博士論文
Gaze Analysis and Visual Saliency Prediction Across Different Age Groups

(年齢の異なるグループに渡る視線の解析と視覚的顕著性の予測)

東京大学
THE UNIVERSITY OF TOKYO

クリシュナ オンカール
Onkar Krishna

Adviser: Kiyoharu Aizawa
Graduate School of Information Science and Technology
The University of Tokyo

This dissertation is submitted for the degree of
Doctor of Philosophy

December, 2017

Acknowledgements

This work has been done during the great time I have spent at Aizawa Yamasaki laboratory, University of Tokyo. I would like express my gratitude to my adviser, Professor Kiyoharu Aizawa for introducing me to this exciting field of applied perception and for his endless support and encouragement.

My thanks also go to my research collaborators Rämä Pia and Andrea Helo for providing me the appropriate resources and fruitful discussions on the topic.

This work would not be the same without the support from all people at Aizawa Yamasaki laboratory, thanks for the great atmosphere, open discussions, fun and friendship. Thanks to all people who read this thesis and have given me their valuable comments to shape it in a better way.

Finally, I would like to thank my parents and my sister for infinite patience and care and all my friends in India who kept in touch and cheered for me from the distance.

This work was financially supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Abstract

Visual attention studies in computer vision research discuss models that have been developed to simulate elements of the scene that are likely to attract human attention. In other words, given an image the model computes a saliency map that encode for conspicuity at every location. These models have wide applicability towards target detection in natural images, advertisement designing, image re-targeting and editing.

The computational models developed so far have considered only adult observers and do not generalize well to the observers of different age groups, i.e. the age-related differences in scene viewing behavior has not been considered while computing saliency.

This dissertation proposes a framework to quantitatively analyze the age-related differences in gaze landings during scene viewing and use the recommendations from this analysis to adopt existing models to the observer's age.

To make the contribution easily comprehensible, the dissertation is divided in three parts:

- **Analysis** As a part of this section we proposed framework to analyze age-impact on two different attributes of fixation; fixation location and duration.
  - **Fixation location based analysis** We developed the measures to quantify four aspects of scene viewing behavior: explorativeness, agreement in explored locations within and between age groups, depth bias, and center bias. Each of these contributes to the detailed understanding of how gaze distribution for scene viewing changes with age.
  - **Fixation duration based analysis**: We check if all the fixations have same role in free viewing? To answer this question the proposed study classified the fixations in ambient and focal for all the age groups based on the fixation duration and saccade amplitude. This is then followed by developing metrics for investigating age-related changes in ambient and focal visual processing mode.

- **Age-adapted Saliency Model** The results of the analysis section are leveraged to develop age-adapted saliency model that can predict salient locations for observers of different age groups. The differences in viewing behavior of the observers in terms of their tendencies to exploit different level of image details, and apparently different foreground-background tendencies are carefully incorporated into the age-adapted saliency model. The proposed age-adapted saliency models are seen to outperform the existing state-of-the-art saliency models in predicting observer's fixations.
- **Application** The prior knowledge about age-related differences in scene viewing is used for a novel application of signboard saliency detection in street videos during free viewing and task viewing. Our proposed signboard saliency detection model can predict relative saliencies of the signboard more accurately than the existing models during free viewing and task viewing scenario. Further we also analyze signboard saliency patterns for two age groups: adults and elderly.

The proposed analysis framework and age-adapted saliency model in this dissertation contribute in understanding the age-impact in scene viewing behavior in context of developing an age-adapted saliency models. Our work is strategically beneficial, as most conventional models of visual attention can be easily tuned to age-related changes in scene viewing by following the recommendations of our analysis result.

Contents

List of Figures

Chapter 1

Introduction

Figure 1.1: First column from left shows the example stimuli used in our study. The second, third, and fourth columns are heat maps representing salient locations attended by children (18 subjects), adults (23 subjects), and elderly (17 subjects), respectively, in an eye-tracking experiment.

1.1 Motivation

This research dates back to May 2015, when we came across a Japanese news article "トイレ貼り付けの洗剤、子どもの誤飲注意" (toilet paste detergent, child mis-detected attention), which reported that 214 children were found to eat toilet paste detergent (a jelly like detergent in the toilet bowl) within a year (see Figure 1.2). This story left us puzzled with many questions, such as "what kind of objects in the vicinity of a children get more attention? Is the same object perceived in a similar fashion by adults? Do children perceive color, intensity, texture, and edges from the surroundings similarly to adults? Does the gaze landing change with age?"

トイレ貼り付けの洗剤、子どもの誤飲注意

ツイートする　シェアする

2014年12月24日 15:44

　トイレの便器の内側に貼り付けて使うタイプの洗剤や芳香剤を、子どもが誤って口に入れる事故が相次いでいることがわかった。日本中毒情報センターによると、１歳前後の子どもがなめたなどの相談が去年１年間で２１４件寄せられ、今年も相次いでいるという。

Figure 1.2: Start of our research inspired by this new article in a Japanese electronic media

All these questions motivated us to study the development of human perception with age. As we reviewed past literature about the human gaze behavior, we came across certain psychological studies that marked an existence of this difference in gaze landings of adult and child observers. This substantiated our hypothesis of age related differences in scene viewing tendencies and led us to the contributions that we describe below as a part of this dissertation.

The human visual system has the ability to view a scene by directing gaze from one location to another in a specific order, as determined by the mechanism of selective attention of the human brain. Over the last few years, researchers in computer vision and cognitive sciences have focused on developing computational attention models that predict conspicuous locations of a scene as perceived by human observers. However, these computational models [13, 28, 70, 22] fail to predict gaze accurately when compared with an actual human gaze as they are not tuned for inter-individual differences in the scene viewings.

In recent years, vision research studies [1, 25] on scene exploration have shown that there exist remarkable differences in the scene-viewing behavior of observers across different age groups. For example, local image features such as color, intensity, luminance, etc, were shown to guide fixation landings early in life, while later, fixation landings are dominated more by top-down processing [1].

In spite of a few studies reporting the developmental changes in the scene viewing behaviors, there have been no studies which systemically analyzed the gaze landing of observers across age groups in the context of developing an age-adapted computational model. So far, computational models have relied on the gaze data collected with adult participants.

Table 1.1: Saliency benchmark dataset

Dataset	Images	Observers	Age range	Duration
MIT300[32]	300	39	18-50	3
FiWI[55]	149	11	21-25	5
NUSEF[47]	758	25	18-35	5
DOVES[65]	101	29	27 (mean)	5
Toronto[24]	120	20	18-22	4

Table 1.1 illustrates that the average age of participants for eye-tracking experiments used by the state-of-the art models that predict the visual saliency. Thus all of visual saliency prediction techniques using these datasets reflect the scene exploration behavior of adult observers only. Due to significant changes in the visual skills during development, it is essential to include the age factors in the computational models.

1.2 Problem Definition

The factors that drive a person's attention can be either related to the scene being observed or to the observer viewing the scene. In a sense, we can also think of fixation as either being 'pulled' to a specific location by the visual properties of the fixated region (i.e. scene-related factors) or 'pushed' to a certain location by cognitive factors (i.e. observer-related factors such as the task to be accomplished and a prior knowledge structure).

The bottom-up features such as color, intensity, and orientation of a scene constitute the scene-related factors. Most of the research in developing computational attention system have been devoted to the scene related factors (i.e. bottom-up image features), which are easier to model compared to the cognitive factors related to observers.

The observer-related factors can be classified as cognitive and physical (see Figure 1.3). The impact of cognitive factors over scene viewing behavior are mainly studied for a given task [59, 24], human tendencies [37], habituation and conditioning [4], and emotions [60], while the impact of physical factors predominately related to the external state of human observers, such as observer's age, eye sight, visual disparity, and gender are least explored.

Physical factors especially the developmental changes in visual perception and eye-movement control [1, 25, 39, 45] has been studied extensively in psychology and neurobiology. But to the best of our knowledge there is no study in computer vision that quantitatively analyzes age-impact on the gaze landings for modeling the age-adapted visual attention systems.

Following voids exist in the present works towards saliency prediction:

- Most of the available eye-gaze data are collected for the adult observers with the mean age of 18-45 years.

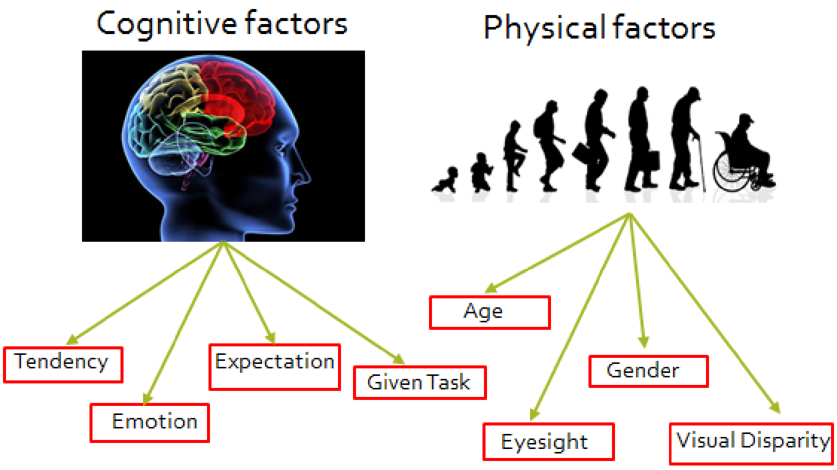


Figure 1.3: Classification of Observer's related factors which influence scene viewing: Cognitive factors and physical factors

- There is no framework for systematic analysis for age-related differences in the scene viewing behaviors.
- Existing computational models predict salient locations pertaining to young adults and ignore the variations that come with age

1.3 Contributions

The contribution of this dissertation is as follows:

- **Analysis framework:** Gaze location and duration are studied to reveal the age-related differences in scene viewing behavior.
  - **Gaze location:** We developed various measures to quantitatively analyze the age-related changes in gaze distribution. We also propose a framework that measures impact of age over the depth bias tendencies (see Figure 1.1), the explorativeness, the agreement score, the intuitive upper performance limits (IUPL), and the center bias tendencies for images belonging to different categories
  - **Gaze duration:** Based on the fixation duration and saccade amplitude, we investigated the undiscovered aspect of the developmental changes in ambient and focal visual processing mode. During scene viewing ambient mode helps in getting the gist of the scene very quickly, whereas the focal mode is responsible for a detailed examination of the salient regions in the scene [3].

- **Age-adapted saliency models:** The analysis results recommended major findings that we use to develop an age-adapted computational model of visual attention. For example, depth object bias tendency showed that the child participants are focused more towards the objects in the foreground compared to the adults and the elderly. This bias can be incorporated into the age-adapted model by including a depth map at different spatial scales for children and elderly age groups

- **Application:** We applied analysis results of age-related differences for a novel application of signboard saliency detection in street videos during free-viewing and task viewing. For this purpose, we first introduce a new eye gaze dataset collected over 30 observers (15 adults and 15 elderly) viewing two street videos in free-viewing and task-viewing mode. Furthermore, we used the quantitative analysis metrics which we developed previously to find the differences in the gaze tendencies for signboards in the street videos during the free-viewing and task-viewing for the adults and the elderly. Finally, video saliency model is developed for detecting relative saliencies of the signboards in free-viewing and task-viewing

The analysis framework that we propose is strategically important as it is easy to incorporate the recommendations into the various existing state-of-the art saliency models to make them age-adaptive. Furthermore, the proposed models not only give state-of-the art saliency prediction results for adult observers but also on observers of other age groups (child and elderly) for whom the existing models yielded less reliable performance.

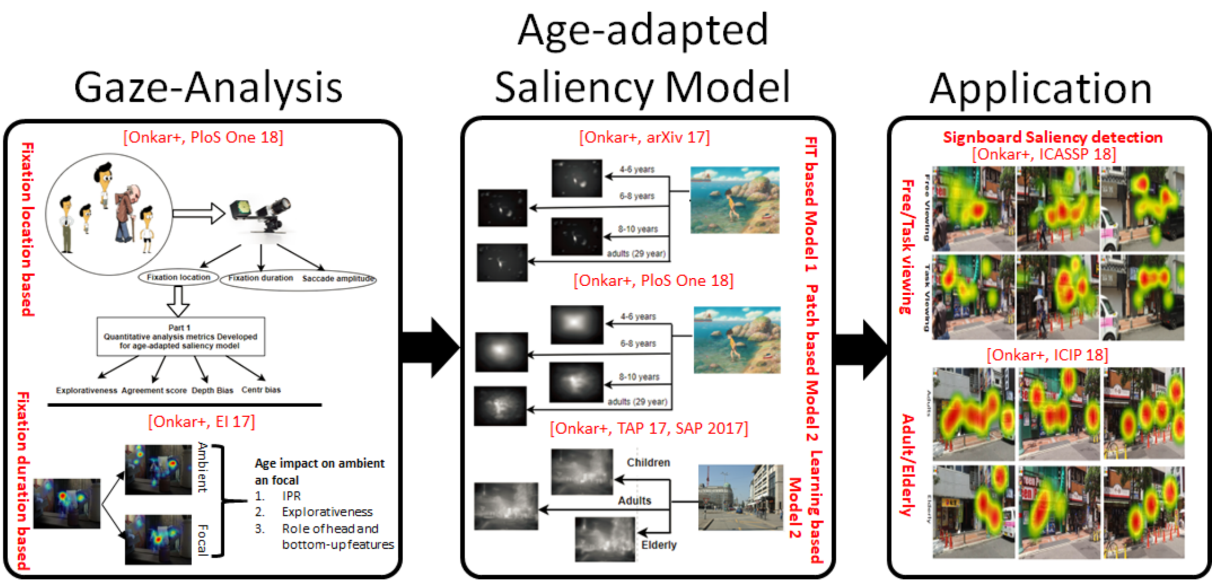


Figure 1.4: Thesis Organization: It consist of three parts, Gaze analysis, Age-adapted saliency model and Application.

1.4 Dissertation Organization

This dissertation is mainly divided into two parts: The first part consists of a quantitative analysis of the age-related differences in fixation landings during scene viewing (Chapter 2 and 3). The second part consists of our proposed age-adapted computational model of saliency prediction based on the analysis reported in the first part (Chapter 4). The thesis organization is shown in Figure 1.4.

The thesis is organized as follows: In Chapter 2, the existing literature is surveyed. Proposed analytical framework is described in Chapters 3 and 4. Chapter 5 consists of three different age-adapted saliency models developed based on the recommendations of analysis results of Chapters 3 and 4. Chapter 6 studies the age-related changes in ambient and focal visual processing. Chapter 7 consists of an application in signboard saliency detection in street videos during free-viewing and task-viewing. I have published these results in [38, 40, 39]

Chapter 2

Related Work

In this chapter we briefly review the existing literature on developmental changes in scene viewing behaviors. We specify the developmental factors missed by these studies that are significantly important to be quantified in order to develop an age-adapted saliency model. We then review prior work on computational modeling of visual attention system for adult observers and discuss the scope of upgrading them to an age-adapted saliency model for complex images belonging to different scene categories.

2.1 Developmental Literature

In this section we present evidences of age-impact on scene viewing behavior from developmental studies establishing the need for quantifying the age-impact on scene viewing to develop an age-adapted saliency model.

There are many studies in psychology and neuroscience investigating the aging effect on eye-movement controls such as focusing on target, maintaining focus on object, moving eyes effectively, right and left eye coordination, and eye-hand coordination [45, 3, 26, 18, 35].

These developmental aspects of eye-movement control have been studied mainly by observing the subject's scene viewing behaviors with the use of artificial stimuli. However, there are only a few studies that have employed natural stimuli to reveal developmental changes in free viewing behavior [1].

Most of the developmental studies have either focused on developmental changes during early stage in life (childhood) or during late stage in life (elderly) [3, 67, 15, 44], and there are only a few studies investigating the age-related changes for the entire life span (children, adult, and elderly) [26, 1].

We mention the major results listed in these studies towards vision development for two types of stimuli; artificial and natural:

2.1.1 Artificial Stimuli Based Study

Development of fixation system: The studies reporting development of fixation system suggest that the ability to fixate on a target is seen since early childhood. But more complex aspect of fixation systems, such as stability and control of fixation, improves with increasing age between 4 to 15 years of age [3, 67].

Development of saccadic control system: The saccade velocity increases during childhood, with peak velocity reached around 10-15 years of age group, and then followed by a decay until 86 years of age [26, 15]. Studies about saccade latency (i.e. reaction time to initiate an eye-movement) have shown that the voluntary eye-movement decreases exponentially from birth to 14-15 years of age [16, 18, 35, 44].

Age-impact on anti-saccade (AS) task: Anti-saccade (AS) task that helps understand cognitive ability to inhibit reflexive saccade was studied for 300 participants of 8-65 years age [16]. The participants with 40-65 years of age demonstrated a moderate deterioration in AS performance. These results are also replicated in several other studies where they concluded that the relationship between age and anti-saccade is curvliner from childhood to adulthood [36, 44]; and linear from adulthood to elderly stage [50, 19, 36]. Table 2.1, lists the vision development facts of these studies for a quick glance for the reader.

2.1.2 Naturalistic Stimuli Based Study

There are only a few studies that have employed the naturalistic stimulus to understand the age-impact on scene viewing behavior [1]. The study in [1] found that bottom-up features of a scene such as color, luminance, contrast, etc. guided viewing during early stage in life (7-9 years), whereas in later stage (more than 72) viewing behavior were dominated by more top-down processing. This result can be helpful in upgrading saliency models developed over guided search theory [66, 7] to make them age-adapted. Specifically the bottom-up and top-down maps need to be tuned according to the observer's age-group, before combining them into a final saliency map. In this study, the total number of fixation landed, inter-fixation distance, and performance in patch matching task have been used to quantify the age-impact on the scene viewing behavior.

The developmental literature presents several evidences of the fact that the age-related differences in natural scene viewing behaviors are significant across children, adult, and elderly participants. Further, these findings motivates us to investigate the age-related differences in scene viewing, with aim of incorporating them in an age-adapted computational model of saliency prediction.

Table 2.1: Details of the developmental studies and there impact on the developmental factors considered in the study

Developmental Studies	**Developmental Factor**	**Target age-group**	**Image-type**	**Age-impact**
Aring et al. 2007	Fixation system	4-15 years	Artificial stimuli	Fixation increases Saccade decreases
Ygge et al. 2005	Fixation stability	4-15 years	Artificial stimuli	Fixation stability increases with age
Fioravanti et al. 1995	Saccadic eye movement	5-13 years	Artificial stimuli	Binocular coordination is significantly different with adults
Irving et al. 2006	Saccadic dynamic	3-86 years	Artificial stimuli	Saccadic latency decreases for 3-14 years and stable to age 50 finally increases up to 80 years
Fischer et al. 1997	Saccadic performance	8-70 years	Artificial stimuli	Age-impacted on prosaccade overlap (PO) and the antisaccade gap (AG)
Klein 2001	Pro and Antisaccade task	6-28 years	Artificial stimuli	Age-impact observed on Antisaccadic task
Luna et al. 2004	Cognitive maturation by Oculomotor tasks	8-30 years	Artificial stimuli	Adult like performance for speed response and working memory at 14, 15, and 19 years respectively
Klein and Feige 2005	Antisaccade task and independent components analysis	7-18 years	Artificial stimuli ECG based study	Pro and antisaccade reaction time become faster in childhood
Nelson et al. 2000	Spatial working memory	8-11 years	Artificial stimuli	Revealed activity in prefrontal coretex
Fukushima et al. 2000	saccadic eye movement,	4-13 years	Artificial stimuli	Antisaccade task showed higher age impact
Açık et al. 2010	Top-down and Bottom-up	7-88 years	Natural stimuli	bottom-up more influential in childhood, top-down influence more elderly
Helo et al. 2014	Fixation duration, saccade amplitude	4-12 years	Stimuli from Children's book	Fixation duration decreases and saccade increases with age

The saliency models developed by the computer vision community mainly exploits the knowledge about fixation location of observers. While the studies reported previously have spent enough discussions over the developmental aspects of gaze attributes like fixation duration, saccade amplitude, blinks, etc., there is no study that has formalized these changes in fixation locations. This is the first gap that we would fill in the present literature in the analysis section of this dissertation.

Next we discuss the state-of-the-art saliency prediction models and where they lack in age-based fixation prediction capabilities.

2.2 Visual Saliency Models

In the last decade, many computer vision researchers have used psychophysical theories and models of attention (such as feature integeration theory (FIT) [63] and Wolfe's Guided Search Model [66]) to create computational models that mimic the adult visual attention system. We briefly review some of these models according to the techniques and/or features they use.

Bottom-up features based models: Itti et al.'s model in [28] based on FIT theory is one of the most well-known models, where bottom-up features of a scene are extracted in parallel by a set of linear center-surrounded operations similar to the visual receptive field. The graph-based visual saliency model in [22] also follows a similar approach to FIT in generating the activation maps of different feature channels at multiple spatial scales. Which are then represented as fully connected graph with the equilibrium distribution in a Markov chain treated as the saliency map.

Combination of bottom-up and top-down features: Torralba (2003)[61] and Torralba (2006)[62] proposed a model using a Bayesian framework that integrates the scene context with a bottom-up saliency map.

Similar to the Bayesian framework, the 'SUN' model of saliency prediction [71] combines bottom-up features represented as self-information with top-down information represented by Difference of Gaussian (DoG) or independent component analysis (ICA) features extracted from images. A boolean map based model [69] was recently developed based on Gestalt psychological studies [37], and outperformed other state-of-the-art models on saliency related datasets.

However, these models do not take into account developmental studies reporting that bottom-up processing dominates during early development, though influences of top-down processing increase with the increasing age [1, 60, 30, 45].

Patch based models: Patch based dissimilarity measures is another line of approach where saliency is estimated in terms of dissimilarity among neighbouring patches. A patch-based saliency estimation method [13] computes the saliency for each patch by measuring the average distance of regional covariance among neighbouring patches. First-order image statistics such as difference of mean value is also incorporated with this algorithm to obtain better results.

