Advanced Web User Monitoring with Real-Time Communications Devices

Lukáš Čegan
University of Pardubice, Faculty of Health Studies,
Department of Informatics, Management and Radiology, Pardubice,
Czech Republic.
lukas.cegan@upce.cz

***Abstract*—This paper presents a new approach to web user monitoring with real-time communications devices. Proposed solution is based on a combination of WebRTC technology, performance data logging and third party services. This solutions allows measurement of a user's web performance data, evaluated user's behavior and makes gender and age classification. The results are an important tool in understanding what people think and feel while browsing a website. Thanks to this expertise we can better tailor web page content to achieve our business goals. Experimental results published in this paper were conducted on real data and show that the proposed solution of capturing and transforming face images from a video stream allows achievement of very high accuracy of gender and age classification by third party services.**

***Index Terms*—Monitoring; Web Performance; WebRTC; Real-Time.**

I. INTRODUCTION

Web user monitoring is a technique that allows us to capture and analyze every transaction of every user of our website. User monitoring gives us insight into performance and helps us answer questions like, how does our website really perform from a specific platform, web browser or country. Generally, web monitoring is implemented as a service that continuously observes your system in action, tracking availability, functionality, and responsiveness. Unfortunately, nowadays passive monitoring of user activity and statistically evaluating the measured data is no longer sufficient. This is because most of the contemporary systems require an active response to each user's step in order to achieve higher user satisfaction (preloading web pages) or to achieve a better business value (luring targeted ads).

A number of proposals on how to solve passive monitoring of user activity and evaluation of the captured data has been published by various authors in recent years. The authors use data to measure and evaluate user experience, to predict the behavior of the user, to target marketing, etc. Many published predicting algorithms used a Markov model [1] as well as the data mining technique [2]. Unfortunately, the Markov model is a stateless mathematical model which is not very accurate in predicting future user's steps. To achieve greater accuracy of prediction authors use a higher order Markov model. However, these solutions are extremely complicated due to their large number of states [3], [4]. Additional published solutions are focused on analysis performance and functional problems in real time [5].

Another reason for the monitoring of web users is marketing. Current monitoring systems capture complete transactions for every user's visit. Unfortunately, the weak point of these systems is to identify the user's gender, user's age and emotions that a website evokes. From the perspective of marketing, they are one of the most valuable parameters by which we can create segments of our web users. Eliminating this bottleneck can be achieved using real-time communications capabilities and the implementation of appropriate algorithms that have been developed in recent years in the field of gender, age and emotion recognition from facial expressions. A variety of approaches have been proposed to recognize human faces from frontal views with varying expression and illumination based on image processing methods and multiple linear regression models [6], [7]. Additional published solutions are focused on fast approximations to support vector decision functions in the field of object detection [8]. Also, some authors present a solution based on neural a network [9], [10].

Successful facial recognition allows subsequent examination of gender, age and emotion. Automatic estimation of demographic attributes is a topic of growing interest with many applications [11]. The recognition of emotion from facial expression has been the focus of a large number of psychological studies over the past several decades. The results of these studies are applied in full automatic solutions for identifying human expressions as well as overcoming facial expression variation and intensity problems [12]. To improve the accuracy of emotion recognition some authors of solutions combine the facial expression recognition approach and the speech recognition (e.g. tone of the voice) approach together.

The paper is organized as follows: The web user monitoring techniques are presented in more detail in Section I. Next, in Section II, a proposed solution for user monitoring with real-time communication devices is described. In Section III practical experiments and test results are described. Finally, the last section IV gives conclusions and future research opportunities.

II. PROPOSED SOLUTION

As the main contributions of this paper, we present a robust solution for web user monitoring that can store user performance data, trace user sessions and segment users. This data helped us find out which areas on the page attract more attention and which make users confused. The core of the proposed system consists of two modules, namely:

- measuring performance agent,
- image and voice processing agent.

The outputs of these modules are processed using the "User session logger" which measured data stored in the database

Hadoop. Then, this data is used for a group of different modules for statistical evaluation. Based on the results of the data evaluation, for example, you can choose the most appropriate web page content (insertion ads) or you can predict with a high degree of accuracy the next user step (used for preloading pages). Architecture of the proposed solution is shown in Figure 1.

