Journal of Forensic Radiology and Imaging 1 (2013) 10–14

Contents lists available at SciVerse ScienceDirect

Journal of Forensic Radiology and Imaging

journal homepage: www.elsevier.com/locate/jofri

Invisible touch—Control of a DICOM viewer with finger gestures using the Kinect depth camera

L.C. Ebert [a,*], G. Hatch [b], M.J. Thali [a], S. Ross [c,d]

[a] University of Zurich, Institute of Forensic Medicine, Forensic Medicine and Imaging and Virtopsy, Winterthurerstrasse 190/52, CH-8057 Zurich, Switzerland
[b] Centre of Forensic Imaging, Office of the Medical Investigator, University of New Mexico, MSC07 4040, 1101 Camino de Salud NE, Albuquerque, New Mexico 87102, USA
[c] University of Bern, Institute of Forensic Medicine, Center for Forensic Imaging and Virtopsy, Bühlstrasse 20, CH-3012 Bern, Switzerland
[d] University of Linköping, Center for Medical Image Science and Visualization (CMIV), SE-581 85 Linköping, Sweden

A R T I C L E I N F O

Article history:
Received 17 July 2012
Received in revised form
25 October 2012
Accepted 5 November 2012
Available online 13 December 2012

Keywords:
Virtopsy
PACS
OsiriX
Kinect
Gesture Detection

A B S T R A C T

With the increasing use of imaging technologies during surgeries and autopsies, new control methods for computer systems are required to maintain sterility. Gesture controlled systems seem to be promising, since they allow for a touch-free control of computer systems. In a previous publication we presented a system which allows the control of the open source Picture Archiving and Communication System (PACS) OsiriX by means of gesture and voice commands. In order to overcome the limitations of this system, we developed a plug-in for OsiriX that allows for gesture control of the DICOM viewer of OsiriX with finger gestures.

1. Introduction

Medical imaging datasets are increasingly used for planning and executing surgery [1] and autopsy [2]. This requires access to medical image viewers in the operating room and autopsy suite. However, using standard input devices such as mouse and keyboard is a potential vector of infectious agents, posing risks to both patients and staff [3]. One solution to this problem is the use of gestures and voice commands to control imaging viewing software. Both are intuitive means of communicating and can allow for touch-free operation of computer systems. Systems that use standard color cameras for gesture identification have been presented, but they require calibration or can be sensitive to changes of lighting [4–6]. Several groups are working on systems combining image viewers with gesture control provided by the Kinect depth camera (Microsoft, Redmond, USA)[7–9]. Some commercial systems are in the making, such as the gesture control system TedCas from TedeSys (Spain)[10]. All these implementations use the skeleton tracking feature included in the Kinect programming interface. Since the skeleton tracking designed for gaming by tracking gross body movement, its accuracy is rather low. We endeavored to overcome this problem via the direct use of the depth images provided by the camera.

Using the Kinect depth camera and a headset (Logitech ClearChat PC wireless headset, Logitech, Switzerland) for voice commands, we created and tested a working prototype of a software application that allowed users to control the open source PACS software OsiriX (OsiriX foundation, Geneva, Switzerland) by means of hand gestures and voice commands [11]. However, our study revealed some drawbacks. The voice recognition worked poorly if there was background noise or the user had a non-American accent. Also, the required headset was considered by many users to be ungainly and distracting. The rather simple method of gesture detection implemented in the first prototype led to involuntary changes if the hands were moved in and out of the view volume. We concluded that more discrete detection of finger gestures might help to overcome these limitations.

In this work, we present a new prototype that enables the user to control the OsiriX DICOM viewer with finger gestures using a low cost depth camera. It runs as a plug-in directly from OsiriX.

2. Materials and methods

2.1. Hardware

For running the OsiriX as well as the Kinect plug-in, we used a MacBook pro (Intel Core i7 2.2 GHz, 8 GB RAM, Apple Inc., USA).
Kinect

A recent development in consumer electronics is the use of a low cost depth camera developed for gesture control of games.

* Corresponding author. Tel.: +41 44 635 6851.
E-mail address: lars.ebert@virtopsy.com (L.C. Ebert).

