

Psychophysical evaluation of individual low-level feature influences on visual attention

David Berga^{1,*}, Xosé R. Fdez-Vidal², Xavier Otazu¹, Víctor Leborán², Xosé M. Pardo²

1 Computer Vision Center, Universitat Autònoma de Barcelona, Spain

2 Centro de Investigación en Tecnoloxías da Información, Universidade Santiago de Compostela, Spain

***corresponding author: David Berga, dberga@cvc.uab.es**

Abstract

In this study we provide the analysis of eye movement behavior elicited by low-level feature distinctiveness with a dataset of synthetically-generated image patterns. Design of visual stimuli was inspired by the ones used in previous psychophysical experiments, namely in free-viewing and visual searching tasks, to provide a total of 15 types of stimuli, divided according to the task and feature to be analyzed. Our interest is to analyze the influences of low-level feature contrast between a salient region and the rest of distractors, providing fixation localization characteristics and reaction time of landing inside the salient region. Eye-tracking data was collected from 34 participants during the viewing of a 230 images dataset. Results show that saliency is predominantly and distinctively influenced by: 1. feature type, 2. feature contrast, 3. temporality of fixations, 4. task difficulty and 5. center bias. This experimentation proposes a new psychophysical basis for saliency model evaluation using synthetic images.

Keywords: visual attention, psychophysics, saliency, task, context, contrast, center bias, low-level, synthetic, dataset

1 Introduction

Visual attention is the cognitive capacity of efficiently selecting relevant visual information from a scene. Researchers record eye movements in psychophysical experiments using eye-tracking technology as means of identifying overt attentional cues around fixation points, [Kowler, 2011]. Registered data of different subjects show different patterns of eye movement depending on reflexive, goal-directed or contextually-specific influences [Rothkopf et al., 2016]. This suggests the existence of two types of general influences in the Human Visual System (HVS), combining both bottom-up and top-down processing [Desimone and Duncan, 1995] [Lamme and Roelfsema, 2000] [Corbetta and Shulman, 2002] [Fecteau and Munoz, 2006] [White and Munoz, 2011]. Bottom-up processing of low-level visual features takes place in the early stages of the HVS, namely, when the nervous system efficiently extracts the basic information of the scene and processes it in the visual cortex. When higher areas of the brain are involved is when the top-down processing occurs, by taking into account internal state of the subject (task, mental state, experiences, etc.).

”Visual salience is the distinct subjective perceptual quality which makes some items in the world stand out from their neighbors and immediately grab our atten-

tion” [Itti, 2007]. According to the saliency map hypothesis, the HVS processes visual information of basic features such as orientation, color and scale, combining them in order to guide our attention [Treisman and Gelade, 1980] [Wolfe et al., 1989]. These visual properties can be represented as distinct feature maps, being processed and then integrated in what is called saliency map (representing the pre-attentive guidance as a master map), which precedes the deployment of attention, mainly processed by bottom-up neuronal mechanisms [Engel, 1977] [Findlay and Gilchrist, 2003] [Borji et al., 2013b]. When the HVS combines both top-down (relevance) and bottom-up (saliency) mechanisms in order to select specific target locations of eye movements (which the superior colliculus [SC] is responsible [Schiller and Tehovnik, 2001] [White and Munoz, 2011]), the resulting map is termed priority map [Egeth and Yantis, 1997].

[Koch and Ullman, 1987] outlined a common framework in which feature maps are processed in parallel, opening up a way to computationally extract and integrate these distinct features captured by the HVS for visual saliency prediction [Itti et al., 1998] [Bruce and Tsotsos, 2005]. Given the basis of this models, a myriad of new computational models, with biological, mathematical and physical inspiration [Judd et al., 2012] [Borji and Itti, 2013] [Zhang and Lin, 2013] [Riche and Mancas, 2016a],

have been able to predict, to some extent, visual fixations while observing still images [Borji et al., 2013a] [Bylinskii et al., 2015] [Riche et al., 2013] [Borji et al., 2013c] [Bruce et al., 2015]. The limits on the prediction capability of these models arise as a consequence of the evaluation from previous datasets, that do not account contextual, perceptual, temporal and task-related biases.

1.1 Contextual Relevance

One of the properties that guide visual attention is the contextual relevance of the observed scene [Palmer, 1975] [Chun and Jiang, 1998] [Henderson et al., 1999] [Parkhurst and Niebur, 2003] [Võ and Henderson, 2011] [Hwang et al., 2011] [de Groot et al., 2016]. Semantically-relevant content or specific high-level features can generate endogenous attentional guidance. For instance, looking at a website promotes specific eye movement patterns that differ from looking at a nature scene image; different scanpath patterns can also be found in eye-tracking experiments while humans observe indoor, outdoor and synthetic images. In most datasets for saliency modeling, observers perform free-viewing tasks with real images labeled in specific scene context categories (either faces, cars...), without taking full account of the top-down priors influenced by the context of the image with respect to feature contrast [Winkler and Subramanian, 2013] [Borji et al., 2013a], which could bias both feature localization and discrimination difficulty [Palmer et al., 2000] [Verghese, 2001] [Wolfe and Horowitz, 2004].

1.2 Contrast Relevance

Eye movement behavior is influenced not only by content and stimulus context, but also by the human perceptual capabilities for distinct contrast adaptation and discrimination [Nothdurft, 2000] [Pashler et al., 2004] [Pestilli et al., 2007] [McDermott et al., 2010] [Hart et al., 2013]. Other perceptually-relevant factors could also be related either to the lighting conditions used in each experiment, the starting point of view when perceiving stimuli, etc. The evaluation of relative distinctiveness between features at distinct regions of the image is needed to be done in order to analyze each image according to its spatial properties and feature specificities. This suggests that each image promotes distinct saliency.

1.3 Temporal Relevance

Eye movements have been shown to have temporal influences, varying its behavior upon viewing time or number of fixations (e.g. showing decreasing saccade amplitude, increasing fixation duration [Follet et al., 2011] [Antes,

1974] or higher inter-participant differences [Tatler et al., 2005] [Rothkegel et al., 2017]), suggesting the idea that saliency influences more early saccades than late viewing saccades [Parkhurst et al., 2002] [Tatler et al., 2005] [Zhaoping, 2012] [Zhaoping and Zhe, 2015]. Most saliency predictions based on eye tracking data do not evaluate the temporal relevance in relation to the saliency elicited by the scene, being for most cases, evaluated spatially across all fixations.

1.4 Task Relevance

Alfred Yarbus' seminal work revealed differences in eye movement patterns [Yarbus, 1967] caused by certain top-down influences such as previous experience, motivation and other endogenous factors. Distinctive studies have also concluded that task priors are decisive in that respect [Buswell, 1935] [Navalpakkam and Itti, 2005] [Tatler et al., 2006] [Castelhano et al., 2009] [Greene et al., 2012] [Borji and Itti, 2014]. Goal-directed tasks proved to be able to condition eye movement behavior enhancing visual attention processing [Posner, 1980] [Jonides, 1981] [Huk and Heeger, 2000] [Kurtz et al., 2017]. That might suggest that visual search tasks could minimize such eye-movement patterns produced by endogenous top-down mechanisms [Horowitz and Wolfe, 1998] [Wolfe, 1998], by increasing induced attention towards salient targets (combining both saliency and relevance to influence eye guidance towards these regions). Thus, for all tasks, there is an induced top-down processing that tune overall visual priority when recording eye-movements [Henderson et al., 1999] [Jonides, 1981] [Desimone, 1998], given both exogenous and endogenous influences. Such design puts forward that there could be a better computational estimation of saliency if such task-related influences were focused uniquely on the regions that pop-out on the scene.

1.5 Center bias

Eye movement datasets built for the assessment of saliency models tend to be center biased, not only because of scene framing (photographies tend to focus the salient region in the center of view) but also because of the specific task and stimuli, whereof top-down modulatory constraints are enough to prevent attentional shift, giving a trend to promote center biases [Borji and Tanner, 2016] [Mannan et al., 1996] [Vincent and Tatler, 2008] [Clarke and Tatler, 2014] [Rothkegel et al., 2017], not only in oculomotor terms but also in tendencies in experimentation of eye movement behavior. As aforementioned, bottom-up and top-down processing of the stimuli will depend on the feature characteristics from the scene. If these are simpler, the contextual influence will be lower, making the indica-

tors of saliency easier to analyze [Tatler, 2007] [Vincent and Tatler, 2008]. There will be an endogenous top-down attentional modulation whether the stimulus is cued or uncued. For concrete salient stimuli, facilitating attentional guidance by inducing specific endogenous cues could enable the selection of specific regions of interest in order to prevent the aforementioned factors that generate these center biases.

Objectives

Acknowledging the aforementioned problems on capturing bottom-up visual saliency, we have decided to create a dataset with synthetic images, lacking the presence of high-level features, promoting saliency uniquely elicited from low-level features (providing as well a synthetic image generator code). An alternative evaluation of saliency proposed, by measuring eye movements upon low-level feature distinctiveness and their temporality. Fixations and saccades will be evaluated individually with the corresponding stimuli on free-viewing and visual search tasks, with different feature types and distinct target-distractor feature contrasts.

In order to vary the level of saliency of specific features in a scene, a parametrization of the distinctiveness between a specific item and a set of distractors or its surrounding background is needed. By parameterizing feature contrast, it is possible to analyze feature search efficiency, its accordance with the Weber Law, and the effects in which search asymmetries apply. Using synthetic images in eye-tracking experiments, the complexity of the image features is reduced by minimizing any top-down contextually-related effect, putting forward an easier and more accurate evaluation of eye movement behavior. By modeling stimulus areas of interest for selected pop-out targets, it is possible to test participants performance on landing inside salient regions and their eye movement patterns (in the extent of fixation duration and saccade amplitude) for distinct feature contrasts, and their temporal evolution. This will allow us to observe whether low-level features influence visual attention in a distinct manner.

Previous experiments that perform psychophysical tests evaluating human visual performance on distinct low-level features (iLab USC [Itti and Koch, 2000], UCL [Zhaoping and May, 2007], VAL Harvard [Wolfe et al., 2010] and ADA KCL [Spratling, 2012]) show that the distinctiveness between a specific region and the rest of distinct regions of an image progressively increases the level of saliency in relation to feature contrast. However, the presence of much less relevant features distorts the overall distinctiveness of a specific region with respect to the rest, thus, affecting to the bottom-up visual guid-

ance towards the salient region. With the aforementioned datasets (Table 1), feature contrast and stimulus conditions has been parametrized with search tasks (using the button trigger for calculating search reaction times) but no eye tracking experimentation has been done.

Table 1. Characteristics of datasets with synthetic/psychophysical pattern images

Dataset	Task	# SS	# PP	PM
MIT*	FV	3	15	✗
CAT2000*	FV	100	18	✗
iLab USC	-	~540	-	✓
UCL	VS & SG	2784	5	✓
VAL Harvard	VS	4000	30	✓
ADA KCL	-	~430	-	✓
SID4VAM (Ours)*	FV & VS	230	34	✓

SS: synthetic stimuli, PP: participants, PM: Parametrization, FV: Free-Viewing, VS: Visual Search, SG: visual segmentation
*: Fixation data is available online

For few eye movement datasets that contain synthetic images (MIT [Judd et al., 2009] and CAT2000 [Borji and Itti, 2015]), no parametrization of feature type or contrast was done. Contrary to other saliency datasets [Bylinskii et al.,], in this study it is possible to evaluate each of these factors individually and exclusively eye movement data is being used for calculating search performance for better accuracy. We will test the following hypotheses:

1. Performance on salient region localization could show differences upon varying the type of features present in our stimuli.
2. If feature contrast is the main factor that contributes to saliency, performance on localization of salient regions should correlate with feature contrast, specially for stimuli that require a serial ‘binding’ step.
3. Acknowledging that saliency is usually evaluated across all fixations on eye tracking experimentation, if a temporal bias exists and is increasing, it is highly possible that the first fixations show higher saliency index than the late ones.
4. If performance on salient region localization with free viewing tasks is lower for stimuli with higher contrast compared to visual search tasks, it will mean that fixations on free viewing tasks are highly guided by endogenous attention.
5. Previous datasets used for saliency prediction do not show how their center biases affects saliency. We will show how eye movement patterns influence the center bias for this dataset and if the bias increases or decreases across viewing time and feature contrast.

Our objective is to allow computer vision researchers to reproduce these influences when modeling eye-movement prediction algorithms. Here we present a dataset in which we evaluate through free-viewing tasks the influence of the features that affect the spatial properties of an image (from the perception of Corners, Segments, Contours and Grouping) and how the relative distinctiveness from a search target is more salient with respect to a set of distractors that differ from specific low-level features (color, orientation, size...). Analyzing low-level features individually would allow us to see which features generate more agreement on saliency measures and are localized faster, in that manner, to allow their modeling according to their distinct neuronal mechanisms. This study can be used for a more plausible and specific saliency modeling given the presented eye-movement patterns, also extrapolable to the analysis of the interactions between these features or to study specific cases of high-level features in future studies.

2 Materials and Methods

2.1 Participants

Thirty four subjects (11 female and 23 male) with normal or corrected-to-normal vision took part in this experiment. Most participants were postdoc scholars and PhD students (aged 21–47 years) from non-related fields of study. No economic compensation for the experiments was given. Participants were allowed to wait until they were comfortable with the eye tracking experimental setup in case they had any kind of visual discomfort in between sessions, and they were allowed to adjust the chair while laying on the chin-rest before the experiment. Participants had to sign a consent form allowing the anonymous usage of the data captured during the experiment.

2.2 Apparatus

The set of stimulus was presented on a LCD monitor (Samsung SyncMaster HMAQ935729) of screen size 340x270 mm, a resolution of 1280x1080 px and a refresh rate of 60 Hz. A color calibrator was used (Xrite i1 Display Pro) in order to set a specific luminance for the monitor of 160 cd m^{-2} , achieving the CIE Illuminant D65 according to the ISO 3664:2000 standard condition (and recommended by Adobe RGB 1998 CIE and ITU-R BT.500-11) with the whitepoint at $x=0.313$, $y=0.329$ and a gamma value of 2.2. The light conditions of the room were set using non-direct adjustable light, measured at 30 lx using a luxmeter (TES1332).

We have used a SMI RED binocular eye tracker with a tracking resolution of $<0.1 \text{ deg}$, gaze position accuracy of $<0.5 \text{ deg}$ and a sampling rate of 50 Hz, set at a distance of 600 mm towards the chin-rest (about 40 pixels per degree of visual angle) and vertically equidistant with respect to the monitor, forming a slope of 19 deg from the horizontal axis. The monitor’s screen was at a vertical distance of 195 mm from the table and the observer’s point of view was adjusted to be centered towards the screen. Fixation and saccade detection was based on SMI iView X Event Detector software, capturing fixations at a minimum duration time of 80 ms and maximum dispersion threshold value of 2 deg and saccades at a peak velocity threshold of 75 deg/s [Salvucci and Goldberg, 2000] [SMI, 2009, p. 243-247].

2.3 Procedure

The experiment was divided in one training and two full sessions. During the training session each participant performed a visual search task with 4 types of stimulus with feature and conjunctive search, combined with present and/or absent search targets, hence to ensure their good performance in the next sessions. The first session had a duration of about 20 minutes and was divided in 8 blocks, each one corresponding to a free-viewing or visual search task. The second session had a duration of about 25 minutes and was divided in 10 blocks, similar to the first session. Each task in a block correspond to a distinct stimulus type (shown in Section 2.4) that was presented in a random order. Stimulus order was also randomized across blocks (to avoid any stimulus-related priming [Kristjánsson and Driver, 2008]), and the location of target distractors was distinct for each case in order to prevent oculomotor biases.

Participants performed two types of tasks: free-viewing and visual search (Figure 1). During free-viewing tasks, they were instructed to freely look at the stimuli during 5000 ms. For the visual search task, they were instructed to look for a specific target previously shown in an instruction slide. In case they could find the target, they had to steer their gaze towards it during a dwell time of 1000 ms (the area of interest was based on the target area with an horizontal and vertical spacing of 1 deg). In case they could not identify the target, they were instructed to press a specific key. Considering that context was distinct for each block (replicating stimulus characteristics from previous studies), we decided to do a template target search task instead of an odd-one-out type of task [Bacon and Egeth, 1994] [Theeuwes, 2004]. Participants had unlimited time for the visual search tasks, in this case for reporting target identification or absence. Transitions between stimuli had a duration of 2000 ms (blank transition without

the presence of an onset cue) with a luminance equal to the stimuli in order to preserve participant’s luminance and chromatic adaptation.

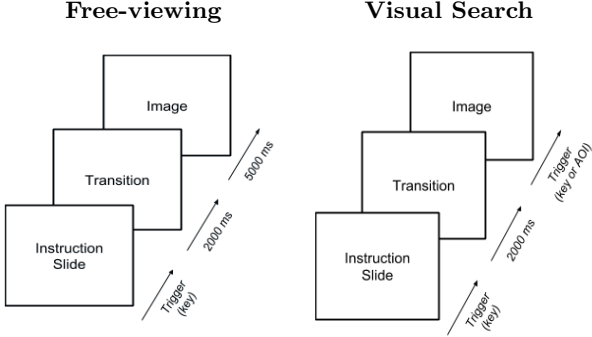


Figure 1. Procedure for the presentation of the stimuli for each task type.