Another patch based method [11] was proposed to estimate the saliency of each patches by measuring the spatially-weighted dissimilarity among them, where the image patches were represented in reduced dimensional space by applying principal component analysis (PCA).

These models are not suitable for age-adapted prediction of salient locations, as the optimal patch size is selected for the highest prediction accuracy over the eye tracking data collected for adult participants only.

Models based on supervised learning on eye tracking datasets: Supervised learning-based models using eye-tracking data collected from adults constitute another technique to build computational models. Authors of [33] proposed a model that simply learns to predict saliency from an eye-tracking dataset containing over 1003 images viewed by 15 adults.

These models have all considered eye data from adults and do not generalize well to observers of all age groups. For example the eye-tracking datasets used for the most of learning methods were collected for adult participants age range 18-50 years. It is important that for a generalized saliency prediction model, the recommendations from past literature and the new recommendations for age impact on fixation locations should be incorporated in these models.

We see how we can analyze impact of age on gaze behavior and upgrade the existing models to become age adapted in the following chapters

Chapter 3

Analysis I: Study of General Gaze Behavior Across Age Groups

We begin detailed discussion of our findings on age-related differences in human gaze behavior. As a part of this chapter, we shed light on these differences for children and adults. For the same we first introduce the eye-gaze data used for this study developed from images taken from children's book and movies. Following this we then describe the quantitative metrics developed to reveal the age-related differences in scene viewing.

The measures developed in this study quantify three aspects of viewing behavior: (1) explorativeness, (2) agreement in explored locations within and between age groups, and (3) center bias. Each of these measures contribute to the detailed understanding of how gaze distribution changes for scene viewing with age.

Explorativeness highlights the spread of fixation locations. Agreement within and between age groups was estimated by using the agreement score which reflects how well observers within same age group or of different age groups agree in terms of explored locations. The last parameter, center bias reveals the age-related differences in the tendency of an observer to look at the center of stimuli.

3.1 Eye-tracking Dataset

We analyzed the eye-tracking dataset collected by our collaborators in [25]. The study was conducted in conformity with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and approved by the ethics committee of the University of Paris Descartes. All young participants or parents in case of children gave written informed consent prior to participation.

3.1.1 Subjects and Stimuli

The eye-tracking data was obtained for 82 observers from different age groups. All observers had normal or corrected-to-normal vision. Participants were assigned to 4 different groups: four-six years, six-eight years, eight-ten years, and adults (mean age, 29 years). We use 4 years, 6 years, 8 years, and adults to refer these groups in order.

The age group assignment was made based on the findings of previous developmental studies suggesting that eye-movement control changes rapidly at the beginning i.e. during childhood, and slowely later [3, 18, 10, 67, 26, 35, 45]. These studies motivated us to investigate the children age groups in relatively smaller intervals to quantify the significant changes in scene viewing behavior and subsequently reflecting these changes in the age-adapted saliency model.

The experiment was conducted on images of 1024×764 pixels. The images were taken from children's books and movies, and characterized to have eventful background. The choice of images might be less interesting for adults than children. However, we decided to use these images since they were suitable for children and they can also be used in eye-gaze study in adults. Previous studies have used paintings [52] and artificial stimulus [49], [34] to revel the eye movement behavior in adults.

Further, to avoid stimulus related bias and to maintain the motivation of our participants, a segment recognition test was performed during the experiment. In which after the presentation of the image (10s) an image segment was presented at the center of the screen during 5 seconds; in 50% of the cases the image segment was valid. Participants had to determine if the segment was part of the previous scene or not by pressing a button. The results of the task performance reported in [25] suggested the high level of engagement for the selected stimuli for all age groups including adult observers, which also confirms the age appropriateness of our selected stimuli.

3.1.2 Apparatus and Procedure

The remote eye-tracking system EyeLink 1000 with a sampling rate of 500 Hz was used to measure eye gaze, and provided us with the raw data that was sampled to obtain fixations and saccades. The spatial resolution of eye tracker was below $0.01°$, and spatial accuracy more than $0.5°$. The random fixations and noise were discarded by processing the raw data by fixation detection algorithm supplied by EyeLink.

During eye tracking experiment pictures were presented at a distance of 60 cm from a screen at a resolution of 1024×728, a five point calibration and validation was performed before starting the viewing task and subjects were asked to explore the scene which was presented for 10 seconds.

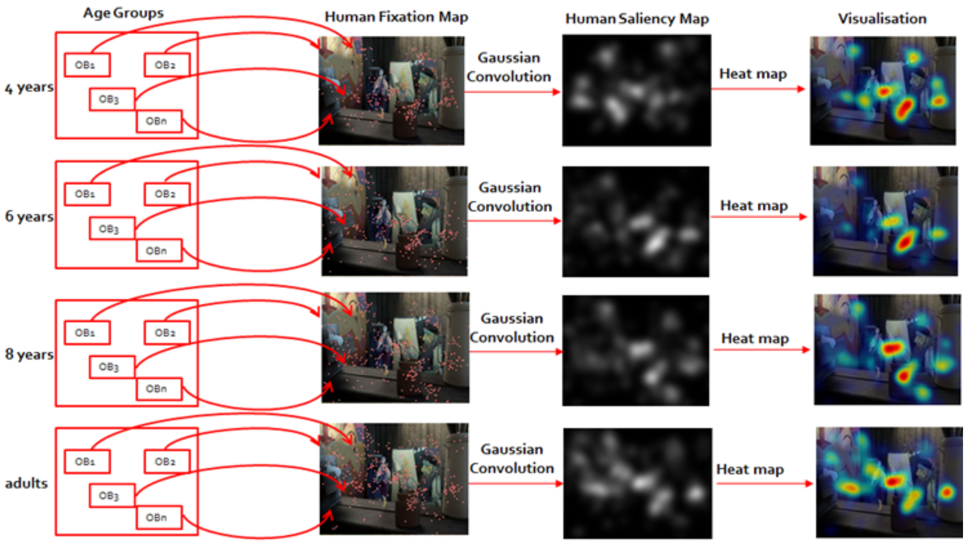


Figure 3.1: The process of generating human fixation map and human saliency maps of an image for different age groups. OB_n stands for the n^{th} observer of an age group. The images used in this figure are similar but not identical to the original image, and is therefore for illustrative purposes only.

3.1.3 Data Representation

For each image fixation landings of all observers were used to generate two maps for different age groups: a human fixation map and a human saliency map. The human fixation map was created as a binary representation of fixation locations, and the human saliency map was obtained by convolving a Gaussian filter across the fixation locations, as in [33]. The visualizations of human fixation and human saliency maps are shown in Figure 3.1. These maps were used to analyze eye-movement behavior.

3.2 Analysis Framework

To quantify the age-related differences in scene-viewing of observers we selected fixation landing locations as a main attribute to analyze. The reason for this selection relies on the fact that existing saliency models consider fixation location as a key gaze attribute in predicting salient regions.

Based on fixation locations, we developed measures to quantify three aspects of scene viewing behavior: (1) explorativeness, (2) agreement in explored locations within and between age groups, and (3) center bias. The selection of these aspects was based on previous studies in adults showing that the accuracy of saliency prediction in computational models improves when fixation spread, agreement between observers, and center bias tendency are included [33, 71, 46].

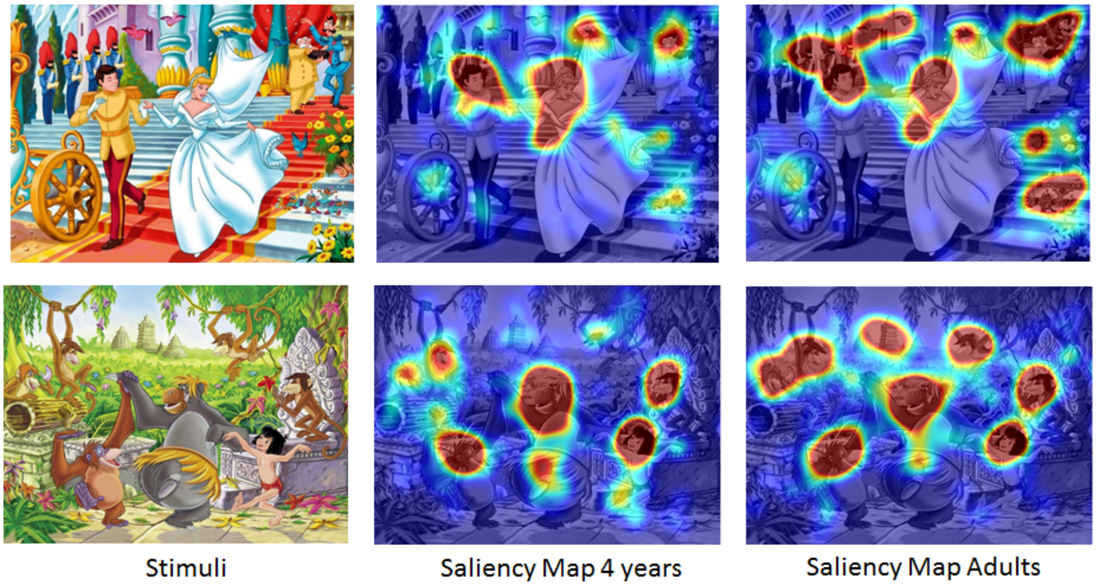


Figure 3.2: Visualizing the differences in explorativeness tendency between 4 years and adults

3.2.1 Explorativeness

To evaluate eye movement behavior during scene exploration across age groups, we conducted an explorativeness analysis. To quantify the explorativeness of observers in a group we calculated first-order entropy of the human saliency map. The selection of human saliency map for explorativeness study is based on the fact that we observed a human saliency map differs between age groups.

The difference in explorativeness can be visualized in Figure 3.2, generated by overlaying the human saliency map over the original image. For the i^{th} image of group g the explorativeness is computed as following:

$$H(U_i^g) = \sum_l h_{U_i^g}(l) * log(L \,/\, h_{U_i^g}(l)) \tag{3.1}$$

Where U_i^g is the human saliency map of the i^{th} image from all observers in a group g for which entropy is calculated and $h_{U_i^g}(l)$ is the histogram entry of intensity value l in image U_i^g, and L is the total number of pixels in U_i^g.

In the context of viewing behavior, a higher entropy corresponds to a more exploratory viewing behavior by an observer, as their saliency points are more scattered in the given scene. Conversely, a lower entropy corresponds to less exploratory behavior. The average behavior of each age group over all images was analyzed based on the average entropy.

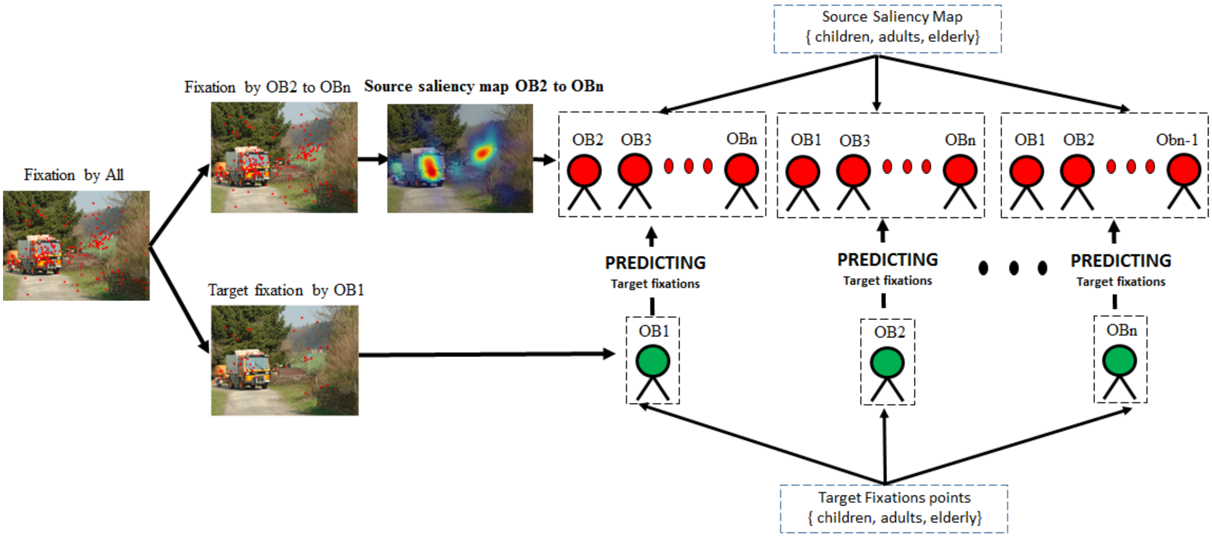


Figure 3.3: Framework to measure inter and intra age group agreement score

3.2.2 Agreement Analysis

Explorativeness falls short of checking for similarity of explored regions within age groups and between age groups. For instance explorativeness score is unable to answer questions such as: Do observers belonging to the same age group explore the same spatial regions of the image? And is there any agreement among observers in terms of explored regions across age groups? It should be noted that poor agreement of fixation landings between adults and children leads to imprecise prediction when using saliency models that are originally developed for adults. This motivated us to conduct an agreement analysis.

The area under the curve (AUC) is the most commonly used metric in the literature for discrete ground truth saliency maps, and we choose it for our analysis. The AUC-based measure analyzed how well the human saliency map of fixation points of all observers of an age group could be used to find the pooled fixation locations of all observers from the group, as well as observers from different groups.

The age group of which the saliency map was used became the source group, and the group for which the fixation locations were being used as target group (see Figure 3.3). Thus, under the intra-age group agreement analysis, the source and target belonged to the same group, and for inter-age group analysis, the source age group was different from the target group.

For this analysis, the human saliency map of the source group was first thresholded to T levels covering different percentages of the most salient areas of the image. To evaluate how well these thresholded maps agreed with the fixation points of the observers in the target group, we then made use of the AUC metrics. This required us to lay down a general formulation of inter-age group metrics - True positive rate (TPR) and false positive rate (FPR) for observers from the source group g_s to find fixation points for observers from the target group g_t.

The intra-age group metric is a special case of inter-age group metrics, when $g_s = g_t$, i.e., within same group. Thus,

$$TPR_{U_n}^{g_s g_t}(I_i) = \frac{TP_{U_n}^{g_s g_t}(I_i)}{TP_{U_n}^{g_s g_t}(I_i) + FN_{U_n}^{g_s g_t}(I_i)} \tag{3.2}$$

$$FPR_{U_n}^{g_s g_t}(I_i) = \frac{FP_{U_n}^{g_s g_t}(I_i)}{TP_{U_n}^{g_s g_t}(I_i) + FN_{U_n}^{g_s g_t}(I_i)} \tag{3.3}$$

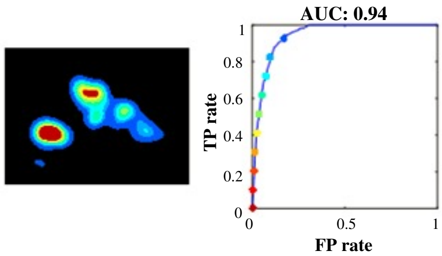
Where the TPR for the i^{th} image is the extent to which the fixation points of observers in group g_t agree to the n^{th} thresholded saliency map U_n of observers from source group g_s. Similarly, FPR deals with non-fixation points that have been considered fixation points. The TPR and FPR for all T-thresholded saliency maps of an image were combined into a vector of T dimension. The area under the ROC curve plotted between TPR and FPR gave us the AUC-score, and an average of these scores across all stimuli of the dataset provided the agreement score of the group.

For a given image, agreement score tells us how accurately the fixation locations of all observers in the group were covered under the differently thresholded saliency maps of observers from the same or different age groups. We can visualize the intra-age and inter-age group agreement analysis in Figure 3.4.

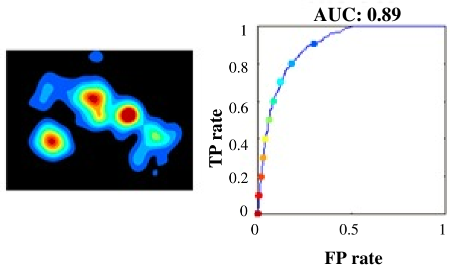
3.2.3 Center Bias

The term "center bias" has been studied using the eye-tracking techniques, and it reflects the human tendency of looking at the center of a given image [58]. Several studies have established the existence of the center bias, but only a few scholars have considered the center bias in their computational models [33, 20].

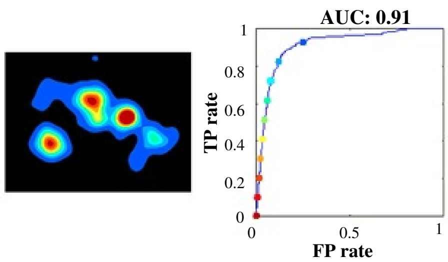
The center bias greatly influences our viewing behavior but to the best of our knowledge no study has investigated the age-related differences toward center bias. In order to reveal differences in center bias across age groups we first computed the center map by taking average of all the saliency maps across age group. Further, the center bias for each age group is measured by measuring the euclidean distance between the centroid of the center-map and the center pixel of the image. We have also used the center map as a saliency map to predict fixations of the same age group. The average prediction performance of the center map for the fixations of the same age group were measured by AUC matrics developed in agreement analysis section.



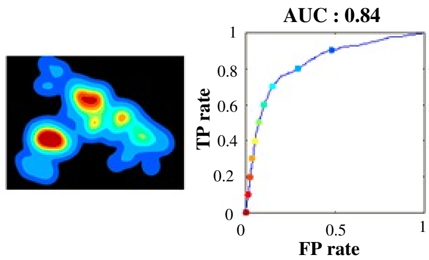
(a) Intra-age group: target fixation points of 4 years by source saliency map of 4 years.



(b) Intra-age group: target fixation points of adults by source saliency map of adults.



(c) Inter-age group: target fixation points of 4 years by source saliency map of adults.



(d) Inter-age group: target fixation points of adults by source saliency map of 4 years.

Figure 3.4: Agreement analysis visualization: The heat map visualizes agreement behavior in predicting the target fixation points by source saliency map and the ROC calculates the quantitative value of the agreement score.

We have also investigated the role of contrast bias by following the study reported in [10]. The result suggested that the image contrast plays an important role in gaze landings but the statistical analysis results showed no significant difference between the age groups. Thus we have only reported age-related changes center bias tendency in this study.

3.3 Results

3.3.1 Explorativeness

When participants of 4 years and adults age groups observed the same set of images of our dataset, the set of least explored scenes were found to be different among observers belonging to different age groups (see Figure 3.5). Thus, exploratory behavior depends on the observer‘s age.

The results of the explorativeness suggested that:

1. Explorativeness increased monotonically with age, $r(29) = 0.99$, $p < 0.001$ (Spear-

(a) (b) (c) (d)

Figure 3.5: Least explored images and their saliency maps: (a,b) 4 year age group, (c,d) adult observers, the top 9 least explored images are ordered from the top left to bottom right.

man correlation). This is illustrated in Figure 3.6b, which plots the entropy of all images for each age group. The histograms of entropy of all images for different age groups are illustrated in Figure 3.6a.

2. One-way ANOVA analysis showed that explorativeness varied significantly among the age groups, $F(3, 29) = 15.8$, $p < 0.001$. Bonferroni correction based Post-hoc test indicated that explorativeness scores of 4 and 6 years age groups were significantly lower than the scores of 8 years and adult groups, all $p < 0.01$. However, no difference was found between 8 years and adults age groups, suggesting that from the age of eight years explorativeness behavior is adult-like.

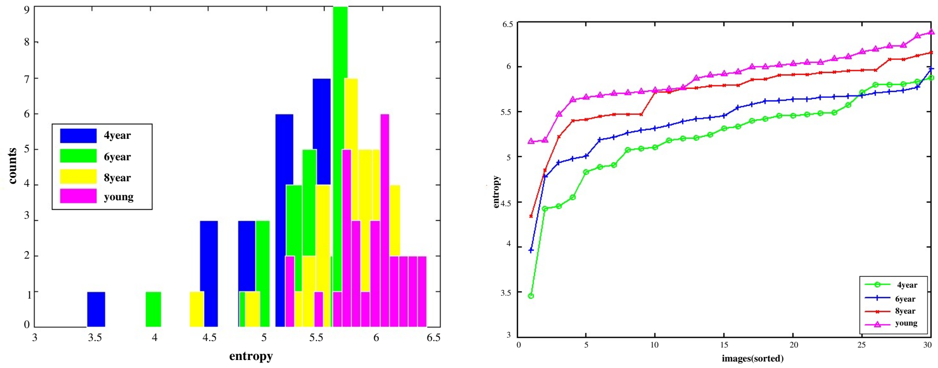
Previously, it has been shown that the spread of the fixation landing i.e. entropy of the saliency map [31] decreases when the image resolution decreases i.e. when there is change in level of details, observers tend to respond by changing in the gaze pattern. Thus, our findings suggest that during scene exploration children older than 8 years of age and adults tended to direct their gazes at different level of details in a given scene. On the contrary, being less explorative, children tended to direct their gazes towards fewer details of the scene.

3.3.2 Agreement Analysis

The main results from the intra-age and inter-age group agreement analysis are as follows:

Intra-age group agreement analysis:

1. There is a negative correlation between intra-age group agreement and observers' age revealing a high intra-age group agreement between children, $r(29) = -0.88$, $p < 0.001$ (see Figure 3.7a).



(a) (b)

Figure 3.6: Explorativeness results: (a) The histogram of entropy indicates that there is a shift from left to right for 4 year to adult age group. (b) Entropy plotted in sorted order for different age groups over all the stimuli

Table 3.1: Agreement score: Average agreement score of human saliency map of observers from the source group in predicting fixation points of target age group.

Target Source	4 years	6 years	8 years	adult
4 year	0.9148	0.8756	0.8695	0.8683
6 year	0.8463	0.9003	0.8509	0.8493
8 year	0.8150	0.8269	0.8870	0.8343
adult	0.8122	0.8265	0.8340	0.8910

2. One-way ANOVA test suggested that the age impacted on agreement score, $F(3, 29) = 65.8$, $p < 0.01$, As shown in Figure 3.7b, Bonferroni correction based Post-hoc showed that the average agreement score of the 4 years age group was highest and significantly different from all other age groups ($p < 0.001$). The score started to decrease as observer's age increased up to 8 years age by showing 8 years old and adults had significantly less intra-age group agreement than 6 years olds, $p < 0.01$.