2212-4780/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jofri.2012.11.006

The Kinect depth camera was first presented in November 2010 as a new input device for the Xbox 360 gaming console.

The Kinect camera system features an infrared light projector and an infrared camera for creating a depth image in addition to a normal color image. This depth image allows a relatively easy implementation of gesture control in contrast to color coded images, since it significantly simplifies the segmentation of single objects. The camera has a sampling rate of 30 frames per second (fps). The view angle of the camera is 57 degrees horizontally and 43 degrees vertically. The camera can be connected to a computer via a standard Universal Serial Bus (USB) connection.

2.2. Software

2.2.1. OsiriX plug-in

OsiriX is an open source PACS system [12]. OsiriX features a DICOM viewer as well as a plug-in interface, which allows for easy extension of the functionality of OsiriX. Our plug-in translates depth images delivered by the Kinect camera into control commands of the OsiriX image viewer. The plug-in was written in Objective-C using XCode (Version 4.1, Apple Inc., USA), the Libfreenect library [13] to access the Kinect depth images and OpenCV [14] for image processing. The plug-in can be started directly from OsiriX and gives a visual feedback on the hand gesture detection in a live view window as displayed on the computer monitor (see Fig. 1). The plug-in runs in real time.

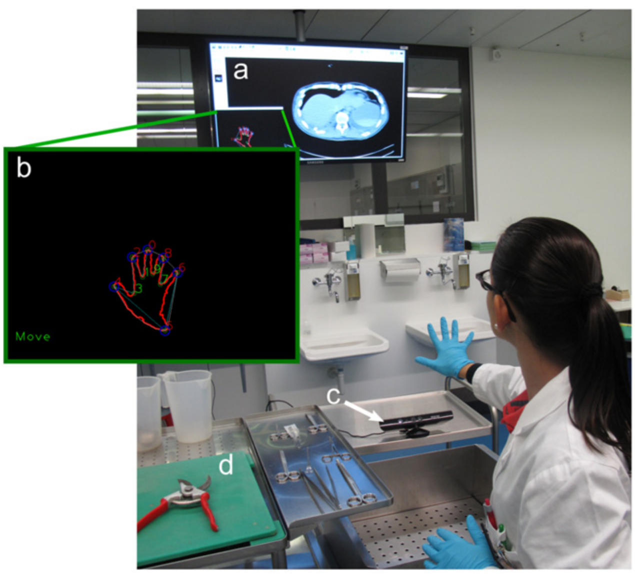


Fig. 1. Setup of the Kinect gesture control system in the autopsy suite. (a): Screen displaying the current dataset and the live view window. (b): Live view window displaying the processed image seen by the camera as well as the gesture currently recognized. (c): Kinect camera. (d): Autopsy table.

2.3. Detection of finger gestures

Finger detection is performed on the cameras depth images. The principle is explained in Fig. 2. In the first step, the object closest to the camera is determined. Based on this information, close objects are cut out (threshold segmentation and binarization) and separated (blob detection). The contour of each blob is calculated and a polygon is calculated that approximates the contour. In order to identify fingertips, all points of the polygon (vertex) that are on the outside of the shape are determined. This is done by calculating the convex hull, which means the smallest surface that surrounds the object without being concave. All contour vertexes that lie on the convex hull as well are considered to be potential fingertips. Next, the angle of each polygon is calculated and assigned to the corresponding vertex. The combination of finger/non finger vertexes and according angles defines the signature of the object. Signatures are then compared to a library of finger postures. Finger postures are recorded over time and the algorithm looks for motion patterns. Once a pattern is recognized, the current state of the OsiriX viewer is altered accordingly. A flowchart displaying how the algorithm works can be seen in Fig. 3.