Mean pupil size was recorded to be 2.98 mm diameter for all samples and there was a standard error of 0.18 mm mm between stimulus type, being almost constant throughout the experiment with no significant stimulus-related luminance imbalance.

A 12 point calibration procedure was performed before each session, in which participants were instructed to gaze a red dot moving along different directions. The calibration showed mean deviations for all sessions of $\sigma_x=0.57$, $\sigma_y=0.75$ deg for the left eye and $\sigma_x=0.56$, $\sigma_y=0.82$ deg for the right eye. Deviations for each participant’s fixation and saccade data were computed using the data of the participant’s eye that presented minimum deviation from calibration points in each session. We did a pilot experiment with 4 participants in order to correctly design the visual search procedure, thus, to test the final experimental design for trigger timing and the difficulty of the tasks. That allowed us also to correctly parametrize the variables corresponding to target-distractor contrast where the target was too hard to identify, this parametrization will be shown in the next section.

2.4 Stimuli

A total of 33 types of stimuli were generated, corresponding to 15 distinct feature evaluations (5 of them using free-viewing tasks and 10 for visual search tasks) at distinct conditions. During free-viewing experiments, we evaluated how spatial properties influence saliency, namely, the capabilities of humans for detecting corners, segmenting and detecting contours as well as localizing groups of objects according to their similarity and spatial distribution (Table 2). This will give some insight of how rapidly humans reflexively perceive and bind spatial properties from the features of an image. In visual search tasks, we evaluated

the speed in detecting specific features and the amount of saliency produced by target-distractor feature contrast characteristics. In that aspect, stimulus were generated with features that pop-out based on their dissimilarities in orientation, color and size. Besides, we also analyzed influences of the guidance prompt from the amount of distractors on the scene, their configuration as well as the influence of background lightness, color and roughness (Table 3).

Stimulus design was inspired by Spratling’s experiments [Spratling, 2012], by generating synthetic images similar to the ones from Li and May’s psychophysical experiments [Zhaoping and May, 2007]. Most stimuli items had a size of 1.5 deg, occupying a region of 2.5 deg including the spacing between distractors. In that manner, stimulus had an available grid of 10×13 distractors. Distractors were black ($lsY = 0, 0, 0$), and background was plane white ($lsY = 0.6548, 0.0175, 1$). We used Spratling’s code and we adapted it in order to also use any distractor shape, displacement and chromatic parameterization, downloadable at <https://github.com/dberga/sig4vam>.

$$\Psi(x) = \left\{ \frac{x-1}{N-1} \mid x \in N \right\}, \quad (1)$$

$$\Psi(x, v) = \{v \cdot \Psi(x)\}, \quad (2)$$

$$\Psi(x, min, max) = \{\Psi(x, min) \cup \Psi(x, max) \mid x \text{ is odd}\}. \quad (3)$$

The parameters of the generated stimuli were set according to "N = 7" contrast values (ranging from 0 to 1), using the Weber’s law uniform fraction in order to set discrete target-distractor evaluation "x = 1...N" for the psychometric function Ψ . For each stimulus type on our experimentation, parameters are set according to specific values of $\Psi(x)$. We have the expression in Equation 1 and 7 contrast values in order to have extreme contrast values (no contrast and maximum contrast) with " $\Psi(1) = 0$ " and " $\Psi(7) = 1$ " as well as a middle value with " $\Psi(4) = 0.5$ ", making the difference between the second lower contrast and the second maximum contrast at the same distance from the extreme contrast values $|\Psi(2) - \Psi(1)| = |\Psi(7) - \Psi(6)|$. In that manner we provide a psychometric function with a constant slope (Weber’s uniform fraction). Absolute values of contrast can be adjusted to fit a specific value of "v" as $\Psi(x, v)$ (Equation 2). For cases that we had higher and lower contrasts with respect the target and overall distractors we adjusted the values for maximum and minimum range of the psychometric function $\Psi(x, min, max)$ as the union of odd values for both sets of $\Psi(x, min)$ and $\Psi(x, max)$ in order to acquire the same set of contrast values (Equation 3).

Acknowledging that each stimulus was distributed according to different contrasts depending on the evaluation parameter, each stimulus was categorized as easy and

hard depending on the assigned contrast (half of them as easy for higher contrasts, and half of them as hard for the case of lower contrasts, with a specific case with minimum or no contrast). One of our interests was to evaluate how low-level features modify the spatial layout between the features on a scene, therefore its spatial properties, affecting visual saliency (in this case with free-viewing experimentation). In order to accurately evaluate low-level feature distinctiveness, visual search tasks were performed, having a search target with a specific low-level contrast with respect to a set of distractors. The stimulus design corresponding distinctively to each feature and task will be explained as follows.

2.4.1 Free-viewing task stimuli

First, we wanted to evaluate the spatial relevance of certain regions of an image. For this stimuli, visual selection cannot be focused on a unique region due to the size and/or spatial organization of the elements in the image. Humans have a limited central vision, namely, they need several fixations over the whole region in order to attend to all of the relevant regions in detail. In that aspect, each of the spatial properties will guide attention towards a single or several spots depending on the analyzed feature. With this type of stimuli it is be able to see the temporal and spatial performance of perceiving boundaries due to corner sharpness, segment angle and spacing as well as preemption and grouping [Rensink and Enns, 1995] [Rensink and Enns, 1998] [Wolfe et al., 2011], effects induced by distractor continuity, proximity and similarity. These preattentive effects are not equally processed in the visual system in the same way as shown for parallel visual search [Wolfe, 1992]. Task was separate for the aforementioned perceptual phenomena with respect to searching for a specific feature, stimuli described on Section 2.4.2.

Corner Angle (1) Troncoso et al’s psychophysical experimentation found that corner salience was higher on sharp corners than on shallow corners or edges [Troncoso et al., 2005] [Troncoso et al., 2009]. This effect could be explained by ON-center receptive field behavior towards corner stimuli [Rodieck, 1965], being sharp corners the ones that produce higher neuronal activity. Original stimuli from Troncoso’s experiment was used, generating corners with a dark-to-white gradient and an upwards angle, corresponding to corner angles of 180, 135, 105, 75, 45, 30 and 15° (shown in Figure 2). The horizontal alignment of the corner was randomized in order to prevent oculomotor anticipation.

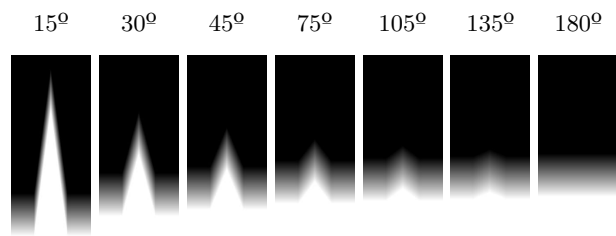


Figure 2. Examples of corner angle slopes (with the sharper at 15° and the smoother at 180°) for dark-to-bright gradient stimuli with upwards angles.

Visual Segmentation Distinctiveness between two homogeneous regions creates higher neural activity near region boundaries than away from them [Li, 1999] [Li, 2000]. In this section is described how an illusory boundary is generated by varying two segment characteristics. This effect is distinct from the concepts of edge or boundary detection (terms used as well in the image segmentation literature) or contour integration. This phenomena proves that illusory boundaries pop-out due to the perceptual breakdown of homogeneity. Here it is studied the influence of angle contrast between these two segments (creating a salient boundary dependent on the segments angle) with an homogeneous single set of bars as well as with superimposed bars. Here is also analyzed the influence of bar spacing and length on detecting the illusory boundary between these two segments.

Segmentation by Angle (2) Visual angle contrast between two segments can induce edge detection and therefore visual saliency towards that illusory edge [Nothdurft, 1991] [Wolfson and Landy, 1995] [Zhaoping and May, 2007]. The resulting saliency would increase with respect to angle contrast from the two segments on the region that separates them [Spratling, 2012]. It is a distinctive effect from orientation feature detection upon a set of distractors, that is described on Orientation Contrast (12), Distractor Heterogeneity (13), Distractor Linearity (14) and Distractor Categorization (15).

$$\Phi(v, a) = \{|\arcsin(\Psi(1\dots N, v)) + a|\}, \quad (4)$$

$$\Delta\Phi(v, a, b) = \min\{|b - \Phi(v, a)|, 180 - |b - \Phi(v, a)|\}, \quad (5)$$

The psychometric values for determining angle values are defined as $\Phi(v, a)$. Here "v" is the incremental factor for adjusting the maximum angle for our set $\Psi(x, v)$ and "a" is the starting angle value for our bar orientation (Equation 4). The angle contrast between a specific angle "b" and our set of angles $\Phi(v, a)$ can be computed with $\Delta\Phi(v, a, b)$, considering that our bar orientations have

Table 2. Description of the generated stimuli for the experiment using the free-viewing task. Stimulus have been divided in "Stimulus type" according to the type of feature or effect that is analyzed and "Stimulus subtypes" for the cases that there are presented distinct conditions using the same feature contrast. The total number of elements has been selected according to the stimulus characteristics, preserving similar spatial properties to the ones presented on the literature.

# of stimuli	Stimulus type	Stimulus subtypes	Parametrized Feature Contrast	Total # of elements
7	Corner Angle (1)		Sharpness Orientation	1
14	Segmentation by Angle (2)	Single Superimposed	Segment Orientation	10×13 (130) 20×26 (520)
7	Segmentation by Spacing (3)		Bar Length and Spacing	10×13 (130)
6	Contour Integration (4)		Bar Continuity	10×13 (130)
14	Perceptual Grouping (5)	Similar Dissimilar	Distractor Proximity	~ 40

upwards and downwards contrast for its comparison (due to its symmetry), contrast is calculated as the minimum from the differences from two quadrants in which these bars can be oriented (Equation 5).

Stimuli was based on Spratling’s visual segmentation, using 2 sets of bars (shown in Figure 3(a,b)) oriented respectively using angles $\Phi(1, 0)$ and " $b = 90^\circ$ ", forming a relative contrast of $\Delta\Phi(1, 0, 90)$. For the case of superimposed bars, we have created a composite of the same bars adding a bar tilted at 45° with respect to each segment. Here are accounted the contrasts between the new superimposed bars and the original segment $\Delta\Phi(1, 45, 90)$. The location of the vertical segment was randomized on the horizontal axis for each stimulus.

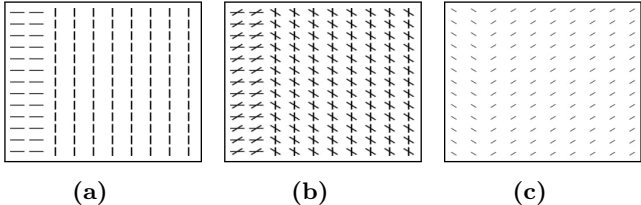


Figure 3. Examples for visual segmentation stimuli. (a) Corresponds to the segmentation by angle with a single segment and (b) to superimposed segments (with both cases with an orientation contrast of $\Delta\Phi=90^\circ$). In (c) segmentation is done distinctively (by changing bar length), using bars oriented at 45° and -45° with a bar length of 1 deg and a segment spacing of 2.5 deg.

Segmentation by Distance (3) Texture discrimination was shown to vary according to the spacing and length of the texture elements [Nothdurft, 1985] [Spratling, 2012], making it harder as element spacing increases (as when segment elements decrease in length or size). Visual segments were modeled using 2 sets of bars, oriented respectively at 45° and -45° (a relative angle contrast of 90°). Here we question how bar length is able to generate a specified distance at the center of the illusory segment.

Segment spacing was calculated as the euclidean distance from the end of the first segment bar to the beginning of the second segment bar, with values of 0 to 2.5 deg (shown in Figure 3(c)), corresponding respectively to a bar length of 1 to 3.6 deg deg in the horizontal axis.

Perceptual organization Perceptual organization has been previously investigated and promoted by Gestalt principles, guided by proximity, similarity, continuity, and closure properties of objects [Koffka, 1935] [Wertheimer, 1938] [Field et al., 1993] [Caputo, 1997] [Rubin, 2001]. Here are described two effects related to perceptual organization, parametrized upon the aforementioned principles.

Contour Integration (4) Continuity within set of features in a scene is able to generate the perception of a contour [Hess et al., 2003], considering that a larger set of collinear bars facilitate its detection. Accounting for saliency being influenced by contour integration [Li and Gilbert, 2002] [Dakin and Baruch, 2009] [Spratling, 2012], a set of stimuli was created with a grid of randomly oriented and equidistant bars and a collinear contour (Figure 4(a)). Contours were generated with a length of 3, 5, 7, 8, 9 and 10 collinear bars, corresponding to 7.5 to 25 deg.

Perceptual Grouping (5) Adding up to the basis of the previous section, we have also studied the relation between perceptual grouping principles and visual attention. According to the literature, the spatial layout can facilitate or prevent contextual cueing [Bhatt et al., 2007] [Conci et al., 2013], in particular, the lower the proximity between a number of randomly distributed objects and a group, the higher the saliency on the grouping region [Nothdurft, 1985] [Ben-Av et al., 1992] [Ben-Av and Sagi, 1995] [Rideaux et al., 2016]. Given that, here the analysis is on the influence of proximity and similarity among objects in a specific spatial organization. To do so, there were generated a set of shapes, randomly

distributed and located at specific distances to a group, with similar and dissimilar shapes Figure 4(b,c). The proximity parameter was the euclidean distance between the group centroid and the rest of distractors, forming a wider gap between the distractors and the group as we increase distance, parametrized as $\Psi(1\dots N, 2.5, 7.5)$. Stimulus shapes were set to be symmetric in order to prevent orientation-variant guidance, with squares as the main shape for both group and distractors in the case of similar object condition, whereas in the case of dissimilar condition were selected triangle shapes for the distractors and squares for the group.

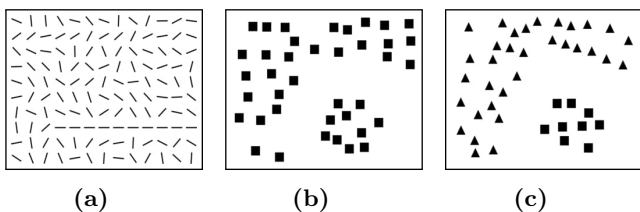


Figure 4. Examples of distinct perceptual organization effects, eliciting Contour Integration (a) formed by 10 collinear bars (corresponding to 25 deg) and Perceptual Grouping for similar (b) and dissimilar shapes (c) with respect a group set at a distance of 7.5 deg from the rest of distractors

2.4.2 Visual Search task stimuli

Visual search tasks were performed with another set of stimuli. In this case, a unique target item with a specific size was used. In that manner, it is possible to change either the amount of distractors, their spatial configuration, the target-distractor feature contrast and background properties of the stimulus. Most target elements overall occupied a small area of interest in order to able to be preserve same fovea-dependent capabilities for each type of stimuli used with this type of task. Hence, using this experimentation we can observe which features pop-out faster and more often (in parallel or "effortlessly"). As small regions away from central vision cannot be detected as in the case of bigger regions shown on Section 2.4.1, the guidance towards the salient target (distinctive from the rest of distractors) minimizes other type of guidance promoted from endogenous factors.

Feature and Conjunctive Search (6) Feature search increases probability and efficiency of saccading towards a specific search target on scene observation due to its unique distinctiveness. The information span processed by the HVS varies depending on the amount of feature distractors to be processed [Treisman and Gelade, 1980] [Pashler, 1988] [Wolfe et al., 1989] [Palmer, 1995] [Wolfe and Ben-

nett, 1997] [Hayward and Burke, 2000] [Nothdurft, 2000]. Given that premise, the amount of objects in a scene would imply a variation of the difficulty towards searching a specific target for the case of serial search (distorting human's sustained attention), but not for the case of parallel search. Previous experiments show that difficulty on visual search is higher with a conjunction of distractors with different image features (such as size, color or orientation). In case distractors vary only by a unique feature, the difficulty of the task would not be as evident as the other case [Nakayama and Silverman, 1986] [Treisman, 1988] [Wolfe, 1994] [Flombaum, 2015]. In order to reproduce feature and conjunctive search, target was a red bar oriented at 45° . For the feature search case, distractors were green and set at 45° (Figure 5). On the case of conjunctive search, half of distractors were green and set at 45° and the other half were red and oriented at -45° .

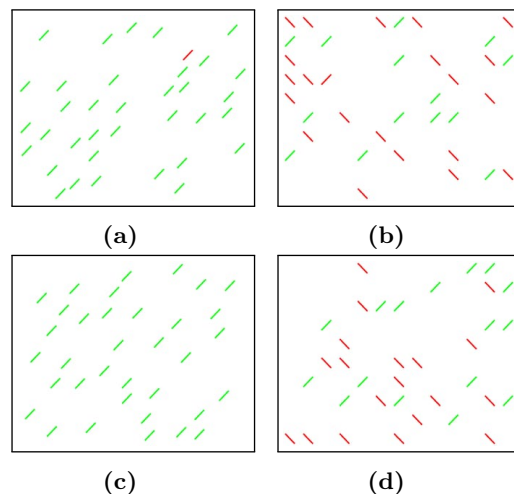


Figure 5. Examples used for feature and conjunctive search. Here are presented the cases of having a red target oriented at 45° and 34 distractors randomly displaced around the scene. For the feature search case (a), all of the distractors are distinct in color (green). For the conjunctive search case (b), 50% of the distractors are distinct in color (green) and the rest are distinctive in orientation (at an orientation of -45°). The same cases (c) and (d) are shown but without the presence of the target.

