EasyChair Preprint

№ 170

Significance of Bottom-up Attributes in Video Saliency Detection Without Cognitive Bias

Jila Hosseinkhani and Chris Joslin

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

May 25, 2018

Significance of Bottom-up Attributes in Video Saliency Detection Without Cognitive Bias

Jila Hosseinkhani
Dept. Systems and Computer Engineering
Carleton University, Ottawa, Canada
jila.hosseinkhani@carleton.ca

Chris Joslin
School of Information Technology
Carleton University, Ottawa, Canada
chris.joslin@carleton.ca

***Abstract*—Saliency in an image or video is the region of interest that stands out relative to its neighbors and consequently attracts more human attention. To determine the salient areas within a scene, visual importance and distinctiveness of the regions must be measured. A key factor in designing saliency detection algorithms for videos is to understand how different visual cues affect the human perceptual and visual system. To this end, we investigated the bottom-up features including color, texture, and motion in video sequences for both one-by-one and combined scenarios to provide a ranking system stating the most dominant circumstances for each feature individually and in combination with other features as well. In this work, we only considered the individual features and various visual saliency attributes investigated under conditions in which we had no cognitive bias. Human cognition refers to a systematic pattern of perceptual and rational judgements and decision-making actions. Since computers do not typically have this ability, we tried to minimize this bias in the design of our experiment. First, we modelled our test data as 2D images and videos in a virtual environment to avoid any cognitive bias. Then, we performed an experiment using human subjects to determine which colors, textures, motion directions, and motion speeds attract human attention more. The proposed ranking system of salient visual attention stimuli was achieved using an eye tracking procedure. This work provides a benchmark to specify the most salient stimulus with comprehensive information.**

***Keywords*—Visual Attention Model, Saliency Detection, Bottom-up Features, Human Visual System, Cognitive Bias, Semantic Analysis.**

I. INTRODUCTION

Human brain receives a massive amount of information when watching virtually any scene. The Human Visual System (HVS) is capable of processing this information rapidly and focusing on the salient regions of the scene. These selected regions which are more interesting to the subject are called salient areas. Human visual attention contains two types of processes: pre-attentive and attentive [1].

Pre-attentive (subconscious) processing rapidly and automatically categorizes an image into regions in a spatially parallel manner to search for significant information across an image. However, the attentive (conscious) processing or focused attention incorporates the goals and desires of the viewer through the process of searching in a serial manner which is time consuming compared to pre-attentive detection [2]. Understanding the processing mechanism of HVS helps us to know how to properly prioritize and combine the visual stimuli as well as the low-level, mid-level, and high-level features in the design of attention models.

Physiological and psychological studies illustrated that the effective factors on visual attention and eye movements are categorized into bottom-up and top-down types [3]. Bottom-up factors capture pre-attentive attention very quickly and have a strong impact on the human visual selection system. On the other hand, top-down factors capture the attention much slower and are influenced by bottom-up factors. Bottom-up and top-down factors are known as low-level and high-level features respectively. In the past two decades, researchers focused on designing visual attention models (VAMs) were inspired by the HVS to reduce the huge volume of data to more visually informative and important data. Saliency detection models or VAMs employ bottom-up and/or top-down factors to search for the salient part of data. Bottom-up based models use low-level attributes such as color, texture, size, contrast, brightness, position, motion, orientation, and shape of objects. Basically, these attributes are rapidly scanned and detected by the human visual system. However, top-down based models exploit high-level context dependent attributes such as face, human, animal, vehicle, text, etc. Both bottom-up and top-down factors can be exploited to design VAMs but because of the complexity and time limitation, few integrated approaches have been proposed that use both factors to detect the salient parts within a scene [4].

To generate the saliency map, different feature maps are usually produced for bottom-up attributes first. Then, these maps are fused to produce the overall activation map indicating the most salient areas. It should be noted that the basic feature maps can be combined to generate top-down attributes as well.

The validation of the saliency maps is usually performed by comparing them with eye movement tracking datasets as the ground-truth data. Studies show that the human visual system is attracted to objects rather than locations [5]. In fact, the pre-attentive part of the HVS firstly segments the scene into objects in a rapid scan [5]. This segmentation is mostly performed based on the low-level attributes.

In this paper, we focus on the study of bottom-up attributes as visual stimuli such as color, texture, motion direction, object velocity, and object acceleration. Our goal is to investigate how they influence the HVS. The aim of this work is to achieve a ranking system to identify the hue range, texture pattern, motion direction, object velocity, and object acceleration that are most likely to be attractive for the HVS in terms of saliency. The

advantage of this work can be described as finding the order of the significance of the bottom-up attributes in determining the salient regions in a scene. The lack of a comprehensive study in this area motivated us to carry out this work.

We will provide a more detailed explanation about existing studies related to the impact of bottom-up attributes on the visual attention system in addition to our designed experiment in next sections. The rest of this paper is organized as follows: Section 2 reviews the related works and provides an overview of how the experiments have been performed to understand HVS response to the bottom-up attributes. Section 3 describes the characteristics of the generated dataset for our experiment and its methodology. Section 4 contains the results of our proposed experiment for each individual attribute, and finally Section 5 concludes the paper.