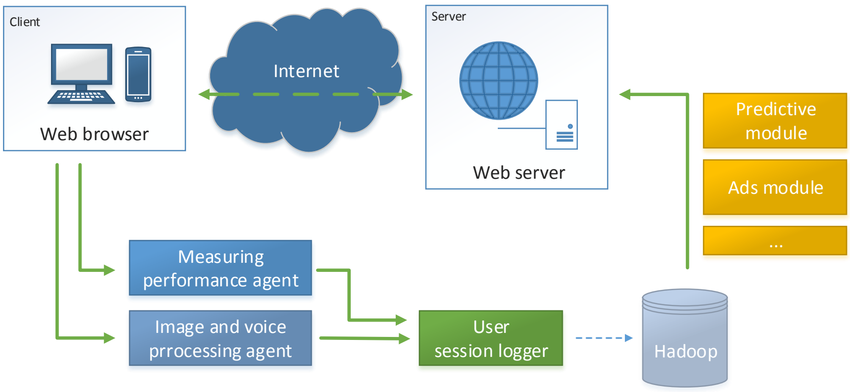

Figure 1: Architecture of the proposed solution

A. Measuring performance agent

The measuring performance agent is responsible for identifying the web user environment settings and it also monitors user interaction with a web page such, as seeing all mouse movements, scrolls, key-strokes and clicks in a time (see Figure 2). The agent collects two type of data. The first category of data includes static data such as:

- type of platform (OS/web browser),
- type of device (smart device, PC, tablet, …),
- screen resolution,
- type of network connection,
- internationalization and localization setting (i18n),
- user position (location).

This data is constant during the entire user session. The second category of data is dynamic data. This data reflects user interactivity with the web page. The category includes:

- date/time stamp and language mutation,
- the predecessor of page (navigation flow),
- page load time and page interaction time,

user interaction with the elements (such as play embedded video, etc.),

- mouse position in time (path of mouse movement) and state of scrollbar,
- interest level/user target.

Real time monitoring is performed using "Navigation Timing" which provides a browser's JavaScript API for obtaining performance data of every HTTP request [13]. The measured data is processed by the "User session logger" module, and then the data is stored in the Hadoop database.

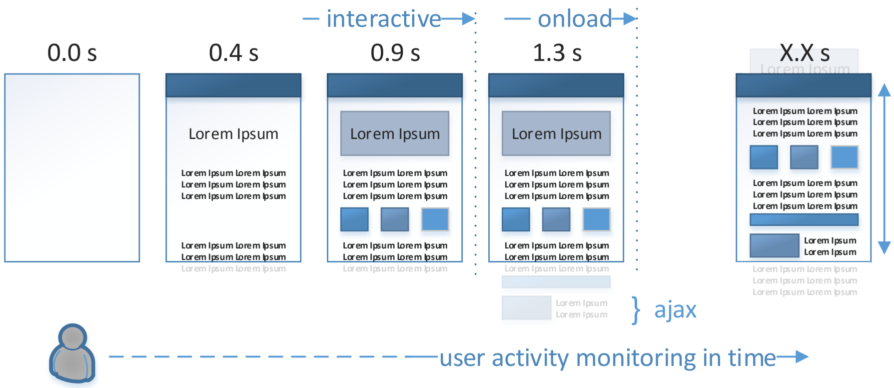

Figure 2: User session monitoring

B. Image and voice processing agent

An image and voice processing agent is responsible for the real-time monitoring of web user and measured data evaluation. The agent uses the WebRTC API to control a user's web camera and microphone. WebRTC is a plugin-free, built-in real-time audio and video capabilities and codecs to web browsers [14]. It is very powerful platform that continues to evolve. Currently, a number of authors also used WebRTC as the appropriate means to establish peet-to-peer connection in a web environment as a solution for a web-based distributed simulation [15]. In our solution the connection between client and server can be realized over HTTP or WebSockets.