The position of the items was randomised, with a set size of $\Psi(1\dots N, 2, 34)$, the amount of distractors ranged from 2 to 34 distractors. Both search conditions without the presence of the target was introduced in order to see if the effects are also reproduced for the case of reporting absence of target. The design of feature and conjunctive search has been defined as keeping similar difficulty between the two conditions, preserving identical targets and displaying each conjunction of distractors maximally

Table 3. Description of the generated stimuli for the experiment using the Visual Search task. The total number of elements has been selected according to the amount of distractors presented on the scene acknowledging that one of the elements is presented to be the search target.

# of stimuli	Stimulus type	Stimulus subtypes	Parametrized Feature Contrast	Total # of elements
28	Feature Search (6)	Feature Conjunctive Feature-absent Conjunctive-absent	Distractor number	3 to 35
14	Search Asymmetries (7)	Bar presence Bar absence	Scale and Distractor number	35 to 520
14	Noise/Roughness (8)	Higher deviation Lower deviation	Surface Roughness	1
28	Color Contrast (9)	Red target and Unsaturated Background Red target and Oversaturated Background Red target and Unsaturated Background Blue target and Oversaturated Background	Distractor Saturation	34
14	Brightness Contrast (10)	Light Background Dark Background	Distractor lightness	34
7	Size Contrast (11)		Target Size	34
7	Orientation Contrast (12)		Target Orientation	34
21	Distractor Heterogeneity (13)	Homogeneous Tilted-right Flanking	Target Orientation	10×13 (130)
28	Distractor Linearity (14)	Linear Nonlinear at 10° of slope Nonlinear at 20° of slope Nonlinear at 90° of slope	Target Orientation	10×13 (130)
21	Distractor Categorization (15)	Steep Steepest Steep-right	Target Orientation	10×13 (130)

dissimilar from each other [Quinlan, 2003].

Search Asymmetries (7) Search asymmetries between two different type of stimuli happen when a specific target of type "a" is found efficiently among distractors of type "b", but not in the opposite case (searching for "b" among distractors of type "a") [Treisman and Souther, 1985] [Treisman and Gormican, 1988] [Hayward and Burke, 2000] [Wolfe, 2001]. Clear evidence was found for plain circles crossed by a vertical bar (Figure 6) at a scale of 5 deg), showing that it was easier to find a circle with a vertical bar among plain circles than vice versa [Wolfe, 2001] [Spratling, 2012].

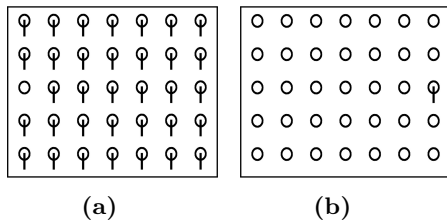


Figure 6. Example of stimulus types in which search asymmetries can apply. The context of circles (b) facilitates the search of a superimposed bar compared to the reverse case (a).

The same type of stimuli was selected (with both con-

ditions: searching a circle crossed by vertical bar among plain circles and searching a plain circle among circles crossed by a vertical bar) filling a grid of distractors according to a specific scale and randomizing the position of the target. The scale values were $\Psi(1...N, 1.25, 5)$, between 1.25 and 5 deg, changing the amount of items to be presented, being in each case from 35 to 520 elements corresponding to arrays of 5×7 , 6×8 , 8×10 , 10×13 , 15×20 and 20×26 objects.

Noise/Roughness (8) For most synthetic stimuli we have considered uniform and plain backgrounds with homogeneous illumination, but in this case the influence of continuous textured background would increase or reduce search time required to detect a specific target depending on the amount of noise present in the scene. Clarke et al. [Clarke et al., 2008] [Padilla et al., 2008] [Clarke et al., 2009] showed that the higher the level of background texture noise of a scene, the higher the level of difficulty of the search task. They represented the background surface as a height map by parameterizing an isotropic and random-phase noise $1/f^\beta$ (being " β " the frequency roll-off magnitude factor of the inverse discrete Fourier transform of the height map and " σ_{RMS} " the deviation of the roughness noise height). The surface was obtained by rendering the height map according to the Lambert's Cosine Law model using a constant light source with slant of 60° and

tilt equal to 90° . Given these previous experiments, each stimulus was a rough surface considering β as the contrast value $\Psi(1...N, 1.5, 1.8)$ with two distinct conditions by using deviations of $\sigma_{RMS} = 0.9$ and 1.1 . A similar target of Clarke’s experimentation was used (Figure 7) with a circular shape and a vertical gradient background corresponding to the height of the surface and a diameter of 0.78 deg (half of size corresponding to the rest of target items of this study, adjusted for preventing too low RT differences between distinct contrasts).

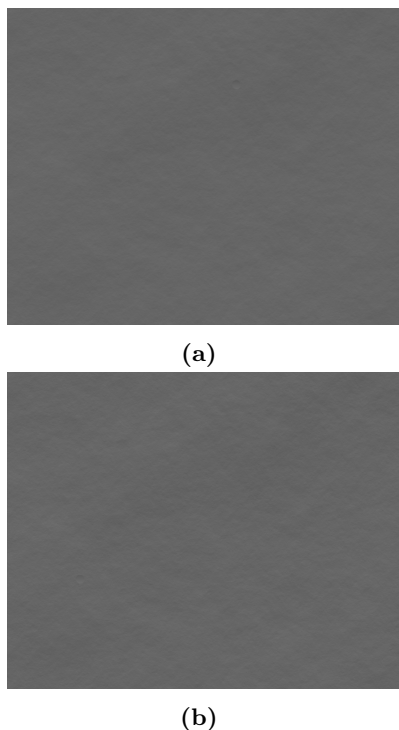


Figure 7. Two examples of a rough surface with $\beta = 1.8$ using height deviations of (a) $\sigma_{RMS} = 0.9$ and (b) $\sigma_{RMS} = 1.1$

Distractor similarity When an object is dissimilar to the rest of objects in a scene, the search of that object is more efficient. That phenomenon is called target-distractor similarity, and has been found to occur when parameterizing specific features such as color, shape or size [Duncan and Humphreys, 1989] [Wolfe, 2010].

Color Contrast (9) In this section the chromatic properties of distractors are changed, as well as the background of the stimuli Figure 8. As shown in previous experiments [D’Zmura, 1991] [Bauer et al., 1996] [Amano and Foster, 2014] [Danilova and Mollon, 2014], color varies spatial and temporal patterns of eye movements, affecting both localization and discrimination of objects. Besides,

search asymmetries happen at different background conditions [Nagy, 1999] [Rosenholtz et al., 2004].

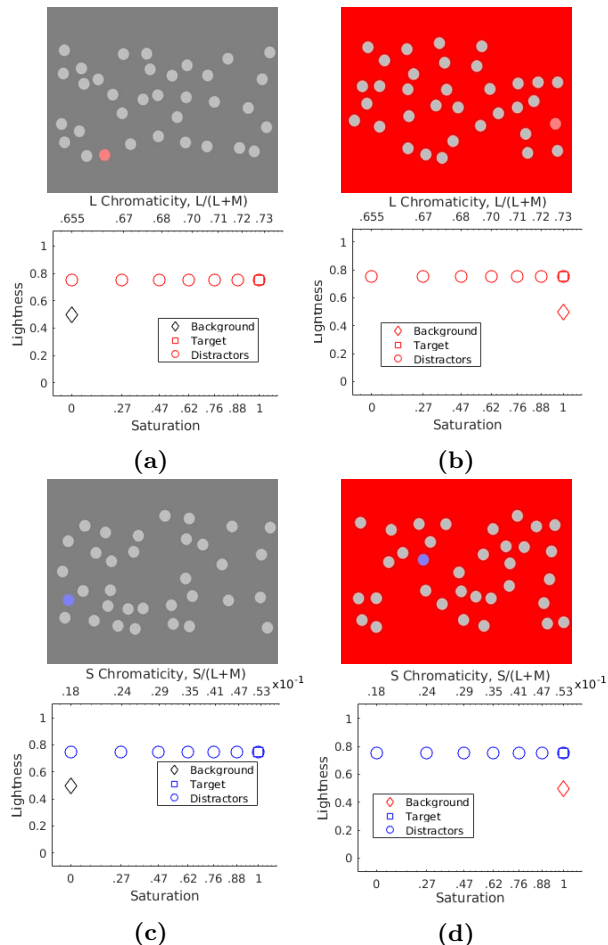


Figure 8. Examples of the 4 conditions at maximum contrast of $\Delta S_{D,T} = 1$, representing the values of Saturation and Lightness on each particular stimuli for each target and background configuration (showing as well the “l” and “s” chromaticities in the lsY space [MacLeod and Boynton, 1979] at 400 nm). In (a) and (b) there are represented the stimulus for grey (unsaturated) and red (oversaturated) background respectively. Similarly, but for blue targets, are represented the cases for both background conditions in (c) and (d).

Taking into account these experiments, we wanted to analyze if these search asymmetries are present when varying saturation of distractors with respect to a search target. We will see if these differences between distractor and search target are affected by changing background saturation at distinct target and distractor hue (using the HSL color space). A set of stimulus was generated with circular shaped items with a similar displacement to Rosenholtz experiment. Stimulus was surrounded with a vertical

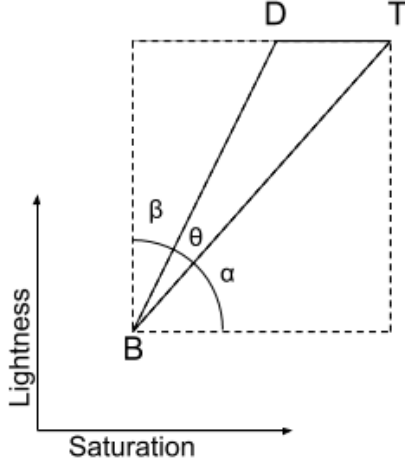


Figure 9. Representation of HSL values for distinct distractors (D), background (B) and search target (T).

padding equal to the presented background in order to prevent monitor-related luminance gradients. Contrast values can be calculated according to the saturation differences between the search target (T) and distractors (D). Two background (B) conditions were defined, corresponding to Grey (achromatic and unsaturated), and Red (chromatic and oversaturated) colors. At isoluminant ($L_{D,T}=0.75$) and isohue conditions ($H_{D,T}=0^\circ$ for red and $H_{D,T}=240^\circ$ for blue distractors), a representative measure of color contrast between the target and distractors can be computed. This measure was named θ , being the angle between the search target and distractors, with the background as the vertex of the intersection. Same trigonometrical properties can apply using the same diagram plotting B,T and D relationships at distinct quadrants (acknowledging that in our case T is oversaturated for both conditions).

In Equation 6, is represented the absolute difference in lightness and saturation between two distinct conditions. In Figure 9 there are the angles that comprise the saturation and lightness contrast between our stimulus objects. Each of these angles represent respectively to the triangles formed by B-T (α), B-D (β) and D-T (θ), being α constant for constant background and target (Equations 7, 8 and 9). Most importantly, θ represents the angle comprising the available contrast between the distractor and the target. Given these angle calculations, it is possible to represent the $\Delta S_{D,T}$ as the absolute saturation difference between D and T (Equation 10), calculated by the parametriza-

$$\begin{aligned}\Delta S_{1,2} &= |S_1 - S_2|, \\ \Delta L_{1,2} &= |L_1 - L_2|,\end{aligned}\tag{6}$$

$$\alpha = \arctan\left(\frac{\Delta L_{T,B}}{\Delta S_{T,B}}\right),\tag{7}$$

$$\theta = (90 - \alpha)(\Psi(1\dots N)),\tag{8}$$

$$\begin{aligned}\beta &= (90 - \alpha)(1 - \Psi(1\dots N)) \\ &\equiv 90 - \theta - \alpha\end{aligned}\tag{9}$$

$$\equiv \arctan\left(\frac{\Delta S_{B,D}}{\Delta L_{B,D}}\right),$$

$$\begin{aligned}\Delta S_{D,T} &= |S_T - S_D| \\ &\equiv |\Delta S_{B,T} - \Delta S_{B,D}| \\ &\equiv |\Delta S_{B,T} - (\Delta S_{B,D} \cdot \tan(\beta))|,\end{aligned}\tag{10}$$

$$S_D = \begin{cases} S_T - \Delta S_{D,T} & \text{if } S_T > S_B \\ S_B - \Delta S_{D,T} & \text{otherwise} \end{cases}\tag{11}$$

tion of β using our psychometric function $\Psi(1\dots N)$. That absolute saturation difference will define the criterion for our distractor saturation S_D as shown in Equation 11. There were generated 4 experimental conditions corresponding to unsaturated and saturated background and red or blue hue. The value of θ is equivalent for saturated and unsaturated background, corresponding to values of 0, 9, 18, 35, 44 and 53° , producing saturation differences ($\Delta S_{D,T}$) of 0, 0.121, 0.246, 0.528, 0.728 and 1.

Brightness Contrast (10) According to previous studies, searching a bright target is harder as luminance of distractors increase [Pashler et al., 2004] [Nothdurft, 2006b] [Spratling, 2012], with a distinct response with respect to chromatic stimuli [Nagy, 1999]. Conversely, salience increases for a dark target when luminance of distractors is increased. It was parametrized as the lightness contrast and stimuli was modeled using the HSL color space, considering an achromatic (unsaturated) and isohue relationship between search target, distractors and background, using the same type of stimuli as in Color Contrast (9). Here the target is gray ($L_T = 0.5$) and background is bright ($L_B = 1$) or dark ($L_B = 0$). In order to parametrize the contrast for this stimuli, we used the absolute lightness difference between search target and distractors. In Figure 10 theta value is 0° for all cases, at 0 saturation, the lightness axis is parametrized.

$$\Delta L_{D,T} = |L_D - L_T| = \begin{cases} \Delta L_{B,T}(1 - \Psi(1...N)) & \text{if } L_T > L_B \\ \Delta L_{B,T}(\Psi(1...N)) & \text{otherwise} \end{cases} \quad (12)$$

$$L_D = \begin{cases} L_T - \Delta L_{D,T} & \text{if } L_T > L_B \\ L_B - \Delta L_{D,T} & \text{otherwise} \end{cases} \quad (13)$$

Lightness differences Equation 12 are calculated by $\Delta L_{D,T}$, corresponding to the absolute difference between target and background lightness ($|\Delta L_{B,T}|$) and adjusted by our psychometric function $\Psi(1...N)$ with distinct distractor lightness values of L_D , depending on the absolute background lightness with respect the target.

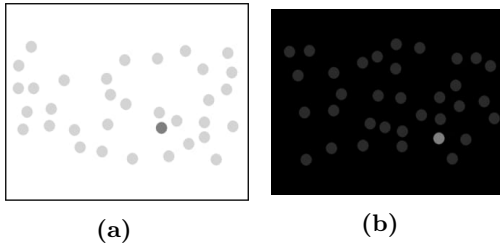


Figure 10. Examples of distinct background conditions, (a) lighter background and distractors at $L_D=0.66$. (b) Dark background with distractors at $L_D=0.17$. For both conditions, the lightness of the search target is grey ($L_T=0.5$). Absolute contrast for these cases is $\Delta L_{D,T}=0.33$.

Size contrast (11) Dissimilarities in size of objects tend to drive increase or decrease search speed when detecting and discriminating salient regions [Sagi and Julesz, 1984] [Goolkasian, 1997] [Tavassoli et al., 2009] [Proulx, 2010]. Here is presented size similarity between symmetric objects (circles, without loss of generality). Each stimuli was generated with a set of 34 objects randomly around the scene (Figure 11(a)). The search target has a distinct size with respect to the distractors, with both cases of smaller and bigger sizes with $\Psi(1...N, 1.25, 5)=[1.25, 1.67, 2.08, 2.5, 3.34, 4.17, 5]$ deg, being the size as the parameter that defines the similarity contrast for this case, corresponding to a scaling factor of 0.5 to 2 with respect to the baseline (2.5 deg).

Orientation contrast (12) For this setting, varying angle of objects is found to increase search efficiency when angle contrast is increased [Duncan and Humphreys, 1989] [Nothdurft, 1993b] [Nothdurft, 1993a] [Kong et al., 2017]. A set of 34 bars were randomly displaced around the scene and oriented at 0° , in which the search target is an equally-shaped bar oriented at a distinct angle (Figure 11(b)). Angle contrast between the search target and the

set of distractors was $\Delta\Phi(1, 0)=[0, 10, 20, 30, 42, 56, 90]^\circ$ Equation 4.

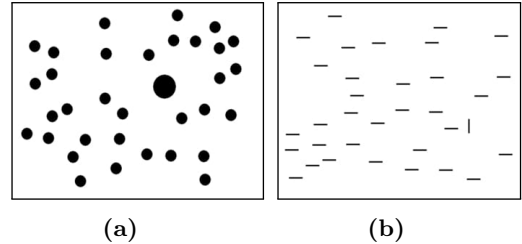


Figure 11. Examples for salient targets with dissimilar size (a) and orientation (b). For (a), the search target has a diameter of 5 deg (a factor of 2 with respect to the rest of distractors). For (b), the orientation of the target is 90° with distractors at 0° , forming an orientation contrast of $\Delta\Phi=90^\circ$.

