MONITORING THROUGH EYE-MOVEMENT DATA IN CONTEXT-AWARE ADAPTIVE SOFTWARE SYSTEMS: A CASE STUDY ON ENDO-NEUROSURGERY TRAINING PROGRAMS

A DOCTOR OF PHILOSOPHY THESIS

in

Software Engineering

Atılım University

by

GONCA GÖKÇE MENEKŞE DALVEREN

DECEMBER 2017

MONITORING THROUGH EYE-MOVEMENT DATA IN CONTEXT-AWARE ADAPTIVE SOFTWARE SYSTEMS: A CASE STUDY ON ENDO-NEUROSURGERY TRAINING PROGRAMS

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

ATILIM UNIVERSITY

BY

GONCA GÖKÇE MENEKŞE DALVEREN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

THE DEPARTMENT OF SOFTWARE ENGINEERING

DECEMBER 2017

Approval of the Graduate School of Natural and Applied Sciences, Atılım University.

____

Prof. Dr. Ali Kara
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

____

Prof. Dr. Ali Yazıcı
Head of Department

This is to certify that we have read the thesis "Monitoring Through Eye-Movement Data in Context-Aware Adaptive Software Systems: A Case Study on Endo-neurosurgery Training Programs" submitted by "Gonca Gökçe Menekşe Dalveren" and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

____

Assoc. Prof. Dr. Nergiz Ercil Çağıltay
Supervisor

Examining Committee Members

Prof. Dr. Ali Yazıcı	______________________
Assoc. Prof. Dr. Nergiz Ercil Çağıltay	______________________
Assoc. Prof. Dr. H. Hakan Maraş	______________________
Assoc. Prof. Dr. Erol Özçelik	______________________
Assist. Prof. Dr. Yavuz İnal	______________________

Date: December 29, 2017

I declare and guarantee that all data, knowledge and information in this document has been obtained, processed and presented in accordance with academic rules and ethical conduct. Based on these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Gonca Gökçe Menekşe Dalveren

ABSTRACT

MONITORING THROUGH EYE-MOVEMENT DATA IN CONTEXT-AWARE ADAPTIVE SOFTWARE SYSTEMS: A CASE STUDY ON ENDO-NEUROSURGERY TRAINING PROGRAMS

Menekşe Dalveren, Gonca Gökçe

Ph.D., Software Engineering Department

Supervisor: Assoc. Prof. Dr. Nergiz Ercil Çağıltay

December 2017, 111 pages

Today, modern software is becoming very complex which needs to be compatible with constant changes in the environment. They required to support autonomic behaviors by monitoring the relevant phenomena of the environment and analyzing the collected data to better understand the possible consequences of the changes in the environment. In other words, by monitoring the relevant phenomena of the environment and analyzing the collected data to better understand the possible consequences of the changes in the environment, these type of software adapt themselves to the environment.

Context defined as anything that can be observed by the software system including end-user, computing, and primary features of identity, location, time, and physical conditions at runtime. Hence, Context aware adaptive software (CAASS) architecture can be implemented at different levels for different purposes by monitoring a wide range of data. However, currently there is no conceptual framework showing the level and scope of the adaptation performed by these systems.

Accordingly, in this study, first the related literature is examined to investigate the main dimensions of CAASS. Afterwards, a conceptual framework is proposed to address the level and scope of adaptation performed by a specific CAASS. The proposed framework has three dimensions namely the definition of the context of the adaptation, definition of the event that is planned to be adapted and finally the plan showing how the adaptation aimed to be performed.

Additionally, a case study is also conducted for endo-neurosurgery education programs through the proposed conceptual framework. Results showed that by monitoring eye-movement events of the surgeons, their skill levels can be estimated with a high precision (91.3%). Accordingly, for this specific case, it is shown that, through the eye-movement events of surgeons, the content can be adapted according to the behaviors of the surgeons. The results of this study show evidences that, by regularly assessing their skill levels and evaluating the difficulty levels of each computer-based simulation scenario through eye movement events of the trainees, order of these scenarios in the curriculum can be adapted to the user skill levels and behaviors under different hand conditions. This will help to create a specific curriculum for each trainee that is adapted dynamically to their skill and knowledge. This study has two main contributions. First it proposes a conceptual model that can be used to evaluate the scope and the level of adaptation for CAASS. This information may help the researchers and the developers to better evaluate and compare the CAASS. The second contribution of this thesis study is the implementation of the proposed model on endo-neurosurgery domain.

The field of endo-neurosurgery education programs have several problems. The main problem of these programs is the skill-based training opportunities. As the training and skill development had to be provided in the operating room, there are several drawbacks of these education programs such as the ethical considerations from the patients' perspective, limited time and cases as well as the risk of patient safety.

Currently, there are not many alternative training opportunities for the surgical training programs. As the skill improvement is very critical for these programs, the individual skill-based training opportunities are required. Even there are some

examples of computer-based simulations for supporting surgical training programs, there are very limited examples of curriculum integrated models. Additionally, there is no instructional model of CAASS for the surgical education programs especially in the endo-neurosurgery education programs.

We believe that, because of its very nature, CAASS approach may provide several benefits for the endo-neurosurgery education programs. However, as the process of creating CAASS for the field of endo-neurosurgery education programs is a very complex, in this thesis study a level of CAASS conceptual model is proposed. The findings of this thesis study is aimed to help future studies to better build CAASS for the field of endo-neurosurgery education programs and to better integrate these systems into the current educational programs.

Keywords: context-aware systems, self-adaptive software systems, eye-tracking, eye-movement events, eye-movement classification algorithms, surgical skill levels, surgical training programs

ÖZ

BAĞLAM-FARKINDA UYARLAMALI YAZILIM SİSTEMLERİNDE GÖZ-HAREKETİ VERİSİ İLE GÖZLEMLEME: ENDOSKOPİK-NÖROŞİRURJİ EĞİTİM PROGRAMLARI İÇİN BİR DURUM ÇALIŞMASI

Menekşe Dalveren, Gonca Gökçe

Doktora, Yazılım Mühendisliği Bölümü

Tez Yöneticisi: Doç. Dr. Nergiz Ercil Çağıltay
Aralık 2017, 111 sayfa

Günümüzde modern yazılımlar, ortamdaki sürekli değişimlerle uyumlu olmaları gerektiğinden dolayı çok karmaşık hale gelmektedir. Çevreyle ilgili olguları izleyebilmek ve çevredeki değişikliklerin olası sonuçlarını daha iyi anlayabilmek için toplanan verileri analiz etme yeteneği aracılığıyla otonomik davranışları desteklemeleri gerekmektedir. Başka bir deyişle, bağlam-farkında uyarlamalı yazılım sistemi (BFYUS), çalışma ortamındaki bu değişikliklere cevaben kendisini çalışma zamanında ayarlamayı amaçlamaktadır.

Bağlam, son kullanıcı, programlama, birincil özellikler (ör. kimlik, yer ve zaman) ve fiziksel koşullar gibi yazılım sistemi tarafından gözlemlenen herhangi bir şey olarak tanımlanır. Bu nedenle BFYUS mimarisi, geniş bir veri yelpazesini izleyerek farklı amaçlar için çeşitli seviyelerde uygulanabilir. Bununla birlikte, şu anda bu sistemlerin uyguladığı adaptasyonun seviyesini ve kapsamını gösteren hiçbir kavramsal çerçeve yoktur.

Bu çalışmada, öncelikle BFYUS'ın ana boyutlarını daha iyi anlamak için ilgili literatür incelenmiştir. Daha sonra, belirli bir BFYUS tarafından gerçekleştirilen

adaptasyonun seviyesine ve kapsamına daha iyi hitap edebilmek için kavramsal bir çerçeve önerilmektedir. Önerilen çerçeve üç boyuta sahiptir: adaptasyon bağlamının tanımı, uyarlanması planlanan olayın tanımı ve nihayetinde adaptasyonun nasıl gerçekleştirileceğini gösteren plan.

Ek olarak, önerilen kavramsal çerçeve aracılığıyla endo-nöroşirurji eğitim programları için bir durum çalışması yürütülmüştür. Sonuçlar, cerrahların göz hareket olaylarını izleyerek, yetenek seviyelerinin yüksek hassasiyetle (%91.3) tahmin edilebileceğini göstermiştir. Buna göre, bu özel durum için, cerrahların göz hareketi olaylarıyla, içeriğin cerrahların davranışlarına göre uyarlanabileceği görülmüştür. Örneğin, beceri düzeylerini düzenli olarak ölçerek ve öğrencilerin göz hareket olaylarıyla her bir senaryonun zorluk seviyelerini değerlendirerek, müfredattaki bilgisayar tabanlı simülasyon senaryolarının düzeni, farklı el koşullarındaki kullanıcı beceri seviyelerine ve davranışlarına göre uyarlanabilir. Bu, her stajyer için beceri ve bilgiye dinamik olarak adapte edilmiş özel bir müfredat oluşturulmasına yardımcı olacaktır. Bu çalışmanın iki ana katkısı vardır. İlk olarak, BFYUS'ın kapsam ve seviyesini değerlendirmek için kullanılabilecek bir kavramsal model önermektedir. Bu bilgi araştırmacılara ve geliştiricilere BFYUS'ı daha iyi değerlendirip karşılaştırmalarına yardımcı olabilir. Bu tez çalışmasının ikinci katkısı önerilen modelin endo-nöroşirürji alanına uygulanmasıdır.

Endo-nöroşirurji alanındaki eğitim programları çeşitli problemlere sahiptir. Bu programların asıl problemi, beceri temelli eğitim fırsatlarıdır. Eğitim ve beceri gelişiminin ameliyathanede sağlanması gerektiğinden dolayı, bu eğitim programlarının, hastaların bakış açısından etik hususlar, sınırlı zaman ve hasta güvenliği riski gibi pek çok dezavantajı vardır.

Şu anda, cerrahi eğitim programları için pek fazla alternatif eğitim olanağı bulunmamaktadır. Bu programlarda beceri geliştirme çok kritik olduğu için, bireysel beceri temelli eğitim olanakları gerekmektedir. Cerrahi eğitim programlarını desteklemek için bilgisayar tabanlı simülasyonlara örnekler olsa bile, müfredata entegre modeller sınırlıdır. Ek olarak, özellikle endo-nöroşirurji eğitim programlarında cerrahi eğitim için BFYUS'ın herhangi bir öğretim modeli yoktur.

Doğası gereği, BFYUS yaklaşımının endo-nöroşirurji eğitim programlarına çeşitli avantajlar sağlayabileceğine inanıyoruz. Bununla birlikte, endo-nöroşirurji eğitim programları için BFYUS oluşturma süreci çok karmaşıktır, bu tez çalışmasında bir BFYUS kavramsal modeli önerilmektedir. Bu tez çalışmasının bulguları, endo-nöroşirurji eğitim programları için daha iyi BFYUS oluşturmak üzere gelecekteki çalışmalara yardımcı olmak ve bu sistemleri mevcut eğitim programlarına daha iyi entegre etmek amacıyla hazırlanmıştır.

Anahtar Kelimeler: Bağlam-farkında sistemler, uyarlamalı yazılım sistemleri, göz izleme, göz hareket olayları, göz hareketi sınıflandırma algoritmaları, cerrahi beceri seviyeleri, cerrahi eğitim programları

To My Family

ACKNOWLEDGMENTS

First and foremost I would like to express my sincere gratitude to my supervisor Assoc. Prof. Dr. Nergiz Ercil Çağıltay for her valuable advises, support, guidance, insight, patience, and encouragement throughout the study.

I owe special thanks to my family, my mother Ayşe Menekşe, my father Memduh Menekşe and my brother Mert Menekşe for their unconditioned support and endless patience. It is great to know that they will be always with me whenever I need a help.

A special thanks to my dearest husband, Yaser Dalveren and his family. He has been my motivation and inspiration for writing this thesis. His endless patience and support helped me throughout my entire thesis. I could never have accomplished this thesis without his love and encouragement.

I express sincere appreciation to my thesis progress jury members, Assoc. Prof. Dr. H. Hakan Maraş and Assoc. Prof. Dr. Erol Özçelik for their criticism, guidance, and suggestions. I am thankful to my thesis jury members, Prof. Dr. Ali Yazıcı and Assist. Prof. Dr. Yavuz İnal for their evaluations and future instructions.

A part of this study was accomplished using Atılım University Software Engineering Simulation laboratory (SimLAB). I would like to additionally thank Assoc. Prof. Dr. Nergiz Ercil Çağıltay for providing the equipment and also for their hands-on laboratory guidance.

I am also grateful to Atılım University, the head of Software Engineering Department, Prof. Dr. Ali Yazıcı for his support.

Finally, this study is conducted for improving the scenario designs of the educational materials, which are developed for endo-neurosurgery education project (ECE: Tubitak 1001, Project No: 112K287) purposes. I would like to thank for the support of TÜBİTAK 1001 program for realizing this research and to the ECE project team and the Hacettepe University Medical School for their valuable support throughout the research.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

LIST OF ABBREVIATIONS

GTE	-	Generic Tutoring Environment
CAT	-	Computer Adaptive Testing
CAASS	-	Context Aware Adaptive Software Systems
ECE	-	Educational Computer-based-simulation Environment
MIS	-	Minimally Invasive Surgery
I-VT	-	Velocity Threshold Identification
I-HMM	-	Hidden Markov Model Identification
I-KF	-	Kalman Filter Identification
I-DT	-	Dispersion Threshold Identification
I-MST	-	Minimum Spanning Tree Identification
I-VVT	-	Velocity and Velocity Threshold Identification
I-VDT	-	Velocity and Dispersion Threshold Identification
AED	-	Adaptive Event Detection
BIT	-	Binocular-Individual Threshold
FN	-	Number of Fixation
FD	-	Fixation Duration
SN	-	Saccade Number
SD	-	Saccade Duration
SAD	-	Saccade Amplitude Degree
PN	-	Pursuit Number
PD	-	Pursuit Duration
PVD	-	Pursuit Velocity Degree
DH	-	Dominant Hand
NH	-	Non-dominant Hand
BH	-	Both Hand

CHAPTER 1

INTRODUCTION

Today, software is the major activator of many applications and devices everywhere in our daily lives. Software systems are required to address management complexity, unexpected circumstances, changing policies and priorities that regulate objectives, and changing conditions (Babaoglu et al., 2005). Software affects the welfare and job satisfaction while providing the best utility anytime and anywhere (Hallsteinsen et al., 2012). With embedded systems, modern software is becoming even more compatible with constant changes in the environment. They are becoming able to support autonomic behaviors through the ability of monitoring the relevant properties of the environment and analyzing the collected data to predict the possible consequences of the changes in the environment (Filieri, Ghezzi, & Tamburrelli, 2012).

Traditionally, a significant portion of the research to address complexity and achieve quality objectives has focused on software development and internal quality attributes. However, in recent years there has been an increasing demand for coping with these problems. The main causes of this tendency are an increase in the level of heterogeneity of software components. Future software systems are expected to work in an extremely dynamic world. Despite of the unexpected changes in factors such as environmental conditions, user expectations, technology, regulations and market opportunities, systems will be expected to work properly (Di Nitto, Ghezzi, Metzger, Papazoglou, & Pohl, 2008). Hence, because of the increased cost of achieving these objectives within complex software systems today, self-adaptive software is becoming an alternative (Robertson, Laddaga, & Shrobe, 2000). Self-adaptive systems and context-aware systems aim to adjust themselves in response to changes at runtime (Esfahani, Elkhodary, & Malek, 2013; Salehie & Tahvildari, 2007) to cope with changes in their environment and according to the user's needs (Hussein, Han, & Colman, 2011; Kephart & Chess, 2003). For this reason there is a need for automatic adaptation to react these changes (Di Nitto et al., 2008). Hence, the need to acquire flexibility and adaptability in complex software systems is obvious and has become a fundamental challenge for modern software engineering. In other words, self-adaptive systems enable algorithms to dynamically adapt to the problem characteristics and cope with changing environmental conditions (Bäck, 2002). Self-adaptability is a technology that brings flexibility and adaptability to information systems. Such systems should be highly adaptable to react to environmental changes while providing fault tolerant, autonomous and acceptable performance (Karsai & Sztipanovits, 1999). These systems should act autonomously by changing the software composition to better fit the current environment while avoiding damage or loss of service (McKinley, Sadjadi, Kasten, & Cheng, 2004). Self-adaptation ensures that the software is operated successfully in dynamic, unpredictable and uncertain environments (Filieri, Hoffmann, & Maggio, 2014). However, self-adaptation by a system is considered as a complex process and depends on several variant variables (da Silva & de Lemos, 2011). Implementing the proposed models on large-scale context-aware systems is not free of obstacles (Raisinghani et al., 2006). In this study, we propose a Context Aware Adaptive Software System (CAASS) as a software that uses context-awareness to adapt to user-specific skills rather than force users to apply a particular assisting technology.

1.1 Context in CAASS

An important feature of adaptive systems is that, they can perceive the context to adapt to the specific capabilities of users (Macik, Cerny, & Slavik, 2014). Context-awareness is defined as "the ability of the computer to sense and act upon information about its environment, such as location, time, temperature or user identity" (Ryan, Pascoe, & Morse, 1999). Context is defined as information that characterizes the conditions of an entity such as a person, place or an object (Abowd et al., 1999). Today, research on self-adaptive systems merely work on how to adapt the system according to the response to context and requirement changes, while context-aware systems are mainly concerned with how to model, process and manage context information, where both approaches need to be evaluated together to generate a single model to better create an adaptive software architecture (Hussein et al., 2011). For example, information and user interface should be adapted to contextual features such as the user, environment, and access device (Viana & Andrade, 2008). However, specifying, designing, verifying and realizing such software systems that evolve at runtime is a challenge (Cheng et al., 2009; Dobson, Sterritt, Nixon, & Hinchey, 2010; Huebscher & McCann, 2008; Kramer & Magee, 2007; Salehie & Tahvildari, 2009).

Self-adaptive software changes its behavior in response to changes in the operating environment, namely context, that is anything that can be observed by the software system including end-user input, external hardware devices, and program instruments (Oreizy et al., 1999). The context is broadly understood as any information related to the user's needs and the working environment that can be changed dynamically and can influence applications and be monitored using the relevant hardware and software mechanisms (Hallsteinsen et al., 2012).

Regarding the abilities and preferences of users, most context models are now trying to adapt to the standard systems of various assistive technologies. Skill-based design (Wobbrock, Kane, Gajos, Harada, & Froehlich, 2011) has been proposed for this purpose. Because of the complex nature of such a process, context parameters could be grouped and selected according to certain criteria and priorities (A. A. Economides, 2009). Researchers suggest that such data may not be satisfactory on its own and may be useful to monitor user's feelings (Sykes & Brown, 2003), such as disappointment levels (Gilleade & Dix, 2004), by making measurements from control pads (Sykes & Brown, 2003) or from more advanced sensors (Charles et al., 2005). As a summary of this literature, the context components of a CAASS can be grouped as shown in Figure 1.1.

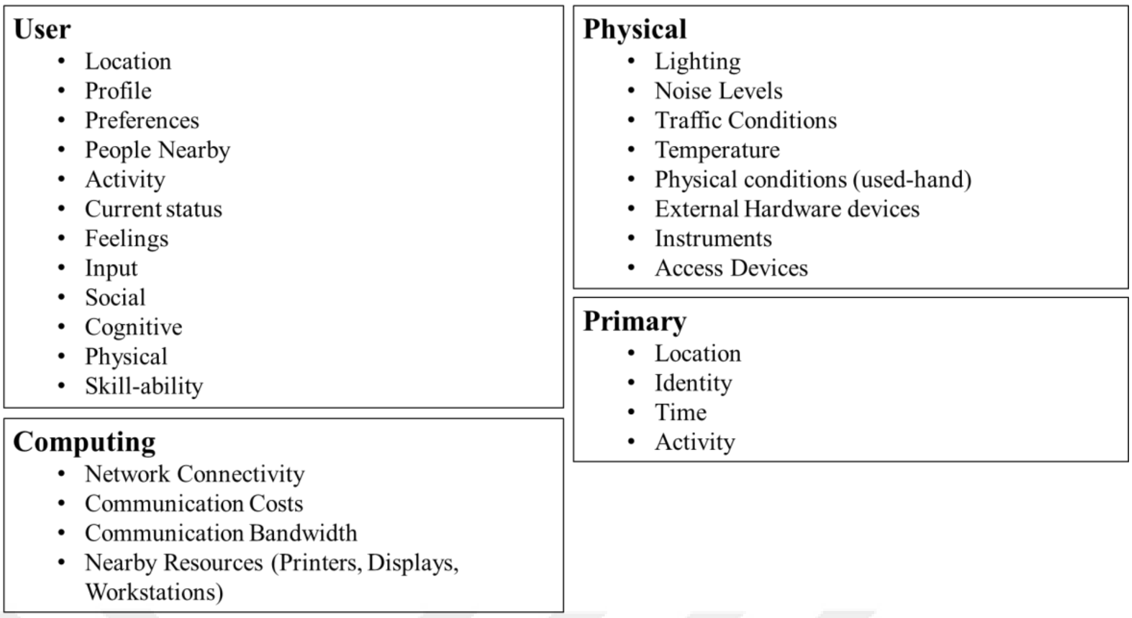


Figure 1.1 Context Components of CAASS

Software systems take the form of self-adaptability by creating applications to automatically and autonomously adapt their behavior to respond to evolving needs, contextual changes, and failures of component services. This is a continuous process conducted by regularly collecting information about the context and performing the adaptation process. The life-cycle of this continuous process is defined in the following.

1.2 Life Cycle of CAASS

As described earlier, the context is a collection of settings and conditions, and knowing a current context is not often easy. For this, a CAASS monitor analyzes the settings and provides relevant information and services to the user; for example, that the information and user interface should be adapted to contextual features like user, environment, and access device (Mizouni, Matar, Al Mahmoud, Alzahmi, & Salah, 2014; Viana & Andrade, 2008). Context-awareness, therefore, aims to increase the utility of the application, taking into account a wide range of contextual features (Ceri, Daniel, Facca, & Matera, 2007). The components of a self-adaptive system are defined as in Figure 1.2 (da Silva & de Lemos, 2011; Salehie & Tahvildari, 2007).

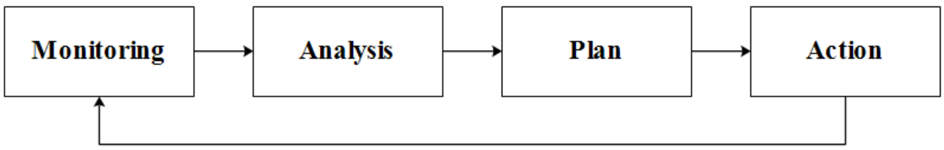


Figure 1.2 Life Cycle of CAASS

1.2.1 Monitoring

It is not possible to develop a software system without considering the various aspects of the environment where organizational and technical aspects are constantly changing (Pahl, 2004). The ability to tailor a software system and its requirements around its lifecycle in a constantly changing environment is of great importance (Pahl, 2004). These systems regularly monitor the domain events, detect significant changes, decide how to react, and act in order to execute the decision (Salehie & Tahvildari, 2007). These decisions can be on different levels and, accordingly, the structure of the self-adaptive software may change. For this, an adaptive system must be able to monitor the environment to determine the changes that have taken place and adapt itself to react to these changes (Pimentel et al., 2012). For monitoring, the participation of users is important but representative examples of intended target groups also should be introduced alongside; subjective observations must be accurately recorded and interpreted, and tests or questionnaires should be constructed using meaningful and valid heuristics (Charles et al., 2005). Fulton reports the difficulties of collecting satisfactory feedback and the difficulties in ensuring it (Bill Fulton & Medlock, 2002). Doing so, early in the development lifecycle can lead to a better design. (Pagulayan, Steury, Fulton, & Romero, 2003) pointed out that, despite high costs and no guarantee of success in the end, efficacy may be significant. The effects of the adaptation on the user can be easily tracked by observing the data, checking how fast a user is progressing, and watching the length of the sessions and other similar data (B Fulton & Romero, 2004). Hence, software compatibility can be defined to address environmental changes (Subramanian & Chung, 2001). Therefore, an adaptive system should be able to monitor the environment and determine the changes that are happening in it (Pimentel et al., 2012).

1.2.2 Analysis

Continuous participation of users is as important as continuous evaluation of the system and observation of changing environments (Pahl, 2004). Among these processes, decision-making is the critical one and a challenging process (McKinley et al., 2004). This is because, there are limited solutions considering variant requirements that need to be considered during the decision-making process (Salehie & Tahvildari, 2005). There are some proposed models to better analyze this dynamically changing data for a better decision-making such as the Box approach (Esfahani et al., 2013).

1.2.3 Plan

An adaptation plan should be generated during run time and should consider the actual state of the system, system elements and relationships among those elements (da Silva & de Lemos, 2011). This plan may need to be generated by considering the specific requirements of different application domains and each solution may be specific for the domain which may not work for other domains (da Silva & de Lemos, 2011).

Different types of information about a user can simultaneously be relevant to a given adaptation decision (Tamminen, Oulasvirta, Toiskallio, & Kankainen, 2004). Also, application should be equipped with the option to decide when and why to make an adaptive change (Di Nitto et al., 2008).

1.2.4 Action

A self-adaptive software evaluates its own behavior and changes this behavior when the evaluation indicates that it is not achieving what the software is proposed to do, or when better functionality or performance is possible (Filieri et al., 2014). These systems are expected to automatically change and improve their own behavior according to the domain knowledge available (Filieri et al., 2014).

Adaptation, then, consists of recognizing the current environment and selecting a configuration that meets the needs of the context (Laddaga, 2006; Laddaga & Robertson, 2004).

1.3 Different Applications of CAASS

Self-adaptive software architecture today is implemented at different levels for addressing various problems. For instance, some self-adaptive software aim to make adaptations on the application level (Salehie & Tahvildari, 2007) and adaptive business intelligence systems (Bäck, 2002). Self-adaptive software is useful for dealing with all forms of embedded software, including robotics (Karuppiah et al., 2000), manufacturing plants, avionics, vehicle control (Musliner, Goldman, Pelican, & Krebsbach, 1999), sensor systems (Reece, 2000), networking and others (Laddaga, 2006). It is also valuable for image and signal processing applications (Laddaga, 2006). Also, there are examples of CAASS for scheduling systems, operating systems, middleware, installation, configuration, system management and planning (Gajos, Weisman, & Shrobe, 2001). Additionally, self-adaptive software is an ideal framework for building pervasive computing systems (Gajos et al., 2001). Also there are other architectures that are designed for context-aware applications (Biegel & Cahill, 2004; Dey, 2001; Indulska & Sutton, 2003; Jameson, 2001; Lonsdale, Baber, Sharples, & Arvanitis, 2004; Petrelli, Not, Zancanaro, Strapparava, & Stock, 2001).

1.3.1 Instructional CAASS

The personalization of knowledge according to the characteristics of a learner is an important factor to better improve instructional systems (Lytras, 2007). As every learner has different characteristics, expectations, background and skills developed prior to the learning process, adaptation of instructional systems based on these individual requirements is critical. The context for adaptive instructional systems includes components such as learner, educational activity, infrastructure, and environment (A. A. Economides, 2009). In general terms, the context information relates to the environment, user or device status (Mizouni et al., 2014).

Content personalization has shown benefits for both users and content providers. Context-awareness can be interpreted as a natural development of personalization that takes the context not only of the user's identity and preferences, but also of the environment, including other users, their applications and their interactions (Ceri et al., 2007). Instructional content needs to be defined in terms of the score, duration, title, presentation type (e.g., skill-based simulation, game, interactive material, text, visual material, etc.), learning outcomes, difficulty level, learner group level definition (K-12, grade 2, Surgical education level 1 etc.) and sequence.