2.4. Gesture commands

If a gesture is recognized, it is displayed in the live view window. The possible gestures include control of the position, magnification and window of the current OsiriX DICOM image as well as alteration of the current position within the dataset. All valid finger gestures are displayed in Fig. 4. Panning the image is possible by presenting an open hand with the fingers and thumb spread apart to the camera. Shifting the position of the hand in plane results in panning of the image, with the image following the direction of the hand shift. To zoom in or out, both hands form a frame. By moving them away from each other, the image is magnified, bringing the hands closer together zooms out. For

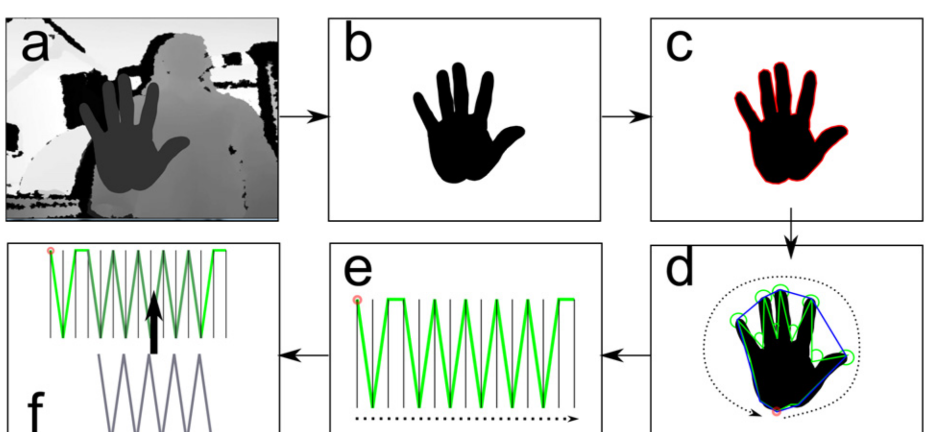


Fig. 2. Hand gesture detection on depth images. (a): Original depth image.(b): Segmentation of objects closest to the camera by thresholding and binarization. (c): Blob detection and extraction of contours (red). (d): Approximation of contour with polygons (green), angles between polygon edges and calculation of convex hull (blue). Polygon points (vertexes) that lie on the outside (on the convex hull) are considered to be fingertips. (e): Counterclockwise creation of signature by determining whether a vertex lies on the convex hull or not (green). Additionally the angles are determined for each vertex. (f): Signature and angles are compared to predefined patterns. The information about a detected pattern is recorded over time and then used to identify the gesture, altering the image viewer accordingly.

navigation, both hands form a V with index and middle finger. Moving the right hand up and down scrolls through the dataset. For changing window and level, we implemented gesture directed presets. If the computer recognizes both hands and the left hand is fully opened (fingers and thumb spread apart), the right hand gesture defines which one of the four presets is to be selected. A computer generated voice gives acoustic feedback about which window level is set.

2.5. System evaluation

To evaluate the system, we gave six medical professionals (four forensic pathologists, one radiologist and one engineer) a standardized instruction into the system. After that, they were allowed to use the system themselves under guidance until they felt comfortable with it. This time was measured. To verify their capability of operating the system, they were shown a CT dataset and asked to find and display a specific anatomical structure or pathology by using gesture control. Finally, the participants were asked to rate the intuitiveness of the gestures on a Likert-type scale (1 = not intuitive at all, 5 = very intuitive) as well as how well they could control the image viewer with the gestures (1 = not usable, 5 = full control).

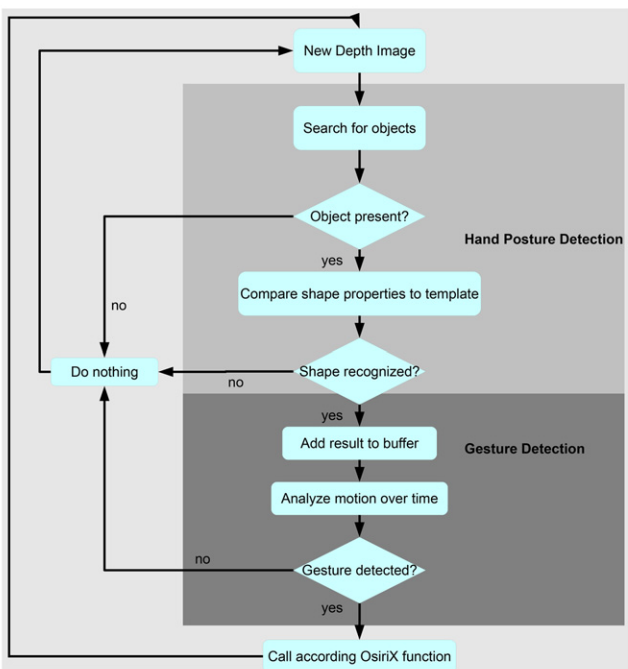


Fig. 3. Flowchart displaying the function of the plug-in.