Distractor Heterogeneity (13) Previous design was to evaluate orientation similarity, given a unique orientation for non-target distractors. Here is presented the phenomenon of distractor heterogeneity. When several sets of distractors are dissimilar with respect to the search target, mutual information between the target and distractors is said to be heterogeneous. In the heterogeneous case, search efficiency is lower, in other terms, target search is harder [Treisman and Gormican, 1988] [Duncan and Humphreys, 1989] [Nothdurft, 1993a] [Bauer et al., 1996] [Rosenholtz, 2001] [Gao, 2008] [Wolfe, 2010]. Distractor orientation heterogeneity, however, can be represented through distinct configurations, either if the set of distractors are tilted to the same direction or towards distinct directions. In this experiment, there is an array of bars oriented at 75° (with a slope of 15° with respect to the vertical quadrant). From two different sets of distractors, are defined three conditions according to the distinct orientation configurations: homogeneous, tilted-right and flanking (Figure 12). For the case of homogeneous distractors, both set of distractors have a unique angle contrast with respect to the target bar. For the case of tilted-right, both set of distractors have an angle tilt of 15 and 30° , $\phi(1, 90, 15, 30)$. For the case of flanking, both sets of distractors have an angle tilt of 15° and -30° respectively, having both positive and negative tilt with respect the search target, $\phi(1, 90, 15, -30)$.

$$\phi(v, a, c_1, c_2) = \{\Phi(v, a, c_1), \Phi(v, a, c_2)\}, \quad (14)$$

$$\Delta\phi(v, a, c_1, c_2) = \{\Delta\Phi(v, a, c_1), \Delta\Phi(v, a, c_2)\}. \quad (15)$$

For this type of stimuli, is defined the angle contrast from the search target to two set of distractors, represented as contrasts from the first set "c₁" and the second

set " c_2 " in Equation 15, being the maximum angle between distractors and search target (considering that bars have angle values on two quadrants for each case) as 90° . Our target angle will have values taken from $\Phi(1, 90)$ being parameterized with contrast values ranging from 0 to 90° , in order to reveal higher angle contrast values, as heterogeneous distractors are harder to be identified.

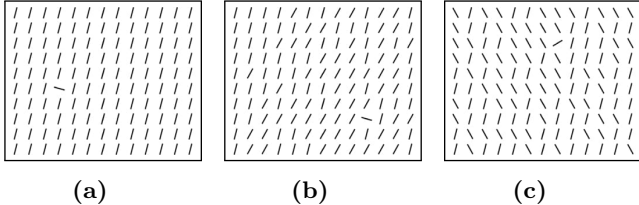


Figure 12. Examples of distinct distractor angle configurations, corresponding to (a) Homogeneous, (b) Tilted-right and (c) Flanking.

Distractor Linearity (14) Orientation collinearity facilitates visual guidance when orientation of target differs from its neighbors, making search efficient and in parallel [Nothdurft, 1993b] [Nothdurft, 1993a] [Wolfe, 2010]. Visual guidance is induced by orientation linearity given an array of bars as defined from the previous stimulus type. Each bar has been oriented with a specific angle, creating a nonlinear pattern for the whole search array (Figure 13). A linear case has also been presented to compare the conspicuity baseline from the other cases.

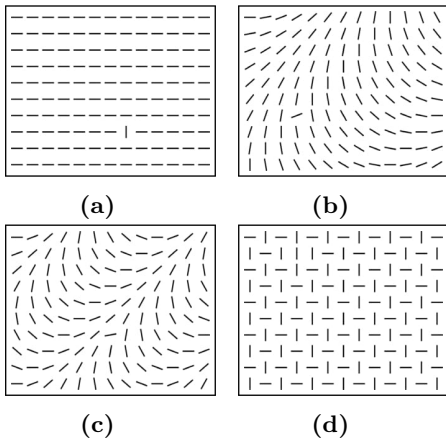


Figure 13. Examples of elicited guidance according to distinct linearity of the distractors. In (a), distractors are set using the same angle contrast of 90° with respect to the target. Conversely, in (b), (c) and (d), nonlinear patterns are set at an accumulative slope of 10, 20 and 90° respectively.

$$\varphi(u, row, col) = u \cdot row + u \cdot col, \quad (16)$$

$$\Delta\varphi(v, u, row, col) = \Phi(v, 0) + \varphi(u, row, col). \quad (17)$$

Angle contrast is calculated as the orientation difference from the corresponding value of nonlinearity pattern " u " at a certain position on the array " row, col " with a maximum angle contrast with respect to search target (Equations 16 and 17).

Distractor Categorization (15) Visual search for an oriented bar can be inefficient with distractors at 2 different orientations [Gao, 2008] [Wolfe, 2010]. However, it was found that not all orientations present on an image are equally coded in pre-attentive vision, different configurations of the orientations of the target and the distractions lead to different discriminability [Treisman, 1991]. Some of these orientation configurations were categorized as "steep", in which search target is identified more efficiently. Other categories of heterogeneous distractors presented harder target search and were dependent on set size. The three categories were modeled, corresponding to "steep", "steepest", "steep-right" as defined by Wolfe et. al. [Wolfe, 1992]. By considering the same orientation contrast between the two sets of angles, target orientation was parametrized in order to reveal at which orientation contrast is the target to both types of distractors that form these categories. We have modeled these three orientation configurations for each distractor pair, corresponding here to $-50, 50^\circ$ for steep, $-30, 70^\circ$ for steepest, $20, 80^\circ$ for steep-right (Figure 14). There was the same amount of distractors for each condition as shown for search on Distractor Heterogeneity (13) and Distractor Linearity (14) in order to uniquely analyze orientation contrast and preserving similar stimulus type conditions. As shown in section Distractor Heterogeneity (13), the orientation values for each set are computed with Equation 14 and the contrast with respect to the target as Equation 15. The maximum orientation contrast was calculated for all conditions at 40° ($v = 90/40$) considering the interference of bar orientation contrast in all quadrants (between the target and both distractor orientations). Target angle had psychometric values of $\phi(v, 90, -50, 50)$ for steep, $\phi(v, 90, -30, 70)$ for steepest and $\phi(v, 90, 20, 80)$ for steep-right.

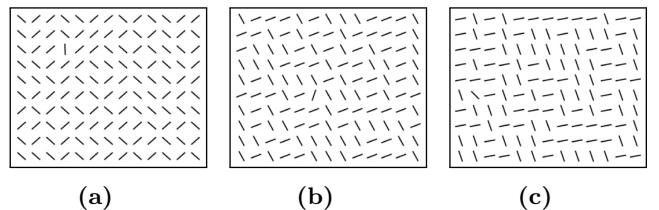


Figure 14. Examples of distinct distractor angle configurations, corresponding to (a) Steep, (b) Steepest and (c) Steep-right.

2.5 Data Analysis

In order to get the spatial relevance of participant’s eye movements we generated binary maps from fixation coordinates. Fixation density maps are computed with a symmetric Gaussian low-pass filtering (with a window size of $[6\sigma \times 6\sigma]$) of the respective binary maps. A value of $\sigma = 1 \text{ deg}$ was used, as recommended by LeMeur and Baccino [LeMeur and Baccino, 2012], corresponding in our case to 40 pixels. The saliency index (SI) is a measure that relates the energy inside a specific region (that can be manually selected, such as a pop-out region) and the one outside that region.

$$SI(S_t, S_b) = \frac{S_t - S_b}{S_b}. \quad (18)$$

We adapted the metric from Spratling’s work [Spratling, 2012] in order to present positive values as a better representation of the SI (Equation 18). The distribution of fixations inside (S_t) and outside (S_b) the area of interest (AOI) will be extracted by cropping the fixation density map using the mask presented on Figure 15.

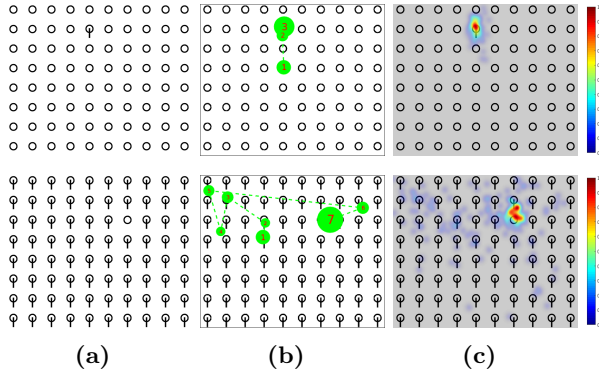


Figure 15. First row shows a grid with 8×10 circles, in which one of them becomes salient because of having a superimposed bar. On the second row, the superimposed bar is located instead on the rest of circles (distractors). (a) Representation of the mask, corresponding to the AOI of the search target in green (S_t) and the background in red (S_b). (b) Example of a scanpath of a single participant, representing each saccade with a green dashed line and each fixation number with a diameter corresponding to its fixation duration. (c) Superposed density map from the accumulation of fixations for all participants for such stimuli. The colorbar represents the probability of the density distribution.

For evaluating the SI for a specific sample, a binary visual mask of the salient region (or AOI) needs to be manually created. Given samples at distinct fixation or saccade number, it is possible to compute gaze-wise SI in order to evaluate the temporal evolution of that measure.

Such metric can provide a gold standard of spatial performance in terms of how a region pops out with respect to the rest using fixation density maps from recorded eye movements. In other words, measuring the SI using the fixation density maps is the same as measuring the distribution of fixations that have been recorded inside a particular region of an image. Same parameters of the SI metric are preserved from previous studies [Soltani and Koch, 2010]. Although other parameters (such as mask area) could be included to better represent the data, this could be a metric to be exploited in future studies.

In order to get the performance of participants on salient region localization, we recorded the reaction time (RT) on landing inside the AOI. For the free-viewing tasks, we recorded this RT from the initial fixation until the gaze landed inside the AOI. Once a fixation was outside the AOI, we recorded the time until the gaze returned to the AOI, being in this case produced by inhibition of return (IOR) mechanisms. For the visual search tasks, we recorded in a similar way the first fixation inside the AOI as well as visual discrimination. For this latter case, dwell fixations were pinpointed as being inside the AOI during 1000 ms in order to report identification of search targets. For the cases in which participants could not find the stimulus target, the RT corresponded to key pressing. We used fixation data for reporting target localization on both free-viewing and search tasks and the dwelling method for reporting target identification for visual search tasks. In that way, it is possible to discard non-representative fixations and saccades that could be present by other methods such as key trigger, that could imply spatial and temporal deviations with respect to both visual localization and identification. Given an image where salient regions are known, if the SI and the RT reproduce similar results at distinct tasks, feature contrasts and stimulus types, the SI could provide a way to spatially measure how salient is an object, considering specific regions as pop-out instead of using fixations across the whole scene as ground truth. The usage of eye tracking experiments and regions of interest for calculating localization RT instead of keyboard triggers reveals a more accurate way for evaluating visual attention, as no temporal delays are presented from the time since the participants see the search target to report that they have seen it. That method also allows to prevent them to attend to other regions outside the experimental source over time, such as looking towards the keyboard, which can impair their perceptual adaptability (in terms of light sensitivity and foveation).

3 Results

A total of 90,100 fixations were recorded over approximately 30 hours of viewing time. The mean number of fixations per stimulus was $M = 12 \pm 1$, corresponding to $M = 15 \pm 1$ for free-viewing (given from 5000 ms of viewing time) and $M = 11 \pm 1$ for visual search task stimulus (given from the total viewing time until the stimulus trigger, corresponding to target identification). Mean fixation duration was $M = 240 \pm 1$ ms and it was not presenting significant differences from the two types of tasks. See that both distributions of Fixation Duration (FD) and Saccade Amplitude (SA) (Figure 16) were skewed to lower values with their upper and lower quartiles at approximately 100 and 300 ms for FD and 2 and 5 deg of SA. We have also plotted the CDF for both variables and results show that most eye movements (80%) have a FD of less than 300 ms and SA tend to be shorter than 10 deg.

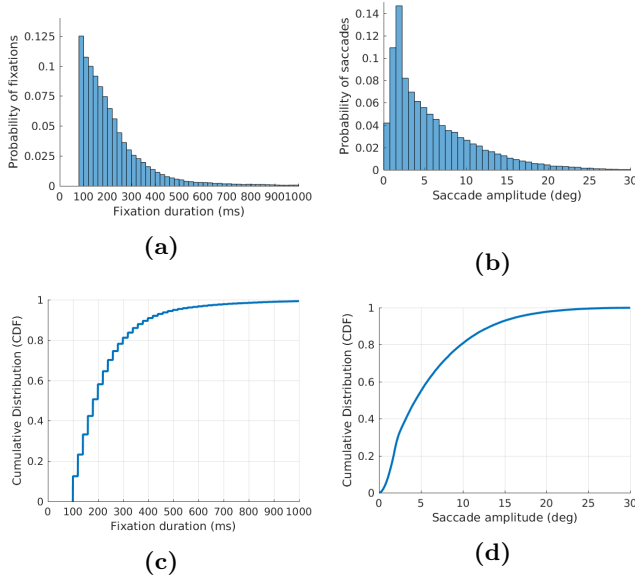


Figure 16. (a) Distribution of Fixation Duration (FD), measured as the absolute fixation time for all samples upon the probability of fixations. (b) Distribution of Saccade Amplitude (SA), measured as the absolute euclidean distance between saccade initiation and saccade landing all samples upon the probability of saccades. (c,d) Cumulative Distribution Functions for FD and SA.

The overall number of fixations was larger for images containing less salient regions, categorized as hard, requiring more fixations for participants in stimulus with less feature contrast. Localization probabilities were calculated, based on the scanpaths in which participants' gaze landed inside the corresponding AOI. Our results report easiest targets more probable to be localized for both free-viewing

($p=6.2 \times 10^{-4}$, $Z=3.4$, $P_{easy}=0.38$, $P_{hard}=0.30$) and visual search tasks ($p=6.9 \times 10^{-88}$, $Z=19.9$, $P_{easy}=0.72$, $P_{hard}=0.47$). After calculating the reaction times for target localization (landing inside the AOI) and identification (reporting presence of target), we discarded samples where $RT > 2\sigma_{RT}$. In that manner we could counteract the impact from oculomotor biases in relation to the localization time with respect objects with approximately that size. As the search targets were smaller for most visual search stimuli, hence less dependent to their respective distance from the stimulus center, we did not discard the respective samples.

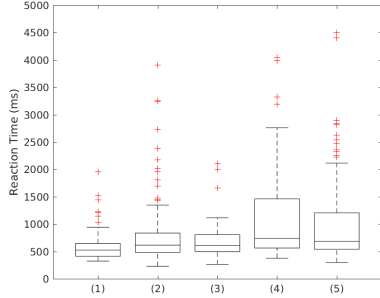
3.1 Performance upon Feature Type (1st Hypothesis)

RTs for AOI localization are evaluated for each stimulus type and task respectively. Since overall data do not follow a normal distribution (through lilliefors test), Kruskal-Wallis tests were performed in order to evaluate task differences for each contrast difficulty (easy vs hard) as well as differences in RT between distinct type of stimuli. As feature contrasts follow distinct contrast values, we want to test if some features have similarities in RT and their interactions. For each stimulus type, the RT is different given the feature type for both free-viewing (Figure 17) and visual search task stimuli (Figure 18).

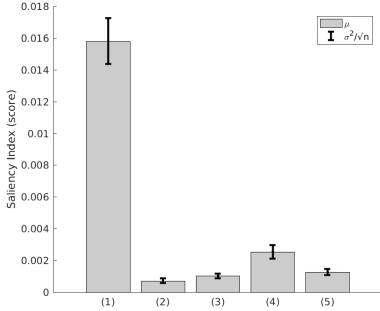
For the former, there were significant differences ($p=1.00 \times 10^{-10}$, $\tilde{\chi}^2=52.7$, $Mdn_{(1)}=523$, $Mdn_{(2)}=615$, $Mdn_{(3)}=604$, $Mdn_{(4)}=736$, $Mdn_{(5)}=684$ ms) between distinct stimulus types RTs, being Corner Angle (1) the fastest stimulus to localize the salient region and Contour Integration (4) the slowest.

For the latter, there were significant differences ($p=1.11 \times 10^{-74}$, $\tilde{\chi}^2=372$, $Mdn_{(6)}=782$, $Mdn_{(7)}=742$, $Mdn_{(8)}=942$, $Mdn_{(9)}=892$, $Mdn_{(10)}=593$, $Mdn_{(11)}=787$, $Mdn_{(12)}=622$, $Mdn_{(13)}=606$, $Mdn_{(14)}=676$, $Mdn_{(15)}=952$ ms) on RTs for searching salient regions, showing highest performance for Orientation Contrast (12) and Distractor Categorization (15) the lowest.

By computing the saliency index from the density maps across all fixations and the stimulus masks, it is possible to spatially evaluate saliency (in terms of number of fixations inside the window, represented as a heat map isotropically distributed using a Gaussian filter) given from each stimulus types.



