Visual and Emotional Salience Influence Eye Movements

YAQING NIU and REBECCA M. TODD, University of Toronto
MATTHEW KYAN, Ryerson University
ADAM K. ANDERSON, University of Toronto

In natural vision both stimulus features and cognitive/affective factors influence an observer's attention. However, the relationship between stimulus-driven (bottom-up) and cognitive/affective (top-down) factors remains controversial: How well does the classic visual salience model account for gaze locations? Can emotional salience counteract strong visual stimulus signals and shift attention allocation irrespective of bottom-up features? Here we compared Itti and Koch's [2000] and Spectral Residual (SR) visual salience model and explored the impact of visual salience and emotional salience on eye movement behavior, to understand the competition between visual salience and emotional salience and how they affect gaze allocation in complex scenes viewing. Our results show the insufficiency of visual salience models in predicting fixation. Emotional salience can override visual salience and can determine attention allocation in complex scenes. These findings are consistent with the hypothesis that cognitive/affective factors play a dominant role in active gaze control.

Categories and Subject Descriptors: D.2.7 [**Software Engineering**]: Distribution and Maintenance—*Documentation*; H.4.0 [**Information Systems Applications**] General; I.7.2 [**Document and Text Processing**]: Document Preparation—*Languages; photocomposition*

General Terms: Algorithms, Design, Experimentation, Human Factors, Measurement, Performance, Theory

Additional Key Words and Phrases: Emotional salience, visual salience, eye movements, attention, top-down, bottom-up

ACM Reference Format:

Niu, Y., Todd, R. M., Kyan, M., and Anderson, A. K. 2012. Visual and emotional salience influence eye movements. ACM Trans. Appl. Percept. 9, 3, Article 13 (July 2012), 18 pages.
DOI = 10.1145/2325722.2325726 http://doi.acm.org/10.1145/2325722.2325726

1. INTRODUCTION

In natural vision human observers sequentially allocate focal attention to subsets of the scene. Such attention shifts are typically associated with eye movement behavior [Rizzolatti et al. 1987]. Previous research shows that both visual stimulus-driven (bottom-up) and cognitive/affective (top-down) factors influence the competition for a share of our limited attention [Corbetta and Shulman 2002].

Bottom-up salience models explain guidance of eye movements based on the concept of a visual salience map [Koch and Ullman 1985]. Shifts of attention and eye movements are initiated toward

Author's addresses: Y. Niu (corresponding author) and R. M. Todd, Affect and Cognition Lab, Department of Psychology, University of Toronto, 100 St. George Street, Toronto, ON, M5S3G3, Canada; email: yaqing0930@hotmail.com; M. Kyan, Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, M5B2K3, Canada; A. K. Anderson, Affect and Cognition Lab, Department of Psychlogy, University of Toronto, 100 St. George Street, Toronto, ON, M5S3G3, Canada.

DOI 10.1145/2325722.2325726 http://doi.acm.org/10.1145/2325722.2325726

the point with the highest salience, which is then inhibited so that attention can be disengaged and be moved to the next most salient location. In this way, the visual salience provides a control mechanism for dynamically targeting eye movements. These visual salience models suggest that low-level feature discontinuities represented in the salience map can explain a significant proportion of where people look. Thus they specify filters that quantify the visual conspicuity which is perceived as significantly distinct from its local background of each part of the scene from two lines of investigation. First, computational models have been developed that use known properties of the visual system to generate a salience map. In these models, the visual properties present in an image generate the visual salience map that explicitly marks regions that are different from their surround such as color, intensity, contrast, and edge orientation [Itti and Koch 2000; Koch and Ullman 1985; Parkhurst et al. 2002; Torralba 2003], contour junctions, termination of edges, stereo disparity, and shading [Koch and Ullman 1985], and dynamic factors such as motion [Koch and Ullman 1985; Rosenholtz 1999]. As a benchmark for the visual salience hypothesis, Itti and Koch's [2000] model is thoroughly cited on behalf of other computational visual salience models.

In a second approach using scene statistics, local scene patches surrounding fixation points are analyzed to determine whether fixated regions differ in some image properties from regions that are not fixated. For example, high spatial frequency content and edge density have been found to be somewhat greater at fixated than nonfixated and two-point intensity correlation is lower for fixated scene patches than control patches [Krieger et al. 2000; Parkhurst and Niebur 2003; Reinagel and Zador 1999]. The Spectral Residual method (SR) [Hou and Zhang 2007] is based on the principle that the human visual system tends to suppress responses to frequently occurring features, while at the same time remaining sensitive to features that deviate from the norm. This method is also well cited because its simplicity of computation. Several studies have confirmed that Itti and Koch's [2000] model predicts salience at fixation significantly greater than chance [Parkhurst et al. 2002]. More recently, however, several studies have suggested that Itti and Koch's [2000] visual salience model predicts fixation location poorly in realistic visual search tasks [Stirk and Underwood 2007; Underwood et al. 2007]. These studies suggest that the visual salience model is a useful predictor of where people fixate with meaningless patterns (such as fractals) and in free viewing, but does a poor job when the viewing task involves active search. For the current study we focus on free viewing of static scenes. Given that Itti and Koch's [2000] model has become the most widely used computational visual salience model, one goal of the current study was to investigate whether the classic Itti and Koch's [2000] salience model performs better than the SR model in predicting fixation placement.

There is also evidence that visual salience does not account for all aspects of a scene that bias attention. For example, semantic meaning and social relevance of elements within a scene also influence allocation of overt attention. A recent study showed that visual salience could not fully account for where observers look within social scenes [Cerf et al. 2008, 2009]. Cerf et al. [2008, 2009] showed that the model that best predicted where observers fixated within scenes was a salience model combined with a face detection model. This combined model outperformed the salience model alone. Birmingham et al. also demonstrated that, when asked to look at a visual scene that includes human faces, participants most frequently fixate on the eyes [Birmingham et al. 2009a, 2009b], a tendency that is not accounted for by computationally modeled bottom-up visual salience. These studies shed light on attentional biases favoring faces and eyes, which cannot be fully explained by the standard bottom-up visual salience models. Thus it is not only visual conspicuity that preferentially commands attention in a complex visual scene.

The emotional salience, or motivational importance, of a stimulus may also influence the relatively reflexive allocation of attention. Many studies have demonstrated that attention is preferentially allocated to emotionally arousing stimuli relative to neutral stimuli [LaBar et al. 2000; Knight M 2007].

This bias favoring emotional stimuli even occurs under direct instructions to ignore the emotionally arousing items [Nummenmaa et al. 2006]. Emotional arousal has also been found to increase viewing duration for both pleasant and unpleasant scenes [Lang et al. 1993] and to capture greater initial attention as well as inhibit subsequent disengagement from a stimulus location [Fox and Russo 2002]. In a recent study, when neutral background scenes were edited to contain a single emotionally salient object and a single visually salient object [Humphrey et al. 2012], more fixations were allocated to emotionally salient than visually salient objects. In this present study, to precisely examine whether emotional salience or visual salience better predict observed gaze allocation, rather than directly comparing pairs of images, or edit the pictures to contain emotional stimulus and visually salient stimulus, we used methods for emotional salient region detection and visual salient region detection within a scene (emotional salience and visual salience in direct competition). Visual and emotional salience may also interact to predict viewing patterns. For example, both affective content and featural distinctiveness have been found to contribute to faster and more accurate visual searches for happy faces relative to other facial expressions [Calvo and Marrero 2009; Calvo and Nummenmaa 2008; Calvo et al. 2008]. Thus, a further goal of the current study was to explore the impact of emotional salience and visual salience on eye movement behavior, to understand the competition between visual stimulus-driven salience and emotional salience and how they affect gaze allocation while viewing complex scenes.

