Bare Hand Natural Interaction with Augmented Objects

Lucas Figueiredo
Voxar Labs
Informatics Center - UFPE
Recife/PE, Brazil
lsf@cin.ufpe.br

Jorge Lindoso
Voxar Labs
Informatics Center - UFPE
Recife/PE, Brazil
jefl@cin.ufpe.br

Rafael Roberto
Voxar Labs
Informatics Center - UFPE
Recife/PE, Brazil
rar3@cin.ufpe.br

Veronica Teichrieb
Voxar Labs
Informatics Center - UFPE
Recife/PE, Brazil
vt@cin.ufpe.br

Ronaldo dos Anjos
Voxar Labs
Informatics Center - UFPE
Recife/PE, Brazil
rfaf@cin.ufpe.br

Edvar Neto
Voxar Labs
Informatics Center - UFPE
Recife/PE, Brazil
excvn@cin.ufpe.br

Manoela Silva
Voxar Labs
Informatics Center - UFPE
Recife/PE, Brazil
mmos@cin.ufpe.br

Extended Abstracts of the
IEEE International Symposium on Mixed and Augmented Reality 2013
Science and Technology Proceedings
1 - 4 October 2013, Adelaide, SA, Australia
978-1-4799-2869-9/13/$31.00 ©2013 IEEE

Abstract

In this work in progress we address the problem of interacting with augmented objects. A bare hand tracking technique is developed, which allied to gesture recognition heuristics, enables interaction with augmented objects in an intuitive way. The tracking algorithm uses a flock of features approach that tracks both hands in real time. The interaction occurs by the execution of grasp and release gestures. Physics simulation and photorealistic rendering are added to the pipeline. This way, the tool provides more coherent feedback in order to make the virtual objects look and respond more likely real ones. The pipeline was tested through specific tasks, designed to analyze its performance regarding the easiness of use, precision and response time.

Author Keywords

bare hand tracking; virtual grasp; augmented objects

ACM Classification Keywords

K.5.2 [Information Interfaces And Presentation]: User Interfaces - Interaction styles.

In this work we address the problem of direct interaction with the augmented content on a scene, which provides to the user natural means to manipulate virtual objects. In tabletop AR applications, the augmented objects are usually rendered within the user arms reach, which encourages the use of tracking algorithms and devices to enable the user with the ability to physically reach these objects. These methods include marker based metaphors [1], attached devices [5], trackable or haptic gloves [19] [9] and bare hand tracking techniques [12].

Our goal is to develop an accessible tracking and gesture recognition tool, which requires no calibration, no attached device and is capable of providing a result that is fast and precise enough to enable the user to directly interact with augmented objects with minimum training. The tracking is aimed (but not restricted) to tabletop applications with augmented objects within the user arms reach. The interaction is performed by a grasp gesture, interpreting the cases in which the user would actually grab the object with two fingers.

The solution requires an inexpensive setup, composed mainly by a computer with CUDA enabled graphics cards [10] and a Kinect device [8]. A real time ray tracer is applied [17]. The proposed tracking technique is evaluated using the photorealistic rendering scenario presented in [6], also integrated with physics simulation. The addition of these features aims to improve the intuitiveness of the application, enabling the virtual objects to act as real ones. In order to evaluate the tracking performance regarding its precision and robustness to fast hand movements, two specific tasks are defined and tested with twenty users.

The document is structured as follows. The next section presents the main related works. Third section details the implementation decisions regarding the tracking, rendering and physics simulation. The case study tests and results are detailed in the fourth section. Last section presents the main conclusions and the assigned future work.

Related Work

Regarding the bare hand tracking, there are model based methods [18] [12]. These methods retrieve the exact articulated positioning of the user hand. However, due to the complexity of the model some issues arise. For example, the range of possible configurations may be limited in order to achieve real time performance. Other possible issue is the occurrence of tracking failures in relatively fast motions due to camera image blur which increases the ambiguity of fingers configuration.

There are also hand tracking methods that are not based on the hand model. They are capable of detecting and following the region that the hand occupies on the scene. Among this type of trackers, the flock of features approach is suitable to track hands as shown in [4] and [13]. This type of tracking aims to be robust to fast movements and changes of the fingers configuration giving the user more freedom during the experience.