(a)



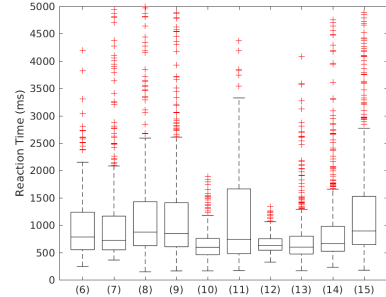
(b)

Figure 17. Plots for salient region localization time **(a)** and saliency index **(b)** corresponding to stimulus types of Corner Angle (1), Segmentation by Angle (2), Segmentation by Distance (3), Contour Integration (4) and Perceptual Grouping (5)

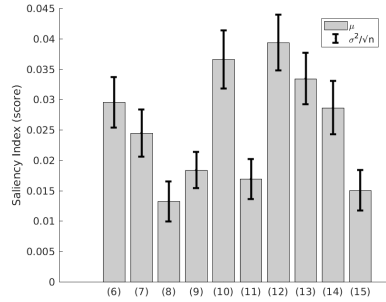
Similarly, there were significant differences on SI depending on stimulus types for free-viewing ($p=3.5 \times 10^{-7}$, $\tilde{\chi}^2=36$, $Mdn_{(1)}=1.69 \times 10^{-2}$, $Mdn_{(2)}=6.7 \times 10^{-4}$, $Mdn_{(3)}=1.1 \times 10^{-3}$, $Mdn_{(4)}=2.2 \times 10^{-3}$, $Mdn_{(5)}=1.2 \times 10^{-3}$ and visual search ($p=4.9 \times 10^{-6}$, $\tilde{\chi}^2=41$, $Mdn_{(6)}=32 \times 10^{-3}$, $Mdn_{(7)}=1.4 \times 10^{-2}$, $Mdn_{(8)}=8.0 \times 10^{-3}$, $Mdn_{(9)}=1.8 \times 10^{-2}$, $Mdn_{(10)}=4.0 \times 10^{-2}$, $Mdn_{(11)}=1.6 \times 10^{-2}$, $Mdn_{(12)}=4.0 \times 10^{-2}$, $Mdn_{(13)}=3.9 \times 10^{-2}$, $Mdn_{(14)}=3.1 \times 10^{-2}$, $Mdn_{(15)}=13 \times 10^{-3}$). Stimulus with higher SI for free-viewing task was Corner Angle (1) and the lower was Visual Segmentation (2-3). For the case of visual search task stimuli, most salient targets were on stimulus presented on Size (12) and Orientation (13) contrast and the least ones on Noise/Roughness (8) and Distractor categorization (12) search.

Given the aforementioned results shown for Figures 17 and 18, RTs were lower (faster) for stimuli with higher SI. The reverse case applies for lower RTs. Target identification (when participants voluntarily report to identify the search target, as explained in Section 2.5) was shown to be slower than target localization ($p < 1.3 \times 10^{-111}$, $Z = -22.4$, $Mdn_{localization} = 726$, $Mdn_{identification} = 1026$ ms), sup-

porting the literature [Sagi and Julesz, 1984] [Nothdurft, 2006a] with an absolute mean time difference of $M=415 \pm 203$ ms.



(a)



(b)

Figure 18. Plots for salient region localization time **(a)** and saliency index **(b)** corresponding to stimulus types of Feature and Conjunctive Search (6), Search Asymmetries (7), Noise/Roughness (8), Color Contrast (9), Brightness Contrast (10), Size Contrast (11), Orientation Contrast (12), Distractor Heterogeneity (13), Distractor Linearity (14) and Distractor Categorization (15).

Discussion

We can observe that saliency is induced through varying distinct features of the images. Fixations from participants are shown to localize salient regions significantly with distinct performance depending on feature type and the amount of fixations are distributed or spread distinctively across these regions. These aforementioned observations might be influenced by distinct processing (and correlates) of the visual features in the HVS.

3.2 Performance upon Feature Contrast (2nd Hypothesis)

Measures of RTs and SI for salient region localization were computed for each stimulus target-distractor contrast. Overall RT data was not normally distributed, but individ-

ual data per stimulus type was normally distributed. Mean RT and error is represented according to the stimulus contrast as well as its mean SI. Spearman’s rank correlation tests show that there was a significant negative correlation between RT and SI ($\rho_{RT,SI}=-.44$, $p_{RT,SI}=2.2 \times 10^{-195}$), suggesting that SI is a plausible measure for representing saliency on a particular region (higher SI and lower RT implies faster localization speed). In that respect both RT and SI were related to stimulus feature contrast (CT) measurements (shown on Section 2.4 and Figures 19,20 and 21). RT was negatively correlated with respect to CT ($\rho_{CT,RT}=-.14$, $p_{CT,RT}=7.1 \times 10^{-21}$). Conversely, SI was correlated with CT $\rho_{CT,SI}=.05$, $p_{CT,SI}=3.4 \times 10^{-3}$). These results show that both measurements were satisfying the Weber Law (RT decreasing with higher CT and SI increasing with respect CT). Individual results for correlations between each contrast measurement satisfy for most cases the aforementioned relationships between CT and RT as well as CT and SI, presented in Table 4.

Table 4. Table of correlations between contrast values (3rd column) with Reaction Time (4th column), or with Saliency Index (5th column)

Feature type	Contrast (CT)	ρ_{RT}	ρ_{SI}
(1) Corner Angle	Slope()	.23*	.53*
(2) Segment. Angle	Angle, $\Delta\Phi()$	-.33*	.11
(3) Segment. Spacing	Spacing(deg)	.65*	-.37*
(4) Contour Integration	Length(deg)	-.25*	-.35*
(5) Perc. Grouping	Distance(deg)	.29*	-.06
(6) Feat. & Conj. Search	Set Size(#)	.15*	-.12*
(7) Search Asymmetries	Set Size(#)	-.33*	-.39*
(8) Noise/Roughness	Freq., $1/f^\beta$	-.56*	.53*
(9) Color Contrast	Sat., $\Delta S_{D,T}$	-.57*	.48*
(10) Brightness Contrast	Light., $\Delta L_{D,T}$	-.41*	.25*
(11) Size Contrast	Size(deg)	-.55*	-.29*
(12) Orientation Contrast	Angle, $\Delta\Phi()$	-.18*	.05
(13) Distr. Heterogeneity	Angle, $\Delta\Phi_{1c}()$	-.04	.17*
(14) Distr. Linearity	Angle, $\Delta\Phi()$	-.07	.01
(15) Distr. Categorization	Angle, $\Delta\Phi_{1c}()$	-.24*	.23*

*: $p < .05$

We have plotted the relationships between RT and CT as well as for SI and CT in order to see how CT varies localization performance for each stimulus feature type individually (Figures 19,20 and 21). In these figures we can observe (in relation to Table 4) which feature targets are perceived in parallel or require a serial ‘binding’ step.

On (1-5) the Weber law applies for stimulus such as Corner Angle, showing slower localization on smoother corners than sharper ones. For Visual Segmentation stimuli, segment localization was faster to be localized when segment angle had a diagonal segment for both single and superimposed segments (due to its own corner angle with respect to other segment bars), being single ones with a trend to be more salient ($p=1.2 \times 10^{-2}$, $\tilde{\chi}^2=6.3$). Segments with 1.5 deg of segment distance and 2.5 deg

of bar length showed faster localization rate compared to wider segments. The Weber law applied as well for contour detection, being larger contours faster to be localized. For Perceptual Grouping, similar shape distractors showed slower localization rates as grouping distance is increased (lower proximity), but it was not so evident for dissimilar distractors, being localized faster and with overall higher SI. The Weber law did not apply for this case, suggesting that at a certain proximity distance (about approximately 5.5 deg) participants fixated into several regions, making them similarly salient. SI results on Corner Angle and Contour Integration had positive correlations with respect RT (contradicting the general case). That would be caused by the size of the masks (from stimulus salient objects), which would be higher for higher stimulus contrasts, with decreasing absolute SI (bigger masks would require more fixations when considering the same spatial conditions). In that aspect, SI must be evaluated considering that the size of the mask is constant, which is not the case for Corner Angle (1) and Contour Integration (4). These center biases might be one of the reasons for the Weber law appliance (presenting less agreement on RT and SI continuity upon feature contrast) as endogenous visual guidance can generate higher inter-participant differences. A more continuous slope for RT and SI observed for stimulus feature contrasts could be acquired by using an onset cue and a constant distance between the initial fixation and the stimulus target, but that method could generate oculomotor biases with respect to the possible positions distinct from the center (that could also vary the temporality of the fixations with respect to the center distance). An alternative solution that would partly solve the problem (as distance from the initial fixation and the stimulus target would still not be totally constant) would be to acquire a larger amount of observations at distinct randomized regions for each stimulus contrast and stimulus type [Wolfe et al., 2010] [Wolfe et al., 2011].

Feature search show a faster localization of the target than conjunctive search (Figure 20), with an almost constant RT with respect to set size (features processed in parallel). Conjunction search reveal slower localization of stimulus targets ($p=2.5 \times 10^{-24}$, $\tilde{\chi}^2=104$) as we increase distractor number (consequently, features being shown to be processed in a serial manner), likewise with lower SI. Similarly, reporting stimulus absence presented similar response time distributions, presenting conjunctive distractors to be more uncertain for reporting absence ($p=1.9 \times 10^{-36}$, $\tilde{\chi}^2=159$) than feature search ones. Searching a target circle among circle distractors with a superimposed bar show lower performance at increasing scale and set size ($p=6.2 \times 10^{-33}$, $\tilde{\chi}^2=143$), reversely, searching a target circle with a superimposed bar among circles shows more constant performance, revealing that search asymme-

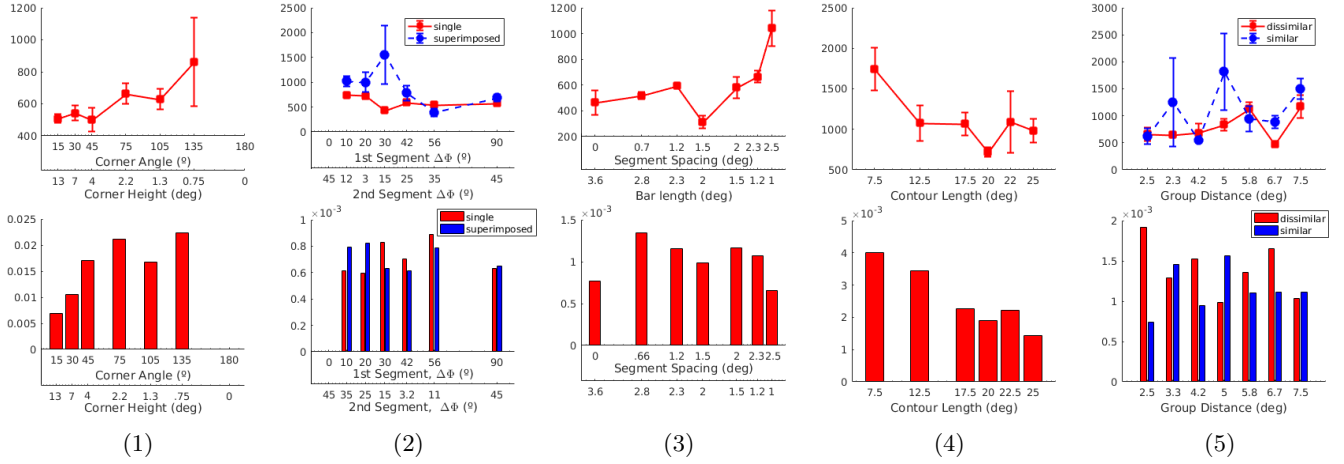


Figure 19. Plots of Reaction Times (top row) and Saliency Index (bottom row). Spearman's rank correlation tests were performed between RT and SI from each stimulus type and participant individually, corresponding on each case to **Corner Angle (1)**: $\rho_{(1)}=8.3 \times 10^{-2}$, $p_{(1)}=.43$, **Visual Segmentation (2,3)**: $\rho_{(2)}=-.22$, $p_{(2)}=5.6 \times 10^{-3}$; $\rho_{(3)}=-5.7 \times 10^{-4}$, $p_{(3)}=.99$, **Contour Integration (4)**: $\rho_{(4)}=-5.1 \times 10^{-2}$, $p_{(4)}=.61$ and **Perceptual Grouping (5)**: $\rho_{(5)}=-.13$, $p_{(5)}=.12$. For this cases, we have discarded samples in which participants had a fixation closer than 5 degrees of eccentricity from the search target, corresponding to the higher visual acuity of the fovea [Strasburger et al., 2011] [Wandell, 1995], as the RT calculation could be impaired by center biases.

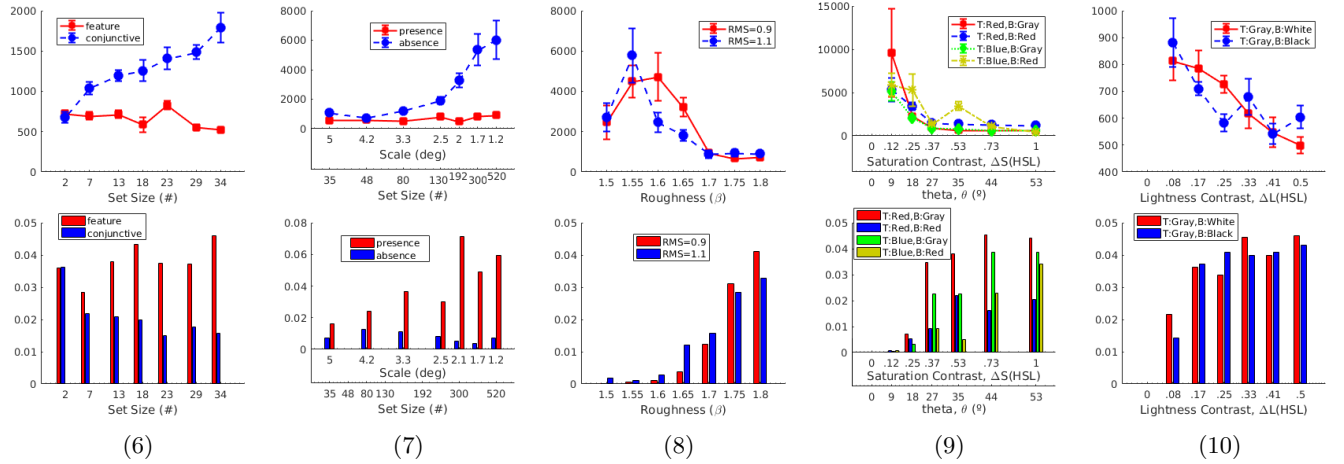


Figure 20. Plots of Reaction Times (top row) and Saliency Index (bottom row). Spearman's rank correlation tests were performed between RT and SI from each stimulus type and participant individually, corresponding on each case to **Feature and Conjunction search (6)**: $\rho_{(6)}=-.59$, $p_{(6)}=4.6 \times 10^{-36}$, **Search Asymmetries (7)**: $\rho_{(7)}=-.45$, $p_{(7)}=3.3 \times 10^{-9}$, **Noise/Roughness (8)**: $\rho_{(8)}=-.68$, $p_{(8)}=5.5 \times 10^{-33}$, **Color Contrast (9)**: $\rho_{(9)}=-.69$, $p_{(9)}=1.5 \times 10^{-72}$ and **Brightness Contrast (10)**: $\rho_{(10)}=-.51$, $p_{(10)}=3.4 \times 10^{-23}$.

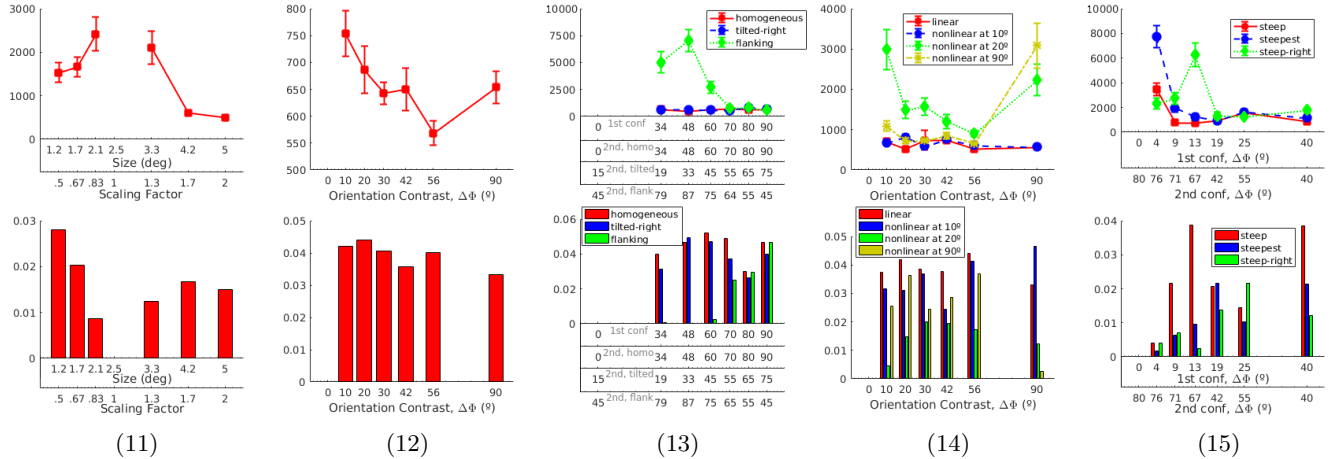


Figure 21. Plots of Reaction Times (top row) and Saliency Index (bottom row). Spearman’s rank correlation tests were performed between RT and SI from each stimulus type and participant individually, corresponding on each case to **Size Contrast (11)**: $\rho_{(11)}=-.14$, $p_{(11)}=9.7 \times 10^{-2}$, **Orientation Contrast (12)**: $\rho_{(12)}=-.41$, $p_{(12)}=2.7 \times 10^{-8}$, **Distractor Heterogeneity (13)**: $\rho_{(13)}=-.57$, $p_{(13)}=3.5 \times 10^{-41}$ and **Distractor Linearity (14)**: $\rho_{(14)}=-.57$, $p_{(14)}=2.1 \times 10^{-53}$ and **Distractor Categorization (15)**: $\rho_{(15)}=-.66$, $p_{(15)}=2.9 \times 10^{-59}$.

tries for this case apply. SI also reveals search asymmetries with respect these two types of stimuli, however, the SI is lower for the former case.