1.3.2 Examples of Instructional CAASS

A flexible e-learning model would take into consideration the learner’s knowledge state and learning preferences (Albano, Gaeta, & Salerno, 2006) to create personalized learning paths (Albano, Gaeta, & Ritrovato, 2007). There are studies conducted on adaptive game designs, arguing that player modeling and adaptive technologies can be used alongside the existing approaches to facilitate player-focused game designs developed to provide a more appropriate level of challenge, straighten the learning curve, and improve the gameplay experience independently according to gender, age and experience (Charles et al., 2005). For instance in a study, by concentrating on diversity in learning and play styles and associating them with personality profiles, it is shown that problems related to players’ age or gender can be avoided (Kerr, 2003). As every player is different, their preference for tempo and style of play also differs, and the ability to play varies widely among individuals (Charles et al., 2005). Even players with a similar level of game playing ability will often find separate aspects of a game to be more difficult for them personally, and the techniques that each player uses to meet the challenges offered by a game can also vary (Charles et al., 2005). This is why adaptive game technology can have an important role to play in next-generation games. This technology can be used to moderate the challenge levels for each individual player, prevent players from being stuck, and assist in further adapting to the preferences of a player (Charles et al., 2005). If all of these activities are carefully arranged, this can lead to an improved understanding of how players can get more satisfaction from a game (Charles et al., 2005). Also, the importance of maintaining the correct level in game-based learning environments is also highlighted (Csikszentmihalyi & Csikszentmihalyi, 1992): "The universal precondition for flow is that a person should perceive that there is something for him or her to do, and that he or she is capable of doing it. Optimal experience requires a balance between the challenges perceived in a given situation and the skills a person brings to it". It is also defined as the balance between difficulty and competence, or complexity and boredom (Charles et al., 2005). For this reason, one of the goals of adaptive design should be to keep the user in a state of flow by increasing the difficulty when the design appears too easy for the user, and decreasing it when it appears too hard (Charles et al., 2005).

Nevertheless, previous studies reported that, without supportive models that take into account the needs or difficulties of individual learners, students may only be of temporary interest in the learning process, and the learning performance is not as good as usually expected (Tseng, Chu, Hwang, & Tsai, 2008). Additionally, personalized learning content or navigation support is considered one of the most important features of educational systems (Tseng et al., 2008). Smart teaching systems are such personalized learning systems that help individual students to improve their learning performance based on personal information (Walonoski & Heffernan, 2006). Adaptive learning systems can be seen as a special type of intelligent teaching system that adapts the presentation of training materials to the needs of the students (Hwang, Sung, Hung, Huang, & Tsai, 2012). (Brusilovsky, 1998) stated that in the development of such systems, two adaptive approaches could be used, namely "adaptive presentation", which provides personalized content for individual learners, and "adaptive navigation support," which leads them to find the learning content with a personalized path.

1.4 A Proposed Framework for Developing CAASS

After analyzing previous studies about CAASS, it can be concluded that the scope and perspective of these systems are very wide. CAASS can be applied to any domain from different perspectives, and it is a complex process. Hence, the architectural design of these systems requires specific design strategies, which may also be specific for the domain that the adaptation is being planned to be performed.

We believe that, during the design stage of a CAASS, three stages need to be planned according to the questions posed in Figure 1.3.

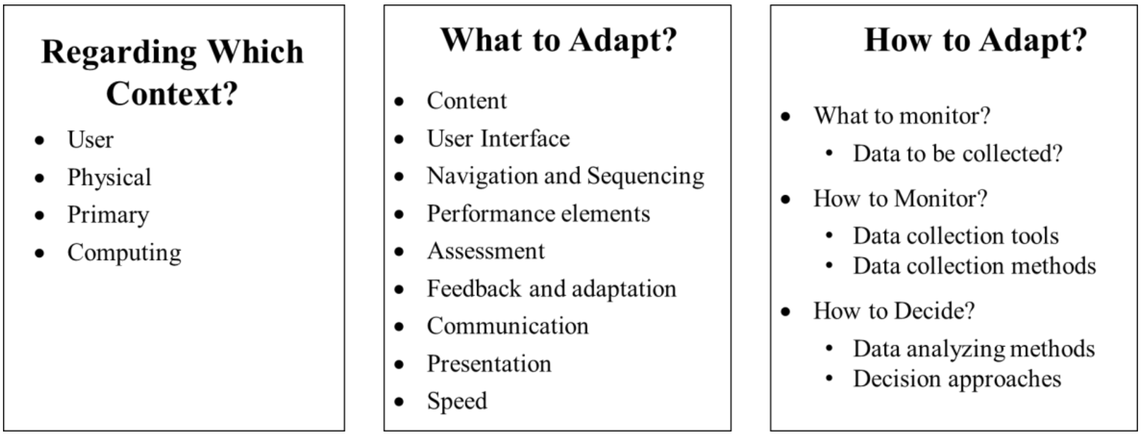


Figure 1.3 Proposed Framework for CAASS

In the first stage, decision has to be made regarding which context the adaptation is being planned for. This question has to be answered by considering the context elements. Depending on the adaptation objectives, different context variables can be targeted to be analyzed.

1.4.1 What to Adapt?

Based on the aim of the adaptation, several different software components could be considered for the adaptation. Some of them are summarized below:

<u>Content and Course Adaptation</u>

Today, numerous studies have been conducted on content and course adaptation (Brusilovsky & Vassileva, 2003; Healey, Hosn, & Maes, 2002; Tretiakov, 2004; Vassileva, 1998; Yau & Joy, 2007a, 2007b). For instance, the course content can be automatically generated specific to the learners by considering their individual goals, previous knowledge and skills, and adapt the course according to the learner progress (Brusilovsky & Vassileva, 2003; Vassileva, 1998). In an ideal scenario, an application is expected to set its behavior based on the current content of its use.

User Interface Adaptation

Researchers reported that information and user interface should be adapted to contextual features, such as user, environment, and access device (Viana & Andrade, 2008). Presentation adaptation (A. A. Economides, 2009) has also received some attention (Kelly & Tangney, 2006; Klett, 2005; Kurzel, Slay, & Chau, 2002; Vassileva, 1998; Wang, Li, & Chang, 2004). For instance, GTE (Generic Tutoring Environment) is adapted the presentation of the content (Vassileva, 1998). The content was presented in a variety of ways based on both students' prior competencies (pre-requisite knowledge and skills) and preferences (Kurzel et al., 2002). The presentation was adapted to facilitate learners' spatial reasoning on geometric topics (Wang et al., 2004). Multiple representations of complex or hidden subjects were also used (Klett, 2005). Different adaptive presentation strategies were used for students with different learning activities (Kelly & Tangney, 2006).

Navigation and Sequencing Adaptation

Several researches have been conducted on navigation and sequencing adaptation (Albano et al., 2007; Albano et al., 2006; Brusilovsky, Eklund, & Schwarz, 1998; Brusilovsky & Vassileva, 2003; Carchiolo, Longheu, & Malgeri, 2002; Eklund & Brusilovsky, 1998; Faraco, Rosatelli, & Gauthier, 2004; Herder & Van Dijk, 2002; Weber & Brusilovsky, 2001). Link annotation was adapted to the individual user in order to help them to find an appropriate path in a learning space (Eklund & Brusilovsky, 1998).

Assessment Adaptation

Assessment adaptation is an important area and several high-stake test organizations use computerized adaptive testing techniques. In computerized adaptive testing, if the examinee answers a question correctly, then the next question is selected from a more challenging pool; otherwise, from an easier pool (A. A. Economides & Roupas, 2007). Material for self-assessment was adapted to the needs of the individual learner (Katerina, 2004). The examinee's confidence in answering the question was also incorporated in adaptive testing (Lamboudis & Economides, 2004). It would be useful for the examinee to know his current status. The amount and timing of this orientation information revealed to the examinees would be adapted to his learning characteristics (A. Economides, 2005). A Computer Adaptive Testing (CAT) system on mobile devices was also developed and evaluated (Triantafillou, Georgiadou, & Economides, 2008).

Feedback Adaptation

Adaptive feedback can be provided to the examinee tailored to her/his needs (A. A. Economides, 2006). This way, the system can try to reduce anxiety during a test (A. A. Economides & Moridis, 2008). A model to measure the student's mood during a test was proposed (Moridis & Economides, 2008a) and validated through experimental data (Moridis & Economides, 2008b).

A learner model- Adaptive Remote Tutor- is also provided for adaptive navigation support, course sequencing, individualized diagnosis of student solutions, and example-based problem-solving support (Weber & Brusilovsky, 2001).

Communication Adaptation

Adaptive communication and collaboration would support learners from diverse cultural origins (A. A. Economides, 2008). Other adaptation approaches considered the users' preferences for informal communication and learning (Groth, Bogdan, Lindqvist, & Sundblad, 2007). Adaptive tools based on teacher's model for authoring, curriculum setting, co-teaching and privileges setting, reward setting, assessment setting and information sharing setting have also proposed (A. A. Economides, 2009; Lin, Young, Chan, & Chen, 2005).

1.4.2 How to Adapt?

In this stage, how the adaptation is being planned to be performed is needed to be decided. In order to establish an appropriate adaptation the software system needs to monitor and assess the context elements that are decided to be considered for the adaptation process. Based on analysis results on the collected data, an appropriate adaptation strategy is needed to be developed. In this stage which data will be collected, which data collection instruments will be used and how the collected data will be analyzed need to be planned.

We believe that this conceptual model for CAASS design provides certain measures about the scope of the specific CAASS. Additionally through this model, the level of adaptation also better evaluated. By applying this conceptual model to the CAASS, the researchers and developers as well as the market analyzers may better understand and compare the scope and level of adaptation of these systems.

The field of endo-neurosurgery education programs have several problems. The main problem of these programs is the skill-based training opportunities. As the training and skill development had to be provided in the operating room, there are several drawbacks of these education programs such as the ethical considerations from the patients' perspective, limited time and cases as well as the risk of patient safety. Currently, there are not many alternative training opportunities for the surgical training programs. The trainees do not have any chance of try-and-error type of learning. As the skill improvement is very critical for these programs, the individual skill-based training opportunities are required for these programs. Even there are some examples of computer-based simulations for supporting surgical training programs, there are very limited examples of curriculum integrated models. Additionally, there is no instructional model of CAASS for the surgical education programs especially for the endo-neurosurgery education programs. We believe that, because of its very nature, CAASS approach may provide several benefits for the endo-neurosurgery education programs.

As the process of creating CAASS for the field of endo-neurosurgery education programs is a very complex process, in this thesis, based on this proposed three stage framework for CAASS, a software requirements collection process is conducted. The methodology of the study is summarized in the next chapter. The findings of this thesis study is aimed to help future studies to better build CAASS for the field of endo-neurosurgery education programs and to better integrate these systems into the current educational programs.

CHAPTER 2

METHODOLOGY

A system can be defined as context-aware, if it uses context either for delivering content, or for performing system adaptations, or for doing both (Ceri et al., 2007). Accordingly, in this study for the development of an instructional CAASS, for surgical residents, their behavior patterns and experience levels were observed from their eye-movement data. The framework is applied as shown in Figure 2.1.

Regarding Which Context?	What to Adapt?	How to Adapt?
• <u>User</u> ➢ Skill level estimation *(CHAPTER 5)* ➢ Experience Level Behaviors *(CHAPTER 7)* • <u>Physical</u> ➢ Hand Condition *(CHAPTER 6)*	• <u>Content</u> ➢ Scenario Difficulty Level *(CHAPTER 4)*	• <u>What to Monitor?</u> ➢ Data to be Collected? ➢ Eye-Movement Events *(CHAPTER 3)* • <u>How to Monitor?</u> ➢ Data Collection Tools ➢ Eye Tracker Devices

Figure 2.1 A Case for Instructional CAASS based on Eye-movements

This study first shows the value of data collected through eye-tracker devices to better build an adaptation process for the CAASS. Eye-tracking device is used to monitor the user behaviors. Eye-movements of surgical residents were recorded with the eye-tracking device. The recorded data is classified into various eye-movement events by open-source eye-movement classification algorithms. As there are several algorithms available to be used for the eye data classification, the classification results of these algorithms are evaluated in the context of surgical training programs. In Chapter 3, the results of these evaluations are given.

In the second phase, the relationship between the mental workload of the participants and the recognized difficulty levels of the content analyzed. Accordingly, as described in Chapter 4, these difficulty levels of the tasks are attempted to be understood through the eye-movement events of the participants. This information is critical to better sequence the content according to the user skill and knowledge level from easier to the harder ones.

In the next stage (Chapter 5), as a context parameter, the user behaviors for the CAASS development is evaluated and how eye-movement events of the users can be used to estimate their skill levels. Here our main assumption is that, by better estimating the skill levels of the users, appropriate content can be better adapted.

As the eye-tracker device records the pupil sizes, through this data, under different hand conditions (dominant-hand, non-dominant hand and both-hand) the behaviors of the surgeons are analyzed. In Chapter 6, these results are provided.

Finally, the behaviors of different experience groups are analyzed through their eye-movement events. This information is believed to be very helpful to adapt the content according to various requirements of different skill level groups. In Chapter 7, details of these results are provided.

As a result this study shows some examples of monitoring, data collection and assessment procedures for creating CAASS for the surgical education programs. We believe that, by understanding three main parts of the CAASS namely context, what-to-adapt and how-to-adapt dimensions, better adaptation algorithms can be developed. In this case, before generating an adaptation algorithm for a CAASS for the endo-neurosurgery education programs, more research is required to understand the behaviors of different skill level groups by collecting other sources of data such as hand-movement behaviors and performance data.

In this study, four different simulation scenarios have been developed for collecting data about surgeons' eye gaze during operations in a virtually simulated environment and performed in different hand conditions. The environments resemble the real world with their visualization and interaction properties (X. Zhang, Jiang, Ordóñez de Pablos, Lytras, & Sun, 2017). These scenarios are developed based on the surgical skill development requirements for endoscopic surgery purposes. As it has been reported there are several potential benefits in developing a gaze-focused approach to understand surgical skills learning and performance (Hermens, Flin, & Ahmed, 2013). Detailed data is also collected in this environment with regard to eye gaze.

Earlier studies report that as their skills on dominant-hand and non-dominant hand are different, surgeons' performance under these different conditions do also vary (Hoffmann, 1997), and that skilled surgeons' hand performances are more stable than the non-experienced ones (Uemura et al., 2014). The main assumption of this study is that surgeons' performances in different hand conditions and their skill levels may also affect their eye-movements. Therefore, we aim to better understand, the differences between the eye-movements of intermediate and novice surgeons while they perform surgical tasks in a computer-simulated environment, and the effect of hand condition (dominant-hand, non-dominant hand and both-hand) on their eye-movements. The results are analyzed using statistical methods to better understand the novice and intermediate surgeons' behaviors and hand condition effects in this environment.

2.1 Participants

A total of 23 participants from neurosurgery and Ear-Nose-Throat (ENT) surgery departments of Hacettepe University Medical School in Ankara, Turkey, voluntarily participated in this study. As it is difficult to access surgeons of a specific field to be volunteer for such initiatives, this number of participants can be usually considered as acceptable. For this reason, studies in this field were conducted with limited number of participants: (M. Wilson et al., 2010) (14 surgeons), (Vine, Masters, McGrath, Bright, & Wilson, 2012) (27 novices), (Uemura et al., 2016) (26 surgeons), (M. R. Wilson, McGrath, et al., 2011) (25 surgeons), (Eivazi et al., 2017) (9 neurosurgeons), (Cope, Mavroveli, Bezemer, Hanna, & Kneebone, 2015) (22 UK surgeons), (J.-Y. Zhang, Liu, Feng, Gao, & Zhang, 2017) (14 participants) and (Zheng, Jiang, & Atkins, 2015b) (14 novices).

Table 2.1 Participant Information

		n	%
Gender	Female	3	13.0
	Male	20	87.0
Wearing Glass	No	17	73.9
	Yes	6	26.1
Surgical Experience	Novice	14	60.9
	Intermediate	9	39.1
Dominant-hand	Right	20	87.0
	Left	3	13.0

A majority of the participants were male (87.0%) and did not use glasses (73.9%). As seen from Table 2.1, the dominant hand of the majority is the right hand (87.0%).

(Silvennoinen, Mecklin, Saariluoma, & Antikainen, 2009) defined the expertise and skill levels in minimally invasive with novices who had begun to gain basic knowledge of endoscopic surgery and an intermediate group that had just started endoscopic surgery operations. Among those 23 participants, 14 of them (2 female) were novices whose average age was 27.71 (SD = 6.96) and who worked as research assistant in the Neurosurgery or ENT departments. None of them had previously performed an endoscopic surgery by themselves. As seen in Table 2.2, on the average, they had observed 9.57 (SD = 13.51) and assisted in 3.57 (SD = 10.64) surgeries.

Table 2.2 Participant Endoscopic Surgery Experience

Participant	Age		Observed		Assisted		Performed	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Intermediate	29.33	1.50	48.33	31.62	32.00	24.19	16.56	16.60
Novice	27.71	6.96	9.57	13.51	3.57	10.64	0.00	0.00

On the other hand, 9 participants (1 female) were intermediates whose average age was 29.33 (SD = 1.50). On average, they had observed 48.33 (SD = 31.62) and assisted in 32.00 (SD = 24.19) surgeries. On average, the intermediate group had performed 16.56 (SD = 16.60) operations as surgeons. The participants' experience levels are varying according to the operations they monitored, assisted and performed and Table 2.2 represents the average number of surgical operations that these surgeons had carried out.

2.2 Apparatus

Four scenarios developed for recording the surgeons' eye-movement data were used in this study. These scenarios were implemented as part of a Tubitak supported research project (ECE: Tubitak 1001, Project No: 112K287). The eye-movements of the surgeons were collected with an eye-tracker during the task performed in different hand conditions by haptic devices. Eye-tracking is the process of using sensors to locate features of the eyes and to estimate where the subject is looking. The data was recorded using Eye Tribe Eye Tracker ("The Eye Tribe," 2017) at 60 Hz with a screen resolution of 1920×1080 pixels (Figure 2.2). The Eye Tribe is a Danish start-up company that produces eye-tracking technology and offers the product to software developers to be incorporated into different applications and programs. The company focuses on a sleek appearance and a portable structure. The Eye Tribe Eye Tracker is an affordable device, thereby making it a potentially available tool for research. According to the (Coyne & Sibley, 2016) the Eye Tribe system is able to significantly differentiate pupil size differences in high and low workload trials and the results are quite promising for human factors researchers.

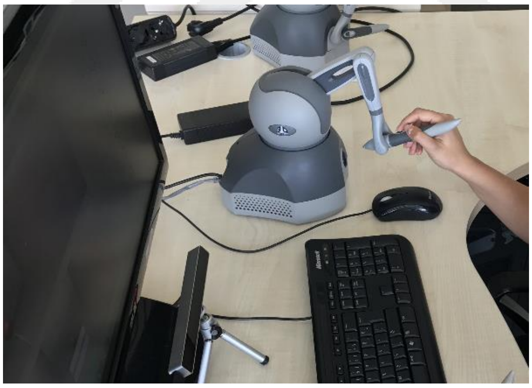


Figure 2.2 Collection of the Data

Haptic technologies offer interaction between people and machines (Ucar, Ustunel, Civelek, & Umut, 2017). Therefore, to perform the tasks required in this research, the Geomagic Touch mid-range professional haptic device is used ("Three-D Systems," 2017). 3D-systems haptic devices present true 3D navigation and force feedback. These device are used in research and 3D modelling. Geomagic Touch gives users opportunity to enhance scientific or medical simulations, improve productivity with interactive training, and easily maneuver mechanical components to produce higher quality designs.

2.3 Scenarios

There were four scenarios developed for simulating different surgical training purposes. The learning outcomes of the scenarios are increasing the 3D perception, gaining depth perception, using the endoscope efficiently, fast-following up of objects, and improving the ability to plan and strategize. Each scenario was performed by surgeons with their dominant-, non-dominant and both- hands. To provide more objectivity and getting rid of the order effect, 12 of the participants were started to perform the tasks by their dominant-hand, and the rest did so with the non-dominant one. Current technologies allow the recreation of real-life operations with adequate fidelity, thus profoundly improving the training environment (Munshi, Lababidi, & Alyousef, 2015). Accordingly, in this study two of the scenarios were simulated as surgical model and can be considered as higher-fidelity; the other two were based on general models which can be considered as lower-fidelity.

Virtually simulated environments resemble the real world with their visualization and interaction properties (Perrenot et al., 2012; X. Zhang et al., 2017). Therefore, surgical simulation scenarios were developed based on surgical skill development requirements for endoscopic surgery purposes. Scenarios were designed to provide a practical alternative for the endoscopic surgery beginners. Hence, the participants were expected to use an operational instrument by their dominant-hand through a haptic device.

<u>Scenario-1</u>

In this scenario, the endoscope is not controlled by the participants. The light source that is in real environment controlled by endoscope is simulated in a fixed position and the camera that is controlled by the endoscope in real environments is coordinated through the surgical instrument in the simulated environment. The operation is conducted in a simulated environment using a surgical instrument under virtually lit conditions. Hence, the scenario is designed to improve the participants' skills on depth perception, camera control, 2D-3D conversion and effective instrument usage. With the help of this instrument, the participants have to catch a red ball in a room. The red ball can appear randomly in different locations (Figure 2.3).

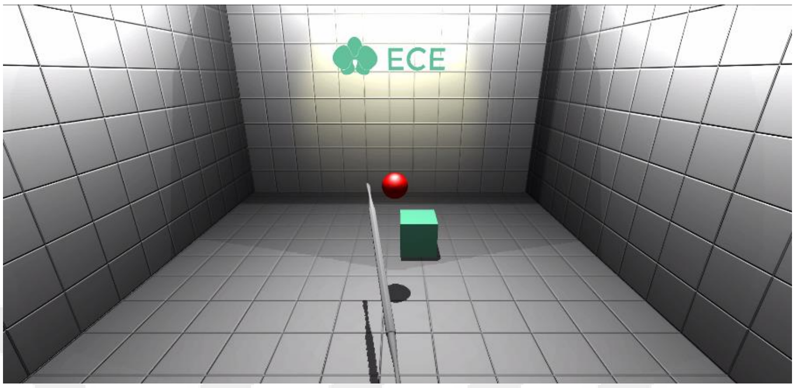


Figure 2.3 Scenario-1: A

As seen from Figure 2.4, after catching the ball and moving it close to the cube, the ball turns green. The participants have to match this ball with a green cube. The cube itself may also appear at random positions in the room. This task is repeated 10 times and the participants' eye-movement data is collected with an eye-tracking device. This scenario is a general simulation model aimed to gain the ability to use the surgical instrument and to develop depth perception.

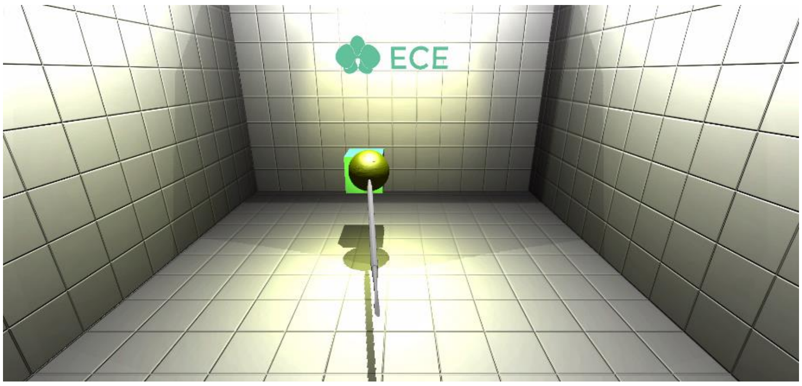


Figure 2.4 Scenario-1: B

Scenario-2

In this scenario, a model is designed based on the inside of a human nose, as depicted in Figure 2.5, which contains tumor like objects. The participants were expected to remove the tumors located at different spots within the model using a surgical tool. There are 10 tumors located in this model. This scenario is a simulated surgical model, which has made it possible for surgeons to feel as if they are in surgical settings. Surgeons can move the endoscopic device through the nose using the haptic device and feel the tissue as the device give force feedback upon collision with any surface. By using the surgical tool in the most accurate way, it is expected to complete the operation by carefully removing the tumors from their locations.

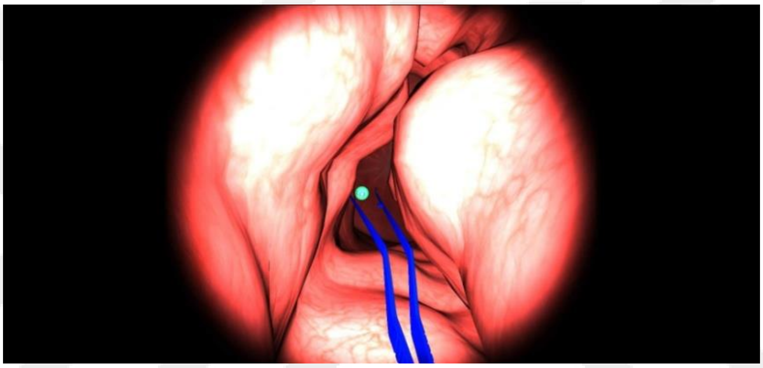


Figure 2.5 Scenario-2

Scenario-3

In this scenario, red balls appeared randomly in one of the boxes in a virtual environment to be focused on as seen in Figure 2.6. Once properly focused on, the red balls which appear randomly in the blue boxes would explode on the condition that focusing is done with the right angle. If the correct angle is achieved, the ball will explode; otherwise it will not. This process is also repeated 10 times. In this scenario the aim is to develop depth perceptions and improve ability to approach a certain point with the correct angle. This scenario is a simulation of a general model.

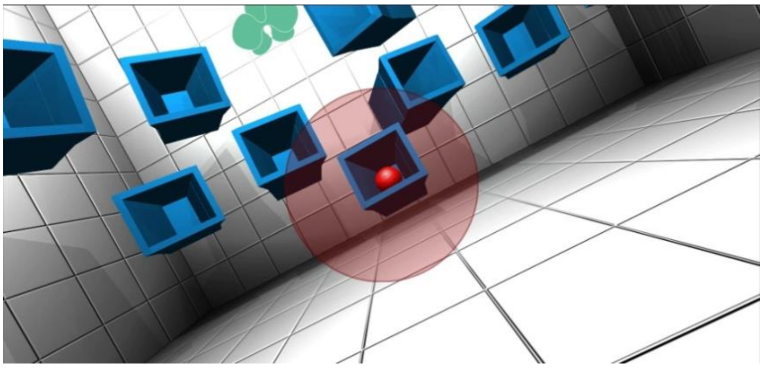


Figure 2.6 Scenario-3

Scenario-4

The aim in this scenario is to continuously follow a moving object (white ball) on a path inside a nose model environment as seen in Figure 2.7. The main task is to complete the route by following the white ball which appears at the beginning of the path within a proper distance and angle. This scenario is a simulated surgical model and designed like a real nose with similar texture, simulating the field vision of a surgeon during an actual operation.

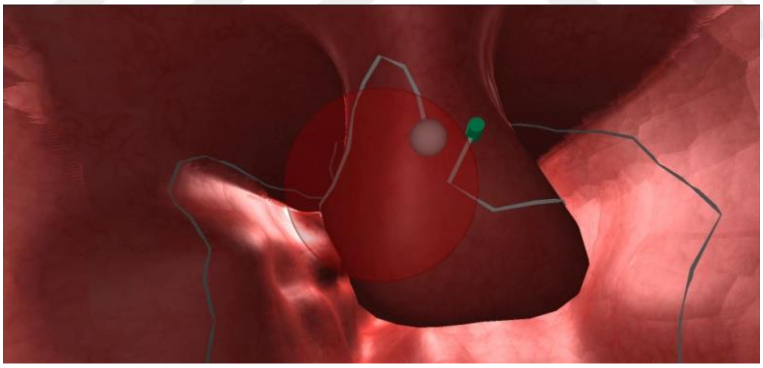


Figure 2.7 Scenario-4

2.4 Procedure

First, general information regarding the participant and his/her dominant hand were recorded. Each participant was seated and centered in front of the monitor at a distance of 70cm and an oral instruction was provided regarding the procedure. Nine calibration points were presented, and calibration is performed. After that, according to the given information, the scenarios would be performed by the participant with dominant- and non-dominant hands. After the completion of one-handed scenarios, two-handed scenarios started, in which the participants were asked to perform the same scenarios using both-hands. The dominant-hand used the operation tool and the non-dominant hand used the camera tool for lighting up the operation area. The scenarios were performed as 1, 3, 2 and 4 representing the scenario numbers as outlined in the preceding section.

The recorded data was classified using an open-source eye-movement classification algorithms (Appendix-A). There are many other commercial classification algorithms, but in this study only the open-source eye-movement classification algorithms were used. Ten open-source eye-movement algorithms were utilized for classifying the eye data as fixations, saccades and pursuits which are defined in the following section.

Then, for the evaluation of the differences based on experience, the events namely number of fixation, fixation duration, saccade number, saccade duration, pursuit number, and pursuit duration were used. These eye-movement events used to distinguish the experience levels of surgeons. The algorithms are based on velocity or position to differentiate the eye-movement events.