Regarding the grasp gesture, there are physically based [15] and heuristic based approaches [3]. In the second one, the real objects (e.g. user hands) are physically represented on the virtual scenario by a set of spheres enabling the user to pick and drop by embracing the augmented objects with thumb and forefinger, for example by using a physically based virtual grasp, however, without defining a precise hand model. Heuristic approaches in contrast do not need to simulate the hand physical presence on the scene. This characteristic can be useful for the physics simulation since the hand presence as a real object may pass through virtual objects and cause physical instability. Moreover, heuristic approaches may simulate the physical aspect of the grasp gesture by considering the position of the user's fingers in relation to the target virtual object as presented in [7].

Developed System

The developed tracker aims to be robust to rapid spatial hand movements as well as to local movements, such as changes of fingers configuration. Firstly, the skin color pixels are segmented using a Bayesian classifier [14] combining the concept of expansive searches from high probable skin pixels to others with less probability [16]. The search is also limited by extracted edges [2] using canny algorithm. The detection procedure is similar to the described in [13], however, the tracker searches only for three open fingers and restricts the sequence of peaks and valleys to respect a pentagonal pattern, which showed to be more robust than search for five open fingers.

Once a color group is detected as a hand, the tracking procedure takes place. Features are extracted and followed through the subsequent frames. The features are treated as a cloud of points with high entropy, however with a control effort to keep them near each other. Distant features are removed or relocated to be nearer to the group. Additionally, at each frame new features are added to refill the lost ones. Considering the entire pipeline, the flock of features tracking is suitable to track hands considering its highly dynamic nature. Changes of hand configuration, spatial movements and scale issues are treated by the algorithm self-healing structure which constantly tracks and updates its feature cloud.

The grasp and release gestures are recognized by estimating the positions of user's thumb and forefinger. The proposed approach does not depend on tracking a specific hand model. The goal is to detect and track the gesture of the grasp hand shape. This is done by searching for a compatible gap in the hand contour nearby the palm point. If there is an augmented object within the region between these fingers points and the evaluated distance is short enough, the object is grabbed. Analogously, if the distance increases the object is released.

The last step is the definition of the 3D aspect of the gesture, which consists of calculating a 3D point for thumb and forefinger. Firstly, a depth map value is assigned to the points. A spiral search for the nearest valid depth information is performed, starting from the palm point. The first found depth is assigned to both thumb and forefinger. This way, it is possible to simulate interactions with the virtual objects over the table. In order to detect if some object is being grabbed, a sphere is created between the grasp fingers. This sphere is not rendered, existing only to detect collisions with virtual objects in order to decide if they must be grabbed by the user. Once a collision is detected, if the distance between the gesture fingers is small enough, the object is grabbed.

The table position and orientation is calculated by collecting three points in the middle region of the sensor image and then estimating a plane. The physics simulation is performed using an abstraction tool above the NVidia PhysX library [11]. Once a grasp gesture is performed the object is attached to the mid point between thumb and forefinger. Moreover, a real time ray tracing technique is applied [17]. Using the depth map info it is possible to assign a 3D position to each pixel of the image. Thus, the renderization considers the interaction of the augmented objects with the real 3D points map. The rendering includes effects like refractions, reflections, soft shadows and occlusions, being all casted from real upon augmented objects and vice-versa. Figure 1 illustrates render results combined with physics simulation, showing objects supported by the physical plane of the detected table, casting shadows upon the real scene, including dynamic content like users hands.