2. METHOD

2.1 Participants

Participants were 50 young adults (24 female, 18–40 years), with normal or corrected-to-normal vision and no history of neurological problems, recruited from the University of Toronto campus. 25 participants (12 female) participated in the main eye tracking experiment. Three subjects were excluded from the eye tracking experiment due to eye tracker drifting error, and eye movement data from 22 participants were used. 25 participants (12 female) performed a separate emotional salience Region Of Interest (ROIs) generation task. All subjects gave written informed consent for participation.

2.2 Stimulus Materials

The method is evaluated on a set of 75 images comprised of 25 negative, 25 positive, and 7 neutral photographs taken from the International Affective Picture System (IAPS) (IAPS Identification Numbers; see Table I), along with a further 18 neutral photographs retrieved from the Internet (Public Domain Images). Due to copyight issues, we only report the statistical results of this IAPS-based evaluation in this work (reported in Figures 2, 5, and 6) . In order to further convey the impact of emotional versus. visual salience on eye movements, we utilize a second set of images for illustration purposes only. The second set is comprised of positive, negative, and neutral public domain photographs, and forms the basis for the visual reports shown in Figures 3, 4, and 7.

Positive, negative, and neutral images were selected to be equivalent in mean log luminance, $F(2,72) < 1$, and RMS contrast, $F(2,72) < 1$, which were computed using the Image Processing Toolbox packaged with Matlab 7.0. Images also did not differ in whether they contained single versus multiple objects, $F(2,72) < 1$, or number of human figures, $F(2,72) < 1$, $p > .6$. Positive and negative images were selected to be equivalent in standardized ratings of emotional arousal (emotional salience). Scene complexity and difficulty of figure ground segregation were also rated by a separate set of participants. Participants were asked to rate how difficult it was to discriminate the focal figure of the scene from the background on a scale of 1–7, as well as the composition of each image on from simple to busy or complex on a scale of 1–7. Negative, positive, and neutral images also did not differ in difficulty of figure ground discrimination, $F(2,72) < 1$, $p > .5$, or scene complexity (scale of 1–7), $F(2,72) < 1$, $p = .5$.

Table I. IAPS Identification Numbers for 25 Negative, 7 Neutral, and 25 Positive Images used in the Experiment

Image Type	IAPS Identification Number
Negative Images	2800, 3030, 3051, 3120, 3130, 3170, 3220, 3280, 3500, 6313, 6250, 6570, 6560, 6540, 6550, 9253, 9402, 9433, 9570, 9921, 9040, 6510, 9400, 9102, 3230
Positive Images	4142, 4210, 4232, 4490, 4669, 4670, 4672, 4666, 4658, 4659, 4660, 4664, 4650, 4651, 4652, 4653, 4656, 4609, 4611, 4608, 4607, 4599, 4490, 4290, 4300
Neutral Images	2200, 2840, 2850, 5510, 7009, 7170, 7175, Other 18 Neutral Images are from Public Domain Images

2.3 Eye Tracking Experiment

2.3.1 *Apparatus.* Eye movement recording experiments were programmed in Experiment Builder and analyzed in DataViewer (SR Research). Eye movements were recorded using an infrared eye tracking desktop monocular system, EyeLink 1000 (SR Research, Mississauga, ON, Canada). Stimuli were shown on a 21W ViewSonic G225f monitor positioned 63 cm away from the participant, with a refresh rate of 140 Hz. Participants sat in front of the computer monitor and a chin rest was used to limit head movements. Throughout the experiment, the observers' right eye position was recorded and sampled at a rate of 1000 Hz. Pictures were presented at a visual angle of $11.17^{o} \times 8.37^{o}$. We used the manufacturer's software for calibration, validation, drift-correction, and determining periods of fixation. A 9-point calibration was performed at the start of the experiment followed by a 0-point calibration accuracy test. An additional drift-correction was performed whenever an observer failed to fixate within about 1.4^{o}- (50 pixels) of an initial central fixation cross within 5 s. In all experiments and conditions, each trial started with a central fixation cross which observers had to fixate for 500 ms to trigger stimulus onset.

2.3.2 *Experimental Procedures.* After informed consent and a brief practice session, participants performed the free viewing task while eye movements were recorded with the EyeLink system. Following calibration and validation, participants were shown each of the 75 images in a randomized sequence. Each image was shown for 2 seconds, and was preceded and succeeded by 2 seconds of black screen to minimize the possibility of proactive or retroactive interference, making each trial 6 seconds in length. Because pilot data indicated that even simple cognitive or memory tasks could alter the participants' eye movement pattern and fixation compared to a free viewing condition, participants were instructed to view the pictures in a natural manner. To guarantee consistent performance and to maintain concentration throughout the entire testing period (up to 20 minutes), participants were given two mandatory breaks after the 25th and the 50th trial.

2.4 Emotional Salience Region of Interest Generation Task

In order to generate regions of interest reflecting the most emotional salient regions of each image used in the task, participants were shown each of the 75 photo stimuli in a randomized sequence. For each image, they were instructed to click the mouse in the center of each of the five parts of each picture that were the most emotionally charged in order of intensity (from most intense to least intense). Participants were instructed as follows, "You will be shown a series of images. We want to know which parts of each image you find to be the most emotionally important or arousing. Please click the mouse in the center of the five parts of each picture that are the most emotionally charged for you in order of intensity (from most intense to least intense). This region could be a person or object or a part or combination of either."

The coordinates of the clicked pixel were processed using two-dimensional convolution with a 50-point Gaussian distribution window using Matlab, and an emotional salience map representing the average emotional salience value across participants was created for each picture stimulus. Then we generated emotionally salient regions based on the emotional salience map by ensuring that salient regions comprised 10% of the total image as shown in Figure 3 and Figure 4. Following the clicking task, participants rated each image for overall emotional salience using a numerical scale from 1 (the image was not emotionally arousing) to 7 (the image was extremely emotionally arousing).

2.5 Computational Visual Salience Models

Both Itti and Koch's [2000] computational visual salience model and the SR model were implemented to determine the visually salient regions in each stimulus image according to each. Itti and Koch's [2000] model was implemented using the salience Tool Box (STB) which is based on Neuromorphic Vision C++ Toolkit (NVT). The SR model was adapted by us to detect salient regions.

Each model was employed to process each image and generated salience maps that visualized the salience values. We then generated visually salient regions controlling the coverage of the salient regions (a region with a salience value higher than threshold was considered a salient region; a region with a salience value lower than threshold was considered a nonsalient region). The salient regions cover 10% of the total image.

2.6 Eye Tracking Data

Fixations were calculated by the built-in software of our eye tracking system. A fixation was defined as anything above 70 ms; micro fixations below 70 ms were discarded. We categorized fixations by their "fixation number" based on a fixations position in the ordered sequence of fixations (i.e., first, second, third). The "initial fixation" is the fixation occurring before stimulus onset, when the subjects are focusing on the centered fixation cross, and is not counted as part of the ordered sequence of fixations.

Saccades were also determined by the eye-tracking system. An eye movement was classified as a saccade when its velocity reached 30 deg/s or when its acceleration reached 8000 deg/s^2. The "saccade planning time" is the duration of time between the stimulus onset and the initiation of the first saccade. Saccade planning times smaller than 50 ms or greater than 600 ms were discarded to remove outliers and artifacts.

The mean number of fixations falling in each category of ROIs, first saccade fixation, runs and dwell time were calculated for emotional salience and visual salience ROIs to test predictions of eye movement behavior generated by each model. The number of runs in salient regions was defined as the number of times the salient region was entered and left by the participant's eye movement trajectory. Salient region dwell time was defined as the total duration across all fixations in salient regions predicted by each model. Repeated measures ANOVAs were employed comparing the predictions of each model (Itti and Koch [2000] Visual salience, SR Visual salience, Emotional Salience) within each image category (positive, negative, neutral).

