Laser pointer interaction for 3D angiography operation

Dina Ahmed^{**}, Ayman Atia^{*}, Essam A. Rashed^{**}, and Mostafa-Samy M. Mostafa^{*}

*HCI-LAB, Department of CS, Faculty of Computers and Information, Helwan University, Helwan, Egypt **Image Science-LAB, Department of Mathematics, Faculty of Science, Suez Canal University, Ismailia, Egypt

Abstract: A primary challenge for creating an interactive display in operating room (OR) is in the definition of control methods that are efficient and easy to learn for the physician. Apart from traditional input methods such as mouse and keyboard, we present a system that utilizes the laser pointer to manipulate directly the 3D medical images of angiography on LCD display. This work is based on laser detection, tracking the spot, and laser gesture recognition using an ordinary laser pointer. The laser spot is captured by a web camera and its location recognized with the help of the computer vision techniques. The system includes two moving operations to execute commands: "circle gesture" with two directions and "line gesture" with four directions. The recorded laser gestures are then recognized using two different algorithms: dynamic time wrapping (DTW) and one dollar (1\$) recognizer. Our experiments show that the DTW algorithm performs better with an overall accuracy of %89.6 and is faster than the 1\$ algorithm. This paper describes the proposed system, its implementation, and an experiment that is performed to evaluate it.

Keywords: Direct interaction, laser pointers, 3D image manipulation, human-computer interaction.

I. Introduction

Large displays are increasingly being critical and prevalent in our daily activities. Previous studies have shown their benefits in several tasks such as navigation [1], multitasking [2]. As well as they encourage communication and collaboration among members of a group in a shared area [3]. User interaction with large displays is one of the important topics in human-computer interaction (HCI) research community [4]. Recently, many applications of a large interactive display have been explored. It had a significant impact on different fields such as medical image processing, advertising, information visualization, public collaboration, and education classrooms [5, 6]. As large display becomes widespread and ubiquitous, it presents some challenges that can not be easily handled by using traditional input devices such as mouse and keyboard [7]. However, users of large displays are often standing, working up close to the screen and may want to walk unrestrictedly in front of it. It is awkward and not practical to ask them to hold the mouse while manipulating the objects on the screen.

Typically, any person in a meeting room, where an object is shown on a large display, uses a laser pointer device to point to some details displayed on the screen. It is useful to identify ways in which improving the use of laser pointers. For instance, employing laser pointer as an interactive tool to do some operations such as open a new application, zooming, and panning images. This idea could be achieved if the laser pointer is somehow used as a device that would be able to go like the mouse on the screen. Several methods have been proposed to interact with large displays using laser pointers. For example, circling gesture [8] and dwell time [9], these methods cover differentfields such as Windows operations [10]. However, it is still difficult to use the laser pointer as an efficient input device because of problems such as the hand jitter problem, detection errors, and latency.

In this paper, we attempt to use the laser pointer in a medical application that can be useful in operating rooms. Hepatic angiography is a study of an X-ray of the blood vessels that supply the liver. The procedure uses a thin and flexible catheter that is placed into a blood vessel through a small cut. A trained doctor called an Interventional radiologist usually performs the procedure. The proposed work is part of a project entitled IAngio, this project is a research activity that aims to develop a three-dimensional (3D) imaging system of hepatic angiography.

In the OR, surgeons can not use devices like mice and keyboards. In such situation, the surgeon usually asks another staff member to help him to interact with the display (e.g. Zoom in/out images). The latency between surgeon decisions and actions to be performed by the other member motivated our work. As a substitute, laser pointer technique is proposed in this paper for this particular problem. An interactive system is presented to help surgeons in the OR to control the reconstructed 3D image on the LCD using laser pointer gestures, which can be executed at a distance from the screen.

The system only uses an inexpensive laser pointer with an on/off button and a web camera to capture the movement of the laser spot on the screen. The main contribution here is to develop an algorithm that has the ability to understand the laser pointer gestures as commands to interact with the objects shown on the screen depending on the presence and the absence of the laser spot, and explore different application that can be used with those gestures.

The rest of the paper is arranged as follows: Section 2 summarizes the related work, Section 3 describes the system overview of our research and illustrates the proposed approach. Section 4 and Section 5 present the experiment and the results.