II. Related Works

Visual attention includes the procedure of selecting the significant and interesting areas across visual information that humans receive in daily life. The selection procedure by the HVS is performed through eye movements. Researchers investigate how visual stimuli influence human eye movements in order to estimate the most salient regions in a scene and consequently design visual attention models to extract those regions.

We provide the concluded results of the existing works and a brief description of the performed experiments in this section. Most of the existing experimental studies focused on investigating differences of 2D and 3D visual data and their impact on the human visual system.

In the literature, few experimental studies have been performed to investigate the impact of the bottom-up factors on the human visual system and eye movements. Previous studies only focused on texture, color, and mostly depth attributes.

For example, Khaustova et al. [17] designed an experimental study to understand how texture complexity, depth, quantity, and visual comfort influence the way people observe 3D content in comparison with 2D content. They utilized uncrossed disparity (i.e. all objects were behind the display plane) for the all stereoscopic content. Two experiments were performed using an eye-tracker and a 3D-TV display. In the first experiment, 51 subjects participated in the test. They found that the objects with crossed disparity are the most salient, even if observers experience discomfort due to the high level of disparity.

The second experiment was designed with the aim of investigating whether depth is a determinative factor for visual attention. In this experiment, 28 observers watched the scenes that contained objects with crossed and uncrossed disparities with different textures. They discovered that texture is more important in comparison with depth for selection of salient objects. They finally concluded that the objects with crossed disparity are significantly more important compared to 2D content. However, objects with uncrossed disparity have the same influence on visual attention as 2D objects.

They reported that there is no relation between the fixation durations and disparities which is in opposition to former studies [17]. They also found that the gaze points were concentrated and centered in the center of the scene during the first 4 seconds of the experiment, but for the other time intervals the gaze points were spread over the entire scene [17].

Hkkinen et al. [18] analyzed the eye movements of participants watching a six-minute movie in both stereoscopic and non-stereoscopic versions. They considered four shots of the movie. The results indicated that viewers tend to look at the actors in the 2D version. In this test, 20 students participated. The short film (6 minutes and 20 seconds long) was presented. They used a Hyundai 46-inch polarizing stereoscopic display with a resolution of 1920 × 1080 pixels. The film was shown with a TriDef stereoscopic player and a Tobii X120 eye movement tracker was utilized. The viewing distance was 140 cm. The participants were trained to compare which of the versions was better. They found that eye movements of subjects are mostly concentrated on the actors and their immediate neighborhood. Based on the eye movement patterns, they inferred that the observers are mostly looking for socially relevant information [18]. These factors are categorized as high-level and content-based features in video sequences. They reported that the eye movements spread more widely in the 3D versions. Also, the objects coming toward the observer caught more of the viewers' attention [18].

Khaustova et al. [19] in another experiment, generated six scenes with different modified parameters using Blender. The modified parameters were: texture complexity and the amount of depth changing the camera baseline and the convergence distance at the shooting side. Their experiment was performed using an eye-tracker and a 3D-TV display. They ensured that each observer had only seen the content of each scene once to avoid memory bias [19]. A Tobii x50 eye-tracker and 42 LG 42LW stereoscopic display with line interleaved technology were used as the setup for this test. In the experiment, the distance from the observer to the eye tracker was around 60 cm and the distance from the observer to the screen was 4.5H (2.34 m). The duration of the experiment was about 10 minutes for each participant. In this test, 135 people (106 males and 39 females from 21 to 60 years old) participated. Each image was tested on 15 observers. Their results illustrate that disparity makes saccade length shorter; however, it does not affect fixation durations [19]. They inferred that texture complexity is significant in salient area selection.

Gelasca et al. [3] designed a subjective experiment to investigate what colors attract human attention more. The goal of this experiment was to quantify the color saliency and to provide a ranking for some of the most common colors. 11 persons participated in the test (3 females and 8 males, aged 19-28). They selected 12 colors including red, pink, magenta, violet, yellow, orange, green, cyan, blue, light blue, maroon, and dark green. The tested colors were chosen in the CIELab color space but there is no available information about their range. The experiment consisted of two cycles. During the first cycle, 20 synthetic images were presented to the subjects containing 12 colored disks. In the first cycle, the task was to choose at first three or four colors which subjects considered the most salient among the displayed colors. Afterwards, the same images were shown but subjects were asked to choose only one or two colored circles which attracted their attention most. They also repeated the experiment for the images containing four colored disks to confirm their results. Table I shows their results as a ranking table for 12 tested colors.

TABLE I
RANKING FOR COLORS ACCORDING TO GELASCA ET AL. [3].

Color	Overall Sum of Hits per Color
Red	128
Yellow	87
Green	84
Pink	60
Orange	44
Blue	32
Cyan	32
Magenta	26
Light Blue	16
Maroon	14
Violet	11
Dark Green	10

The results of color saliency ranking.

Based on the results, they divided colors into two overall groups. The colors that had much more priority were red, yellow, green and pink. The colors with lower saliency were reported as light blue, maroon, violet and dark green.