3. RESULTS

3.1 Analysis 1: Comparing Two Visual Salience Model Predictions

In the first analysis, we compared how well the classic Itti and Koch [2000] and SR salience models accounted for gaze locations during free viewing of the stimuli. To test the visual salience models, subjects' gaze locations were used to validate the predictions of attention allocation by each of the two models. For this analysis, both the Itti and Koch's [2000] and SR salience models were used to determine the visual salient regions in each of our test images. Using each model, we generated a salience map for each image in which the brighter areas represent higher degree of salience (see Figure 1). Figure 1(a) shows the original image, Figure 1(b) shows the salience map generated by Itti and Koch's [2000] model, Figure 1(c) shows the salience map generated by the SR salience model.

We then generated salient regions on each image based on the salience maps generated by each model respectively (Figure 1(d) and Figure 1(e)). To compare the performance, we controlled the salience threshold preference to ensure the salience regions of both models covered 10% of the image (10% of the image regions' salience value higher than the threshold preference). Here we can see the ROIs generated by each model for the example image shown in the curves in Figure 1(f) and Figure 1(g).

To better understand the relationship between participants' fixations and the ROIs of both salience models, two measures of the participants' eye movements were examined: number of salient region fixations and number of first saccade fixations to regions predicted by each model.

Number of salient region fixations was defined as the number of participant-generated fixations that fell into salient regions for each image. This measure reflects all fixations in each type of salient region (Figure 2(a)) illustrating the number of fixations falling ROIs based on each of the two models. A repeated-measures ANOVA revealed a significant effect of model on number of fixations $F(1,21) = 23.162$, $p < .001$, $\eta^2 = 0.524$, with more fixations falling in regions predicted by the SR model than by Itti and Koch's [2000] model.

First fixation location was defined and analyzed as the number of participant-generated first saccades that landed in salient regions generated by each model. We chose to analyze the first fixation data (i.e., the first fixation after the experimenter determined fixation at center) because salience is expected to be most influential early on in scene viewing, and so analyzing first fixation helps to better understand the role of salience in determining fixation position.

Figure 2(b) shows the mean number of first fixations falling in ROIs based on each model. An ANOVA revealed a significant mean difference in first fixations falling into these two types of ROI, $F(1,21) = 64.56$, $p < .001$, $\eta^2 = 0.755$, with more first fixations falling in regions predicted by the SR model, suggesting that the SR model is a better predictor of human first fixations than the Itti and Koch's [2000] model.

In summary, from overall fixation and the first fixation results, we observe that the SR model provides overall better performance than Itti and Koch's [2000] model. In light of this result, we used the SR-generated visual salience ROIs in subsequent comparisons with participant-identified emotional salience ROIs.

3.2 Analysis 2: Visual Salience vs. Emotional Salience

To evaluate whether affective factors or visual factors have a greater influence on viewers' attention allocation, we generated ROIs based on the SR visual salience and emotional salience maps respectively. For illustration purposes only, Figure 3 shows the visual result of emotional salience maps generated on public domain images via the emotional salience ROI generation task. An example of 5 pixels identified by clicking each picture at the center of the region that participants find the most emotionally meaningful is shown in Figure 3(a). Figure 3(b) illustrates the resulting emotional salience map. As in the visual salience maps, the ROIs comprised 10% of the image size. This allowed us to compare the performance of visual salience and emotional salience in predicting eye movement behavior (Figure 4). Emotional salience ROIs are shown in Figure 4(a) and visual salience ROIs are shown in Figure 4(b). We note, the emotional salience ROIs are more focal, similar to that of Itti and Koch's [2000] model,

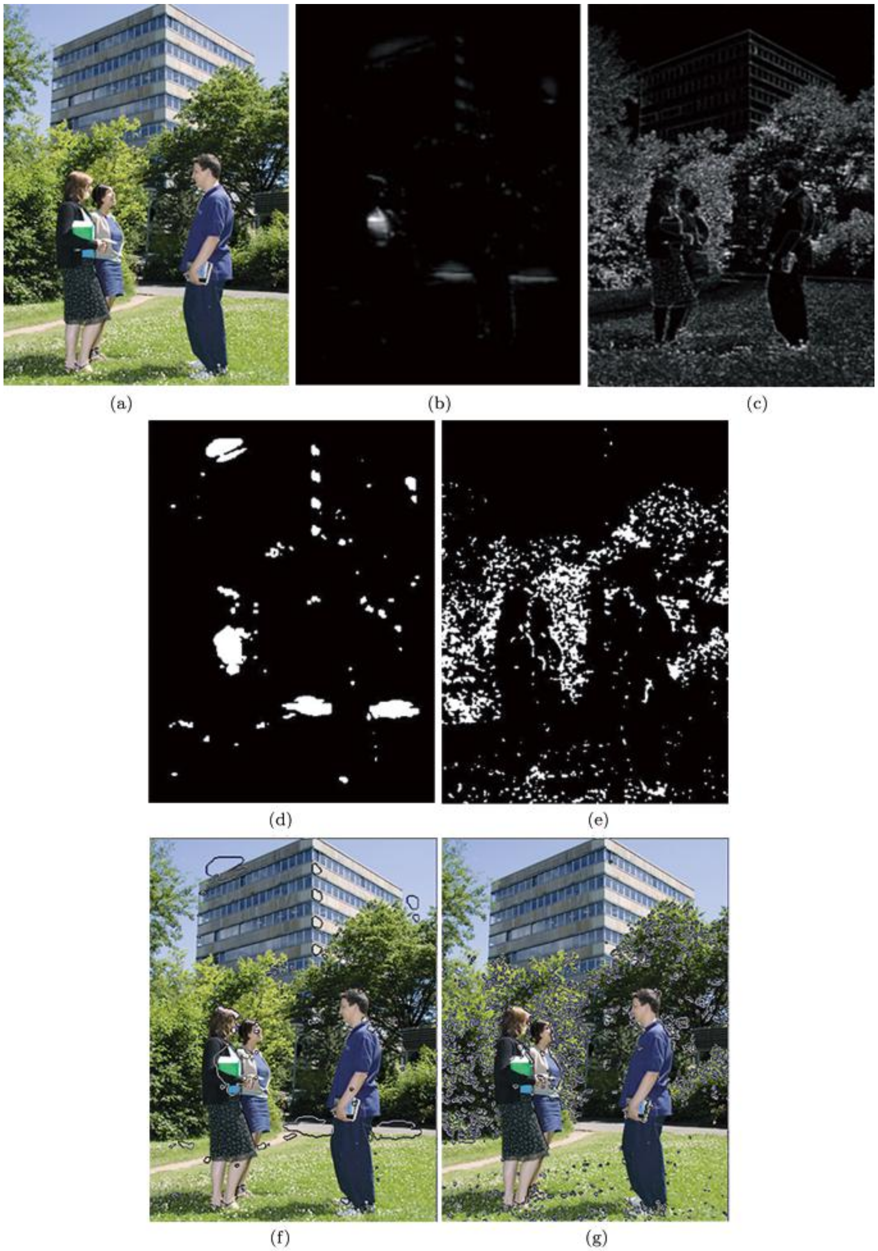


Fig. 1. Application of two visual salience models to a target image (not taken from the IAPS for copyright reasons). (a) The original image, no demarcation of visual salience. (b) The visual salience map generated by Itti and Koch's [2000] model. (c) The visual salience map generated by SR salience model. (d) ROIs based on the visual salience map generated by Itti and Koch's [2000] model. (e) ROIs based on the visual salience map generated by SR salience model. (f) ROIs generated by Itti and Koch's [2000] model, applied to target image. (g) ROIs by SR model, applied to target image.