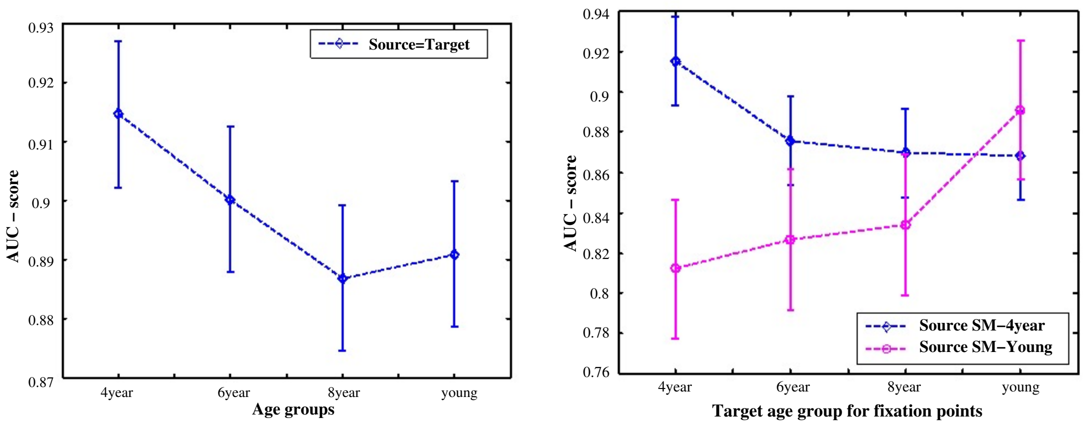
Similar to the explorativeness results, the agreement score suggested that scene-viewing tendency matures at the age of eight. This can be understood by the fact that 8 years and adults were the most explorative, and there salient regions may not be consistent with one another at higher level of the details.

Inter-age group agreement analysis:

1. Table 3.1 shows that the agreement scores of inter-age group analysis was lower than those of intra-age group analysis for all ages. Thus, it was even more evident that the age has an impact on visual behavior as the same age group maps predicted the fixations more precisely.

2. The most important observation of the inter-age analysis was that the agreement score of adults predicting all others (4, 6, 8 years) was significantly less than the agreement scores of diagonal colored boxes of the Table 3.1 (prediction by same age groups). Spearman's correlation based post-hoc analysis indicated significant differences in performance of adults predicting 4 year, 6 year, and 8 years than the prediction by the same age-group, $p < 0.01$.

Thus, ignoring the age factor and using conventional models developed and learned over adults can not give optimal performance for other age groups. This calls for the modification of existing models to make them adapt to age. Figure 3.7b shows the comparison of agreement score for saliency maps of 4 years and adults in finding the target fixations of different age groups.



(a) (b)

Figure 3.7: Agreement analysis between observers within age group and across age groups: (a) Intra-age group agreement scores, which reflects that kids agree more in explored locations than younger adults. (b) Agreement analysis results for the source saliency map of 4 years and adults in finding the target fixation of different age groups.

3.3.3 Center Bias

As shown in Figure 3.8, age-related differences in bias towards the center map across age groups suggested that the 4 years age group had the highest bias among all age groups. It decreased with increasing age, where adult-like behavior was exhibited at 8 years of the age. The results of One way ANOVA analysis indicated the significant age-impact on the center-map bias, $F(3, 29) = 8.15$, $p < 0.03$.

Further, post-hoc analysis indicated that both adults and 8 year were significantly different from 4 years and 6 years age groups, $p < 0.01$, similarly, 4 and 6 years also shown significant different with each other $p < 0.03$. The highest euclidean distance for adults suggested the lowest center bias in adults among the age groups (168, 182, 181, and 226 are the euclidean distance in pixels for children, adult and elderly participants)

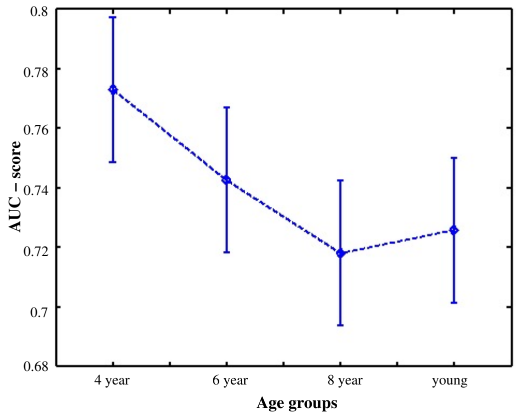


Figure 3.8: Age-based changes in center bias tendency across age groups

3.4 Discussion and Conclusion

This chapter focused towards analyzing the age-impact on three basic aspects of gaze distribution behavior: explorativeness, agreement within and between age-groups, and center bias tendency.

Explorativeness results: A significant impact of age was observed on explorativeness of observers belonging to four different age groups (4, 6, 8 years and adults). Our results showed that explorativeness increased monotonically with age, and adult-like behavior was achieved at the age of 8 years. These results support previous studies comparing explorative performance in young and old adults suggesting that explorativeness change with age [23, 41]. A new evidence about how explorativeness changes during childhood was added to the available facts .

Agreement analysis results: Agreement score results within the age groups showed that observers of 8 years and adult age group had the lowest agreement with each other. This is in accordance to previous study in [46] where the fixation density map of an adult subject was used to predict the fixation density map of other subjects. The result showed a lower inter-similarity score suggesting the lower agreement between adult observers.

Comparing the results of explorativeness and agreement analysis, it is interesting to note that the trend followed by the intra-group agreement analysis was opposite to that exhibited by the explorativeness analysis. This makes sense: as explorativeness decreases, observers tends to focus on lesser details of the scene [31], mostly the ones that were the salient areas of an image. This suggests that the fixation points of the observers of the least explorative age group would mostly be consistent with one another, and would be mostly localized at salient locations and, hence, the agreement score would be high.

Center bias results: All the age groups showed significant effect of center bias tendency in scene viewing. However, children of 4 years of age had the highest center bias among different age groups, whereas the adults exhibited lowest center bias tendency for the same set of images. In previous studies center bias tendency has been reported only in adult observers [33, 46], but in our study we provide novel evidence that the strength of the center bias tendency varies across different age groups.

One possible explanation is that the center bias is driven by the content of the scene i.e. bottom-up saliency as computed by different saliency models [70, 47, 28, 33, 63]. The higher center bias in children observers can be inferred from the previous studies in [1] which found that bottom-up saliency maps dominated more to the fixation landings in children than adult observers, suggesting higher center bias in a children than adults.

Based on the analysis results of our study, we concluded that ignoring the age factor and using existing models developed for adults in predicting gaze behavior of other age groups cannot give the optimal prediction accuracy.

As a part of the next chapter we show how these parameters show similar patterns for different sets of images taken from “man-made”, “nature”, and “fractals” datasets. We further discuss how observers for different age groups differ in tendencies to fixate towards objects at different depths. Having discussed all these parameters and their generalizability, we then discuss the framework of how our analysis results of this chapter and the one that follows can help in making the existing saliency models adapt to observers age.

Chapter 4

Analysis II: Depth Bias in Children, Adults and Elderly

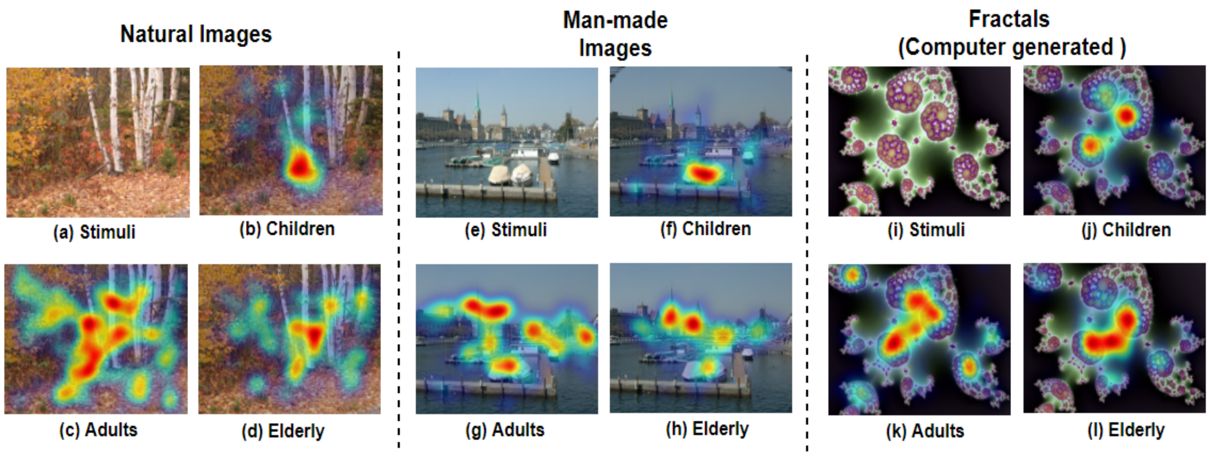


Figure 4.1: Sample stimulus from natural, man-made, and fractals categories and their heat map generated from the recorded eye-gaze data of children, adults and elderly participants.

4.1 Introduction

Till now we have been discussing the age-related differences in scene viewing behavior for child and adult observers. As a part of this chapter, we extend our analysis and check the trends for elderly observers too. For the same, the eye tracking dataset developed by Açiket al. [1] was used which consisted of eye-gaze recording of children, adult and elderly observers while they viewed images of three different categories: “man-made”, “nature”, and “fractals”. This will not only allow us to get insight of the scene viewing tendency of elderly observers, but also help to check if the age-based trends for the parameters that we introduced in the previous chapter (explorativeness, agreement score and center bias) are generalized to any image category.

Further, the fixation landings for man-made dataset inspired us to not only carry out the study for explorativeness, agreement score and center bias but also study of viewing tendencies towards different depths in an image (i.e. foreground and background) for different age groups. Depth bias, and intuitive upper performance limits (IUPL) are the new aspects analyzed for this dataset.

Finally we conclude the chapter by laying down the age-adaptive framework over the results on the parameters discussed in the previous chapter and introduced as a part of this chapter. This framework can be included in the existing saliency models to make them age adapted.

4.2 Dataset

4.2.1 Participants, Stimuli, and Apparatus

The dataset used in this study was collected by Açiket al. [1] and the data can be retrieved from (see Figure 4.1 for sample stimulus). Fifty-eight participants participated in this study, comprising children (18 participants, age-range 7 to 9 years, mean age 7.6), adults (23 participants, age-range 19 to 27 years, mean age 22.1), and the elderly (17 participants, age-range 72 to 88 years, mean age 80.6) groups.

All participatns reported normal or corrected-to-normal vision, including the elderly participants. Acik et al.[1] declared that all participants or their parents signed a written consent to participate in the experiment. The experiment was conducted in compliance with the Declaration of Helsinki as well as national and institutional guidelines for experiments with human participants.

In this study, we used 192 color images belonging to 3 different categories; "naturals", "man-made", and "fractals" (64 in each categories). "Naturals" image category is representing natural scenes having trees, flowers, and bushes, this image category did not contain any artificial objects. "Man-made" category includes urban scenes such as street, road, building and construction sites, and the "fractals" are the computer generated shapes taken from the different web database such as Elena's Fractal Gallery. Stimulus randomizations were balanced across pairs of participants. All images had a resolution of 1280×960. EyeLink 1000 was used to record the gaze in its remote and hand-free mode.

4.2.2 Eye Movement Recording

Eye gaze was recorded while the observers viewed the images displayed for 5 seconds. The scene was subsequently replaced by a circular patch and participants had to determine if the patch was part of the previous scene or not. The patch recognition task was included to maintain the motivation of participants; thus the recognition results were not used in this study. Target stickers were placed on each observer's head to compensate for head movements. Observers viewed the stimuli from a distance of 65 cm on a 20-inch LCD monitor display (width: 40 cm). All observers were instructed to explore the scene. Fixation and saccade were identified via a fixation detection algorithm supplied by EyeLink.

4.2.3 Data Representation

Similar to the eye-gaze data in the previous study (chapter 3), we generated human fixation maps, human saliency maps and heat maps from the gaze data of this study. For each age group the human fixation map for an image stimulus was generated by combining the fixation landings of all observers of that age group. Further, the human saliency maps were generated from human fixation maps by convolving a Gaussian, similar to Velichkovsky et al. [52]. The heat map was obtained from the human saliency maps to visualize the age differences in the region of interest.

4.3 Method

In this section, we discuss two new measures developed for quantitative analysis of age-related changes in scene-viewing behavior of images belonging to "naturals", "man-made", and "fractals" categories. These are the depth bias tendency, and intuitive upper performance limits (IUPL). Human saliency maps across age groups served as a basis for all the metrics developed for this purpose. We also propose a framework that measures age-impact over the depth bias tendency, and Intuitive Upper Performance Limits (IUPL) for images belonging to the different categories.

Figure 4.2: Depth bias analysis framework

4.3.1 Depth Bias

The human visual system has the tendency of focusing earlier on the objects placed in the foreground than the objects in the background of the scene [29]. Gautier and Le Meur [20] investigated the influence of disparity on saliency for 3D conditions. Their results indicate that foreground features play an important role in attention deployment. Itti and Koch [27] suggested that stereo disparity could be used as additional cue in saliency detection. However, this tendency of the human visual system was mainly explored for three-dimensional media but there is no study investigating the role of depth bias for two-dimensional media. Most importantly, to the best of our knowledge, for the first time we reported the role of scene depth in gaze landings across different age groups.

In the presented study, we visualized the age-related differences in gaze landings to the most salient foreground and the background regions in the scene. However, it is important to develop a framework to quantify this tendency across age groups. This tendency was observed in all those images in which some interesting item was placed in the foreground and the background of the scene, there was 20 such images in man-made category.

We propose a novel framework to quantify the depth bias tendency across age groups (Figure 4.2). As a first step a combined saliency map for each image is generated by linearly integrating the human saliency map of children, adults, and the elderly groups. The combined saliency map is generated to find the most salient locations independent of an observer's age group. In the second step, we threshold the combined saliency map to get t_1 and t_2 percentage of the most salient locations of the scene. We fix t_1 and t_2 to five and ten, respectively, as we discovered that a small percentage of the most salient locations is sufficient to represent most of the fixation landings in the scene. In the third step, the most salient locations of the thresholded maps are manually labelled to foreground and background based on the near and far places in the scene. As shown in Figure 4.2, the thresholded map gives the most salient regions of the image (four different regions). Further, these regions are labelled as F or B according to their respective locations in the foreground or background of the scene (as shown in the ML map in Figure 4.2).

Finally, the percentage of the fixations landing on the most salient foreground and background locations is calculated for each image to obtain an average score for individual age groups from the two thresholded maps. In summary, to measure the depth bias across age groups, we labelled most salient locations as foreground and background and then calculated average of the percentage of fixations landed on the foregrounds and backgrounds across all the images.

4.3.2 Intuitive Upper Performance Limit (IUPL)

The final goal of our study is to reflect the age-related changes reported in the eye gaze-data analysis into the proposed age-adapted model of saliency prediction. Keeping that in mind, we have used the metrics reported in [57] to estimate the intuitive upper performance limit of any age-adapted saliency model developed for children, adults, and elderly. The UPL score also gives us an idea that how well a saliency model can predict fixations for images belonging to "natural", "man-made", and "fractals" image categories. The UPL metric was developed to answer the following question: Can a model outperform a computational model that fully replicate the human visual system? Making it more clear, in this section we presented a robust metric UPL to measure the intuitive upper performance limits of a computational model of visual attention developed for these age groups.

The intuitive upper performance limit of an age group is defined as the ability of fixations from half of the observers to predict fixations of other half of the observers. In order to do that first we generate human saliency map over fixations of first half of the observers by following the similar procedure mentioned in data representation section. Further, the human saliency map obtained after this step is used to predict the fixation points of rest of the observers. The area under the curve (AUC) based metrics are then used to give the measure of consistency between the fixations of two halves. The framework to measure intuitive upper performance limits (IUPL) is shown in Figure 4.3.

Higher UPL score of an age group suggests higher agreement among fixations of a group of observers with rest of the observers. Similarly the lower UPL score of an age group, suggests the lower agreement between two group of observers.

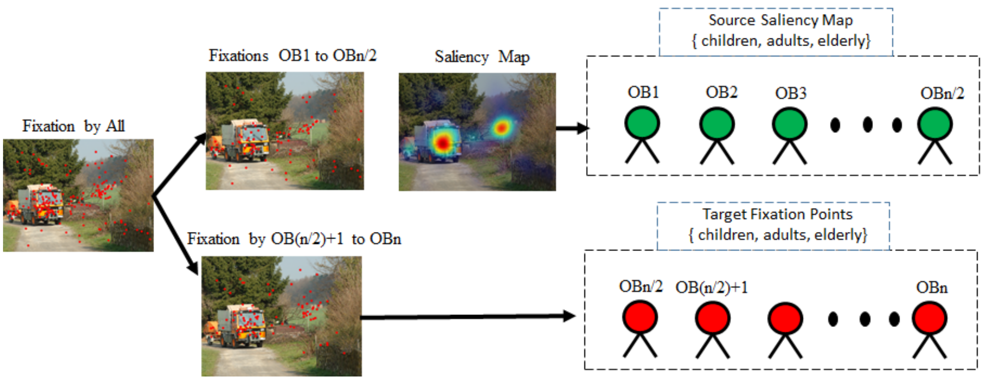


Figure 4.3: Framework to measure intuitive upper performance limit

4.4 Results and Discussion

4.4.1 Depth Bias

The results are shown in Figure 4.4, which show that:

- Average percentage of fixations landed on the most salient foreground locations in the images is significantly higher for child observers than for adults and the elderly (Figure 4.4a). The one-way ANOVA analysis suggest a significant influence of age over gaze landings in the foreground regions of the scene ($F(2, 19) = 2.99, p < 0.05$), Further, a post-hoc analysis confirms that the foreground bias tendency in children is significantly different from that of adult and elderly observers ($p < 0.05$).

- The percentage of fixation landings on most salient background places (Figure 4.4b) gives an interesting observation for the gaze behavior of elderly observers. It was found that elderly observers have a significantly higher tendency for fixating over items placed in the scene background than children and adult observers. The one-way ANOVA analysis confirmed the age-impact over the background bias ($F(2, 19) = 3.56, p < 0.03$), and post-hoc analysis revealed the behavior of elderly observers is significantly different to adults ($p < 0.03$) and children ($p < 0.03$).

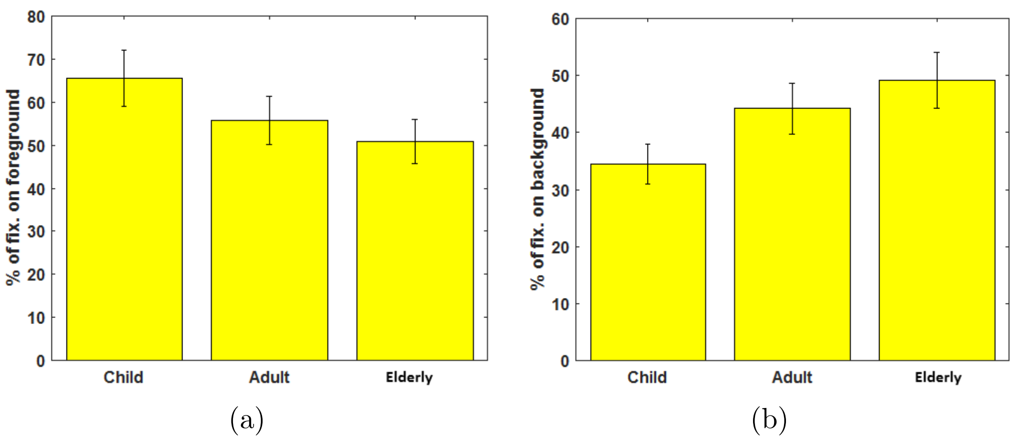


Figure 4.4: (a) Percentage of the gaze landed on foreground. (b) Percentage of the gaze landed on background

4.4.2 Explorativeness

The explorativeness score is measured by the metrics developed in Chapter 3. The explorativeness result suggests following tendency for children, adults, and elderly observers:

- Result showed significant impact of age over explorativeness tendency across all the image categories ($F(2, 191) = 179.45, p < 0.001$). Post-hoc results across all the image categories showed that the child observers possess the lowest exploratory tendency compared to adults ($p < 0.001$), who are highly explorative for the same set of images. Elderly observers showed less exploratory tendency than adults but higher than that of children ($p < 0.001$) (see Figure 4.5).

- Explorativeness tendency reflects the level of details an observer visualizes in a scene. With the highest exploratory behavior, adults have a tendency to visualize finer to coarser details in the scene, as inferred from [31], which states that a decrease in spatial scale results in a decrease in the exploratory behavior of the observers.

The age-related differences in explorativeness tendency on three different class of images show that this exploratory behavior is independent of the type of stimuli being observed.

4.4.3 Agreement Analysis Results: Inter-Individual Similarity

The agreement score on the new dataset is measured by the metrics developed in Chapter 3. The only difference is that contribution of individual observer's source saliency map in predicting target fixations of rest of the observers are reported for this dataset.

The inter and intra-individual similarity results are shown in Table 4.1 and Figure 4.6.

- One-way ANOVA confirmed the age influence over intra-age-group agreement between observers independent of the class of the image ($F(2, 192) = 41.39$, $p < 0.001$). Further, post-hoc revealed that the children are in better agreement with each other for explored locations than adults and elderly observers for all the image categories, $p < 0.001$ (Figure 4.6a and Table 4.1).

- Inter-age-group agreement analysis suggests that fixations of a specific age group can be best predicted by an observer of the same age group. As highlighted in Table 4.1, the diagonal entries in each image class are highest compared to other elements in the same row, which suggest a fixation is well predicted by an observer of the same age-group than the observer of different age-group. For instance, as shown in Figure 4.6c for natural images, a child can predict well the fixations of other children better than an adult can predict children observers ($F(2, 63) = 115.95, p < 0.001$).