The Weber law is present for the case of background roughness, showing a decrease in search performance and SI at low beta values (rougher surfaces). Both conditions of height deviation ($\sigma_{RMS}=0.9, 1.1$) present similar performance with both metrics, with a trend of better search efficiency for higher RMS values. When searching a target with higher saturation contrast with respect distractors, both search performance and SI is higher than with lower saturation contrasts. Background conditions present a trend to drive search asymmetries, showing faster localization RTs and SI for unsaturated backgrounds. In that aspect, achromatic backgrounds presented faster localization for both red ($p=3.8 \times 10^{-9}$) and blue distractors ($p=2.1 \times 10^{-4}$). SI is shown to be higher for red hue in contrast to blue hue for search target and distractors. Lightness contrast also conforms with the Weber law, similarly to saturation contrast but with higher overall performance. Lighter backgrounds with darker search targets present a trend to have higher SI with respect to darker backgrounds with lighter salient objects.

Results on size similarity reveal increased search efficiency with respect to size contrast, with a tendency of perceiving bigger objects as more salient than smaller ones for both localization time and saliency index as in Figure 21. Similarity on orientation also shows increased search efficiency with respect angle contrast, with diagonal angles localized faster than vertical or horizontal ones. Orientation contrast has been found to have high search efficiency,

specially with diagonal angles and homogeneous angle organization for distractors. In contrast, heterogeneous set of distractor angles present a lower search efficiency with respect to the homogeneous ones. Homogeneous distractors were significantly localized faster for heterogeneity at distinct angle quadrant configurations (flanking) $p=1.0 \times 10^{-9}$ but not for heterogeneous distractors with angle configurations at the same quadrant (tilted-right) $p=.63$. Another distinct type of orientation-related guidance is distractor linearity, presenting differences depending on each slope condition ($p=1.3 \times 10^{-35}$, $\chi^2=165$). Non-linear orientation patterns at a slope increment of 20° present lower search efficiency than the ones at 10° and 90° . The latter case presents a slightly lower search efficiency at vertical or horizontal orientations due to its similar orientation interactions between the target and one of the distractor sets. Results suggest that both the amount of distractor sets and each of their orientation contrasts with respect to search target might be the source of overall distinctiveness for non-linear orientation patterns. Results related with orientation pattern categorization report overall higher SI and a trend for faster localization rate for steep orientation organization than steepest ($p=8.4 \times 10^{-2}$) and significant with respect to steep-right ($p=1.2 \times 10^{-5}$), confirming that search asymmetries apply for this case considering that three conditions possess the same orientation contrast between the two distractor sets.

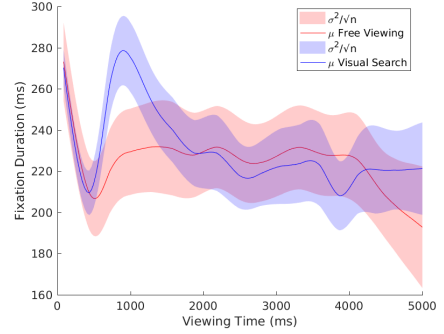
Discussion

In this study we show that feature contrast is correlated to saliency using distinct measures and feature types, being saliency higher at higher feature contrasts. By using visual search tasks and synthetic images, there is a better control of exogenous cues by reducing endogenously-dependent guidance. It is about to consider that the SI is a good measure for evaluating saliency for specific areas of interest.

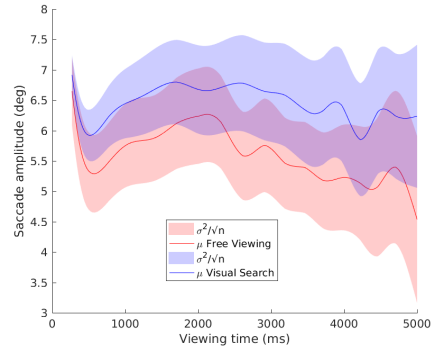
3.3 Attention changes nonlinearly over time (3rd Hypothesis)

Values of FD and SA were grouped for each gaze as functions of viewing time. In Figure 22(a,b), during the first 1 to 2 seconds, fixations have a larger duration for visual search tasks. For the visual search task, fixations have a duration with a peak at 274ms during the beginning of the experiment and progressively drop during the end of the stimulus view to 217 after 5000 ms of viewing time. In free-viewing tasks, FD remains stable after the first and second fixation at approximately 202 ms. For the SA on both tasks there is a peak for the first saccade between 6.5 and 7 deg. During the first and second gaze, SA drops to a value between 5.5 and 6 deg and increase during 1 second to amplitudes between approximately 6 and 6.5 deg. During the last gazes, after 2 seconds of viewing time, SA progressively drops during the rest of viewing time. Such behavior occurs similarly for both visual search and free-viewing cases, these patterns might also be related to endogenous factors commented previously. These distinct eye movement patterns might be related to how participants approach targets depending on task priors and show an overview of how relevant is to account for temporal properties when evaluating eye movements.

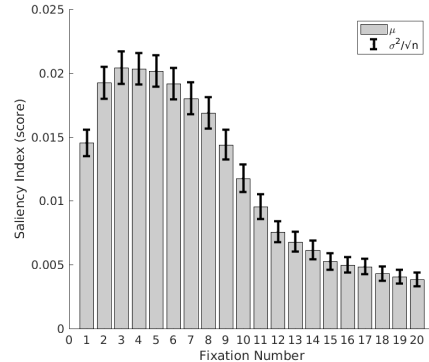
SI was computed using the density maps across fixation number Figure 22(c), it decreases with respect to fixation number, being the first fixations (from the 1st to the 5th) the ones that have higher SI (accounting for fixations inside the salient region). Inhibition of return (IOR) mechanisms might be responsible for the aforementioned effects. IOR was present and we believe that it may have influenced both types of tasks. To know that, mean return saccade time was computed, corresponding to the time spent from the first fixation inside the AOI to the second fixation that returned inside the AOI, which was $M=16.6 \pm 0.9 \times 10^2$ ms, corresponding to $M=14.1 \pm 0.5 \times 10^2$ ms for Free-Viewing and $M=18.6 \pm 1.5 \times 10^2$ ms for Visual Search tasks respectively.



(a)



(b)



(c)

Figure 22. (a) Temporal evolution (from 0 to 5000 ms) of fixation duration. (b) Temporal evolution (from 0 to 5000 ms) of saccade amplitude. For both plots, samples corresponding to free-viewing task fixations and saccades are represented in red and blue for the case of Visual Search. (c) Mean saliency index upon fixation number.

Discussion

The temporal evolution of fixation and saccade behavior reveal distinct patterns of eye movements upon viewing time, confirming the evidence that visual attention is an active process and its modeling involving temporality requires further investigation. Scanpath prediction

could allow the reproduction of the aforementioned effects, regarding in that aspect both bottom-up and top-down processing of visual features that distinctively guide visual attention [Boccignone and Ferraro, 2004] [Kubota et al., 2012] [Chang et al., 2014] [LeMeur and Liu, 2015] [Aboudib et al., 2015] [Adeli et al., 2016] [Wang et al., 2016] [Wloka et al., 2017] [White et al., 2017]. In that aspect, as saliency decreases over time, saliency evaluation measures should be done in that line.

3.4 Task influences perceived attention (4th Hypothesis)

Distinct eye movement behavior in terms of FD and SA was presented depending on each task type (Section 3.3). Task priors also influenced the localization performance in relation to feature contrast.

First, Wilcoxon signed-rank tests were performed to evaluate the amount of fixations between easy and hard targets and was found to be lower for easy than for hard targets in the visual search task ($p=2.1 \times 10^{-147}$, $Z=-26$, $Mdn_{easy}=4$, $Mdn_{hard}=7$), but there was no difference for the case of the free-viewing task ($p=.069$, $Z=-0.1$, $Mdn_{easy}=15$, $Mdn_{hard}=16$). There were differences in FD between the easy and hard targets for visual search ($p=3.6 \times 10^{-36}$, $Z=13$, $Mdn_{easy}=199$, $Mdn_{hard}=179$ ms), but it was not occurring for free-viewing tasks ($p=.57$, $Z=.57$, $Mdn_{easy}=199$, $Mdn_{hard}=199$ ms). Same phenomena was presented for SA, in which there was a significant difference depending on the stimulus contrast difficulty for visual search ($p=1.7 \times 10^{-56}$, $Z=-16$, $Mdn_{easy}=3.9$, $Mdn_{hard}=4.7$ deg) but not for the case of free-viewing ($p=.069$, $Z=-1.8$, $Mdn_{easy}=3.9$, $Mdn_{hard}=4.1$ deg). These results evidence less dependence from low-level feature contrasts for free-viewing tasks in contrast to visual search tasks, acknowledging that participants are not always exogenously guided to gaze towards salient regions for free-viewing tasks, namely, that endogenous factors are prevailing more in this kind of task, making saliency less accurate spatially and temporally.

Second, we observed the correlations of RT, SI and feature contrast (FC), described in Section 3.2. Here we define FC as ψ values for considering a generalized feature contrast, as CT values vary between blocks, but FC values do not. For visual search stimuli, RT was negatively correlated with SI ($\rho_{RT,SI}=-.59$, $p_{RT,SI}=.00$), FC was negatively correlated with RT ($\rho_{FC,RT}=-.08$, $p_{FC,RT}=6.4 \times 10^{-7}$) but positively correlated with SI ($\rho_{FC,SI}=.05$, $p_{FC,SI}=2.4 \times 10^{-3}$). For free-viewing stimuli, there was a significant negative correlation between RT and SI ($\rho_{RT,SI}=-.16$, $p_{RT,SI}=1.2 \times 10^{-4}$), a negative correlation between FC and RT ($\rho_{FC,RT}=.26$, $p_{FC,RT}=6.5 \times$

10^{-10}) but the relationship with respect feature contrast and SI was non-significant ($\rho_{FC,SI}=-.04$, $p_{FC,SI}=.32$). Distinct behavior is presented on the regression lines shown in Figure 23 the relationships from RT and SI with respect to CT (here represented as a unique contrast value, although calculated for each contrast measurement separately as in Table 4) for both tasks.

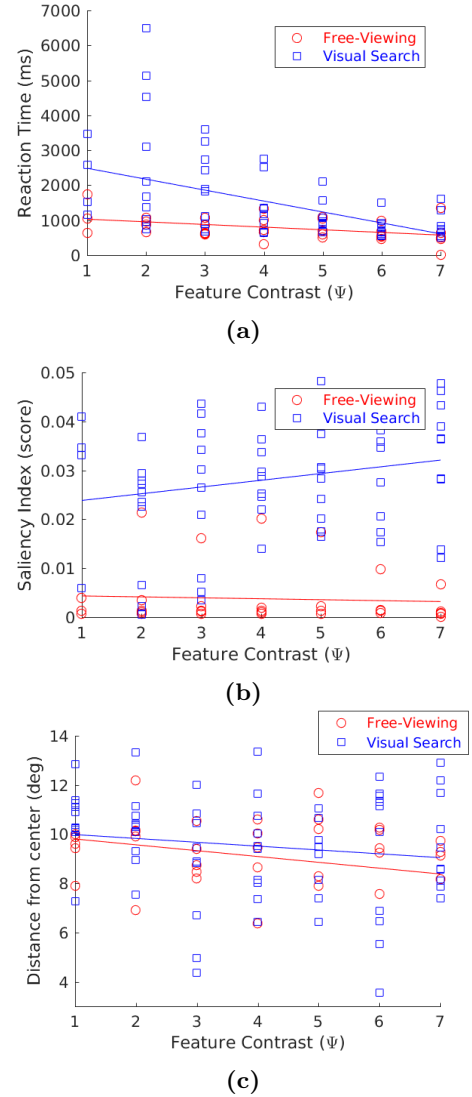


Figure 23. Scatter plots of Reaction Time (a), Saliency Index (b) and Distance from center (c) upon feature contrast (Ψ). We represented the mean of each feature type separately and we have plotted the regression line for both tasks.

Discussion

Salient region localization performance varies with respect to feature contrast depending on the task. Fixation

duration and saccade amplitude are affected more by stimulus contrast on Visual Search than Free-Viewing tasks. Moreover, the center bias seems to be more present for Free-Viewing tasks. Further analysis of interest would be the evaluation of absolute task differences in localization performance. In that respect, we could present the same stimuli with several observations for each feature contrast and distinct cueing, so that to see the absolute influences from endogenous guidance for each distinct feature type and contrast.

3.5 Center biases are endogenous (5th Hypothesis)

The center bias was represented by grouping fixations for all samples and representing the density map shown in Figure 25. From such baseline, it is possible to estimate the mean euclidean distance from every fixation to the baseline center (DC). This baseline shows increasing spreadity and area with respect to fixation number and consequently with respect time. In Figure 24 there is the DC as a function of viewing time (centroid was computed as a unique point corresponding to the initial fixation baseline). From this plot, we can observe that participants move their eyes away from the center of the stimulus after the first and second fixation, between 10 and 11 deg. After 2 seconds of viewing time, mean distance from baseline center is nearly constant for the visual search case but not for the free-viewing case. For the latter, fixations get become closer to the baseline center showing increasing patterns of center bias for this task, similarly to SI (Figure 22), which increases during the first fixations and drops on late fixations.

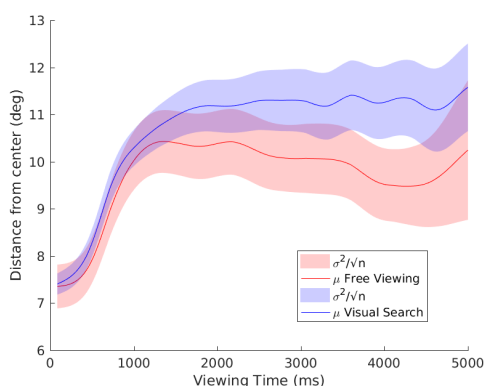


Figure 24. Representation of the center bias as the mean euclidean distance between fixation localization and the baseline center.

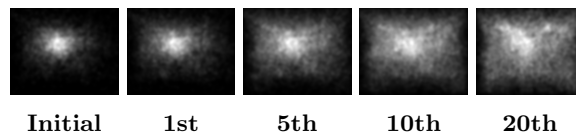


Figure 25. Representation of the density map for all fixations grouped together across all stimuli.

DC was negatively correlated with FD ($\rho=-.08$, $p=3.9 \times 10^{-134}$) and positively correlated with SA ($\rho=.08$, $p=1.4 \times 10^{-148}$). Here, short fixations and large saccades might be eye movement patterns highly related to saliency (as being negatively correlated to the center bias). By computing the mean per stimulus for the case of FC, it is possible to compare how DC was affected by feature contrast Figure 23c. For Free-Viewing task, DC was significantly negatively correlated with FC ($\rho=-.13$, $p=2.6 \times 10^{-3}$). For Visual Search task, DC was not significantly correlated with FC ($\rho=-.03$, $p=.07$). Acknowledging that stimulus targets were randomized, feature contrast was decreasing the center bias (increasing DC) more on Visual Search tasks than for Free-Viewing tasks, supporting the literature stated in Section 1.5.

Table 5. Table of correlations between Feature Contrast (FC) with Distance from baseline center (DC)

Feature type	$\rho_{\psi, DC}$
(1) Corner Angle	-.004
(2) Segment. Angle	-.32*
(3) Segment. Spacing	-.16
(4) Contour Integration	.012
(5) Perc. Grouping	-.07
(6) Feat. & Conj. Search	-.34*
(7) Search Asymmetries	.24*
(8) Noise/Roughness	-.29*
(9) Color Contrast	-.12*
(10) Brightness Contrast	.66*
(11) Size Contrast	-.31*
(12) Orientation Contrast	-.18*
(13) Distr. Heterogeneity	.21*
(14) Distr. Linearity	-.28*
(15) Distr. Categorization	-.03

*: $p < .05$

We have added in Table 5 the correlations between DC and FC for each feature individually. Most cases of singleton search (i.e. 6-15) show a significant negative correlation between DC and FC, meaning, when the feature contrast is higher, the center bias is lower.

Discussion

Short saccades and large fixation durations are shown to be correlated with eye movement behavior related to the center bias. Temporality of fixations show a non-linear evolution of the center bias, showing more dispersion with respect to viewing time. Moreover, distance from center

In that aspect, saliency would not only need to be evaluated by adjusting metric performances using metrics that account for the aforementioned center biases [Zhang et al., 2008] [Borji et al., 2013c] [Bethge et al., 2015] [Wloka and Tsotsos, 2016] [Nuthmann et al., 2017], but also upon the importance of temporality on fixation and saccade characteristics, by computing each metric upon gaze number on each stimulus fixational data. Thus, saliency metrics should account for feature contrast and minimize the contextual effects in order to accurately reproduce eye movement behavior.