II. Related Work

Interaction with large displays falls mainly under two categories: remote and up-close interaction. The classic problem with up-close interactions is that users are required to walk up to the screen to interact with objects that are within arms reach. Anastasia et al. [11] proposed the Vacuum system, as a potential solution to the problem of access to remote areas of the display that are difficult to reach. It brings remote objects closer to the user for viewing and manipulation. However, when users step back from the display to view the contents of the entire screen, they can no longer interact with the objects until they step forward to touch the screen.

One of the solutions is to use remote interactive techniques that allow users to interact with large displays from a distance. Instead of learning completely different new methods to interact using new scenarios, it would be better to adopt the natural channels of communication that are familiar to people in their daily life, e.g. gesture, eye-gaze, speech, and physical tools [12, 13, 14, 15]. One way explored is through the use of laser pointer [16] as an input device. As it allows direct interactions, and the users can perform all the operations by small wrist motion [17, 18]. However, all of these systems adopted a compatible mouse interface. Where, the user holds the laser pointer with a fixed posture for a certain amount of time on a specific spot on the display to click anything. This technique is called the "holding technique".

Myers [19] showed that it is difficult for a user to keep a laser pointer fixed on a specific area of a screen for a long time as the beam is unsteady, and users cannot turn the device on or off at will. Shizuki et al. [20] attempted to move away from this method where the proposed system depends on the crossing technique where the users execute commands through using the four peripheral screen areas. Hisamatsu et al. [21] also combined the crossing technique with another one that's based on moving gesture, i.e. the user can draw circles by laser pointer to select objects. Bi et al. [22] used laser pointers with buttons as interaction devices for one or more displays. Such systems are more suitable for multi-user interactions. However, the additional electronics requirements reduce the advantage of laser pointers: low cost and ubiquity.

One of the important issues in the laser pointer system is its ability to detect the laser spot. Threshold method is considered one of the popular ways in the detection process. It attempts to extract the object of interest by finding the brightest spot in the image [23]. However, Shin et al. [24] claimed that: 1) high contrast level may cause the laser spot to appear as a glow, which prevents the correct detection. 2) A change in the environment lighting may cause the algorithm to misinterpret the spot position, and 3) laser spot may not be the brightest spot on the display due to lighting effects or the existence of other bright colors. Sugawara and Miki [25] used a special laser pointer called WDM with a visible laser beam and an invisible infrared laser beam. Then, they attached a band pass filter to the camera so that only infrared wavelength can pass through. A hybrid technique [26] may also be used to solve the problem of the threshold method, but the algorithm may be too slow for a real-time application.

Background subtraction is another simple and popular method for laser spot detection [27]. One of the disadvantages of this method is that any foreground object remains static for a period will disappear as the algorithm will misunderstand it as a background object.

In this paper, two methods are evaluated to select the one that is well suited to our system for the laser detection process: thresholding and background subtraction. The proposed system is based on several steps: laser spot detection, tracking laser spot, and laser gesture recognition that will be described in detail in the following section.

III. System Overview

The proposed system is composed of a large screen, a laser pointer and a laptop PC connected to a web camera to detect the laser pointer position (see Figure 1(a)). For the camera, the system requires a low-cost web camera yet adequate for initial tests, it can deliver up to 30 frames per second. The camera is used to capture the image of the screen and then in this captured image, the system detects the laser pointer and identifies its corresponding position on the screen. The image is shown over the large display (see Figure 1(c)), and the surgeon controls the reconstructed 3D image through performing laser gestures for zooming, rotating, and flipping (see Figure 1(b)).



b) Laser gestures to browse medical images of the patient (c) The reconstructed 3D image

Figure 1: Laser pointer control system.

3.1 System Details

In this section, the proposed system details are discussed that endow computers with the ability to understand the context in which laser gestures are made. After the image has been captured by the web camera, it is considered as the input of the laser detection algorithm. Laser spot detection step is performed to identify which one of the detected foregrounds is the actual laser spot. First, thresholding process is performed to separate the image into a foreground object and background. Every frame is converted from RGB to HSV color space, where the color of interest (i.e. red or green color) is filtered between a minimum and maximum threshold values as seen in Figure 2. However, the threshold approach failed to detect the laser spot with other bright colors (see Figure 3(a)). Second, the background subtraction method is performed where a frame is compared to another frame in which a significant difference between those frames is identified as the foreground object. This method achieves good results in distinguishing the laser spot as illustrated in Figure 3(b). After testing the two approaches, the background subtraction is perfectly suited to the proposed system based on the assumption that the background of our environment is entirely static.