The Hidden Markov Model Identification (I-HMM), Velocity Threshold Identification (I-VT), and Kalman Filter Identification (I-KF) are velocity-based algorithms (Komogortsev & Karpov, 2013). The Minimum Spanning Tree Identification (I-MST), and Dispersion Threshold Identification (I-DT) are position-based algorithms (Andersson, Larsson, Holmqvist, Stridh, & Nyström, 2017). These algorithms are used to classify fixations and saccades, while the others are used for smooth pursuit events such as Velocity and Velocity Threshold Identification (I-VVT), Velocity and Movement Pattern Identification (I-VMP) and Velocity and Dispersion Threshold Identification (I-VDT) (Andersson et al., 2017).

Additionally, there are two more algorithms used for eye-movement classifications. The Adaptive Event Detection algorithm (AED) proposed by Nyström and Holmqvist (Nyström & Holmqvist, 2010) is a velocity-based algorithm and developed for classifying fixation and saccade eye-movement types. Binocular-Individual Threshold (BIT) algorithm is developed by (van der Lans, Wedel, & Pieters, 2011) is a velocity-based algorithm to classify fixations from the data with individual-specific thresholds.

2.5 Eye Movement Events

In recent psycho-physiological researches, the eye-tracking technique has been widely used to obtain reaction parameters from eye-movement data to analyze cognitive processes underlying visual behavior (Bailey & Iqbal, 2008). Eye-tracking provides a valuable source of physiological data for the allocation of information processing resources through ocular activity, which are closely linked to the underlying neural networks in the brain (Bröhl et al., 2017). To understand surgeons' behaviors while they were performing computer simulated surgical tasks, under different hand conditions, specific eye movement events were used. These are fixation, saccade, smooth pursuit events and pupil size.

<u>Fixation</u>

Fixation is a slow period event when the eye-movement is almost still with small dispersion and velocity. Eye-movement classification algorithms can be able to classify fixation events into number of fixation and fixation duration.

<u>Saccade</u>

When the eye makes fast movements between various locations, this is called 'saccade'. Saccade numbers and durations are classified by eye-movement algorithms.

<u>Smooth Pursuit</u>

Eye-movements are generally used for static stimuli; for dynamic stimuli, an event called 'smooth pursuit' occurs while tracking a dynamic object, and the pursuit number and pursuit duration measures are used to observe the differences based on experience and hand conditions.

Pupil Size

There is extensive evidence as to the relationship between pupil size and cognitive functions (Bailey & Iqbal, 2008; Beatty, 1982; Bröhl et al., 2017; Just, Carpenter, & Miyake, 2003; Pomplun & Sunkara, 2003; Verney, Granholm, & Marshall, 2004). Hence, pupillary response is known as a measure of cognitive activity and attention (Bailey & Iqbal, 2008). Pupil size has been suggested as a metric for assessing workload during complex visual tasks, which are associated with higher pupil dilation (Bailey & Iqbal, 2008). Consequently, pupil size variances are taken into consideration and their relation with experience and mental workload was examined.

2.6 Analysis Method

Data analysis was performed using SPSS for the Windows software package (Version 23; IBM Corporation, New York, USA) with 95% confidence level. Because the normality assumptions are violated and the sample size is 23, the non-parametric Mann Whitney and Friedman test techniques were used in this study (McCrum-Gardner, 2008). Mann Whitney is a test technique used to compare two independent groups in terms of a quantitative variable and is used for observing the differences between intermediate and novice surgeons. Friedman is an alternative to bi-directional variance analysis and is used to determine the difference between two major masses (McCrum-Gardner, 2008). This test technique is used to observe the effect of hand conditions (dominant, non-dominant and both) on eye-movement events among surgeons. The logistic regression analysis was conducted in this study to predict of the skill levels of surgeons. Hierarchical clustering method is applied to better understand if there is a clustering among the classification results of 10 classification algorithms utilized in this study.

CHAPTER 3

EVALUATION OF TEN OPEN-SOURCE EYE-MOVEMENT CLASSIFICATION ALGORITHMS

With today's eye-movement tracking technology several benefits for various studies can be reached. Hence, eye-movement classification is widely used for many fields such as neurology, psychology, ophthalmology and in commercial areas and classifying eye-movements is crucial for understanding visual attention and provides evidence regarding certain brain states and psychological functions (Bedell & Stevenson, 2013; Komogortsev, Gobert, Jayarathna, Koh, & Gowda, 2010). Later, these evidences can be used for diagnoses, treatment and training purposes (Jarodzka, Holmqvist, & Gruber, 2017). Eye-movement classification is also important for clinical applications such as Alzheimer's (Crawford et al., 2005), HIV-1 infected patients with eye-movement dysfunction (Sweeney, Brew, Keilp, Sidtis, & Price, 1991), and schizophrenia (Bolding et al., 2014; Flechtner, Steinacher, Sauer, & Mackert, 1997). Furthermore, they are applied in commercial purposes such as Web-page navigation, shopping and human computer interaction (Larsson, Nyström, & Stridh, 2013).

The classification of basic eye-movement events such as fixations, saccades and pursuits from noisy eye data is essential for researchers, who use eye-trackers for recording eye data in their studies. Fixations are the slow periods when the eye is nearly still. Saccades occur when the eye makes rapid shifting movements between different positions. If research is carried out using static stimuli, then these two eye-movement events are the most commonly used ones. When using dynamic stimuli, such as moving objects in a dynamic scene, the eyes will track these objects. In this case, smooth pursuit eye-movement event occurs. It is up to researchers whether manually identify these eye-movement events or use any of the commercially or freely available algorithms for classification but working on this issue manually is very sluggish compared to classification algorithms, so, today classification algorithms are the only practical solution for classification (Andersson et al., 2017).

In the literature, there are not many studies conducted to comparatively evaluate these algorithms. The evaluation of these algorithms is done based on eye-movement data collected from 23 surgeons while they were performing 4 different virtual-simulation scenarios in a 3D dynamic and interactive environment by using their dominant-, non-dominant and both-hands successively.

3.1 Current Eye-Movement Classification Algorithms

Nowadays, several eye-movement classification algorithms are being used; however, as some are not open-source or are only commercially available as part of software suites made by companies developing eye-tracking devices, it is not possible to evaluate them. Consequently, because of this situation researchers today prefer to use open-source eye-movement classification algorithms. The selected algorithms are chosen according to the criteria being up-to-date and independent of any eye-tracker device. Accordingly, in this study 10 different open-source, up-to-date eye-movement classification algorithms are used to make evaluations. The investigated algorithms in this study were implemented with MATLAB.

Eye-movement classification algorithms classify eye data by considering different features that are gathered by eye-tracker devices. There are different types of algorithms to classify events such as fixations, saccades and smooth pursuits. These types are separated as velocity-based algorithms and dispersion-based algorithms. Velocity-based algorithms are Velocity Threshold Identification (I-VT), Hidden Markov Model Identification (I-HMM), Kalman Filter Identification (I-KF) and the dispersion-based algorithms are Dispersion Threshold Identification (I-DT), Minimum Spanning Tree Identification (I-MST) (Komogortsev & Karpov, 2013).Velocity and Velocity Threshold Identification (I-VVT) algorithm is the modified version of the I-VT algorithm to identify smooth pursuits from fixations, Velocity and Movement Pattern Identification (I-VMT) firstly utilizes velocity threshold for classifying saccades, same as I-VVT algorithm, and investigates the eye-movement samples to split smooth pursuits from fixations, Velocity and Dispersion Threshold Identification (I-VDT) is used as a ternary classification algorithm (Komogortsev & Karpov, 2013) and also this algorithm is a modified combination of velocity-based and dispersion-based algorithms.

Velocity Threshold Identification (I-VT) is based on the velocity of the eye-movements to separate fixation events from saccade events and works with a velocity threshold for classifying the fixations and saccades and if the samples velocities are below the threshold, then the algorithm defines these samples as fixation; and if the sample velocities are higher than the threshold, than the algorithm set that sample a saccade and this velocity threshold principle is the basis for other algorithms (Komogortsev & Karpov, 2013). Velocity and Velocity Threshold Identification (I-VVT) is designed to classify fixations, saccades and smooth pursuits and contains a filter function which filters noisy saccade-like events according to minimum amplitude and duration (Komogortsev & Karpov, 2013).Velocity and Movement Pattern Identification (I-VMT) examines the movement patterns to detach smooth pursuits from fixations and this movement pattern is examined in a temporal window. In which the magnitude of movement is computed by analyzing angles created by every pair of the adjacent positional points and the horizontal coordinate axis, after the value representing the magnitude of the motion is calculated, a threshold is compared; the values above the threshold are marked as smooth pursuit and the below values are marked as fixation (Komogortsev & Karpov, 2013).

Velocity and Dispersion Threshold Identification (I-VDT) is also used for the classification of fixations, saccades and smooth pursuits. Similar to the algorithms I-VVT and I-VMT, it separates the saccades firstly, then, smooth pursuit dimensions are separated from fixations; what using an adapted I-DT method (Komogortsev & Karpov, 2013). The Hidden Markov Model Identification (I-HMM) similar to the I-VT algorithm, the I-HMM is based on velocity and has two additional algorithms; the first one is re-classifying the fixations and saccades according to probabilistic parameters, and the second one updates the parameters (Komogortsev, Gobert, et al., 2010). Kalman Filter Identification (I-KF), is a recursive predictor that computes future states with estimating a series of dynamic system states from noisy measurements (Komogortsev, Jayarathna, Koh, & Gowda, 2010). Because the actual data is usually noisy and may cause data loss, the Kalman Filter minimizes the error between the state of the system and the state of the real system, only the previous time step and the estimated new condition are required to calculate the new situation estimate (Komogortsev, Jayarathna, et al., 2010). If the value is less than the set threshold and the minimum time threshold is met, it is specified as a fixation, and if it is above the threshold value, it is specified as a saccade (Komogortsev, Jayarathna, et al., 2010).

One of the dispersion-based algorithms, Dispersion Threshold Identification (I-DT), is commonly used for classifying eye-movements into fixations and saccades and this algorithm uses x and y coordinates of the eye and thresholds for classification, namely maximum fixed dispersion threshold and minimum fixed time threshold (Komogortsev, Gobert, et al., 2010). Minimum Spanning Tree Identification (I-MST), the other dispersion-based algorithm, creates a minimum spanning tree by taking a predetermined number of eye position points (Salvucci & Goldberg, 2000). MST is defined as a spanning tree with the least distance between all spanning trees in this node set, I-MST breaks the MST into determinations and thresholds based on predetermined distance thresholds (Salvucci & Goldberg, 2000). The advantage of using an I-MST is that the algorithm can correctly identify the anchor points if a large part of the signal is missing, as a result of this property, I-MST is claimed to be a highly flexible and controllable eye-movement detection tool (Salvucci & Goldberg, 2000).

An adaptive event detection algorithm (AED) proposed by (Nyström & Holmqvist, 2010) is an algorithm that tends to alter the velocity threshold based on the noise level of the subject, also, it describes post-saccadic releases which are defined as 'glissades' with fixation and saccades. As a different eye-movement 'glissade' is a wobbling movement at the end of saccades, as a different type of eye-movements, the AED algorithm is a velocity-based algorithm which is developed for classifying fixation, saccade and glissade eye-movement events and provides graphical representations of the these events (Nyström & Holmqvist, 2010). Binocular-

Individual Threshold (BIT) algorithm (van der Lans et al., 2011), is a velocity-based algorithm to classify fixations from the eye-movement data of both eyes with individual-specific thresholds. To verify fixations, the algorithm uses the velocity thresholds of both eyes. The BIT algorithm has advantages over the existing algorithms in that it contains binocular viewing and uses the information about fixations and co-variations between the movements of both eyes to identify saccades; it estimates rather than pre-sets the velocity threshold to identify fixations and saccades, and it permits the threshold to vary between eye-movement directions, tasks and individuals. Also, each record exceeding the threshold value contains the stochasticity which is spontaneous in the eye-movements so as not to be labeled as saccade (van der Lans et al., 2011). The other important feature is that BIT algorithm is independent of machine and sampling frequency and can be easily adapted to the data from varying eye-trackers with different sensitivity and sampling frequency (van der Lans et al., 2011).

3.2 The Limitations for Eye-Movement Classification Algorithms

The need for a single algorithm used in all systems and the presence of many algorithms addressing the same problem implies that eye-movement classification is not a mundane issue and assessing the performance of different algorithms is an important undertaking (Andersson et al., 2017). Notably, selecting the most appropriate from amongst means that a thorough evaluation method has to be designed. As this study is not the first one attempting to evaluate algorithm performance, it is necessary to consider the benefits and drawbacks of the previously established methods.

One of the approaches has been to establish an optimal or rational relationship between stimuli and an individual's viewing behavior. For instance, researchers offered a trial to the participants in order to look at a single moving target that jumped a number of times (Komogortsev, Gobert, et al., 2010). Given the known number of jumps, positions, and amplitudes, it is possible to calculate how the ideal eye-movement behavior looks; then, the gaze data parsed by the algorithms is compared with this ideal viewing pattern, and the more similar the algorithm is, the better (Komogortsev, Gobert, et al., 2010).

The detection of eye-movement events is not a completely solved problem, because there is no consensus on how to evaluate the algorithms, which means that further refinement of the algorithms is hindered as it is not clear whether differences are due to the algorithms or the evaluation process itself (Andersson et al., 2017). What is more, it is not clearly known what is meant when we talk about an event; for example, there is no theoretically motivated threshold for the eye to be classified as a saccade to sufficiently move in a particular direction or to classify anything under it as another event. Classification algorithms focus on a rigorous oculomotor definition of fixations and saccades. Even in the definition of a fully oculomotor eye-movement, it is difficult to identify the end point of the fixation and the start point of a smooth pursuit. This point is arbitrary and more or less determined by the sensitivity of the system. Many algorithms have some form of adjustments that need to be set by the researcher, such as minimum fixation time, saccade speed threshold, and so on. If there is an explicit and theoretically applied threshold, then in this case it will already be coded as a 'constant' in the algorithm. Setting the thresholds depends on the researcher. There could be new results that deviate from previous ones, selected algorithms, selected thresholds, or both, or something else entirely. It is common knowledge that different parameter values for these algorithms produce different classification results (Andersson et al., 2017). There are many algorithms, which have not been compared to one another sufficiently. Often, a modest assessment is made while presenting a new algorithm, but it usually examines only a few algorithms and mainly highlights the new algorithm.

3.3 Threshold Values of Eye-Movement Classification Algorithms

The evaluated algorithms were used with small changes in their default settings. An ideal algorithm does not require any parameter setting from the user, and automatically adjusts the thresholds: and then, it categorizes all samples of the data stream in a comprehensive way (Andersson et al., 2017). There are same parameter values for all algorithms these are screen size value (1920x1080), distance to the screen (70cm) and sampling frequency (60Hz). In the following section all other parameters for each algorithm are listed.

3.3.1 Velocity Threshold Identification (I-VT)

Velocity Threshold Identification (I-VT) algorithm uses the saccade detection threshold for classifying saccades and fixations. This parameter, is set to 70°/s by default. If the movement speed from one eye position to the next is below this value, it means that the eye-movement belongs to fixation; otherwise it is a saccade.

3.3.2 Velocity and Velocity Threshold Identification (I-VVT)

Velocity and Velocity Threshold Identification (I-VVT) algorithm uses two threshold parameters; one is saccade detection threshold and the other is fixation detection threshold. The saccade detection threshold adjusts the value of the speed threshold used to distinguish between saccades and fixations. This parameter was set to 70°/s by default. If the movement speed from an eye position to the next is greater than this value, it means that the eye-movement belongs to the saccade; otherwise, it is a fixation or smooth pursuit.

The determination should also be done with another threshold. Fixation detection threshold; which sets the value of the speed threshold used to distinguish between fixations and smooth pursuits. If the speed of the movement from one eye position to the next is larger than this value, it means that the eye-movement is a smooth pursuit; otherwise, it is a fixation. This parameter was set to 20°/s by default.

3.3.3 Velocity and Movement Pattern Identification (I-VMT)

The Velocity and Movement Pattern Identification (I-VMT) algorithm uses three threshold parameters, namely saccade detection threshold, temporary window length and range threshold value. The saccade detection threshold field on the I-VT tab adjusts the value of the speed threshold used to determine the difference between saccades and fixations. This parameter was set to 70°/s by default. If the movement speed from an eye position to the next is greater than this value, it means that the eye-movement is belongs to saccade; otherwise, this is either a fixation or a smooth pursuit. Determination should be done in the next stage of classification. The temporary window length field specifies how much time is needed for dispersion calculation during data processing. This parameter was set at 0.5 by default. The range threshold value sets the threshold for the distribution of the selected samples in a range of 0 to 1. This parameter was set at 0.1 by default.

3.3.4 Dispersion Threshold Identification (I-DT)

Dispersion Threshold Identification (I-DT) algorithm uses two threshold parameters, first one is dispersion duration threshold, and the second is dispersion threshold. The dispersion duration threshold specifies how much time we must use to calculate the distribution during data processing. This field sets the threshold value for the distribution of the selected samples in degrees. If the distribution is less than this value, it means that fixation has been detected; otherwise, it is a saccade. The default values for this classification were set at 100ms as dispersion duration threshold and 1.35° as the dispersion threshold.

3.3.5 Velocity and Dispersion Threshold Identification (I-VDT)

The Velocity and Dispersion Threshold Identification (I-VDT) algorithm uses three threshold parameters as, saccade detection threshold, dispersion duration threshold and dispersion threshold. The saccade detection threshold on the I-VT tab adjusts the value of the speed threshold used to determine the difference between saccades and fixations. If the movement speed from an eye position to the next is greater than this value, it means that the eye-movement is a saccade; otherwise, this is either a fixation or a smooth pursuit. Determination should be done in the next stage of classification. This parameter, saccade detection threshold was set at 70°/s by default. The dispersion duration threshold in the I-DT tab specifies how much time is needed to calculate the distribution during data processing. This parameter was set at 100ms by default. The dispersion threshold field in the I-DT tab sets the threshold value for the distribution of the selected samples in degrees. If the distribution is less than this value, it means that a fixation has been detected; otherwise, it is a saccade. This parameter was set at and 1.35° by default.

3.3.6 Hidden Markov Model Identification (I-HMM)

The Hidden Markov Model Identification (I-HMM) algorithm uses three threshold parameters, saccade detection threshold, Viterbi sample size and Baum-Welch reiteration. The saccade detection threshold is identical to the I-VT classifier. The Viterbi sample size specifies the number of samples the classifier uses as a data set. If the threshold value set too high, it means that there is not enough machine precision to calculate the statistical parameters. The Baum-Welch reiteration specifies the number of iterations of the Baum-Welch algorithm. It makes sense to set this value equal at 4 or 5 reiteration. Thresholds were set to the default values in this way as saccade detection threshold at 70°/s, Viterbi sample size at 200 and Baum-Welch reiteration at 5.

3.3.7 Kalman Filter Identification (I-KF)

The Kalman Filter Identification (I-KF) algorithm uses three threshold parameters, Chi threshold, sampling window size and deviation. The value for $\chi 2$-distribution threshold set by the Chi threshold field. If the $\chi 2$-distribution values are lower than this threshold, than a fixation has been detected, if not, it is a saccade. The sampling window size specifies the number of samples for which the $\chi 2$-distribution is calculated. The deviation field sets the deviation value between the anticipated and computed values for the $\chi 2$-distribution calculation. The parameter thresholds are set as defaults, namely chi-square threshold of 15, a window size of 5 samples, and a deviation value of 1000.

3.3.8 Minimum Spanning Tree Identification (I-MST)

The Minimum Spanning Tree Identification (I-MST) algorithm uses two threshold parameters, saccade detection threshold and window size. The saccade detection threshold adjusts the distance between two adjacent eye focus positions in degrees. If this distance is less than the threshold, it means that the eye-movement is a fixation.

If this distance is higher than the threshold value, it means that eye-movement is a saccade. The window size field sets the number of instances the classifier uses during data processing. If it is set too high, the sensing time for the saccades can be significantly increased. It is reasonable to set this value slightly lower than the average fixation duration. The parameter thresholds were set as defaults 0.6° for saccade detection threshold and 200 samples for the windows size parameter.

3.3.9 An Adaptive Event Detection Algorithm (AED)

The default values were used except the minimum saccade duration, which was set at 20ms because the algorithm was design for 1250Hz of data with high precision the window length (F) didn't scale satisfactorily when using 60Hz data. Therefore, increasing F to the nearest odd integer (F=3) was sensible. To do this, there was also a need to increase the value of the minimum saccade duration from 10ms to 20ms.

3.3.10 Binocular-Individual Threshold (BIT)

This algorithm is a parameter-free fixation identification algorithm that automatically identifies relative and individual-specific speed thresholds by optimally using statistical properties of eye data.

3.4 Results

In this study first the common eye-movement events of those 10 algorithms were analyzed (Appendix-A). Data were collected from 4 scenarios and in different hand conditions. The classification results of these algorithms were analyzed to better understand the commonalities and differences among them.

To evaluate and compare the different classification methods, each method was considered with respect to several characteristics such as classified eye-movement events, classification methods and classification results.

3.4.1 Algorithms and Eye-Movement Events

As seen in Table 3.1, 10 algorithms have some common and specific eye-movement events. For instance, the glissade duration (GD) is only detected by the AED algorithm and BIT algorithm classifies number of fixation (FN), fixation duration (FD) and saccade number (SN) events. Likewise, the algorithms I-DT, I-HMM, I-KF, I-MST, I-VDT, I-VMT, I-VT and I-VVT classify FN, FD and SN events. In addition to these, I-DT, I-HMM, I-KF, I-MST, I-VDT, I-VMT, I-VT and I-VVT algorithms classify saccade duration (SD), saccade amplitude degree (SAD). Also, I-VDT, I-VMT, I-VVT algorithms classify pursuit number (PN), pursuit duration (PD) and pursuit velocity degree (PVD). FD and SD are the eye-movement events that the AED algorithm is common with other algorithms. Also, the release dates of the algorithms are presented in Table 3.1. Accordingly, the initial algorithms I-VT, I-HMM, I-DT and I-MST were proposed by (Salvucci & Goldberg, 2000). The algorithms I-KF, I-VVT, I-VMT, I-VDT (Komogortsev, Jayarathna, et al., 2010), AED (Nyström & Holmqvist, 2010) and BIT (van der Lans et al., 2011) were developed afterwards. Algorithms have different classification methods, I-VT, I-KF, I-HMM, AED and BIT are velocity based algorithms and I-DT and I-MST algorithms are dispersion-based. The I-VVT and I-VMT algorithms are modified versions of the I-VT velocity based algorithm and they can be able to identify smooth pursuit eye-movement event. Also, the I-VDT algorithm can be able to identify smooth pursuit event and this algorithm is also a modified version of the velocity based and dispersion-based algorithm.

Table 3.1 Algorithms and Events

Name	Year	Velocity Based	Dispersion Based	FN	FD	SN	SD	SAD	GD	PN	PD	PVD
AED	2010	✓			✓		✓		✓			
I-DT	2000		✓	✓	✓	✓	✓	✓				
I-HMM	2000	✓		✓	✓	✓	✓	✓				
I-KF	2009	✓		✓	✓	✓	✓	✓				
I-MST	2000		✓	✓	✓	✓	✓	✓				
I-VDT	2011	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
I-VMT	2011	✓		✓	✓	✓	✓	✓		✓	✓	✓
I-VT	2000	✓		✓	✓	✓	✓	✓				
I-VVT	2011	✓		✓	✓	✓	✓	✓		✓	✓	✓
BIT	2011	✓		✓	✓	✓						

FN: Fixation Number; FD: Fixation Duration; SN: Saccade Number; SD: Saccade Duration; SAD: Saccade Amplitude Degree; GD: Glissade Duration; PN: Pursuit Number; PD: Pursuit Duration; PVD: Pursuit Velocity Degree

3.4.2 Differences among the Classification Results of the Algorithms

A non-parametric Friedman test of differences among repeated measures was conducted for the eye-movement classification algorithm effect on the eye-movement events. According to the results of Friedman test for ten eye-movement classification algorithm on all eye-movement events, a significant difference is found ($p < 0.001$). The eye-movements events mean ranks for all algorithms according to the results that algorithms produce are shown for different scenarios under different hand conditions. The Friedman test mean ranks and rendered Chi-square values for each algorithm and eye-movement event are shown for each scenario in Appendix-B.

3.4.3 Hierarchical Clustering Results

As seen from Table 3.1, Fixation Duration (FD) is the only common eye-movement event among all algorithms, a hierarchical clustering method is applied to better understand if there is a clustering among the classification results of these 10 algorithms. According to the results of hierarchical clustering based on mean FD for all algorithms and hand-conditions, two clusters are recognized (Figure 3.1). Cluster-1 includes only BIT algorithm and Cluster-2 includes other 9 algorithms.

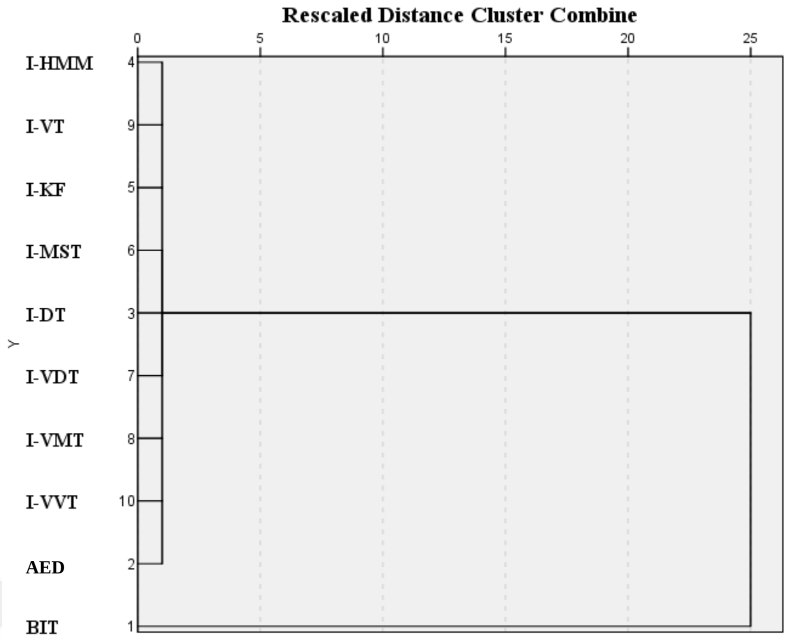


Figure 3.1 Dendrogram for 10 Algorithms Using Ward Linkage

The measured mean values of FD for the Cluster-1 and Cluster-2 are given in the Table 3.2.

Table 3.2 Values of Each Two Cluster

CLUSTER	FD (ms)
I	54286.87
II	6314.26

Since the number of common eye-movement events is more for the other algorithms without the BIT and AED algorithms (Table 3.1), these two algorithms are excluded and the cluster analysis is performed again with the common eye-movement events (FN, FD, SN, SD and SAD) for 8 algorithms (I-DT, I-HMM, I-KF, I-MST, I-VDT, I-VMT, I-VT and I-VVT). Three clusters are encountered with these 8 algorithms. According to the 3 clustered structure result, I-DT and I-VDT algorithms are in Cluster-1, I-HMM, I-KF, I-MST and I-VT algorithms are in Cluster-2, and I-VMT and I-VVT algorithms are in Cluster-3. Dendrogram graph of this clusters is provided in Figure 3.2.

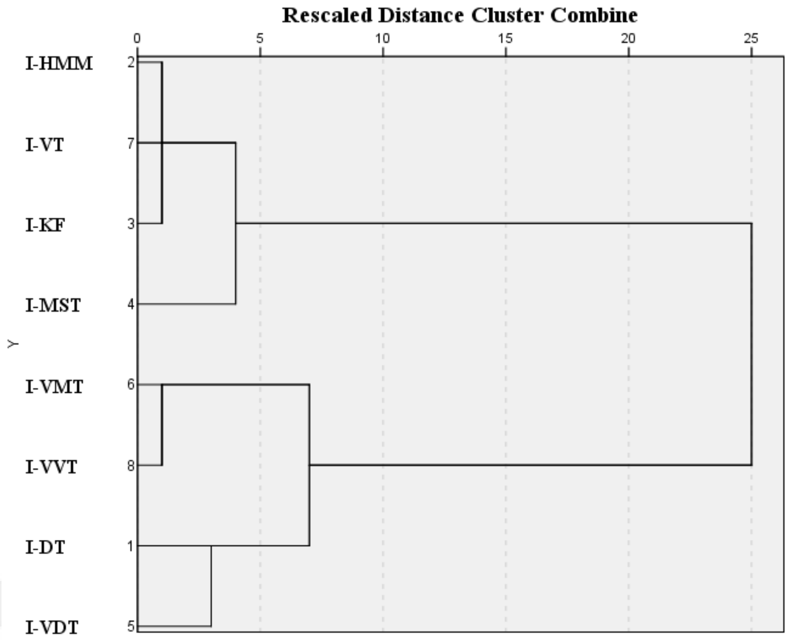


Figure 3.2 Dendrogram for 8 Algorithms Using Ward Linkage

The measured mean values obtained according to the algorithms in Cluster-1, Cluster-2 and Cluster-3 are given in the Table 3.3.