2.6. Statistics

We calculated the average time required to be able to use the system as well as the average answers (mean and standard deviation) using Microsoft Excel (Mircosoft Corp., Redmond, USA).

3. Results

On average, 4,5 min ($\pm$ 2.2) were required to be able to use the system. Participants rated the intuitiveness of the gestures with 3.8 out of 5 ($\pm$ 1.2). Asked how well they could control die images, they rated 3.8 out of 5 ($\pm$ 0.4).

4. Discussion

Touch-free manipulation of radiological images is a desirable function in clinical[15] as well as in forensic medicine. To our knowledge, we are the first to present a prototype that allows a user to control a DICOM viewer with finger gestures using the Kinect camera. Prior gesture controlled image viewing systems did not use finger gestures, integrated with proprietary PACS systems or were based on standard color cameras, which makes them susceptible to changes of lighting or color conditions. Professional PACS systems can cost many tens of thousands of dollars, which seriously limits the acquisition of such technology by small centers or non-imaging specialists. The chosen system is inexpensive; at a cost of approximately 2000$ for a suitable Apple computer, 150$ for a Kinect camera and the free OsiriX software the system is ready to use and has comparable, if not superior functionality for image review. In cases where a 64-bit version is required (i.e. if large datasets have to be loaded), around 400$ are charged for a single user license. The use of the low cost, widely available equipment significantly reduces barriers to acquiring or implementing touch-free navigation.

We developed an OsiriX plug-in, that allows the navigation and manipulation of an image stack with finger gestures, based on the experiences with our last prototype. A demonstration of the

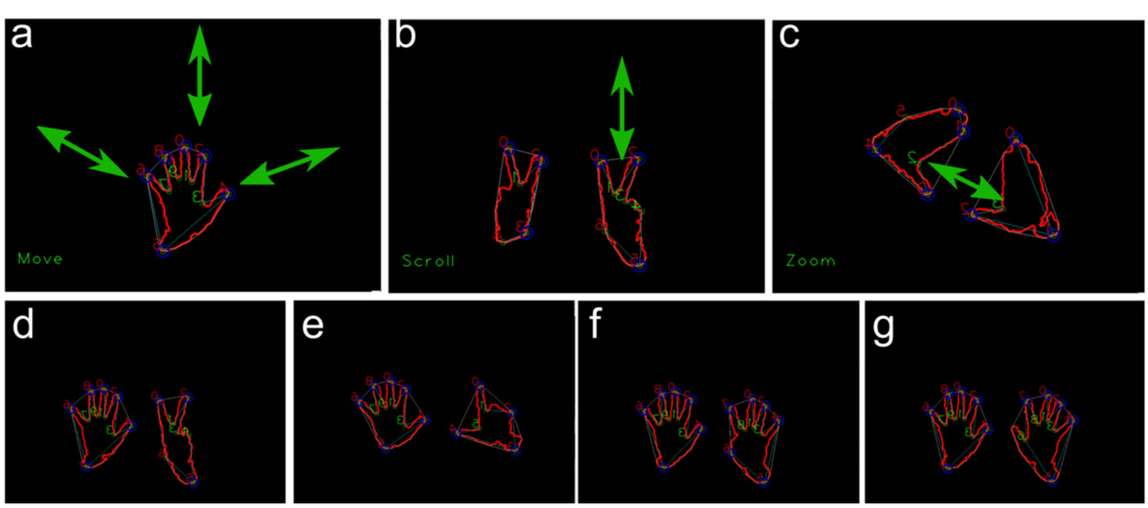
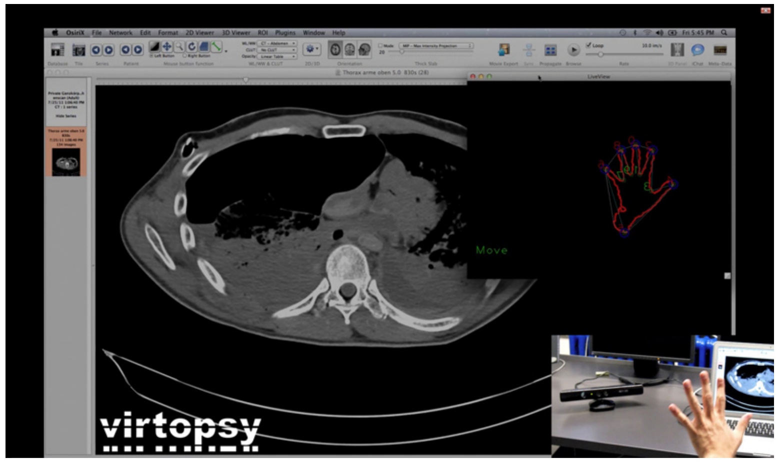