III. EXPERIMENTAL METHODOLOGY

The main purpose of this work is to achieve a ranking system among different bottom-up stimuli to be able to design a saliency detection algorithm. We hypothesized that color, texture, motion, and depth influence visual attention as the following formula:

$$S = \alpha C \ \Theta \ \beta T \ \Theta \ \gamma M \ \Theta \ \theta D \qquad (1)$$

where S indicates the saliency value, C, T, M, and D illustrate color, texture, motion, and depth respectively. Also, α, β, γ, θ are the weights that show the amount of importance of each stimulus in absorbing attention. It should be noted that $C = [c_1, c_2, c_3, \dots]$, $T = [t_1, t_2, t_3, \dots]$, $M = [m_1, m_2, m_3, \dots]$, and $D = [d_1, d_2, d_3, \dots]$ are vectors in a ranked manner to order each individual stimulus. Θ represents the operator that determines how to combine different feature maps. This operator can be a summation, a multiplication, an averaging, an optimization problem, etc. We will introduce the best way to combine different bottom-up attributes maps later in our algorithm design as a future work.

To this end, we designed an experiment to find those weights and orders for both individual assessment of stimuli as well as their combination states. We asked human subjects to watch image/video content datasets and their eye movements were recorded using an eye tracker device. Then, we studied and analyzed the fixation and saccade points to estimate the pattern of the HVS to discover which stimuli are more effective on the attention system and how they change its operation. We obtained a ranking system to explain which colors, textures, motion directions, and motion speeds are most attractive for human vision. Therefore, we found a priority quantity for different ranges of each individual feature. The answer of this question is useful to examine the efficiency of the designed VAMs for saliency detection. Therefore, we will know which attributes to focus on more in designing our future VAM. We designed a set of experiments to investigate each attribute individually. Then, we tested a limited number of the permutations of the attributes.

In this study, we had 25 participants (11 females and 14 males) within ages 18-35. Participants only interacted with a 55-inch LG 3D-TV screen and a tripod mounted eye tracker device. We used an SMI eye-tracker iView 120 Hz as the eye tracking equipment. Participants were expected to watch 39 images and 74 videos, while their eye movements were recorded. A set of 2D images and video sequences were created using Adobe Illustrator and Adobe After Effects to show the participants. This experiment was a free-viewing task and participants did not need to perform other tasks at the same time. We combined all images and videos together to generate a video with a duration of 15 minutes.

We assessed participants' vision in advance to avoid having vision issues such as color blindness, or low visual acuity. Also, we provided an instruction to participants on how to act during the experiment before they started the test. To avoid misunderstandings and recording false data as much as possible, we trained them by showing a few sample images and videos before starting the actual test.

Our designed experiment contains five stages for each participant:

1) A Visual Test to Check any Vision Issues.

All the subjects were checked by pretests to assess their:

- Visual acuity using Snellen chart
- Color blindness using Ishihara graphs

A Snellen chart was prepared to check participants' visual acuity. We used an online Ishihara test website (i.e. Ishihara 38 Plates CVD Test) including 38 plates available at [20].

In our study, the subjects who did not pass either of the two vision pretests were not allowed to continue to the eye-tracking experiment.

2) A Verbal Explanation of the Instruction.

We instructed each subject on how to behave during the calibration stage, the training stage, and the actual test itself. We emphasized that they should keep their position without any movement during the test because their distance from the eye tracker device and the TV must be fixed within the entire experiment. In addition, we asked them not to rotate their head while watching the video as well as concentrating on the video.

In this way, we were assured to avoid many problems which might happen during the test and make it inefficient by producing invalid data.

3) A Calibration Procedure.

The eye tracker requires a calibration for each subject in order to learn the characteristics of their eyes. The SMI iView 120 Hz has an automatic calibration software. We used a 9-fixation point calibration setting. Subjects who did not provide proper data at the calibration stage were not allowed to participate in the test. According to our experience, people with high prescription glasses or lower eyelids may have difficulty passing the calibration stage. Therefore, these situations may have caused to the failure in calibration stage.

4) A Training Period.

We trained each subject before starting the main visual attention test to avoid facing misunderstanding during the

actual test as much as possible. We prepared different image/video sets for the training stage and showed them to subjects to help them get more familiar with the test practically. These images or videos were not considered in our final results.

5) A Visual Attention Test.

In this stage, we presented our produced datasets to the participants and asked them to do a free-viewing task by simply observing the images or videos. Then, we recorded their eye gaze and eye movements using an SMI eye-tracker. The distance between subjects and the screen was 217.60 cm which is around 3.2 times the display height based on the recommendations listed in ITU-R BT.2022 which is a standard guide on general viewing conditions for subjective assessment of image/video dataset on flat displays [21]. The distance between the eye tracker device and subjects was 60 cm.

The rest of this section provides detailed explanation about different parts of our experiment. We broke down the experiment into individual parts to test bottom-up attributes including color test, texture test, motion direction and velocity test, and contrast test.

A. Color Test

In order to obtain reliable results, ideally a high number of different colors should be considered. However, this would make the task too complicated with huge number of different permutations of the color positions and causes eye fatigue in participants. Therefore, we selected 12 main colors in the HSV color space that is more compatible with the human vision and perceptual system.

