A Usability Evaluation of Two Computer Vision-Based Selection Techniques

Jacob Eisenstein JACOBE@CSAIL.MIT.EDU
MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge MA, 02139 USA

1. Introduction

Conventional human-computer interaction techniques have been optimized for one specific usage case: a stationary individual whose full attention is devoted to the computer. The standard input hardware – a mouse and keyboard – are well-suited for this kind of interaction, but they function poorly in other usage cases. Desktop computers do not mesh well with other activities, such as (non-virtual) socialization, performing household chores, or watching TV (Norman, 1999).

Ongoing research in *Communication Appliances* focuses on designing technology for remote communication that can be seamlessly integrated into other daily activities. An interaction scheme that treats a computer as a workstation is not likely to be successful in this domain. New interaction techniques must be developed and evaluated.

The problem of selecting from a menu of choices is ubiquitous, not only in desktop computer user interfaces, but also in appliances such as microwave ovens, stereos, and thermostats. Such household items typically provide a custom interface of physical buttons, either on the device itself or on a remote control. But anecdotal experience tells us that as the functionality of these devices becomes more complex, such physical interfaces become increasingly difficult to use (Nielsen, 2004).

Computer-vision based interaction has been proposed as a solution (Freeman et al., 2000). Using a camera to track either an object or the user's body, there is no need to type or use a mouse; the user need not even approach the computer at all. At the same time, there is no complicated hardware interface to learn, and no remote control to find buried under the couch cushions. While vision-based interaction offers promising solutions for this usage case, our understanding of the usability principles of such interfaces is still at an early stage. This paper describes an empirical evaluation of two different computer vision-based interaction techniques for the problem of making a selection from a menu.

2. Selection Techniques

Two selection techniques have been implemented, using the OpenCV computer vision library. The first – *motion selectors* – is modeled after the Sony EyeToy. The user triggers a selector by waving a hand or otherwise creating motion within the selector. A meter increases with the amount of motion detected, and the selector is triggered when a threshold is reached. While the Sony EyeToy uses simple image differencing to detect motion, the system implemented for this experiment uses optical flow detection, which was thought to be more accurate.

Figure 1 demonstrates the motion selectors. This set of images are taken from the video that the participant himself actually observed while performing this experiment. As discussed above, the visual feedback of the system state is considered an important part of this interaction technique. In part (a), the desired target selector flashes, indicating the participant should select it. In part (b), the participant moves his hand towards the selector. Once the participant has reached the selector, he moves his hand within it (part (c)), and the activation level rises. When the activation level reaches the top in part (d), the selector is activated, and flashes to indicate this.

The second selection technique – *tracking selectors* – requires that the user place a real color ball within the selector. The ball is tracked using the camshift algorithm,with the backprojection computed by a set of histograms at several levels of precision in the YCrCb color space.

The system's guess of the location of the tracked object is indicated by an ellipse. The tracker is enable to assess its level of confidence in its own estimate, using goodness-of-fit metrics based on the size, shape, and color of the estimated location and boundary of the tracked object. A high level of certainty is indicated to the user by coloring the ellipse green; a low level is indicated by coloring the ellipse yellow.

Figure 2 shows several images of the tracker selectors, again taken directly from the video feed seen by the participant. In part (a), the system is not yet sure of the location of the tracked ball, as indicated by the large ellipse, which is colored yellow. As the participant moves towards

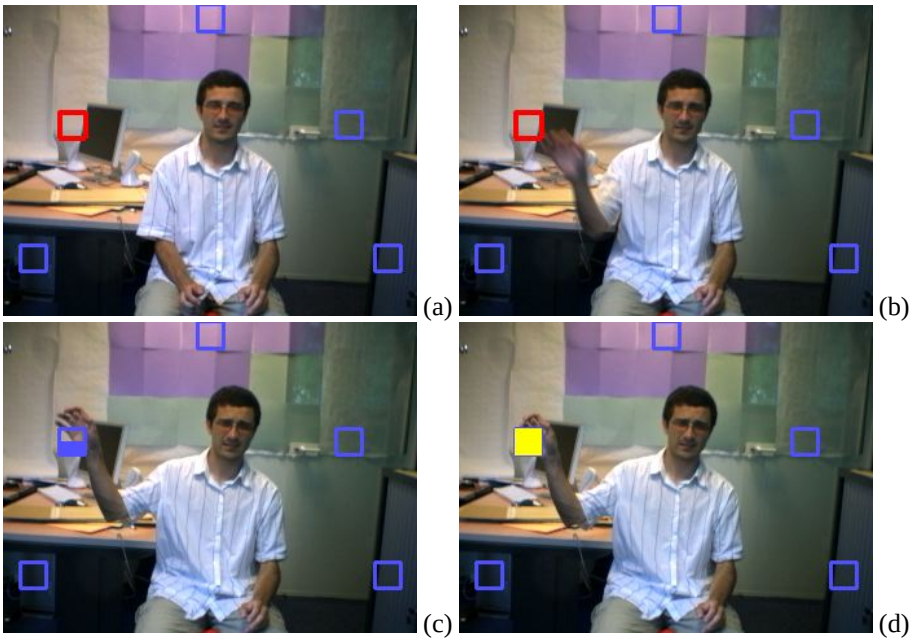

Figure 1. Motion selectors

the target, the tracker focuses in on the location, and ellipse becomes green (part (b)). In part (c), the system momentarily becomes confused, possibly because the tracked object is being moved very quickly, and the ellipse again turns yellow to indicate this. Without being instructed to do so, the user holds the ball still for an instant, enabling the system to recover. The selector is triggered instantaneously when the user moves the tracked object into it in part (d).