Fig. 4. Possible finger gestures. (a): Panning. (b): Scrolling. (c): Zooming. (d–g): Window presets for lung-, bone-, soft tissue- and brain windows.



Video 1. Demonstration of the system, showing how different gestures can be used to control the image viewer. A video clip is available online. Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jofri.2012.11.006.

system can be seen in Video 1. By eliminating voice input, a major source of error has been removed. Since the new plug-in requires specific finger gestures to trigger the functions of the PACS-system, the interaction is more robust in areas with background noise. The prior iteration of our gesture control did not distinguish between differently shaped objects for gesture control, which made it prone to accepting false commands. The more complex algorithm for finger gesture recognition therefore reduced the amount the number of false commands significantly.

Compared to the prior hand gesture and voice controlled system [11], finger gestures are considered more intuitive (3.8/5 vs. 3.4/5) by the users. Users state that they have a better control by using the new system (3.8/5 vs. 3.4/5), especially considering the input inaccuracy of voice control (3.0/5). Since no voice commands have to be memorized, the time required to learn the system is relatively short.

During autopsy, the system could assist pathologists to find the locations of foreign bodies, small bone fractures or other pathologies that are visible in CT scans.

The system works well with liquid coated rubber gloves. An infrared tracking camera used for surgical navigation (Axios CamBar 2, Axios3d Services GmbH, Germany) had no effect on the Kinect camera. However, different tracking cameras work with different wavelengths and should therefore be tested separately. The effect of the Kinect infrared laser on the measurement accuracy of such tracking cameras should be investigated as well.

Since finger gestures can be used at a close range to the camera, it allows the usage of gesture control even in rooms with tight spatial constraints. In general, the algorithm works best between 50 and 90 cm distance between camera and hands.

The use of shape recognition alone on depth images allows for a reliable detection of finger gestures only if the hand is perpendicular to the camera. The more the hand is angulated, the more difficult it is for the algorithm to recognize the gestures. The visual feedback of our live view helps the user to identify this problem and correct the gesture accordingly. Since the resolution of the depth camera images is rather low, in some cases the algorithm has difficulty separating the fingers, for instance if they are held too close together. A new generation of depth cameras with higher resolutions (planned for release by Microsoft in 2012) might help to overcome this issue. Another limitation of the current system is that it is not possible to change between patients or series due to the limited number of gestures that can currently be detected.

By combining body posture and finger gesture detection, it might be possible to significantly increase the reliability and intuitiveness of the system.

Future development may include the improvement of the finger gesture recognition and the validation of the prototype for use in operating rooms and autopsy suites. We envision giving the user complete control over OsiriX (i.e. switching between series, volume rendering etc.) instead of just controlling axial views in a single series. The use of more advanced pattern recognition methods [16,17] might help to improve the robustness of the gesture recognition.

5. Conclusion

With the use of finger gesture detection, it is possible to control the basic functionality of a medical image viewer. Finger gesture control has conceivably eliminated the major problems encountered during testing of a preview iteration of the remote control plug-in, namely voice recognition that struggled with accents and gesture detection that proved too simplistic. This system, made up of low cost and widely available equipment places touch-free image navigation within reach for nearly any potential user. Future applications are primarily envisioned in operating rooms and autopsy suites, but could include any situation where touch-free radiologic image review is desirable.

