# Modelling of Human Visual Attention

Patrik Polatsek*
Institute of Computer Engineering and Applied Informatics
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava
Ilkovičova 2, 842 16 Bratislava, Slovakia
patrik.polatsek@stuba.sk

## Abstract

In recent decades, visual attention modelling became a prominent research area. In order to simulate human attention, a computational model has to incorporate various stimulus-driven and goal-directed attention mechanisms. This work explores how low- and mid-level features such as color, motion, depth and shape influence visual attention in our own eye-tracking experiments. To measure these effects, we utilized various state-of-the-art as well as novel computational models which estimate saliency of a specific feature. In order to deeper understand the process of selective attention in everyday actions, we conducted several experiments in real environments recorded from the first-person perspective. Our results showed that egocentric attention is very individual and differs from 2D image viewing conditions, partially due to binocular cues that enhance viewer's perception. We therefore suggest to employ specialized models for egocentric vision. Finally, we found out that high-level factors such as individual's emotions and task-based analysis of visualizations influence human gaze behavior too.

## Categories and Subject Descriptors

I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding—*Perceptual reasoning*; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Color, Depth cues, Motion, Shape*

## Keywords

visual attention, saliency model, egocentric attention

## 1. Introduction

*Visual attention* is a set of cognitive processes that selects relevant information and filters out irrelevant information from the environment [4]. Therefore, it plays an important role in the control of head and eye movements. Scene scan is performed by a sequence of rapid movements called *saccades* and *fixations*. During a fixation, the eye is relatively still to acquire visual information from the focus of interest [32, 17].

Attention is influenced by both *bottom-up* factors such as salient stimuli and *top-down* factors including individual's goals and prior knowledge. Psychologists assume that bottom-up and top-down processes work together to organize and interpret visual information from the environment. This is referred to as *perception* [8].

Visual saliency has been researched in many research areas including psychology, neurobiology, image processing and computer vision [5]. In general, there are two different approaches how to define saliency of an image [29]. We can *measure saliency*, e.g. using eye-trackers or we can *predict saliency* by computational saliency models. They compute a *saliency map* from an image, representing a topographical map of conspicuousness [5].

The primary goal of this work is to individually study various aspects of visual attention using novel eye-tracking experiments and computational saliency models. Since the majority of these factors has been explored marginally so far, we recorded fixation data in image viewing conditions and real environments to deeper understand human visual system and increase the performance of saliency models.

## 2. Related Work

Recent decades of visual attention research have brought many computational models that can be grouped by various criteria, including [5, 26]:

- **factors influencing attention:** *bottom-up* factors and *top-down* factors,
- **temporality:** *spatial* models based solely on a current scene and *temporal* models based on the accumulated prior knowledge or motion analysis,
- **stimuli type:** *static* stimuli such as intensity, color and depth and *dynamic* stimuli such as motion and flicker,
- **task type:** *free viewing*, *visual search tasks* and other more *complex tasks* (e.g. driving),

*Recommended by thesis supervisor: Assoc. Prof. Vanda Benešová
Defended at Faculty of Informatics and Information Technologies, Slovak University of Technology in Bratislava on [to be specified].

© Copyright 2019. All rights reserved. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from STU Press, Vazovova 5, 811 07 Bratislava, Slovakia.

- **saliency units:** *location-based* models that assign saliency values to each location defined by pixels and macro-blocks and *object-based* models that either extract salient objects from location-based saliency or directly compute saliency at object level,

- **size of information** used in saliency estimation: *local* information based only on a subregion of an image, *global* information based on a whole image.

## 3. Egocentric Motion Saliency Modelling

Since egocentric saliency has not been widely explored so far, we evaluated own spatiotemporal superpixel-based saliency model on a natural shopping task recorded by eye-tracking glasses.

Our model uses a superpixel segmentation [1] to at least partially implement object-based attention. Each superpixel is described by static (intensity, color and orientation) and dynamic (motion) features in multiple scales. Since human gaze is also directed to unexpected, surprising stimuli, saliency estimation includes motion surprise too.

Each video frame is decomposed into intensity, red, green, blue and yellow colors and orientation of gradients by applying Sobel filter. Motion between consecutive frames is calculated by an optical flow algorithm [10]. The distribution of each feature within a superpixel is represented by a histogram.

To follow the multi-scale approach of Itti et al. [22], we generate a Gaussian pyramid, but we employ superpixels instead of pixels. We compare histograms on finer and coarser pyramid scales to compute spatial and temporal saliency. To estimate motion surprise, we compare prior knowledge about motion field with the actual frame at a given location.

The performance of our model has been compared with the spatial location-based model by Itti et al. [22] and the spatiotemporal superpixel-based model by Liu et al. [28] (see examples in Figure 1) using AUC[1] and NSS[2] scores. The evaluation dataset contains gaze data of two participants at a shopping mall who were asked to find specific products.

We found out that static saliency predominates over motion saliency despite of multiple moving objects in subjects' views. Varying performance of computational models (Table 1) indicates that both types of saliency do not affect their attention equally and thereby it could be also guided by depth information and top-down features, e.g. object detection and detection of biological motion. In addition to these factors, egocentric saliency modelling should incorporate the effect of surprise from static as well as dynamic stimuli.

