Usability Evaluation Framework for Computer Vision Based Interfaces

Muhammad Raza Ali and Tim Morris

***Abstract*—**Human computer interaction has progressed considerably from the traditional modes of interaction. Vision based interfaces are a revolutionary technology, allowing interaction through human actions, gestures. Researchers have developed numerous accurate techniques, however, with an exception to few these techniques are not evaluated using standard HCI techniques. In this paper we present a comprehensive framework to address this issue. Our evaluation of a computer vision application shows that in addition to the accuracy, it is vital to address human factors

***Keywords*—** Usability evaluation, cognitive walkthrough, think aloud, gesture recognition.

I. INTRODUCTION

Human computer interaction is one of the most diverse and dynamic disciplines in computer science. The principles, guidelines and techniques of human computer interaction are used to evaluate and improve applications, websites and computerized services.

The discipline has evolved rapidly over the past two or three decades, mainly due to the ever increasing range of applications. The computers have become an integral part of all aspects of our lives. The interaction between human and computer is longer limited to keyboard and mouse. The novel interaction techniques have emerged and these techniques can be divided into two major categories:

1. Haptic interfaces: These are advanced touch based interfaces [1] that are now used extensively in smart phones and other handheld devices e.g. iPhone.

2. Vision based interfaces: Another area in interactive technologies is computer vision based interfaces. These interfaces allow human-machine interaction through actions, gestures or even expressions. The research covers application areas involving single [2] and multiple users [3] .In this paper our discussion focuses on vision based gestural interfaces.

A. Research in Gesture Recognition

Among computer vision researchers, recognition of gestures has been a key area of interest [4], due to its potential scope of application in human computer interaction it has also generated interested in the HCI community[5].However, researchers have been primarily focussed on aspects like improvement in the accuracy of gesture recognition algorithms, developing better imaging hardware, addressing key challenges etc. As a result we now see numerous gesture based systems with high reported accuracy covering a decent range of applications from gaming [6] and education [7] to surveillance [8]. However, very few computer vision systems have been evaluated using standard HCI techniques. This is important as these systems are aimed at human users. Some researchers [8, 9] have pointed out this issue but have not proposed a framework that can serve in developing standardized framework for evaluating a vision based interface from HCI point of view. In this paper, we present a framework, based on standard HCI techniques for evaluating a vision based interfaces. This is the first time such a framework is presented and we use this framework to evaluate our vision based gestural application. The motivation for this work comes from the work done in evaluating handheld devices. Not many users realize that efficient, robust and responsive handheld interfaces have been developed after numerous elaborate and tedious usability studies [10], not just superior hardware components. The main purpose of the paper is to highlight the importance of HCI based evaluation techniques and encourage computer vision researchers and developers to employ these techniques. This is vital if vision based interfaces are to become a more established mode of computer interaction.

II. OVERVIEW OF THE PAPER

The paper is divided into following sections:

Usability Evaluation Framework for Gestural Interfaces (Section III): This section discusses the proposed usability evaluation framework for gestural interfaces. We describe various components of the framework and provide rationale for the inclusion of selected techniques.

Evaluation of the Gestural Interface (Section IV): In this section we provide an overview of our application. We also discuss the methodology for evaluating the interface using the framework proposed in section III.

Results (Section V): The section discusses results of our usability evaluation studies and important lessons that can be learned.

M.R. Ali is a PhD candidate with the Advanced Interfaces Group, School of Computer Science, University of Manchester, UK. e-mail: alim@ cs.man.ac.uk).

T. Morris is a Lecturer in the School of Computer Science, University of Manchester, UK. He is part of the Advanced Interfaces Group e-mail: tmorrris@cs.man.ac.uk).

III. USABILITY EVALUATION FRAMEWORK FOR GESTURAL INTERFACES

The proposed framework comprises of five stages. Before the details are discussed it would be useful to give an overview of the techniques and tools used in the framework. There are two types of testing in the framework i.e. user testing and expert testing. The framework is shown in fig 1.