Disclosure statement

We declare, that we have no conflict of interest.

Funding

none.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jofri.2012.11.006.

References

[1] W Korb, S Bohn, O Burgert, A Dietz, S Jacobs, V Falk, et al., Surgical PACS for the digital operating room. Systems engineering and specification of user requirements, Studies in Health and Technology Informatics 119 (2006) 267–272.

[2] MJ Thali, K Yen, W Schweitzer, P Vock, C Boesch, C Ozdoba, et al., Virtopsy, a new imaging horizon in forensic pathology: virtual autopsy by postmortem

multislice computed tomography (MSCT) and magnetic resonance imaging (MRI)–a feasibility study, Journal of Forensic Sciences 48 (2) (2003) 386–403.
[3] B Hartmann, M Benson, A Junger, L Quinzio, R Röhrig, B Fengler, et al., Computer keyboard and mouse as a reservoir of pathogens in an intensive care unit, Journal of Clinical Monitoring and Computing 18 (1) (2004) 7–12.
[4] JP Wachs, HI Stern, Y Edan, M Gillam, J Handler, C Feied, et al., A gesture-based tool for sterile browsing of radiology images, Journal of the American Medical Informatics Association 15 (3) (2008) 321–323.
[5] C Grätzel, T Fong, S Grange, C. Baur, A non-contact mouse for surgeon-computer interaction, Technology and Health Care 12 (3) (2004) 2–245.
[6] Kipshagen T, Graw M, Tronnier V, Bonsanto M, Hofman UG. Touch- and marker-free interaction with medical software. IFMBE Proceedings of the World Congress on Medical Physics and Biomedical Engineering, Springer, vol. 25, no. I–XIII , 2009.
[7] Xbox Kinect in the Hospital Operating Room, Sunnybrook Health Science Center, Canada. 〈http://www.youtube.com/watch?v=f5Ep3oqicVU〉, (accessed 03.20.2012).
[8] J Tan, J Unal, K. Link Kinect Sensor Allows Surgeons to Manipulate 3D CT Images in Midair, Wake Forest University Baptist Medical Center. 〈http://www.youtube.com/watch?v=id7OZAbFaVI〉, (accessed 03.20.2012).
[9] KineMed – Gestural interface for Philips iSite, UPMC, Pittsburgh, USA. 〈http://www.youtube.com/watch?v=Lt5Dv250rPM〉, (accessed 03.20.2012).
[10] TedCas: Kinect in Operating Rooms and Hospitals, Tedesys Spain. 〈http://www.youtube.com/watch?v=Ekg4p2QySoY〉, (accessed 03.20.2012).
[11] LC Ebert, G Hatch, G Ampanozi, MJ Thali, S. Ross, You Can't Touch This: Touch-free Navigation Through Radiological Images, Surgical Innovation (2011), Epub ahead of print.
[12] A Rosset, L Spadola, O. Ratib, OsiriX: an open-source software for navigating in multidimensional DICOM images, Journal of Digital Imaging 17 (3) (2004) 205–216.
[13] Libfreenect: 〈https://github.com/OpenKinect/libfreenect〉, (accessed on 11.30.2011).
[14] OpenCv: 〈http://opencv.willowgarage.com/wiki/〉, (accessed on 11.30.2011).
[15] R Johnson, K O'Hara, A Sellen, C Cousins, A. Criminisi Exploring the Potential for Touchless Interaction in Image Guided Interventional Radiology, Proceedings of ACM Conference on Computer-Human Interaction (CHI). ACM Conference on Computer–Human Interaction, 7 May 2011.
[16] SS Ge, Y Yang, TH. Lee, Hand gesture recognition and tracking based on distributed locally linear embedding, Image and Vision Computing 26 (12) (2008) 1607–1620.
[17] E Stergiopoulou, N. Papamarkos, Hand gesture recognition using a neural network shape fitting technique, Engineering Applications of Artificial Intelligence 22 (8) (2009) 1141–1158.