Figure 3 illustrates the inner structure of the Image and voice processing agent, which consists of three basic blocks. The first block "Transcoder" is responsible for converting the video and audio streams to the output format (digital encoding, compression, fragmentation, etc.). The next block "Gateway" ensures the routing of the input data to the target service. Within the agent one internal service based on the OpenCV library is available, which provides application logic such as face detection from video stream, capture images, image transformation, etc. (third block). Moreover, input video and audio signal can be routed for a more comprehensive analysis to third-party services like a Sky Biometry, Animetrics, SensibleVision and many others.

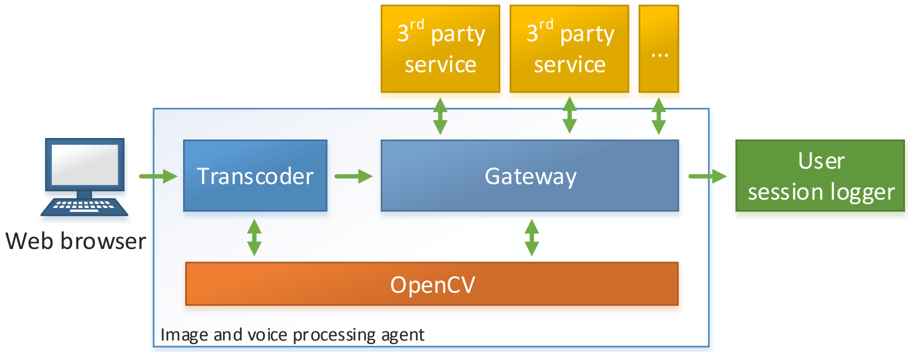

Figure 3: Image and voice processing agent

III. EXPERIMENTAL RESULTS AND DISCUSSION

We have the performed an experiment to evaluate the performance of the proposed solutions. Experiment was performed on a reference group of twenty users of different age and gender. Performance evaluation was realized in two different scenarios. In scenario A, the images of user's face were captured at a fixed time after the detection from the video stream and the images were not subsequently adjusted. In scenario B, the images of user's face were captured dynamically in an appropriate position (front position) and images were automatically adjusted (the brightness, contrast and color scheme). Furthermore, in both scenarios, the images were transmitted via the API for processing to third-party applications. In our experiment, we used Alchemi, EyeDea, Animetric(only gender), Sky Biometry, Face++. The resulting measured values are shown in Table I. The column "gender detection" represents the user's gender recognition accuracy, column "age +/- 5 years" represents the accuracy of the user's age within the range of plus or minus 5 years. The last column "age group" shows the accuracy of the user classification in the correct age group <18, 18-24, 25-34, 35-44, 45-54, 55-64, >64.

Table 1
The degree of accuracy of detection and classification [%]

Service	Scenario A Gender Detection	Scenario A Age +/- 5 years	Scenario A Age group	Scenario B Gender Detection	Scenario B Age +/- 5 years	Scenario B Age group
Animetrics	70	-	-	85	-	-
Alchemi	75	55	40	90	65	50
EyeDea	90	60	45	100	70	55
Sky Biometry	100	70	55	100	70	60
Face++BEA	85	50	40	95	65	55

IV. CONCLUSION AND FUTURE WORK

This paper presented an approach to web user monitoring with real-time communication devices (web camera, microphone). The proposed solution allows us to effectively identify web user's environment settings and it also allows us to monitor and measure user interaction with a web page. Moreover, an image and voice processing agent allow us to recognize the gender and age of web users using WebRTC and third party applications. Evaluation of data set showed a high accuracy combined with real-time performance. This makes the approach particularly suitable for services that evaluate user behavior for marketing purpose, for measuring the impact of the changes in a website on the target group of users or to evaluate websites performance parameters. A more serious limitation of user monitoring with WebRTC is that the user must allow the use of communication devices at its facility which currently is not usual for security reasons. Currently we are working on improving the algorithm to capture the appropriate position of the user's face for more accurate gender and age detection. As future work, we are planning implement to our solution module for emotion detection, which would, in an automated manner, determine what kind of emotions a website raises.