4 General Discussion

Given the presented results, we emphasize that saliency is influenced by a variety of factors when observing eye movement behavior. In this study is presented a dataset considering all the aforementioned factors, by evaluating eye movements for distinct feature types, contrasts, temporality, task and representing the center biases. First, scene context (here defined as different feature types) is known to affect attention with specific performance, significantly determining efficiency of localizing and/or identifying salient regions. Second, saliency measures are shown to be correlated to feature contrast and distinctively depending on feature type. Third, fixation and saccade characteristics are presented to evolve non-linearly over time, making saliency decrease with respect saccade number and/or viewing time. Fourth, visual search tasks show higher performance in comparison to free-viewing on our saliency measurements and they have a higher correlation with respect saliency and feature contrast. Fifth, the central bias is shown to be correlated to short saccades and long fixation durations.

Eye movements are a behavioral output that imply processing of both endogenous and exogenous factors, namely, that have both top-down tuning and bottom-up interactions at different levels of the HVS. Thus, eye movement prediction might require recurrent processing of information from the ventral and dorsal pathways of the HVS, generating a unique representation for eye movement control (visual priority) [Lamme and Roelfsema, 2000] [Corbetta and Shulman, 2002] [Fecteau and Munoz, 2006]. If the unique factor to be evaluated is early saliency, stimulus in which features are processed fast and in parallel would be more relevant when evaluating eye movement prediction (showing less inter-participant differences as a consequence of higher SI), namely, the ones with salient regions that are reflectively selected and separated from the background (with higher contrasts with respect the rest of the scene).

Further considerations

Current literature acknowledges that temporal patterns of saccades have been shown to be fovea-dependent and lately classified as focal and ambient, being ambient fixations responsible for early saccades (sensitive to peripheral signals) and the latter for later saccades (being these ones foveal) [Unema et al., 2005] [Pannasch et al., 2008] [Follet et al., 2011] [Eisenberg and Zacks, 2016]. Similarly with saccade latencies, a bimodal latency distribution distinguishes regular from express saccades [Saslow, 1967] [Schiller et al., 1987] [Sommer, 1997] [van Zoest and Donk, 2006]. We have to acknowledge that the usage of an eye tracker with higher sampling rate (e.g. above 250 Hz) would improve accuracy in this type of experimentation, especially for a possible microsaccadic analysis. Distinct eye movement behavior is presented to be dependent as well for saccade length, pupil dilation and eye vergence [Privitera et al., 2014] [Miura et al., 2001] [Fallah and Reynolds, 2012] [Wang et al., 2014] [Puig et al., 2013]. All of these factors should be considered in future visual attention modeling considering their relationships with the two-stream hypothesis [VanEssen and Gallant, 1994] [Bell et al., 2013] [Trevarthen, 1968] [Sheth and Young, 2016] in order to specify the experimental conditions for a better evaluation of uniquely bottom-up visual attention.

Future work

Future experimentation for low-level feature analysis in eye movements would be to explore covert attention influences varying some of the presented feature contrasts at distinct eccentricities [Carrasco and Yeshurun, 1998] [Carrasco et al., 2006] [Carrasco, 2006]. Another observation of interest would be the evaluation of task differences in localization performance. In that respect, to present the same stimuli with several observations for each feature contrast and distinct cueing would reveal absolute influences from endogenous guidance. Our study could be extended by analyzing the influence of dynamic scenes on saliency modeling [Leboran et al., 2017] [Riche and Mancas, 2016b] using synthetic videos with both static or dynamic camera. In that direction, it would be able to see the interaction between low-level visual features and temporally-variant features. Another remark would be to see the impact of the target template search in comparison to the odd-one out type of tasks, in this case, but for stimuli with similar display conditions but distinct feature type.

Physiological evidence could provide an explanation for the low-level feature processing, including both bottom-up and top-down computations reproducing the presented effects not only spatially but temporally. Computations made by the visual cortex that process these low-level features (in reference to the mechanisms that respond

distinctively to color, orientation and spatial sensitivities as well as their interactions) might be responsible for most if not all of the effects presented in this study. Further analysis on mid and high-level features would require further study in terms of their relation to psychophysical effects on eye movements as well as their biological foundations [Kruger et al., 2013].

5 Acknowledgments

This work was funded by the Spanish Ministry of Economy and Competitiveness (DPI2017-89867-C2-1-R and TIN2015-71130-REDT), Agencia de Gestió d’Ajuts Universitaris i de Recerca (AGAUR) (2017-SGR-649), and CERCA Programme / Generalitat de Catalunya.

References

- [Aboudib et al., 2015] Aboudib, A., Gripon, V., and Coppin, G. (2015). A model of bottom-up visual attention using cortical magnification. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE.
- [Adeli et al., 2016] Adeli, H., Vitu, F., and Zelinsky, G. J. (2016). A model of the superior colliculus predicts fixation locations during scene viewing and visual search. *The Journal of Neuroscience*, 37(6):1453–1467.
- [Amano and Foster, 2014] Amano, K. and Foster, D. H. (2014). Influence of local scene color on fixation position in visual search. *Journal of the Optical Society of America A*, 31(4):A254.
- [Antes, 1974] Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, 103(1):62–70.
- [Bacon and Egeth, 1994] Bacon, W. F. and Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, 55(5):485–496.
- [Bauer et al., 1996] Bauer, B., Jolicoeur, P., and Cowan, W. B. (1996). Distractor heterogeneity versus linear separability in colour visual search. *Perception*, 25(11):1281–1293.
- [Bell et al., 2013] Bell, A. H., Pasternak, T., and Ungerleider, L. G. (2013). *The new visual neurosciences*. The MIT Press, Cambridge, Massachusetts.
- [Ben-Av and Sagi, 1995] Ben-Av, M. B. and Sagi, D. (1995). Perceptual grouping by similarity and proximity: Experimental results can be predicted by intensity autocorrelations. *Vision Research*, 35(6):853–866.
- [Ben-Av et al., 1992] Ben-Av, M. B., Sagi, D., and Braun, J. (1992). Visual attention and perceptual grouping. *Perception & Psychophysics*, 52(3):277–294.
- [Bethge et al., 2015] Bethge, M., Kümmerer, M., and Wallis, T. (2015). How close are we to understanding image-based saliency?
- [Bhatt et al., 2007] Bhatt, R., Carpenter, G. A., and Grossberg, S. (2007). Texture segregation by visual cortex: Perceptual grouping, attention, and learning. *Vision Research*, 47(25):3173–3211.
- [Boccignone and Ferraro, 2004] Boccignone, G. and Ferraro, M. (2004). Modelling gaze shift as a constrained random walk. *Physica A: Statistical Mechanics and its Applications*, 331(1-2):207–218.
- [Borji and Itti, 2013] Borji, A. and Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):185–207.
- [Borji and Itti, 2014] Borji, A. and Itti, L. (2014). Defending yarbus: Eye movements reveal observers’ task. *Journal of Vision*, 14(3):29–29.
- [Borji and Itti, 2015] Borji, A. and Itti, L. (2015). Cat2000: A large scale fixation dataset for boosting saliency research. *CVPR 2015 workshop on "Future of Datasets"*. arXiv preprint arXiv:1505.03581.
- [Borji et al., 2013a] Borji, A., Sihite, D. N., and Itti, L. (2013a). Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing*, 22(1):55–69.
- [Borji et al., 2013b] Borji, A., Sihite, D. N., and Itti, L. (2013b). What stands out in a scene? a study of human explicit saliency judgment. *Vision Research*, 91:62–77.
- [Borji and Tanner, 2016] Borji, A. and Tanner, J. (2016). Reconciling saliency and object center-bias hypotheses in explaining free-viewing fixations. *IEEE Transactions on Neural Networks and Learning Systems*, 27(6):1214–1226.
- [Borji et al., 2013c] Borji, A., Tavakoli, H. R., Sihite, D. N., and Itti, L. (2013c). Analysis of scores, datasets, and models in visual saliency prediction. In *2013 IEEE International Conference on Computer Vision*. IEEE.
- [Bruce et al., 2015] Bruce, N. D., Wloka, C., Frosst, N., Rahman, S., and Tsotsos, J. K. (2015). On computational modeling of visual saliency: Examining what’s right, and what’s left. *Vision Research*, 116:95–112.
- [Bruce and Tsotsos, 2005] Bruce, N. D. B. and Tsotsos, J. K. (2005). Saliency based on information maximization. In *Proceedings of the 18th International Conference on Neural Information Processing Systems, NIPS’05*, pages 155–162, Cambridge, MA, USA. MIT Press.
- [Buswell, 1935] Buswell, G. (1935). *How People Look at Pictures: A Study of the Psychology of Perception in Art*. University of Chicago Press.
- [Bylinskii et al., 2015] Bylinskii, Z., DeGennaro, E., Rajalingham, R., Ruda, H., Zhang, J., and Tsotsos, J. (2015). Towards the quantitative evaluation of visual attention models. *Vision Research*, 116:258–268.
- [Bylinskii et al.,] Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., and Torralba, A. Mit saliency benchmark. <http://saliency.mit.edu/>.
- [Caputo, 1997] Caputo, G. (1997). Object grouping contingent upon background. *Vision Research*, 37(10):1313–1324.
- [Carrasco, 2006] Carrasco, M. (2006). Covert attention increases contrast sensitivity: psychophysical, neurophysiological and neuroimaging studies. In *Visual Perception - Fundamentals of Vision: Low and Mid-Level Processes in Perception*, pages 33–70. Elsevier.
- [Carrasco et al., 2006] Carrasco, M., Giordano, A. M., and McElree, B. (2006). Attention speeds processing across eccentricity: Feature and conjunction searches. *Vision Research*, 46(13):2028–2040.
- [Carrasco and Yeshurun, 1998] Carrasco, M. and Yeshurun, Y. (1998). The contribution of covert attention to the set-size and eccentricity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2):673–692.
- [Castelhano et al., 2009] Castelhano, M. S., Mack, M. L., and Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3):6–6.
- [Chang et al., 2014] Chang, H.-C., Grossberg, S., and Cao, Y. (2014). Where’s waldo? how perceptual, cognitive, and emotional brain processes cooperate during learning to categorize and find desired objects in a cluttered scene. *Frontiers in Integrative Neuroscience*, 8.

- [Chun and Jiang, 1998] Chun, M. M. and Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36(1):28–71.
- [Clarke et al., 2008] Clarke, A., Green, P., Chantler, M., and Emrith, K. (2008). Visual search for a target against a $1/f(\beta)$ continuous textured background. *Vision Research*, 48(21):2193–2203.
- [Clarke and Tatler, 2014] Clarke, A. D. and Tatler, B. W. (2014). Deriving an appropriate baseline for describing fixation behaviour. *Vision Research*, 102:41–51.
- [Clarke et al., 2009] Clarke, A. D. F., Chantler, M. J., and Green, P. R. (2009). Modeling visual search on a rough surface. *Journal of Vision*, 9(4):11–11.
- [Conci et al., 2013] Conci, M., Muller, H. J., and von Muhlenen, A. (2013). Object-based implicit learning in visual search: Perceptual segmentation constrains contextual cueing. *Journal of Vision*, 13(3):15–15.
- [Corbetta and Shulman, 2002] Corbetta, M. and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3):201–215.
- [Dakin and Baruch, 2009] Dakin, S. C. and Baruch, N. J. (2009). Context influences contour integration. *Journal of Vision*, 9(2):13–13.
- [Danilova and Mollon, 2014] Danilova, M. V. and Mollon, J. D. (2014). Symmetries and asymmetries in chromatic discrimination. *Journal of the Optical Society of America A*, 31(4):A247.
- [de Groot et al., 2016] de Groot, F., Huettig, F., and Olivers, C. N. L. (2016). When meaning matters: The temporal dynamics of semantic influences on visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 42(2):180–196.
- [Desimone, 1998] Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 353(1373):1245–1255.
- [Desimone and Duncan, 1995] Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18(1):193–222.
- [Duncan and Humphreys, 1989] Duncan, J. and Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3):433–458.
- [D’Zmura, 1991] D’Zmura, M. (1991). Color in visual search. *Vision Research*, 31(6):951–966.
- [Egeth and Yantis, 1997] Egeth, H. E. and Yantis, S. (1997). VISUAL ATTENTION: Control, representation, and time course. *Annual Review of Psychology*, 48(1):269–297.
- [Eisenberg and Zacks, 2016] Eisenberg, M. L. and Zacks, J. M. (2016). Ambient and focal visual processing of naturalistic activity. *Journal of Vision*, 16(2):5.
- [Engel, 1977] Engel, F. (1977). Visual conspicuity, visual search and fixation tendencies of the eye. *Vision Research*, 17(1):95–108.
- [Fallah and Reynolds, 2012] Fallah, M. and Reynolds, J. H. (2012). Contrast dependence of smooth pursuit eye movements following a saccade to superimposed targets. *PLoS ONE*, 7(5):e37888.
- [Fecteau and Munoz, 2006] Fecteau, J. and Munoz, D. (2006). Saliency, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, 10(8):382–390.
- [Field et al., 1993] Field, D. J., Hayes, A., and Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local “association field”. *Vision Research*, 33(2):173–193.
- [Findlay and Gilchrist, 2003] Findlay, J. M. and Gilchrist, I. D. (2003). *Active Vision*. Oxford University Press.
- [Flombaum, 2015] Flombaum, J. (2015). Visual search for features and conjunctions. *J. Vis. Exp.*
- [Follet et al., 2011] Follet, B., Meur, O. L., and Baccino, T. (2011). New insights into ambient and focal visual fixations using an automatic classification algorithm. *i-Perception*, 2(6):592–610.
- [Gao, 2008] Gao, D. (2008). *A discriminant hypothesis for visual saliency: computational principles, biological plausibility and applications in computer vision*. PhD thesis, UC San Diego.
- [Goolkasian, 1997] Goolkasian, P. (1997). Size scaling and spatial factors in visual attention. *The American Journal of Psychology*, 110(3):397.
- [Greene et al., 2012] Greene, M. R., Liu, T., and Wolfe, J. M. (2012). Reconsidering yarbus: A failure to predict observers’ task from eye movement patterns. *Vision Research*, 62:1–8.
- [Hart et al., 2013] Hart, B. M., Schmidt, H. C. E. F., Roth, C., and Einhäuser, W. (2013). Fixations on objects in natural scenes: dissociating importance from saliency. *Frontiers in Psychology*, 4.
- [Hayward and Burke, 2000] Hayward, W. G. and Burke, D. (2000). What pops-out in pop-out?. In *41st Annual Meeting of the Psychonomic Society*. IEEE.
- [Henderson et al., 1999] Henderson, J. M., Phillip A., J. W., and Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 25(1):210–228.
- [Hess et al., 2003] Hess, R., Hayes, A., and Field, D. (2003). Contour integration and cortical processing. *Journal of Physiology-Paris*, 97(2-3):105–119.
- [Horowitz and Wolfe, 1998] Horowitz, T. S. and Wolfe, J. M. (1998). Visual search has no memory. *Nature*, 394(6693):575–577.
- [Huk and Heeger, 2000] Huk, A. C. and Heeger, D. J. (2000). Task-related modulation of visual cortex. *Journal of Neurophysiology*, 83(6):3525–3536.
- [Hwang et al., 2011] Hwang, A. D., Wang, H.-C., and Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research*, 51(10):1192–1205.
- [Itti, 2007] Itti, L. (2007). Visual saliency. *Scholarpedia*, 2(9):3327.
- [Itti and Koch, 2000] Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12):1489–1506.
- [Itti et al., 1998] Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259.
- [Jonides, 1981] Jonides, J. (1981). *Voluntary versus automatic control over the mind’s eye’s movement Attention and performance IX*. Erlbaum Associates.
- [Judd et al., 2012] Judd, T., Durand, F., and Torralba, A. (2012). A benchmark of computational models of saliency to predict human fixations. *CSAIL Technical Reports*.
- [Judd et al., 2009] Judd, T., Ehinger, K., Durand, F., and Torralba, A. (2009). Learning to predict where humans look. In *2009 IEEE 12th International Conference on Computer Vision*. IEEE.
- [Koch and Ullman, 1987] Koch, C. and Ullman, S. (1987). Shifts in selective visual attention: Towards the underlying neural circuitry. In *Matters of Intelligence*, pages 115–141. Springer Netherlands.
- [Koffka, 1935] Koffka, K. (1935). *Principles of gestalt psychology*. Harbrace J, US.