1. **User Testing:** This types of testing involves human users. We employ a simple yet very technique called *think aloud.* The investigator sets task(s) for the user to perform. The technique as the name suggests, asks the user to describe what they are doing and thinking as they perform the task. The description provided by the user is recorded by the investigator on paper, audio or video. The users' comments reveal interesting points about the evaluated system, interface or website[12]. A variant of technique is often referred to as *concurrent thinking aloud* i.e. task completion is accompanied by user describing the process. There is another version of this technique which is termed as *retrospective think aloud.* In this the user completes the task and after completion provides feedback. We recommend the use of concurrent thinking aloud. The former is best for understanding the user responses at every step of the interaction, while with the latter user may not recall some of things experienced during the task.
2. **Expert Testing:** Some testing techniques do not require the involvement of users. One such technique is called the walkthroughs. These techniques are 'inspection-based methods' where an expert evaluates by setting a series of tasks that the prospective user will perform. We employ a technique called the cognitive walkthrough [11]. This is an extremely useful technique in which for a given task the expert asks a series of questions, through the answers to these questions any usability issues are identified. For our evaluation we employ cognitive walkthrough using a three question format (Sharp et. al.).
   Q1. Will users know what to do?
   Q2. Will users see how to do it?
   Q3. Will users understand from feedback that the action was correct or not?
3. **Subjective Scales:** These scales are used for obtaining user feedback in the form of a questionnaire or a performa. As the name suggests it is aimed at gathering user opinions and understanding the user experience. We rely on a subjective scale called the Likert scale[13].The scale comprises of short statements pertaining to various aspects of the system and each statement accompanied by a scale indicating the user's agreement or disagreement with the statement. For our evaluation we have used ten statements with a 5-point discrete scale. Please refer to the appendix I for the list of statements.
4. **Prototypes:** Prototyping is useful in usability evaluation studies as it helps the developers and designers to identify any flaws that could lead to unexpected or undesirable behaviour when using the actual system. The testing is to be done using two prototypes; lo-fidelity and hi-fidelity prototypes. The lo-fidelity prototype is developed early on in the development process and is very simple and in-expensive to create. The lo-fidelity prototype is generally created from everyday materials like paper, cardboard etc. The hi-fidelity prototype is the fully functional prototype of the application. This is useful in uncovering in addition to design flaws, any functional flaws.

A. The Usability Evaluation Framework

The components\stages of the usability evaluation framework are discussed below:

Stage 1: Develop the lo-fidelity and the hi-fidelity prototypes of the application.

Stage 2.a: Conduct expert and user evaluations.

Stage 2.b: Obtain written user feedback.

This feedback should not be open ended. The best way in our opinion is to use Likert scale. Prepare statements that cover all aspects of the application. The ten statements we presented can be used or adapted for this purpose.

Stage 3. Analysis of the findings from both types of testing.

This involves that first we look for similar trends across all users; this will help in identifying the most obvious design and implementation issues. Less occurring issues are also noted as they might point to a critical design flaw.

Stage 4. Identifying the key issues.

As the complete data is available, findings from both sets of evaluations and subjective feedback are combined. Issues identified are to be prioritized in the order of importance and urgency. It would be better to split them in to categories e.g. design, functional and theoretical (the latter deals with any flaws discovered with computer vision technique).

Stage 5. Refining the prototypes.

Based on the recommendations in stage 4 the design and implementation issues are resolved in the next version of the prototypes. This may require some modification to the original algorithm or simply some design modifications. We recommend that after first evaluation cycle the lo-fidelity prototyping can be discarded, unless there is a major revamp of the application and the new lo-fidelity prototype would be significantly different from the first one. Only this case it can reveal any new useful information.