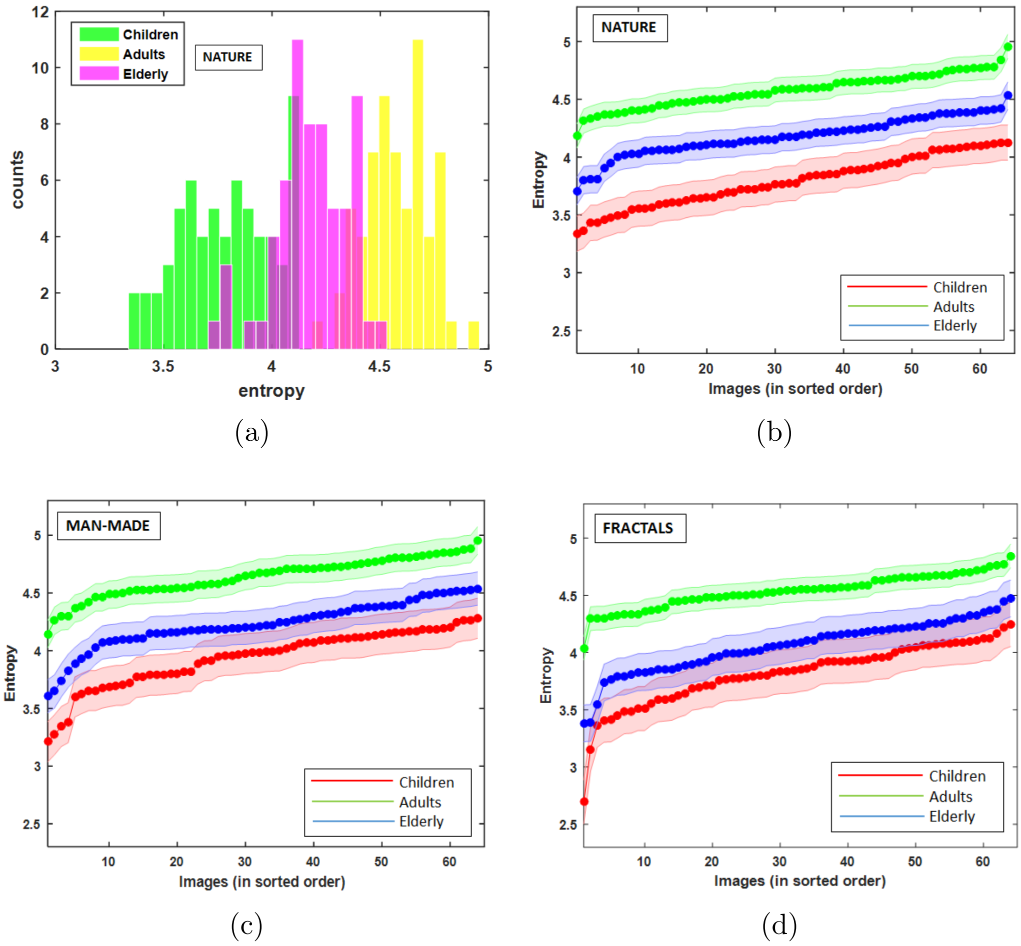


Figure 4.5: The age-impact over explorativeness tendency: (a) Histogram representation entropy values; showing a shift of explorativeness from left to right with increasing age (b,c,d) The figures show that age-impact over explorativeness tendency is independent of the image-type

4.4.4 Intuitive Upper Performance Limit (IUPL)

The Intuitive Upper Performance Limit results are shown in Figure 4.7a.

- One-way ANOVA confirmed the age influence over IUPL for all the image categories, for the man-made category $F(2, 63) = 14.93, p < 0.001$. Further, post-hoc revealed that the IUPL of children were highest and significantly different to adults and elderly participants for man-made images ($p < 0.001$).

- Similarly for the natural and man-made categories the IUPL was significantly higher for children and elderly participants than the adults ($p < 0.001$). The IUPL results presents a very important finding that the prediction accuracy of a computational model developed for children will always be higher than a model developed for adults in all the image categories.

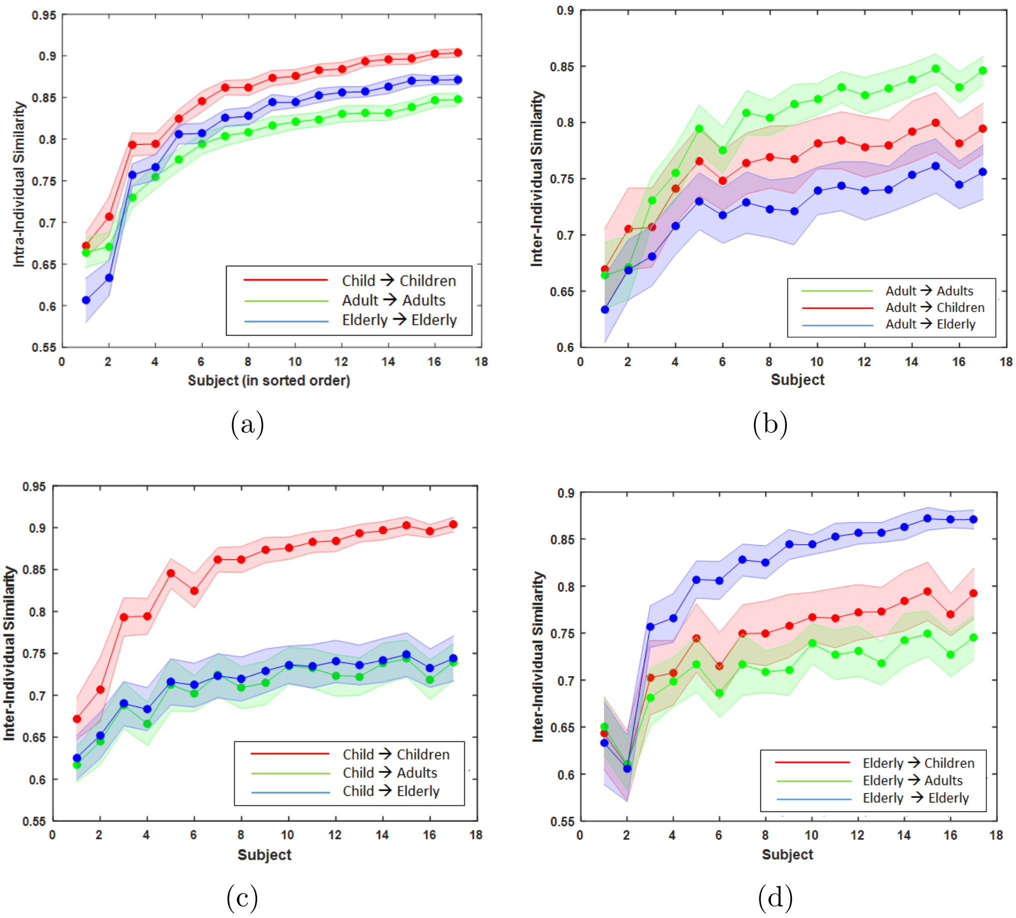


Figure 4.6: Inter-individual similarity: (a) Similarity score of individual participants predicting others from the same age group. (b,c,d) Similarity score of individual participants predicting others from different age group.

The prior knowledge about the upper-bound of the proposed age-adapted saliency model will help in optimizing the prediction accuracy for each age groups.

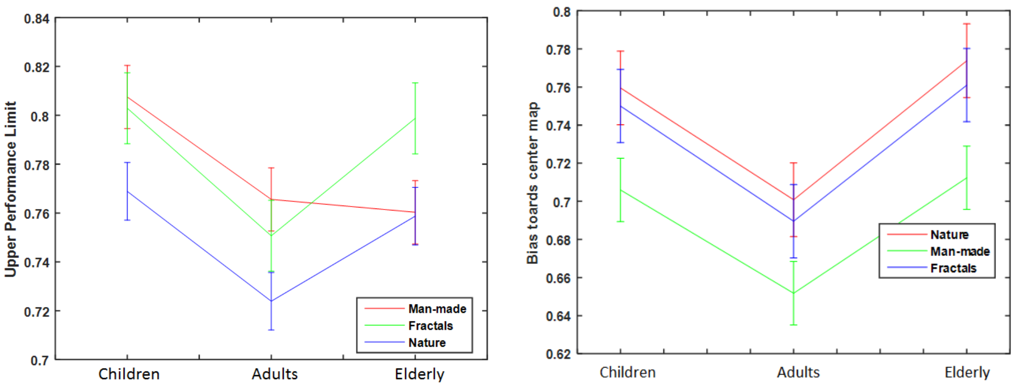
4.4.5 Center Bias

We can visualize different tendency of gaze landings towards center of scene for different age groups as shown in center map of different age groups (Figure 4.8). Agreement score (AUC score) of fixations of different age groups towards their center map revealed the age-impact over this tendency independent of the image categories $F(2, 63) = 34.10, p < 0.001$ (see Figure 4.7b). The post-doc analysis revealed that children have significantly different tendency towards center map that adults for all the scene categories ($p < 0.001$).

Further, the result of Euclidean distance from the centriod of the explored region by different age group and center pixel of the scene showed that children have highest center bias but elderly participants showed similar tendency. (Nature- 150, 210, and 184 are the

Table 4.1: Average of the AUC score of children, adults and elderly in predicting observers of same and different age groups for different image types

Class	Nature			Man-made			Fractals		
Age	Child	Adult	Elder	Child	Adult	Elder	Child	Adult	Elder
Child	0.8390	0.6937	0.7253	0.8483	0.7093	0.7172	0.8579	0.7257	0.7559
Adult	0.7548	0.7874	0.7193	0.7693	0.8085	0.7310	0.7870	0.8037	0.7507
Elder	0.7432	0.6860	0.8063	0.7410	0.7094	0.8093	0.7693	0.7173	0.8326



(a) (b)

Figure 4.7: (a) The Intuitive Upper Performance Limit (IUPL) of of different age groups for natural, man-made and fractals image categories. (b) Center bias score across age groups for all the scene categories.

euclidean distance in pixels for children, adult and elderly participants, man-made- 203, 267, and 244, and fractals- 189, 214, and 202).

4.5 Recommendations for Age-adapted Saliency Model

The analysis results recommended five major findings that we use to develop an age-adapted computational model of visual attention. These recommendations are as follows:

- Explorativeness result showed that children were least explorative among the three age groups i.e. they are mostly focused on the coarser details of the scene, whereas being highly explorative adults evaluated finer detail too. This can be included in age-adapted model by selecting different spatial scale for different age groups.

- Agreement analysis suggests that the saliency maps of adults fail to appropriately predict fixation landings of child and old observers. Thus it is advisable to train the learning based models over the age-specific gaze data to optimize the prediction performance for different age groups.

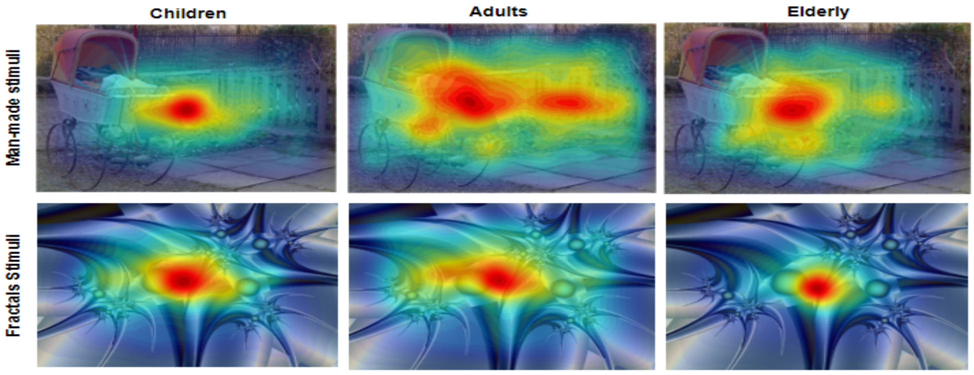


Figure 4.8: The heat map of the center-map overlay on a sample image for different age groups

- Result of age-related changes in central bias tendency motivated us to investigate the scope of reflecting these change in computational model of visual attention by tuning the center feature map of existing algorithms.
- Depth object bias tendency showed that child observers are focused more towards the objects in foreground compared to adult and old observers. This can be incorporated in age-adapted model by including a depth map at different spatial scales for children and old age groups.
- The results of Intuitive Upper Performance Limit gave us clue about the prediction accuracy of any saliency model developed for different age groups.

Chapter 5

Age-adapted Saliency Models

5.1 Introduction

Chapter 3 and 4 presented a detailed analysis of two different type of eye-gaze dataset in context of developing an age-adapted saliency model. Chapter 3 quantified the age-impact on explorativeness, agreement score, and center bias tendencies for children and adult observers. Chapter 4 introduced depth bias and intuitive upper performance limits for an eye gaze dataset collected from children, adult and elderly. Further, based on the recommendations received from the analysis results in Chapter 3 and 4, we develop age-adapted saliency models in this chapter.

Several computational models for visual saliency have been developed in past to provide important insights into the underlying mechanisms of the human visual attention system. All existing models predict regions of interest in images considering the gaze behavior of adults. The evidences from Chapter 3 and 4 suggest that the gaze behavior of the observers vary with age. This motivates us to use the framework discussed in the chapter 4 to optimize the existing saliency models towards giving better prediction performance for observers of all age groups.

In the proposed work, we optmize three different models of visual saliency prediction, the Itti's model [28], the patch based model [11] and learning based model with depth bias [33]. We chose these models carefully in light of the fact that they had different modeling architectures. Age-adapted model based on Itti and Patch model is developed on the recommendations received for children and adults eye-gaze data reported in Chapter 3, whereas the learning based model is developed by incorporating the depth bias tendency in children and elderly observers as reported in Chapter 4. We verified that the proposed age-adapted framework was generalizable, and could be applied to different type of existing model as, most of them follow the same basic structure with minor variations.

5.2 Itti's Based Age-adapted Model

Most existing bottom-up models follow the basic multi-scale feature selection architecture proposed by Itti et al [28]. In these models, we observe the following basic structure: (a) Basic visual features such as color, intensity, and orientation, are extracted over multiple scales of the image, where each scale represents a different level of detail in the scene. (b) All features are investigated in parallel, to obtain the conspicuity map for each feature channel. (c) These features are integrated to obtain the saliency map.

However these models are built over the saliency results from adult observers and ignore the variations that come with age. To make these models age-adapted we re-iterate the three basic steps that were laid as analysis results in Chapter 4:

- The observers from different age groups tend to explore different levels of details which calls for choosing appropriate set of feature scales for different age groups.
- One major difference between Itti's feature integration and our's is that in Itti's model all feature scales are integrated linearly while we choose weighted integration of feature maps. These weights are optimized separately for different age groups.
- Observers of different age have different center bias tendency which needs to be incorporated in the existing models.

We now discuss each of these steps in detail.

5.2.1 Multi-scale Feature Subset Selection: Proposed 1

We used the multi-scale feature extraction technique proposed in the famous Itti et al.'s saliency model [28]. The different scales represented the different levels of detail in scenes, from finer details to coarser object-level details. As stated earlier for more explorative observers, all levels of details were important whereas for the less explorative observers, only coarser-level details were important.

Observers from different age groups showed different levels of explorativeness. Thus, to make our model adapt to age-related differences in scene viewing behavior, we focused on a feature scale selection mechanism, where we identified the subsets of the feature maps that best represented the different levels of details viewed by the observers of different age groups.

We now discuss the steps to extract features for our age-adapted saliency model. For an input image, eight spatial scales were first developed using a Gaussian pyramid. The features were then extracted using the "center-surround" operations with the same settings as in [28] to yield six intensity maps $\mathcal{I}_i$, 12 color maps - six for $\mathcal{RG}_i$ and six for $\mathcal{BY}_i$

Table 5.1: Average prediction accuracy (AUC-score) by multi-scale feature subset selection: where as a∼b means from scale a to scale b are selected (Proposed 1). Scale 6 is the coarsest level and scale 1 is the finest

Age \ Scale	1∼6	2∼6	3∼6	4∼6	5∼6	6∼6
4 year	0.7074	0.7180	0.7288	0.7280	**0.7381**	0.7366
6 year	0.6839	0.6940	0.6978	**0.7183**	0.7044	0.7071
8 year	0.6640	0.6655	**0.6722**	0.6664	0.6572	0.6566
adult	**0.6628**	0.6573	0.6541	0.6512	0.6470	0.6495

each and 24 orientation maps - $\mathcal{O}_i(\theta)$ i.e., sets of six maps computed for four orientation $\theta \in \{0, 45, 90, 135\}$. The 6 maps for different feature represents different level of detail in scene.

The feature maps were then combined into three "conspicuous maps", $\bar{\mathcal{I}}$ for intensity, $\bar{\mathcal{C}}$ for color, and $\bar{\mathcal{O}}$ for orientation. However, as stated above, unlike Itti et al.'s model, this point-wise combination was not conducted over all six maps; we also chose subsets of six maps for each age group. The point wise combination of feature map was:

$$\bar{\mathcal{I}} = \bigoplus_{i=s}^{6} \mathcal{N}(\mathcal{I}_i) \tag{5.1}$$

$$\bar{\mathcal{C}} = \bigoplus_{i=s}^{6} [\mathcal{N}(\mathcal{RG}_i) + \mathcal{N}(\mathcal{BY}_i)] \tag{5.2}$$

$$\bar{\mathcal{O}} = \sum_{\theta \in \{0,45,90,135\}} \bigoplus_{i=s}^{6} \mathcal{N}(\mathcal{O}_i(\theta)) \tag{5.3}$$

where $\mathcal{N}$ represents the normalization and s is the starting index from where maps were taken.

We developed six cases by varying s to 1,2,3,4,5, and 6. If $s = 1$, the subset of feature scale starting from scale 1 (finer) to scale 6 (coarser) had to be combined. Similarly if $s = 6$, only the feature scale 6 was used. Without using the trend toward explorativeness found in the analysis section, we evaluated the model over all such possible subsets for all groups, and defined the subset for each age that best represented the gaze levels (finer to coarser) of the observers in a given age group.

As shown in prediction results in Table 5.1, children age groups (4 years, 6 years, and 8 years) are performing better than adults in the existing settings which makes use of all scales (1∼6). However, the prediction accuracy of children age groups are not optimized on the existing scale (1 ∼ 6). The predictive performance of children get optimized if used coarser scales and ignore finer ones (as in Table 3, scale $5 \sim 6$, $4 \sim 6$, and $3 \sim 6$ are optimized scale selection for 4 year, 6 year, and 8 year age groups respectively) while for adults, prediction accuracy was highest if we chose all scales (similar to [28]).

It is interesting to note that this result is consistent with our analysis result, i.e., children are less explorative than adults and, hence, require only coarser scales to predict their fixations. Age related differences in center bias tendency was also incorporated in this model by including a differently weighted center-map as explained in age-adapted model for center bias section.

5.2.2 Feature Combination Optimization: Proposed 1.1

In this section we proposed another modification in existing models based on our agreement analysis recommendation reported in Chapter 5. The choice of linear integration of feature maps used in previous section was ill-suited because different features contribute differently to the final saliency map. Some state-of-the art models addressed [33] this by learning the optimal weights of feature integration in a supervised manner. These optimal weights are, however, not suitable for our age-adapted mechanism, as they are learned only over eye-tracking data collected for adults. To fit this into our scenario, we learn these optimal weights over features extracted from age-specific subsets of the dataset.

We divided the dataset into a training set with 20 images and a test set with the remaining images. Color, intensity, and orientation features were extracted for the training images. We then selected P strongly positive and negative samples, each corresponding to the top and least-rated salient locations of the human saliency map of all observers generated from ground truth eye-tracking data.

Our agreement analysis result suggests that intra-age group fixation point prediction was better than inter-age group performance. In other words, the fixation points of the observers were better predicted by the saliency maps of observers of the same group rather than those of observers of other groups. Thus, the P positive and negative samples to be chosen were age group specific, i.e., the positive and negative samples for all age groups were differently chosen for training.

We fixed value P to 10; choosing more samples only involved adding redundancy and yielded no performance improvement. For a given set of features and labels (positive and negative samples) for an age group, liblinear SVM was used to learn the model parameters to predict salient locations on the training images. Thus, we obtained model parameters for predefined features over all age groups.

For a given test image, we first collected its features as described in the multi-scale feature selection mechanism, and further predicted saliency values at each pixels as,

$$S_g(I_i) = w_g X^T(I_i) + b_g \tag{5.4}$$

Table 5.2: Average prediction accuracy by combining scale based subset selection, non-linear integration and age-adapted center bias (Proposed 1.1).

Age \ Scale	1~6	2~6	3~6	4~6	5~6	6
4 year	0.7387	0.7409	0.7374	0.7414	**0.7434**	0.7388
6 year	0.7203	0.7218	0.7155	0.7201	**0.7295**	0.7160
8 year	0.6766	**0.6833**	0.6698	0.6590	0.6655	0.6728
adult	**0.6646**	0.6637	0.6613	0.6549	0.6533	0.6585

Where w_g and b_g are model parameters learned for each age group g and $X(I_i)$ is feature vector for the i^{th} test image, this vector is composed of intensity ($\bar{\mathcal{I}}$), color ($\bar{\mathcal{C}}$), and orientation ($\bar{\mathcal{O}}$) features. Based on the saliency values we classified the local pixel as salient or not.

Integrating the feature maps over the optimally set weights learned over the age-specific dataset suggests further improvement in prediction accuracy for all age groups including adults, as shown in Table 5.2. Age-related differences in center bias tendency were also considered while evaluating the performance of the proposed model. The method of incorporating center bias in the age-adapted model is explained in following section. The improvement in prediction performance for our Proposed 1 and Proposed 1.1 model is shown in Figure 5.1.

5.2.3 Age-adapted Model for Center Bias

Humans have the tendency to observe at the center of a given scene. This behavior can be incorporated with existing saliency models by simply defining saliency to include weight factor C, which is inversely propositional to the distance to the center of the pixel under consideration.

$$C(i) = 1 - d(c, p_i)/D \tag{5.5}$$

Where $d(c, p_i)$ is the distance between the pixel under consideration p_i and center pixel c and D is the maximum distance used as a normalization factor. Further center bias $C(i)$ is updated based on the results of analysis reflect the age-related variations. $w_k C(i)$ is the updated center bias weight factor, where w_k is the strength of the center bias tendency for different age groups.