Figure 2: Laser spot detection.

Second, some morphological operations are performed such as "Erode" and "Dilate" to remove noise and improve the appearance of the laser spot. After the algorithm detects the presence of the laser spot on the image, it returns its position corresponding to the screen grid. The moments method used to calculate the position of the center of the laser spot. The first order spatial moments calculated around x-axis and y-axis and the zero order central moments of the binary image. Where, the zero order central moments of the binary image are equal to the white area of the image in pixels (i.e. the noise of the binary image also should be at the minimum level to get accurate results). For high-quality results, there are few constraints in the system to be setup. The proposed system uses the following basic assumptions:

- 1. The web camera is supposed to be fixed and motionless.
- 2. The brightness and exposure settings of the camera are reduced in order to improve the detectability of . the laser spot



Figure 3: The two approaches with other bright colors.

One of the challenges that is considered to be a part of the laser gesture recognition system is about identifying the start and the end points of the intended gesture. As the user moves from one laser gesture to another, he performs several unintended movements with his hand that links the two consecutive intended gestures. The system may attempt to recognize this inescapable intermediate move as a meaningful one. To solve this problem, each gesture is recognized with the laser-on and laser-of events. The start point is recognized when the laser button is pressed. In contrast, the end point is recognized when the user releases the laser pointer button. However, the user may attempt to perform laser on/off events only to point to some details on the screen, and he/she does not mean to interact with the large display.

To differentiate between intended and unintended gestures and based on our user analysis empirical study, we assume that every intended laser gesture takes about 1-2 seconds to be performed. Thus, any gesture exceeds that time is considered as an unintended gesture.

Finally, to recognize the laser gestures, two different algorithms are tested in order to choose the best algorithm for the proposed system. The first algorithm is called dynamic time warping (DTW) algorithm. The DTW is a time series alignment algorithm developed basically for speech recognition. It aligns two sequences of feature vectors by warping the time axis iteratively until an optimal match between the two sequences is found. The second one is called 1\$ recognizer algorithm. The 1\$ uni-stroke recognizer is a 2-D single-stroke recognizer designed for rapid prototyping of gesture-based user interfaces. It is an instance-based nearest-neighbor classifier with the Euclidean scoring function, i.e., a geometric template matcher.

In the recognition process, the system receives a continuous stream of coordinates, the trajectory of each intended laser gesture is recorded. The data that is collected from the web camera is used to recognize gestures such as the circle in the two rotational directions and the line with the four directions: up, down, left, and right. Every gesture will then be compared with a set of predefined training data, or prototypes (in the form of templates) to recognize which gesture is being signed. A label corresponding to the recognized gesture will be displayed on the large display.

IV. Experiment

In order to understand how laser pointer performs in interventional procedures, the experimental evaluation has been performed. The main research questions are:

- 1. Whether the laser pointer system is a potential substitute for the conventional input system in ORs.
- 2. Whether the laser pointer system can facilitate the surgeons in dealing with the 3D models that is shown on the large display.

Thus, our goal of this experiment was to measure the total time, including error correction, that is spent on executing the interactions with the large display using the laser pointer compared to other techniques that are executed in the OR in a conventional way, i.e. the surgeon asks another staff member to help him to interact with the display using mouse and keyboard. Second, measuring the accuracy of recognizing proposed gestures for controlling the 3D model. The classification accuracy is evaluated in the experiment between two algorithms: DTW and 1\$ recognizer.

4.1 Experiment Setup

The laboratory setup composed of a USB camera with 30 fps and resolution of 640X480. A green laser pointer (wavelength 532 nm, output power about 500 mW) is used, and the data is collected on a laptop PC running Microsoft Windows 7 with 2.4 GHz processor and 8 GB RAM. The software was written in OpenCV C++. The experiment was performed by 15 subjects, there were 5 males and 10 females aged between 19 and 25. All participants were frequent users of mouse-and-windows based systems, but none of them had any previous experience with interactions through laser pointers. The users were asked to stand close to the large display at a distance of 1.5 m (see Figure 5).