[1]A saliency map is treated as a binary classifier. Saliency at fixations and some non-fixated pixels are extracted. Fixations with saliency above a gradually increasing threshold and non-fixations above the threshold are considered as TPs and FPs, respectively. The ROC is plotted and the area under the curve (AUC) is computed [9].

[2]The Normalized Scanpath Saliency (NSS) score equals to the average saliency at fixation locations in a saliency map normalized to have a zero mean and a unit standard deviation [9].

**Figure 1: Saliency maps estimated by our model (top right), Itti et al. [22] (bottom left) and Liu et al. [28] (bottom right). Fixation is labelled with a red circle.**

**Table 1: Individual AUC/NSS scores.**

| Subject | Our | [22] | [28] |
|---|---|---|---|
| #1 | .661/0.59 | .595/0.35 | .649/0.59 |
| #2 | .749/0.87 | .756/0.91 | .742/0.94 |

## 4. Visual Attention to Color

Color is the fundamental component of visual attention. Saliency is usually associated with color contrasts. Beside this bottom-up perspective, some recent works indicate that psychological aspects should be considered too [14]. However, relatively little research has been done on potential impacts of color psychology on attention. To our best knowledge, a publicly available fixation dataset specialized on color feature does not exist. We therefore conducted a novel eye-tracking experiment with color stimuli and made it publicly available. We studied whether color differences can reliably model color saliency or particular colors are preferably fixated.

In our experiment we showed simple colored objects on a uniform background to 15 participants. We used colors such as red, green, blue, yellow, cyan, magenta, pink and orange (Figure 2).

Fixation data showed a strong correlation with color contrasts in the LAB color space across displayed images. However, individual correlations across participants revealed that attention of some participants was influenced by color contrasts only negligibly (Figure 3). Furthermore, we found only slightly higher fixations of red and yellow colors associated to danger and warning. On the other hand, cyan objects were largely ignored with a remarkable gap in fixations (Figure 4).

## 5. Visual Attention to Shape

Beside simple feature saliency, such as intensity, color and orientation, attention is influenced by object shape and

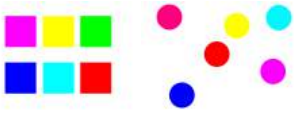

**Figure 2: Stimuli in the color experiment.**

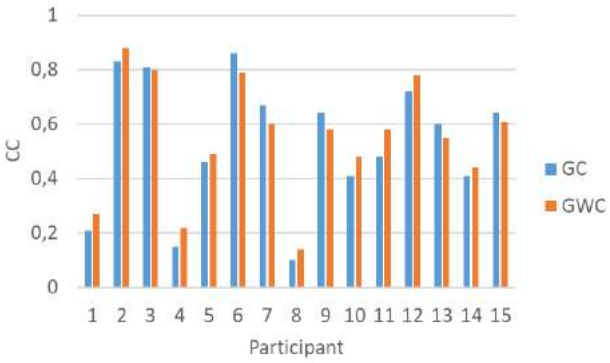


**Figure 3: Individual average correlations between fixations and color contrasts (GC) and spatially weighted color contrasts (GWC).**

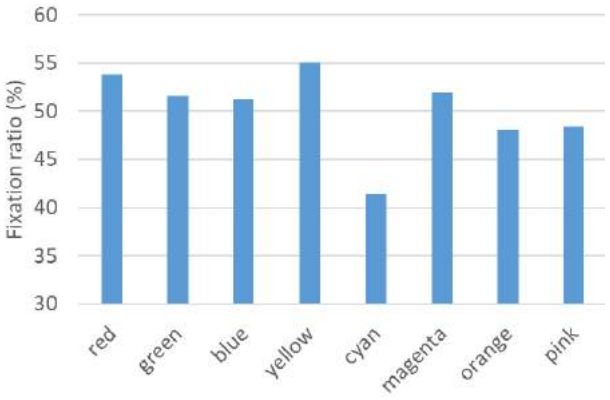


**Figure 4: Fixation ratio of each color.**

size too. We therefore explored whether and to what extent global and local shape characteristics and their mutual differences affect visual saliency. To our best knowledge, an eye-tracking dataset focused solely on shape has not been available so far. Therefore, we created such a dataset and made it publicly available.

We showed silhouettes of abstract shapes and real-world objects on a uniform background to 73 students in our experiment. Each scene contains either 12 shapes organized in a circle or 2 shapes on both image sides (Figure 5).