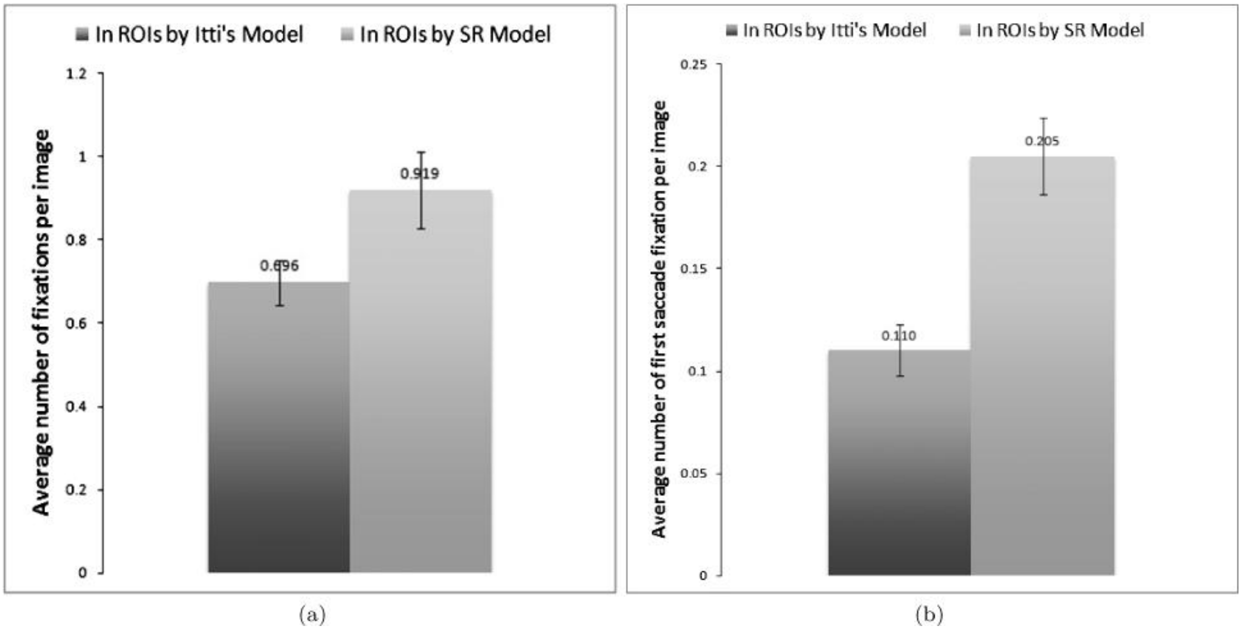


Fig. 2. Average number of participant-generated (a) fixations and (b) first saccade fixations, from all images, as a function of visual salience models. Error bars represent Standard Error of the Mean (SEM).

compared to the more diffuse ROIs generated by the SR model. More image examples from public domain images with participant-generated fixations (Figure 7) are shown in the appendix section.

For Analysis 2, we employed ROIs generated by competing models to examine the participants' eye movement behavior using four measures: (a) number of fixations in salient regions, (b) number of first fixations in salient regions, (c) number of runs in salient regions, and (d) dwell time in salient regions. Separate repeated-measures ANOVAs were performed on all four measures with the salience model (Visual vs. Emotional Salience) and emotional category (positive, negative, neutral) as within-subject factors. All results showing a main effect of salience model are reported first in Section 3.2.1. All results showing a main effect of emotional category and salience model and emotional category interactions are reported in Section 3.2.2.

3.2.1 *Participant-Identified Emotionally Salient Regions versus Visually Salient Regions.* Previous studies have shown that low-level visual salience guides our eye movements in free viewing of an image or scene by producing a pop-out effect. When viewing emotionally meaningful stimuli, both visually salient and emotionally salient ROIs compete for the viewers' attention. To evaluate whether emotional versus visual salience can better predict overt attention, in the second analysis we compared predictions generated by the SR model with those based on ROIs created using participant-identified emotionally salient regions for each image. In all analyses, Bonferroni-adjusted alpha levels are used for all contrasts.

(i) Number of Fixations. The average number of fixations that fell in visual salient ROIs and emotional salient ROIs was compared. In Figure 5(a) we can see the average fixation number in emotional salient ROIs and visual salient ROIs. An ANOVA revealed a difference in mean number of fixations between the two types of ROIs, F(1,21) = 57.112, p < .001, $\eta^2 = 0.731$. Overall, more fixations fell in emotional salience ROIs, suggesting that emotionally salient regions of a scene are more likely to correspond with the drawing of attention than visually salient regions.

(ii) Number of First Fixations. In Figure 5(b) we can see the number of first fixations in emotionally versus visually salient ROIs. An ANOVA revealed a significant mean difference between the two models $F(1,21) = 60.836$, $p < .001$, $\eta^2 = 0.743$. Reliably more first saccade fixations corresponded to emotionally than visually salient ROIs.

(iii) Number of Runs. We chose to analyze this measure because it captures the degree to which the salient regions attract eye movement. The average number of runs that fell within visually salient and emotionally salient ROIs was compared. In Figure 5(c) we can see the number of runs in both emotionally and visually salient ROIs. An ANOVA revealed a significant difference between these two types of ROIs, $F(1,21) = 86.795$, $p < .001$, $\eta^2 = 0.805$. More runs were made in emotional salient ROIs than in visual salient ROIs.

(iv) Dwell Time. The average dwell time in emotionally versus. visually salient ROIs was compared. Figure 5(d) illustrates dwell time in emotionally and visually salient ROIs. An ANOVA revealed a significant difference between these two types of ROI, $F(1,21) = 79.741$, $p < .001$, $\eta^2 = 0.792$. Participants' gaze lingered longer in emotionally salient ROIs than in visually salient ROIs.

From the preceding comparisons, it is clear that emotionally salient regions overwhelmingly elicited greater attention allocation than visually salient regions. This finding suggests that overt attention was drawn toward emotional salient ROIs due to their emotive nature rather than because they were visually unexpected.

3.2.2 *Influence of Emotional Category on Eye Movement Behavior.* The main effect of salience model reported in Section 3.2.1 suggests that the attentional pull of emotional salience overrides that of visual salience. Here we focus on how the emotion category of each image influences eye movement behavior, and whether this influence differs for fixations in regions predicted by the emotional salience and visual salience models.

(i) Number of Fixations. There was a main effect of emotion category, $F(2,42) = 7.95$, $p = .001$, $\eta^2 = .28$. Overall, negatively valenced images elicited more fixations than did neutral ($p = .005$) or positive ($p = .02$) images. This finding was qualified by a salience model by emotion category interaction $F(2, 42) = 11.70$, $p < .001$, $\eta^2 = .36$, suggesting this effect was driven by fixations in the emotional salience ROIs (Figure 6(a)). Pairwise comparisons revealed that, in emotional salient ROIs, negative images elicited more fixations than neutral ($p < .001$) as well as positive ($p = .02$) images.

(ii) First Fixation. There was a main effect of emotion category, $F(2,42) = 3.74$, $p < .05$, $\eta^2 = .15$. Overall, negatively valenced images elicited more first fixation than did neutral ($p = .005$) or positive ($p = .02$) images. There is no effect of salience model interaction with emotion category.

(iii) Runs. There was a main effect of emotion category, $F(2,42) = 10.94$, $p < .001$, $\eta^2 = .34$. Overall, negatively valenced images elicited more runs than did neutral ($p = .01$) or positive ($p = .001$) images. This finding was qualified by a salience model by emotion category interaction $F(2, 42) = 8.73$, $p = .001$, $\eta^2 = .29$, suggesting this effect was driven by runs in the emotional salience ROIs (Figure 6(c)). Pairwise comparisons revealed that, in emotional salient ROIs, negative images elicited more fixations than neutral ($p < .001$) as well as positive ($p < .001$) images.

(iv) Dwell Time. There was a main effect of emotion category, $F(2,42) = 5.612$, $p < .05$, $\eta^2 = .211$. Overall, negatively valenced images drew significantly higher fixation duration than did neutral ($p = .06$) or positive ($p = .02$) images. This finding was qualified by a salience model by emotion category interaction $F(2, 42) = 32.74$, $p < .001$, $\eta^2 = .61$, suggesting this effect was driven by dwell time in the emotional salience ROIs (Figure 6(d)). Pairwise comparisons revealed that, in emotional salient ROIs, negative images drew higher fixation duration than neutral ($p < .001$) as well as positive ($p < .001$) images, also positive images drew higher fixation duration than neutral images ($p = .001$).

