AUTOMATIC ASSESSMENT OF PRODUCT DESIGN USABILITY – A DESIGN METHODOLOGY

Babak Taati and Alex Mihailidis
Intelligent Assistive Technology and Systems Lab
University of Toronto & Toronto Rehabilitation Institute

ABSTRACT

Usability assessment enables improving environmental design so to facilitate independent living among older adults. A multi-phase system design methodology is presented which was used in developing video analysis techniques for product usability assessment. The multi-phase scheme incorporates the output of each phase during the design of following phases so to ease the need for manual annotation.

INTRODUCTION

The prevalence of Alzheimer's disease and other forms of dementia increase with age. With the greying population, the total costs of dementia, including informal care, is estimated at ~1% of the global GDP [1]. Improving the usability of everyday products for people with dementia helps in preserving their independence and reduces the burden of care.

Comparative usability analysis enables selecting the most appropriate product designs for a target population. Unfortunately, usability studies are costly, laborious, time consuming, and also often rely on subjective data. These factors prohibit exhaustive usability studies over the vast array of products of everyday living. It is thus desirable to automate usability assessment so to ameliorate such limiting factors. Ideally, automatic usability assessment would be objective, fast to perform, and generic, i.e. the same approach could be applied to a wide range of products with minimal modifications.

Towards this goal, the objective of this project is to apply artificial intelligence techniques to automate the process of assessing product design usability via computer-based video analysis. Bathroom faucets are chosen as a proof of concept product as they often cause confusion among older adults with dementia [3] while their proper use is critical to self-care activities and personal hygiene.

The developed algorithms analyze recorded videos of human subjects—i.e., older adults at different stages of Alzheimer's disease, while they wash their hands using various faucet types (dual knob, dual lever, single lever). These algorithms aim to automatically estimate factors such as the rate of operation difficulties, the amount and type of caregiver assistance, and the average time it takes to complete actions such as turning on the water or adjusting the flow.

This paper presents the overall multi-phase design methodology applied in developing the automatic video analysis system and some of the lessons learned during the development process. Details on the algorithms developed for each phase can be found at [5], [6], and [7].

Beyond usability analysis, human activity monitoring has other applications within smart home environments, e.g. in cognitive assistive technology and safety monitoring. Previous work in this area includes providing reminder prompts to older adults with dementia [3] and automatic fall detection and emergency response [2].

OVERVIEW

Automatic video analysis typically involves extracting a set of *features* from a video sequence and applying a *classifier*, or multiple classifiers, to categorize the information contained in each video and infer useful knowledge. In a handwashing video, the features encode the flow of water, the motion of the hands, and their interaction with the faucet actuators. The classifiers attempt to *temporally segment* each video, i.e. distinguish the action being performed at each video frame, and also identify the segments which contain operation difficulties.

Computer vision and audio signal processing techniques are used here for feature extraction, e.g. by computing the optical flow, tracking the users' hands, and detecting the sound and appearance of water flow. This information is sufficient for categorizing actions and distinguishing operation difficulties. Turning the water off for instance typically involves reaching for the faucet, placing the hand or hands on the faucet actuator(s), and rotating it clockwise in the case of a dual knob faucet. Following a correct turning-off operation, the water flow also stops. So a classifier observing this sequence of events via the extracted features should be able to distinguish the action of turning the water off from other actions (e.g. turning it on or rinsing hands). It should also be able to recognize operation difficulties such as reaching for the wrong location or pushing a knob instead of rotating it.

Handcrafting classifiers based on such ad hoc observations is laborious and would not scale to the wide range of activities of daily living and associated products. Supervised machine learning methods are used instead to *learn* the governing patterns within each task using a set of labeled *training data* in order to mimic the same categorization in future test cases. Given a set of relevant features and labeled training data, such generic classifiers could be employed in the analysis of a wide range of activities.