To investigate shape saliency, we proposed three groups of computational models. They employ shape descriptors and matchers to detect salient shapes (object-based models) or salient boundary contours (location-based models). The first group of models estimates object saliency by global geometrical properties ignoring the spatial context. Higher saliency values are assigned to larger, asymmetrical, irregular and highly curved objects represented by area size, perimeter length, equivalent diameter, eccentricity, aspect ratio, extent, rectangularity, solidity and

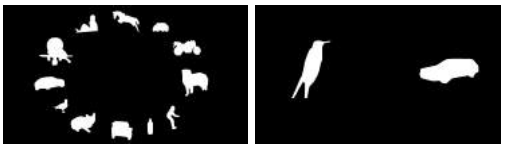

**Figure 5: Stimuli in the shape experiment.**

| Objects | A | ED | P |
|---|---|---|---|
| *2* | 0.39 | 0.40 | 0.31 |
| *12* | 0.29 | 0.30 | 0.54 |

**Table 2: Correlation between fixations and shape properties such as area size (A), equivalent diameter (ED) and perimeter length (P). All correlations are significant ($p < .001$).**

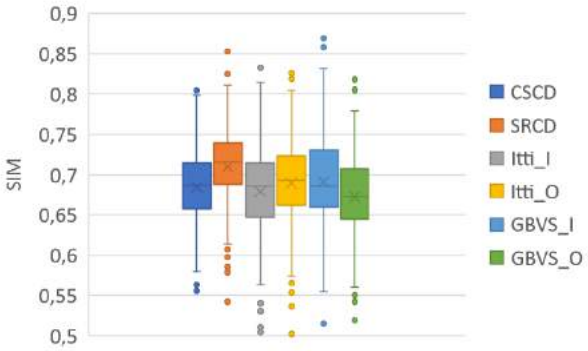


**Figure 6: SIM scores of location-based saliency models – centroid distance in the spatial (CSCD) and the frequency domain (SRCD), intensity and orientation saliency by Itti et al. [22] (Itti_I, Itti_O) and Harel et al. [18] (GBVS_I, GBVS_O).**

circularity. The second group of models assigns higher saliency values to shapes that are different from the other ones. The unique shapes are defined by the simple characteristics employed in the first group of models, Hausdorff distance, shape context, centroid distance in the spatial and the frequency domain and boundary moments [37, 3]. In contrast, the third group estimates contour saliency. These models consider contours differ from their surroundings as salient. The first contour model denoted **CSCD** follows the center-surround approach [22]. It therefore builds a Gaussian pyramid from the centroid distance signature and compares finer and courser pyramid levels. The second contour model denoted **SRCD** applies a Fourier transform on the centroid distance. It is based on a work of Hou and Zhang [21] that introduced the spectral residual approach to create a saliency map. Beside own models, we evaluated the shape saliency model based on the Jaccard index [11] and two standard models [22, 18] that compute intensity and color saliency which could participate in shape perception too.

We found out that attention is directed to larger objects (see significant positive correlations in Table 2), but we did not observe a clear trend for fixating asymmetrical and complex objects. The results also showed that saliency from global shape contrast only slightly affects attention. Finally, our analysis revealed a significant effect of contour saliency. Comparing contour models using SIM metric[3] as visualized in Figure 6, our **SRCD** significantly outperforms other location-based models (see example in Figure 7). In addition, human-like figures seem to be often more salient than silhouettes of non-living objects, particularly human and animal heads.

[3]The similarity metric (SIM) is defined as $\sum_x (S(x), F(x))$, where saliency $S$ and continuous fixation map $F$ are normalized to the sum of 1 [9].

**Figure 7: Fixation heatmap and saliency map estimated by SRCD.**

## 6. Visual Attention to Egocentric Depth

Computational models usually do not employ depth information to estimate saliency. However, depth is an important aspect of visual attention in egocentric vision. In contrast to previous studies that investigated depth on 2D and 3D images, our experiments whose fixation data are publicly available took place in a natural environment. Since binocular vision enhances depth discrimination [2, 31], we explored in our eye-tracking experiment whether saliency of depth in natural 3D environment differs from pictorial depth.

We recorded fixations of 28 students in a room with identical balls arranged in an octagonal layout whose distance from an observer varied, up to 4 m (6 depth planes with a step of 30 cm). Participants were shown 5 scene types that differ in depth level steps between adjacent objects – the only one step of 1; alternating zero steps and steps of 2; only steps of 1; steps of 1, but one step of 3 and steps varying from 0 up to 4 (Figure 8). We represented each object by the relative depth and the global depth contrast.

We cannot confirm the finding of experiments with 2D and 3D images [23, 25, 35] which concluded that fixations are biased towards areas close to a viewer. Surprisingly, our results indicate that attention is more strongly directed towards distant objects in a real environment, but this relationship between saliency and depth is non-linear (Figure 9). Despite of this bias we believe there is a threshold distance when object saliency starts to decrease since their size is relatively too small to grab attention. Furthermore, high-contrast objects in depth channel are salient even though this effect is also not linear (Figure 10).

Therefore, we conclude that human gaze behavior in real environments differs from stereoscopic image displays and standard 3D saliency models incorporating depth channel are not suitable for egocentric video sequences.