- [Kong et al., 2017] Kong, G., Alais, D., and der Burg, E. V. (2017). Orientation categories used in guidance of attention in visual search can differ in strength. *Attention, Perception, & Psychophysics*.
- [Kowler, 2011] Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research*, 51(13):1457–1483.
- [Kristjánsson and Driver, 2008] Kristjánsson, Á. and Driver, J. (2008). Priming in visual search: Separating the effects of target repetition, distractor repetition and role-reversal. *Vision Research*, 48(10):1217–1232.
- [Kruger et al., 2013] Kruger, N., Janssen, P., Kalkan, S., Lappe, M., Leonardis, A., Piater, J., Rodriguez-Sanchez, A. J., and Wiskott, L. (2013). Deep hierarchies in the primate visual cortex: What can we learn for computer vision? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1847–1871.
- [Kubota et al., 2012] Kubota, H., Sugano, Y., Okabe, T., Sato, Y., Sugimoto, A., and Hiraki, K. (2012). Incorporating visual field characteristics into a saliency map. In *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12*. ACM Press.
- [Kurtz et al., 2017] Kurtz, P., Shapcott, K. A., Kaiser, J., Schmiedt, J. T., and Schmid, M. C. (2017). The influence of endogenous and exogenous spatial attention on decision confidence. *Scientific Reports*, 7(1).
- [Lamme and Roelfsema, 2000] Lamme, V. A. and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11):571–579.
- [Leboran et al., 2017] Leboran, V., Garcia-Diaz, A., Fdez-Vidal, X. R., and Pardo, X. M. (2017). Dynamic whitening saliency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(5):893–907.
- [LeMeur and Baccino, 2012] LeMeur, O. and Baccino, T. (2012). Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior Research Methods*, 45(1):251–266.
- [LeMeur and Liu, 2015] LeMeur, O. and Liu, Z. (2015). Saccadic model of eye movements for free-viewing condition. *Vision Research*, 116:152–164.
- [Li and Gilbert, 2002] Li, W. and Gilbert, C. D. (2002). Global contour saliency and local colinear interactions. *Journal of Neurophysiology*, 88(5):2846–2856.
- [Li, 1999] Li, Z. (1999). Visual segmentation by contextual influences via intra-cortical interactions in the primary visual cortex. *Network: Computation in Neural Systems*, 10(2):187–212. PMID: 10378191.
- [Li, 2000] Li, Z. (2000). Pre-attentive segmentation in the primary visual cortex. *Spatial Vision*, 13(1):25–50.
- [MacLeod and Boynton, 1979] MacLeod, D. I. A. and Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America*, 69(8):1183.
- [Mannan et al., 1996] Mannan, S., Wooding, D., and Ruddock, K. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, 10(3):165–188.
- [McDermott et al., 2010] McDermott, K. C., Malkoc, G., Mulligan, J. B., and Webster, M. A. (2010). Adaptation and visual salience. *Journal of Vision*, 10(13):17–17.
- [Miura et al., 2001] Miura, K., Suehiro, K., Yamamoto, M., Kodaka, Y., and Kawano, K. (2001). Initiation of smooth pursuit in humans. *Experimental Brain Research*, 141(2):242–249.
- [Nagy, 1999] Nagy, A. L. (1999). Interactions between achromatic and chromatic mechanisms in visual search. *Vision Research*, 39(19):3253–3266.
- [Nakayama and Silverman, 1986] Nakayama, K. and Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, 320(6059):264–265.
- [Navalpakkam and Itti, 2005] Navalpakkam, V. and Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, 45(2):205–231.
- [Nothdurft, 1985] Nothdurft, H. (1985). Sensitivity for structure gradient in texture discrimination tasks. *Vision Research*, 25(12):1957–1968.
- [Nothdurft, 1991] Nothdurft, H. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Research*, 31(6):1073–1078.
- [Nothdurft, 1993a] Nothdurft, H.-C. (1993a). The conspicuousness of orientation and motion contrast. *Spatial Vision*, 7(4):341–363.
- [Nothdurft, 1993b] Nothdurft, H.-C. (1993b). The role of features in preattentive vision: Comparison of orientation, motion and color cues. *Vision Research*, 33(14):1937–1958.
- [Nothdurft, 2000] Nothdurft, H.-C. (2000). Saliency from feature contrast: additivity across dimensions. *Vision Research*, 40(10-12):1183–1201.
- [Nothdurft, 2006a] Nothdurft, H.-C. (2006a). Saliency and target selection in visual search. *Visual Cognition*, 14(4-8):514–542.
- [Nothdurft, 2006b] Nothdurft, H.-C. (2006b). Saliency-controlled visual search: Are the brightest and the least bright targets found by different processes? *Visual Cognition*, 13(6):700–732.
- [Nuthmann et al., 2017] Nuthmann, A., Einhäuser, W., and Schütz, I. (2017). How well can saliency models predict fixation selection in scenes beyond central bias? a new approach to model evaluation using generalized linear mixed models. *Frontiers in Human Neuroscience*, 11.
- [Padilla et al., 2008] Padilla, S., Drbohlav, O., Green, P. R., Spence, A., and Chantler, M. J. (2008). Perceived roughness of $1/f^\beta$ noise surfaces. *Vision Research*, 48(17):1791–1797.
- [Palmer, 1995] Palmer, J. (1995). Attention in visual search: Distinguishing four causes of a set-size effect. *Current Directions in Psychological Science*, 4(4):118–123.
- [Palmer et al., 2000] Palmer, J., Verghese, P., and Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, 40(10-12):1227–1268.
- [Palmer, 1975] Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, 3(5):519–526.
- [Pannasch et al., 2008] Pannasch, S., Helmert, J. R., Roth, K., Walter, H., and Herbold, A.-K. (2008). Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, 2(2).
- [Parkhurst et al., 2002] Parkhurst, D., Law, K., and Niebur, E. (2002). Modeling the role of saliency in the allocation of overt visual attention. *Vision Research*, 42(1):107–123.
- [Parkhurst and Niebur, 2003] Parkhurst, D. and Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, 16(2):125–154.
- [Pashler, 1988] Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, 44(4):369–378.
- [Pashler et al., 2004] Pashler, H., Dobkins, K., and Huang, L. (2004). Is contrast just another feature for visual selective attention? *Vision Research*, 44(12):1403–1410.

- [Pestilli et al., 2007] Pestilli, F., Viera, G., and Carrasco, M. (2007). How do attention and adaptation affect contrast sensitivity? *Journal of Vision*, 7(7):9.
- [Posner, 1980] Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1):3–25.
- [Privitera et al., 2014] Privitera, C. M., Carney, T., Klein, S., and Aguilar, M. (2014). Analysis of microsaccades and pupil dilation reveals a common decisional origin during visual search. *Vision Research*, 95:43–50.
- [Proulx, 2010] Proulx, M. J. (2010). Size matters: Large objects capture attention in visual search. *PLoS ONE*, 5(12):e15293.
- [Puig et al., 2013] Puig, M. S., Zapata, L. P., Aznar-Casanova, J. A., and Supér, H. (2013). A role of eye vergence in covert attention. *PLoS ONE*, 8(1):e52955.
- [Quinlan, 2003] Quinlan, P. T. (2003). Visual feature integration theory: Past, present, and future. *Psychological Bulletin*, 129(5):643–673.
- [Rensink and Enns, 1995] Rensink, R. A. and Enns, J. T. (1995). Preemption effects in visual search: Evidence for low-level grouping. *Psychological Review*, 102(1):101–130.
- [Rensink and Enns, 1998] Rensink, R. A. and Enns, J. T. (1998). Early completion of occluded objects. *Vision Research*, 38(15-16):2489–2505.
- [Riche et al., 2013] Riche, N., Duvinage, M., Mancas, M., Gosselin, B., and Dutoit, T. (2013). Saliency and human fixations: State-of-the-art and study of comparison metrics. In *2013 IEEE International Conference on Computer Vision*. IEEE.
- [Riche and Mancas, 2016a] Riche, N. and Mancas, M. (2016a). Bottom-up saliency models for still images: A practical review. In *From Human Attention to Computational Attention*, pages 141–175. Springer New York.
- [Riche and Mancas, 2016b] Riche, N. and Mancas, M. (2016b). Bottom-up saliency models for videos: A practical review. In *From Human Attention to Computational Attention*, pages 177–190. Springer New York.
- [Rideaux et al., 2016] Rideaux, R., Badcock, D. R., Johnston, A., and Edwards, M. (2016). Temporal synchrony is an effective cue for grouping and segmentation in the absence of form cues. *Journal of Vision*, 16(11):23.
- [Rodieck, 1965] Rodieck, R. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5(12):583–601.
- [Rosenholtz, 2001] Rosenholtz, R. (2001). Visual search for orientation among heterogeneous distractors: Experimental results and implications for signal-detection theory models of search. *Journal of Experimental Psychology: Human Perception and Performance*, 27(4):985–999.
- [Rosenholtz et al., 2004] Rosenholtz, R., Nagy, A. L., and Bell, N. R. (2004). The effect of background color on asymmetries in color search. *Journal of Vision*, 4(3):9.
- [Rothkegel et al., 2017] Rothkegel, L. O. M., Trukenbrod, H. A., Schütt, H. H., Wichmann, F. A., and Engbert, R. (2017). Temporal evolution of the central fixation bias in scene viewing. *Journal of Vision*, 17(13):3.
- [Rothkopf et al., 2016] Rothkopf, C. A., Ballard, D. H., and Hayhoe, M. M. (2016). Task and context determine where you look. *Journal of Vision*, 7(14):16.
- [Rubin, 2001] Rubin, E. (1915/2001). *Figure and Ground*. Psychology Press.
- [Sagi and Julesz, 1984] Sagi, D. and Julesz, B. (1984). Detection versus discrimination of visual orientation. *Perception*, 13(5):619–628.
- [Salvucci and Goldberg, 2000] Salvucci, D. D. and Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the symposium on Eye tracking research & applications - ETRA '00*. ACM Press.
- [Saslow, 1967] Saslow, M. G. (1967). Latency for saccadic eye movement. *JOURNAL OF THE OPTICAL SOCIETY OF AMERICA*, 57(8).
- [Schiller et al., 1987] Schiller, P. H., Sandell, J. H., and Maunsell, J. H. (1987). The effect of frontal eye field and superior colliculus lesions on saccadic latencies in the rhesus monkey. *Journal of Neurophysiology*, 57(4):1033–1049.
- [Schiller and Tehovnik, 2001] Schiller, P. H. and Tehovnik, E. J. (2001). *Vision : from neurons to cognition*. Elsevier Science, Amsterdam New York.
- [Sheth and Young, 2016] Sheth, B. R. and Young, R. (2016). Two visual pathways in primates based on sampling of space: Exploitation and exploration of visual information. *Frontiers in Integrative Neuroscience*, 10.
- [SMI, 2009] SMI (2009). *iViewX System Manual (IVX-2.4-0908)*. SensoMotoric Instruments GmbH.
- [Soltani and Koch, 2010] Soltani, A. and Koch, C. (2010). Visual saliency computations: Mechanisms, constraints, and the effect of feedback. *Journal of Neuroscience*, 30(38):12831–12843.
- [Sommer, 1997] Sommer, M. A. (1997). The spatial relationship between scanning saccades and express saccades. *Vision Research*, 37(19):2745–2756.
- [Spratling, 2012] Spratling, M. (2012). Predictive coding as a model of the v1 saliency map hypothesis. *Neural Networks*, 26:7–28.
- [Strasburger et al., 2011] Strasburger, H., Rentschler, I., and Juttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, 11(5):13–13.
- [Tatler, 2007] Tatler, B. W. (2007). 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.
- [Tatler et al., 2005] Tatler, B. W., Baddeley, R. J., and Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research*, 45(5):643–659.
- [Tatler et al., 2006] Tatler, B. W., Baddeley, R. J., and Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research*, 46(12):1857–1862.
- [Tavassoli et al., 2009] Tavassoli, A., van der Linde, I., Bovik, A., and Cormack, L. (2009). Eye movements selective for spatial frequency and orientation during active visual search. *Vision Research*, 49(2):173–181.
- [Theeuwes, 2004] Theeuwes, J. (2004). Top-down search strategies cannot override attentional capture. *Psychonomic Bulletin & Review*, 11(1):65–70.
- [Treisman, 1988] Treisman, A. (1988). Features and objects: The fourteenth bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology Section A*, 40(2):201–237.
- [Treisman, 1991] Treisman, A. (1991). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, 17(3):652–676.
- [Treisman and Gormican, 1988] Treisman, A. and Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95(1):15–48.

- [Treisman and Souther, 1985] Treisman, A. and Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, 114(3):285–310.
- [Treisman and Gelade, 1980] Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136.
- [Trevarthen, 1968] Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychologische Forschung*, 31(4):299–337.
- [Troncoso et al., 2009] Troncoso, X., Macknik, S., and Martinez-Conde, S. (2009). Corner salience varies linearly with corner angle during flicker-augmented contrast: a general principle of corner perception based on vasarely’s artworks. *Spatial Vision*, 22(3):211–224.
- [Troncoso et al., 2005] Troncoso, X. G., Macknik, S. L., and Martinez-Conde, S. (2005). Novel visual illusions related to vasarely’s ‘nested squares’ show that corner salience varies with corner angle. *Perception*, 34(4):409–420.
- [Unema et al., 2005] Unema, P. J. A., Pannasch, S., Joos, M., and Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, 12(3):473–494.
- [van Zoest and Donk, 2006] van Zoest, W. and Donk, M. (2006). Saccadic target selection as a function of time. *Spatial Vision*, 19(1):61–76.
- [VanEssen and Gallant, 1994] VanEssen, D. C. and Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron*, 13(1):1–10.
- [Verghese, 2001] Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*, 31(4):523–535.
- [Vincent and Tatler, 2008] Vincent, B. T. and Tatler, B. W. (2008). Systematic tendencies in scene viewing.
- [Võ and Henderson, 2011] Võ, M. L.-H. and Henderson, J. M. (2011). Object–scene inconsistencies do not capture gaze: evidence from the flash-preview moving-window paradigm. *Attention, Perception, & Psychophysics*, 73(6):1742–1753.
- [Wandell, 1995] Wandell, B. (1995). *Foundations of vision*. Sinauer Associates, Sunderland, Mass.
- [Wang et al., 2014] Wang, C.-A., Boehnke, S. E., Itti, L., and Munoz, D. P. (2014). Transient pupil response is modulated by contrast-based saliency. *Journal of Neuroscience*, 34(2):408–417.
- [Wang et al., 2016] Wang, Y., Wang, B., Wu, X., and Zhang, L. (2016). Scanpath estimation based on foveated image saliency. *Cognitive Processing*, 18(1):87–95.
- [Wertheimer, 1938] Wertheimer, M. (1923/1938). *Laws of organization in perceptual forms*. Harcourt, Brace & Jovanovitch, London.
- [White and Munoz, 2011] White, B. and Munoz, D. P. (2011). *The Oxford Handbook of Eye Movements*. Oxford University Press.
- [White et al., 2017] White, B. J., Berg, D. J., Kan, J. Y., Marino, R. A., Itti, L., and Munoz, D. P. (2017). Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video. *Nature Communications*, 8:14263.
- [Winkler and Subramanian, 2013] Winkler, S. and Subramanian, R. (2013). Overview of eye tracking datasets. In *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE.
- [Wloka et al., 2017] Wloka, C., Kotseruba, I., and Tsotsos, J. K. (2017). Saccade sequence prediction: Beyond static saliency maps.
- [Wloka and Tsotsos, 2016] Wloka, C. and Tsotsos, J. (2016). Spatially binned ROC: A comprehensive saliency metric. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
- [Wolfe, 1998] Wolfe, J. (1998). *Attention*. Psychology Press.
- [Wolfe, 1992] Wolfe, J. M. (1992). “effortless” texture segmentation and “parallel” visual search are not the same thing. *Vision Research*, 32(4):757–763.
- [Wolfe, 1994] Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, 1(2):202–238.
- [Wolfe, 2001] Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception & Psychophysics*, 63(3):381–389.
- [Wolfe, 2010] Wolfe, J. M. (2010). Guided search 4.0: A guided search model that does not require memory for rejected distractors. *Journal of Vision*, 1(3):349–349.
- [Wolfe and Bennett, 1997] Wolfe, J. M. and Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, 37(1):25–43.
- [Wolfe et al., 1989] Wolfe, J. M., Cave, K. R., and Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3):419–433.
- [Wolfe and Horowitz, 2004] Wolfe, J. M. and Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5(6):495–501.
- [Wolfe et al., 2010] Wolfe, J. M., Palmer, E. M., and Horowitz, T. S. (2010). Reaction time distributions constrain models of visual search. *Vision Research*, 50(14):1304–1311.
- [Wolfe et al., 2011] Wolfe, J. M., Reijnen, E., Horowitz, T. S., Pedersini, R., Pinto, Y., and Hulleman, J. (2011). How does our search engine “see” the world? the case of amodal completion. *Attention, Perception, & Psychophysics*, 73(4):1054–1064.
- [Wolfson and Landy, 1995] Wolfson, S. and Landy, M. S. (1995). Discrimination of orientation-defined texture edges. *Vision Research*, 35(20):2863–2877.
- [Yarbus, 1967] Yarbus, A. L. (1967). *Eye Movements and Vision*. Springer US.
- [Zhang and Lin, 2013] Zhang, L. and Lin, W. (2013). *Selective Visual Attention*. John Wiley & Sons (Asia) Pte Ltd.
- [Zhang et al., 2008] Zhang, L., Tong, M. H., Marks, T. K., Shan, H., and Cottrell, G. W. (2008). SUN: A bayesian framework for saliency using natural statistics. *Journal of Vision*, 8(7):32.
- [Zhaoping, 2012] Zhaoping, L. (2012). Gaze capture by eye-of-origin singletons: Interdependence with awareness. *Journal of Vision*, 12(2):17–17.
- [Zhaoping and May, 2007] Zhaoping, L. and May, K. A. (2007). Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. *PLoS Computational Biology*, 3(4):e62.
- [Zhaoping and Zhe, 2015] Zhaoping, L. and Zhe, L. (2015). Primary visual cortex as a saliency map: A parameter-free prediction and its test by behavioral data. *PLOS Computational Biology*, 11(10):e1004375.

6 Supplementary Material

Code for the Stimulus Generation

<https://github.com/dberga/sig4vam>

Dataset Images, Masks and Fixation Data

http://www.cvc.uab.es/neurobit/?page_id=53