Table 3.3 Values of Each Three Cluster

CLUSTER	FN	FD (ms)	SN	SD (ms)	SAD (deg)
I	15.17	6527.02	24.82	2261.56	4.39
II	12.17	9042.51	13.87	1599.52	5.18
III	11.75	3801.77	18.02	1367.94	5.04

Three clusters have been found according to the hierarchical clustering analysis and it is seen that the clusters differs according to the methods and the threshold values of the algorithms. Only the BIT algorithm specifies the threshold values itself from eye-movement data. Accordingly, the BIT algorithm clustered different from all other algorithms when the hierarchical clustering analysis was conducted based on the only common measure FD. The BIT and AED algorithms are excluded to repeat the cluster analysis by increasing common measures. Hierarchical clustering analysis was repeated with 5 common measures for the remaining 8 algorithms. The results show that according to classification methods, algorithms appear in different clusters.

3.5 Discussion and Conclusion

The results of this study show that, as each algorithm uses different methods of classification, the threshold values are important and need to be set carefully. For example, eye-trackers, computer properties and algorithm thresholds used in classification can be different. However, open-source algorithms should be able to work with data obtained from different types of hardware. Yet, it is difficult to obtain such data from different sources to be brought to a working state with open-source algorithms. Because eye-movement classification algorithms are complex and contain many structures, it is particularly challenging for non-software researchers.

In order to make AED algorithm workable, it was necessary to have software infrastructure to understand the code. Since the algorithm is designed for a 1250Hz eye-tracker, it needs to be modified in order to work with the data received from eye-trackers operating at different Hz values. Accordingly, the code was modified to make it adaptable to the eye data gained from the Eye Tribe Eye Tracker which has a 60Hz sampling frequency.

The algorithms I-DT, I-HMM, I-KF, I-MST, I-VDT, I-VMT, I-VT and I-VVT are integrated in a single software (Appendix-A). Although, the software is an open-source, it is password-protected. The combination of algorithms with different features in a single interface is very useful and provides great convenience. The algorithms can be used one's own data by making changes through interface. Also, this software provides graphically representation of the results. This also brings about significant convenience for non-software-based researchers. Based on the results of this study the following important points can be highlighted:

All algorithms classify different eye-movement events. The most common eye-movement event among those is the FD. The second common ones are the FN, FD, SN, SD and SAD. Different algorithms give different classification results. This is because of the methods and the threshold values they use. There is no consensus for defining the threshold values for algorithms which significantly affects the classification results.

According to hierarchical clustering based on mean of FD produced by all 10 algorithms, two clusters are recognized: cluster 1 includes only BIT algorithm and cluster 2 includes other 9 algorithms. It can be concluded that, BIT algorithm individually specifies the threshold values from the provided eye data. Therefore, the results of BIT algorithm becomes different from others. Main reason behind this difference is most probably the assigned threshold values. BIT automatically calculates threshold values whereas the default threshold values are taken for the other algorithms.

As BIT and AED algorithms have restricted eye-movement events, another hierarchical clustering analysis is performed by excluding them. As a results, 3 clusters are recognized. According to hierarchical clustering based on means of FN, FD, SN, SD and SAD produced by 8 algorithms (I-DT, I-HMM, I-KF, I-MST, I-VDT, I-VMT, I-VT and I-VVT), 3 clusters have been recognized: Cluster-1 consisted of I-DT which is a dispersion-based algorithm and I-VDT algorithm which is the modified combination of the velocity-based I-VT and dispersion-based I-DT algorithms. Dispersion-based algorithms are based on the x and y coordinates of the eye data for classifying eye-movements into fixation and saccade events. I-VDT algorithm differs from the modified velocity-based algorithms (I-VVT and I-VMT) because this algorithm separates smooth pursuits from fixations by employing a modified dispersion threshold identification method. Therefore, the results of I-VDT is different from the other modified algorithms I-VVT and I-VMT similarly reported in (Komogortsev & Karpov, 2013). Cluster-2 includes I-HMM, I-KF, I-MST, and I-VT algorithms. Except I-MST algorithm remaining all algorithms in this cluster are velocity-based algorithms and they use velocity of the eye-movements to separate fixation events from saccade events. I-MST algorithm has an advantage of correctly identifying the anchor points if a large part of the signal is missing and this property makes I-MST highly flexible and controllable eye-movement detection tool (Salvucci & Goldberg, 2000) and this algorithm produce similar results to the velocity based algorithms. The threshold values may cause to that situation but the reasons behind this should be investigated further. The algorithms I-VMT and I-VVT are take part in Cluster-3, the common property of these two algorithms is they are both modified versions of the velocity-based algorithm I-VT. Firstly, these two algorithms utilizes velocity threshold for classifying saccades then investigates the eye-movement samples to detach smooth pursuits from fixations. These algorithms are more suitable to the dynamic stimuli such as video-viewing, where objects are moving (Andersson et al., 2017; Komogortsev & Karpov, 2013).

According to the results of this study, it can be concluded that if the threshold values were not specified than the BIT algorithm is an appropriate algorithm for classifying the eye-movement events. Because this algorithm can individually specifies the threshold values for the eye-movement data which is gathered by eye-tracking device. The other algorithms can be applicable if specific threshold values are known. Hence, the threshold values are critical for the event detection in these algorithms. As a conclusion, for better interpretation of eye research, appropriateness of the algorithm based on the research specific features need to be considered.

CHAPTER 4

USING EYE-MOVEMENT EVENTS TO DETERMINE SCENARIO DIFFICULTY LEVELS

Technology-enhanced educational environment, provide several benefits to improve surgical education programs. For instance, simulation is one of the technologies that allows trainees to perform clinical activities interactively by recreating such operations in a computer-based system without exposing patients to the associated risks (Maran & Glavin, 2003; Munshi et al., 2015). However, still there is a need for research to develop strategies for improving the curriculum integration of these systems and for creating standardized approaches. In this respect, the mental workload theory and the eye-tracking technology are two important concepts that can be implemented in surgical education programs.

The mental workload concept has long been accepted as an essential aspect of individual performance within complex systems (Xie & Salvendy, 2000). It is reported that mental workload can change the performance of individuals (Zheng, Cassera, Martinec, Spaun, & Swanström, 2010) and further affect the competence of the whole system (Xie & Salvendy, 2000). Accordingly, system developers need certain models to assess the mental workload imposed on individuals at an early stages so that alternative system designs can be appraised (Xie & Salvendy, 2000). At the same time, mental workload can negatively affect performance and increase the probability of errors (Zheng et al., 2010), and researchers have spent a great deal of effort developing measures and probes of mental workload (Ahlstrom & Friedman-Berg, 2006). For instance, Moray stated that adjusting the allocation of mental workload could reduce human errors, improve system safety, and increase productivity (Moray, 1988).

Eye-tacking provides a valuable source of information, and events such as fixations, blinks, and pupil diameter can be used to assess the mental workload (Tsai, Viirre, Strychacz, Chase, & Jung, 2007). Accordingly, there are several studies conducted on the assessment of mental workload by using eye-tracking technology. A precise evaluation of mental workload will be essential for developing systems that manage user attention (Atkins, Tien, Khan, Meneghetti, & Zheng, 2013; Iqbal, Zheng, & Bailey, 2004). Researchers have used eye-movement events found to correlate with cognitive demands (Ahlstrom & Friedman-Berg, 2006). For instance, (Benedetto et al., 2011) examined the changes in blink duration and blink rate in a simple driving task and stated that blink events reflect the effects of visual workload. Another study evaluates the mental workload by developing combined measures based on various physiological indices (Ryu & Myung, 2005). To determine the mental workload, three physiological signals were recorded; these are: alpha rhythm, eye blink interval, and heart rate variability (Ryu & Myung, 2005). The study of (de Greef, Lafeber, van Oostendorp, & Lindenberg, 2009) describes an approach for objective assessment of mental workload by analyzing the differences in pupil diameter and several aspects of eye-movement under different levels of mental workload. Eye-movement events are also used in medicine for diagnoses, treatment and training purposes (Jarodzka et al., 2017) and for clinical applications such as Alzheimer’s (Crawford et al., 2005), HIV-1 infected patients with eye-movement dysfunction (Sweeney et al., 1991), and schizophrenia (Flechtner et al., 1997). Studies show that these events provide crucial information about how users interact with complex visual displays (Marshall, 2002). The field of radiology and visual search (Nodine & Kundel, 1987) and laparoscopic surgery training (Law, Atkins, Kirkpatrick, & Lomax, 2004; G. Tien, Atkins, Zheng, & Swindells, 2010) are among the cases in medicine where eye-tracking approach has been adopted. To provide an example, according to the study (Zheng et al., 2015b), participants perform a simulated laparoscopic procedure, and when the task difficulty is increased, the task completion time and pupil size also increase as a result.

Previous studies were conducted mostly on pupil size changes, but there are other eye-movement events, fixation for example, that can be informative for understanding mental workload. Fixation occurs when eye-movements are nearly still and in order to assemble necessary information. Accordingly, in this study number of fixation and fixation duration events are used to validate the mental workload imposed by different scenarios. As changes in eye-movement events, such as number of fixation and fixation duration, with changes in mental workload are likely affected due to the nature of the scenarios (Tsai et al., 2007), understanding the surgeon's mental workload while performing surgical operations is crucial for assessing task difficulties. Hence, this study attempts to understand the mental workload changes of the participants through their eye-movement events, namely number of fixation and fixation duration, while performing different surgical tasks. The authors believe that, this information will be very helpful to better design, order and adapt related computer-based simulation technologies according to the individual requirements and progress of the trainees.

4.1 Results

Four different computer-based simulation scenarios are performed with dominant-hand, non-dominant hand and both-hands. During this process, eye-movement data is recorded by an eye-tracker. The results were analyzed using statistical methods aimed to better understand the surgeons' behaviors in these different simulation scenarios.

The recorded data was classified using an open-source eye-movement classification algorithm (Binocular-Individual Threshold-BIT). BIT algorithm, developed by (van der Lans et al., 2011) was utilized which is a velocity-based algorithm to classify fixations from the data with individual-specific thresholds. For the evaluation of differences based on scenario difficulties, the number of fixation and fixation duration events were used.

In all, 276 (23 surgeons, 4 scenarios, and 3 hand conditions) datasets were recorded, significantly increasing the accuracy of the results in this work. To evaluate and compare the differences among the difficulty levels of the scenarios, the eye-movement events, number of fixation, fixation duration, were analyzed.

4.1.1 Number of Fixation

A non-parametric Friedman test of differences among the repeated measures was conducted for the scenario difficulty level effect on the number of fixation. The effect of the scenario was significant (in all $p < .05$) on the number of fixation according to the results. While the hand condition is fixed, the results of the analysis of the repeated measurements differ according to the scenarios. Based on the Friedman test for different measurement groups, there is a statistically significant difference between the number of fixation when using the dominant-hand ($\chi 2$ = 37.08, $p < 0.05$) for different scenarios. Scenario-1 has the lowest mean rank for the number of fixation (1.57), while Scenario-2 has the highest (3.78). When using the non-dominant hand ($\chi 2$ = 50.18, $p < 0.05$) for different scenarios, Scenario-1 has the lowest mean rank for the number of fixation (1.26) while Scenario-2 has the highest (3.70) number of fixation. According to the test results when using both-hands ($\chi 2$= 52.74, $p < 0.05$) for different scenarios, Scenario-1 has the lowest mean rank for the number of fixation (1.07) while Scenario-2 has the highest mean rank (3.80) for the number of fixation . According to the results of the three hand conditions for the number of fixation measure, the scenario that makes number of fixation larger is reported (Figure 4.1). Generally, in Scenario-2 the number of fixation becomes larger compared to the other scenarios.

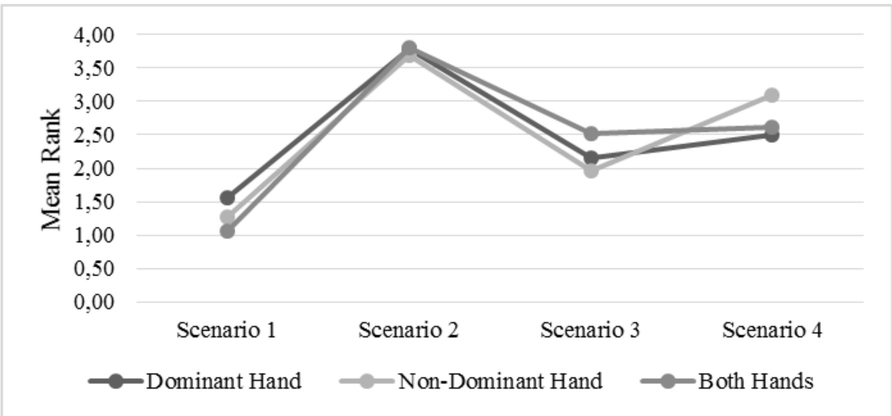


Figure 4.1 Number of Fixation Differences among Scenarios

4.1.2 Fixation Duration

A non-parametric Friedman test of differences among the repeated measures was conducted for the scenario effect on fixation duration. The effect of scenario was significant (in all $p < .05$) on the fixation duration according to the results. While the hand condition is fixed, the results of the analysis of the repeated measurements differ according to the scenarios. According to Friedman test for different measurement groups, there is a statistically significant difference between the fixation duration when using the dominant-hand ($\chi 2$ = 52.41, $p < 0.05$) for different scenarios. Scenario-1 has the lowest mean rank for the fixation duration (1.04) while Scenario-2 has the highest mean rank for the (3.70) fixation duration. When the non-dominant hand is used ($\chi 2$= 54.49, $p < 0.05$) for different scenarios, Scenario-1 has the lowest mean rank for the fixation duration (1.04) while Scenario-4 has the highest mean rank for the (3.52) fixation duration. In the both-hand condition ($\chi 2$ = 65.56, $p < 0.05$), Scenario-1 has the lowest mean rank for the fixation duration (1.00) while Scenario-2 has the highest mean rank for the (3.96) fixation duration. According to the results of the three hand conditions, the scenario that makes the fixation duration longer is reported (Figure 4.2). In Scenario-2 and Scenario-4 the fixation duration becomes larger compared to the other scenarios.

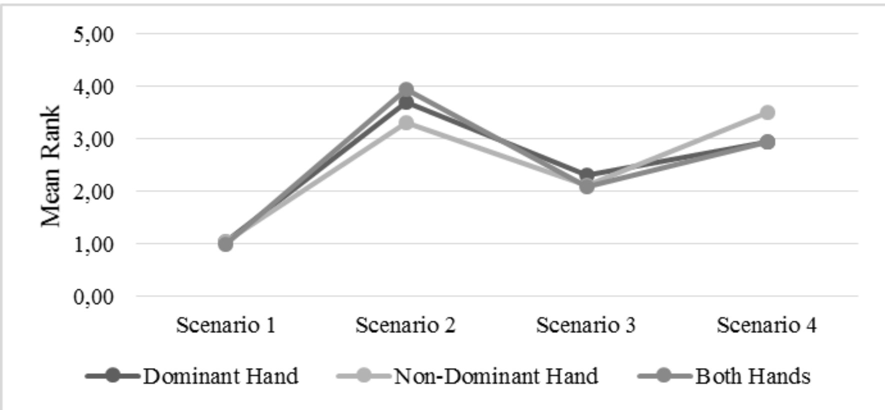


Figure 4.2 Fixation Duration Differences among Scenarios

4.2 Discussion and Conclusion

This research describes an approach for an objective assessment of mental workload by analyzing the differences in the number of fixation and fixation duration under different levels of mental workload while surgeons perform simulated scenarios. The eye-movement data was collected with an eye-tracking device and classified into number of fixation and fixation duration events with an eye-movement classification algorithm. There are many other eye-movement classification algorithms, but in this study an open-source eye-movement classification algorithm, BIT, was used. The reason behind this choice was that BIT algorithm is eye-tracker independent and easy to implement and use. The aim of this study is to examine whether the number of fixation and fixation duration events can, indeed, be indicators for mental workload and whether there are any among the imposed mental workloads within different scenarios. According to the results, the number of fixation and fixation duration both show a significant increase if the mental workload increases. For understanding the differences between the scenarios, four of them were developed in this study; two were simulated surgical models and the other two were general models. The results can be summarized as highlighted below:

- In the dominant-hand condition, Scenario-1 has the lowest mean rank for the number of fixation (1.47) and fixation duration (1.04) while Scenario-2 has the highest mean rank for the number of fixation (3.78) and fixation duration (3.70).

- When using the non-dominant hand, Scenario-1 has the lowest mean rank for the number of fixation (1.26) and fixation duration (1.04), while Scenario-2 has the highest mean rank for number of fixation (3.70) and Scenario-4 has the highest mean rank for fixation duration (3.52).

- When using both-hands, Scenario-1 has the lowest mean rank for the number of fixation (1.07) and fixation duration (1.00), whereas Scenario-2 has the highest mean rank for number of fixation (3.80) and fixation duration (3.96).

In general, it can be concluded that in the scenarios that are designed by using the models that simulate the operational area (Scenario 2 & 4), the fixation duration and number of fixation values become higher compared to the other group of scenarios (Scenario 1 & 3). When the hand condition is evaluated descriptively, it can be concluded that in all scenarios, mostly under the non-dominant hand and both-hand conditions, the fixation duration and number of fixation events are higher compared to the dominant-hand condition. However, the effect of the hand condition on these events needs to be researched in further detail.

In previous studies, it has been stated that the pupil diameter and fixation time both show a general significant increase if the mental workload increases (de Greef et al., 2009). Another study stated that the pupil size increased in response to task difficulty (Nakayama, Takahashi, & Shimizu, 2002). (Iqbal et al., 2004) also stated that more difficult tasks demand longer processing times, induce higher subjective ratings of mental workload, and reliably evoke greater pupillary response at corresponding subtasks than a less difficult task. Additionally, (Zheng et al., 2015b) stated that the pupil size of surgeons is influenced depending on the task difficulties increasing as the difficulty level elevates. It is also reported that the fidelity level is a crucial factor affecting the mental workload (Munshi et al., 2015). In support to these studies, our results show that the scenarios based on simulated tasks using surgical models (higher level of fidelity) increase surgeons' mental workloads. Hence, it can be concluded that eye-movement events, such as number of fixation and fixation duration, can be used to increase our knowledge of the mental workload of surgical trainees.

Additionally, as there are very limited studies analyzing the eye-movement behaviors of endo-neurosurgery residents, there are no standards in classifying the simulation content according to the level of surgical skills. Similarly, the metrics that can be used to evaluate the skill levels of these residents are also very limited and there are no standards on these metrics, either. Hence, the results of this study encourage researchers to develop other standardized approaches for using objective metrics in surgical skill performance, and provides additional insights about the threshold values of the novice or intermediate level of endo-neurosurgery residents' eye-movement events. Additionally, the results may guide instructional system designers in this field to better organize the content of computer-based simulation scenarios based on the eye-movement behaviors of the trainees.

CHAPTER 5

DETERMINING SURGICAL SKILL LEVELS

Surgical skills assessment is a critical process to ensure competence and prevent clinical errors while developing effective instructional methods for such evaluation. Currently, the assessment of surgical residents' skill levels is based on a subjective approaches and conventional methods and, as such, open to bias and questionable rationality (Eubanks et al., 1999; Feldman, Hagarty, Ghitulescu, Stanbridge, & Fried, 2004; Resnick, Taylor, & Maudsley, 1991; Richstone et al., 2010; Wanzel, Ward, & Reznick, 2002).

It is commonly acknowledged that the skill levels of surgeons vary (Uhrich, Underwood, Standeven, Soper, & Engsberg, 2002). Several studies have reported that more effective skill level assessment techniques and training evaluation systems could improve skill-based training programs and, accordingly, patient health care (Reiley, Lin, Yuh, & Hager, 2011). Assessment is crucial for adapting the content and level of training programs to the surgical residents' needs and qualifications, providing appropriate feedback for both the trainees and educators, and improving the curriculum (Cagiltay, Ozcelik, Sengul, & Berker, 2017). However, traditional assessment methods have many shortcomings, such as subjective evaluation, need for an expert, and need for a standardized methods to assess surgical skills (Ahmidi, Ishii, Fichtinger, Gallia, & Hager, 2012; Moorthy, Munz, Sarker, & Darzi, 2003). Objective evaluation of surgical residents is very important and necessary, but difficult to achieve. In this context, there is a need for objective metrics to evaluate the skill levels of surgical residents. To fill this gap, new technologies can be used advantageously by recording the eye-movements of surgeons and analyzing such data to provide a cost-effective, automated, and objective basis for assessing a surgeon's skill level (Ahmidi et al., 2012).

Eye-tracking provides objective metrics about human behavior (Bröhl et al., 2017; Yarbus, 1967). Currently, eye-tracking systems have many beneficial properties and it is easy to record and analyze eye-movement data with these systems (T. Tien et al., 2015). Hence, eye-tracking is being used for assessing and understanding the differences between skill levels in the medical domain. It has been reported that surgical skill levels can be objectively evaluated by eye-tracking metrics through virtually simulated and live environments (Richstone et al., 2010). Additionally, it has been reported that the differences in performances are, in fact, differences in the information-processing capabilities of the left and right hemispheres of the brain, implying that when visual control is required, the dominant-hand will perform better than the non-dominant hand and both-hands (Hoffmann, 1997). For instance, (J. Vickers, 1995) examined the change in eye-movements between expert and novice basketball players in foul shooting, and reported that earlier in a shot, expert basketball players' visual system programs their motor control system, which means that they do not need to follow the entire shooting process with their eyes, yet, the novice ones use their visuals to adjust their shots until releasing the ball. Also, (Kasarskis, Stehwien, Hickox, Aretz, & Wickens, 2001) reports differences between the eye-movements of expert and novice pilots while simulating the landing operation, and it was shown that experts' fixation times are shorter than novices because the former assemble the necessary information more rapidly. Virtually simulated environments allow surgical residents to perform clinical operations interactively in computer-based systems without risks (Gallagher & Satava, 2002; Lababidi, Alyousef, & Munshi, 2015; Maran & Glavin, 2003). Such, virtual scenarios are developed for surgical-skill development and assessment requirements for surgery purposes (Gallagher & Satava, 2002; McNatt & Smith, 2001).

Accordingly, in this study a surgical simulation scenario is developed for surgical residents. Even though virtual simulations provide alternatives to improve education and assessment in surgery education programs and allow for several objective assessment measures (Oostema, Abdel, & Gould, 2008), there are not many existing tools for assessing the overall performance of surgical residents (Cagiltay et al., 2017). Therefore, in this study a surgical scenario is performed by surgical residents and their eye-movement is were collected while performing the required tasks. Data is classified with an open-source eye-movement classification algorithm Binocular Individual Threshold (BIT) (van der Lans et al., 2011). This algorithm classifies the eye-movement data into number of fixation and fixation duration events. Fixation is an eye-movement event when movements are nearly still, and this event occurs to assemble the necessary information. Also, the dominant-hand is an important factor for performing surgical tasks.

5.1 Results

The eye-movement data of surgical residents were collected with an eye-tracking device while they performed Scenario-1. The recorded data was classified using an open-source eye-movement classification algorithm, namely BIT, designed to classify number of fixation and fixation duration events (van der Lans et al., 2011). Hence, to understand the differences between the skill levels, number of fixation and fixation duration events were used. The results were analyzed using statistical methods to predict the skill levels of surgeons based on their eye-movement behaviors. The logistic regression analysis was conducted in this study to predict of the skill levels of surgeons.

In this study, the number of fixation and fixation duration events are analyzed. Twenty-three (9 intermediate, 14 novice) surgeons performed the simulation scenario. The results show that the number of fixation (Figure 5.1) and fixation duration (Figure 5.2) of novices are higher than the intermediate surgeons.

Figure 5.1 Number of Fixations of Novice and Intermediate Surgeons

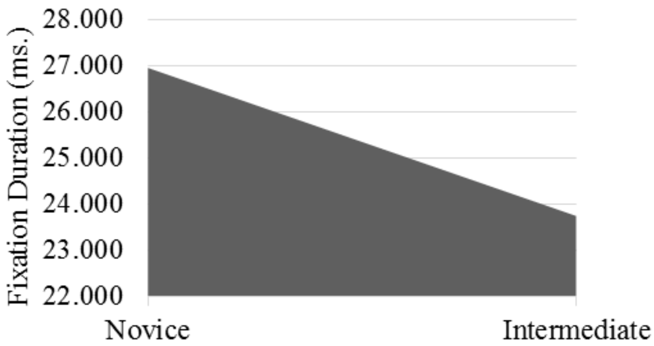


Figure 5.2 Fixation Durations of Novice and Intermediate Surgeons

Logistic regression was performed to ascertain the effects of number of fixation, fixation duration and the dominant-hand on the prediction of the participants' skill level. According to the results, the logistic regression model was statistically significant, $\chi2(3) = 15.661$, $p = .001$. The model explained 66.9% (Negelkerke R2) of the variance in skill level and correctly classified 91.3% of cases.

5.2 Discussion and Conclusion

Assessment of surgical residents' skill levels is critical because objective evaluations are necessary to monitor the progress of surgeons in surgical education programs. Also, these measurements are vital to prevent medical errors (Richstone et al., 2010). However, assessment of surgical skill levels is typically made by an observer and such evaluations do not always yield objective results (Richstone et al., 2010). An ideal evaluation of such skill has to be objective, devoid from bias, and should require neither the presence of evaluators nor a review of lengthy operative performances (Richstone et al., 2010). The results of this study show that it is possible to objectively assess the skill levels of novice and intermediate surgeons using eye-movement events such as number of fixation and fixation duration. These events could be used to objectively monitor the acquisition of skills throughout training programs. The use of eye-movement events is more objective than the approaches currently used, and does not involve the time or expense of an expert evaluator. Therefore, from this study it can be seen that eye-movement events can reliably distinguish intermediate surgeons from novices. In the near future, the data obtained from the eye-movements of surgical residents is likely to take the place of subjective evaluation methods, and assessments will become more objective in this way.

CHAPTER 6

DETERMINING MENTAL WORKLOAD FROM PUPIL SIZES CONSIDERING HAND CONDITION

When compared with traditional methods, endoscopic surgery provides several benefits. It has been reported that people who underwent endoscopic surgery instead of traditional open procedures experience a faster recovery period and less pain (Feng, Rozenblit, & Hamilton, 2007). In these operations, smaller size incisions are used, and unlike traditional methods, surgeons need to use both-hands effectively. For this reason, the surgeons' skills in performing tasks with their dominant-, non-dominant and both-hands are very critical for the patients' safety and quality of operations.

According to the mental workload theory, there is a correlation between the rate at which knowledge is handled by the human operator, and the rate at which decisions are taken (Moray, 2013). In the literature, it is stated that mental workload is related to several physiological measures, namely heart rate, blink frequency and duration, pupil size, electro-dermal activity, respiratory frequency and other variables derived from EEG (Hogervorst, Brouwer, & van Erp, 2014). Factors such as hand condition are most probably effective on the mental workload of surgical residents. Parallel to the mental workload theory, psycho-physiological studies show that eye-tracking is broadly used to collect response parameters from the eye-movement data to analyze cognitive processes underlying visual behavior (Berger, Winkels, Lischke, & Höppner, 2012). Additionally, cognitive load theory describes how the mental effort of learners is influenced by the design of the learning material (Mayer & Moreno, 2003; Sweller, 1988, 1994).

Studies show that human pupillary response is satisfactory evidence for the relationship between pupil size and a wide range of important cognitive variables, including mental workload (Andreu-Perez, Solnais, & Sriskandarajah, 2016; Bradshaw, 1967; Hess & Polt, 1964; Kahneman, 1973; Kahneman, Beatty, & Pollack, 1967; Klingner, Tversky, & Hanrahan, 2011; Menekse, Cagiltay, Ozcelik, & Maras, 2017; Simpson, 1969); hence, there exists a relationship between cognitive load and pupil diameter (Beatty & Lucero-Wagoner, 2000). Supportively, other studies report that pupil size is an important indicator of the brain function (Joshi, Li, Kalwani, & Gold, 2016; Murphy, O'connell, O'sullivan, Robertson, & Balsters, 2014; Rajkowski, 1993; Varazzani, San-Galli, Gilardeau, & Bouret, 2015).