**Figure 8: Example scene in the depth experiment.**

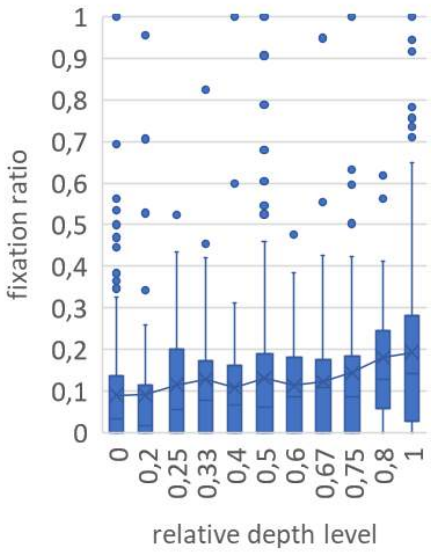


**Figure 9: Distribution of participants' fixations on depths, ranging from 0 (closest) to 1 (farthest).**

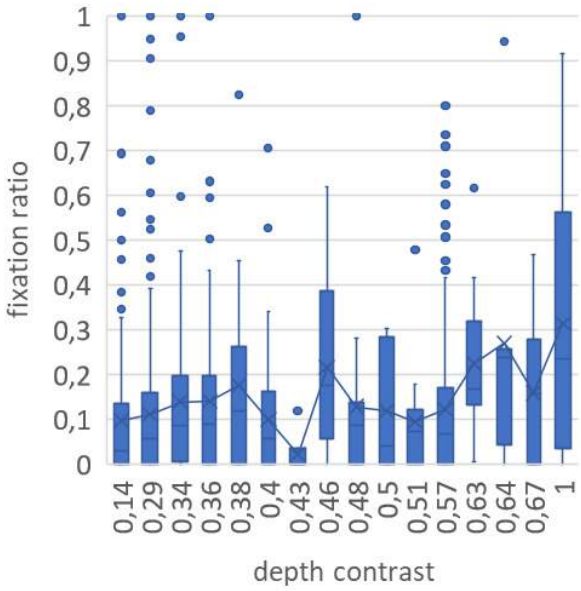


**Figure 10: Distribution of participants' fixations on depth contrasts, normalized between 0 and 1.**

## 7. Static Feature-Based Egocentric Visual Attention

Our experiments explored the effects of static features on attention separately (Section 4, 5 and 6). The aim of this section is to find out how static stimuli compete for our attention in everyday actions. In contrast to prior works, we analyzed scene depth too.

Our experiments employed eye-tracking glasses to record users' gaze and Kinect device to capture depth of objects. 6 students were asked to freely walk and explore a laboratory room which contains only static stimuli for 15 up to 30 sec.

We analyzed low-level features such as intensity, color, orientation and depth, mid-level features such as object shape and contours and the center bias. We predicted their saliency effects by conventional [22, 18] and novel computational models that decompose overall attention to separate feature saliency maps:

1. Intensity (I), color (C) and orientation (O):
   (a) Itti et al's model [22] (denoted **Itti** with suffixes **_I**, **_C** and **_O**, respectively),
   (b) Harel et al's model [18] (denoted **GBVS** with suffixes **_I**, **_C** and **_O**, respectively),
   (c) own superpixel model which correlates superpixel histograms at finer and coarser pyramid layers for intensity and orientation saliency (the highest saliency is estimated for uncorrelated superpixels) and compares their color distances for color saliency (denoted **SPX** with suffixes **_I**, **_C** and **_O**, respectively; Section 3).
2. Depth (D):
   (a) simple model which linearly increases saliency with shorter object distance to a viewer so that the closest objects are most salient (denoted **D_lin**),
   (b) experimental model based on our depth experiment (denoted **D_nlin**; Section 6),
   (c) simple contrast model which defines saliency linearly as global contrast of superpixels[4] (denoted **DC_lin**),
   (d) experimental contrast model which weights the contrast of superpixels[4] based on our depth experiment (denoted **DC_nlin**; Section 6),
   (e) our superpixel model which correlates superpixel histograms of depth as in **SPX_I** (denoted **SPX_D**; Section 3).
3. Shape (S):
   (a) perimeter model that assigns higher saliency to larger regions defined by perimeter length (denoted **S_p_intra**; Section 5),
   (b) perimeter model based on region contrasts[4] (denoted **S_p_inter**; Section 5),
   (c) equivalent diameter model that assigns higher saliency to larger regions defined by equivalent diameter (denoted **S_e_intra**; Section 5),
   (d) equivalent diameter model which measures region contrasts[4] (denoted **S_e_inter**; Section 5),
   (e) CSCD model based on the centroid signature in the spatial domain (denoted **S_CSCD**; Section 5),
   (f) SRCD model based on the centroid signature in the frequency domain (denoted **S_SRCD**; Section 5).
4. Center-bias modelled by Gaussian at the image center (denoted **Center**).

[4]We defined global region contrast as $S(r_i) = \sum_{i \neq j} D_d(r_i, r_j) exp(-D_s(r_i, r_j))$, where $r_i$ denotes the i-th region, $D_d$ is the distance between average region depths and $D_s$ is the normalized Euclidean distance between region centroids.

Comparing NSS scores[2] listed in Table 3 revealed the strongest influence of intensity, color and orientation contrasts on egocentric attention. Besides, we observed significantly different gaze behavior among participants. This could be explained by top-down guidance of attention that we ignored in our analysis such as object identification and surprising stimuli over time.

Intensity saliency is best modeled by our **SPX_I**, whereas color and orientation saliency were predicted with highest accuracy by **GBVS** [18]. On the hand, object distances and global depth contrasts had a much lower effect on egocentric attention. However, depth-weighting approach based on our shape experiment in Section 5 (**D_nlin**) significantly improved saliency estimation of standard depth weighting (**D_lin**) for 4 subjects. All these depth models are outperformed by local contrast **SPX_D** model. Shape saliency was successfully estimated only by contour models – **S_CSCD** and **S_SRCD**. Our experiment also confirmed a strong center bias of egocentric gaze behavior.