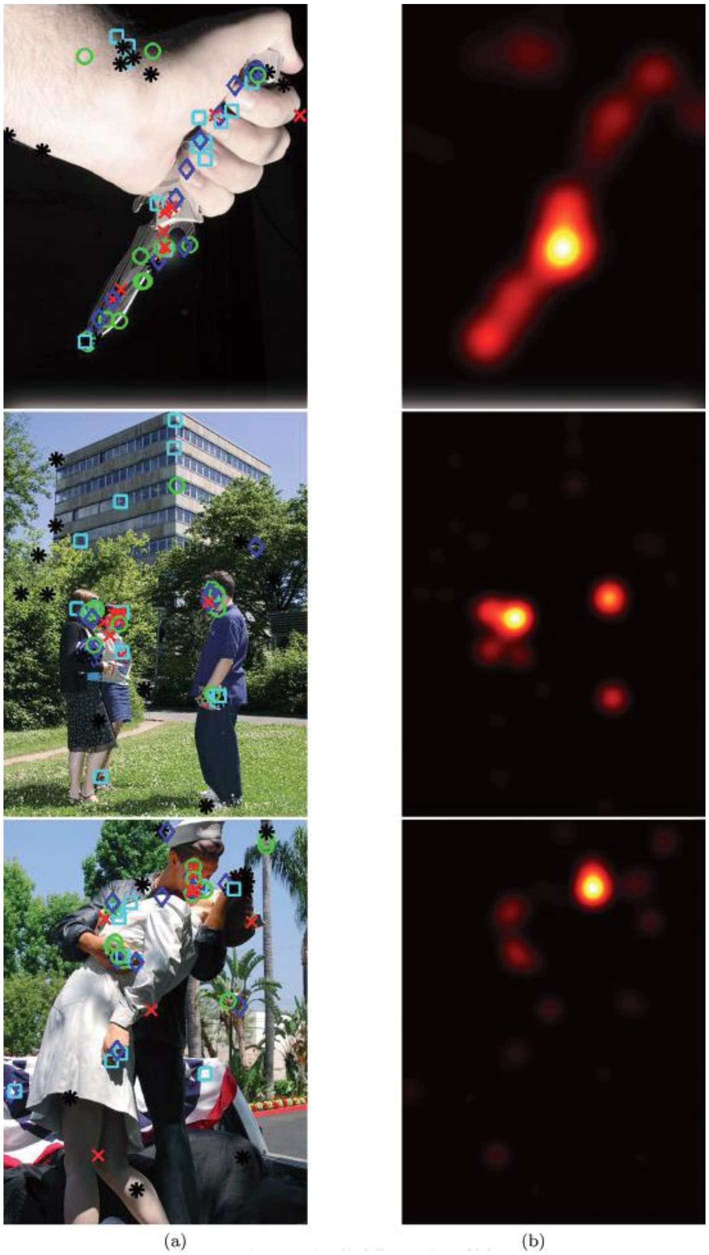


Fig. 3. Generation of emotional salience ROIs (not taken from the IAPS for copyright reasons). Column (a) From top to bottom, images categorized as: negative, neutral, positive. Shapes overlying the images denote spots that participants identified, via mouse clicks, as emotionally salient. Different shapes denote participants' order of preference. Column (b) Emotional salience maps, generated from participants responses to images in Column (a). (Bottom Left Photo: http//www.sxc.hu/photo/1392165).

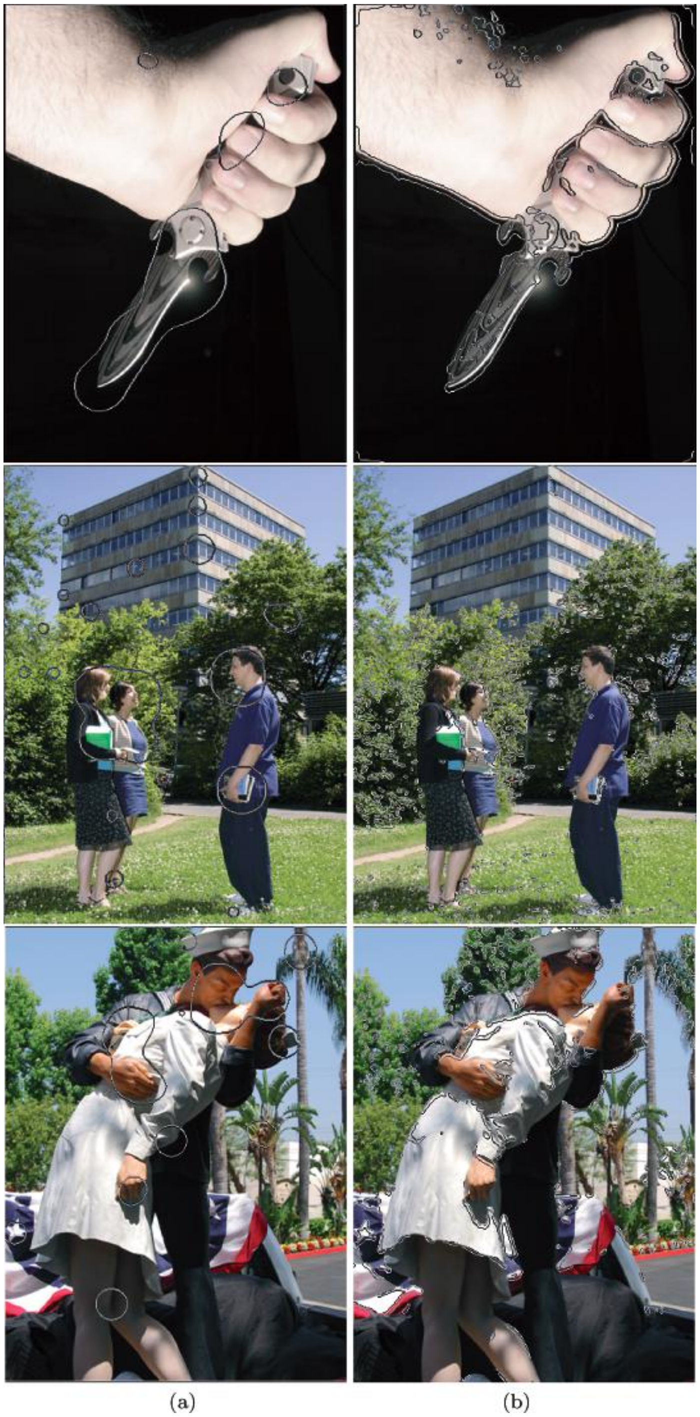
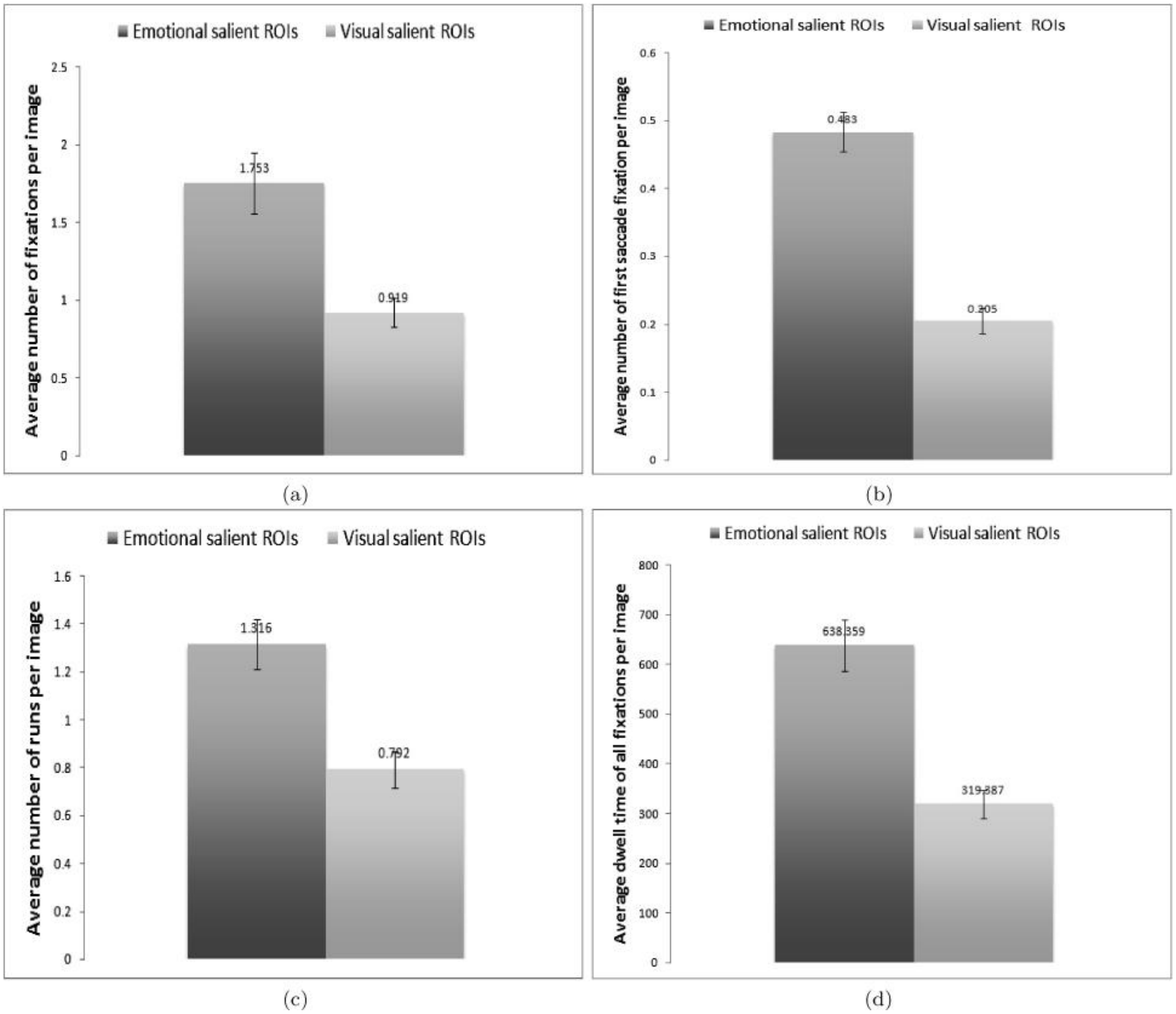