Today’s eye-tracking technology provides priceless physiological data to better understand the consumption of resources through ocular activity, which is closely related to the neural networks underlying the brain (Andreu-Perez et al., 2016; Jarodzka et al., 2017); it has also been reported that this technology can be used for high-temporal-resolution tracking of cognitive workload (Wierda, van Rijn, Taatgen, & Martens, 2012). Therefore, current advances in the consistency of the eye-tracking methodology in addition to the increasing accessibility of affordable simulation and modeling technology have widened research prospects in a range of areas and applications (Andreu-Perez et al., 2016). In addition, pupil size changes have been suggested as a metric for evaluating mental workload while complex visual tasks are performed (Andreu-Perez et al., 2016).

For instance, studies report that changes in pupil diameter might reflect neuronal activity and cognitive functions throughout some parts of the brain (Joshi et al., 2016; Zekveld, Heslenfeld, Johnsrude, Versfeld, & Kramer, 2014). In another study, pupil size changes have recommended as a biomarker in early-phase detection of Alzheimer’s (Andreu-Perez et al., 2016).

Additionally, the relationship between pupil size and information processing load in a variety of cognitive tasks is also well proven (Jiang, Atkins, Tien, Bednarik, & Zheng, 2014; Zekveld et al., 2014), while pupil size dynamics have been shown to be a reliable measure to investigate the cognitive processes involved in sentence processing and memory functioning (Fernández, Biondi, Castro, & Agamenonni, 2016). As the human mental capacity is a finite resource, the extent of achievement of a complex task relies on the task requirements to be performed (Cassenti & Kelley, 2006). It is reported that the pupil response pattern is distinguishable depending on different levels of task difficulty as the pupil diameter increases in harder tasks (Jiang et al., 2014). Studies conducted in this area show that, based on psychomotor evidence regarding surgeons' performances, task specific training curricula can be designed to improve the related skills (Zheng, Jiang, & Atkins, 2015a). Especially, eye-tracking applications are being increasingly used in different areas. The accessibility of affordable devices from monitor screens, to goggles and computer peripherals has extended this application area in a variety of disciplines such as medicine, commerce and education (Andreu-Perez et al., 2016).

Several eye-tracking studies conducted on physiological signs have been used to evaluate surgeons' mental workloads (Zheng et al., 2015a). The results of past research suggest that, when the surgical tasks become difficult, the mental workload increases as evident from the pupil size changes. Additionally results of earlier studies about surgeons also show that, due to the long tool shaft and the lack of depth information when projecting the scene inside the body to a two-dimensional screen, the task requirements in laparoscopic surgery are considered as higher (more demanding) than in open surgery (Jiang et al., 2014). Also, when the task difficulty level increases in the surgical laparoscopic procedures, the subjects' peak pupil size also increases (Zheng et al., 2015a). However, in the literature there are not many studies conducted to better understand how the pupil size changes correlate with the level of the task difficulty in an eye-hand coordination movement (Jiang et al., 2014). Therefore, in this study the changes of surgeons' physiological signals are examined related with the mental workload by considering their eye-movements in different hand conditions.

This study is an experimental one and conducted on a computer-simulated surgical task with 23 surgical residents. In detail, this study attempts to better understand the influence of the hand condition (dominant-, non-dominant, or both-hands) on the eye-movements of surgeons while they perform a skill-based surgical task in a computer simulated environment. The results of this study expected to guide instructional system designers to better address the skill development requirements especially for educational CAASS.

6.1 Results

In order to fully grasp the influence of the hand condition (dominant-, non-dominant, or both-hands) on the performance of skill-based surgical tasks in a computer-based simulated environment, right eye pupil sizes, left eye pupil sizes and average both-eye pupil sizes of 23 surgical residents are examined. An endoscopic surgery was performed with an endoscope and several long, thin instruments through small incisions. According to the medical requirements, in this study a simulation scenario of an endoscopic procedure was modeled in an educational computer-based-simulation environment (ECE). The pupil sizes of the surgical residents were collected with an eye tracker while they were performing the Scenario-2 performed with the haptic devices under different hand conditions (dominant-, non-dominant, both hand).

A non-parametric Friedman test of differences among repeated measures was conducted for hand condition effect on left eye pupil size and rendered a Chi-square value of 11.57 which was significant ($p < 0.010$), as seen from Table 6.1. Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.017$. There were no significant differences ($Z = -1.765$, $p = 0.078$) between the mean of left eye pupil sizes of the participants while they were performing the tasks under the dominant hand condition (18.37, $SD = 3.22$), and non-dominant hand condition (19.27, $SD = 2.13$). Similarly, there were no significant differences ($Z = -2.103$, $p = 0.035$) between the mean of left eye pupil sizes of the participants while they were performing the tasks under the dominant hand condition (18.37, $SD = 3.22$), and both-hands condition (19.89, $SD = 2.76$). Finally, there were no significant differences ($Z = -1.583$, $p = 0.113$) between the mean of left eye pupil sizes of the participants while they were performing the tasks under the non-dominant hand condition (19.27, $SD = 2.13$) and both-hands condition (19.89, $SD = 2.76$). This result indicates that the hand condition significantly effects the left eye pupil size however there is no significant difference on the left eye pupil size among different hand conditions.

A non-parametric Friedman test of differences among repeated measures was conducted for hand condition effect on right eye pupil size and rendered a Chi-square value of 16.44 which was significant ($p < 0.010$), as seen from Table 6.1. Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.017$. There were no significant differences ($Z = -1.491$, $p = 0.126$) on the mean of the right eye pupil sizes of the surgical residents, while they were performing the computer simulated surgical tasks under the dominant hand condition (18.56, SD = 2.73), and non-dominant hand condition (19.12, SD = 2.10). However, there was a statistically significant ($Z = -3.535$, $p = 0.000$) reduction on the mean of the right eye pupil sizes of the surgical residents while they were performing the computer simulated surgical tasks under the dominant hand condition (18.56, SD = 2.73), compared to under the both-hands condition (20.34, SD = 2.60). Similarly, there was a statistically significant ($Z = -2.801$, $p = 0.005$) reduction on the mean for the right eye pupil sizes of the surgical residents while they were performing the computer simulated surgical tasks under the non-dominant hand condition (19.12, SD = 2.10) compared to under the both-hands condition (20.34, SD = 2.60).

A non-parametric Friedman test of differences among repeated measures was conducted for hand condition effect on average of both-eyes pupil sizes and rendered a Chi-square value of 11.57 which was significant ($p < 0.010$), as seen from Table 6.1. Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.017$. No significant differences ($Z = -1.522$, $p = 0.128$) on the mean for average of left and right eye pupil sizes was found while the surgical residents were performing the computer simulated surgical tasks under the dominant hand condition (18.47, SD = 2.87), and non-dominant hand condition (19.23, SD = 2.01). Similarly, no significant differences ($Z = -2.192$, $p = 0.028$) on the mean of the average of left and right eye pupil sizes was found while the surgical residents were performing the computer simulated surgical tasks under the non-dominant hand condition (19.23, SD = 2.01) and the both-hands condition (20.12, SD = 2.62). However, there was a statistically significant (Z = -2.499, p = 0.012) reduction in the mean of the average of left and right eye pupil sizes of the surgical residents while they were performing the computer simulated surgical tasks under the dominant hand condition (18.47, SD = 2.87), when compared to the both-hands condition (20.12, SD = 2.62).

Table 6.1 Friedman Test Statistics for Pupil Size

	Left Eye	Right Eye	Both Eyes
N	23.00	23.00	23.00
Chi-Square	11.57	16.44	11.57
df	2.00	2.00	2.00
p	0.00	0.00	0.00
		Mean Rank	
Dominant-hand	1.52	1.48	1.48
Non-dominant hand	1.96	1.87	2.04
Both-hands	2.52	2.65	2.48

The results indicate that, when the task was performed under different hand conditions, there is a significant difference among the left, right and the average of both eyes pupil sizes. As seen from Figure 6.1, under the both-hand condition, the left eye and the right eye pupil sizes as well as the average pupil sizes of both eyes are larger when compared to the dominant-hand and non-dominant hand conditions.

Figure 6.1 Mean Ranks of Pupil Size in Different Hand Conditions

6.2 Discussions and Conclusion

The main aim of the study was to better understand the influence of hand conditions on the mental workload of surgical residents. For this purpose, while performing computer-simulated surgical tasks, the participants' pupil sizes were recorded and analyzed. Surgeons performed the same tasks under dominant-hand, non-dominant hand, and finally both-hand conditions. The Friedman statistical analysis method was used to better understand the effect of different hand conditions on pupil sizes. The results of this study show that, while performing computer-simulated surgical tasks, the hand condition (dominant-hand, non-dominant hand and both-hand) significantly affects the pupil sizes (left eye pupil size, right eye pupil size and average of both eye pupil sizes) of the participants. According to this result, under both-hand conditions, the pupil sizes of the surgical residents become larger than other conditions (non-dominant and dominant-hand). Earlier studies reported that pupil sizes grow in direct proportion to the mental workload (Beatty & Lucero-Wagoner, 2000; Hess & Polt, 1964; Zheng et al., 2015a).

Also it is noted that mental workload is generally defined as the ratio between the capacity of a person and task demands, and that mental workload is high when the task demands exceed capacity (Strang, Best, & Funke, 2014). Mental workload has been described as a subjective perception of the association between mental processing ability and the amount of processing required to perform a task (Strang et al., 2014). Knowledge of a person's mental workload will be useful in assessing and designing systems or working conditions, such as monitoring and helping people at work (Brouwer et al., 2012). Hence, under the both-hand condition the simulated surgical tasks are regarded as harder than under other hand conditions, indicating that under the both-hand conditions the mental workload increases.

According to the results, it is seen that surgical residents have more difficulty when using both-hands. Since in real operations it is necessary to use both-hands in a coordinated fashion, surgical residents' skills under both-hand conditions need to be assessed and improved systematically during surgical education programs. As the pupil sizes appear to provide an objective assessment of the mental workload among surgical residents, this information can be used to better evaluate and guide skill improvements during simulation-based surgical training programs. The pupil size measures are capable of providing instructional system design alternatives for individualized training programs. This study encourages instructional system developers of simulation-based surgical education programs to utilize pupil size data for better guiding and assessing the trainees' skill improvements.

This study was conducted on surgical residents, a majority of whom used their right hands as the dominant-hand. Since there were only three participants whose dominant-hand was their left hand, a comparison could not be conducted in this study. However, in future attempts the dominant-hand being left or right can also be evaluated as a factor. Additionally, similar studies can also be conducted with people from other domains, such as pilots, users of intricate machines and systems, and so on.

CHAPTER 7

EFFECT OF EXPERIENCE LEVEL AND HAND CONDITION ON EYE-MOVEMENT EVENTS

Eye-tracking technology provides objective measures about human behaviors (Bröhl et al., 2017; Yarbus, 1967). Currently, eye-tracking systems have many useful features and it is easy to collect and analyze data with these systems (T. Tien et al., 2015). Eye-tracking is being used by many researchers to investigate the behavior patterns of experts in various fields such as aviation (Schriver, Morrow, Wickens, & Talleur, 2008), arts (Vogt & Magnussen, 2007), sports (North, Williams, Hodges, Ward, & Ericsson, 2009) and driving (Crundall, Underwood, & Chapman, 1999). For instance, (J. Vickers, 1995) examined the change in eye-movements between expert and novice basketball players in foul shooting. It is reported that, earlier in a shot, expert basketball players' visual system programs their motor control system which means that they do not need to follow the entire shooting process with their eyes. However, novice basketball players use their visuals to adjust their shots until they release the ball (J. Vickers, 1995). Another study (J. Vickers, 2003) differentiates the eye-movements of expert and novice golf players, and reporting that expert players make fewer saccadic eye-movements. Also, (Kasarskis et al., 2001) report the differences between the eye-movements of expert and novice pilots while simulating the landing operation. According to that work experts' fixation times are shorter than the novices' because the former tended to acquire the necessary information faster.

Studies also show that there are several benefits associated with eye-tracking technology in the field of medicine. For instance, it is reported that expert radiologists generally do not scan the edges of the lungs and, instead, look at the other regions because there are fewer lesions in the corners (Nodine & Mello-Thoms, 2000). According to them, novice radiologists are not aware of this and, hence, examine the corner regions of the lungs as well (Nodine & Mello-Thoms, 2000).

Significant differences are reported in gaze patterns between novice and expert surgeons while watching surgical videos (Khan et al., 2012). As a result of a meta-analysis, it is reported that experts' fixation duration are shorter than non-experts' (Gegenfurtner, Lehtinen, & Säljö, 2011). However, their number of fixations on target locations are higher than non-experts' where the latter split their time between focusing on the target and tracking the tool (M. R. Wilson, McGrath, et al., 2011). It is also reported that the number of fixations of experts are higher on task-relevant areas, and fewer on task-redundant areas (Gegenfurtner et al., 2011) as well as during the entire process (Dogusoy-Taylan & Cagiltay, 2014) than that of the non-experts'.

Additionally, as a results of a literature review, it is reported that recording the eye-movements of surgeons may be beneficial both for skill assessment and training purposes (Hermens et al., 2013). It is also well-known that there is a need for better assessment methodologies for the skill-based training programs (Cagiltay et al., 2017). Other studies also report the potential benefits of captured gaze patterns to improve medical education either as part of an assessment system or in a gaze-training application (Di Stasi et al., 2017; Eivazi et al., 2017). Eye-tracking metrics, through virtually simulated and live environments are also suggested as objective measures for surgical skill level assessment purposes (Richstone et al., 2010).

Studies found in the literature show that eye-tracking systems are also used in laparoscopic/endoscopic surgery that is Minimally Invasive Surgery (MIS) in general. The validation of MIS training systems continues and these systems provide information on surgeons' skill levels, taking into account different measurements, namely time, motion economics and number of errors (Law et al., 2004). In addition to these measurements, the use of eye-movements in order to develop instructional modules for surgical training programs is under investigation by researchers (Hermens et al., 2013). Hence, understanding the behaviors and differences between intermediate and novice surgeons while performing surgical tasks is important for developing better assessment and instructional materials to support the surgical training programs.

However, a review study also reported that the vast majority of virtual-reality psychomotor skills tasks show construct validity for one or more metrics, mainly time and motion; however, there is a need for standardized proficiency scores should facilitate virtual reality-based laparoscopic psychomotor skills curricula (Sinitsky, Fernando, & Berlingieri, 2012). Earlier studies also reported that more work has to be done to understand how experienced surgeons attempt to overcome the perceptual difficulties inherent in the laparoscopic environment (M. R. Wilson, McGrath, et al., 2011). Similarly, based on the results of a systematic review conducted in 2013, it is concluded that eye-tracking provides reliable quantitative data as an objective assessment tool with potential applications in surgical training to improve performance. For this reason, this field remains as a promising area of research with the possibility of future implementations in surgical skills assessment (T. Tien et al., 2014).

MIS requires the ability to use both-hands at the same time. In MIS operations, surgeons use a surgical tool with one hand (usually the dominant one) and an endoscope (camera and light source) with the other (usually the non-dominant one), implying that they need to improve their eye-hand and left-right hand coordination skills. Accordingly, using dominant-hand, non-dominant hand or both-hands create mental workload at different levels which can be assessed by pupil size, number of fixation, fixation duration, saccade number, saccade duration, pursuit number and pursuit duration. Accordingly, it has been suggested that the differences in performances are due to information-processing capabilities of the left and right hemispheres of the brain (Hoffmann, 1997). It is also suggested that the times for hand movements should be similar when movements are made ballistically, and that when visual control is required, the preferred (dominant) hand will perform better (Hoffmann, 1997). When the non-preferred (non-dominant) hand and both-hands are used, then the times for performing the tasks increases (Hoffmann, 1997).

Earlier studies also report that expert behavior is manifested in distinct eye-movement patterns of proactivity, reactivity and suppression depending on the nature of the task and the presence of abnormalities at any given moment (Bertram, Helle, Kaakinen, & Svedström, 2013). Additionally, it is reported that gaze entropy and velocity were significantly higher when surgeons performed the most complex surgical procedure (Diaz-Piedra, Sanchez-Carrion, Rieiro, & Di Stasi, 2017). In complex environments, the number of fixations increase and fixations durations decrease (Vine et al., 2014). In the literature, there are promising results showing that training interventions designed to guide optimal gaze control may facilitate the performance of surgeons and improve computer-based simulation training programs (Behan & Wilson, 2008; Causer, Vickers, Snelgrove, Arsenault, & Harvey, 2014; Gegenfurtner, Lehtinen, Jarodzka, & Säljö, 2017; J. N. Vickers & Williams, 2007; Vine, Chaytor, McGrath, Masters, & Wilson, 2013; Vine et al., 2012; M. R. Wilson, Vine, et al., 2011).

However, in the literature there are very limited studies analyzing the mental workload of surgeons while performing surgical tasks in different hand conditions. Therefore, in this experimental study four simulated applications of a surgical task are presented with an eye-tracking approach to understand the mental workload in different hand conditions: the dominant-hand, the non-dominant hand, and both-hands. In the literature studies mostly conducted on the differences between the experts and novices in other domains but in this study it is aimed to understand the differences between intermediate and novice surgeons which is more difficult to differentiate.

Also, few studies have been attempted to better understand the eye-movements of the surgeons in the field of endo-neurosurgery. For this reason, in this study an endo-neurosurgery simulation environment is designed and developed to examine their eye and hand coordination skills. Accordingly, this work presents an eye-tracking initiative to determine the experience levels of surgeons and the effect of hand conditions (dominant, non-dominant and both-hand) from their gaze behaviors while performing interactive endo-neurosurgical tasks in different hand conditions.

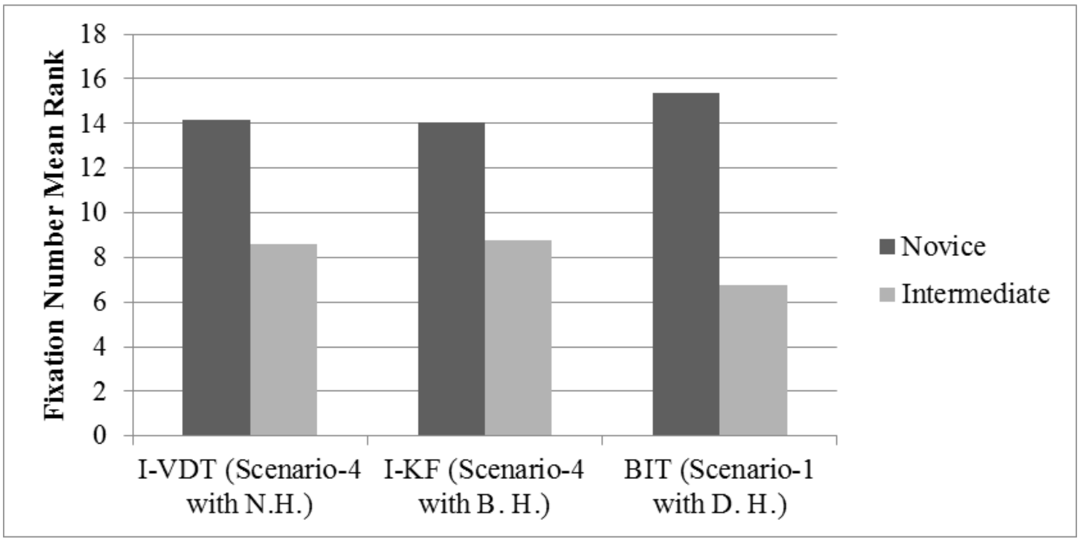
7.1 Results

Four simulation scenarios have been used for collecting data about intermediate and novice surgeons' eye gaze during operations in a virtually simulated environment and performed in different hand conditions. The main assumption of this study is that surgeons' performances in different hand conditions and their skill levels may also affect their eye-movements. Therefore, we aim to better understand, firstly, the differences between the eye-movements of intermediate and novice surgeons while they perform surgical tasks in a computer-simulated environment, and secondly the effect of hand condition (dominant-hand, non-dominant hand and both-hand) on their eye-movements. The results are analyzed using statistical methods to better understand the novice and intermediate surgeons' behaviors and hand condition effects in this environment. The recorded data was classified using open-source eye-movement classification algorithms.

As a result, 276 (23 surgeons, 4 tasks, and 3 hand conditions) datasets were recorded, which significantly increases the accuracy of the results in this work. To evaluate and compare the difference between intermediates and novices, various eye-movement events were analyzed: number of fixation, fixation duration, saccade number, saccade duration, pursuit number, pursuit duration, and pupil size.

7.1.1 Number of Fixation

According to the Mann Whitney test for novices and intermediates, there is a statistically significant difference in the number of fixation between the two groups as seen in Figure 7.1. Novice surgeons fixated longer than the intermediate surgeons according to results of the BIT algorithm for Scenario-1 with dominant-hand ($U = 16$, $p < 0.05$), the I-KF algorithm for Scenario-4 with both-hand ($U = 34$, $p < 0.05$), and the I-VDT algorithm for Scenario-4 with non-dominant hand ($U = 32.5$, $p < 0.05$). It is seen from the results that there is a difference between intermediate and novice surgeons based on their skill levels.



D.H.: Dominant-Hand; N.H.: Non-dominant hand; B.H.: Both-Hand

Figure 7.1 Number of Fixation Differences between Intermediate and Novice Surgeons

A non-parametric Friedman test of differences among repeated measures was conducted for the hand condition effect on number of fixation. For Scenario-1, the effect of hand condition was significant (in all $p < 0.05$) on the number of fixation according to the results of the algorithms I-HMM $\chi2(2) = 15.09$, I-MST $\chi2(2) = 9.689$, I-VDT $\chi2(2) = 21.478$, I-VMT $\chi2(2) = 30.484$, I-VT $\chi2(2) = 9.512$, I-VVT $\chi2(2) = 34.308$. For the algorithms BIT, I-DT, and I-KF the effect of hand condition was not significant (in all $p > 0.05$). For Scenario-2, the effect of hand condition was significant (in all $p < 0.05$) on the number of fixation according to the results of algorithms I-DT $\chi2(2) = 28.637$, I-HMM $\chi2(2) = 9.867$, I-KF $\chi2(2) = 9.191$, I-MST $\chi2(2) = 10.023$, I-VDT $\chi2(2) = 35.043$, I-VMT $\chi2(2) = 30.522$, I-VT $\chi2(2) = 10.747$, I-VVT $\chi2(2) = 35.670$. For the BIT algorithm, the effect of hand condition was not significant ($p > 0.05$). For Scenario-3, the effect of hand condition was significant (in all $p < 0.05$) on the number of fixation according to the results of algorithms BIT $\chi2(2) = 21.217$, I-DT $\chi2(2) = 25.438$, I-HMM $\chi2(2) = 19.727$, I-KF $\chi2(2) = 23.889$, I-MST $\chi2(2) = 26.847$, I-VDT $\chi2(2) = 19.126$, I-VMT $\chi2(2) = 4.963$, I-VT $\chi2(2) = 22.091$, I-VVT $\chi2(2) = 8.156$. For Scenario-4, the effect of hand condition was significant (in all $p < 0.05$) on the number of fixation according to the results of algorithms BIT $\chi2(2) = 13.889$, I-DT $\chi2(2) = 13.622$, I-HMM $\chi2(2) = 9.489$, I-KF $\chi2(2) = 7.622$, I-VDT $\chi2(2) = 23.143$, I-VMT $\chi2(2) = 13.241$, I-VT $\chi2(2) = 11.565$, I-VVT $\chi2(2) = 16.637$. For the I-MST algorithm, the effect of hand condition was not significant ($p > 0.05$). The AED algorithm was not reported for number of fixation because this algorithm does not measure this eye-movement event.

Table 7.1 Hand Condition Effect on Number of Fixation

	I-DT	I-HMM	I-KF	I-MST	I-VDT	I-VMT	I-VT	I-VVT	BIT
Scenario 1		NH		DH	DH	NH	DH	DH	
Scenario 2	BH	BH	BH	NH	BH	BH	NH	BH	
Scenario 3	BH	BH	BH	BH	BH		BH	BH	BH
Scenario 4	BH	NH	BH		BH	BH	NH	BH	BH

DH: Dominant-hand; NH: Non-dominant hand; BH: Both-hand

According to the results of each eye-movement classification algorithm for the number of fixation measure, the hand conditions that make this measure larger is reported in Table 7.1. Generally, in both-hand condition the number of fixation measure is becomes larger compared to the dominant- and non-dominant hand conditions as can be seen from Table 7.1.

7.1.2 Fixation Duration

According to the Mann Whitney test for novices and intermediates, there is a statistically significant difference in the fixation duration between the two groups as seen in Figure 7.2. Novice surgeons' fixation durations are longer than the intermediate surgeons according to the results of the I-VVT algorithm for Scenario-1 with dominant-hand ($U = 30$, $p < 0.05$) and the I-KF algorithm for Scenario-4 with non-dominant hand ($U = 30.5$, $p < 0.05$). Only the results of I-VMT algorithm for Scenario-3 with both-hand the fixation durations of intermediates are longer than the novice surgeons ($U = 26$, $p < 0.05$). It is seen from the results that there is a difference in fixation durations between intermediate and novice surgeons based on their skill levels.

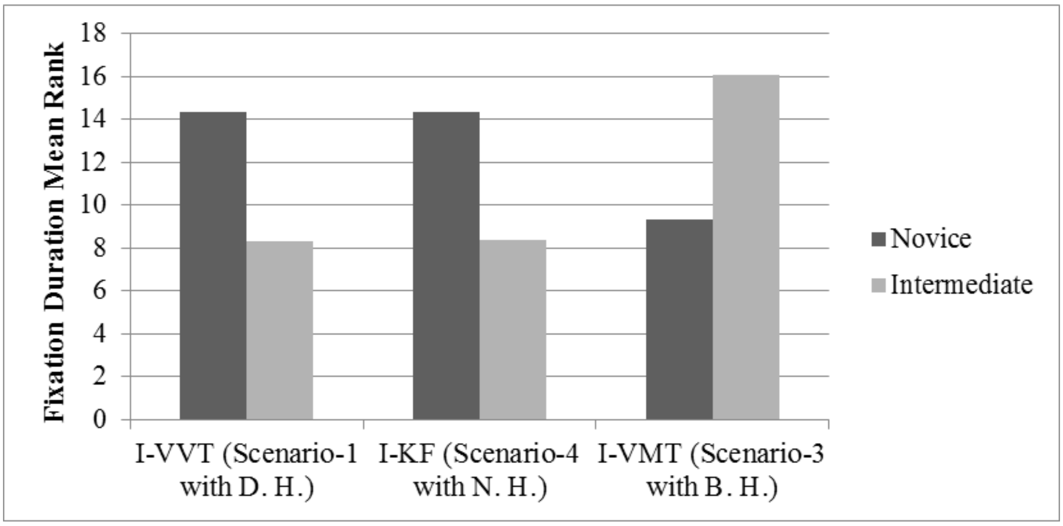


Figure 7.2 Fixation Duration Differences between Intermediate and Novice Surgeons

A non-parametric Friedman test of differences among repeated measures was conducted for the hand condition effect on fixation duration. For Scenario-1, the effect of hand condition was significant (in all $p < 0.05$) on the fixation duration according to the results of algorithms BIT $\chi2(2) = 31.391$, I-DT $\chi2(2) = 36.261$, I-HMM $\chi2(2) = 36.261$, I-KF $\chi2(2) = 36.261$, I-MST $\chi2(2) = 37.130$, I-VDT $\chi2(2) = 35.565$, I-VMT $\chi2(2) = 32.957$, I-VT $\chi2(2) = 36.261$, I-VVT $\chi2(2) = 32.957$. For algorithm AED, the effect of hand condition was not significant ($p > 0.05$). For Scenario-2, the effect of hand condition was significant (in all $p < 0.05$) on fixation duration according to the results of algorithms BIT $\chi2(2) = 34.696$, I-DT $\chi2(2) = 35.565$, I-HMM $\chi2(2) = 34.696$, I-KF $\chi2(2) = 34.696$, I-MST $\chi2(2) = 34.696$, I-VDT $\chi2(2) = 36.261$, I-VMT $\chi2(2) = 35.043$, I-VT $\chi2(2) = 35.043$, I-VVT $\chi2(2) = 35.565$. For algorithm AED, the effect of hand condition was not significant ($p > 0.05$). For Scenario-3, the effect of hand condition was significant (in all $p < 0.05$) on fixation duration according to the results of algorithms BIT $\chi2(2) = 21.913$, AED $\chi2(2) = 40.261$, I-DT $\chi2(2) = 17.304$, I-HMM $\chi2(2) = 17.043$, I-KF $\chi2(2) = 12.522$, I-MST $\chi2(2) = 21.217$, I-VDT $\chi2(2) = 11.565$, I-VT $\chi2(2) = 13.652$, I-VVT $\chi2(2) = 10.522$. Only for algorithm I-VMT, the effect of hand condition was not significant ($p > 0.05$). For Scenario-4, the effect of hand condition was significant (in all $p < 0.05$) on fixation duration according to the results of algorithms BIT $\chi2(2) = 31.391$, I-DT

$\chi2(2)$ = 21.913, I-HMM $\chi2(2)$ = 21.478, I-KF $\chi2(2)$ =18.348, I-MST $\chi2(2)$ = 22.261, I-VDT $\chi2(2)$ = 21.478, I-VMT $\chi2(2)$ = 10.738, I-VT $\chi2(2)$ = 21.478, I-VVT $\chi2(2)$ = 27.217. Only for algorithm AED, the effect of hand condition was not significant ($p > 0.05$).