3. Experiment

An experiment was conducted to compare the speed, accuracy, and likability of the two techniques. Twelve people participated in this study, and none was previously familiar with either selection technique.

Participants were told that the goal of the experiment was to see how fast they could select a targeted button. The targeted button was indicated by making it blink. Participants were told to go as fast as possible, as long as errors were kept within reason. Participants underwent an initial training period in which they were allowed as many trials as they felt they needed to learn how to use each selection technique.

For both selection techniques, the buttons were laid out in an semi-ellipse, such that no button would come closer than 15 pixels from the edge of the screen. The entire experimental procedure was automated. Participants were instructed to keep their hands in their laps until one of the buttons started blinking; then they were to activate that button as quickly as possible. Afterwards, they were to return their hands to their lap. The system also enforced rest periods to prevent fatigue.

After this initial training period, six experimental blocks were conducted. Each block consisted of twenty trials of a single selection technique. Alternating blocks of each interaction technique were used, and the ordering was counterbalanced across participants. Within each block, the number of selectors was varied between 2,5,11, and 21; an equal number of each type was given within each block, and the order was determined randomly. Similarly, the location of the target button was varied, and each participant experienced an equal distribution across the semi-ellipse of buttons. The following results were logged: intended target, actual selection, and elapsed time for the trial.

4. Results

A two-factor within-participant analysis was used to analyze the effects of the selection technique and the number of selectors on two dependent variables: error rate, and selection time. Results are reported as statistically significant when $p < .05$.

The choice between motion selectors and tracking selectors did not significantly impact on the error rate, but the number of selectors did. The interaction between the number of selectors and the selection technique was also observed to have a statistically significant effect on error rate. These results are described in Table 1.

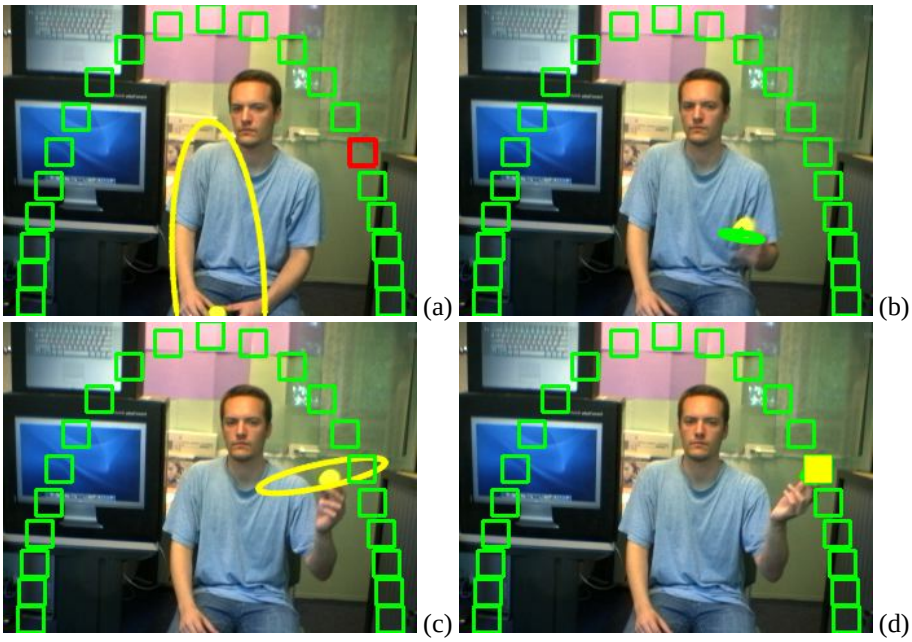

Figure 2. Tracking selectors

num. selectors →	2	5	11	21
motion selectors	0	0	7	37
tracking selectors	0	0	1	20

Table 1. distribution of errors

	tracking	motion	no pref
Which is faster?	10	1	1
Which is more accurate?	9	3	0
Which did you prefer?	6	5	1

Table 2. qualitative results

Users were significantly faster when using the tracking selectors, compared to the motion selectors. The average selection time was 1.94 seconds with the motion selectors, compared to 1.63 seconds with the tracking selectors. Although the average selection time increased monotonically with the number of selectors, this effect was not found to be significant. There were also no significant interactions between the two factors.

As described in Table 2, most of the participants believed the tracking selectors were faster and more accurate, but were more evenly divided as to which method they preferred.

5. Discussion

The tracking-based selection technique was not observed to be worse on any measure than the motion-based selection technique. It was observed to be faster, and to produce fewer errors when the number of selectors was very large.

As suggested by the qualitative results, speed and accuracy are not the only considerations that shape the participants' preferences. Both techniques have strengths and weaknesses which are more difficult to quantify in a controlled experiment. The tracking-based technique is robust to background movement, but was found to be not particularly robust to lighting changes. The motion-based technique does not require to user to keep a tracked object, but we had to go to great lengths to eliminate background motion that might confuse the system. A longer-term longitudinal study of how each system is actually used is necessary to assess the role of these factors.

6. Acknowledgements

This research was performed under the supervision of Wendy Mackay at the In Situ research group of INRIA.

References

Freeman, W. T., Beardsley, P. A., Kage, H., Tanaka, K.-I., Kyuma, K., & Weissman, C. D. (2000). Computer vision for computer interaction. *SIGGRAPH Comput. Graph.*, *33*, 65–68.

Nielsen, J. (2004). Remote control anarchy. http://www.useit.com/alertbox/20040607.html.

Norman, D. (1999). *The invisible computer*. The MIT Press.