Fig. 4. Comparison of ROIs. Column (a) Emotional salient ROIs. (b) Visual salient ROIs. (Bottom Left Photo: http//www.sxc.hu/photo/1392165).



(a)

(b)

(c)

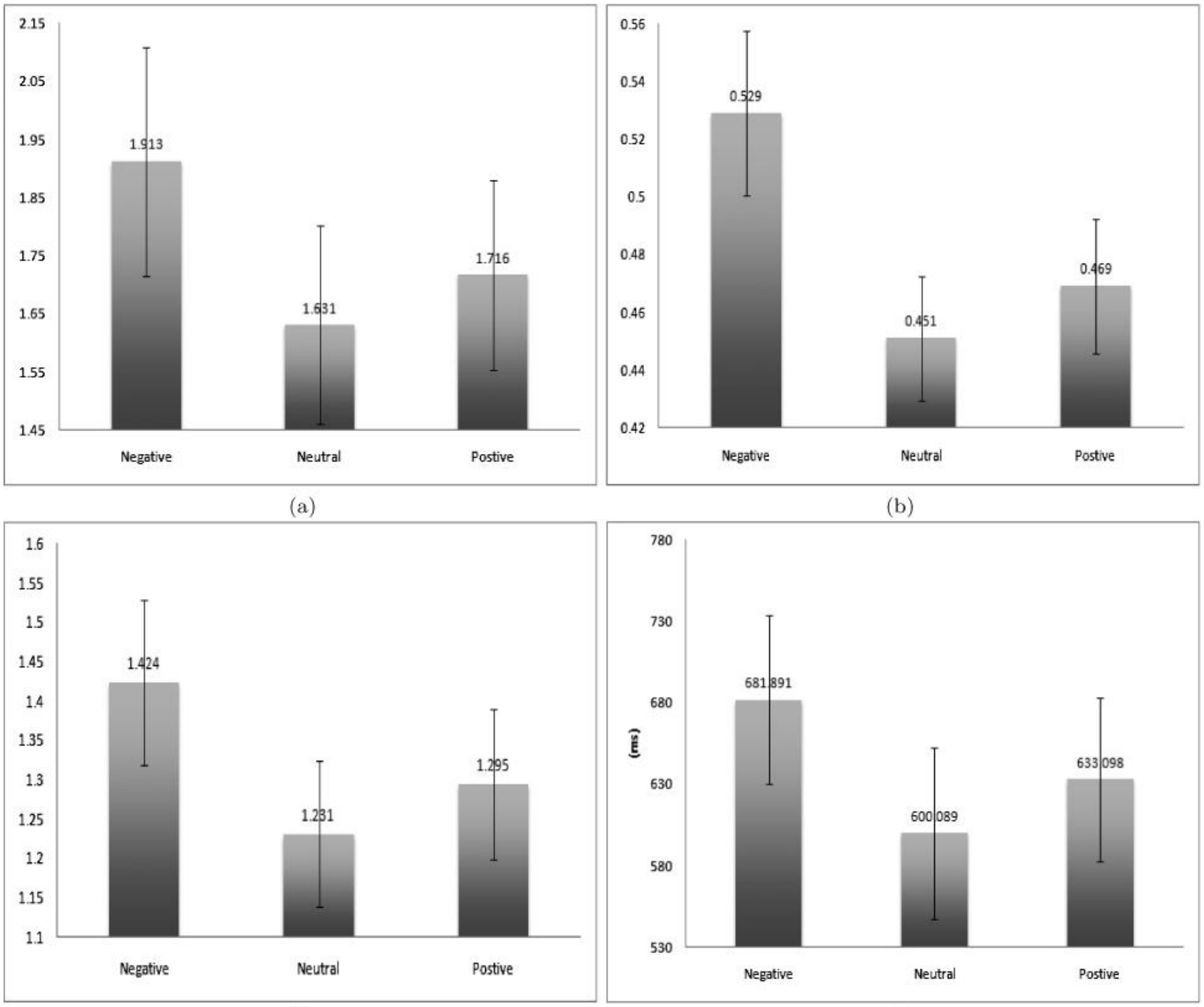
(d)

Fig. 5. Attention allocation measures comparison as a function of ROI type.

From the preceding comparisons, our attention to emotive regions in a scene is influenced by the emotional valence and arousal of such stimuli. ANOVAs revealed a significant effect of emotional arousal, with more fixations, more eye movement activity, and longer duration in emotionally salient regions of images that were arousing versus neutral overall. ANOVAs also revealed a significant effect of emotional valence, with more fixations, more eye movement activity, and longer duration in negatively versus positively valenced regions. This finding was consistent across salient region fixations, runs, and dwell time measures.

4. DISCUSSION

In natural vision both visual stimulus features and cognitive/affective factors influence an observer's attention. However, the relationship between stimulus-driven (bottom-up) and cognitive/affective



(a) (b)

(c) (d)

Fig. 6. Attention allocation measures comparison as a function of emotion type in emotional salient ROIs. (a) Fixations count. (b) First fixation count. (c) Run count. (d) Dwell time.