5.3 Age-adapted Patch-based Saliency Model: Proposed 2

Another approach that we choose to verify the generalizability of our age-adapted framework is the patch-based model for saliency prediction [11]. This model is also tested for the eye-gaze data reported in Chapter 3. This technique follows the given basic structure: (a) Image is first divided into patches of the same size. (b) The set of features are extracted from these patches. (c) Finally, the spatial dissimilarity among neighbouring patches is evaluated to generate the saliency map.

To make these patch models age-adapted, we introduce some modifications. For this, we represent different features extracted from a patch by using the subset of eigenvalues obtained after SVD decomposition of the feature matrix. We elaborate this before explaining how to render this newly constructed model age-adapted.

5.3.1 SVD Decomposition Based Representation of Features

We first construct the feature matrix. The first step in feature matrix construction is to extract non-overlapping patches of size $t \times t$ from a given image I of size $M \times N$. Thus, the total number of patches $n_p = M \times N/t \times t$. Further, each patch is represented by a column vector of features f_i, where i indexes the patch. f_i is obtained by combining three color of features (L^*, a^*, b^*) and two intesity features (I_x, I_y). This generates a feature vector for each patch that appears as $[L_1, L_2, ..., L_t, a_1, a_2, ..., a_t,\ b_1, b_2, ..., b_t, I_{x_1}, I_{x_2}, ..I_{x_t},$ $I_{y_1}, I_{y_2}, ..., I_{y_t}]$. Finally, feature matrix X, $X = [f_1, f_2, ..., f_{n_p}]$ for the entire image is obtained by combining the feature vectors of all patches.

Once the feature matrix representation is ready, we generate the covariance matrix representation of feature matrix X, $C = X'X^T$. Principle component analysis was used to diagonalizes covariance matrix C by solving the following eigen vector problem:

$$\lambda V = CV \tag{5.6}$$

Where V are the eigen vectors of C and λ represents the corresponding eigenvalues. The eigenvectors are ranked in descending order of eigen values. Choosing d eigenvectors corresponding to the d largest eigenvalues gives us the basis along the directions of maximum variance in features. Thus, the resultant matrix can be represented as $E = [V_1, V_2, ..V_d]^T$.

5.3.2 Saliency Measurement

In the final step, saliency can be measured based on the dissimilarity between patches, which can be simply defined as euclidean distance between patches in reduced dimension.

$$S(R_i) = \omega(i) \sum_{j=1}^{L} \frac{\sum_{s=1}^{d} |x_{si} - x_{sj}|}{1 + dist(p_i, p_j)} \tag{5.7}$$

where i, j are the i^{th} and j^{th} patches of an image and $\omega(i)$ can be defined as a weight factor to adjust the center bias.

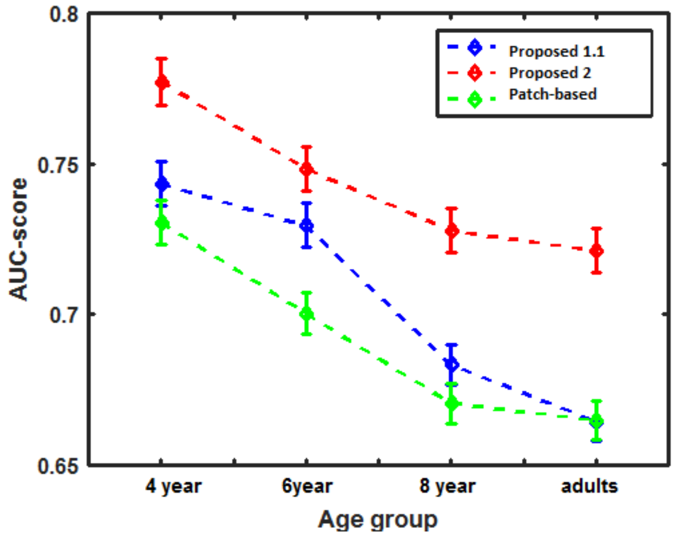


Figure 5.1: Comparison of age-adapted proposed saliency models with baseline models of computational attention system

Similarly to the previous model, the age-adapted framework is incorporated into this model by selecting a different subset of patch sizes for different age groups and incorporating the age-adapted center bias. We can select patch sizes from the set $\{64, 32, 16, 8\}$, which varies from coarser to finer scale. The result of this model is shown in Table 5.3. As expected, all scales are suitable for adults, whereas children are more sensitive to fewer scales.

Table 5.4 lists the fixation prediction accuracies of some state-of-the art saliency models executed unaltered for our age specific gaze dataset over observers of different age groups. From Figure 5.1 and Table 5.4, it is clear that our modification of Itti's model and patch-based models that leverage the age-adapted framework outperformed existing models. Our modified age-adapted algorithms improves the prediction performance for adult observers as well. We believe that difference in the fixation prediction accuracies was evidence for the fact that our algorithm not only personalizes saliency models to achieve optimal performance according to the observer's age group but also improves the prediction performance of adults.

Table 5.3: Average prediction accuracy by proposed patch based method, scale 1,2,3,4 are correspond to a, 16, 8 patch sizes respectively (Proposed 2).

Age \ Scale	1∼4	2∼4	3∼4	4∼4
4 year	0.7678	0.7767	**0.7773**	0.7772
6 year	0.7400	**0.7483**	0.7480	0.7482
8 year	0.7195	**0.7279**	0.7272	0.7269
adult	**0.7212**	0.7113	0.7208	0.7188

Table 5.4: Comparison table of our proposed (Proposed 1.1 and Proposed 2 models) age-adapted models with available computational models of saliency prediction.

Age \ Model	Proposed 1.1	Proposed 2	Itti's	GBVS	Judd's	Patch
4 year	0.7434	**0.7773**	0.6218	0.7184	0.7296	0.7306
6 year	0.7295	**0.7483**	0.6147	0.6969	0.7033	0.7003
8 year	0.6833	**0.7279**	0.6027	0.6722	0.6721	0.6706
adult	0.6646	**0.7212**	0.6062	0.6707	0.6660	0.6649

5.4 Age-adapted Learning Based Saliency Model with Depth Bias

In this section we propose a learning based model based on the recommendations of our analysis results for developing an age-adapted saliency model with depth bias for children, adults and elderly. The general structure of our proposed model is similar to the saliency model in [33] in which the final saliency map is the weighted sum of the features extracted from the input image, where the weights are learned from the adults gaze data. We modified the previous learning based algorithms to reflect the age-related changes in saliency prediction following the below mentioned steps: (a) We include an extra feature map for depth information at different spatial scale and orientation, generated by steerable pyramid [56]. (b) Instead of using the center map independent of the observers age group and image category, we use a weighted mechanism for tuning them in accordance with age-related changes in center bias tendency. (c) Age-related differences in explorativeness tendency was incorporated by selecting different scales of features for different age groups. (d) Finally, the model weights are learned over the age-specific gaze data instead of using adults gaze data.

5.4.1 Feature Used in Machine Learning

We use a learning approach to train a classifier directly from age-adapted eye tracking data. The general structure is that the proposed model learns the weight for combining different features extracted from a input image to generate the final saliency map.

Features that we used for training and testing can be categorized in low-level, mid-level, and high-level features. Low-level feature includes, color features, itti's saliency map (intensity, color, and orientation channel) [28], torralba's saliency map [61], and the local energy of steerable pyramid [56]. In mid-level features horizon feature is extracted similar to [51], as the human's have tendency of naturally directing their gaze on horizon locations in a scene viewing. For high-level features, object and car detection algorithms were applied in the proposed model. Figure 5.3 shows the feature maps extracted from a sample image.

The center bias is included by a adding center map generated by Gaussian blob placed at center of the scene as in [33]. The most important contribution of this work is that for the first time we included a depth feature maps extracted from input 2D scenes in context of the age-related diffrences in depth bias tendecies. The depth maps from single images are estimated using a method proposed in [43] as shown in Figure 5.2.

(a) Original Image (b) Depth Map

Figure 5.2: Input image and its disparity map generated by the algorithme proposed by Liu et al. [43] (black area represent the nearest locations and white area is farthest locations in the scene).

5.4.2 Training and Testing: Age-adapted Model

We used Liblinear support vector machine to train our model over different features extracted from the scene. The model weight learned in [33] was not well suited for the age-adapted mechanism as the depth bias features maps were not included in the proposed model. To fit the training and testing in our age-adapted framework, we follow the recommendations of our analysis results. We can see in Figure 5.3 different features

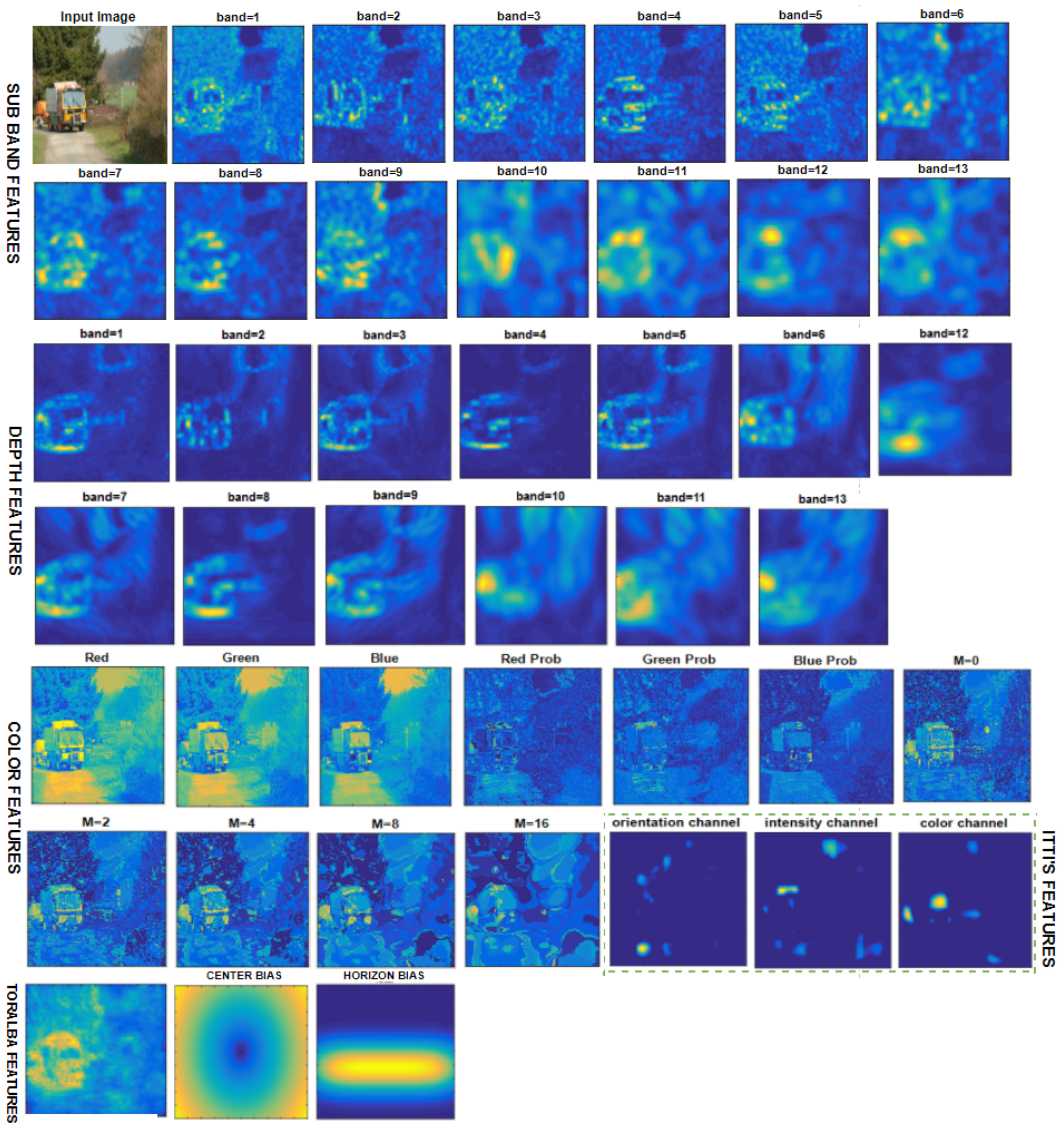


Figure 5.3: Feature maps used in our model

are extracted at several spatial scales (finer to coarser), where band 2 to 4 is scale 1 (finer), band 5 to 8 is scale 2 and band 9 to 12 is scale 3 (coarser). Further, based on the results of explorativeness study, first we focus on feature scale selection, where we found that being less explorative only a few scales are required to represent the level of details explored by children and old observers, conversely, for adults, all the scales are used to represent the level of details they explored in a scene. From our experiments we found that optimal scale for children was 3, whereas scale 2 and 3 for old and scale 1, 2, and 3 for adult observers.

We split the dataset with 192 images into 120 training and 72 test image set. Further, different features were extracted at all the scales, and only a few scales were chosen to represent the age-adapted changes in scene viewing behavior. The use of center map

Table 5.5: Comparison table of our proposed age-adapted models with available computational models of saliency prediction. IUPL in showing the intuitive upper performance limit of any saliency molde proposed for different age groups

Dataset	Age	Itti	GBVS	Patch	Judd	Age-adapted	IUPL
Manmade	Children	0.7140	0.7322	0.6765	0.7365	0.7456	0.8075
	Adults	0.6723	0.6681	0.6083	0.6627	0.6960	0.7656
	Elderly	0.6956	0.7160	0.6679	0.7211	0.7402	0.7603
Nature	Children	0.6224	0.7427	0.7359	0.7657	0.7648	0.7689
	Adults	0.6021	0.6807	0.6513	0.6774	0.6991	0.7239
	Elderly	0.6150	0.7267	0.7186	0.7433	0.7498	0.7587
Fractals	Children	0.6903	0.7610	0.7111	0.7627	0.7831	0.8029
	Adults	0.6582	0.6998	0.6439	0.6855	0.6960	0.7507
	Elderly	0.6758	0.7485	0.7139	0.7562	0.7708	0.7988

feature was altered by including a weight factor for tuning the age-related differences in center bias tendency as reported in our analysis. Top 10 strong positive and negative samples from top 5% of the human saliency map and similarly 10% strong negative pixels from bottom 20

Further, as suggested by the agreement score analysis, strong positive and negative samples from the age-specific human saliency maps were used for training as it was seen that saliency maps of adults were poor predictors for the fixation landings of children and old observers and it is advisable to use the saliency maps of the child observers and old observers to predict fixation landings of observers of these individual groups respectively.

For a given set of feature and levels (positive and negative samples) selected over the age-based human saliency map, SVM is used to learn the optimal weights for depth features i.e. model parameters for each age group in order to predict salient locations for different age groups. Further, for a given image of test dataset, a final saliency map is generated by combining different feature extracted at different scales over an age-specific optimal weights learned for newly included depth map during training the model. Although the prediction accuracy is mentioned for different image categories as shown in Table 5.5 but the model parameters are learned over all the images independent of the class of the image.

$$SM_g(I_i) = w_g F^T(I_i) + b_g \tag{5.8}$$

Where w_g and b_g are optimal weights for age group g, i.e., children, adults, and old separately, $F(I_i)$ is feature vector for the i^{th} test image, this vector is of different size for different age group due to the proposed optimal feature scale mechanism. Based on the saliency values we classified the local pixel as salient or not.

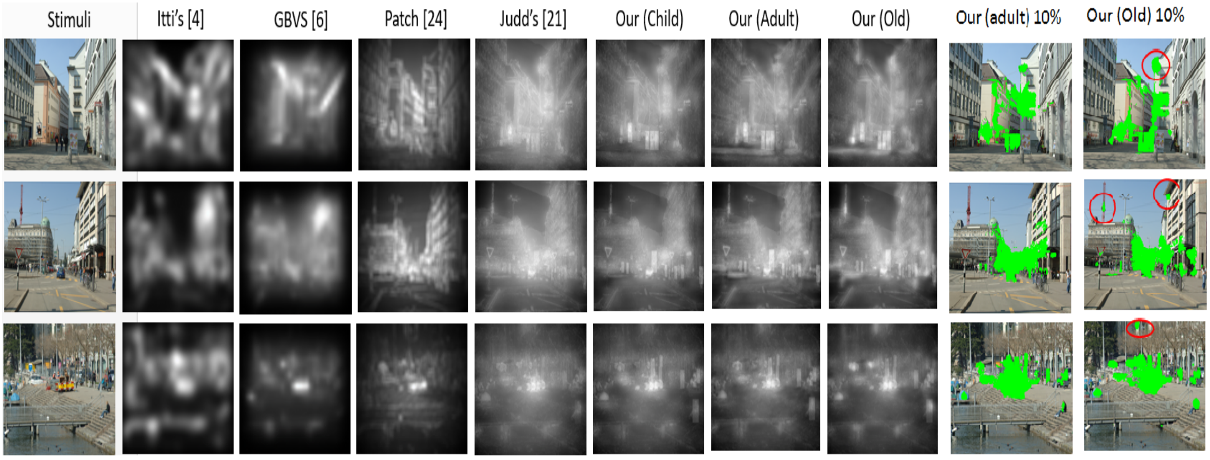


Figure 5.4: Comparison of the saliency map generated from the state-of-the-art methods and and our proposed age-adapted approach, according to the hypothesis elderly focused more to the salient locations in the background and that is reflected in our proposed saliency model as shown in the red circles over the the 10% map.

Performance of our proposed age-adapted model was evaluated using AUC metrics [33]. The evaluation score is listed in Table 5.5 and the resulted saliency map for sample images are shown in Figure 5.4. We compared our model with several state-of-the art methods reported in [28, 22, 29]. The result shows that the proposed age-adapted model outperform the existing saliency models for all age groups independent of the image categories. It is also interesting to note that children observers show better prediction accuracy independent of the saliency model, however the age-adapted model optimized the prediction accuracy each age groups, which suggest that our modifications especially including depth map in learning based model works well for adult observes as well.

5.5 Conclusion

As a part of this chapter, we incorporated the evidences of age based differences in gaze distribution learned in chapter 3 and 4 to the existing state-of-the-art saliency prediction models and make them more generic, i.e. not biased to young adults only but generalized to observers of all age groups

It was seen that the framework was strategically important and easy to upgrade the existing models. Further the upgraded models outperformed the existing models.

Till now we have been discussing developmental changes for one crucial aspect of vision - fixation locations. Next, we switch our discussion to a new direction of gaze behavior which is the fixation duration. We discuss impact of age during two visual processing modes - ambient and focal as a part of the next chapter.

We hold the discussion of age adapted saliency models here and shall return to these upgraded age adapted saliency models when we discuss there application towards sign board detection application in the later part of the dissertation.

Chapter 6

Age-related Changes in Ambient and Focal Visual Processing

6.1 Introduction

In Chapter 3 and 4 we have analyzed fixation locations distribution for the eye gaze data collected for children, adults, and elderly. The impact of age on other gaze attributes like fixation duration and saccade amplitude is totally hidden till now. As a part of this chapter we take a step forward in this direction and uncover age-impact on these aspects of gaze behavior.

Human's gaze behavior during scene viewing is mainly characterized by fixation duration and saccade amplitude at fixated locations. The systematic tendencies of gaze deployment were initially identified in studies of Buswell (1935) [6], and Antes (1974) [2], where it was observed that fixation duration increases and saccade amplitudes decrease over viewing time.

More recently Unema et al., (2005) [64], reported that shorter fixations were associated with longer saccades and conversely, longer fixation was associated with shorter saccades, suggesting the existence of two processing modes: ambient and focal modes, respectively. Ambient mode generally answers to the question of 'where' in the scene and the focal mode is concerned with 'what' is the nature of objects being viewed in the scene [17].

Developmental aspects of the eye movement behavior have been extensively studied in past [5, 54, 9]. However, development of scene perception is relatively unexplored but it has been recently addressed by Helo et al., 2014, [25]. Their results suggest that as the age of an observer increases, fixation duration decreases and saccade amplitude increases during scene exploration. In addition, the study revealed the existence of ambient and focal processing modes in as early as 2 years of age. In another study, Alper et al., 2010, [1] found that the bottom-up features of the images guide the fixation landings more in younger observers than adults.

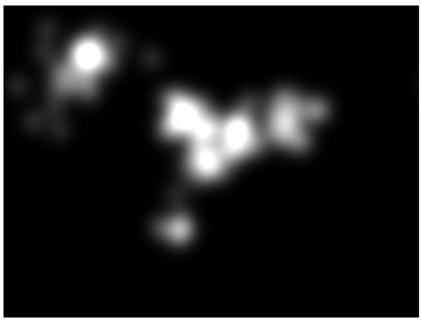
Even though these studies investigated some developmental changes in scene viewing behavior while considering the fixation duration and saccade amplitudes, they did not study the age-impact over distribution of gaze locations during ambient and focal modes. Moreover, the parametric measures used in these studies do not suffice to make comparative comments on the nature of explored region across different age groups.

This study aims to investigate age-associated inter-individual differences in the gaze behavior during ambient and focal visual processing modes. The three major objectives of the current study were, 1) We investigate the maturation of gaze distribution during ambient and focal modes. Explorativeness and bit-rate analysis metrics were used to analyze the gaze distribution. 2) We investigated the influence of bottom-up features of the scene during ambient and focal processing in different age groups. 3) We studied whether a face or a head (manually labeled) in a scene attracts gaze differently during scene viewing across age groups and two modes of viewing.