Each user is asked to use three input ways: the traditional way, i.e. input with a mouse and keyboard, and the laser pointer per each classifier (DTW and 1\$ recognizer) in order to complete two tasks as shown in Figure 4:

Task 1: Rotating the 3D model over X and Y axis using four gestures: right-line, left-line, up-line, and down-line.

Task 2: Zooming in/out the 3D model using two gestures: circle-clockwise and circle-counterclockwise.



Figure 4: Laser command mappings.

Users have a single training session before performing the experiment to familiarize themselves with the laser pointer technique. Each user was briefed on the goals of the experiment, and the system was demonstrated by the experimenter. Then, each of them should complete 4 sessions per each technique. Each session consisted of all six commands and has a different order for the performing gestures.



Figure 5: Laboratory experiment using the traditional method and the laser pointer.

V. Results and Discussion

Experimental data shows that mouse-based interactions are faster than the laser pointer either with the DTW or the 1\$ (Figure 6). However, the average time cost to complete the total task with the mouse and the keyboard is just two seconds faster than that with a laser pointer using the DTW recognizer. Considering that, the latency problem or the slowness of the spot recognition can be solved by using better and faster cameras connections, we are satisfied with the performance of the laser pointer with DTW recognizer relative to the traditional method.

Laser pointer interaction for 3D angiography operation



Figure 6: Total time cost for the three input ways including error correction.

The averaging accuracy of DTW is %89.6, it means that the system misinterprets the laser pointer gestures at a rate of %10. Generally, it is noticed that there are some obvious sources of error in the DTW system. For instance, some circle gestures and line gestures (i.e. usually left-line and right-line) are misclassified as down-line or up-line. The expected reason is that some users perform lines with a slightly skewness down or up. Similarly, circles with a very small diameter also misunderstood by the system. In the experiment of the DTW, we observed the hand jitter problem but, it was not a major problem as reported in existing literature [9]. It did not have a significant effect in our experiment because users were not asked to hold the laser steady for a long time. That is, drawing circles or lines is relatively easier than keeping a laser spot fixed on a specified location for a period of time.



Figure 7: Time cost in seconds.

As shown in Figures 7 and 8, the laser pointer with the 1\$ recognizer including error correction is almost twice as slow as the mouse and the averaging accuracy of 1\$ is %75. The algorithm performs well and achieves good results in differentiating between the line shape and the circle shape, but it also has some limitations. The 1\$ can not distinguish gestures whose identities depend on aspect ratios, locations, or specific orientations. This means that separating up-lines from down-lines or left-lines from right-lines is not possible without modifying the conventional algorithm. This explains why the 1\$ comes last in the total time results. When the users make mistakes, they need extra time to correct their errors as we give them two trial for error correction.

Figure 9 displays the number of errors for all participants in each session for both algorithms. It is observed that the errors in DTW in session 3 and 4 are much less than those in session 1 and 2 which indicates that a user can be expected to improve in the usage of the laser pointer with the DTW. However, we can not expect that in the 1\$ classifier. There was no discernible difference in user performance between any of the four sessions. After calculating the number of errors for all participants in each session, the largest number occurs during session 2.

The obtained data from participants gestures are then analyzed using an ANOVA test. First, a one-way ANOVA test is performed in regards to the time of the total task for the three input ways. The test indicates that there is a significant difference between the three algorithms (p<0.05). However, the ANOVA analysis only indicates that if there is a significant difference between at least one pair of the group means. It does not indicate what the pair or pairs are significantly different. To find which method is of better performance, a Tukey HSD test needs to be performed. A Tukey test is interested in examining mean differences where any two means that are greater than HSD are significantly different. From ANOVA results, the mean values of the three input ways: mouse,1\$, and DTW are M1, M2, and M3, respectively.

