ACM symposium on User interface software and technology:33-42, 2009.
[27] M. Kaschny, S. Buron, U. von Zadow and K. Sostmann. Medical education on an interactive surface. ACM International Conference on Interactive Tabletops and Surfaces:267-268, 2010.
[28] H. Jang, J. Kim and C. Lee. Augmented reality cooking system using tabletop display interface. Proceedings of the 5th International Symposium on Ubiquitous Virtual Reality: 1, 2007.
[29] J. Fernquist, K.S. Booth, A.K. Mackworth, R. Kellett and C. Girling. Using Multi-touch Tabletops to Create and Compare Neighbourhood Designs that Satisfy Constraints. 2nd International Conference on Computational Sustainability, 2010.
[30] B. Buxton. 31.1: Invited Paper: A Touching Story: A Personal Perspective on the History of Touch Interfaces Past and Future. SID Symposium Digest of Technical Papers, 41(1):444-448, 2012.
[31] J.Y. Han. Multi-touch interaction wall. ACM SIGGRAPH 2006 Emerging technologies:25, 2006.
[32] H. Benko, A.D. Wilson and P. Baudisch. Precise selection techniques for multi-touch screens. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems:1263-1272, 2006.