(a) Ambient heat map

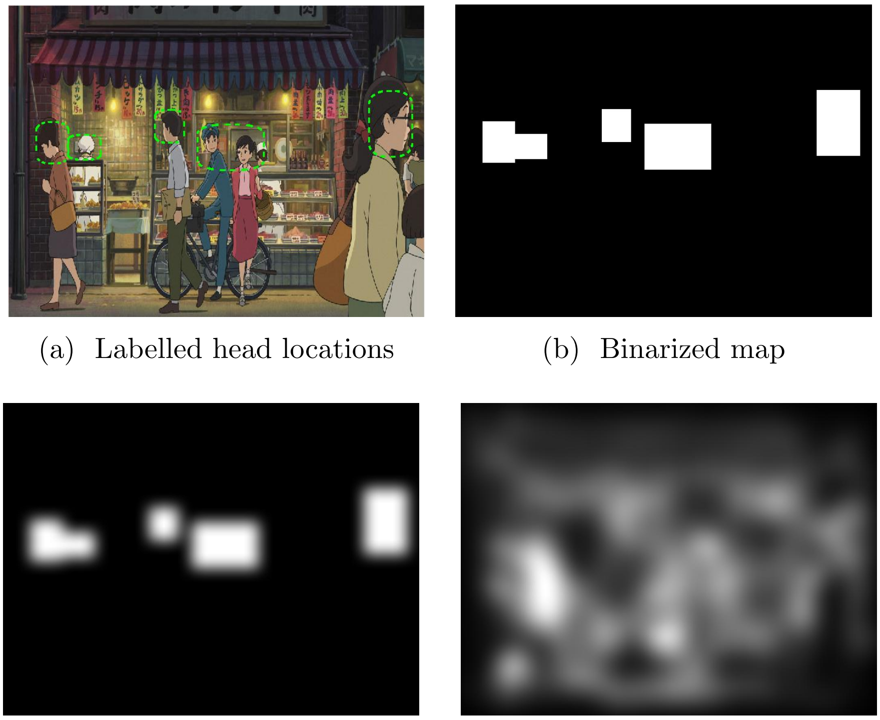
(b) Focal heat map

(c) Ambient saliency map



(d) Focal saliency map

Figure 6.1: Data Representation: Ambient and focal saliency maps obtained by convolving the Guassian at fixation locations



(a) Labelled head locations (b) Binarized map

(c) Head map (d) Bottom-up Saliency map

Figure 6.2: Head map and Saliency map: Head map generated by Gaussian convolution of the binarized map obtained from the manually labelling of the head locations on the data. Bottom-up features are represented by bottom-up saliency map.

6.2 Gaze Data and Fixation Classification

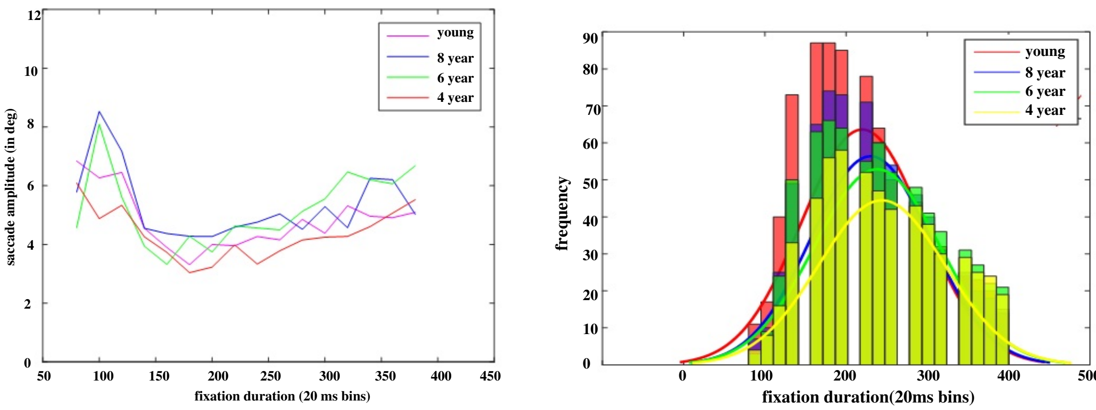
6.2.1 Gaze Data

We analyzed the eye-tracking dataset collected in [25]. The detailed description of the dataset has been covered in Chapter 3. However we summarize the data here for a quick glance for the readers.

We used eye-tracking data of 50 observers from different age groups. All the observers had normal vision or corrected to normal vision. Based on the participant's age, they were assigned to four age groups 4-6 year, 6-8 year, 8-10 year and adult (mean age 29 years). We use 4 year, 6 year, 8 year and young adult to refer to these groups in order. The experiment was conducted on images of the dimension of 1024×764 pixels. These images were taken from children's books and movies and characterized to have an eventful background.

Remote eye tracking system Eyelink 1000 with a sampling rate of 500Hz was used to measure eye gaze and provided the raw data which was sampled to obtain fixations and saccades. Pictures were viewed from distance of 60 cm on the screen having a resolution of 1024×728 pixels. Each stimulus was presented for 10 seconds in each trail.

Data Representation: For a given age group we used classified fixation points (ambient and focal) by all the observers to generate the saliency map and heat map for all images as shown in Figure 6.1. Ambient and focal human saliency maps were obtained by convolving a Gaussian filter across the fixation locations, as in [33].



(a) (b)

Figure 6.3: (a) Saccade amplitude as function of fixation duration: Irrespective of the age groups shorter fixations are associated with the longer saccade and relatively longer fixations associated with shorter saccades. (b)Frequency plot of the number of fixation in each bin: Adults distribution is more concentrated towards the average fixation of the group than children which indicates that ambient and focal tendency changes with age

6.2.2 Data Pre-processing: Fixation Classification

In order to understand the maturation of fixation distribution during two modes, fixations were classified in ambient and focal following the same steps as reported in [25, 64]. The focal and ambient population of the fixations is decided based on the relationship between fixation duration and saccade amplitude. To examine this relationship we plotted average value of saccade amplitude within bins of the size 20 ms for all the fixations up to 400 ms.

As can be clearly seen from Figure 6.3a the fixations across all age groups can be broken down into two distinguishable subsets, one characterized by significantly larger saccades compared to the other. Further to statistically validate the results, MANOVA analysis was performed. The results reported that this observation was significant across all the age groups (**adults-** $F(1,14) = 15.31$,$p < 0.001$, **8 year-** $F(1,14) = 7.46$,$p = 0.016$, **6 year-** $F(1,14) = 5.51$,($p < 0.043$), **4 year-** $F(1,14) = 6.07$,$p < 0.027$).

A bar graph of number of fixations in each bin was plotted to understand the distribution of a number of fixations across bins. As can be seen in Figure 6.3b the fixations distribution gradually become less spread with increasing age which implies, the fixations converge towards the average value of fixation duration with increasing age of an observer. In other words, the percentage of fixations with exceptionally high and low fixation duration decreases with increasing age.

6.3 Quantitative Measures to Analyze Age-impact on Ambient-Focal Processing Modes

In this section, we quantitatively analyze ambient-focal dichotomy for observers across age groups and its impact on gaze locations in terms of explorativeness and data-rate analysis. Consecutively, we studied the influence of bottom-up features during two scene exploration strategies and finally investigated the role of human heads in attracting observer's attention during the two modes.

6.3.1 Explorativeness: Ambient and Focal

Explorative behavior of an observer is his tendency to probe a given scene. The observer's probing tendency differs drastically during ambient and focal visual processing strategies. This difference becomes even more prominent when we observer the gaze behavior of observers from different age groups.

Ambient and focal saliency maps are plausible map to be used to analyze explorative behavior of an observer as the human saliency maps for the two modes differ with age. First order entropy of ambient and focal saliency maps are used to quantify the explorativeness of ambient and focal modes across age groups. Formulation of the ambient and focal explorativeness is as following:

$$H(A_i^g) = \sum_l h_{A_i^g}(l) * log(L \,/\, h_{A_i^g}(l)) \tag{6.1}$$

$$H(F_i^g) = \sum_l h_{F_i^g}(l) * log(L \,/\, h_{F_i^g}(l)) \tag{6.2}$$

Where A_i^g and F_i^g are the ambient and focal saliency maps respectively, generated for i^{th} image over fixation landings of all observers in a group g. $h_{A_i^g}(l)$ and $h_{F_i^g}(l)$ are the histogram entries of intensity value l for saliency maps A_i^g and F_i^g respectively, L is the total number of pixels in A_i^g. The more scattered a saliency map is i.e., the more the number of non zero pixels in the map, the more is its entropy.

In the context of viewing behavior, higher entropy corresponds to the higher explorative viewing behavior of observer as they have more number of scattered saliency points in the scene. Similarly, lower entropy corresponds to the less explorative behavior.

6.3.2 Data-rate Analysis

Exporativeness gives an idea about explored regions but does not consider the time taken to explore the corresponding regions during ambient and focal modes. The study reported in [64] reveals that ambient fixations are associated with significantly shorter fixation duration than the focal fixations which suggests that the visual information processed during ambient and focal modes accord to different time intervals.

The fundamental difference in ambient-focal strategies suggests the requirement of a metric which can measure the rate of visual information processing during ambient and focal scene processing. The data rate metric is developed for this purpose. It is defined as the ratio of a total number of bits required to represent the ambient and focal saliency maps of an image to the total time associated with ambient and focal fixations.

6.3.3 Influence of Bottom-up Features on the Ambient and Focal Processing Modes

Bottom-up features of a scene include features such as color, intensity, and orientation. These are best represented by the Itti's saliency map [27], which extracts these features in parallel by a set of the linear center surrounded operations similar to the visual receptive field, and then combines their normalized value at the later stage to obtain the conspicuous location of an image.

To study the influence of the bottom up features over the two visual processing modes, the bottom-up saliency maps are first thresholded to T different levels representing different percentage of the most salient area of the map. These are then used to predict ambient and fixation landings for different age groups and the prediction performance is measured using AUC score [42].

6.3.4 Role of Human Heads in Seeking Observer's Attention During the Two Modes

A study in [33] on eye movement data of fifteen observers (18-35 years) over 1003 natural images revealed that observers frequently fixated on faces in scenes. This result helps in the development of a better computational model of visual attention by considering faces as feature maps.

In order to develop age adopted saliency models, it becomes important to investigate the role of human heads/faces across age groups during two processing strategies. For the same, we generated a head map for the scene.

To generate head maps first the head locations in a scene are manually labeled for all the images in the dataset, further these manually labeled locations are binarized, and finally Gaussian blurring is used to minimize the effect of the sharp boundaries around the head locations in the binary map. The results of these steps for a sample image are shown in Figure 6.2. These maps are used to evaluate the prediction performance for ambient and focal fixation landings. AUC based measures are used for this purpose.

6.4 Results

We identified fixations in two classes across age groups and formalized the metric to quantify the gaze distributions in terms of explorativeness and bit rate. This section report the results of these analyses in order to understand developmental changes occurring in these two processing modes.

6.4.1 Ambient and Focal Explorativeness

Explorativeness was measured for different age groups during different processing modes. It was found that the explorative behavior does not only vary for focal and ambient processing modes but it is also influenced by an observer's age.

The comparative results of explorativeness are as follows:

- **Explorativeness is more in Focal mode than Ambient:** Explorativeness measure showed that observers were more explorative during the focal mode than in the ambient mode. The same observation was found in all age groups. Paired t-test showed that the explorativeness measure was statistically significant between the ambient and focal modes in all age groups ($p < 0.001$-for all age groups) as in Figure 6.4a.

- **Focal explorativeness increases with age** The result showed that focal explorativeness increased with age. Explorativeness of adults were significantly higher than the children ($p < 0.001$) (ANOVA analysis results reported in Figure 6.4b). However the changes in explorative behavior for ambient processing mode was not significant during childhood as can be seen in Figure 6.4c.

6.4.2 Ambient and Focal Data Rate Analysis

Having analyzed the explorative behavior of observers, their capability of quickly grabbing the scene context during ambient mode and detailed processing of scene during focal mode was measured in terms of data rate metrics. Similar to the explorativeness analysis comparisons was performed between two visual processing modes and over different age groups. The data rate analysis results is as following:

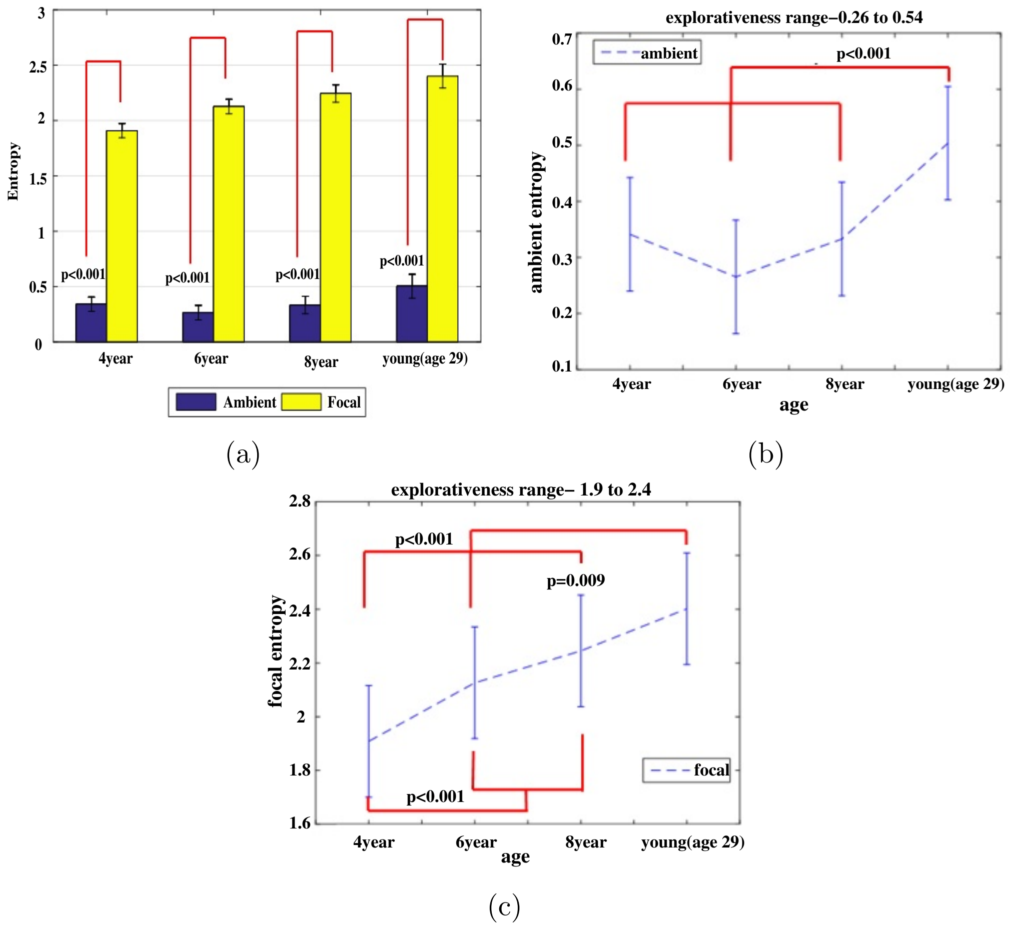


Figure 6.4: Ambient and focal explorativeness is measure by the entropy of the respective saliency maps. (a) Ambient vs. focal: The bar graph shows the focal explorativeness is significantly higher than ambient. (b) Ambient explorativeness vs age: Age impact is not significant for ambient explorativeness during childhood. (c) Focal explorativeness vs age: fixations distribution are significantly different among age groups.

- **Information processing rate is higher in ambient mode** The rate of the visual information processing was higher in ambient mode than in the focal mode (ambient range- 1.26 to 1.51, focal range- 0.27 to 0.32) and it was true for all age groups. The statistical values are indicated in Figure 6.5a.

- **Focal rate increases with observer's age** It is also observed that the rate of visual information processing during ambient mode was same across age groups. On the contrary, the information processing during focal processing increased with age. The statistical values are indicated in Figure 6.5c.

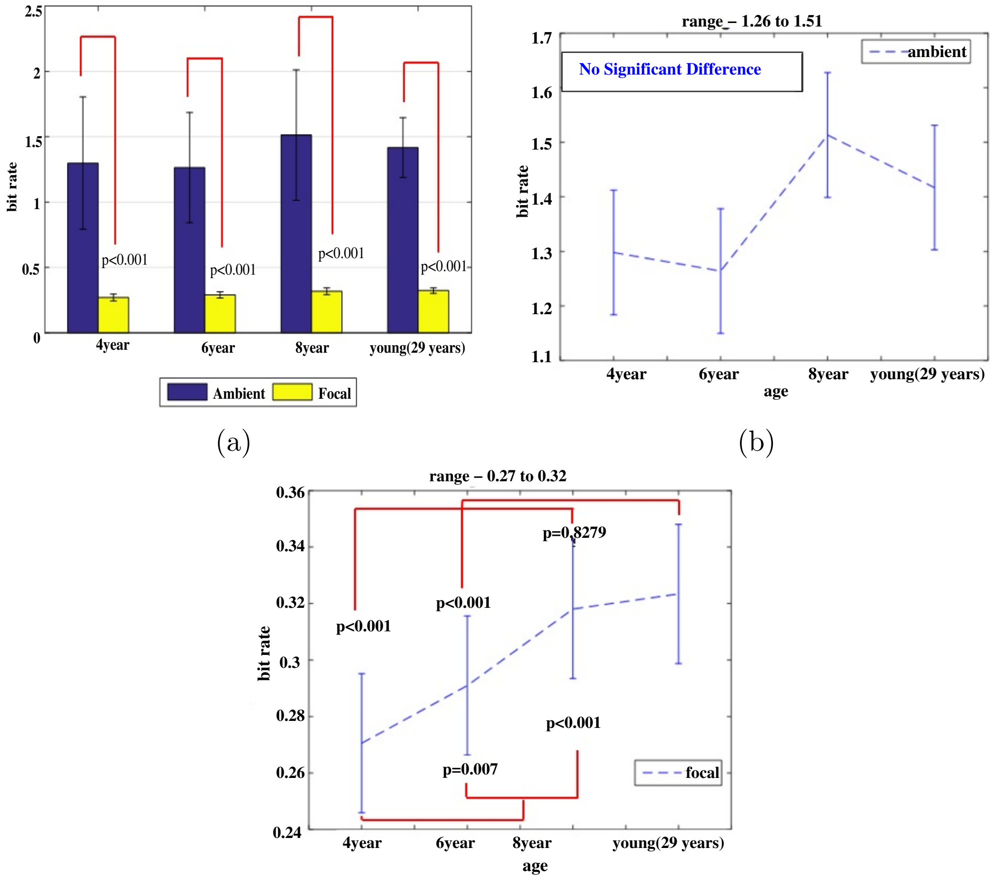


Figure 6.5: Results of the bit rate analysis: (a) Ambient vs. focal bit rate: Ambient information processing rate is significantly higher than the focal. (b) Ambient rate vs age: There is no significant impact of the age during ambient scene processing rate. (c) Focal rate vs. age: The focal rate of scene processing significantly increases with observer's age.

6.4.3 Role of the Bottom-up Features and Head Locations in Attracting the Gaze During Ambient and Focal Mode

Our experiment indicated that the focal mode was more bottom-up than ambient for all age groups, i.e., bottom-up saliency maps are more suitable to predict the focal maps as compared to the ambient maps. In order to validate these results the paired t-test was performed between AUC values of the ambient and focal processing for all the age group; statistical values for which are indicated in Figure 6.6 .

Further analysis revealed that the head locations in scenes attracted the observer's attention equally under the two processing modes for all age groups, except for the observers of four years of age. The statistical significance values are indicated in Figure 6.6.

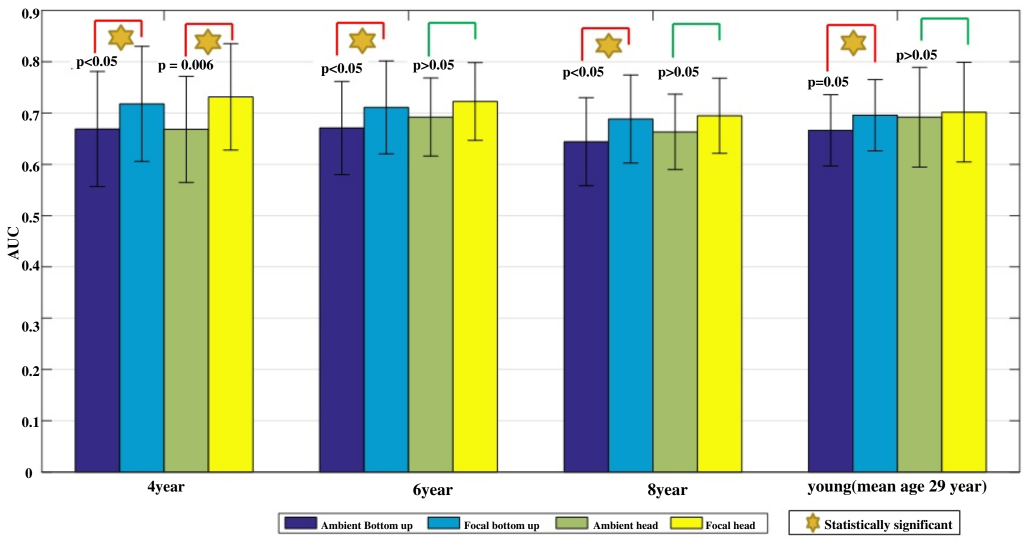


Figure 6.6: Bottom-up and head map influence over ambient and focal modes for different age. AUC score 0.5 is for random performance

6.5 Discussion and Conclusion

In the current work, our aim was to investigate developmental changes in gaze distribution during the two visual processing modes, ambient and focal, across four age groups. The explorativeness and bit rate analysis metrics were formulated to quantify the gaze behavior of observers. Explorativeness measures the spread of fixations while bit rate helps to quantify the fast grabbing and detailed analysis capability of observers during ambient and focal modes, respectively.