Table 7.2 Hand Condition Effect on Fixation Duration

	AED	I-DT	I-HMM	I-KF	I-MST	I-VDT	I-VMT	I-VT	I-VVT	BIT
Scenario 1		NH	NH	NH	NH	NH	NH	DH	NH	DH
Scenario 2		BH	BH	BH	BH	BH	BH	BH	BH	BH
Scenario 3	DH		BH		BH	BH	NH	BH	NH	BH
Scenario 4		BH	BH	BH	BH	BH	BH	BH	BH	BH

According to the results for the fixation duration measure, the hand conditions that make this measure larger are reported in Table 7.2. Generally, in both-hand condition and non-dominant hand condition the fixation duration measure becomes larger compared to the dominant-hand condition.

7.1.3 Saccade Number

According to the Mann Whitney test for novices and intermediates, there is a statistically significant difference in the saccade numbers between these two groups as seen in Figure 7.3.

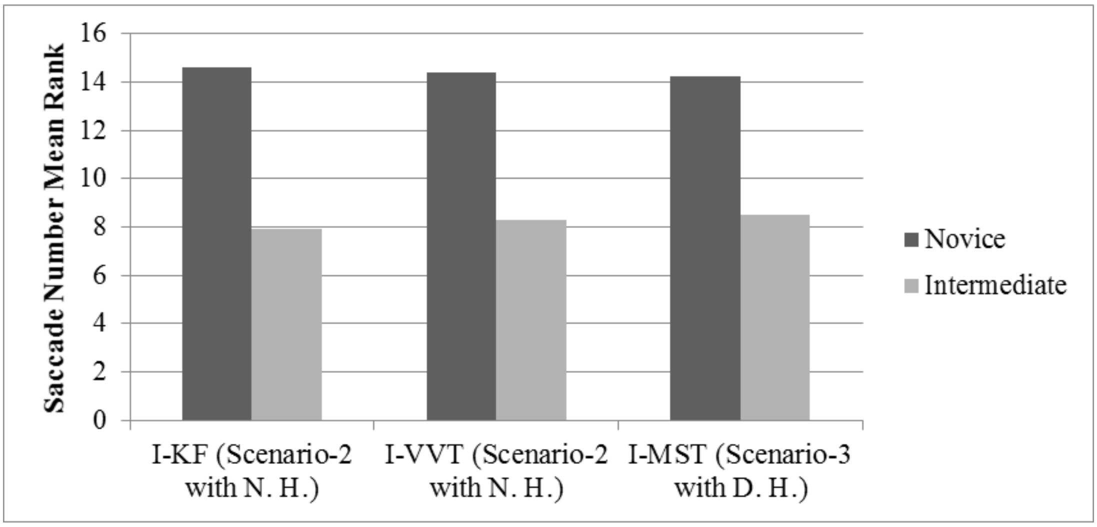


Figure 7.3 Saccade Number Differences between Intermediate and Novice Surgeons

Novice surgeons' saccade numbers are larger than the intermediate surgeons according to the results of the I-KF algorithm for Scenario-2 with non-dominant hand ($U = 26.5$, $p < 0.05$) the I-MST algorithm for Scenario-3 with dominant-hand ($U = 31.5$, $p < 0.05$), and the I-VVT algorithm for Scenario-2 with non-dominant hand ($U = 29.5$, $p < 0.05$). It is seen from the results that there is difference in saccade numbers between intermediate and novice surgeons based on their skill levels.

A non-parametric Friedman test of differences among repeated measures was conducted for the hand condition effect on the saccade number. For Scenario-1, the effect of hand condition was significant (in all $p < 0.05$) on saccade number according to the results of algorithms I-DT $\chi2(2) =14.174$, I-HMM $\chi2(2) = 25.622$, I-MST $\chi2(2) = 9.841$, I-VDT $\chi2(2) = 11.495$, I-VMT $\chi2(2) = 10.705$, I-VT $\chi2(2) = 15.187$, I-VVT $\chi2(2) = 11.596$. For algorithm BIT and I-KF, the effect of hand condition was not significant (in all $p > 0.05$). For Scenario-2, the effect of hand condition was significant (in all $p < 0.05$) on the saccade number according to the results of algorithms I-DT $\chi2(2) = 9.692$, I-HMM $\chi2(2) = 11.231$, I-KF $\chi2(2) = 21.303$, I-MST $\chi2(2) = 12.289$, I-VDT $\chi2(2) = 17.297$, I-VMT $\chi2(2) = 14.112$, I-VT $\chi2(2) = 18.870$, I-VVT $\chi2(2) = 18.870$. For algorithm BIT, the effect of hand condition was not significant ($p > 0.05$). For Scenario-3, the effect of hand condition was significant (in all $p < 0.05$) on the saccade number according to the results of algorithms BIT $\chi2(2) = 7.143$, I-DT $\chi2(2) = 25.473$, I-HMM $\chi2(2) = 22.422$, I-KF $\chi2(2) = 16.822$, I-MST $\chi2(2) = 15.846$, I-VDT $\chi2(2) = 19.609$, I-VMT $\chi2(2) = 19.846$, I-VT $\chi2(2) = 15.846$, I-VVT $\chi2(2) =24.261$. For Scenario-4, the effect of hand condition was significant (in all $p < 0.05$) on the saccade number according to the results of algorithms BIT $\chi2(2) = 10.522$, I-DT $\chi2(2) = 19.143$, I-HMM $\chi2(2) = 11.798$, I-KF $\chi2(2) = 8.022$, I-MST $\chi2(2) = 6.494$, I-VDT $\chi2(2) = 6.489$, I-VMT $\chi2(2) =8.273$, I-VT $\chi2(2) =12.154$. For algorithm I-VVT, the effect of hand condition was not significant ($p > 0.05$). The AED algorithm was not reported for saccade number because this algorithm does not measure this eye-movement event.

For the saccade number measure, the hand conditions that make this measure larger according to the results of each eye-movement classification algorithm are reported in Table 7.3. Generally, in both-hand condition and non-dominant hand condition the saccade number measure becomes larger compared to the dominant-hand condition.

Table 7.3 Hand Condition Effect on Saccade Number

	I-DT	I-HMM	I-KF	I-MST	I-VDT	I-VMT	I-VT	I-VVT	BIT
Scenario 1	BH	BH	BH	BH	BH	BH	BH		
Scenario 2	NH	NH	BH	NH	NH	NH	NH	NH	
Scenario 3	BH	BH	BH	BH	BH	BH	BH	BH	BH
Scenario 4	NH	NH	NH	NH	NH	NH	NH		DH

7.1.4 Saccade Duration

According to the Mann Whitney test for novices and intermediates, there is a statistically significant difference in the saccade durations between these two groups as seen in Figure 7.4. Novice surgeons' saccade durations are larger than the intermediate surgeons based on the results of the AED algorithm for Scenario-1 with Dominant-hand ($U = 31.5$, $p < 0.05$) and Scenario-1 with Non-dominant hand ($U = 34.5$, $p < 0.05$). It is seen from the results that there is a difference in the saccade durations between intermediate and novice surgeons based on their skill levels.

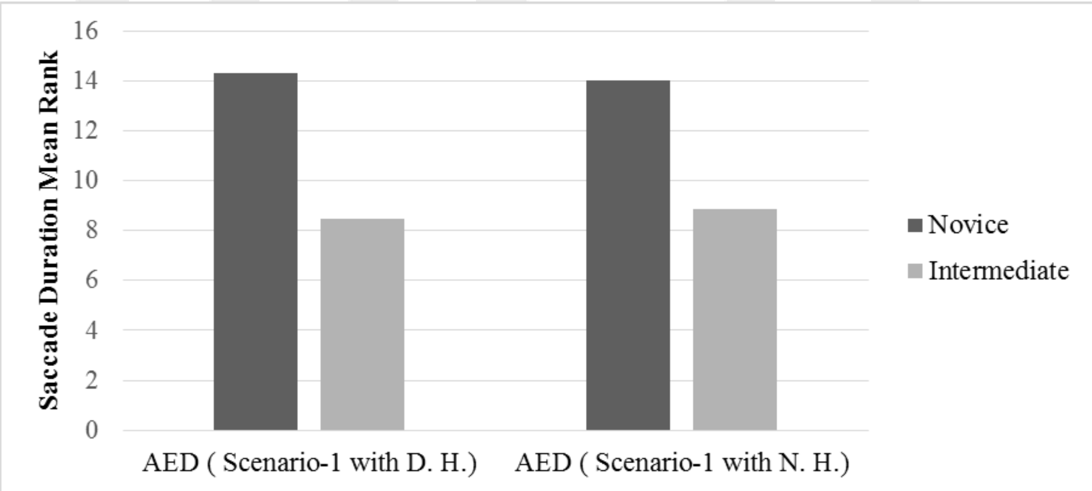


Figure 7.4 Saccade Duration Differences between Intermediate and Novice Surgeons

A non-parametric Friedman test of differences among repeated measures was conducted for the hand condition effect on saccade duration. For Scenario-1, the effect of hand condition was significant (in all $p < 0.05$) on the saccade duration according to the results of algorithms AED $\chi2(2) = 26.584$, I-DT $\chi2(2) = 16.435$, I-HMM $\chi2(2) = 8.600$, I-KF $\chi2(2) = 7.187$, I-MST $\chi2(2) = 9.708$, I-VDT $\chi2(2) = 8.330$, I-VMT $\chi2(2) = 11.217$, I-VT $\chi2(2) = 11.934$, I-VVT $\chi2(2) = 11.217$. For Scenario-2, the effect of hand condition was significant (in all $p < 0.05$) on the saccade duration according to the results of algorithms AED $\chi2(2) = 36.609$, I-DT $\chi2(2) = 14.957$, I-HMM $\chi2(2) = 17.478$, I-KF $\chi2(2) = 18.879$, I-MST $\chi2(2) = 15.913$, I-VDT $\chi2(2) = 14.957$, I-VMT $\chi2(2) = 12.783$, I-VT $\chi2(2) = 18.467$, I-VVT $\chi2(2) = 19.934$. For Scenario-3, the effect of hand condition was significant (in all $p < 0.05$) on the saccade duration according to the results of algorithms AED $\chi2(2) = 38.289$, I-DT $\chi2(2) = 23.363$, I-HMM $\chi2(2) = 20.957$, I-KF $\chi2(2) = 22.522$, I-MST $\chi2(2) = 11.565$, I-VDT $\chi2(2) = 16.435$, I-VMT $\chi2(2) = 19.913$, I-VT $\chi2(2) = 12.087$, I-VVT $\chi2(2) = 18.087$. For Scenario-4, the effect of hand condition was significant (in all $p < 0.05$) on the saccade duration according to the results of algorithms I-DT $\chi2(2) = 19.143$, I-MST $\chi2(2) = 10.422$. However, the effect of hand condition was not significant (in all $p > 0.05$) on the saccade duration according to the results of algorithms AED, I-HMM, I-KF, I-VDT, I-VMT, I-VT, and I-VVT. The BIT algorithm was not reported for the saccade duration because this algorithm does not classify this eye-movement event.

Table 7.4 Hand Condition Effect on Saccade Duration

	AED	**I-DT**	**I-HMM**	**I-KF**	**I-MST**	**I-VDT**	**I-VMT**	**I-VT**	**I-VVT**
Scenario 1	DH	BH	BH	BH	BH	BH	BH	BH	
Scenario 2	DH	NH	NH	BH	NH	NH	NH	NH	NH
Scenario 3	NH		BH		BH	BH	BH	BH	BH
Scenario 4		NH		NH	NH				

According to the results of each eye-movement classification algorithm for the saccade duration measure the hand conditions that makes this measure larger is reported in Table 7.4. Generally, in both-hand condition and non-dominant hand condition the saccade duration measure is becoming larger comparing to the dominant-hand condition.

7.1.5 Pursuit Number

As seen from Figure 7.5, based on the Mann Whitney test for novices and intermediates, there is a statistically significant difference in the pursuit numbers between these two groups. The novice surgeons' pursuit numbers are larger than the intermediate surgeons according to the results of the I-VMT algorithm for Scenario-1 with dominant-hand ($U = 32$, $p < 0.05$), the I-VMT algorithm for Scenario-2 with non-dominant hand ($U = 31$, $p < 0.05$) and the I-VVT algorithm for Scenario-2 with non-dominant hand ($U = 29.5$, $p < 0.05$). Only in the I- VVT algorithm for Scenario-1 with dominant-hand ($U = 31$, $p < 0.05$) are the pursuit numbers of intermediates more than the novice surgeons. It is seen from the results that there is a significant difference in pursuit numbers between intermediate and novice surgeons based on their skill levels.

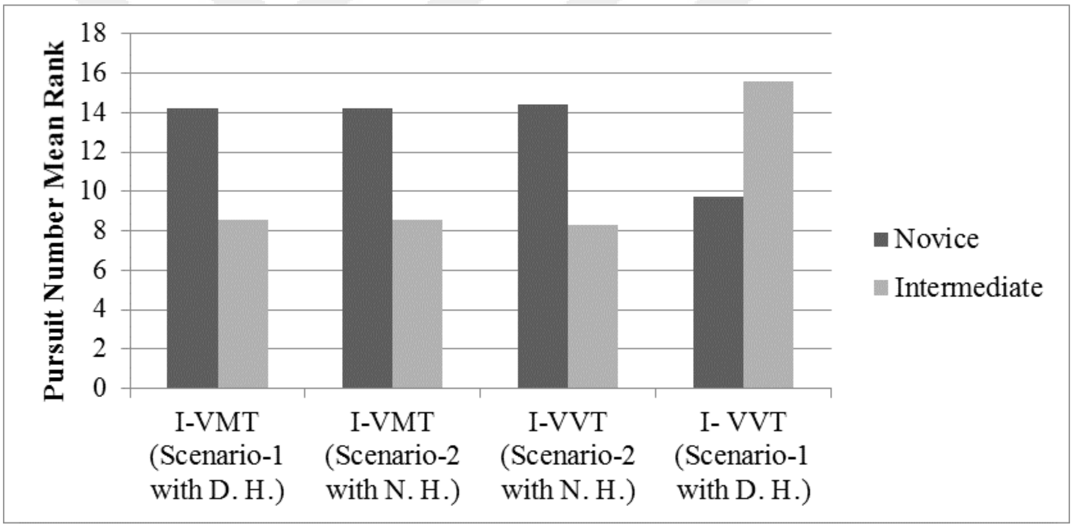


Figure 7.5 Pursuit Number Differences between Intermediate and Novice Surgeons

A non-parametric Friedman test of differences among repeated measures was conducted for the hand condition effect on pursuit number. For Scenario-1, the effect of hand condition was significant (in all $p < 0.05$) on the pursuit number according to the results of algorithms I-VDT $\chi2(2) = 16.242$, I-VMT $\chi2(2) = 6.000$ and I-VVT $\chi2(2) = 32.435$. For Scenario-2, the effect of hand condition was significant (in all $p < 0.05$) on the pursuit number according to the results of algorithms I-VMT $\chi2(2) =$

18.087 and I-VVT $\chi2(2) = 27.478$. However, the effect of hand condition was not significant ($p > 0.05$) on the pursuit number according to the results of the I-VDT algorithm. For Scenario-3, the effect of hand condition was significant (in all $p < 0.05$) on the pursuit number according to the results of algorithms I-VDT $\chi2(2) = 18.747$, I-VMT $\chi2(2) = 21.478$, and I-VVT $\chi2(2) = 26.308$. For Scenario-4, the effect of hand condition was significant (in all $p < 0.05$) on the pursuit number according to the results of algorithms I-VDT $\chi2(2) = 17.868$, I-VMT $\chi2(2) = 12.202$, and I-VVT $\chi2(2) = 22.220$. Algorithms BIT, AED, I-DT, I-HMM, I-KF, I-MST, and I-VT were not reported for the pursuit number because they do not classify this eye-movement event.

Table 7.5 Hand Condition Effect on Pursuit Number

	I-VDT	**I-VMT**	**I-VVT**
Scenario 1	BH	BH	DH
Scenario 2		NH	BH
Scenario 3	BH	BH	BH
Scenario 4	NH	NH	NH

Table 7.5 reports the larger pursuit number measures according to the hand conditions. Generally, in both-hand condition and non-dominant hand condition the pursuit number measure becomes larger compared to the dominant-hand condition.

7.1.6 Pursuit Duration

According to the Mann Whitney test for novice and intermediate surgeons, there is a statistically significant difference in the pursuit durations between these two groups as seen from Figure 7.6.

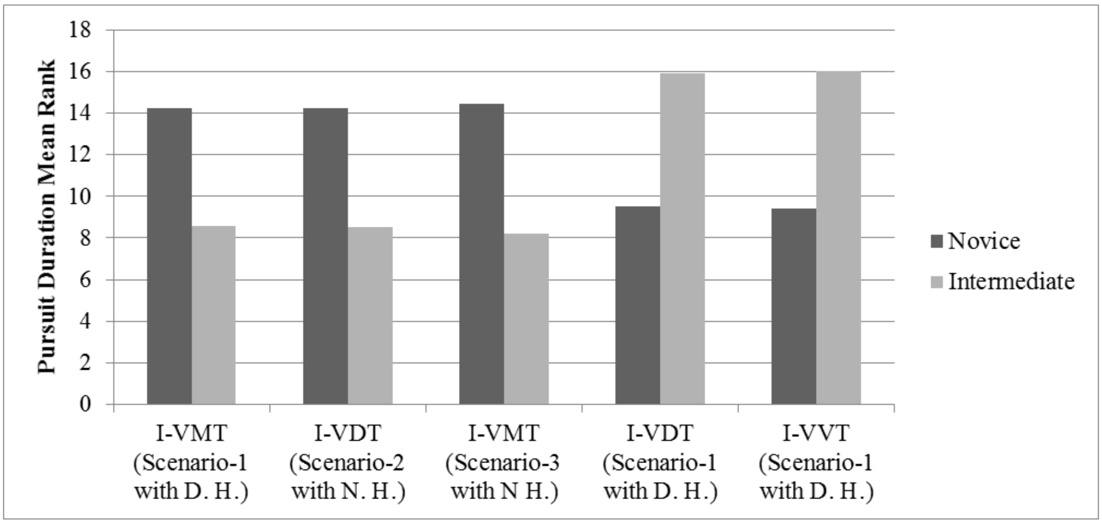


Figure 7.6 Pursuit Duration Differences between Intermediate and Novice Surgeons

Novice surgeons' pursuit durations are larger than the intermediate surgeons based on the results of the I-VMT algorithm for Scenario-1 with dominant-hand ($U = 32$, $p < 0.05$), the I-VMT algorithm for Scenario-3 with non-dominant hand ($U = 29$, $p < 0.05$) and the I-VDT algorithm for Scenario-2 with non-dominant hand ($U = 31.5$, $p < 0.05$). However, the pursuit durations of intermediate surgeons are longer than the novices according to the results of I-VDT algorithm for Scenario-1 with dominant-hand ($U = 28$, $p < 0.05$) and the I-VVT algorithm for Scenario-1 with dominant-hand ($U = 27$, $p < 0.05$). It can be concluded from the results that there is a significant difference in pursuit durations between intermediate and novice surgeons based on their skill levels.

A non-parametric Friedman test of differences among repeated measures was conducted for hand condition effect on the pursuit duration. For Scenario-1, the effect of hand condition was significant (in all $p < 0.05$) on the pursuit duration according to the results of algorithms I-VDT $\chi2(2) = 10.174$, I-VMT $\chi2(2) = 33.652$, and I-VVT $\chi2(2) = 28.261$. For Scenario-2, the effect of hand condition was significant (in all $p < 0.05$) on the pursuit duration according to the results of algorithms I-VDT $\chi2(2) = 14.174$, I-VMT $\chi2(2) = 19.043$, and I-VVT $\chi2(2) = 15.913$. For Scenario-3, the effect of hand condition was significant (in all $p < 0.05$) on the pursuit duration according to the results of algorithms I-VDT $\chi2(2) = 21.478$, I-VMT $\chi2(2) = 24.609$, and I-VVT $\chi2(2) = 21.478$. For Scenario-4, the effect of hand condition was significant (in all $p < 0.05$) on the pursuit duration according to the results of algorithms I-VDT $\chi2(2) = 15.913$, I-VMT $\chi2(2) = 8.696$, and I-VVT $\chi2(2) = 11.934$. The algorithms BIT, AED, I-DT, I-HMM, I-MST, I-KF, and I-VT were not reported for pursuit durations because they do not classify this eye-movement event.

Table 7.6 Hand Condition Effect on Pursuit Duration

	I-VDT	**I-VMT**	**I-VVT**
Scenario 1	BH	NH	DH
Scenario 2	NH	BH	BH
Scenario 3	BH	BH	BH
Scenario 4	NH	BH	NH

According to the results of each eye-movement classification algorithm for the pursuit duration measure, the hand conditions that make this measure larger are reported in Table 7.6. Generally, in both-hand condition and non-dominant hand condition, the pursuit duration measure tends to become larger compared to the dominant-hand condition as can be seen the table.

7.1.7 Pupil Size

According to the Mann Whitney test for novices and intermediates, there is a statistically significant difference in the pupil sizes between the intermediate and novice surgeons as seen from Figure 7.7. Novice surgeons right eye pupil sizes become larger than the intermediate surgeons according to the pupil size measurement results of Scenario-1 with dominant-hand ($U = 31$, $p < 0.05$). It is seen from the results that there is difference in pupil sizes between intermediate and novice surgeons based on their skill levels.

Figure 7.7 Pupil Size Differences between Intermediate and Novice Surgeons

A non-parametric Friedman test of differences among repeated measures was conducted for the hand condition effect on the pupil size. For Scenario-1, the effect of hand condition was significant (in all $p < 0.05$) on the pupil size of left eye $\chi2(2) = 26.174$, the pupil size of the right eye $\chi2(2) = 37.130$ and the average pupil size of both left and right eye $\chi2(2) = 35.826$. For Scenario-2, the effect of hand condition was significant (in all $p < 0.05$) on the pupil size of the left eye $\chi2(2) = 11.565$, the pupil size of the right eye $\chi2(2) = 16.435$ and the average pupil size of both left and right eye $\chi2(2) = 11.565$. For Scenario-3, the effect of hand condition was significant (in all $p < 0.05$) on the pupil size of the right eye $\chi2(2) = 7.913$ and the average pupil size of both left and right eye $\chi2(2) = 6.870$. However, the effect of hand condition was not significant (in all $p > 0.05$) on the left eye pupil size. For Scenario-4, the effect of hand condition was significant (in all $p < 0.05$) on the pupil size of the left eye $\chi2(2) = 18.087$, the pupil size of the right eye $\chi2(2) = 22.261$ and the average pupil size of both left and right eye $\chi2(2) = 19.826$.

According to the pupil sizes measured by the eye-tracker, the hand conditions that make this measure larger are reported in Table 7.7. Generally, in both-hand condition the pupil sizes are become larger compared to the dominant- and non-dominant hand conditions.

Table 7.7 Hand Condition Effect on Pupil Size

	LE	RE	AVG
Scenario 1	BH	BH	BH
Scenario 2	BH	BH	BH
Scenario 3		BH	BH
Scenario 4	NH	BH	BH

LE: Left eye Pupil Size; RE: Right eye Pupil Size; AVG: Average of Left and Right eye Pupil Size

7.2 Discussions

This study has two purposes: the first one is to examine the differences between the eye-movements of intermediate and novice surgeons in a simulated virtual environment; the second one is to analyze the hand condition effect on these surgeons' eye-movements. The study simulates endo-neurosurgery tasks and records the gaze data of the surgeons. Such data can vary depending on the experience of the subjects, thereby providing useful information by which to train basic MIS skill development. Specifically, the present work focusses on the common eye-movements, namely number of fixation, fixation duration, saccade number, saccade duration, pursuit number, pursuit duration and pupil sizes.

According to the results of this study, it is found that novice surgeons tend to fixate more on different objects than the intermediates. Also, they spend more time fixating on an object. This result supports that of earlier studies reporting that non-experts split their time between focusing on the targets and tracking the tools (M. R. Wilson, McGrath, et al., 2011); in addition the fixation durations of experts is shorter than non-experts (Gegenfurtner et al., 2011). The main reason for this behavior could be that, as the experience level increases, the ability to perceive information is also develops, thus reducing the fixation duration among intermediate surgeons is becoming shorter. One study (Kasarskis et al., 2001) showed the differences in performance between expert and novice pilots based on their eye-movements while conducting a landing operation in a simulator. Unsurprisingly, expert pilot's performances were better than the novices, and it was also found that the fixation durations of experts were shorter than the novice pilots. This example highlights that expert pilots assemble the necessary information quickly than the novice ones.

Another study (J. Vickers, 1995) showed that expert basketball players tend to have shorter fixation durations than novice players. Expert players use their visual system to program their motor system, implying that they do not need to fixate at a point longer (J. Vickers, 1995). On the other hand, novice players tend to fixate at a point until shooting the ball (J. Vickers, 1995). Therefore, as it is found in this study, expert players are likely to have shorter fixation durations because they can gather the required information easily and quickly than the novices.

Saccade numbers and saccade durations also differ between intermediate and novice surgeons according to the results of the analysis. Novice surgeons make more saccadic movements compared to the intermediate surgeons. Also, the saccade durations of novice surgeons are longer than the intermediates. Supportively, the study (J. Vickers, 2003) shows this by comparing the eye-movements of expert and novice golf players, concluding that novice players tended to make more saccades. Whereas expert players made fewer such movements between different locations in order to reduce the memory impairment of distance clues the most (J. Vickers, 2003). Another study showed that, as well as performing the tasks better, expert surgeons made more efficient eye-movements with fewer number of saccades between different objects (Koh, Park, Wickens, Ong, & Chia, 2011). Also expert's saccadic rates were lower and the peak velocities were higher according to the novice surgeons. In other studies (G. Tien, Atkins, & Zheng, 2012; G. Tien et al., 2010; T. Tien et al., 2015), experts were shown to distribute their attention more effectively by looking at the necessary locations of a sight at the right time, and attention of the experts tends to be more compact and locally defined, with the result that the focus of attention does not change as often as the novices.

According to the results of this study, the differences in the pursuit numbers between intermediate and novice surgeons were significant. Novice surgeons tend to follow more objects than the intermediates. According to the study (Law et al., 2004), the gaze behaviors of expert and novice surgeons are distinctly different. It is stated that there were differences between expert and novice surgeons eye gaze behavior types, such that novices tend to make more saccades than experts (Law et al., 2004). Also, there is a difference between the target gaze and tool following behavior, experts tend to fixate on the target more frequently than novices, but for the tool following behavior experts tend to follow the object less frequently than the novices (Law et al., 2004). Due to their unfamiliarity with surgical tools, novice surgeons need to take more visual feedback to perform surgical tasks. As a result, they tend to follow the operational tools more frequently than the expert surgeons (Law et al., 2004). This practices represent an opportunity to learn how to coordinate hand and tool movements.

Furthermore, pupil sizes show vary depending on the experience levels. The pupil size of novice surgeons is larger than that of the intermediate ones while performing a surgical task. The reason for this is that pupil sizes grow as the mental workload increases. While performing the same simulation scenario, the pupil sizes of the novices increase, because of additional workload compared to the intermediate surgeons. This measure is also an important finding as a way to determine the experience levels of individuals, for example if they are novices, intermediates, or experts.