A post-hoc analysis using a Tukeys HSD testshowed that HSD=13. 18, with M1 - M2 =109.36, M2 - M3 =108.43, and M3 - M1 =0.93. Thus, the total time of the 1 recognizer was significantly larger than those with the other input methods (i.e. no difference was found between the traditional method using the mouse and the laser pointer using DTW recognizer).



Figure 8: Accuracy of laser pointer with two algorithms.

Second, another one-way ANOVA test is performed in regards to recognition rate (accuracy) of the laser pointer with the two algorithms: DTW and 1\$. The test indicates that there is a significant difference between the two algorithms (p<0.05). Thus, participants within the DTW method group generated significantly more accurate gestures than the 1\$ recognizer.



Figure 9: Number of errors for all participants in each session.

VI. Conclusion

A real-time system is introduced to track and recognize laser pointer gestures for controlling the reconstructed 3D image on the large display using preset laser pointer gestures, which can be executed at a distance from the display. The system includes three fundamental steps: laser spot segmentation, laser spot tracking and gesture recognition. The conclusion that can be drawn from these tests is that mouse-based interactions are clearly faster than the laser pointer with 1\$ recognizer that shows poor performance with the line gestures and overall accuracy of %75. The laser pointer with DTW performs about the same as the mouse in the

completion time with an overall accuracy of %89.6. Considering that, better cameras can improve the total time that is spent on executing the interactions with the large display, the laser pointer system could be a potential substitute for the conventional input system in ORs, and the surgeons can perform interactions without needing help from another staff member.

Future work includes firstly testing the proposed system in a real-world scenario such as the OR. However, the OR is a very critical place so the accuracy of the laser pointer with DTW should increase. Secondly, we plan to extend this work to support multiple users with personalized service. Also, we hope to find out whether other techniques can be useful in ORs, replacing the laser pointing or augmenting it.