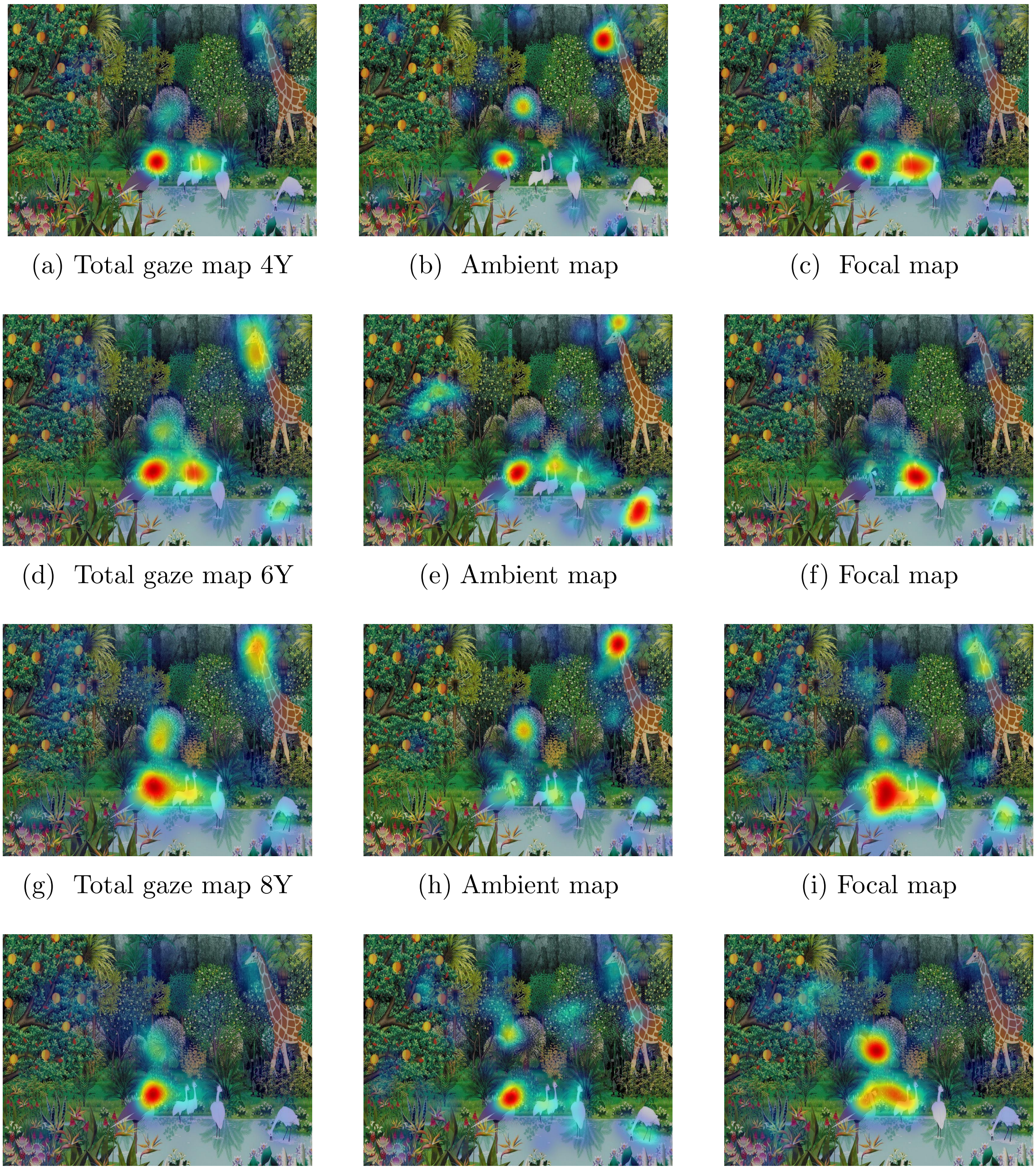
Our results revealed that an observer is more explorative during the focal than the ambient mode of scene viewing and that the explorativeness during the focal mode increases with age of an observer. Similar results were reported in [12] where development of visual cognitive functions during scene viewing was analyzed by examining the exploratory eye movements of 4 to 16 years old observers using two metrics - a total number of gaze points (TNGP) and eye-scanning length (TESL) of gaze points. These result suggested that both TNGP and TESL increases with age suggesting the useful biologic markers for estimating the development of visual cognitive function

Ambient maps are characterized by fewer fixations with shorter fixation duration and the converse is true for focal maps([25, 17, 64]). However, there is no evidence to the relationship between the proportion of fixations and total fixation duration in two modes. Data-rate metrics was developed to parameterize this relation and to examine the developmental changes in this behavior. The results of our study suggested that an observer has higher information processing rate during the ambient mode as compared to the focal mode and this is true for all age groups. Further, information processing rate in focal mode increases with age.

Next major contribution of our work was to demonstrate the impact of bottom-up features over eye movements during ambient and focal modes across the age groups. The results suggested that focal mode is more bottom-up than ambient across all age groups. The results of this study are in agreement with previous findings reported in [17], which also suggested that focal visual processing mode is more bottom-up than ambient for young adults. Similarly, the influence of bottom-up features was investigated by [25] during the early and late phases of viewing, which showed that early phase was more influenced by bottom-up than late phase in older groups.

Finally, we investigated whether an observer is more likely to allocate attention to face or head locations during the ambient or focal processing mode. The results suggested that except for the youngest children (4-years) head locations had a similar impact on attracting the gaze locations during ambient and focal modes of viewing. One similar study reported in [8] revealed that faces attract gaze independent of a task.

Till now we have been discussing the impact of age on gaze behavior of observers in terms of fixation locations, duration, and saccade amplitude, as part of the previous chapters and the present chapter. In the next chapter we conclude our contributions by discussing a practical application of our analysis of age impact on gaze behavior. We discuss the impact of age on sign board saliency during free viewing and task viewing applications.



(a) Total gaze map 4Y (b) Ambient map (c) Focal map

(d) Total gaze map 6Y (e) Ambient map (f) Focal map

(g) Total gaze map 8Y (h) Ambient map (i) Focal map

(j) Total gaze map adults (k) Ambient map (l) Focal map

Figure 6.7: Ambient and focal map for sample image 1 for different age groups

(a) Total gaze map 4Y

(b) Ambient map

(c) Focal map

(d) Total gaze map 6Y

(e) Ambient map

(f) Focal map

(g) Total gaze map 8Y

(h) Ambient map

(i) Focal map

(j) Total gaze map adults

(k) Ambient map

(l) Focal map

Figure 6.8: Ambient and focal map for sample image 2 for different age groups


(a) Total gaze map 4Y


(b) Ambient map


(c) Focal map


(d) Total gaze map 6Y


(e) Ambient map


(f) Focal map


(g) Total gaze map 8Y


(h) Ambient map


(i) Focal map


(j) Total gaze map adults


(k) Ambient map


(l) Focal map

Figure 6.9: Ambient and focal map for sample image 3 for different age groups

Chapter 7

Application: Signboard Saliency Detection in Street Videos

Figure 7.1: Visualization of different tendencies of gaze landings around the signboards in free viewing and task viewing scenario.

7.1 Introduction

Chapter 3 to 6 discuss age-related differences in scene viewing and proposed computational models that can detect salient regions for different age groups. In this chapter we use the knowledge about age-related differences in scene viewing developed so far towards a novel application of signboard saliency detection in street videos.

Up-til now we have only analyzed the age-impact together with bottom and top-up features in scene viewing behavior. In this chapter we have studied the impact of a given task over gaze landings around signboard in street videos and how this behavior varies for adults and elderly observers. Furthermore, we used our previous recommendations like selective scale policies to upgrade the existing video saliency models to predict signboard saliencies in free viewing and task viewing.

Figure 7.2: Segmented signboards of restaurants in the street video and their corresponding saliency rankings, generated from the eye-gaze data collected over thirty participants, (task viewing ranking: board 13, 5, 14, 11, 15, 12, 2, 1, 6, 7, 10, 16, 3, 8, 4, 9.)

The human visual system has developed the ability to process a scene by selecting the most relevant parts of the scene unconsciously. This mechanism is called selective attention, and has been developed in order to quickly spot danger, which was key to human survival. Only within the last 10-15 years, researchers in computer vision exploited the concept of human selective attention for computational models of saliency prediction. However, the saliency models developed so far are limited to predict conspicuous location in static images [27, 22, 32, 13, 20], whereas only a few saliency detection algorithms are proposed for video signals [14, 21, 68].

Accurately predicting salient regions in video signals is critical in many video processing applications such as object detection, video re-targeting, robot navigation, and video compression. However, the video saliency models developed so far still lack in their performance in mimic the human visual system.

There are several studies in vision research that have revealed that gaze distribution during scene viewing is highly dependent on the given task (e.g. free viewing and task viewing) [53, 4, 60]. However, most of these studies have been conducted on static images, and the impact of cognitive factors (e.g. a given task, intentions) on video signals is largely ignored. This study quantitatively analyzes how different viewing strategies (free viewing and task viewing) influence the gaze landings around the signboards in a street video, and then propose an algorithm which can predict the relative saliency of signboards during free viewing and task viewing more accurately. Further we also provide a glance on how free viewing and task viewing behaviors change for different age groups.

Most of the saliency detection algorithms are based on the Feature Integration Theory (FIT) of which the main idea is to compute bottom-up features (spatial and temporal) in parallel and to fuse their saliency to get a so-called master saliency map. However, some of the saliency detection algorithms are based on the guided search model [66] where the master saliency map is generated by combining the saliency map obtained from top-down information together with bottom-up saliency maps.

Following the guided search model, most of the recent video saliency detection algorithms are based upon integrating the spatial information with the temporal information [14, 21, 68]. The main difference between algorithms lies in the way bottom-up and top-down information is represented. The study reported in [68] integrated top-down information represented as semantic cues, such as faces and speech, with bottom-up saliency maps obtained from classical saliency algorithms. Similarly, the study in [48] represented bottom-up information as a color histogram of the video frames, while the temporal map calculated the motion between images by applying RANSAC.

The proposed study analyzes how well a bottom-up saliency prediction algorithm can be used in a novel application of predicting the relative ranking of the saliency of the signboards in street videos.

Our main contribution is three-folded: First, we introduce a new eye gaze dataset collected over 30 observers viewing two street videos in both free viewing and task viewing scenarios. For our dataset we assigned the observers a task of searching a food zone to have lunch. Secondly, we propose a metric for quantitative analysis of the collected eye gaze data to find differences in tendencies of gaze distribution for signboards in street videos during the free viewing and task viewing. We extend this analysis to give a glance on how these tendencies change with age. Finally, we propose a modification to an existing algorithm to improve the prediction accuracy of signboard saliency in street videos.

7.2 Experiment and Analysis

7.2.1 Participants and Video Stimuli

A total of 30 participants attended the experiment including 15 university students (3 female, 12 male, age range 21-31, mean age 24.1) and 15 (4 female, 11 male, age range 66-80, mean age 73.1) elderly participants were recruited from retirement job centers in Tokyo. All the subjects reported normal or corrected to normal vision. Participants were not familiar with the experimental procedure and had not seen the video stimuli before the experiment. All of the participants signed a consent form before the start of the experiment.

Two different street videos, each with a duration of two minutes thirty seconds, were used in the experiment. Each video consisted of 4,473 frames in total. Tobii x2-60 eye-tracker was used for recording the eye-gaze data, whereas the fixations and saccades were detected by the default Tobii fixation filter (I-VT fixation filter). Video stimuli were displayed at a 20 inch LCD monitor of 40 cm width, and all the participants were viewing the stimuli at a 65cm distance from the monitor surface.

7.2.2 Procedure and Task

Thirty participants who took part in the study were divided into two groups of 15 participants each. To avoid repeating the same video in free viewing and task viewing mode for any one participant, one group of the participants watched the video in free viewing mode, whereas the other group watched the same video with the given task of finding a place to have lunch at. Each trial started with the gaze calibration followed by the eye tracking while viewing the videos. Before the video stimuli began, participants were instructed to either view freely or to fulfill the task.

7.2.3 Analysis: Free viewing and Task viewing

The different tendencies of the gaze landings during free viewing versus task viewing can be seen in the heat maps shown in Figure 7.1. Further, in order to determine the relative saliencies of the signboards of restaurants in the street video during the free viewing and task viewing mode, the distribution of fixation locations around the signboards has been analyzed.

To perform this study, first, we manually labeled two instances of each signboard appearing in the video and then interpolated the label for the rest of the frames containing the same signboard. There were a total of 16 signboards labeled in the street video for the full duration of their appearance. Few examples the labelled signboards are shown in Figure 7.3.

Further, two different approaches were adopted to analyze the gaze distribution around the signboards labeled in the previous step. First, we simply measured the total number of fixations that have landed on each signboard during the free viewing and the task viewing scenario for the whole duration of the signboard's appearance. For the second one, we used explorativeness metrics to measure the differences in the scene exploration tendency during the free viewing and the task viewing scenario.

The relative ranking of the signboard saliencies based on the gaze counting during free viewing is shown in Figure 7.2. One way ANOVA was conducted to measure the statistical difference between the signboard's saliencies in the highly salient, salient, average salient and least salient category. A significant difference was reported for gaze landings among the signboards belonging to the four different categories, $F(3, 29) = 3.56, p < 0.001$.

Further, similar to the explorativeness analysis over still images, entropy-based metrics were developed to measure the explorativeness during free viewing and task viewing. In

(a)

(b)

(c)

Figure 7.3: Labeling of signboards in street videos, further, labels are lineally interpolated for whole duration of appearance

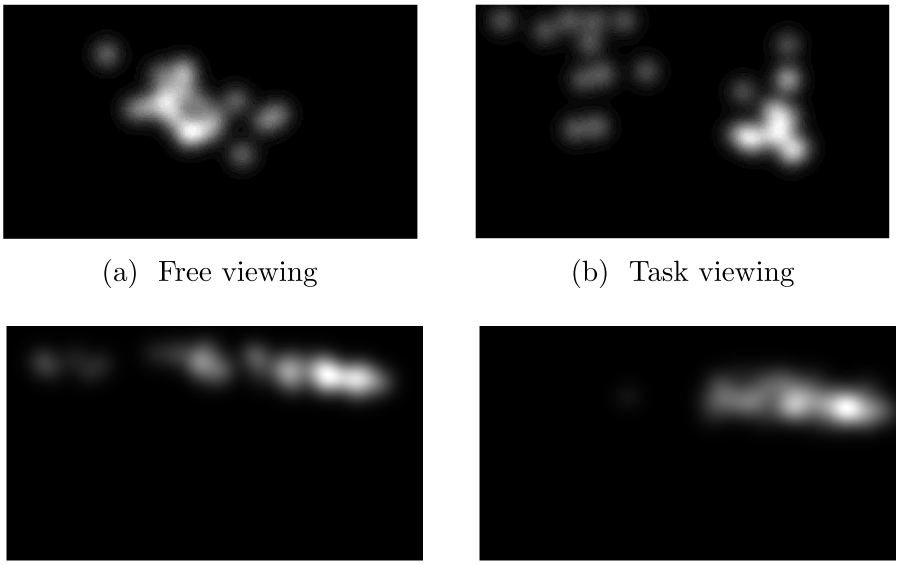
order to do that, we first generated saliency maps by convolving a Gaussian similar to [33] over a combined fixations recorded over 30 frames. As an output of this step, we have generated 149 saliency maps (total 4473 frames divided by 30), showing the area explored during each second of the video. As shown in Figure 7.4a and 7.4b the average of all these 149 maps shows the average rate of exploration during free viewing and task viewing of the street videos. Finally, the explorativeness is measured by measuring the entropy of the 149 saliency maps. Formulation of the explorativeness is as follows:

$$H(I_j) = \sum_{l} h_{I_j}(l) * log(L \,/\, h_{I_j}(l)) \tag{7.1}$$

where I_j is the saliency map of the total gazes recorded during one second of viewing for which entropy is calculated and $h_{I_j}(l)$ is the histogram entry of intensity value l in image I_j, and L is the total number of pixels in I_j.

The average score of the entropy value suggests higher explorativeness during task viewing than free viewing (task viewing - 1.94, free viewing - 1.40). The one-way ANOVA showed an effect of a task in scene exploration tendencies, $F(1, 148) = 22.13, p < 0.001$. Similarly, we have measured the explorativeness tendency for the labeled locations of restaurant signboards only (see Figure 7.4c and 7.4d), where the results showed comparable explorativeness in both viewing scenarios.

We can see from Figure 7.4a and 7.4b, that the center bias tendency is very different during free viewing and task viewing. The center bias for two different viewing modes can further be measured by measuring the euclidean distance between the centroid of the



(a) Free viewing (b) Task viewing

(c) Free viewing (d) Task viewing

Figure 7.4: (a, b) Average tendency of the rate of video exploration during free viewing and task viewing. (c, d) Different tendency of explorativeness around board 14 only during free or task viewing

average maps and the center pixel of the image. The higher euclidean distance for task viewing suggests the lower center bias in the task viewing scenario compared to the free viewing scenario (203 and 267 are the euclidean distances in pixels for free viewing and task viewing).

The proposed analysis suggests three major findings that can be applied in upgrading the existing video saliency algorithm to improve the prediction performance for the relative rankings of signboard saliencies. First, we generated the ground truth rankings of the signboards based on the eye-gaze data collected in the free viewing and the task viewing scenario. Secondly, the explorativeness results indicate a higher exploration tendency during task viewing compared to free viewing. Lastly, we discovered a higher center-bias for free viewing than task viewing.

7.3 Proposed Model for Signboard Saliency Detection

Before proposing any algorithm for signboard saliency detection, we evaluated the performance of existing video saliency algorithms in predicting the fixations landed on the signboards. We selected a few state-of-the art algorithms [27, 22] with their motion included version, which can be aligned with the goal of signboard saliency detection in street videos. The performance of these algorithms are evaluated to answer the following two

Table 7.1: Prediction accuracy of different algorithms for the gaze over the full duration for free viewing and task viewing

	GBVS	s_map	m_map	e_map	Itti
Free	0.7969	0.7626	0.6996	0.7429	0.7961
Task	0.7717	0.7350	0.6836	0.7045	0.7483

questions: (1) How accurately can they predict the gaze points landing anywhere within the frame during free viewing and task viewing? (2) How well can these algorithms predict the gazes only for the signboards in those two viewing scenarios?

As shown in Table 1 the performance of [27] and [22] in predicting the gazes for the whole frame is better than the `s_map` [13] `m_map`[20], and `e_map` [14]. Similarly, the result of prediction accuracy for signboards only showed that GBVS and Itti's perform better in predicting signboard saliencies for the least and highest salient signboards as labeled in Table 2.

Motivated by the prediction accuracy of Itti's model, especially for the signboards belonging to the least and highest salient categories, we applied the recommendations from the analysis results to upgrade Itti's model with motion features [27] to predict the signboard saliencies more accurately for free viewing and task viewing scenarios. Another reason for selecting Itti's model was based on the fact that most existing bottom-up models follow the same basic architecture proposed by Itti et al.

In these models the following basic structure can be observed: (a) Basic visual features such as color, intensity, orientation and motion are extracted over multiple scales of the image, where each scale represents a different level of detail in the scene. (b) All features are investigated in parallel, to obtain the conspicuity map for each feature channel. (c) These features are integrated to obtain the saliency map.

The analysis result suggested that participants explored the video more during task viewing than the free viewing. Thus, to make our model adapt to free viewing and task viewing scenarios, we focused on a feature scale selection mechanism, where we identified the subsets of the feature map scales that best represented the different levels of details viewed by the observers during free viewing and task viewing. Further, we illustrated step combining feature maps at different scales to generate the final saliency map.

$$Intensity = \bigoplus_{i=s}^{6} \mathcal{N}(Intensity_i) \tag{7.2}$$

$$Color = \bigoplus_{i=s}^{6} [\mathcal{N}(\mathcal{RG}_i) + \mathcal{N}(\mathcal{BY}_i)] \tag{7.3}$$

Table 7.2: The signboard saliency scores (AUC score) for the lowest salient signboards (determined by gaze data) generated by different algorithms in free viewing and task viewing.

Board Ranking (Task)		12	15	13	11	Avg.
Board Ranking (Free)		14	16	13	12	
Board Name		3	8	9	16	
GBVS[2]	Free	0.73	0.53	0.65	0.70	0.65
	Task	0.76	0.63	0.52	0.70	0.65
Itti's[1]	Free	0.79	0.61	0.56	0.63	0.64
	Task	0.71	0.72	0.57	0.71	0.67
s_map[18]	Free	0.67	0.61	0.68	0.60	0.64
	Task	0.69	0.65	0.58	0.77	0.67
m_map[19]	Free	0.61	0.65	0.61	0.61	0.62
	Task	0.63	0.62	0.54	0.56	0.58
e_map[6]	Free	0.70	0.62	0.54	0.61	0.61
	Task	0.65	0.56	0.48	0.60	0.57
Ours	Free	0.70	0.51	0.49	0.63	0.58
	Task	0.71	0.55	0.45	0.56	0.56

$$Orientation = \sum_{\theta \in \{0,45,90,135\}} \bigoplus_{i=s}^{6} \mathcal{N}(Orientation_i(\theta)) \tag{7.4}$$

Where $\mathcal{N}$ represents the normalization and s is the starting index from where maps were taken, scale 1 (finer) to scale 6 (coarser). Similarly to the motion features, we have selected a subset of scales for representing the explorativeness in the two different viewing scenarios.

The experimental result shows that the prediction accuracy of the signboard saliency detection during free viewing is highest for the following subset of the coarser scales $s = 4, 5, 6$, conversely the task viewing prediction accuracy improved for the finer scales $s = 1, 2, 3$. This result is in accordance with our previous findings of the explorativeness' differences in free viewing and task viewing. Further, we tuned the center bias weights in the existing model, based on the analysis' results that free viewing has a higher center bias than task viewing. As shown in Table 2 and Table 3, the weight tuning for the center bias together with the subset selection for explorativeness in free viewing and task viewing have slightly improved the relative saliency prediction for both highly and least salient signboards .

Table 7.3: The signboard saliency scores (AUC score) for the highest salient signboards (determined by gaze data) generated by different algorithms in free viewing and task viewing (higher score is better).

