

Automatic Segmentation of a Screen Recording for Scene Identification

Jakub Benjamín VRBA*

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
xvrba@fiit.stuba.sk*

Users are daily interacting with several application or websites, during their work with computer. Usability of these products represents how difficult is to use them. Nowadays it is really important to concentrate on creating interfaces, that are easy to use, because when users are having difficulties with reaching their goals, other solution is waiting just around the corner [1]. User experience is a term describing how users subjectively perceive these systems [2].

Usability testing allows us to evaluate to what extent are interfaces usable. Traditional method of testing applications demands user to be spectated and his behavior is evaluated by researcher manually. This approach may be applicable only with small number of participants – qualitative research. For quantitative test, in which tens of users participate, would be this type of evaluation challenging. In consequence, it is needed to gather data about user interaction from eye or mouse trackers and evaluate them afterwards.

In a phase before analysis, it is necessary for researcher to identify scenes in screen recordings. Scene represents time segment of a screen recording video, during which the examined application remains on a single screen and where monitored areas of interest can be considered static. Scenes are crucial for mapping individual occurrences of a particular application screen into a single entity, which allows the computation of aggregate eye-tracking metrics. At the present time, scenes are annotated manually, that means researcher must go through every recording.

Purpose of this thesis is automatic segmentation of a screen recording for scene identification, for the sake of making this process faster and easier. Our specific goal are usability tests, which were performed on mobile devices. Our data contains screen recordings of these devices and also users activity (fingers). Image processing is ideal technique for implementing this automation. To be specific two approaches were identified:

* Supervisor: Jakub Šimko, Institute of Informatics, Information Systems and Software Engineering

1. supervised approach
2. unsupervised approach.

In supervised approach input from researcher is needed. Specifically, he needs to identify scenes from one video and these scenes will be found on other recordings. That means he does not need to go through all recordings, but one. For this purpose methods of image comparison are used, to determine how similar two images are. On the other hand the unsupervised approach is based on cut detection methods. In this case user will be presented with multiple scenes, where he will select which static image should be considered as a scene and which should be not.

In conclusion, this approach is not automatic too, but the problem of annotating scenes is too complex to be fully automated. The main reason behind this statement is, that what is considered to be a scene is very subjective and depends on what the researcher want to study (in one research it could be ads, in the other it could be menu). This is also related to definition of a scene, where area of interest cannot be interfered and area of interests differ from research to research and cannot be generalized.

References

- [1] Steve Krug. *Don't Make Me Think, Revisited: A Common Sense Approach to Web Usability (3rd Edition)*. New Riders, 2014.
- [2] Jakob Nielsen. *Usability Engineering*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.