Providing ground truth labels on the training set is also a laborious task since it requires manual annotation of several videos. To counter this issue, the final system is developed through multiple phases. At each phase, manual annotation of a handful of videos is used to develop an automatic estimator which eases the annotation process in the next phase. Thus the output of each phase helps with the development of later phases. Following this procedure, the system moves from fully manual annotation, through semi-manual stages, and towards a fully automatic scheme.

DESIGN METHODOLOGY

In the **first phase**, a user-friendly software tool was developed that allowed for loading, playing, and annotating videos. A tracking module was embedded in the software which enabled real-time tracking of regions based on their color distribution. In analyzing handwashing videos, the tracker was used to follow the subjects' hands and how they interacted with the faucets.

The software also provided a graphical user interface for color calibration (to initiate the tracker) and also for annotating actions and states. Action annotation involves marking the beginning and end frames of various actions (e.g. turning the water on or rinsing hands) and also annotating various types of operation difficulties and various levels of caregiver assistance during each segment. State annotation enables the user to specify various states at each frame of the video, e.g. whether or not the water is flowing. Care was given to the design and implementation of a software structure that ensured generality so it could be used for applications beyond the study of water faucets and in assessing the usability of other objects of daily living.

In the **second phase**, an automatic water flow detection algorithm was developed that could identify the flow of water in a bathroom sink based on the combination of audio and video signal processing. Audio features encode the "hissing" sound of the water flow, while discarding other background noise (speech, music, etc), and video features observe the flow of water (or soap water and bubbles) in the sink from the view an overhead camera.

The state annotator GUI developed in the first phase was used to mark ground truth training data for the classifier by manually labeling the water flow in 16 handwashing videos. Marking the labels was easy via the state annotation user interface developed during the first phase and simply involved watching each video and pressing a key every time the water was turned on or off. Combining and fusing an exhaustive list of all audio and video features along with a simple linear classification method (namely, logistic regression) provided the best results in detecting the flow of water. The overall classifier was able to predict the flow of water with high accuracy levels even in presence of distracting factors such as background noise or partial occlusion of the sink region by the subject's hands [6]. After a simple (non-causal) time smoothing (median filtering over 1 s) the average leave-one-out cross validation per frame prediction accuracy was over 92%.

The **third phase** was dedicated to action recognition through temporal segmentation, i.e. assigning a unique action label to each frame of a video. In the study of water faucet designs, temporal segmentation assigns one of five pre-determined labels to each video frame. Three of these labels (*Turn-On*, *Adjust*, and *Turn-Off*) correspond with *actions of interest* indicating the subjects' interaction with the faucet (i.e. turning the water on or off, or adjusting the flow or temperature). The two remaining labels indicate time segments with no user-faucet interaction, while the water is either running (*Is-On*) or not running (*Is-Off*), e.g. when the subject is rinsing or drying their hands.

Histograms of optical flow and gradient of flow, a particle filter for hand-tracking, and the water detection algorithm from the previous phase were used to extract features. A total of 60 manually segmented handwashing videos provided the training sequences with ground truth labels. Hidden Markov Support Vector Machines (HM-SVM) were used to predict the state of the handwashing process for each video frame with high accuracy (~94% average leave-one-out cross-validation per frame accuracy) [5]. Figure 1 illustrates sample segmentation results along the time axis and a sample video frame within each of the three identified segments of interest.

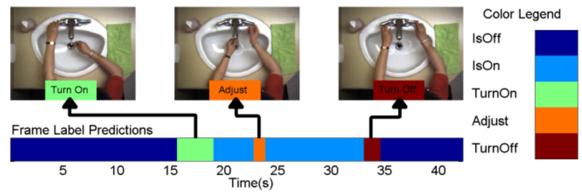

Figure 1: Sample Temporal Segmentation Results

A major challenge during this phase was that the user annotations of actions along the temporal axis contain a certain level of uncertainty associated with subjective decisions. In a handwashing video for instance, while it is intuitively clear when the subject is rinsing his or her hands or turning off the tap, it is not precisely clear when the act of rinsing terminates and the act of turning off the tap starts. In manual annotation, one person might indicate the switch from rinsing to turning-off as soon as the subject's hands leave the water area and start approaching the tap, whereas another annotator might mark the same transition slightly later when the subject's hands reach the faucet actuator. While the difference is subtle and brief, typically less than one second per transition, it leads to confusion when a supervised sequence labeling algorithm attempts to extract a rule based on manually annotated training data. Consequently, the HM-SVM sequence labeling algorithm had to be modified to take into account this uncertainly during its training phase.