We created 10 two dimensional images using Adobe Illustrator for our color test. Each image consists of 12 colored disks of the same size located on the circumference of a circle like a clock dial. The background was gray with luminance of (120, 0, 50) in HSV because gray is a neutral color and has an average intensity difference with most of the colors. In this way, we can avoid high contrast that may cause any bias in saliency of the colors. Our selected colors include red, yellow, green, cyan, blue, magenta, orange, turquoise, pink, dark red, dark green, and dark blue. To introduce our selected colors, we illustrated their HSV characteristics in the Table II. A sample of our created image is shown in Figure 1.

TABLE II
HSV COLOR TABLE

Color Number	Color Name	H °	S %	V %
01	Dark Blue	240	100	50
02	Orange	30	100	100
03	Green	120	100	100
04	Blue	240	100	100
05	Cyan	180	100	100
06	Red	0	100	100
07	Turquoise	180	100	50
08	Pink	300	100	100
09	Purple	300	100	50
10	Yellow	60	100	100
11	Dark Green	120	100	50
12	Dark Red	0	100	50

HSV color space indicates Hue, Saturation, and intensity Value/Brightness respectively.

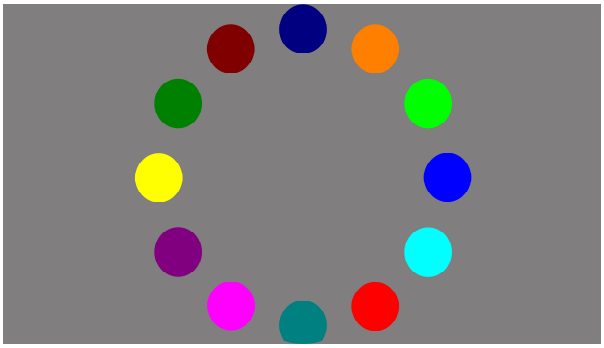

Fig. 1. A sample image for experimental color test.

The colored disks were displaced along the circle circumference within 10 images in a way that different colors can be scattered along the circle and no high contrast happens. To avoid categorizing colored disks based on their HSV characteristic, we made sure not to have very similar colors beside each other. The human visual system usually pays more attention to the center of scene which is known as center bias. This fact is used in photography and film making strategies. According to this fact, we located the colored disks close to the center of the image with equal distance from the center that leads us to reach a circular circumference.

According to the ITU-R BT.2022 document, each image was presented to each participant for 10 seconds and we embedded a plain gray image between two consecutive images for 3 seconds to clear the participants' gaze point.

B. Texture Test

We selected three different levels of texture from a complexity perspective in our experiment including low, medium, and high complexity. The low complexity texture is known as the absence of a pattern on objects and low contrast. The medium-level contains simple geometrical patterns, Finally, the high-level appears with more complex geometrical patterns, higher contrast, and dense edges.

We created 10 two dimensional images for our texture test in a similar manner to the color test. Each image consists of 12 textured disks located on the circumference of a circle and the background is gray. Textures are chosen from gray-scale images within similar intensity ranges to achieve more similar texture patterns from an intensity/brightness perspective. Otherwise, we may have contrast leading to bias in the saliency detection stage. A sample image for the texture test is shown in Figure 2.

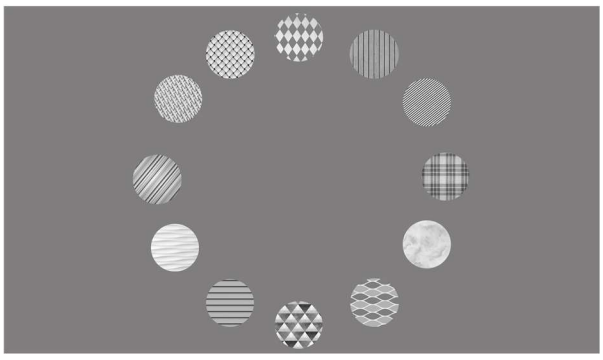

Fig. 2. A sample image for experimental texture test.

We presented each image to participants for 10 seconds and a plain gray image was embedded between two consecutive images for three seconds.

C. Motion Direction and Velocity Test

We utilized four main directions of the motion including horizontal on both sides (toward left and right), vertical on both sides (upward and downward), diagonal with an angle of 45 degrees in both sides (toward up and down), and diagonal of 135 (i.e. -45) degrees in both directions (upward and downward). We only used linear directions and no curvature movement was considered to avoid encountering complexity and having massive numbers of states in our experimental study.

We exploited velocity and acceleration in our motion test to investigate their influence on the human attention system. To this end, we used two different motion patterns such as motion with the constant speed, and motion with acceleration. We used white circles as the moving objects in a gray background. All the circles have the same size but different movement directions or speeds.

Rendered videos for this test are divided into three main groups: 1) moving circles with the same speed in different directions, 2) moving circles in the same direction with different speeds and occasionally different accelerations, and 3) moving circles in different directions with different speeds/accelerations. We rendered 14 video sequences for first group; 7 and 5 video sequences for the second and third groups respectively.