Board ranking (Task)		8	7	1	4	Avg.
Board ranking (Free)		4	5	3	1	
Board Name		1	2	13	14	
GBVS[2]	Free	0.79	0.87	0.75	0.76	0.79
	Task	0.83	0.82	0.64	0.64	0.73
Itti's[1]	Free	0.82	0.79	0.88	0.84	0.83
	Task	0.82	0.75	0.71	0.71	0.74
s_map[18]	Free	0.81	0.74	0.88	0.79	0.80
	Task	0.75	0.72	0.66	0.68	0.70
m_map[19]	Free	0.66	0.67	0.57	0.64	0.63
	Task	0.69	0.62	0.61	0.57	0.62
e_map[6]	Free	0.76	0.72	0.75	0.74	0.74
	Task	0.77	0.72	0.61	0.61	0.67
Ours	Free	0.81	0.86	0.79	0.81	0.81
	Task	0.83	0.84	0.73	0.75	0.78

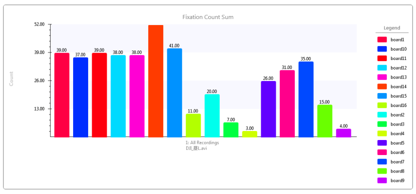
7.4 Analysis and Future Direction: Adult and Elderly Signboard Saliency

So far we have only been studying the signboard saliencies in free viewing and task viewing scenarios without any consideration given to the observer's age. As a part of this section we analyze the collected eye-gaze data to understand the age-impact on signboard saliencies. The analysis reported here is limited to only basic fixation attribute (i.e. number of fixation)

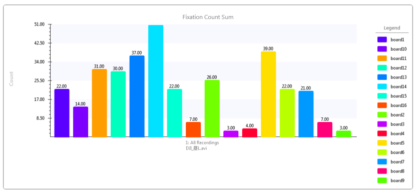
It can be visualized from the heat-map of the sample frames that scene exploration was very different in adults and elderly in Figure 7.5. The statistical evidence is reported in Figure 7.6, which simply shows the total fixations landed on different signboards for adults and elderly observers in free viewing and task viewing scenario. Figure 7.7 shows that adults and elderly have more agreement in billboard rankings in task viewing than in free viewing.

Age-adapted saliency modeling for videos is out of the scope of this dissertation and we thus introduce a possible future research direction and encourage researchers to use the age-adapted saliency study reported in this dissertation for signboard saliency detection in street videos for adults and elderly observers.

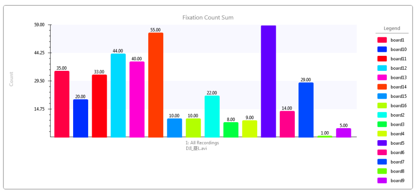
Figure 7.5: Visualization of different tendencies of gaze landings around the signboards in free viewing for adults and elderly participants



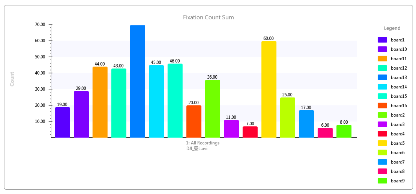
(a) Free viewing



(b) Task viewing

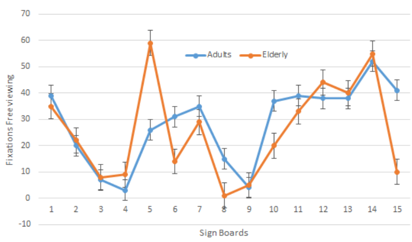


(c) Free viewing

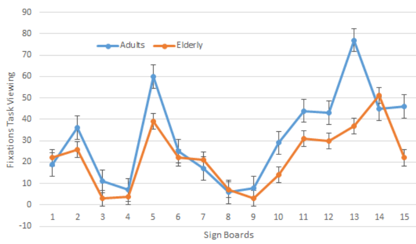


(d) Task viewing

Figure 7.6: (a, b) Total gaze landing for different signboards from adults during free viewing and task viewing. (c, d) Total gaze landing for different signboards from elderly during free viewing and task viewing



(a) Free viewing



(b) Task viewing

Figure 7.7: (a) Total gaze landing for different signboards from adults and elderly in free viewing. (b)Total gaze landing for different signboards from adults and elderly in task viewing.

7.5 Conclusion

As a part of this chapter we presented a novel application of video saliency detection for ranking signboards within a street video based on the relative signboard saliencies. We contributed a collection of eye-gaze data for 2 street videos for both free viewing and task viewing scenarios, and encourage the researchers to work towards developing age adapted saliency models for videos. We proposed a quantitative analysis method based on the rate of the explorativeness and center bias metrics. The analysis results were used in upgrading the basic saliency model for predicting signboard saliencies more accurately for free viewing and task viewing.

Chapter 8

Conclusion and Future Direction

In this work, we have investigated the variations in gaze landing behavior with observer's age during scene viewing. The initial motivation for the study was a news article that reported children eating jelly from the pot in Japan which triggered us to think that what attracts children's visual perception and how different are their scene viewing strategies from adults. Inspired by related findings in vision psychology, which have shown that people tend to vary in their gaze behavior with age, we introduced certain analysis metrics to study age-related changes in scene viewing behavior for observers of different age groups (Chapter 3 and 4) and proposed to use these metrics to develop an age adapted computational model of saliency prediction (Chapter 5).

Primarily we introduced four metrics to formalize developmental changes in scene viewing. First of these was the "explorativeness" measure that learns the spread of fixation landings across age groups. It was seen that the explorativeness increases with age with children being less explorative and adult being the most explorative i.e. The children tend to focus on fewer details in the scene compared to adults that explore more. Thus while developing a computational model, it is important that we choose appropriate scales of feature maps according to the level of details the age group for which the fixations need to be predicted.

Next, "agreement score" metric was introduced which measures the uniformity between the fixations of observers within the age group and between the age groups. It was found that the fixation landings of an age group are significantly more consistent with those of observers who belong to the same group compared to the observers of other age groups. This suggests that while learning parameters for saliency model for an age group, the learning model should only consider fixation points of the observers of the same age group as their training input.

Finally the age-impact on two bias metrics - center bias and depth bias were introduced. The result on center bias tendency revealed that center bias decreases with increasing age from Children to adults and further it decreases for elderly participants.

Most of the state-of-the art saliency models include center map as a feature map to reflect center bias tendency in saliency prediction. It has been seen that including center bias in the existing model increases the prediction accuracy considerably. Further, the results of the age-impact on center bias tendency were used to upgrade the existing models to reflect the age-related changes by tuning the center feature map of existing algorithms.

Gaze landings for man-made dataset showed that whereas child observers focus more on the scene foreground, i.e., locations that are near, elderly observers tend to explore the scene background, i.e., locations farther in the scene. Considering this result a framework is proposed in this dissertation to quantitatively measure the depth bias tendency across age groups. The depth bias can be incorporated in age-adapted model by including a depth map at different spatial scales for children and elderly age groups.

These trends in the scene viewing behavior measured from these analysis metrics and the corresponding recommendations for saliency prediction are incorporated in existing saliency models to upgrade them to become more generalized and not biased to adult observers only. It was seen that the age-adapted models outperformed the existing state-of-the-art saliency prediction models with prediction accuracy of all age groups more than the previous models.

In addition to the analysis of impact of age over fixation locations we also developed some insights for the fixation duration and saccade amplitude. We studied the age impact on scene viewing behavior during two visual processing modes - focal and ambient. The result showed that focal is more explorative than ambient and focal explorativeness changes with age. Further, the data-rate analysis metrics revealed that information processing rate during ambient mode is significantly faster than in focal mode. Furthermore, the results of this study showed that bottom-up saliency maps are more suitable to predict focal map than ambient. Finally, analysis revealed that head locations in a scene attract observers attention equally during ambient and focal modes for all all age groups.

Finally the the knowledge about age-related differences in scene viewing tendency were used in a novel application of signboard saliency detection in street videos (Chapter 7). The contribution of this work is three folded: First, we introduced a new eye gaze dataset collected over 30 observers viewing two street videos in both free viewing and task viewing scenarios, the tasks was to look for a restaurant for lunch. Secondly, we propose a metric for quantitative analysis of the collected eye gaze data to find differences in tendencies of gaze distribution for signboards in street videos during the free viewing and task viewing. Finally, we propose a modification to an existing algorithm to improve the relative saliency prediction for both highly and least salient signboards.

Our analysis framework was seen to generalize well to images of different categories - images from children books, fractal (i.e. computer generated), man-made (urban scene), and nature images. The framework was strategically beneficial as it could be easily incorporated into most of the existing saliency models rendering them to become age-adapted.

The research towards the study of age impact on visual behavior of observers does not end here and has a lot of future scope and potential applications in many practical domains. Some of the immediate applications that we see and which are beyond the scope of this thesis are - 1) Study of impact of age on visual processing while viewing videos - till now we have only been restricted to still images in our works. The study should be extended to video stimuli too. 2) Learn the early existence of disease like autism by comparing the viewing pattern of children with the learned behavior of our model. 3) Age-related differences in gaze landings of adult and elderly can be further extended to analyze the viewing differences in adult and elderly drivers. Some other applications of this research can be in designing the books for elementary school children, where having prior knowledge about the scene viewing tendency can help in placing the important content in such a way that it can increase the efficient learning.

We encourage researchers in the field of applied perception and computer vision communities to exploit this newly discovered age parameter and tune the models for even better results.

References

[1] Alper Açık, Adjmal Sarwary, Rafael Schultze-Kraft, Selim Onat, and Peter König. Developmental changes in natural viewing behavior: bottom-up and top-down differences between children, young adults and older adults. *Frontiers in psychology*, 1, 2010.

[2] James R Antes. The time course of picture viewing. *Journal of experimental psychology*, 103(1):62, 1974.

[3] Eva Aring, Marita Andersson Grönlund, Ann Hellström, and Jan Ygge. Visual fixation development in children. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 245(11):1659–1665, 2007.

[4] Christian Balkenius. Attention, habituation and conditioning: Toward a computational model. *Cognitive Science Quarterly*, 1(2):171–214, 2000.

[5] Oliver Braddick and Janette Atkinson. Development of human visual function. *Vision research*, 51(13):1588–1609, 2011.

[6] Guy Thomas Buswell. *How people look at pictures.* University of Chicago Press Chicago, 1935.

[7] Kyle R Cave and Jeremy M Wolfe. Modeling the role of parallel processing in visual search. *Cognitive psychology*, 22(2):225–271, 1990.

[8] Moran Cerf, E Paxon Frady, and Christof Koch. Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of vision*, 9(12):10–10, 2009.

[9] A Chandna. Natural history of the development of visual acuity in infants. *Eye*, 5(1):20–26, 1991.

[10] Ming-Ming Cheng, Niloy J Mitra, Xiaolei Huang, Philip HS Torr, and Shi-Min Hu. Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):569–582, 2015.

[11] Lijuan Duan, Chunpeng Wu, Jun Miao, Laiyun Qing, and Yu Fu. Visual saliency detection by spatially weighted dissimilarity. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 473–480. IEEE, 2011.

[12] Chiyomi Egami, Kiichiro Morita, Takashi Ohya, Youhei Ishii, Yushiro Yamashita, and Toyojiro Matsuishi. Developmental characteristics of visual cognitive function during childhood according to exploratory eye movements. *Brain and Development*, 31(10):750–757, 2009.

[13] Erkut Erdem and Aykut Erdem. Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of vision*, 13(4):11–11, 2013.

[14] Yuming Fang, Zhou Wang, Weisi Lin, and Zhijun Fang. Video saliency incorporating spatiotemporal cues and uncertainty weighting. *IEEE Transactions on Image Processing*, 23(9):3910–3921, 2014.

[15] F Fioravanti, P Inchingolo, S Pensiero, and M Spanio. Saccadic eye movement conjugation in children. *Vision research*, 35(23):3217–3228, 1995.

[16] Burkhart Fischer, Monica Biscaldi, and Stefan Gezeck. On the development of voluntary and reflexive components in human saccade generation. *Brain research*, 754(1):285–297, 1997.

[17] Brice Follet, Olivier Le Meur, and Thierry Baccino. New insights into ambient and focal visual fixations using an automatic classification algorithm. *i-Perception*, 2(6):592–610, 2011.

[18] Junko Fukushima, Teppei Akao, Sergei Kurkin, Chris RS Kaneko, and Kikuro Fukushima. The vestibular-related frontal cortex and its role in smooth-pursuit eye movements and vestibular-pursuit interactions. *Journal of Vestibular Research*, 16(1, 2):1–22, 2006.

[19] Junko Fukushima, Tatsuo Hatta, and Kikuro Fukushima. Development of voluntary control of saccadic eye movements: I. age-related changes in normal children. *Brain and Development*, 22(3):173–180, 2000.

[20] Josselin Gautier and Olivier Le Meur. A time-dependent saliency model combining center and depth biases for 2d and 3d viewing conditions. *Cognitive Computation*, 4(2):141–156, 2012.

[21] Chenlei Guo and Liming Zhang. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE transactions on image processing*, 19(1):185–198, 2010.

[22] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In *Advances in neural information processing systems*, pages 545–552, 2007.

[23] Alan A Hartley, James Kieley, and Craig RM Mckenzie. Allocation of visual attention in younger and older adults. *Attention, Perception, & Psychophysics*, 52(2):175–185, 1992.

[24] Mary Hayhoe and Dana Ballard. Eye movements in natural behavior. *Trends in cognitive sciences*, 9(4):188–194, 2005.

[25] Andrea Helo, Sebastian Pannasch, Louah Sirri, and Pia Rämä. The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision research*, 103:83–91, 2014.

[26] Elizabeth L Irving, Martin J Steinbach, Linda Lillakas, Raiju J Babu, and Natalie Hutchings. Horizontal saccade dynamics across the human life span. *Investigative ophthalmology & visual science*, 47(6):2478–2484, 2006.

[27] Laurent Itti and Christof Koch. Computational modelling of visual attention. *Nature reviews. Neuroscience*, 2(3):194, 2001.

[28] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998.

[29] Lina Jansen, Selim Onat, and Peter König. Influence of disparity on fixation and saccades in free viewing of natural scenes. *Journal of Vision*, 9(1):29–29, 2009.

[30] Scott P Johnson. Development of visual perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5):515–528, 2011.

[31] Tilke Judd, Fredo Durand, and Antonio Torralba. Fixations on low-resolution images. *Journal of Vision*, 11(4):14–14, 2011.

[32] Tilke Judd, Frédo Durand, and Antonio Torralba. A benchmark of computational models of saliency to predict human fixations. 2012.

[33] Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th international conference on*, pages 2106–2113. IEEE, 2009.

[34] Canan Karatekin. Eye tracking studies of normative and atypical development. *Developmental review*, 27(3):283–348, 2007.

[35] Christoph Klein. Developmental functions for saccadic eye movement parameters derived from pro-and antisaccade tasks. *Experimental Brain Research*, 139(1):1–17, 2001.

[36] Christoph Klein and Bernd Feige. An independent components analysis (ica) approach to the study of developmental differences in the saccadic contingent negative variation. *Biological Psychology*, 70(2):105–114, 2005.

[37] Kurt Koffka. *Principles of Gestalt psychology*, volume 44. Routledge, 2013.

[38] Onkar Krishna and Kiyoharu Aizawa. Age-adapted saliency model with depth bias. In *Proceedings of the ACM Symposium on Applied Perception*, page 5. ACM, 2017.

[39] Onkar Krishna, Kiyoharu Aizawa, Andrea Helo, and Rama Pia. Gaze distribution analysis and saliency prediction across age groups. *arXiv preprint arXiv:1705.07284*, 2017.

[40] Onkar Krishna, Toshihiko Yamasaki, Andrea Helo, Rämä Pia, and Kiyoharu Aizawa. Developmental changes in ambient and focal visual processing strategies. *Electronic Imaging*, 2017(14):224–229, 2017.

[41] David LaBerge. Spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance*, 9(3):371, 1983.

[42] Olivier Le Meur and Thierry Baccino. Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior research methods*, 45(1):251–266, 2013.

[43] Fayao Liu, Chunhua Shen, and Guosheng Lin. Deep convolutional neural fields for depth estimation from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5162–5170, 2015.

[44] Beatriz Luna, Krista E Garver, Trinity A Urban, Nicole A Lazar, and John A Sweeney. Maturation of cognitive processes from late childhood to adulthood. *Child development*, 75(5):1357–1372, 2004.

[45] Beatriz Luna, Katerina Velanova, and Charles F Geier. Development of eye-movement control. *Brain and cognition*, 68(3):293–308, 2008.

[46] Chih-Yao Ma and Hsueh-Ming Hang. Learning-based saliency model with depth information. *Journal of vision*, 15(6):19–19, 2015.

[47] Keng Teck Ma, Terence Sim, and Mohan S Kankanhalli. Vip: A unifying framework for computational eye-gaze research. In *HBU*, pages 209–222. Springer, 2013.

[48] Yu-Fei Ma, Xian-Sheng Hua, Lie Lu, and Hong-Jiang Zhang. A generic framework of user attention model and its application in video summarization. *IEEE transactions on multimedia*, 7(5):907–919, 2005.

[49] David J Madden and Wythe L Whiting. Age-related changes in visual attention. *Recent advances in psychology and aging*, pages 41–88, 2004.

[50] Charles A Nelson, Christopher S Monk, Joseph Lin, Leslie J Carver, Kathleen M Thomas, and Charles L Truwit. Functional neuroanatomy of spatial working memory in children. *Developmental psychology*, 36(1):109, 2000.

[51] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3):145–175, 2001.

[52] Sebastian Pannasch, Johannes Schulz, and Boris M Velichkovsky. On the control of visual fixation durations in free viewing of complex images. *Attention, Perception, & Psychophysics*, 73(4):1120–1132, 2011.

[53] Rajesh PN Rao, Gregory J Zelinsky, Mary M Hayhoe, and Dana H Ballard. Eye movements in iconic visual search. *Vision research*, 42(11):1447–1463, 2002.

[54] André Roucoux, Christine Culee, and Marguerite Roucoux. Development of fixation and pursuit eye movements in human infants. *Behavioural brain research*, 10(1):133–139, 1983.

[55] Chengyao Shen and Qi Zhao. Webpage saliency. In *European Conference on Computer Vision*, pages 33–46. Springer, 2014.

[56] Eero P Simoncelli and William T Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *Image Processing, 1995. Proceedings., International Conference on*, volume 3, pages 444–447. IEEE, 1995.

[57] J Stankiewicz, Nathan J Anderson, and Richard J Moore. Using performance efficiency for testing and optimization of visual attention models. In *Proceedings of SPIE*, volume 7867, page 78670Y, 2011.

[58] Benjamin W Tatler. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of vision*, 7(14):4–4, 2007.

[59] Benjamin W Tatler, Nicholas J Wade, Hoi Kwan, John M Findlay, and Boris M Velichkovsky. Yarbus, eye movements, and vision. *i-Perception*, 1(1):7–27, 2010.

[60] JG Taylor and N Fragopanagos. Modelling the interaction of attention and emotion. In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*, volume 3, pages 1663–1668. IEEE, 2005.

[61] Antonio Torralba. Modeling global scene factors in attention. *JOSA A*, 20(7):1407–1418, 2003.

[62] Antonio Torralba, Aude Oliva, Monica S Castelhano, and John M Henderson. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological review*, 113(4):766, 2006.

[63] A Treisman. The perception of features and objects in attention: Selection, awareness and control, 1993.

[64] Pieter JA Unema, Sebastian Pannasch, Markus Joos, and Boris M Velichkovsky. Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual cognition*, 12(3):473–494, 2005.

[65] Ian Van Der Linde, Umesh Rajashekar, Alan C Bovik, and Lawrence K Cormack. Doves: a database of visual eye movements. *Spatial vision*, 22(2):161–177, 2009.

[66] Jeremy M Wolfe, Kyle R Cave, and Susan L Franzel. Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human perception and performance*, 15(3):419, 1989.

[67] JN YGGE, Eva Aring, Ying Han, Roberto Bolzani, and Ann HellstrÖm. Fixation stability in normal children. *Annals of the New York Academy of Sciences*, 1039(1):480–483, 2005.

[68] Yun Zhai and Mubarak Shah. Visual attention detection in video sequences using spatiotemporal cues. In *Proceedings of the 14th ACM international conference on Multimedia*, pages 815–824. ACM, 2006.

[69] Jianming Zhang and Stan Sclaroff. Saliency detection: A boolean map approach. In *Proceedings of the IEEE international conference on computer vision*, pages 153–160, 2013.

[70] Libo Zhang, Lin Yang, and Tiejian Luo. Unified saliency detection model using color and texture features. *PloS one*, 11(2):e0149328, 2016.

[71] Lingyun Zhang, Matthew H Tong, Tim K Marks, Honghao Shan, and Garrison W Cottrell. Sun: A bayesian framework for saliency using natural statistics. *Journal of vision*, 8(7):32–32, 2008.

Publications

Journals

[1] Onkar Krishna, Andrea Helo, Pia Rama, and Kiyoharu Aizawa. Gaze Distribution Analysis and Saliency Prediction Across Age Groups. In *arXiv preprint arXiv:1705.07284*, 2017, (Accepted, Plos One Journal).

[2] Onkar Krishna, Kiyoharu Aizawa, and Go Irie. Computational Attention System for Children, Adults and Elderly. In *Transaction on Applied Perception (ACM TAP)*, 2017, (Under review)

Conferences

[3] Onkar Krishna, Toshihiko Yamasaki, Andrea Helo, Rama Pia, and Kiyoharu Aizawa. Developmental Changes in Ambient and Focal Visual Processing Strategies. In *Electronic Imaging*, pages 224-229, 2017. (Accepted as oral)

[4] Onkar Krishna, and Kiyoharu Aizawa. Age-adapted saliency model with depth bias. In *Proc. of the ACM Symposium on Applied Perception*, page 5, 2017. (Accepted as oral)

[5] Onkar Krishna, Saskia Reimerth, and Kiyoharu Aizawa. Signboard Saliency Detection in Street Videos. In *Acoustics, Speech and Signal Processing (ICASSP)*, 2018. (Accepted)

[6] Onkar Krishna, and Kiyoharu Aizawa. Billboard Saliency Detection in Street Videos for Adults and Elderly. In *IEEE International Conference on Image Processing (ICIP)*, 2018. (Submitted)

Awards

[7] Onkar Krishna. Received Electronic Imaging travel grant to attend EI student showcase. A Jury selected 17 best papers from applicants and awarded the grants accordingly, 2017.