The results also indicate that the above mentioned effects did not remain constant over time. Simple low-level features seem to affect attention more rapidly, whereas complex features such as shape have a delayed effect.

## 8. Emotionally-Tuned Visual Attention

While psychological studies [36] have confirmed a connection between emotional stimuli and visual attention, there is a lack of evidence, how much influence individual's mood has on visual information processing of emotionally neutral stimuli. In contrast to prior studies, we explored if bottom-up low-level saliency could be affected by positive mood. While recent saliency models aimed

**Table 3: Individual NSS scores (the highest value is bold; the second and the third highest values are underlined).**

| Model | #1 | #2 | #3 | #4 | #5 | #6 |
|---|---|---|---|---|---|---|
| Itti_I | 0.97 | 1.52 | 1.79 | 1.58 | 1.68 | 1.77 |
| GBVS_I | 0.99 | 1.65 | 1.80 | 1.62 | 1.61 | 1.66 |
| SPX_I | 1.22 | 1.85 | 1.78 | **1.71** | 1.82 | 1.69 |
| Itti_C | 0.99 | 1.11 | 1.38 | 1.11 | 1.69 | 1.33 |
| GBVS_C | 1.09 | 1.39 | 1.27 | 1.48 | 1.85 | 1.56 |
| SPX_C | 0.73 | 0.90 | 1.20 | 1.48 | 1.73 | 1.59 |
| Itti_O | 1.11 | 1.87 | 1.65 | 1.66 | 1.85 | 1.79 |
| GBVS_O | 1.20 | **2.00** | **2.02** | 1.70 | **2.04** | **1.95** |
| SPX_O | 0.94 | 1.10 | 1.27 | 1.19 | 1.14 | 1.51 |
| D_lin | 0.14 | -0.10 | 0.29 | 0.50 | -0.23 | 0.09 |
| D_nlin | 0.45 | 0.66 | 0.28 | -0.03 | 0.88 | 0.52 |
| DC_lin | 0.49 | 0.30 | -0.19 | -0.28 | 0.64 | -0.06 |
| DC_nlin | 0.60 | 0.16 | -0.29 | -0.33 | 0.56 | -0.10 |
| SPX_D | **1.27** | 1.12 | 1.17 | 1.27 | 1.18 | 1.23 |
| S_p_intra | 0.26 | -0.35 | -0.51 | -0.21 | -0.50 | -0.44 |
| S_p_inter | 0.30 | -0.33 | -0.48 | -0.15 | -0.45 | -0.41 |
| S_e_intra | 0.14 | -0.42 | -0.58 | -0.36 | -0.60 | -0.54 |
| S_e_inter | 0.18 | -0.40 | -0.55 | -0.29 | -0.56 | -0.51 |
| S_CSCD | 0.79 | 0.91 | 0.89 | 0.85 | 0.99 | 0.80 |
| S_SRCD | 0.86 | 1.22 | 1.18 | 1.06 | 1.18 | 1.20 |
| Center | 0.98 | 1.25 | 0.99 | 1.08 | 1.29 | 1.03 |

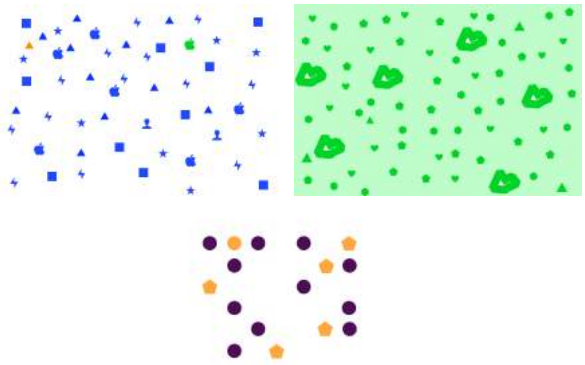

**Figure 11: Visual search tasks. FO-task: Find an orange triangle (left). FOA-task: Find a triangle (right). FU-task: Find a unique object (target is the orange circle; bottom).**

to predict attention for affective stimuli [27, 13], conventional saliency models have been never evaluated under a particular emotional state of observers.

Our study examined attention of 10 participants who were randomly induced to experience either a positive mood or a neutral mood by recalling own personal memories. Therefore, they were asked to recall either a happy event or the route they took to an experimental room. After the mood induction, participants were shown natural and synthetic images with valence-neutral stimuli. They were instructed to freely explore the images (**V**), memorize their contents (**M**) or find a target object as quickly as possible. They searched for a single target among non-targets based on specific criteria (**FO**), any of objects that meets the criteria (**FOA**) or a unique object without explicit target description (**FU**), as shown in Figure 11. Participants assessed their mood in the I-PANAS-SF [24].