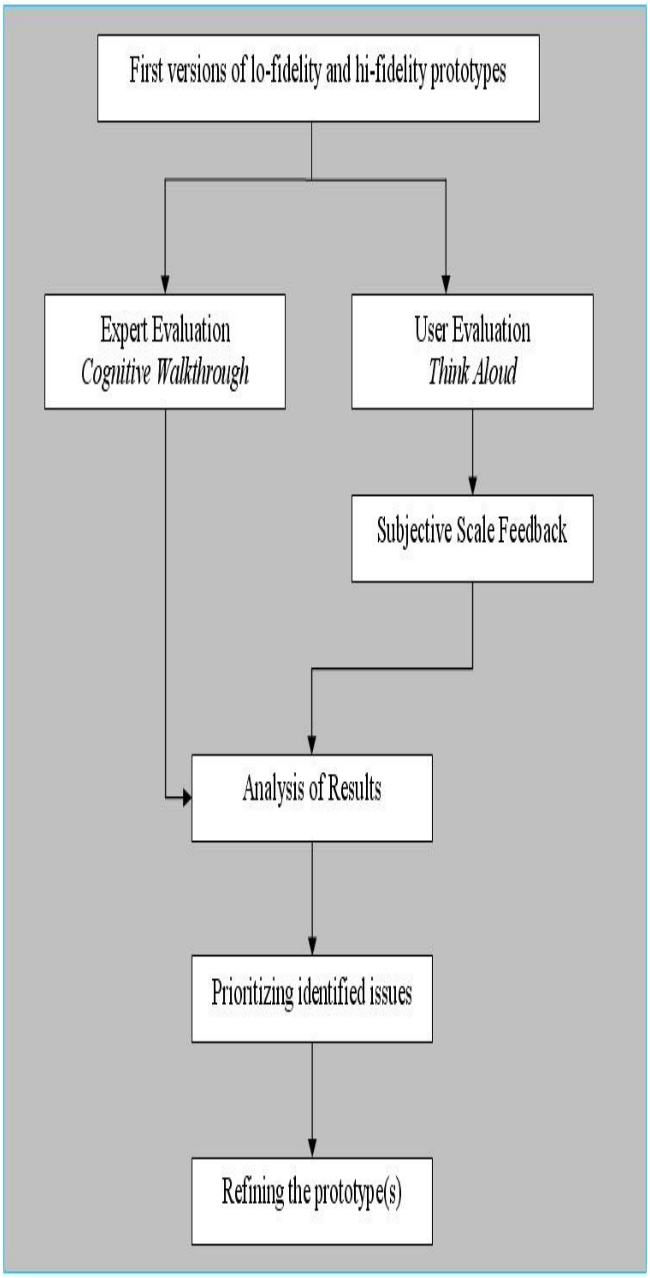

Fig. 1 The usability evaluation framework for compute vision based interactive systems

IV. USABILITY EVALAUTION OF A GESTURE BASED APPLICATION

The evaluation framework presented above is used to evaluate a simple gesture based application we developed. Before discussing the methodology it would be useful to give an overview to application.

A. Overview of the Application

The basic concept of the application is to move smaller on-screen objects (boxes) to appropriately labelled target locations. This all is to be accomplished through hand gestures. The layout of the application area is shown in fig 2.

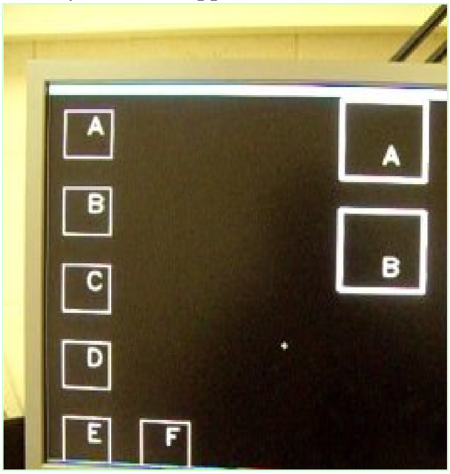

Fig. 2 Application area

The smaller boxes on the left are each selected by a separate gesture and moved using the 'hold' gesture. In total seven gestures are recognized by the application i.e. our gesture vocabulary. The example of valid gestures is shown in fig 3.