ACKNOWLEDGMENT

This paper is published thanks to the financial support Faculty of Health Studies, University of Pardubice specific research project.

REFERENCES

[1] DongshanX. and Juni S., "A New Markov Model for Web Access Prediction," *IEEE Computing in Science and Eng*. vol. 4, pp. 34-39, November 2012.

[2] ShafiqAlam, Gillian Dobbie, Patricia Riddl. "Exploiting Swarm Behaviour of Simple Agents for Clustering Web Users' Session Data," *Data Mining and Multi-agent Integration*, pp. 61-75, 2009, Springer US.

[3] Deshpande M. and Karypis G., "Selective Markov Models for Predicting Web Page Accesses," *ACM Trans. Internet Technology*, vol. 4, pp. 163-184. May 2004.

[4] Borges J. andLevene M., "Evaluating Variable-Length Markov Chain Models for Analysis of User Web Navigation Sessions," *IEEE Trans. On knowledge and Data Engineering*, vol. 19, no. 4, April 2007.

[5] Cegan L., "Intelligent Preloading of Websites Resources Based on Clustering Web User Sessions," *IT Convergence and Security (ICITCS)*, pp. 1- 4, Kuala Lumpur. IEEE. Aug 2015.

[6] Talele K.T., Kadam S., "Face detection and geometric face normalization," *TENCON 2009 - 2009 IEEE Region 10 Conference*, Singapore, ISBN 978-1-4244-4546-2, 2009.

[7] Lijing Zhang, Yingli Liang, *The 2nd IEEE International Conference on Advanced Computer Control(ICACC 2010)*, Shenyang, China, pp. 490 - 494, 2010, ISBN 978-1-4244-5845-5

[8] Kienzle, W., Bakir, G., Franz, M., Scholkopf, B. "Face detection – efficient and rank deficient ,"*Advan. in neural inform. process. systems* vol. 17, pp. 673–680, 2005.

[9] ZulhadiZakaria, Shahrel A. Suandi, "Face Detection Using Combination of Neural Network and Adaboost ," *TENCON 2011 - 2011 IEEE Region 10 Conference*, Bali, pp. 335 - 338, 2011ISBN 978-1-4577-0256-3.

[10] Xiaoning Liu, GuohuaGeng, Xiaofeng Wang,"Automatically face detection based on BP neural network and Bayesian decision, Natural Computation (ICNC)," *Sixth International Conference on Natural Computation*, Yantai, Shandong, pp.1590-1594. ISBN 978-1-4244-5958-2.

[11] Bauckhage C., Jahanbekam A., Thurau C., "Age Recognition in the Wild," *Pattern Recognition (ICPR)*, Istanbul, pp. 392-395, 2010, ISBN 978-1-4244-7542-1.

[12] Mliki H., Fourati N., Smaoui S., Hammami M., "Facial Expression Recognition System," *Computer Systems and Applications (AICCSA), 2013*, Ifrane, ISSN: 2161-5322.

[13] Wang Z. and Jain A. Navigation Timing. Technical Report https://dvcs.w3.org/hg/webperf/rawfile/tip/specs/NavigationTiming/Overview.html, W3C, January 2013.

[14] Alan B. Johnston, Daniel C. Burnett. "WebRTC: APIs and RTCWEB Protocols of the HTML5 Real-Time Web," Third Edition, pp. 350, *Digital Codex LLC; 3 edition*, March 11 2014, ISBN 978-0985978860

[15] Kartak S., Kavicka A., "WebRTC Technology as a Solution for a Web-Based Distributed Simulation," *Proceedings of the European Modeling and Simulation Symposium,* pp. 343–349, 2014, Genova: Università di Genova.

[16] PavelJetensky, "Human Hand Image Analysis Extracting Finger Coordinates and Axial Vectors: Finger axis detection using blob extraction and line fitting," *In 2014 Radioelektronika (RADIOELEKTRONIKA), 24th International Conference IEEE. Bratislava (Slovak Republic)*, pp. 73-77, 2014, ISBN 978-1-4799-3713-4.