IV. Results and Data Analysis

In this section, we consider the results of each part of the test individually. The scan path, fixation points, and saccade paths of all participants were analyzed to extract the area that participants gazed at more as salient parts. It should be noted that at the analysis stage, we used the following computational equations to obtain the number of fixations A_n for N participants. If we assume f_{i_j} implies i^{th} feature such as color, texture, and motion at the j^{th} gazing order, then vector F can be stated as:

$$F = [f_1, f_2, f_3, \dots, f_P]^T \in \{C, T, M\} \quad (2)$$

where $C = [c_1, c_2, c_3, \dots, c_Q]^T$, $T = [t_1, t_2, t_3, \dots, t_L]^T$, and $M = [m_1, m_2, m_3, \dots, m_O]^T$. In our study 12 colors, 12 texture patterns, and 8 motion directions used. After analyzing participants' scan-path, fixation, and saccade points, we reached a matrix of features for all participants. In this matrix, S_j shows the summation of a feature f_{i_j} in jth column and each column indicates the order of attention. Therefore, the total number of fixations for each feature A_n can be defined as equation (4).

$$S_j = (\sum_{i=1}^{N} f_{i_j}) \quad (3)$$

$$A_n = \sum_{j=1}^{M} 2^{(1-j)} \cdot S_j = \sum_{j=1}^{M} \sum_{i=1}^{N} 2^{(1-j)} \cdot f_{i_j} \quad (4)$$

where N and M illustrate the number of participants and the number of order of the attention respectively. We assigned weights for each order based on powers of $\frac{1}{2}$ to emphasize the importance of the order of fixation.

A. Color Test

According to fixation maps of all 24 participants (11 females and 13 males), the graph of Figure 3 was extracted. This graph indicates that red, yellow, dark red, pink, and cyan are the most salient colors. Paying attention to the ranking table of the 12 used colors leads us to conclude that warm colors such as red, dark red, and pink and bright colors such as yellow and cyan usually are more salient for the HVS. Dark blue, dark green, purple, and turquoise are the least salient colors because they were fixated on less and mostly at the end of the watching time duration for each image.

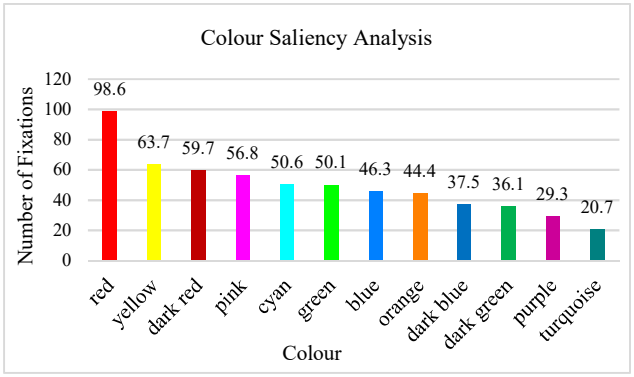

Fig. 3. Color saliency graph to show the order of importance among 12 selected colors.

B. Texture Test

We provided a look-up table (i.e. Table III) to show the labels of each texture used in our test. It helps understand better the saliency ranking graph for different patterns.

TABLE III
Texture Look up Table

Texture Label	Pattern Sample	Texture Label	Pattern Sample
SQ3		D4	
DM2		Spl12	
DM1		Spl14	
Spl2		D7	
DM8		H2	
D3		V1	

Texture labeling.

Figure 4 shows the ranking of saliency for 12 chosen texture patterns. According to this graph, DM8, SQ3, spl2, and DM2 absorbed more attention along with all 25 participants (11 females and 14 males). Therefore, more complex textures which contain dense edges such as DM8, SQ3, and spl2 and patterns with higher contrast such as DM8 will become more interesting and outstanding for human vision. We conclude that areas with dense edges are more important than areas with high contrast in the intensity as textures. However, high contrast parts with smaller areas of the intensities which makes an entirely homogeneous and repeated pattern such as D7, cannot stand out as a salient pattern.

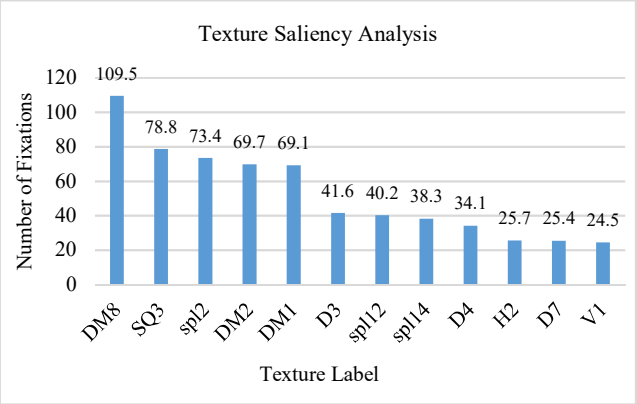

Fig. 4. Texture saliency graph to show the order of importance among 12 selected texture patterns.

On the other hand, Spl12, V1, H2, and D7 are the least salient patterns in our experiment which are more simple and ordinary patterns. Based on our observes and evidences, we concluded that simple patterns with lower edges and lower intensity difference are not attractive to human attention system while dense and compact edges are distinctive.