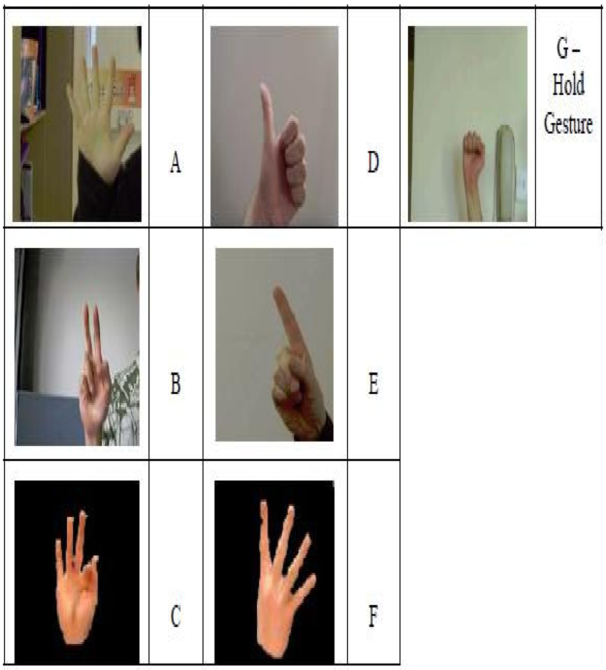

Fig. 3 The seven gesture vocabulary

Traditionally, the computer vision based interfaces are evaluated from the point of view gesture recognition accuracy, real time performance etc. Our proposed computer vision technique gave high recognition rates (> 96%) and worked in real time. The reader can refer to [14] for the details. However, we also evaluated it from HCI point of view using the framework presented above. It is worth mentioning that this is the first time a computer vision application has been evaluated using techniques like think aloud, cognitive walkthrough and prototyping.

B. Evaluation Methodology

For the purpose of evaluation 13 participants were involved in the user testing phase of the study. The participants were postgraduate students; ten from computer science department and one each with business administration, social science and humanities background.

The user testing involved evaluation using the lo-fidelity and hi-fidelity prototypes. The users were asked to perform two tasks on each of both prototypes. First task involved moving two objects to the target location and the second task involved moving three objects. Therefore, 4 tests were accomplished. In lo-fidelity testing the investigator acted on the behalf of the system handling printouts that represent various states of the system. In hi-fidelity testing the participants interacted with the actual application. The process of selecting and moving an object is indicated in the fig 4. Following the think aloud technique the participant was asked to describe what they are doing and thinking, while the investigator made notes. Once both stages are completed the user feedback is obtained based on Likert scale. An expert conducted cognitive walkthrough of both lo-fidelity and hi-fidelity prototypes

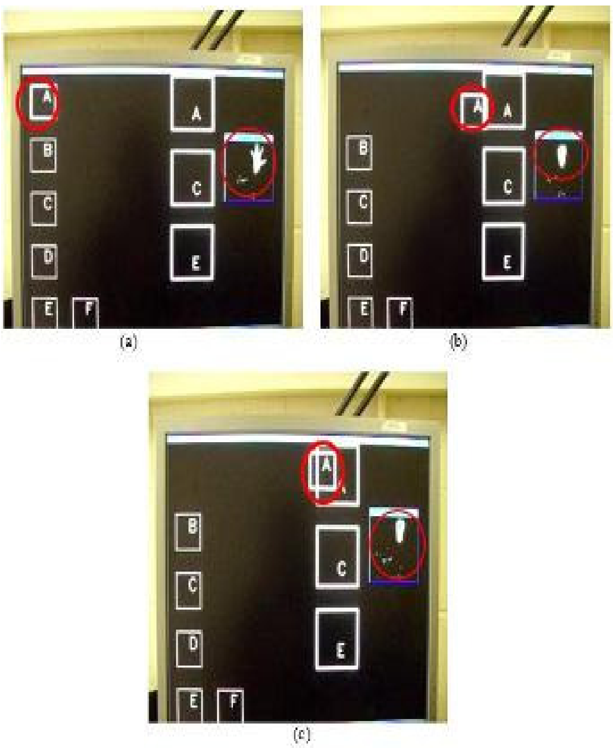

Fig. 4 Interaction with a hi-fidelity prototype, a) Gesture A is made to select object A. Selected object is highlighted, b) Hold Gesture is used to move selected object towards appropriate target, c) The selected object is placed inside the target

V. RESULTS AND DISCUSSIONS

The usability evaluation of the application revealed some interesting points regarding the application. The findings that are discussed below are based on the data acquired through both types of testing and subjective scale feedback. The findings provide the rationale for using a proper evaluation framework and making the HCI techniques an integral part of development process of computer vision based interfaces.