Though the mood induction affected only slightly how participants assessed their affective state, we found some differences between both groups of subjects. We assumed that experiences of positive emotions broaden individual's attention, as proposed by the broaden-and-built theory [15]. However, we found that happy memories could even distract individuals from visual search. While we did not observe significantly similar fixation patterns for participants within as well as between induced emotion (Figure 12), the results suggest the interaction between induced mood and bottom-up saliency [22]. Comparing NSS scores[2], we found that attention bias towards bottom-up features varies across task type (Figure 13). Positive conditions result in stronger saliency when freely exploring images (**V**). However, the saliency effect is suppressed when solving **FU** and **M**, compared to neutral conditions. We therefore speculate that broadening attention in terms of bottom-up processing might be associated with a low level of engagement in the task.

## 9. Visual Attention during Task-Based Analysis of Information Visualization

The way users observe a visualization is affected by salient stimuli in a scene as well as by domain knowledge, interest, and tasks. While recent saliency models manage to predict users' visual attention in visualizations during exploratory analysis, there is little evidence how much in-

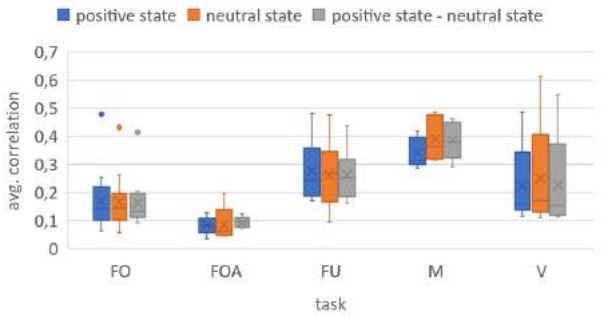


**Figure 12: Fixation similarities of subjects with the same emotion and between different emotions.**

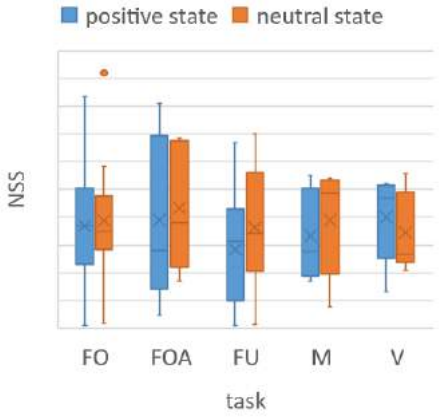


**Figure 13: NSS scores of bottom-up saliency model [22].**

fluence bottom-up saliency has on task-based visual analysis. In contrast to previous studies, we conducted an eye-tracking study whose aim is to determine user's gaze behavior in visualizations when solving three low-level analytical tasks and made it publicly available.

We analyzed attention of 47 students who were instructed to solve data visualization tasks as quickly as possible. Subjects were shown visualizations from the MASSVIS database such as bar charts, maps, area charts, point charts, tables and line charts that were originally used in the memorability experiment [6]. We designed three low-level analytical tasks for each chart – retrieve value of a specific data element (**RV**), filter data elements based on specific criteria (**F**) and find an extremum attribute value within a dataset (**FE**). To analyze visual performance, we defined task-dependent AOIs that need to be attended to correctly answer the question. We listed their optimal viewing strategy in Table 4 (see examples in Figure 14). We compared participants' fixations from this confirmatory (task-based) analysis to the memorability experiment (**Mem**) [6] with conditions closer to free exploration.

To estimate bottom-up saliency, we generated saliency maps from 12 saliency algorithms including convolutional neural models and **DVS** model [30] which combines Itti et al.'s saliency [22] with text saliency, and could thereby increase the performance for information visualizations significantly.


**Figure 14: Target-dependent AOIs of sub-parts of visualizations. Red, green and blue outlines define the target data points, their item labels and value labels respectively. RV-task: What is the attendance of Universal Studios Hollywood? (left) F-task: Which German states have an unemployment rate of more than 12%? (middle) FE-task: In which country do people anticipate to spend the least money for personal Christmas gifts? (right)**

**Table 4: Optimal viewing order of task-dependent AOIs.**

| Task | Step 1 | Step 2 | Step 3 |
|---|---|---|---|
| *RV* | search item label | map to the item | read the value label |
| *F* | search value label(s) | map to the item(s) | read the item label(s) |
| *FE* | search value label(s) | map to the item(s) | read the item label(s) |
| | search item(s) | | read the item label(s) |

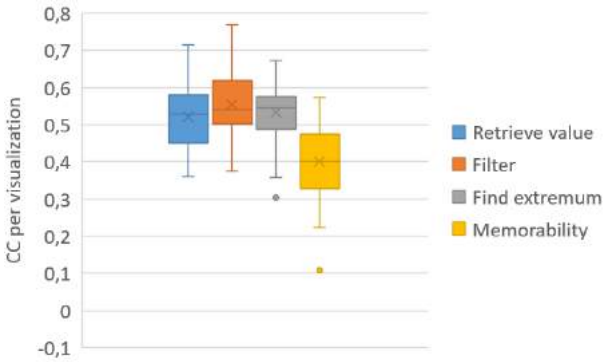


**Figure 15: Similarity between fixations of the same type of activity.**

We found out that fixation patterns of users solving the same analytical task are more coherent than for **Mem** (Figure 15). However, first fixation times of task-related AOIs revealed that subjects used the optimal viewing strategy in Table 4 only for **RV**. Analytical tasks therefore seem to have a measurable top-down guidance for the users where to look, but not necessarily in which order. Furthermore, comparing fixations of different types of activity surprisingly showed that **Mem** much closer resembles **FE** than other tasks (Figure 16). A possible explanation is that users were intentionally seeking for extrema as representative values to memorize the content of the visualization.