What is more, hand condition has a significant effect on the eye-movement behaviors of surgeons. According to the results, under the both-hand and non-dominant conditions the pupil sizes, numbers of fixation, saccade and pursuit events and also their durations increase compared to the dominant-hand condition. Supportively, in the literature it is stated that the non-dominant hand has larger force production variability and needs more corrections, thereby requiring greater movement times (Annett, Annett, Hudson, & Turner, 1979). Another study stated that the dominant-hand performance was significantly faster and always superior to the non-dominant hand (Hoffmann, 1997).

7.3 Conclusion

Parallel to earlier studies on eye-movement, the results of this study show that there are significant differences in the eye behaviors' of surgeons having different experience levels. It has been suggested that by better understanding the differences in the eye behaviors of surgeons having different skill levels, appropriate assessment tools and instructional systems can be developed in order to train and improving the skill levels of surgeons.

However, the critical issue for integrating these technologies into the curriculum of surgical training programs is the appropriateness of the simulation tools which appropriately adapts these strategies into such programs. Our results show that, when the tasks are performed under the both-hand condition and the non-dominant hand condition, pupil sizes, number of fixations, saccades, pursuit events and their durations increase when compared to the dominant-hand condition. Most probably, this is because of the level of task complexity. Hence, there is an urgent need to develop strategies to better understand the level of task complexity as well as the acceptable threshold values for different performance indicators of the tasks specific to different surgical skill levels.

Current computer-based simulation technologies are provide solutions to implement such strategies; however, there is a need to adapt this strategy to software development processes. In other words, the software development methodology for surgical simulation software for training purposes needs to add one more step for evaluating the simulated tasks with surgeons having different skill levels and developing task-specific threshold values for each performance indicator (e.g., task completion time, error rates, eye-movement events, hand movements, so on.) assessed in the simulation software. Afterwards, these task-specific threshold values can be integrated to the simulation software to better guide and assess the trainees' skill development. Hence, once such standardized threshold values are arranged specific to the tasks being performed, the potential increases as to use of this information to better guide and train surgeons and improve the performance of the surgical education programs.

CHAPTER 8

DISCUSSION AND CONCLUSION

Main contribution of this study is twofold. Firstly, a conceptual framework for the Context Aware Adaptive Software System (CAASS) is proposed. Secondly, a case study on this framework is conducted.

The framework consists of three main components that need to be defined for a specific CAASS namely context of adaptation, what-to-adapt and how-to-adapt. Accordingly, first the context, such as user, physical, primary and computing parameters, that is being addressed by the specific CAASS is need to be defined. Secondly, the description of the entities that are aimed to be adapted, is need to be defined. A single CAASS may address a unique adaptation entity, or may be aimed to address several adaptation entities such as content, network performance and user interface. Accordingly the target adaptation entity/entities should be clearly defined in this stage. The last component is the definition showing how the adaptation process is performed. Here, what events are aimed to be monitored, how the data is planned to be collected and analyzed as well as how the decisions are aimed to be made for the adaptation process are all should be described. Hence again, a specific CAASS may be designed to monitor several different forms of data and apply different analysis processes. These methods are important to differentiate and analyze the structures of CAASS. We believe that by applying this conceptual framework, the structure and behavior of CAASS can be designed and understood in a better way.

The case study is conducted in the field of endo-neurosurgery to better understand the surgical residents' behaviors and develop guidance for an appropriate CAASS. Their eye-movement data was collected with an eye-tracker device and classified into eye-movement classification algorithms. The obtained data from eye-movement classification algorithms was investigated for understanding the task difficulties, surgical skill levels, mental workload, effect of hand condition and effect of experience levels on eye-movement events. These are the core elements for this case study for monitor issues in a CAASS. Monitoring and analyzing these context elements can help developers to better understand and develop CAASS for the endo-neurosurgery domain. Accordingly, the results of this study shows that gathering detailed information from surgical simulation environments, can provide insights about behaviors of the surgeons. It is shown that, scenario difficulty levels and surgical skill levels can be assessed from eye-movement events. Also, a significant difference between the intermediate and novice surgeons' eye-movement behaviors has been detected. The hand condition significantly effects the eye-movements and pupil sizes of surgical residents. Eye-movement data can be used to assess the task difficulty, skill levels and mental workloads of surgeons. These findings can be used as an objective measure for better guiding the trainees and adapting the scenarios to the surgeons' performance in surgical education programs. The results of this study show promising findings on developing CAASS for the endo-neurosurgery training programs. In other words, by monitoring eye-movement events of the surgeons a CAASS can be developed to adapt the sequence and difficulty levels of the training scenarios according to the skill-levels of the trainees. Through such a CAASS, surgeons' eye-movements can be collected. While the surgeons are performing the surgical tasks, their eye data can be collected and monitored. Afterwards, the system may dynamically adapt the scenarios by considering their difficulty levels and the eye behaviors of the surgeons. In other words, by analyzing the data obtained from the eye-movements, mental workloads of surgeons can be assessed. Accordingly, their skills that need to be developed can be detected and an appropriate task schedule can be developed for each individual. This will in turn help the surgical programs to better provide individualized technology enhanced skill-based training.

CHAPTER 9

LIMITATIONS AND FUTURE WORK

This study has some limitations such as the procedure applied can be repeated with higher number of participants and more simulation scenarios can be developed to address different skills and skill levels of surgical residents. Also, the accuracy depends on properties of the eye tracker, more accurate eye-tracking devices can be used to provide high statistical power to the analyses. According to the results of this study it is seen that based on the methods and threshold values of eye-movement classification algorithms their classification results varies. Thus, eye-movement classification algorithm threshold values can be specified for better evaluating these algorithms. What is more, the luminance conditions, time of the day and dominant hand may affect the pupil sizes of the surgical residents while performing the scenarios. Also, the order of the hand conditions (dominant or non-dominant and both hand) may have an effect on the eye-movements of the surgical residents.

This study shows that a CAASS can be developed by monitoring surgical residents' eye-movement behaviors. It is shown that, through this data surgeons' behaviors, and skill levels can be differentiated. Additionally, this data provides information to measure the task difficulty levels objectively. However, monitoring for such a CAASS can also be supported by some other data such as hand movements and task progress. Accordingly, the future studies may also be conducted to better understand the evidences from other data sources. This is potentially may improve the adaptation level of a CAASS for the surgical training programs. Furthermore, these findings can be used to develop a CAASS for surgical training programs. Our long-term goal is to develop a comprehensive CAASS by using these results. We believe that such a system potentially will improve the skill levels of surgical residents which in turn will reduce the error rates in the operating room and prevent harm to the patients.

REFERENCES

Abowd, G., Dey, A., Brown, P., Davies, N., Smith, M., & Steggles, P. (1999). *Towards a better understanding of context and context-awareness.* Paper presented at the Handheld and ubiquitous computing.

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics, 36*(7), 623-636.

Ahmidi, N., Ishii, M., Fichtinger, G., Gallia, G. L., & Hager, G. D. (2012). *An objective and automated method for assessing surgical skill in endoscopic sinus surgery using eye- tracking and tool- motion data.* Paper presented at the International forum of allergy & rhinology.

Albano, G., Gaeta, M., & Ritrovato, P. (2007). IWT: an innovative solution for AGS e-Learning model. *International Journal of Knowledge and Learning, 3*(2-3), 209-224.

Albano, G., Gaeta, M., & Salerno, S. (2006). E-learning: a model and process proposal. *International Journal of Knowledge and Learning, 2*(1-2), 73-88.

Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., & Nyström, M. (2017). One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior research methods, 49*(2), 616-637.

Andreu-Perez, J., Solnais, C., & Sriskandarajah, K. (2016). EALab (Eye Activity Lab): a MATLAB toolbox for variable extraction, multivariate analysis and classification of eye-movement data. *Neuroinformatics, 14*(1), 51-67.

Annett, J., Annett, M., Hudson, P., & Turner, A. (1979). The control of movement in the preferred and non-preferred hands. *The Quarterly journal of experimental psychology, 31*(4), 641-652.

Atkins, M. S., Tien, G., Khan, R. S., Meneghetti, A., & Zheng, B. (2013). What do surgeons see: capturing and synchronizing eye gaze for surgery applications. *Surgical innovation, 20*(3), 241-248.

Babaoglu, O., Jelasity, M., Montresor, A., Fetzer, C., Leonardi, S., van Moorsel, A., & van Steen, M. (2005). The self-star vision. *Self-star properties in complex information systems*, 397-397.

Bäck, T. (2002). Adaptive business intelligence based on evolution strategies: some application examples of self-adaptive software. *Information Sciences, 148*(1), 113-121.

Bailey, B. P., & Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction (TOCHI), 14*(4), 21.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin, 91*(2), 276.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. *Handbook of psychophysiology, 2*, 142-162.

Bedell, H. E., & Stevenson, S. B. (2013). Eye movement testing in clinical examination. *Vision research, 90*, 32-37.

Behan, M., & Wilson, M. (2008). State anxiety and visual attention: The role of the quiet eye period in aiming to a far target. *Journal of Sports Sciences, 26*(2), 207-215.

Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., & Montanari, R. (2011). Driver workload and eye blink duration. *Transportation research part F: traffic psychology and behaviour, 14*(3), 199-208.

Berger, C., Winkels, M., Lischke, A., & Höppner, J. (2012). GazeAlyze: a MATLAB toolbox for the analysis of eye movement data. *Behavior research methods, 44*(2), 404-419.

Bertram, R., Helle, L., Kaakinen, J. K., & Svedström, E. (2013). The effect of expertise on eye movement behaviour in medical image perception. *PloS one, 8*(6), e66169.

Biegel, G., & Cahill, V. (2004). *A framework for developing mobile, context-aware applications.* Paper presented at the Pervasive Computing and Communications, 2004. PerCom 2004. Proceedings of the Second IEEE Annual Conference on.

Bolding, M. S., Lahti, A. C., White, D., Moore, C., Gurler, D., Gawne, T. J., & Gamlin, P. D. (2014). Vergence eye movements in patients with schizophrenia. *Vision research, 102*, 64-70.

Bradshaw, J. (1967). Pupil size as a measure of arousal during information processing. *Nature, 216*(5114), 515-516.

Brouwer, A.-M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural engineering, 9*(4), 045008.

Bröhl, C., Theis, S., Rasche, P., Wille, M., Mertens, A., & Schlick, C. M. (2017). Neuroergonomic analysis of perihand space: effects of hand proximity on eye-tracking measures and performance in a visual search task. *Behaviour & Information Technology*, 1-8.

Brusilovsky, P. (1998). Methods and techniques of adaptive hypermedia. In *Adaptive hypertext and hypermedia* (pp. 1-43): Springer.

Brusilovsky, P., Eklund, J., & Schwarz, E. (1998). Web-based education for all: a tool for development adaptive courseware. *Computer Networks and ISDN Systems, 30*(1-7), 291-300.

Brusilovsky, P., & Vassileva, J. (2003). Course sequencing techniques for large-scale web-based education. *International Journal of Continuing Engineering Education and Life Long Learning, 13*(1-2), 75-94.

Cagiltay, N. E., Ozcelik, E., Sengul, G., & Berker, M. (2017). Construct and face validity of the educational computer-based environment (ECE) assessment scenarios for basic endoneurosurgery skills. *Surgical endoscopy*, 1-11.

Carchiolo, V., Longheu, A., & Malgeri, M. (2002). Adaptive formative paths in a web-based learning environment. *Educational Technology & Society*, *5*(4), 64-75.

Cassenti, D. N., & Kelley, T. D. (2006). *Towards the shape of mental workload.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Causer, J., Vickers, J. N., Snelgrove, R., Arsenault, G., & Harvey, A. (2014). Performing under pressure: Quiet eye training improves surgical knot-tying performance. *Surgery, 156*(5), 1089-1096.

Ceri, S., Daniel, F., Facca, F. M., & Matera, M. (2007). Model-driven engineering of active context-awareness. *World Wide Web, 10*(4), 387-413.

Charles, D., Kerr, A., McNeill, M., McAlister, M., Black, M., Kcklich, J., Stringer, K. (2005). *Player-centred game design: Player modelling and adaptive digital games.* Paper presented at the Proceedings of the digital games research conference.

Cheng, B. H., De Lemos, R., Giese, H., Inverardi, P., Magee, J., Andersson, J., Cukic, B. (2009). Software engineering for self-adaptive systems: A research roadmap. In *Software engineering for self-adaptive systems* (pp. 1-26): Springer.

Cope, A. C., Mavroveli, S., Bezemer, J., Hanna, G. B., & Kneebone, R. (2015). Making meaning from sensory cues: A qualitative investigation of postgraduate learning in the operating room. *Academic Medicine, 90*(8), 1125-1131.

Coyne, J., & Sibley, C. (2016). *Investigating the Use of Two Low Cost Eye Tracking Systems for Detecting Pupillary Response to Changes in Mental Workload.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Crawford, T. J., Higham, S., Renvoize, T., Patel, J., Dale, M., Suriya, A., & Tetley, S. (2005). Inhibitory control of saccadic eye movements and cognitive impairment in Alzheimer's disease. *Biological psychiatry, 57*(9), 1052-1060.

Crundall, D., Underwood, G., & Chapman, P. (1999). Driving experience and the functional field of view. *Perception, 28*(9), 1075-1087.

Csikszentmihalyi, M., & Csikszentmihalyi, I. S. (1992). *Optimal experience: Psychological studies of flow in consciousness*: Cambridge university press.

da Silva, C. E., & de Lemos, R. (2011). A framework for automatic generation of processes for self-adaptive software systems. *Informatica, 35*(1).

de Greef, T., Lafeber, H., van Oostendorp, H., & Lindenberg, J. (2009). Eye movement as indicators of mental workload to trigger adaptive automation. *Foundations of augmented cognition. Neuroergonomics and operational neuroscience*, 219-228.

Dey, A. K. (2001). Understanding and using context. *Personal and ubiquitous computing, 5*(1), 4-7.

Di Nitto, E., Ghezzi, C., Metzger, A., Papazoglou, M., & Pohl, K. (2008). A journey to highly dynamic, self-adaptive service-based applications. *Automated Software Engineering, 15*(3), 313-341.

Di Stasi, L. L., Díaz-Piedra, C., Ruiz-Rabelo, J. F., Rieiro, H., Carrion, J. M. S., & Catena, A. (2017). Quantifying the cognitive cost of laparo-endoscopic single-site surgeries: gaze-based indices. *Applied Ergonomics, 65*, 168-174.

Diaz-Piedra, C., Sanchez-Carrion, J. M., Rieiro, H., & Di Stasi, L. L. (2017). Gaze-based Technology as a Tool for Surgical Skills Assessment and Training in Urology. *Urology, 107*, 26-30.

Dobson, S., Sterritt, R., Nixon, P., & Hinchey, M. (2010). Fulfilling the vision of autonomic computing. *Computer, 43*(1).

Dogusoy-Taylan, B., & Cagiltay, K. (2014). Cognitive analysis of experts' and novices' concept mapping processes: An eye tracking study. *Computers in human behavior, 36*, 82-93.

Economides, A. (2005). *Adaptive orientation methods in computer adaptive testing.* Paper presented at the E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education.

Economides, A. A. (2006). Adaptive feedback characteristics in CAT. *International Journal of Instructional Technology and Distance Learning, 3*(8), 15-54.

Economides, A. A. (2008). Culture-aware collaborative learning. *Multicultural Education & Technology Journal, 2*(4), 243-267.

Economides, A. A. (2009). Adaptive context-aware pervasive and ubiquitous learning. *International Journal of Technology Enhanced Learning, 1*(3), 169-192.

Economides, A. A., & Moridis, C. N. (2008). *Adaptive self-assessment trying to reduce fear.* Paper presented at the Advances in Computer-Human Interaction, 2008 First International Conference on.

Economides, A. A., & Roupas, C. (2007). Evaluation of computer adaptive testing systems. *International Journal of Web Web-Based Learning and Teaching Technologies, 2*(1).

Eivazi, S., Hafez, A., Fuhl, W., Afkari, H., Kasneci, E., Lehecka, M., & Bednarik, R. (2017). Optimal eye movement strategies: a comparison of neurosurgeons gaze patterns when using a surgical microscope. *Acta Neurochirurgica, 159*(6), 959-966.

Eklund, J., & Brusilovsky, P. (1998). *The value of adaptivity in hypermedia learning environments: A short review of empirical evidence.* Paper presented at the Proceedings of Second Adaptive Hypertext and Hypermedia Workshop at the Ninth ACM International Hypertext Conference Hypertext.

Esfahani, N., Elkhodary, A., & Malek, S. (2013). A learning-based framework for engineering feature-oriented self-adaptive software systems. *IEEE transactions on software engineering, 39*(11), 1467-1493.

Eubanks, T. R., Clements, R. H., Pohl, D., Williams, N., Schaad, D. C., Horgan, S., & Pellegrini, C. (1999). An objective scoring system for laparoscopic cholecystectomy. *Journal of the American College of Surgeons, 189*(6), 566-574.

The Eye Tribe. (2017). Retrieved from http://theeyetribe.com/theeyetribe.com/about/index.html

Faraco, R. A., Rosatelli, M. C., & Gauthier, F. A. (2004). *Adaptivity in a learning companion system.* Paper presented at the Advanced Learning Technologies, 2004. Proceedings. IEEE International Conference on.

Feldman, L. S., Hagarty, S. E., Ghitulescu, G., Stanbridge, D., & Fried, G. M. (2004). Relationship between objective assessment of technical skills and subjective in-training evaluations in surgical residents. *Journal of the American College of Surgeons, 198*(1), 105-110.

Feng, C., Rozenblit, J. W., & Hamilton, A. J. (2007). *A hybrid view in a laparoscopic surgery training system.* Paper presented at the Engineering of Computer-Based Systems, 2007. ECBS'07. 14th Annual IEEE International Conference and Workshops on the.

Fernández, G., Biondi, J., Castro, S., & Agamenonni, O. (2016). Pupil size behavior during online processing of sentences. *Journal of integrative neuroscience, 15*(04), 485-496.

Filieri, A., Ghezzi, C., & Tamburrelli, G. (2012). A formal approach to adaptive software: continuous assurance of non-functional requirements. *Formal Aspects of Computing, 24*(2), 163-186.

Filieri, A., Hoffmann, H., & Maggio, M. (2014). *Automated design of self-adaptive software with control-theoretical formal guarantees.* Paper presented at the Proceedings of the 36th International Conference on Software Engineering.

Flechtner, K.-M., Steinacher, B., Sauer, R., & Mackert, A. (1997). Smooth pursuit eye movements in schizophrenia and affective disorder. *Psychological medicine, 27*(6), 1411-1419.

Fulton, B., & Medlock, M. (2002). *Beyond psychological theory: Getting data that improve games.* Paper presented at the Game Developer's Conference 2002 Proceedings.

Fulton, B., & Romero, R. (2004). *User-testing in a hostile environment: overcoming resistance and Apathy in your game company.* Paper presented at the Game Developer’s Conference, San Jose CA.

Gajos, K., Weisman, L., & Shrobe, H. (2001). *Design principles for resource management systems for intelligent spaces.* Paper presented at the International Workshop on Self-Adaptive Software.

Gallagher, A. G., & Satava, R. (2002). Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. *Surgical endoscopy, 16*(12), 1746-1752.

Gegenfurtner, A., Lehtinen, E., Jarodzka, H., & Säljö, R. (2017). Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis. *Computers & Education*.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review, 23*(4), 523-552.

Gilleade, K. M., & Dix, A. (2004). *Using frustration in the design of adaptive videogames.* Paper presented at the Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology.

Groth, K., Bogdan, C., Lindqvist, S., & Sundblad, Y. (2007). Simple and playful interaction for informal communication and learning. *International Journal of Knowledge and Learning, 3*(2-3), 191-208.

Hallsteinsen, S., Geihs, K., Paspallis, N., Eliassen, F., Horn, G., Lorenzo, J., Papadopoulos, G. A. (2012). A development framework and methodology for self-adapting applications in ubiquitous computing environments. *Journal of Systems and Software, 85*(12), 2840-2859.

Healey, J., Hosn, R., & Maes, S. H. (2002). *Adaptive content for device independent multi-modal browser applications.* Paper presented at the International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems.

Herder, E., & Van Dijk, B. (2002). Personalized adaptation to device characteristics. *AH, 2,* 598-602.

Hermens, F., Flin, R., & Ahmed, I. (2013). Eye movements in surgery: A literature review. *Journal of Eye Movement Research, 6*(4).

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 143*(3611), 1190-1192.

Hoffmann, E. R. (1997). Movement time of right-and left-handers using their preferred and non-preferred hands. *International Journal of Industrial Ergonomics, 19*(1), 49-57.

Hogervorst, M. A., Brouwer, A.-M., & van Erp, J. B. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload.

Huebscher, M. C., & McCann, J. A. (2008). A survey of autonomic computing—degrees, models, and applications. *ACM Computing Surveys (CSUR), 40*(3), 7.

Hussein, M., Han, J., & Colman, A. (2011). *An architecture-based approach to context-aware adaptive software systems*. Retrieved from

Hwang, G.-J., Sung, H.-Y., Hung, C.-M., Huang, I., & Tsai, C.-C. (2012). Development of a personalized educational computer game based on students' learning styles. *Educational Technology Research and Development, 60*(4), 623-638.

Indulska, J., & Sutton, P. (2003). *Location management in pervasive systems.* Paper presented at the Proceedings of the Australasian information security workshop conference on ACSW frontiers 2003-Volume 21.

Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004). *Task-evoked pupillary response to mental workload in human-computer interaction.* Paper presented at the CHI'04 extended abstracts on Human factors in computing systems.

Jameson, A. (2001). Modelling both the context and the user. *Personal and ubiquitous computing, 5*(1), 29-33.

Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in Educational Science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research, 10*(1).

Jiang, X., Atkins, M. S., Tien, G., Bednarik, R., & Zheng, B. (2014). *Pupil responses during discrete goal-directed movements.* Paper presented at the Proceedings of the 32nd annual ACM conference on Human factors in computing systems.

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron, 89*(1), 221-234.

Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science, 4*(1-2), 56-88.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063): Prentice-Hall Englewood Cliffs, NJ.

Kahneman, D., Beatty, J., & Pollack, I. (1967). Perceptual deficit during a mental task. *Science, 157*(3785), 218-219.

Karsai, G., & Sztipanovits, J. (1999). A model-based approach to self-adaptive software. *IEEE Intelligent Systems and Their Applications, 14*(3), 46-53.

Karuppiah, D., Deegan, P., Araujo, E., Yang, Y., Holness, G., Zhu, Z., Riseman, E. (2000). Software mode changes for continuous motion tracking. In *Self-Adaptive Software* (pp. 161-180): Springer.

Kasarskis, P., Stehwien, J., Hickox, J., Aretz, A., & Wickens, C. (2001). *Comparison of expert and novice scan behaviors during VFR flight.* Paper presented at the Proceedings of the 11th International Symposium on Aviation Psychology.

Katerina, G. (2004). *WASA: An intelligent agent for Web-based self-assessment.* Paper presented at the Proceedings of IADIS International Conference Cognition and Exploratory Learning in Digital Agent.

Kelly, D., & Tangney, B. (2006). Adapting to intelligence profile in an adaptive educational system. *Interacting with computers, 18*(3), 385-409.

Kephart, J. O., & Chess, D. M. (2003). The vision of autonomic computing. *Computer, 36*(1), 41-50.

Kerr, A. (2003). Girls/Women Just Want to Have Fun-A Study of Adult Female Players of Digital Games.

Khan, R. S., Tien, G., Atkins, M. S., Zheng, B., Panton, O. N., & Meneghetti, A. T. (2012). Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation? *Surgical endoscopy, 26*(12), 3536-3540.

Klett, F. (2005). *The challenge in learning design concepts: Personalization and adaptation in virtual arrangements.* Paper presented at the Information Technology Based Higher Education and Training, 2005. ITHET 2005. 6th International Conference on.

Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology, 48*(3), 323-332.

Koh, R. Y., Park, T., Wickens, C. D., Ong, L. T., & Chia, S. N. (2011). Differences in attentional strategies by novice and experienced operating theatre scrub nurses. *Journal of Experimental Psychology: Applied*, *17*(3), 233.

Komogortsev, O. V., Gobert, D. V., Jayarathna, S., Koh, D. H., & Gowda, S. M. (2010). Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on Biomedical Engineering, 57*(11), 2635-2645.

Komogortsev, O. V., Jayarathna, S., Koh, D. H., & Gowda, S. M. (2010). *Qualitative and quantitative scoring and evaluation of the eye movement classification algorithms.* Paper presented at the Proceedings of the 2010 Symposium on eye-tracking research & applications.

Komogortsev, O. V., & Karpov, A. (2013). Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. *Behavior research methods*, *45*(1), 203-215.

Kramer, J., & Magee, J. (2007). *Self-managed systems: an architectural challenge.* Paper presented at the 2007 Future of Software Engineering.

Kurzel, F., Slay, J., & Chau, Y. (2002). *Towards an adaptive multimedia learning environment.* Informing Science,

Lababidi, H., Alyousef, S., & Munshi, F. (2015). Low-versus high-fidelity simulations in teaching and assessing clinical skills.

Laddaga, R. (2006). *Self adaptive software problems and projects.* Paper presented at the Software Evolvability, 2006. SE'06. Second International IEEE Workshop on.

Laddaga, R., & Robertson, P. (2004). *Self adaptive software: A position paper.* Paper presented at the SELF-STAR: International Workshop on Self-* Properties in Complex Information Systems.

Lamboudis, D., & Economides, A. (2004). Adaptive exploration of user knowledge in computer based testing. *WSEAS Transactions on Communications, 3*(1), 322-327.

Larsson, L., Nyström, M., & Stridh, M. (2013). Detection of saccades and postsaccadic oscillations in the presence of smooth pursuit. *IEEE Transactions on Biomedical Engineering, 60*(9), 2484-2493.

Law, B., Atkins, M. S., Kirkpatrick, A. E., & Lomax, A. J. (2004). *Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment.* Paper presented at the Proceedings of the 2004 symposium on Eye tracking research & applications.

Lin, C.-B., Young, S. S.-C., Chan, T.-W., & Chen, Y.-H. (2005). Teacher-oriented adaptive Web-based environment for supporting practical teaching models: a case study of "school for all". *Computers & Education, 44*(2), 155-172.

Lonsdale, P., Baber, C., Sharples, M., & Arvanitis, T. N. (2004). A context awareness architecture for facilitating mobile learning. *Learning with mobile devices: Research and development*, 79-85.

Lytras, M. D. (2007). Teaching in the knowledge society: an art of passion. *International Journal of Teaching and Case Studies, 1*(1-2), 1-9.

Macik, M., Cerny, T., & Slavik, P. (2014). Context-sensitive, cross-platform user interface generation. *Journal on Multimodal User Interfaces, 8*(2), 217-229.

Maran, N. J., & Glavin, R. J. (2003). Low- to high- fidelity simulation–a continuum of medical education? *Medical education, 37*(s1), 22-28.

Marshall, S. P. (2002). *The index of cognitive activity: Measuring cognitive workload.* Paper presented at the Human factors and power plants, 2002. proceedings of the 2002 IEEE 7th conference on.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist, 38*(1), 43-52.

McCrum-Gardner, E. (2008). Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery, 46*(1), 38-41.

McKinley, P. K., Sadjadi, S. M., Kasten, E. P., & Cheng, B. H. (2004). Composing adaptive software. *Computer, 37*(7), 56-64.

McNatt, S., & Smith, C. (2001). A computer-based laparoscopic skills assessment device differentiates experienced from novice laparoscopic surgeons. *Surgical endoscopy, 15*(10), 1085-1089.

Menekse, G. G. D., Cagiltay, N. E., Ozcelik, E., & Maras, H. (2017). *Simulation-Based Environments for Surgical Practice.* Paper presented at the

International Conference on Control, Decision and Information Technologies (CoDIT), Barcelona.

Mizouni, R., Matar, M. A., Al Mahmoud, Z., Alzahmi, S., & Salah, A. (2014). A framework for context-aware self-adaptive mobile applications SPL. *Expert Systems with applications, 41*(16), 7549-7564.

Moorthy, K., Munz, Y., Sarker, S. K., & Darzi, A. (2003). Objective assessment of technical skills in surgery. *BMJ: British Medical Journal, 327*(7422), 1032.

Moray, N. (1988). Mental workload since 1979. *International Reviews of Ergonomics, 2*, 123-150.