1. One of the important characteristics of a user interface is the *feedback* of the system [15]. The user must understand the state of the system. In order to indicate the selection of an object, the selected object's borders are highlighted. We as developers presumed that it would provide adequate feedback. However, some users did not realize the selection has been made due to the white color (see fig 4). This could be resolved by using different color to indicate emboldened borders.
2. The developers and researchers need to ensure that potentially confusing gestures should be avoided. From the computer vision perspective the ability of the algorithm to distinguish between confusing gestures is desirable, but it might not be advisable from the usability point of view. For example in our gesture vocabulary some users had to refer to crib sheet as they confused gestures 'B' and 'C'.
3. Another important aspect is to minimize the learning curve for a computer vision interface. The mechanism of interaction need to be consistent with the mental model of the user. In our application, once the selected object reaches the target the user was to make the next gesture to select the next object i.e. there is no explicit release gesture. This confused some users as they expected an explicit release gesture. This is an important observation as it reveals that the release or drop gesture is consistent with the user's conceptual model of the 'drag and drop' operation.
4. An encouraging aspect of the evaluation was the feedback received regarding the concept of using vision technologies for human computer interaction. The participants used the words like 'intuitive', 'interesting' to describe the concept. Although, some users remarked that they would require more practice to get used to it. Only one user among all participants said that he would not like to use hand gestures for manipulation of on-screen objects.
5. In our opinion the most important findings are the observations related to the 'hold' gesture. These observations further stress that development focus in computer vision, similar to other disciplines in computer science; need to move towards the user [16, 17]. In other words, developers and researchers cannot assume anything from the user of point of view. The reasons the hold gesture was included are:
   - The system consistently correctly recognized this gesture. The 'hold' gesture consistently showed high recognition rates (> 97%). This accuracy was observed for both image datasets

and real time data.

- The gesture is simple and intuitive, as established during the evaluation.

Despite these important characteristics the hold gesture did not provide the ease of manipulation as expected. Important observations and our analyses are presented below:

- Some of the users assumed more flexibility with the hold gesture and understood that the system will be able to recognize the gesture at any angle. This assumption made the manipulation error prone and tedious. This is interesting, and the developers and researchers need to remember that an apparent simplicity of a gesture will require greater flexibility in recognition. We addressed this issue by including more examples to train our recognition system. This issue was only discovered through user evaluation.
- This can also be explained by way users select and move\manipulate object using devices like a mouse as users do tend to concentrate more when selecting object.

Before we conclude this section it would be useful to discuss the results obtained through the subjective scale. The ten statements listed in Appendix 1 are formulated to address three different aspects of vision based interaction i.e. HCI, vision based HCI and computer vision based related statements. The user responses are obtained on a 5-point scale shown in fig 5.

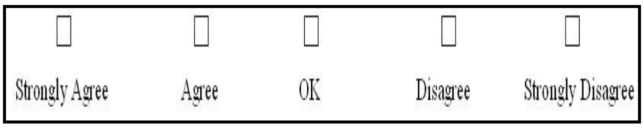

Fig. 5 The 5-point Likert Scale

Statements 1, 2 and 3 relate to HCI aspects of the application. The user feedback for statements 1 and 2 indicates that most of users find ergonomically feasible. The layout of the interaction area is considered visually clear. Feedback on statement 3 which deals with feedback of the system confirms our earlier finding as most users have a neutral view on usefulness of the feedback.

The statements 5, 6, 7 and 8 of the Likert scale deal with the aspects specific to vision based human computer interaction. The responses for statements 5 and 6 indicate that the conceptual simplicity of vision based interaction. Users found it simple to understand the overall concept of a gestural interface and quite straightforward to remember the elements of the gesture vocabulary. User opinion is relatively divided for statements 7 and 8 i.e. preferred mode of interaction and tediousness of manipulation respectively. Most the users do agree that vision based HCI can be a preferred mode of interaction for future.

Finally, statements 4, 9 and 10 deal with the issues specific to the performance of computer vision technique. The user opinion is less favorable regarding these statements. This finding is interesting keeping in view high recognition rates and immediate selection of the required object during evaluations. Therefore, theoretically users should not notice the lag in switching gestures nor have to adjust their hand pose. However, some users did experience this difficulty. This again emphasizes the importance of conducting usability evaluations using standard HCI techniques. The performance of a vision based system cannot be assumed based only on recognition rates or real time performance.