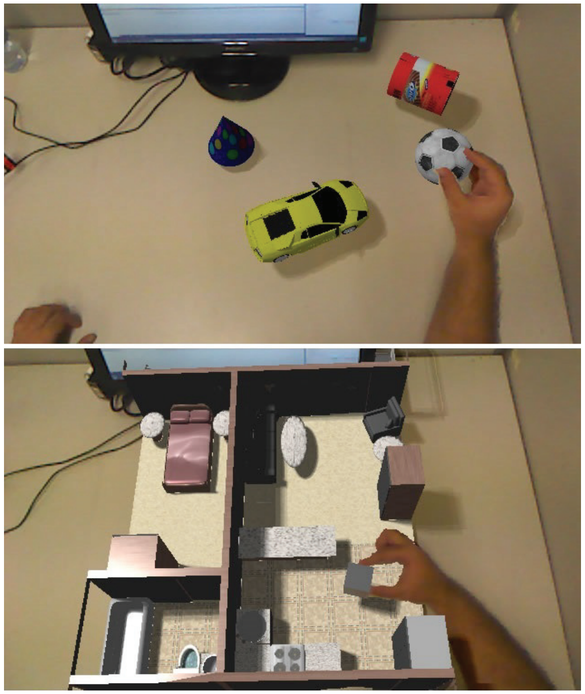

Figure 1: Case studies using our tool. Top: children toys over the table. These can be used to teach types of objects, such as spheres, cylinders, cones, and their physical interactions. Bottom: a setup for manipulating virtual furniture exploring the concept of interior design.

Case Study and Results

The application setup consisted on a Kinect device fixed with a wall support in height approximately 40 cm above users head. The Kinect device pointed directly to the work table. The user table had 72 cm of height and the users were seated while experimented the application. The visualization occurred on a 22 monitor. The application runs on a Intel Core i5 2.8GHz processor, having 4GB of RAM and a Geforce 440 GT graphics card from NVidia. The overall frame rate of the application during the tests was approximately 25 frames per second.

Twenty users participated on the formal tests, being seventeen men and three women. The age of the users varied from eighteen to thirty nine years old. Two tasks were designed and the users were randomized in two equally sized groups according to the order they experimented the tasks. Before the users performed the required tasks, a familiarization time of at most two minutes was given to use the technology without restrictions.

Task 1 (T1) was designed to evaluate the activity of picking an augmented object, move and place it on another point of the scene. The user was asked to pick a virtual cube from one side of the table and move it to the opposite side in a distance of at least 50 cm. Two boundaries were indicated on the table to guide the user about how far he/she should move the object. The task should be repeated six times by each user, each time moving the object from the left side to the right side and vice-versa. Each execution was timed for further analysis.

Task 2 (T2) aimed to evaluate the tracker precision. The user was asked to align a red cube inside another green cube with the exact same size but being intangible. The user should take as much time as needed to align them and at the moment he/she considered that the maximum alignment was reached, three measures were recorded: the time spent performing the task, the distance between the cubes and the difference of the orientation angle.

The main results are presented in Table 1. The average time of execution for T1 was 6.9 seconds followed by a standard deviation of 3.6. The minimum time a user achieved for this task was 1.2 seconds while the maximum was 39. However the user who took 39 seconds on its second attempt improved considerably in the remaining attempts by achieving times of 11, 15,5 and 5 seconds, respectively. T1 evidenced that all users were able to move the virtual objects from one point to another in a relatively short time. Considering the execution of the same task with real objects, which is about 1.5 seconds, the performance with the proposed system is slower. Still, it allows the execution of the task with virtual objects in a small time. By its turn, T2 presented an average performance time of 32 seconds and the average errors in the alignment regarding the distance (translation error) and angle (rotation error) were 2.3 cm and 5 degrees respectively. The maximum and minimum errors were of 5 and 0.6 cm, and 21 and 0.1 degrees. This way, the overall precision revealed results that would restrict the range of applications to low precision tasks requiring no more than 1 to 3 cm of precision.

Table 1: Main results of the specified tasks. Time is measured in seconds, the distance in centimeters and the angle in degrees.

	Time T1	Time T2	Distance T2	Angle T2
Average	6.88	32.17	2.29	5.05
Deviation	3.59	30.48	1.19	4.94

Conclusion

The presented work proposes and evaluates a solution for direct interaction with augmented objects on tabletop applications through hand tracking and gesture recognition. A flock of features tracker is implemented as well as the heuristic based gesture recognition. The solution is integrated with physical simulation and photorealistic rendering. It provides the user with more realistic augmented objects. The interaction is evaluated considering two tasks aimed to understand the tool behavior regarding the grasp and release execution time and precision. The results revealed that the proposed tool is still suitable for applications with more loose requirements regarding precision.