In the **fourth phase**, the isolated actions of interest go through further processing in order to identify those in which the subject had any difficulties in operating the faucet. To train a classifier for this task, human annotation is only required for those portions of handwashing videos containing the actions of interest, which typically occupy a small fraction (~3-10%) of the overall video lengths. As such, temporal segmentation results from the previous phase reduced the burden of manual annotation by an order of magnitude. Table 1 presents the average percentage of time was spend on the three actions of interest, using three major faucet types, averaged over 281 handwashing videos.

Table 1: *Average Percentage of Time Spent on Each Action of Interest during a Typical Handwashing Video of an Older Adult with Alzheimer's Disease*

	Turn On	*Adjust*	*Turn off*
Dual Knob	10.1%	3.8%	6.6%
Dual Lever	7.7%	2.6%	4.4%
Single Lever	10.2%	2.8%	3.2%

The features used for this classification task included optical flow and gradient of flow information, computed on overlapping cells across the entire image, and the water flow features. Signal processing and machine learning algorithms (e.g. dynamic time warping and structured support vectors machines with modified loss functions) were used to facilitate the classification challenge.

Preliminary experiments show that the method can reliably identify the relative advantage (or disadvantage) of one faucet design over another in terms of the rate of operation difficulties [7]. For instance, both the automatic video analysis technique and manual annotation identified that older adults had more difficulty in turning the water on using the single lever faucet than when using dual knob of dual lever faucets.

Remaining algorithmic challenges and the focus of current work include automatic identification of caregiver assistance and providing reliable estimates on action completion times.

CONCLUSIONS AND FUTURE WORK

State-of-the-art computer vision and machine learning algorithms were applied to develop video analysis algorithms for product usability assessment. The lessons learned throughout the system development process include taking into account the inherent uncertainly in manually labeled sequences and employing a multi-phase design methodology to ease the burden of manual annotation.

Future phases include automatic identification and categorization of caregiver assistance and estimating action completion times. Future work also involves extending the work beyond the study of water faucets and into the usability analysis of other products of everyday living.

ACKNOWLEDGEMENTS

This work was partially funded by research grants from NIDRR through the RERC on Universal Design and by Toronto Rehabilitation Institute.

REFERENCES

[1] World Alzheimer report. Report by the Alzheimer's Disease International, 2010.

[2] T. Lee and A. Mihailidis. An intelligent emergency response system: preliminary development and testing of automated fall detection. *Journal of Telemedicine and Telecare*, 11:194–198, 2005.

[3] A. Mihailidis, J. Boger, T. Craig, and J. Hoey. The COACH prompting system to assist older adults with dementia through handwashing: An efficacy study. BMC Geriatrics, 8(28), 2008.

[4] A. Mihailidis, J. Boger, K. Fenton, T. Craig, and Z. Li. The impact of familiarity on the usability of everyday products for older adults with alzheimers disease. Intl Conf Alzheimer's Disease and Related Disorders, 2008.

[5] J. Snoek, B. Taati, Y. Eskin, A. Mihailidis, "Automatic Segmentation of Video to Aid the Study of Faucet Usability for Older Adults", Proc. IEEE Intl W. Computer Vision and Pattern Recognition for Human Communicative Behavior Analysis, 2010

[6] B. Taati, J. Snoek, D. Giesbrecht, A. Mihailidis, "Water Flow Detection in a Handwashing Task", Proc. 7th Canadian Conf. Computer and Robot Vision, 2010.

[7] B. Taati, J. Snoek, A. Mihailidis, "Automatic Detection of Difficulties in Operating Everyday Products for Product Design Usability Assessment", submitted to Computer Vision and Pattern Recognition, 2010.