(top-down) factors remains controversial. How well does the classic visual salience model account for gaze locations? Can emotional salience counteract strong visual stimulus signals and shift attention allocation irrespective of bottom-up features? This study aimed to explore the impact of visual salience and emotional salience on eye movement behavior, to understand the competition between visual salience and emotional salience and how they affect gaze allocation in complex scenes viewing.

The visual salience hypothesis has generated a good deal of interest and in many ways has become the dominant view in the computational vision literature. Computational models have been developed which use known properties of the visual system to generate a salience map (landscape of visual salience) across an image. In these models, the visual properties present in an image generate a visual salience map that explicitly marks regions that are different from their surround based on color, intensity, contrast, and edge orientation [Itti and Koch 2000; Koch and Ullman 1985; Parkhurst et al. 2002; Torralba 2003] , contour junctions, termination of edges, stereo disparity, and shading [Koch and Ullman 1985], and dynamic factors such as motion [Koch and Ullman 1985; Rosenholtz 1999]. As a benchmark for the visual salience hypothesis, Itti and Koch's [2000] model is thoroughly cited on behalf of other computational visual salience models. Thus in the first analysis we examined whether the classic Itti and Koch's [2000] salience model performs better than other visual salience models in predicting fixation placement. As opposed to the classic method, the newer SR method [Hou and Zhang 2007] is based on the principle that the human visual system tends to suppress responses to frequently occurring features, while at the same time remaining sensitive to features that deviate from the norm. We compared the power of Itti and Koch's [2000] salience map and the SR salience map to predict human fixation locations during a free viewing task. Results showed that more fixations fell into salient regions generated by the SR model than Itti and Koch's [2000] model. Because visual salience is expected to be especially strong for early fixations [Henderson et al. 1999; Parkhurst et al. 2002], we also examined how frequently each model predicted the region of the first saccade landed as a fixation. Again, first saccade was landed most frequently to regions predicted by the SR model ROIs than Itti and Koch's [2000]. Thus, by both metrics the results suggested that the SR model does a better job for predicting attention allocation than the Itti and Koch's [2000] model formerly considered the gold standard for visual salience. Computational visual salience models have a wide range of applications, like adaptive content delivery, region-of-interest-based image compression, video summarization, image segmentation, image quality assessment, object recognition, and content-aware image scaling. The visual salience model comparison result gives us suggestions for salience-based image processing applications. The newer SR model provides superior results and does not rely on the choice of specific parameters which are needed to optimize performance in Itti and Koch's [2000] model.

Previous studies have shown that low-level visual salience helps guide eye movements in free viewing [Parkhurst et al. 2002]. Yet it is not only visual conspicuity that can produce a pop-out effect in the inspection of an image. There is also evidence that higher-level aspects of a stimulus, such as semantic meaning, can bias attention in favor of socially relevant stimuli such as faces/eyes [Cerf et al. 2008, 2009; Birmingham et al. 2009a, 2009b] Cerf et al. [2008, 2009] showed that the model that better predicted where observers committed fixations within scenes was a salience model combined with a face-detection model. Birmingham et al. [2008a, 2009b] showed that when asked to look at a visual scene that includes human faces, people frequently fixate on the eyes. When semantic meaning is further associated with emotional arousal, commonly feared or pleasant stimuli (e.g., a murder scene, erotica) can prioritize attention relative to neutral stimuli. For example, when emotionally arousing scenes are paired with neutral scenes, more fixations are allocated to the arousing scenes after controlling for visual salience [LaBar et al. 2000; Nummenmaa et al. 2009]. When neutral background pictures were edited to contain emotionally and visually salient objects [Humphrey et al. 2012], fixations were more likely to be on emotionally salient object. In the present study, rather than directly comparing pairs of images that differed in levels of overall arousal, we examined within a complex scene (emotional salience and visual salience in direct competition) the degree to which attention allocation is due to emotional salience and how much to low-level visual characteristics. Also rather than edit the pictures to contain emotional stimulus and visually salient stimulus, we use emotional salient region detection and visual salient region detection methods in emotional images.

To justify our choice of emotional salience ROIs generation task, we did a pilot study using a different subject-determined emotional salience ROIs task. In the pilot task participants were asked to click as rapidly as possible on the 5 parts of each image that caught their interest in order of interest. They were instructed to "go with their guts," and not "overthink" their choices. Comparison of the two tasks revealed that the ROIs created by the pilot task were highly correlated with those chosen in the emotional salience task despite different subjects in both studies, suggesting that what is considered interesting is what is most affectively charged and both tasks predicted fixation patterns better than visual salient maps. With regard to the decision to have participants select a single pixel rather circle whole objects, we wished to locate coordinates based on a central pixel in order to more precisely predict the xy coodinates of fixations. We also did not want to prespecify the size or scale of the region that would be chosen. For example, there is evidence that participants fixate on eyes rather than entire faces [Birmingham et al. 2009b]. Moreover, having participants circle entire objects could lead to overlarge regions of interest that would not be comparable to ROIs determined by visual salience. Stimulus materials were selected from the IAPS as well as the Internet. Positive, negative, and neutral images were equated in brightness, contrast, and scene complexity. Positive and negative images were also matched on standardized emotional arousal ratings. Whereas negative high-arousing picture contents included accidents, mutilation, human threat, positive high-arousing images were all within the category of erotica. Although affective salience relates to subjective impressions elicited by emotion rather than image categories, the fact that there was greater similarity between images in the positive category than in the negative and neutral categories may have influenced the results.

Visual and emotional salience may also interact to predict viewing patterns. Visual salience can serve as a cue for emotionally meaningful stimuli and emotional salience may enhance a viewer's perception of aspects of visual salience. For example, both affective content and featural distinctiveness have been found to contribute to faster and more accurate visual searches for happy faces relative to other facial expressions [Calvo and Marrero 2009; Calvo and Nummenmaa 2008; Calvo et al. 2008]. We investigated which aspect of salience biases competition for overt attention, that is, eye fixations when visual salience encounters emotionally meaningful aspects of an image. To evaluate whether affective factors or visual stimulus-driven factors better predict observed gaze allocation, in the second analysis we compared predictions generated by the SR visual salience model with measures indexing participant-identified emotionally meaningful regions of each image. Visual salience does have an effect on eye movements when one is inspecting an emotionally arousing scene, but the capacity of emotional salience to override visual salience can be plausibly observed from our results. More fixations were observed in emotionally salient regions than visually salient regions. Moreover, the same phenomenon occurred for the first saccade fixation, which suggests that the processing of affective features occurs at a relatively early stage of perception. Run results further revealed more eye movement activity in emotionally salient regions than visually salient regions and dwell time results show us participants gaze' also rests longer in emotionally than visually salient regions. These findings were not only robust in scenes that were high in emotional arousal, but also held true within low arousal scenes. The findings of this study go beyond the results showing higher quantity of fixation as in Humphrey's [Humphrey et al. 2012], they also showed more focused fixations and longer duration. Overall, the stronger capture of attention by the emotionally salient regions of the images could be interpreted as a high-level affective factor override of low-level visual salience. The SR models use visual stochastics, but emotional salience likely relies on content stochastics. The latter appears to win out of the former in directing the majority of looking behavior. We also found a negativity effect, showing greater likelihood of fixating on emotionally salient regions within negative relative to positive scenes. This finding suggests that negatively valenced scenes have an even stronger impact on attention allocation to emotionally salient regions compared to scenes that are equally arousing but positively valenced. Thus, our attention to emotive regions in a scene is influenced by the valence of such stimuli. This finding of negativity effect can be interpreted in the light of previous findings from our lab that negative, but not positive, affect enhances selective visual attention [Rowe et al. 2007; Schmitz et al. 2009]. Here, it is possible that negative affect generated by the negative arousing images increased selective attention in a form of "weapon focus" on the most emotionally salient items in the scene.