References

- P. Dias, J. Parracho, J. Cardoso, B. Q. Ferreira, C. Ferreira, and B. S. Santos, "Developing and evaluating two gestural-based virtual environment navigation methods for large displays," in Distributed, Ambient, and Pervasive Interactions, pp. 141–151, Springer, 2015.
- [2] X. Bi, S.-H. Bae, and R. Balakrishnan, "Walltop: Managing overflowing windows on a large display," Human{Computer Interaction, vol. 29, no. 2, pp. 153–203, 2014.
- [3] K. Cheng, J. Li, and C. M^{*}uller-Tomfelde, "Supporting interaction and collaboration on large displays using tablet devices," in Proceedings of the International Working Conference onAdvanced Visual Interfaces, pp. 774–775, ACM, 2012.
- [4] C. Ardito, P. Buono, M. F. Costabile, and G. Desolda, "Interaction with large displays: A survey," ACM Computing Surveys (CSUR), vol. 47, no. 3, p. 46, 2015.
- [5] L. Hespanhol, M. Tomitsch, K. Grace, A. Collins, and J. Kay, "Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays," in Proceedingsof the 2012 International Symposium on Pervasive Displays, p. 6, ACM, 2012.
- [6] S. H. Hirano, M. T. Yeganyan, G. Marcu, D. H. Nguyen, L. A. Boyd, and G. R. Hayes, "vsked: evaluation of a system to support classroom activities for children with autism," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1633–1642, ACM, 2010.
- [7] G. Robertson, M. Czerwinski, P. Baudisch, B. Meyers, D. Robbins, G. Smith, and D. Tan, "The large-display user experience," Computer Graphics and Applications, IEEE, pp. 44–51, 2005.
- [8] K. Cheng and K. Pulo, "Direct interaction with large-scale display systems using infrared laser tracking devices," in Proceedings of the Asia-Paci c symposium on Information visualisation-Volume 24, pp. 67–74, Australian Computer Society, Inc., 2003.
- D. R. Olsen Jr and T. Nielsen, "Laser pointer interaction," in Proceedings of the SIGCHIconference on Human factors in computing systems, pp. 17–22, ACM, 2001.
- [10] P. Dey, A. Paul, D. Saha, S. Mukherjee, and A. Nath, "Laser beam operated windows op-eration," in Communication Systems and Network Technologies (CSNT), 2012 InternationalConference on, pp. 594–599, IEEE, 2012.
- [11] A. Bezerianos and R. Balakrishnan, "The vacuum: facilitating the manipulation of distant objects," in Proceedings of the SIGCHI conference on Human factors in computing systems, 361–370, ACM, 2005.
- [12] L. Stifelman, A. Elman, and A. Sullivan, "Designing natural speech interactions for the living room," in CHI'13 Extended Abstracts on Human Factors in Computing Systems, pp. 1215–1220, ACM, 2013.
- [13] M. Vidal, K. Pfeuffer, A. Bulling, and H. W. Gellersen, "Pursuits: eye-based interaction with moving targets," in CHI'13 Extended Abstracts on Human Factors in Computing Systems, 3147–3150, ACM, 2013.
- [14] L. Hespanhol, M. Tomitsch, K. Grace, A. Collins, and J. Kay, "Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays," in Proceedingsof the 2012 International Symposium on Pervasive Displays, p. 6, ACM, 2012.
- [15] J. Y. Lee, M. S. Kim, J. S. Kim, and S. M. Lee, "Tangible user interface of digital products in multi-displays," The International Journal of Advanced Manufacturing Technology, vol. 59, no. 9-12, pp. 1245–1259, 2012.
- [16] R. Jota, M. A. Nacenta, J. A. Jorge, S. Carpendale, and S. Greenberg, "A comparison of ray pointing techniques for very large displays," in Proceedings of Graphics Interface 2010, 269–276, Canadian Information Processing Society, 2010.
- [17] J.-Y. Oh and W. Stuerzlinger, "Laser pointers as collaborative pointing devices," in GraphicsInterface, vol. 2002, pp. 141–149, Citeseer, 2002.
- [18] R. Sukthankar, R. G. Stockton, and M. D. Mullin, "Smarter presentations: Exploiting ho-mography in camera-projector systems," in Computer Vision, 2001. ICCV 2001. Proceedings.Eighth IEEE International Conference on, vol. 1, pp. 247–253, IEEE, 2001.
- [19] B. A. Myers, R. Bhatnagar, J. Nichols, C. H. Peck, D. Kong, R. Miller, and A. C. Long, "Interacting at a distance: measuring the performance of laser pointers and other devices," in Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 33–40, ACM, 2002.
- [20] B. Shizuki, T. Hisamatsu, S. Takahashi, and J. Tanaka, "Laser pointer interaction techniques using peripheral areas of screens," in Proceedings of the working conference on Advancedvisual interfaces, pp. 95–98, ACM, 2006.
- [21] T. Hisamatsu, B. Shizuki, S. Takahashi, and J. Tanaka, "A novel click-free interaction tech-nique for large-screen interfaces," APCHI2006, 2006.
- [22] X. Bi, Y. Shi, and X. Chen, "Upen: a smart pen-liked device for facilitating interaction on large displays," in Horizontal Interactive Human-Computer Systems, 2006. TableTop 2006.First IEEE International Workshop on, pp. 7–pp, IEEE, 2006.
- [23] C.-S. Wang, L.-P. Hung, S.-Y. Peng, and L.-C. Cheng, "A laser point interaction system in-tegrating mouse functions," World Academy of Science, Engineering and Technology, vol. 67, pp. 657–662, 2010.
- [24] J. Shin, S. Kim, and S. Yi, "Development of multi-functional laser pointer mouse through im-age processing," in Multimedia, Computer Graphics and Broadcasting, pp. 290–298, Springer, 2011.
- [25] S. Sugawara and T. Miki, "A remote lecture system with laser pointer for the internet and broadband networks," in Applications and the Internet Workshops, 2004. SAINT 2004 Work-shops. 2004 International Symposium on, pp. 61–66, IEEE, 2004.
- [26] F. Ch'avez, F. Fern'andez, R. Alcal'a, J. Alcal'-Fdez, G. Olague, and F. Herrera, "Hybrid laser pointer detection algorithm based on template matching and fuzzy rule-based systems for domotic control in real home environments," Applied Intelligence, vol. 36, no. 2, pp. 407–423, 2012.
- [27] O. Barnich and M. Van Droogenbroeck, "Vibe: A universal background subtraction algorithm for video sequences," Image Processing, IEEE Transactions on, vol. 20, no. 6, pp. 1709–1724, 2011.