We report the average AUC scores[1] of saliency models in Table 5. In contrast to **Itti** [22], the performance of **DVS** [30] confirmed that bottom-up saliency strongly in-

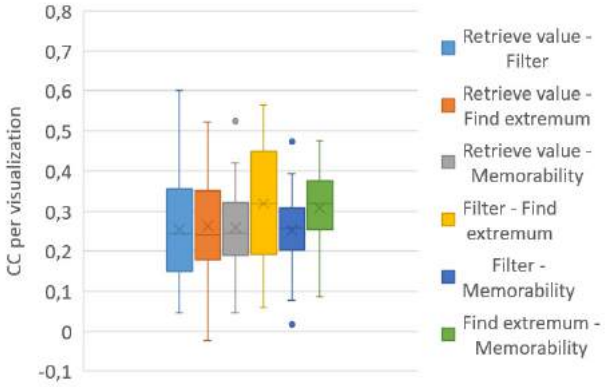


**Figure 16: Similarity between fixations of different types of activity.**

fluences fixations of users when freely exploring the visualization, but has a significantly lower effect on visual attention when performing a low-level analytical task. A potential explanation for this significantly worse performance could be that users direct their attention more towards the data areas than the text areas when performing low-level analytical tasks, than when trying to memorize the visualization. Finally, targets with extreme values in **FE** are neither more efficiently searched nor more salient in saliency maps [22] than targets of the other two tasks.

To improve existing saliency models and tailor them more towards task-based visual analysis, we therefore recommend to merge classic image-based saliency. The model should localize and identify visualization elements, compare their features and estimate their relationships.

## 10. Conclusions

This thesis explored various bottom-up and top-down factors of visual attention. We performed novel eye-tracking experiments, proposed computational saliency models and discussed human gaze behaviour. Visual attention modelling needs to have specialized fixation datasets which could improve saliency prediction. Therefore, we made our fixation databases available to the public.

Our experiments examined attentional factors separately. Most of them showed a high diversity of their effects on visual attention and visual performance, particularly in nat-

**Table 5: The average AUC scores for each task. We evaluated 12 saliency models, denoted Itti [22] (implementation by Harel [18]), AIM [7], GBVS [18], SUN [39], CAS [16], Sign [20], BMS [38], eDN [34], SAMv and SAMr [12] (feature maps extracted by the convolutional neural model based on VGG-16 [33] and ResNet-50 [19], respectively), DVS [30] (with the optimal weight of text saliency for MASSVIS database) and TextS [30] (text saliency of the DVS model separately).**

| Task | Itti | AIM | GBVS | SUN | CAS | Sign | BMS | eDN | SAMv | SAMr | TextS | DVS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *RV* | .684 | .646 | .608 | .593 | .595 | .576 | .621 | .596 | .630 | .632 | .647 | **.702** |
| *F* | .690 | .645 | .642 | .593 | .604 | .622 | .651 | .595 | .618 | .631 | .624 | **.692** |
| *FE* | .679 | .654 | .599 | .602 | .601 | .600 | .638 | .568 | .637 | .647 | .651 | **.705** |
| *Mem* | .686 | .675 | .553 | .622 | .637 | .589 | .652 | .554 | .653 | .664 | .696 | **.738** |

ural environments from the first-person perspective which significantly differ from image viewing conditions.

In future work, the above mentioned experimental findings should be merged to create a computational model that could reliably predict attention in natural scenes and specialized domains, such as information visualizations or medical imaging. Because of the individuality in visual information processing, a possible solution could be to learn fixation preference from fixation data of a particular user solving a particular task using deep neural networks.

**Acknowledgements**. This work was supported by the Slovak Scientific Grant Agency, VEGA 1/0874/17, and from the Research and Development Operational Programme for the project "University Science Park of STU Bratislava", ITMS 26240220084, co-funded by the European Regional Development Fund.

## References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, et al. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.

[2] R. Allison, B. Gillam, and E. Vecellio. Binocular depth discrimination and estimation beyond interaction space. *Journal of Vision*, 7(9):817–817, 2007.

[3] S. Belongie, J. Malik, and J. Puzicha. Matching shapes. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 454–461. IEEE, 2001.

[4] M. D. Binder, N. Hirokawa, and U. Windhorst. *Encyclopedia of neuroscience*. Springer Berlin, Germany, 2009.

[5] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):185–207, 2013.

[6] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2306–2315, 2013.

[7] N. Bruce and J. Tsotsos. Saliency based on information maximization. In *Advances in neural information processing systems*, pages 155–162, 2006.

[8] C. Bundesen and T. Habekost. *Principles of Visual Attention: Linking Mind and Brain*. Oxford University Press Oxford, 2008.

[9] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand. What do different evaluation metrics tell us about saliency models? *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[10] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011.