As future work, one objective is to use more deeply the depth information provided by the sensor, using inpaint algorithms to fulfill the gaps on the map and assigning more precise 3D positions for the fingertips. The physical design of the gesture is another point for improvement, by developing a spring model with spheres attached to the fingertips, approximating the gesture behavior to the real one.

Acknowledgements

The authors would like to thank CNPq, CAPES, FACEPE and Reuni-UFPE for partially funding this research.

References

[1] J. Almgren, R. Carlsson, H. Erkkonen, J. Fredriksson, S. Moller, H. Rydgard, M. Osterberg, and M. Fjeld. Tangible user interface for chemistry education: Visualization, portability, and database. In *Visualization, Portability, and Database. Proc. SIGRAD 2005*, pages 19–24, 2005.

[2] D. Ghimire and J. Lee. A Robust Face Detection Method Based on Skin Color and Edges. 9(1):141–156, 2013.

[3] O. Hilliges, D. Kim, S. Izadi, M. Weiss, and A. Wilson. Holodesk: direct 3d interactions with a situated see-through display. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 2421–2430, New York, NY, USA, 2012. ACM.

[4] M. Kolsch and M. Turk. Hand tracking with flocks of features. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 1187 vol. 2–, 2005.

[5] F. Liarokapis, L. Macan, G. Malone, G. Rebolledo-Mendez, and S. d. Freitas. A pervasive augmented reality serious game. In *Proceedings of the 2009 Conference in Games and Virtual Worlds for Serious Applications*, VS-GAMES '09, pages 148–155, Washington, DC, USA, 2009. IEEE Computer Society.

[6] J. E. F. Lindoso, L. S. Figueiredo, V. Teichrieb, R. Roberto, and R. F. A. Filho. Integrated pipeline for natural interaction with photorealistic rendering. In *SVR 2013 ()*, Cuiab/MT, may 2013.

[7] H. Maekawa and J. Hollerbach. Haptic display for object grasping and manipulating in virtual environment. In *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on*, volume 3, pages 2566–2573 vol.3, 1998.

[8] C. Microsoft. Kinect - xbox.com. http://www.xbox.com/en-US/kinect, apr 2013.

[9] M. Moehring and B. Froehlich. Effective manipulation of virtual objects within arm's reach. In *Virtual Reality Conference (VR), 2011 IEEE*, pages 131–138, 2011.

[10] C. NVIDIA. Nvidia cuda c programming guide. http://bit.ly/131aIar, feb 2013.

[11] C. NVIDIA. Projects using physx sdk. http://physxinfo.com/, mar 2013.

[12] I. Oikonomidis, N. Kyriazis, and A. Argyros. Full dof tracking of a hand interacting with an object by modeling occlusions and physical constraints. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2088–2095, 2011.

[13] Z. Pan, Y. Li, M. Zhang, C. Sun, K. Guo, X. Tang, and S. Zhou. A real-time multi-cue hand tracking algorithm based on computer vision. In *Virtual Reality Conference (VR), 2010 IEEE*, pages 219–222, 2010.

[14] S. L. Phung, A. Bouzerdoum, and D. Chai. Skin segmentation using color pixel classification: Analysis and comparison. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(1):148–154, Jan. 2005.

[15] M. Prachyabrued and C. W. Borst. Virtual grasp release method and evaluation. *Int. J. Hum.-Comput. Stud.*, 70(11):828–848, Nov. 2012.

[16] J. Ruiz-del solar and R. Verschae. SKINDIFF Robust and Fast Skin Segmentation. Technical Report March, Computational Vision Group Department of Electrical Engineering Universidad de Chile, 2006.

[17] A. Santos, D. Lemos, J. Lindoso, and V. Teichrieb. Real time ray tracing for augmented reality. In *Virtual and Augmented Reality (SVR), 2012 14th Symposium on*, pages 131–140, 2012.

[18] B. Stenger. Template-based hand pose recognition using multiple cues. In *Proceedings of the 7th Asian conference on Computer Vision - Volume Part II*, ACCV'06, pages 551–560, Berlin, Heidelberg, 2006. Springer-Verlag.

[19] R. Y. Wang and J. Popović. Real-time hand-tracking with a color glove. *ACM Trans. Graph.*, 28(3):63:1–63:8, July 2009.