Moray, N. (2013). *Mental workload: Its theory and measurement* (Vol. 8): Springer Science & Business Media.

Moridis, C. N., & Economides, A. A. (2008a). A computer method for giving adequate feedback to students current mood. *IEEE Technology and Engineering Education (ITEE), 3*(3), 104-107.

Moridis, C. N., & Economides, A. A. (2008b). *Modeling Student's Mood during an Online Self-assessment Test.* Paper presented at the World Summit on Knowledge Society.

Munshi, F., Lababidi, H., & Alyousef, S. (2015). Low-versus high-fidelity simulations in teaching and assessing clinical skills. *Journal of Taibah University Medical Sciences, 10*(1), 12-15.

Murphy, P. R., O'connell, R. G., O'sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human brain mapping, 35*(8), 4140-4154.

Musliner, D. J., Goldman, R. P., Pelican, M. J., & Krebsbach, K. D. (1999). Self-adaptive software for hard real-time environments. *IEEE Intelligent Systems and Their Applications, 14*(4), 23-29.

Nakayama, M., Takahashi, K., & Shimizu, Y. (2002). *The act of task difficulty and eye-movement frequency for the'Oculo-motor indices'.* Paper presented at the Proceedings of the 2002 symposium on Eye tracking research & applications.

Nodine, C. F., & Kundel, H. L. (1987). Using eye movements to study visual search and to improve tumor detection. *Radiographics, 7*(6), 1241-1250.

Nodine, C. F., & Mello-Thoms, C. (2000). The nature of expertise in radiology. *Handbook of Medical Imaging. SPIE.*

North, J. S., Williams, A. M., Hodges, N., Ward, P., & Ericsson, K. A. (2009). Perceiving patterns in dynamic action sequences: Investigating the processes underpinning stimulus recognition and anticipation skill. *Applied Cognitive Psychology*, *23*(6), 878-894.

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior research methods, 42*(1), 188-204.

Oostema, J. A., Abdel, M. P., & Gould, J. C. (2008). Time-efficient laparoscopic skills assessment using an augmented-reality simulator. *Surgical endoscopy, 22*(12), 2621-2624.

Oreizy, P., Gorlick, M. M., Taylor, R. N., Heimhigner, D., Johnson, G., Medvidovic, N., Wolf, A. L. (1999). An architecture-based approach to self-adaptive software. *IEEE Intelligent Systems and Their Applications, 14*(3), 54-62.

Pagulayan, R. J., Steury, K. R., Fulton, B., & Romero, R. L. (2003). Designing for fun: User-testing case studies. In *Funology* (pp. 137-150): Springer.

Pahl, C. (2004). Adaptive development and maintenance of user-centric software systems. *Information and Software Technology, 46*(14), 973-986.

Perrenot, C., Perez, M., Tran, N., Jehl, J.-P., Felblinger, J., Bresler, L., & Hubert, J. (2012). The virtual reality simulator dV-Trainer® is a valid assessment tool for robotic surgical skills. *Surgical endoscopy*, *26*(9), 2587-2593.

Petrelli, D., Not, E., Zancanaro, M., Strapparava, C., & Stock, O. (2001). Modelling and adapting to context. *Personal and ubiquitous computing, 5*(1), 20-24.

Pimentel, J., Lucena, M., Castro, J., Silva, C., Santos, E., & Alencar, F. (2012). Deriving software architectural models from requirements models for adaptive systems: the STREAM-A approach. *Requirements Engineering, 17*(4), 259-281.

Pomplun, M., & Sunkara, S. (2003). *Pupil dilation as an indicator of cognitive workload in human-computer interaction.* Paper presented at the Proceedings of the International Conference on HCI.

Raisinghani, M. S., Benoit, A., Ding, J., Gomez, M., Gupta, K., Gusila, V., Schmedding, O. (2006). Ambient intelligence: Changing forms of human-computer interaction and their social implications. *Journal of digital information, 5*(4).

Rajkowski, J. (1993). Correlations between locus coeruleus (LC) neural activity, pupil diameter and behavior in monkey support a role of LC in attention. *Soc. Neurosc., Abstract, Washington, DC, 1993*.

Reece, S. (2000). Self-adaptive multi-sensor systems. In *Self-Adaptive Software* (pp. 224-241): Springer.

Reiley, C. E., Lin, H. C., Yuh, D. D., & Hager, G. D. (2011). Review of methods for objective surgical skill evaluation. *Surgical endoscopy*, *25*(2), 356-366.

Resnick, R., Taylor, B., & Maudsley, R. (1991). In-training evaluation—it's more than just a form. *Ann R Coll Phys Surg Can, 24,* 415-420.

Richstone, L., Schwartz, M. J., Seideman, C., Cadeddu, J., Marshall, S., & Kavoussi, L. R. (2010). Eye metrics as an objective assessment of surgical skill. *Annals of surgery, 252*(1), 177-182.

Robertson, P., Laddaga, R., & Shrobe, H. (2000). *Introduction: the first international workshop on self-adaptive software.* Paper presented at the International Workshop on Self-Adaptive Software.

Ryan, N., Pascoe, J., & Morse, D. (1999). Enhanced reality fieldwork: the context aware archaeological assistant. *Bar International Series, 750,* 269-274.

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics, 35*(11), 991-1009.

Salehie, M., & Tahvildari, L. (2005). *Autonomic computing: emerging trends and open problems.* Paper presented at the ACM SIGSOFT Software Engineering Notes.

Salehie, M., & Tahvildari, L. (2007). *A weighted voting mechanism for action selection problem in self-adaptive software.* Paper presented at the Self-Adaptive and Self-Organizing Systems, 2007. SASO'07. First International Conference on.

Salehie, M., & Tahvildari, L. (2009). Self-adaptive software: Landscape and research challenges. *ACM transactions on autonomous and adaptive systems (TAAS), 4*(2), 14.

Salvucci, D. D., & Goldberg, J. H. (2000). *Identifying fixations and saccades in eye-tracking protocols.* Paper presented at the Proceedings of the 2000 symposium on Eye tracking research & applications.

Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise differences in attentional strategies related to pilot decision making. *Human Factors, 50*(6), 864-878.

Silvennoinen, M., Mecklin, J.-P., Saariluoma, P., & Antikainen, T. (2009). Expertise and skill in minimally invasive surgery. *Scandinavian Journal of Surgery, 98*(4), 209-213.

Simpson, H. (1969). Effects of a Task- Relevant Response On Pupil Size. *Psychophysiology*, *6*(2), 115-121.

Sinitsky, D. M., Fernando, B., & Berlingieri, P. (2012). Establishing a curriculum for the acquisition of laparoscopic psychomotor skills in the virtual reality environment. *The American Journal of Surgery, 204*(3), 367-376. e361.

Strang, A. J., Best, C., & Funke, G. J. (2014). *Heart Rate Correlates of Mental Workload in a Large-Scale Air-Combat Simulation Training Exercise.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Subramanian, N., & Chung, L. (2001). *Software architecture adaptability: an NFR approach.* Paper presented at the Proceedings of the 4th International Workshop on Principles of Software Evolution.

Sweeney, J. A., Brew, B. J., Keilp, J. G., Sidtis, J. J., & Price, R. W. (1991). Pursuit eye movement dysfunction in HIV-1 seropositive individuals. *Journal of Psychiatry and Neuroscience, 16*(5), 247.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science, 12*(2), 257-285.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction, 4*(4), 295-312.

Sykes, J., & Brown, S. (2003). *Affective gaming: measuring emotion through the gamepad.* Paper presented at the CHI'03 extended abstracts on Human factors in computing systems.

Tamminen, S., Oulasvirta, A., Toiskallio, K., & Kankainen, A. (2004). Understanding mobile contexts. *Personal and ubiquitous computing, 8*(2), 135-143.

Three-D Systems. (2017). Retrieved from https://www.3dsystems.com/haptics-devices/touch

Tien, G., Atkins, M. S., & Zheng, B. (2012). *Measuring gaze overlap on videos between multiple observers.* Paper presented at the Proceedings of the symposium on eye tracking research and applications.

Tien, G., Atkins, M. S., Zheng, B., & Swindells, C. (2010). *Measuring situation awareness of surgeons in laparoscopic training.* Paper presented at the Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications.

Tien, T., Pucher, P. H., Sodergren, M. H., Sriskandarajah, K., Yang, G.-Z., & Darzi, A. (2014). Eye tracking for skills assessment and training: a systematic review. *journal of surgical research, 191*(1), 169-178.

Tien, T., Pucher, P. H., Sodergren, M. H., Sriskandarajah, K., Yang, G.-Z., & Darzi, A. (2015). Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair. *Surgical endoscopy, 29*(2), 405-413.

Tretiakov, A. (2004). *A unified approach to mobile adaptation of educational content.* Paper presented at the Advanced Learning Technologies, 2004. Proceedings. IEEE International Conference on.

Triantafillou, E., Georgiadou, E., & Economides, A. A. (2008). CAT-MD: Computerized adaptive testing on mobile devices. *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT), 3*(1), 13-20.

Tsai, Y.-F., Viirre, E., Strychacz, C., Chase, B., & Jung, T.-P. (2007). Task performance and eye activity: predicting behavior relating to cognitive workload. *Aviation, space, and environmental medicine, 78*(5), B176-B185.

Tseng, J. C., Chu, H.-C., Hwang, G.-J., & Tsai, C.-C. (2008). Development of an adaptive learning system with two sources of personalization information. *Computers & Education, 51*(2), 776-786.

Ucar, E., Ustunel, H., Civelek, T., & Umut, I. (2017). Effects of using a force feedback haptic augmented simulation on the attitudes of the gifted students towards studying chemical bonds in virtual reality environment. *Behaviour & Information Technology, 36*(5), 540-547.

Uemura, M., Jannin, P., Yamashita, M., Tomikawa, M., Akahoshi, T., Obata, S., Hashizume, M. (2016). Procedural surgical skill assessment in laparoscopic training environments. *International journal of computer assisted radiology and surgery, 11*(4), 543-552.

Uemura, M., Tomikawa, M., Kumashiro, R., Miao, T., Souzaki, R., Ieiri, S., Hashizume, M. (2014). Analysis of hand motion differentiates expert and novice surgeons. *journal of surgical research, 188*(1), 8-13.

Uhrich, M., Underwood, R., Standeven, J., Soper, N., & Engsberg, J. (2002). Assessment of fatigue, monitor placement, and surgical experience during simulated laparoscopic surgery. *Surgical endoscopy, 16*(4), 635-639.

van der Lans, R., Wedel, M., & Pieters, R. (2011). Defining eye-fixation sequences across individuals and tasks: the Binocular-Individual Threshold (BIT) algorithm. *Behavior research methods, 43*(1), 239-257.

Varazzani, C., San-Galli, A., Gilardeau, S., & Bouret, S. (2015). Noradrenaline and dopamine neurons in the reward/effort trade-off: a direct electrophysiological comparison in behaving monkeys. *Journal of Neuroscience, 35*(20), 7866-7877.

Vassileva, J. (1998). DCG+ GTE: Dynamic courseware generation with teaching expertise. *Instructional Science, 26*(3), 317-332.

Verney, S. P., Granholm, E., & Marshall, S. P. (2004). Pupillary responses on the visual backward masking task reflect general cognitive ability. *International Journal of Psychophysiology, 52*(1), 23-36.

Viana, W., & Andrade, R. M. (2008). XMobile: A MB-UID environment for semi-automatic generation of adaptive applications for mobile devices. *Journal of Systems and Software, 81*(3), 382-394.

Vickers, J. (1995). Gaze control in basketball foul shooting. *Studies in Visual Information Processing, 6*, 527-541.

Vickers, J. (2003). *Toward defining the role of gaze con-trol in complex targetting skills.* Paper presented at the Visual Search 2: Proceedings Of The 2nd International Conference On Visual Search.

Vickers, J. N., & Williams, A. M. (2007). Performing under pressure: The effects of physiological arousal, cognitive anxiety, and gaze control in biathlon. *Journal of motor behavior, 39*(5), 381-394.

Vine, S. J., Chaytor, R. J., McGrath, J. S., Masters, R. S., & Wilson, M. R. (2013). Gaze training improves the retention and transfer of laparoscopic technical skills in novices. *Surgical endoscopy, 27*(9), 3205-3213.

Vine, S. J., Masters, R. S., McGrath, J. S., Bright, E., & Wilson, M. R. (2012). Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills. *Surgery, 152*(1), 32-40.

Vine, S. J., McGrath, J. S., Bright, E., Dutton, T., Clark, J., & Wilson, M. R. (2014). Assessing visual control during simulated and live operations: gathering evidence for the content validity of simulation using eye movement metrics. *Surgical endoscopy, 28*(6), 1788-1793.

Vogt, S., & Magnussen, S. (2007). Expertise in pictorial perception: eye-movement patterns and visual memory in artists and laymen. *Perception, 36*(1), 91-100.

Walonoski, J., & Heffernan, N. (2006). *Detection and analysis of off-task gaming behavior in intelligent tutoring systems.* Paper presented at the Intelligent Tutoring Systems.

Wang, H.-C., Li, T.-Y., & Chang, C.-Y. (2004). *Adaptive presentation for effective Web-based learning of 3D content.* Paper presented at the Advanced Learning Technologies, 2004. Proceedings. IEEE International Conference on.

Wanzel, K. R., Ward, M., & Reznick, R. K. (2002). Teaching the surgical craft: from selection to certification. *Current problems in surgery, 39*(6), 583-659.

Weber, G., & Brusilovsky, P. (2001). ELM-ART: An adaptive versatile system for Web-based instruction. *International Journal of Artificial Intelligence in Education (IJAIED), 12,* 351-384.

Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences, 109*(22), 8456-8460.

Wilson, M., McGrath, J., Vine, S., Brewer, J., Defriend, D., & Masters, R. (2010). Psychomotor control in a virtual laparoscopic surgery training environment: gaze control parameters differentiate novices from experts. *Surgical endoscopy, 24*(10), 2458-2464.

Wilson, M. R., McGrath, J. S., Vine, S. J., Brewer, J., Defriend, D., & Masters, R. S. (2011). Perceptual impairment and psychomotor control in virtual laparoscopic surgery. *Surgical endoscopy, 25*(7), 2268-2274.

Wilson, M. R., Vine, S. J., Bright, E., Masters, R. S., Defriend, D., & McGrath, J. S. (2011). Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: a randomized, controlled study. *Surgical endoscopy, 25*(12), 3731-3739.

Wobbrock, J. O., Kane, S. K., Gajos, K. Z., Harada, S., & Froehlich, J. (2011). Ability-based design: Concept, principles and examples. *ACM Transactions on Accessible Computing (TACCESS), 3*(3), 9.

Xie, B., & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single-and multi-task environments. *Work & stress, 14*(1), 74-99.

Yarbus, A. L. (1967). Eye movements during perception of complex objects. In *Eye movements and vision* (pp. 171-211): Springer.

Yau, J., & Joy, M. (2007a). *Architecture of a context-aware and adaptive learning schedule for learning Java.* Paper presented at the Advanced Learning Technologies, 2007. ICALT 2007. Seventh IEEE International Conference on.

Yau, J., & Joy, M. (2007b). *A Context-aware and Adaptive Learning Schedule framework for supporting learners' daily routines.* Paper presented at the Systems, 2007. ICONS'07. Second International Conference on.

Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., & Kramer, S. E. (2014). The eye as a window to the listening brain: neural correlates of pupil size as a measure of cognitive listening load. *Neuroimage, 101*, 76-86.

Zhang, J.-Y., Liu, S.-L., Feng, Q.-M., Gao, J.-Q., & Zhang, Q. (2017). Correlative Evaluation of Mental and Physical Workload of Laparoscopic Surgeons

Based on Surface Electromyography and Eye-tracking Signals. *Scientific Reports, 7*(1), 11095.

Zhang, X., Jiang, S., Ordóñez de Pablos, P., Lytras, M. D., & Sun, Y. (2017). How virtual reality affects perceived learning effectiveness: a task–technology fit perspective. *Behaviour & Information Technology, 36*(5), 548-556.

Zheng, B., Cassera, M. A., Martinec, D. V., Spaun, G. O., & Swanström, L. L. (2010). Measuring mental workload during the performance of advanced laparoscopic tasks. *Surgical endoscopy, 24*(1), 45.

Zheng, B., Jiang, X., & Atkins, M. S. (2015a). Detection of Changes in Surgical Difficulty Evidence From Pupil Responses. *Surgical innovation, 22*(6), 629-635.

Zheng, B., Jiang, X., & Atkins, M. S. (2015b). Detection of changes in surgical difficulty: evidence from pupil responses. *Surgical innovation, 22*(6), 629-635.

APPENDIXES

Appendix-A. Algorithm Information

Binocular-Individual Threshold (BIT) Algorithm

The MATLAB source code of BIT algorithm can be downloaded from the webpage (http://www.bm.ust.hk/~mark/staff/rlans.html).

An Adaptive Event Detection (AED) Algorithm

The source code of AED algorithm can be downloaded from authors' webpage (http://www.humlab.lu.se/en/person/MarcusNystrom)

Other Algorithms

The algorithms I-DT, I-HMM, I-KF, I-MST, I-VDT, I-VMT, I-VT and I-VVT can be downloaded from (http://cs.txstate.edu/~ok11/emd_offline.html). However researchers are expected to send an e-mail to the author for explaining their research purpose and asking their permission. The author provides a software password and the algorithms become available for the research purposes.

Appendix-B. Detailed Eye Movement Classification Algorithm Results

Table 8.1 Algorithm Classification Results for Each Eye Movement Event of Scenario-1

ALGORITHM	DOMINANT HAND MEAN RANKS for								NON-DOMINANT HAND MEAN RANKS for								BOTH HAND MEAN RANKS for							
	FN	**FD**	**SN**	**SD**	**SAD**	**PN**	**PD**	**PVD**	**FN**	**FD**	**SN**	**SD**	**SAD**	**PN**	**PD**	**PVD**	**FN**	**FD**	**SN**	**SD**	**SAD**	**PN**	**PD**	**PVD**
AED		1,0		1,0						1,0		1,0						1,0		1,0				
I-DT	4,1	6,2	8,0	8,5	1,3				3,4	6,5	7,4	7,9	2,0				5,8	5,3	7,8	8,7	1,6			
I-HMM	4,5	7,1	4,0	6,1	6,1				5,1	6,7	5,9	6,8	5,2				4,8	7,6	7,8	5,4	4,8			
I-KF	1,8	5,9	4,6	8,3	5,9				2,0	5,5	4,4	8,5	7,0				2,8	6,1	2,6	7,8	6,9			
I-MST	1,8	8,9	2,3	2,3	2,3				1,7	8,9	2,1	2,5	2,5				1,8	9,0	1,5	2,2	5,1			
I-VDT	7,2	4,2	3,5	3,8	5,4	1,3	1,0	1,8	6,9	4,1	3,5	4,0	5,4	1,2	1,0	1,8	6,3	4,1	3,9	4,0	5,3	1,8	1,3	2,0
I-VMT	3,8	2,3	3,1	3,3	6,7	1,7	2,9	1,2	4,3	2,2	2,7	3,2	5,7	1,8	3,0	1,2	2,3	2,3	3,1	3,3	6,3	1,3	3,0	1,1
I-VT	5,2	6,7	5,2	4,9	4,9				5,1	7,2	4,7	4,8	4,5				5,7	7,0	5,3	5,8	3,5			
I-VVT	8,4	2,7	7,2	6,8	3,5	3,0	2,1	3,0	7,9	2,9	6,6	6,5	3,6	3,0	2,0	3,0	6,4	2,9	6,1	6,8	2,5	2,9	1,7	2,9
BIT	8,2	10,0	7,1						8,7	10,0	7,9						9,0	9,6	6,8					
x^2	154,56	192,89	102,42	168,25	103,35	37,08	42,35	39,91	151,43	194,32	112,50	163,00	79,59	39,39	46,00	39,39	134,06	189,89	135,88	163,87	91,98	32,09	37,13	38,35

FN: Fixation Number; FD: Fixation Duration; SN: Saccade Number; SD: Saccade Duration; SAD: Saccade Amplitude Degree; GD: Glissade Duration; PN: Pursuit Number; PD: Pursuit Duration; PVD: Pursuit Velocity Degree

Table 8.2 Algorithm Classification Results for Each Eye Movement Event of Scenario-2

	DOMINANT HAND								NON-DOMINANT HAND								BOTH HAND							
ALGORITHM	**MEAN RANKS for**								**MEAN RANKS for**								**MEAN RANKS for**							
	FN	**FD**	**SN**	**SD**	**SAD**	**PN**	**PD**	**PVD**	**FN**	**FD**	**SN**	**SD**	**SAD**	**PN**	**PD**	**PVD**	**FN**	**FD**	**SN**	**SD**	**SAD**	**PN**	**PD**	**PVD**
AED		1,0		1,0						1,0		1,0						1,0		1,0				
I-DT	5,7	5,1	8,3	8,8	1,4				5,1	4,9	8,7	8,9	1,4				6,7	5,0	8,4	9,0	1,1			
I-HMM	6,8	7,7	3,2	6,7	7,0				6,9	7,8	2,7	6,6	7,3				4,8	7,4	4,2	6,7	5,3			
I-KF	5,1	6,1	4,0	8,2	6,5				5,7	6,1	3,9	8,0	6,6				3,1	6,4	4,2	7,9	6,0			
I-MST	2,8	9,0	1,1	2,0	2,7				3,8	9,0	1,0	2,0	3,1				1,1	9,0	1,1	2,0	4,5			
I-VDT	3,8	3,8	4,3	3,9	5,4	1,3	2,0	1,9	3,1	3,7	4,6	4,0	4,8	1,8	2,0	2,0	5,9	4,0	3,2	3,7	6,3	1,9	1,2	2,0
I-VMT	1,3	2,5	2,7	3,1	6,6	1,8	3,0	1,1	1,4	2,9	3,0	3,0	6,2	1,3	3,0	1,0	2,1	2,2	2,5	3,3	6,9	1,1	3,0	1,0
I-VT	7,8	7,2	6,1	5,0	3,7				7,8	7,2	6,3	5,1	3,8				5,2	7,2	5,8	5,1	3,6			
I-VVT	2,5	2,7	7,3	6,3	2,8	2,9	1,0	3,0	2,2	2,5	7,4	6,3	2,7	2,9	1,0	3,0	7,1	2,8	7,1	6,3	2,3	3,0	1,8	3,0
BIT	9,0	10,0	8,0						9,0	10,0	7,3						9,0	10,0	8,5					
x^2	165,74	202,23	1860,72	181,00	123,05	28,26	46,00	40,78	168,60	201,79	162,38	180,89	119,29	30,02	46,00	42,09	1160,62	202,02	168,89	179,52	110,45	42,00	38,17	46,00

FN: Fixation Number; FD: Fixation Duration; SN: Saccade Number; SD: Saccade Duration; SAD: Saccade Amplitude Degree; GD: Glissade Duration; PN: Pursuit Number; PD: Pursuit Duration; PVD: Pursuit Velocity Degree

Table 8.3 Algorithm Classification Results for Each Eye Movement Event of Scenario-3

ALGORITHM	DOMINANT HAND MEAN RANKS for								NON-DOMINANT HAND MEAN RANKS for								BOTH HAND MEAN RANKS for							
	FN	**FD**	**SN**	**SD**	**SAD**	**PN**	**PD**	**PVD**	**FN**	**FD**	**SN**	**SD**	**SAD**	**PN**	**PD**	**PVD**	**FN**	**FD**	**SN**	**SD**	**SAD**	**PN**	**PD**	**PVD**
AED		1,00		1,00						1,00		1,00						1,00		1,00				
I-DT	7,28	5,13	8,17	8,83	1,22				7,28	4,96	8,87	8,96	1,26				7,00	5,00	8,26	9,00	1,13			
I-HMM	4,63	6,91	4,09	6,85	5,43				4,46	7,33	4,87	6,87	5,67				4,07	7,04	4,98	6,85	5,67			
I-KF	2,89	6,17	3,57	7,80	6,13				3,09	6,57	4,00	7,72	6,22				3,02	6,17	3,15	7,85	7,74			
I-MST	1,39	9,00	1,41	2,17	5,22				1,35	8,98	1,39	2,30	2,85				1,33	8,91	1,37	2,13	2,76			
I-VDT	5,83	3,96	4,43	3,91	5,15	1,91	1,87	1,98	6,30	3,78	3,46	3,93	5,65	2,00	1,91	2,00	6,28	4,00	4,07	3,93	5,07	2,00	1,78	2,00
I-VMT	2,20	2,22	2,46	2,96	6,50	1,15	3,00	1,02	2,33	2,87	2,41	3,02	7,00	1,00	2,91	1,00	2,15	2,13	2,43	3,04	6,67	1,00	3,00	1,00
I-VT	5,48	7,70	5,59	5,00	3,83				5,39	7,13	5,59	4,93	4,33				4,72	7,87	5,98	4,98	4,04			
I-VVT	6,30	2,91	7,17	6,48	2,52	2,93	1,13	3,00	5,85	2,39	7,28	6,26	3,02	3,00	1,17	3,00	7,52	2,87	7,09	6,22	2,91	3,00	1,22	3,00
BIT	9,00	10,00	8,11						8,96	10,00	7,13						8,91	10,00	7,67					
x^2	152,55	200,96	148,23	176,30	90,67	37,62	40,78	45,52	149,38	2199,71	150,76	173,97	105,63	46,00	35,04	46,00	204,01	204,39	147,03	179,21	131,23	46,00	38,17	46,00

FN: Fixation Number; FD: Fixation Duration; SN: Saccade Number; SD: Saccade Duration; SAD: Saccade Amplitude Degree; GD: Glissade Duration; PN: Pursuit Number; PD: Pursuit Duration; PVD: Pursuit Velocity Degree

Table 8.4 Algorithm Classification Results for Each Eye Movement Event of Scenario-4

ALGORITHM	DOMINANT HAND MEAN RANKS for								NON-DOMINANT HAND MEAN RANKS for								BOTH HAND MEAN RANKS for							
	FN	**FD**	**SN**	**SD**	**SAD**	**PN**	**PD**	**PVD**	**FN**	**FD**	**SN**	**SD**	**SAD**	**PN**	**PD**	**PVD**	**FN**	**FD**	**SN**	**SD**	**SAD**	**PN**	**PD**	**PVD**
AED		1,0		1,0						1,0		1,0						1,0		1,0				
I-DT	7,2	5,0	8,2	9,0	1,2				6,2	5,2	7,9	8,7	1,2				6,6	5,0	8,0	8,9	1,2			
I-HMM	5,7	7,2	3,6	6,8	6,9				6,1	7,2	3,5	6,2	6,6				5,8	7,7	3,1	6,2	6,5			
I-KF	4,7	6,0	2,3	8,0	7,8				3,8	5,8	2,7	8,2	7,3				4,0	6,0	2,7	8,1	7,1			
I-MST	2,9	9,0	1,8	2,3	2,6				2,6	9,0	1,3	2,1	2,7				2,8	9,0	1,3	2,1	2,8			
I-VDT	5,1	4,0	4,4	3,8	5,0	1,9	1,8	2,0	4,4	4,1	4,2	3,8	5,5	1,9	1,8	2,0	5,0	4,0	4,4	3,9	5,5	1,8	1,8	2,0
I-VMT	1,0	2,3	3,0	2,9	6,1	1,1	3,0	1,0	1,2	2,2	3,5	3,2	6,0	1,2	3,0	1,0	1,0	2,1	3,5	3,0	6,1	1,2	3,0	1,0
I-VT	6,1	7,9	6,0	5,0	3,7				6,8	7,7	6,0	5,1	3,6				7,0	7,3	6,1	5,0	3,9			
I-VVT	3,6	2,7	7,0	6,2	2,7	3,0	1,2	3,0	4,9	2,8	7,0	6,8	3,0	3,0	1,2	2,9	3,7	3,0	7,2	6,8	2,9	3,0	1,2	3,0
BIT	8,7	9,8	8,7						9,0	10,0	8,8						9,0	10,0	8,7					
x^2	132,25	203,67	160,82	179,70	145,82	43,29	40,51	44,09	140,04	202,29	161,36	178,63	126,57	38,26	39,39	42,09	143,67	205,54	166,14	181,44	119,00	35,63	38,17	44,09

FN: Fixation Number; FD: Fixation Duration; SN: Saccade Number; SD: Saccade Duration; SAD: Saccade Amplitude Degree; GD: Glissade Duration; PN: Pursuit Number; PD: Pursuit Duration; PVD: Pursuit Velocity Degree