C. *Motion Test*

We tested four main motion directions including horizontal, vertical, diagonal of 45 and 135 degrees by rendering 2D videos with Adobe Illustrator and Adobe After Effects. We rendered 26 videos with time durations of 3-7 seconds each by incorporating both motion direction and motion speed. Generated videos for investigating different directions can be divided into four different categories: 1) movements in one main direction with two different sides, 2) movements in two main directions which contain both sides and result four directions in total, 3) movements in three main directions which result six different directions considering both sides, and 4) movements in all directions.

Four main directions on both sides will give eight different directions. To this end, four 2D videos were rendered using Adobe After Effects. In this group, each main direction is compared in two statuses like upward and downward to consider which orientation of these direction are more significant for human subjects.

According to our results among 23 participants (10 females and 13 males), horizontal movement toward the right side is more attractive than the left side. Also, vertical movement downward seems to be more salient than upward. Diagonal movements for 45 degrees downward is more salient compared to upward orientation. For 135 degrees is the opposite way i.e. upward orientation became more salient.

Table IV shows the results of the tests for the second category of movement directions. We rendered six 2D video sequences to compare pairs of main movements on both sides including vertical and horizontal, vertical and diagonal 45 degrees, vertical and diagonal 135 degrees, horizontal and diagonal 45 degrees, horizontal and diagonal 135 degrees, and finally diagonal 45 with diagonal 135.

TABLE IV
RANKING FOR TWO DIFFERENT DIRECTIONS

Compared Directions	Salient Direction
Horizontal vs. Vertical	Vertical
Horizontal vs. Diagonal +45	Horizontal
Horizontal vs. Diagonal -45	Horizontal
Vertical vs. Diagonal +45	Vertical
Vertical vs. Diagonal -45	Vertical
Diagonal +45 vs. -45	Almost same

According to table IV, vertical movements are more interesting than horizontal ones, however both vertical and horizontal directions stand out compared to diagonal directions of both 45 and 135 degrees. Among different diagonal orientations upward 45, downward 135, upward 135, and downward 45 took more attention respectively.

We rendered four 2D videos to compare three main movement directions with each other. The contents of these six videos include the following combination of directions: (horizontal, diagonal 45 and 135), (vertical, diagonal 45 and 135), (horizontal, vertical, diagonal 45), and (horizontal, vertical, diagonal 135). Figures 5-8 illustrate the graphs of the ranking for those four videos in comparison with all participants.

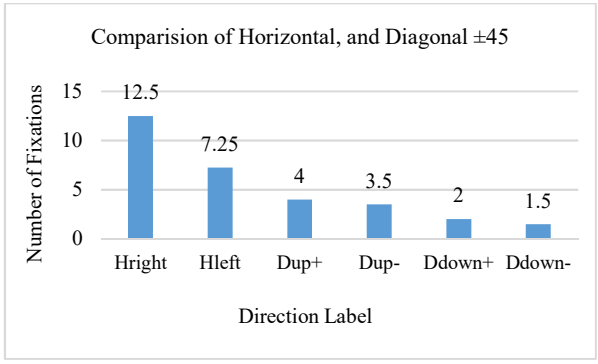

Fig. 5. Resulted graph to compare motion directions in horizontal, diagonal of +45 degrees, and -45 degrees.

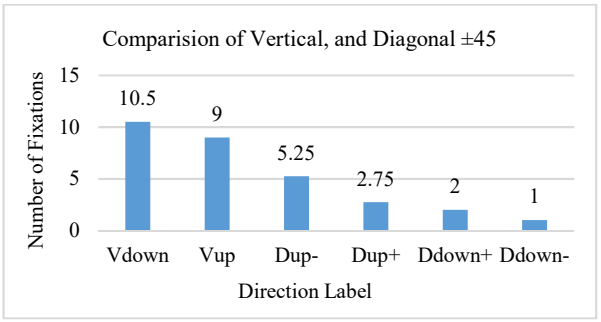

Fig. 6. Resulted graph to compare motion directions in vertical, diagonal of +45 degrees, and -45 degrees.

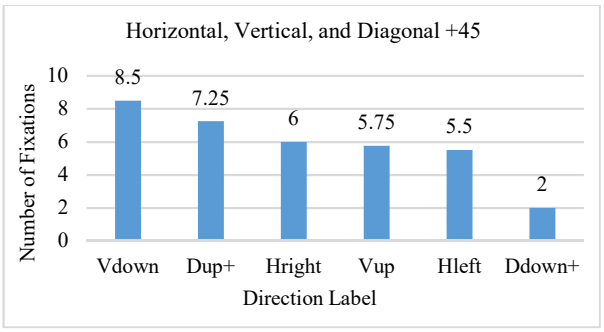

Fig. 7. Resulted graph to compare motion directions in horizontal, vertical, and diagonal of +45 degrees.