[11] Y. Chen, C.-P. Yu, and G. Zelinsky. Adding shape to saliency: A proto-object saliency map for predicting fixations during scene viewing. *Journal of Vision*, 16(12):1309–1309, 2016.

[12] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara. Predicting human eye fixations via an lstm-based saliency attentive model. *CoRR*, abs/1611.09571, 2016.

[13] X. Ding, L. Huang, B. Li, C. Lang, Z. Hua, and Y. Wang. A novel emotional saliency map to model emotional attention mechanism. In *International Conference on Multimedia Modeling*, pages 197–206. Springer, 2016.

[14] A. J. Elliot. Color and psychological functioning: a review of theoretical and empirical work. *Frontiers in Psychology*, 6:368, 2015.

[15] B. L. Fredrickson. The broaden-and-build theory of positive emotions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1449):1367, 2004.

[16] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. *IEEE transactions on pattern analysis and machine intelligence*, 34(10):1915–1926, 2012.

[17] E. B. Goldstein and J. Brockmole. *Sensation and perception*. Cengage Learning, 2016.

[18] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *Advances in neural information processing systems*, pages 545–552, 2007.

[19] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[20] X. Hou, J. Harel, and C. Koch. Image signature: Highlighting sparse salient regions. *IEEE transactions on pattern analysis and machine intelligence*, 34(1):194–201, 2012.

[21] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[22] L. Itti, C. Koch, and E. 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.

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

[24] J. Karim, R. Weisz, and S. U. Rehman. International positive and negative affect schedule short-form (i-panas-sf): Testing for factorial invariance across cultures. *Procedia-Social and Behavioral Sciences*, 15:2016–2022, 2011.

[25] C. Lang, T. V. Nguyen, H. Katti, K. Yadati, M. Kankanhalli, and S. Yan. Depth matters: Influence of depth cues on visual saliency. In *Computer vision–ECCV 2012*, pages 101–115. Springer, 2012.

[26] J. Li and W. Gao. *Visual saliency computation: A machine learning perspective*, volume 8408. Springer, 2014.

[27] H. Liu, M. Xu, J. Wang, T. Rao, and I. Burnett. Improving visual saliency computing with emotion intensity. *IEEE transactions on neural networks and learning systems*, 27(6):1201–1213, 2016.

[28] Z. Liu, X. Zhang, S. Luo, and O. Le Meur. Superpixel-based spatiotemporal saliency detection. *IEEE transactions on circuits and systems for video technology*, 24(9):1522–1540, 2014.

[29] M. Mancas. *Computational attention towards attentive computers*. Presses univ. de Louvain, 2007.

[30] L. E. Matzen, M. J. Haass, K. M. Divis, Z. Wang, and A. T. Wilson. Data visualization saliency model: A tool for evaluating abstract data visualizations. *IEEE transactions on visualization and computer graphics*, 24(1):563–573, 2018.

[31] S. Palmisano, B. Gillam, D. G. Govan, R. S. Allison, and J. M. Harris. Stereoscopic perception of real depths at large distances. *Journal of vision*, 10(6):19–19, 2010.

[32] A. C. Schütz, D. I. Braun, and K. R. Gegenfurtner. Eye movements and perception: A selective review. *Journal of vision*, 11(5):9, 2011.

[33] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[34] E. Vig, M. Dorr, and D. Cox. Large-scale optimization of hierarchical features for saliency prediction in natural images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2798–2805, 2014.

[35] J. Wang, P. Le Callet, S. Tourancheau, V. Ricordel, and M. P. Da Silva. Study of depth bias of observers in free viewing of still stereoscopic synthetic stimuli. *Journal of Eye Movement Research*, 5(5), 2012.

[36] J. R. Zadra and G. L. Clore. Emotion and perception: The role of affective information. *Wiley interdisciplinary reviews: cognitive science*, 2(6):676–685, 2011.

[37] D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern recognition*, 37(1):1–19, 2004.

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

[39] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell. Sun: A bayesian framework for saliency using natural statistics. *Journal of vision*, 8(7):32–32, 2008.

## Selected Papers by the Author

P. Polatsek, W. Benesova. Bottom-up saliency model generation using superpixels. *Proceedings of the 31st Spring Conference on Computer Graphics*, pages 121–129, ACM, 2015.

P. Polatsek, W. Benesova, L. Paletta, R. Perko. Novelty-based spatiotemporal saliency detection for prediction of gaze in egocentric video. *IEEE Signal Processing Letters*, 23.3, pages 394–398, 2016.

V. Olesova, W. Benesova, P. Polatsek. Visual attention in egocentric field-of-view using RGB-D data. *Ninth International Conference on Machine Vision (ICMV 2016)*, Vol. 10341, International Society for Optics and Photonics, 103410T, 2017.

P. Polatsek, M. Waldner, I. Viola, P. Kapec, W. Benesova. Exploring visual attention and saliency modeling for task-based visual analysis. *Computers & Graphics*, 72, pages 26–38, 2018.

P. Polatsek, M. Jakab, W. Benesova, M. Kužma. Computational models of shape saliency. *Eleventh International Conference on Machine Vision (ICMV 2018)*, Vol. 11041, International Society for Optics and Photonics, 110412B, 2019.