論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Personalized Saliency in Non-immersive and Immersive Environments for Practical Applications
著者(和文)	Erum Zaib Sumaira
Author(English)	Sumaira Erum Zaib
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12294号, 授与年月日:2022年12月31日, 学位の種別:課程博士, 審査員:山村 雅幸,瀧ノ上 正浩,小野 功,青西 亨,関嶋 政和
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12294号, Conferred date:2022/12/31, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	Computer Science, Artificial Intelligence　系 コース	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of (Engineering)
学生氏名： Student's Name	Sumaira Erum Zaib	指導教員（主）： Academic Supervisor(main)	Professor Masayuki Yamamura
		指導教員（副）： Academic Supervisor(sub)	

要旨（英文 800 語程度）
Thesis Summary（approx.800 English Words ）

Saliency is the ability of being important, noticeable, and standing out among others. Saliency that takes the heterogeneity of individuals into consideration is called personalized saliency. Currently personalized saliency models in non-immersive environments, such as smartphones and computers, use deep learning methods which are resource heavy on-device and pose network limitations and privacy concerns in cloud. In this thesis, we propose using gradient boosted tree regression and just the color spaces of user fixations to determine personalized saliency without collecting invasive personal data. In immersive environments such as virtual reality, we determine that the use of auditory and emotional saliency is much more useful for practical application purpose. Therefore, we propose using different machine learning algorithms along with heart rate from user's personal smartwatch to successfully induce and amplify emotions for personalized saliency. Using devices already owned by users and these techniques, we successfully bridge the gap between personalized saliency research and practical applications in both non-immersive and immersive environments.

In chapter 1, a deep understanding of saliency and personalized saliency is established. Bottom-up, top-down and combination model approaches used to determine and predict saliency are studied. Different kinds of saliency including visual, emotional, and auditory saliency is also understood. The research motivations and objectives of this thesis are laid down at the end of this chapter.

In chapter 2, the current state of research regarding saliency is discussed. For visual saliency, different universal saliency models and personalized saliency models have been proposed for both detection and prediction purposes. However, a trend towards utilizing deep learning, more specifically convolutional neural networks, is noticed. Although impressive results have been achieved using deep learning, we establish that they cannot be used for applications in consumer devices, such as smart phones, due to expensive resource needs. Efforts have been made to utilize cloud services to relieve storage and computational needs but at the expense of privacy and network costs.
For emotional saliency, there is a lot of work being done to understand, induce and predict emotions. However, this research is not easily utilized by consumers due to the usage of specialized gadgets and sensors that are either not available or expensive for everyday consumers. Furthermore, these studies are carried out in controlled environments and the results might be different in real life conditions. Research also shows heart rate to be an acceptable parameter for determining emotion. For personalization purposes, it is established that colors, gender, and environment play a huge role in determining what is considered salient for an individual.

In chapter 3, a gradient boosted tree regression model that extracts personalized saliency map from the universal saliency map is proposed for prediction of personalized saliency in non-immersive environments. The model is designed to be least invasive with personal information and utilizes only the fixations of the individuals. Based on the color spaces of fixation pixels, the proposed model can predict visual personalized saliency with an average accuracy of 0.80 (AUC Judd metric). Discussion on how image content and individual gaze behavior impacts saliency prediction is also presented.

In chapter 4, first an understanding is created that visual saliency is not appropriate to determine attention in immersive environments. Research shows that in such environments, what people look at is not enough to determine if it matches up with what people think. Also due to the inaccessibility

of eye tracking technology, using emotional saliency rather than visual saliency in immersive environments is proposed. Using virtual reality headset, a smart watch and android smartphone emotional saliency (more specifically fear) is determined. The results show no statistically significant differences in control group and experimental group. Correlation between heart rate from subjects ' own smart watch and self-reported results from self-assessment manikin test also establish that smart watch parameters, at least heart rate, can be utilized be developers to create personalized applications using emotions. Discussion on how gender and environment effect the experience of emotions is also presented.

In chapter 5, applications of personalized saliency in non-immersive and immersive environments are put forth. Although universal saliency is already being utilized for image and video compression, this application can also be extended towards personalized saliency. Other applications include dynamic user interface design and search optimization. Personalized saliency in immersive environments has important applications in emotion regulation and mental therapy for treatment of anxiety, stress, and other disorders. It can also be used in training for situations that might be dangerous in real-life. Other social applications include empathy invocation, immersive gaming, and tourism.
Furthermore, limitations of this study in both immersive and non-immersive environments are presented that can be used for improvement in future works. In conclusion, this thesis discusses why personalized saliency is important and proposes frameworks to calculate them successfully in both immersive and non-immersive environments for practical applications.