VI. CONCLUSION

Vision based interfaces are considered to be one of the technologies that will define the future of HCI. We have argued that research and development of vision based systems should include standard usability evaluation techniques. We have presented an evaluation framework based on established user and expert evaluation techniques. We also present a ten-statement; five-point Likert scale that can be used or adapted for obtaining user feedback on a computer vision based interactive application. Based on our evaluations we highlight the need for inclusion of HCI techniques for development of computer vision applications.

APPENDIX

The following ten statements constituted our Likert Scale. Each statement is accompanied by a 5-point

1. The movement\hold gesture is user-friendly.
2. Layout of the interaction area is clear
3. Feedback of the system is good
4. The lag in switching gestures is minimal/not noticeable
5. It is straightforward to learn the concept of gesture based interaction.
6. It is easy to remember the gesture vocabulary
7. Gesture based interaction as preferred mode of interaction for future applications.
8. Manipulation is not at all tedious.
9. I did not have to adjust my gesture pose/style in order to interact.
10. The system did recover quickly from an incorrect selection (if applicable).

REFERENCES

[1] V. Hayward, O.R. Astley, M.C-Hernandez, D. Grant and G. Torre, “Haptic devices and interfaces”. *Sensor Review,* vol 24(1), pp:16-29.

[2] Y. Wu and T. Huang, “Vision-based gesture recognition: a review”. In *International Gesture Workshop (GW99*), 1999.

[3] Y. Yin and D. R. Toward, Natural interaction in the real world: real-time gesture recognition. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI '10)*, 2010.

[4] R. Hassanpour, S. Wong and A. Shahbahrami, “Vision based hand gesture recognition for human computer interaction: A review”. In *IADIS International Conference on Interfaces and Human Computer Interaction*, 2008.

[5] R. Kjeldsen and J. Hartman, “Design issues for vision-based computer interaction systems”. In *Proceedings of Perceptible User Interfaces,* 2001.

[6] K. Li, Y. Du and Z. Fu. TreeHeaven: A table game using vision-based gesture recognition. In *RDURP’11*, 2011.

[7] H. Ren and G. Xu, “Human action recognition in smart classroom”. In *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 417-422, 2000.

[8] G. Iannizzotto, F.L. Rosa, C. Costanzo, and P. Lanzafame, “A multimodal perceptual user interface for video-surveillance environments”. In *Proceedings of Multimodal Interfaces (ICMI’05),* pp. 45-52, 2005.

[9] N. Stefanov, A. Galata, R. Hubbold, R., A. Group., and O. Road, “Real-time hand tracking with variable-length markov models of behaviour”. In *IEEE Int. Workshop on Vision for Human-Computer Interaction (V4HCI), in conjunction with CVPR,* 2005.

[10] S. Weiss, *Handheld Usability.* Wiley, 2002.

[11] P. Polson and C. Lewis, “Theory based design for easily learned interfaces”. Human Computer Interaction, vol. 5(2), pp. 191-220, 1990.

[12] B. Schneiderman, C. Plaisant, M. Cohen and S. Jacobs, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*.. Addison Wesley, 2009.

[13] R. Likert, “A technique for the measurement of attitudes”. *Archives of Psychology*, vol. 22(140), pp. 1-55, 1932.

[14] M.R.Ali and T. Morris, “Real Time Single and Multiuser Gesture Recognition Based on Skin Colour and Optical Flow”. To appear in *BMVW,* 2012.

[15] J. Nielsen, Heuristic Evaluation. In *Nielsen, J. and Mack, R.L.* (Eds), *Usability Inspection Methods*, John Wiley & sons, N.Y. 1994.

[16] C. Abras, D. Maloney-Krichmar and J. Preece. User-Centered Design. In: Bainbridge W, editor. *Encyclopedia of Human-Computer Interaction*. Sage; Thousand Oaks: 2004

[17] D.E. Norman, Cognitive engineering. In D. A. Norman and S. W. Draper (eds*) User Centered Systems Design*, Hillsdale, NJ: Lawrence Erlbaum Associates Inc., 1986.