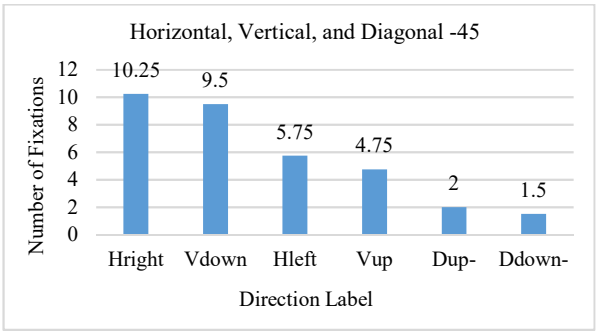

Fig. 8. Resulted graph to compare motion directions in horizontal, vertical, and diagonal of -45 degrees.

Finally, one video was rendered to compare all four main directions (i.e. 8 directions considering both sides) together. The graph in Figure 9 shows the results of ranking for all directions.

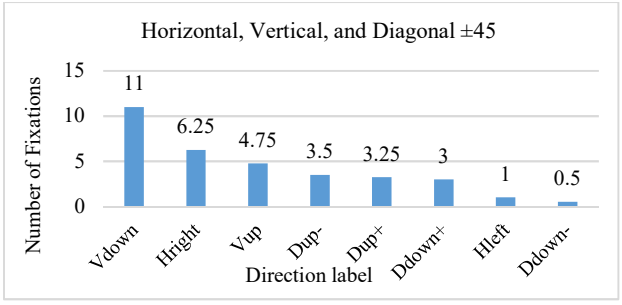

Fig. 9. A graph to compare motion in all selected directions.

Based on this graph, vertical downward is the most salient direction followed by horizontal toward the right side. Vertical upward absorbed subjects' attention more than other directions. Horizontal left side and diagonal 135 downward are the least salient ones.

The rest of this section assesses the saliency of movement direction for the rendered video which included velocity and acceleration. For this purpose, we rendered 12 different videos. These videos are divided into two main groups: 1) objects with the same movement directions but different speeds and/or accelerations and 2) objects with different directions and speeds simultaneously.

According to our results, we observed that the fastest objects usually stand out more. However, if the speed exceeded a threshold and became too fast, subjects could not follow and track those objects because they did not have enough time to focus on them. Therefore, they ignore objects which move too quickly. On the other hand, the slowest objects are attractive to the HVS as well because subjects have more time to see and track those objects within a scene. In addition, any rare movements which includes abrupt acceleration absorb more attention than other objects.

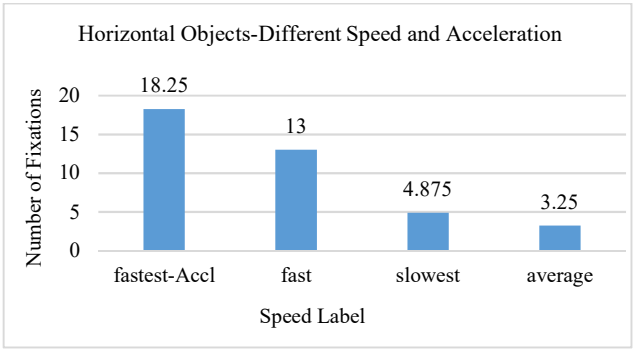

Fig. 10. Resulted graph to compare motion speed for four objects in horizontal, direction with different speed and acceleration.

We noticed that speed and acceleration are much more significant compared to the motion direction. In all videos, participants fixated on very fast, very slow, and any rare motions regardless of the motion direction. Based on the eye tracking results, we can conclude that motion speed outweighs motion direction in saliency detection. For example, Figure 10 is the result of moving in only the horizontal direction with different velocities and accelerations.

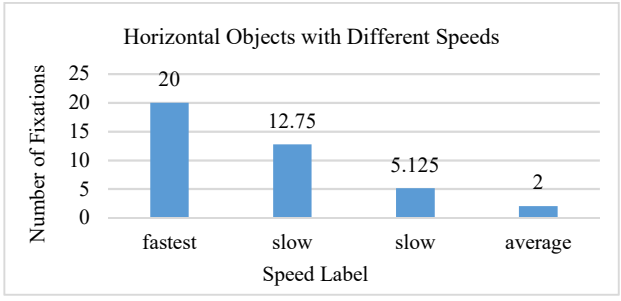

Fig. 11. Resulted graph to compare motion speed for four objects in

horizontal, direction with simply different speeds.

The graph in Figure 11 resulted from the situation where four objects were moving in horizontal directions with different velocities without having acceleration.

In the motion test, the objects which are moving vertical across the display become more salient. When all objects have the same velocity, those circles which move vertically will pass the width of display quickly than other objects. Therefore, they may be fixated on more than the other objects.

In the scenario in which moving circles have different velocity/acceleration, objects become salient candidates when they have different motion compared to others i.e. moving faster or slower than other objects or having abrupt acceleration specially while changing their direction. Also, objects moving in different directions from the majority of other objects will absorb more attention.

According to our results, the consistency among participants in motion tests is very high. We conclude that motion is a very important stimulus for the HVS compared to other stimuli such as color, brightness, texture, and depth. In the images with only color and/or texture attributes, the inter-subject's consistency is not high enough and we conclude that in the still images the salient area is more subjective. However, in video sequences involving motion, the salient areas are more predictable.

V. Conclusion and Discussion

In this paper, we proposed an experimental study devoid of cognitive bias to formulize the saliency within a scene. For this purpose, we investigated bottom-up attributes such as color, texture, motion direction, and motion speed as stimuli for the human visual system. We designed this experiment to understand the order of importance among bottom-up attributes in absorbing human attention while observing a scene.

According to our results, warm colors such as red and pink, and bright colors such as yellow and cyan are more salient. Textures with dense and compact are more outstanding. In addition, textures with obvious high contrast within their pattern are likely to be salient. Vertical movements are more likely to be fixated on and the motions with very high or very low speed or any unique acceleration are more salient for subjects.

We concluded that motion speed and motion direction are the most important factors in guiding human attention toward specific objects or areas in the video dataset. Color contrast is more important than color and texture stimuli.

In our future work, we will consider the combination of these attributes to understand human vision behavior in the presence of different permutations of these attributes. Considering combination of these features will lead us to obtain more comprehensive and reliable results.

Acknowledgement

The authors would like to acknowledge that this research was supported by the NSERC Strategic Project Grant: "Hi-Fit: High Fidelity Telepresence over Best Effort Networks.'

References

[1] W. Osberger, Perceptual Vision Models for Picture Quality Assessment and Compression Applications, Ph.D. thesis, Queensland University of Technology, Brisbane, Australia, 1999.

[2] C. G. Healey, and J. T. Enns," Attention and Visual Memory in Visualization and Computer Graphics," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 18, Issue. 7, pp. 1170-1188, 2012.

[3] E. D. Gelasca, D. Tomasic, T. Ebrahimi," Which Colors Best Catch Your Eyes: A Subjective Study of Color Saliency," *SPIE International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2005.

[4] K. Duncan and S. Sarkar, "Saliency in Images and Videos: a Brief Survey," *IET Computer Vision*, Vol. 6, pp. 514-523, 2012.

[5] A. Banitalebi-Dehkordi, E. Nasiopoulos, M. T. Pourazad, and P. Nasiopoulos, "Benchmark Three-dimensional Eye Tracking Dataset for Visual Saliency Prediction on Stereoscopic Three-dimensional Video," *SPIE Journal of Electronic Imaging*, Vol. 25, Issue 1, 2016.

[6] P. Correia and F. Pereira," Video Object Relevance Metrics for Overall Segmentation Quality Evaluation," *EUSIPCO Conference on Estimation of Video Objects Relevance*, pp. 1-11, 2000.

[7] W. Osberger and A.M. Rohaly, "Automatic Detection of Regions of Interest in Complex Video Sequences," *in Proceedings of Human Vision and Electronic Imaging*, Vol. 6, pp. 361-372, 2001.

[8] F. Birren, Le Pouvoir, de la Couleur, Les Editions de Homme, 1998.

[9] M. Dick, S. Ullman, and D. Sagi, "Parallel and Serial Processes in Motion Detection," *Science*, Vol. 237, pp.400-402, 1987.

[10] R. B. Ivry, "Asymmetry in Visual Search for Targets Defined by Differences in Movement Speed," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 18, No. 4, pp. 1045-1057, 1992.

[11] K. F. Van Orden, J. Divita, and M. J. Shim," Redundant Use of Luminance and Flashing with Shape and Color as Highlighting Codes in Symbolic Displays," *Human Factors*, Vol. 35, No. 2, pp. 195-204, 1993.

[12] W. Osbergera, A. Maederb, and N. Bergmanna, "A Technique for Image Quality Assessment Based on a Human Visual System Model," *Signal Processing Conference*, 1998.

[13] N. H. Mackworth and A. J. Morandi, "The Gaze Selects Informative Details Within Pictures," *Perception and Psychophysics*, Vol. 2, No. 11, pp. 547-552, 1967.

[14] A. L. Yarbus, "Eye Movements and Vision," *Springer*, Plenum Press, New York, 1967.

[15] S. E. Jenkins and B. L. Cole, "The Effect of the Density of Background Elements on the Conspicuity of Objects," *Vision Research*, Vol. 22, pp. 1241-1252, 1982.

[16] D. Navon, "Forest Before Trees: The Precedence of Global Features in Visual Perception," *Cognitive Psychology*, pp. 353-383, 1977.

[17] D. Khaustova, J. Fournier, E. Wyckens, "An Investigation of Visual Selection Priority of Objects with Texture and Crossed and Uncrossed Disparities," *SPIE Human Vision and Electronic Imaging*, 2014.

[18] J. Hkkinen, T. Kawaid, J. Takataloc, R. Mitsuyad, and G. Nymanc, "What Do People Look at When They Watch Stereoscopic Movies?," *SPIE Stereoscopic Displays and Applications*, 2010.

[19] D. Khaustova, J. Fournier, and E. Wyckens, "How Visual Attention is Modified by Disparities and Textures Changes?," *SPIE HVEI*, 2013.

[20] http://www.color-blindness.com/ishihara-38-plates-cvd-test/.

[21] Recommendation ITU-R BT.2022, General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays, 2012.