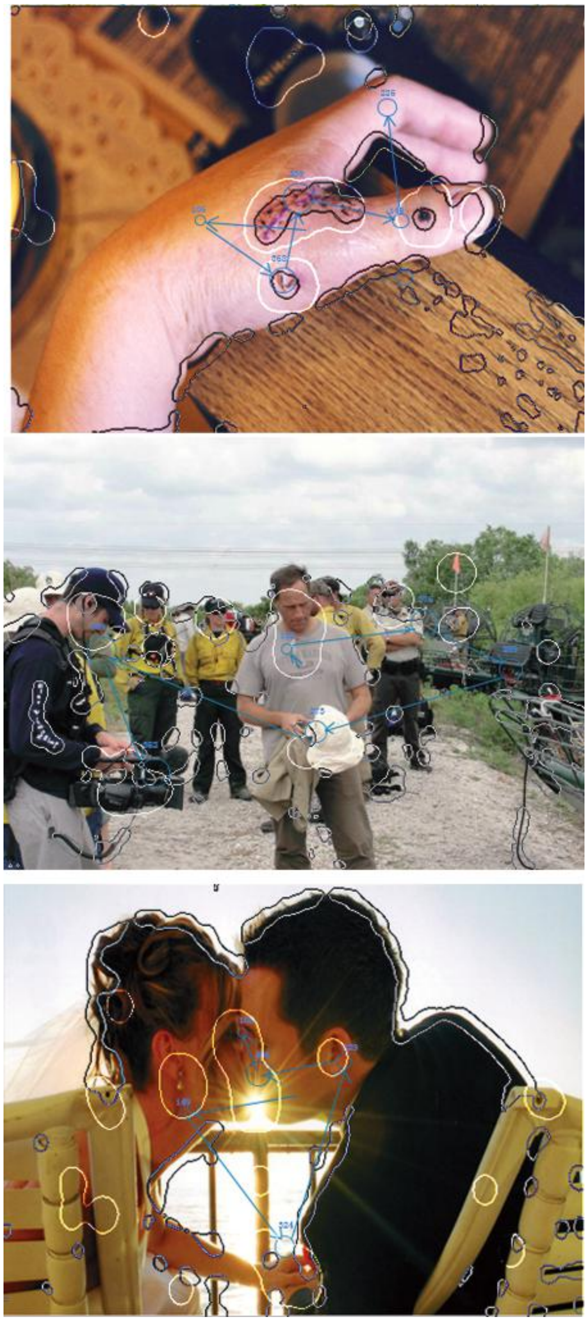


Fig. 7. More image examples with participant-generated fixations. Middele Photo. http://www.public-domain-image.com/people-public-domain-images-pictures/crowd-public-domain-images-pictures/a-little-down-time.jpg.html; (Bottom Photo: http://www.sxc.hu/photo/405013).

APPENDIX

We can see in Figure 7 that the white outlines demonstrate the emotional salient regions and the black outlines demonstrate the visual salient regions. We can also see data from one of the participants whose eye movement scan the path in blue. Note that the size of the circle denotes the fixation duration and the arrow illustrates sequences of fixations.

REFERENCES

BIRMINGHAM, E., BISCHOF, W. F., AND KINGSTONE, A. 2009a. Get real! Resolving the debate about equivalent social stimuli. *Vis. Cogn. 17*, 67, 904–924.

BIRMINGHAM, E., BISCHOF, W. F., AND KINGSTONE, A. 2009b. Saliency does not account for fixations to eyes within social scenes. *Vis. Res. 49*, 24, 2992–3000.

CALVO, M. G. AND MARRERO, H. 2009. Visual search of emotional faces: The role of affective content and featural distinctiveness. *Cogn. Emo. 23*, 4, 782–806.

CALVO, M. G. AND NUMMENMAA, L. 2008. Detection of emotional faces: Salient physical features guide effective visual search. *J. Exper. Psychol. 137*, 3, 471–494.

CALVO, M. G., NUMMENMAA, L., AND AVERO, P. 2008. Visual search of emotional faces. Eye-Movement assessment of component processes. *Exper. Psychol. 55*, 6, 359–370.

CERF, M., FRADY, E. P., AND KOCH, C. 2009. Faces and text attract gaze independent of the task: Experimental data and computer model. *J. Vis. 9*, 2, 1–15.

CERF, M., HAREL, J., E., AND W., KOCH, C. 2008. *Predicting Human Gaze Using Low-Level Saliency Combined with Face Detection*. Vol. 20. MIT Press.

CORBETTA, M. AND SHULMAN, G. L. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci. 3*, 3, 201–215.

FOX, E. AND RUSSO, R. 2002. Attentional bias for threat: Evidence for delayed disengagement from emotional faces. *Cogn. Emot. 16*, 3, 355–379.

HENDERSON, J. M., WEEKS, P. A., AND HOLLINGWORTH, A. 1999. The effects of semantic consistency on eye movements during complex scene viewing. *J. Exper. Psychol. Hum. Percep. Perform. 25*, 210–228.

HOU, X. AND ZHANG, L. 2007. Saliency detection: A spectral residual approach. In *Proceedings og the Conference on Computer Vision and Pattern Recognition (CVPR)*.

HUMPHREY, K., UNDERWOOD, G., AND LAMBERT, T. 2012. Salience of the lambs: A test of the saliency map hypothesis with pictures of emotive objects. *J. Vis. 12*, 1.

ITTI, L. AND KOCH, C. 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vis. Res. 40*, 1489–1506.

KNIGHT M, SEYMOUR TL, G. J. B. C. N. K. M. M. 2007. Aging and goal-directed emotional attention: Distraction reverses emotional biases. *Emot. 7*, 4, 705–714.

KOCH, C. AND ULLMAN, S. 1985. Shifts in selective visual attention: Towards the underlying neural circuitry. *Hum. Neurobiol. 4*, 219–227.

KRIEGER, G., RENTSCHLER, I., HAUSKE, G., SCHILL, K., AND ZETZSCHE, C. 2000. Object and scene analysis by saccadic eye-movements: An investigation with higher-order statistics. *Spatial Vis. 13*, 201–214.

LABAR, K. S., MESULAM, M., GITELMAN, D. R., AND WEINTRAUB, S. 2000. Emotional curiousity: Modulation of visuospatial attention by arousal is preserved in aging and early-stage alzheimer's disease. *Neuropsychol. 38*, 1734–1740.

LANG, P. J., GREENWALD, M. K., BRADLEY, M. M., AND HAMM, A. 1993. Aging and goal-directed emotional attention: Distraction reverses emotional biases. *Psychophysiol. 30*, 261–273.

MANNAN, S. K., R. K. H. W. D. S. 1996. The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vis. 10*, 165–188.

NUMMENMAA, L., HYONA, J., AND CALVO, M. G. 2006. Eye movement assessment of selective attentional capture by emotional pictures. *Emot. 6*, 2, 257–268.

NUMMENMAA, L., HYONA, J., AND CALVO, M. G. 2009. Emotional scene content drives the saccade generation system reflexively. *J. Exper. Psychol. 35*, 2, 305–323.

PARKHURST, D., LAW, K., AND NIEBUR, E. 2002. Modeling the role of salience in the allocation of overt visual attentio. *Hum. Neurobiol. 42*, 107–123.

PARKHURST, D. AND NIEBUR, E. 2003. Scene content selected by active vision. *Spatial Vis. 6*, 125–154.

Reinagel, P. and Zador, A. M. 1999. Natural scene statistics at the centre of gaze. *Netw. Comput. Neural Syst. 10*, 1–10.

Rizzolatti, G., Riggio, L., Dascola, I., and Umilta, C. 1987. Opposing influences of affective state valence on visual cortical encoding. *Neuropsychol. 25*, 31–40.

Rosenholtz, R. 1999. A simple saliency model predicts a number of motion popout phenomena. *Visi. Res. 39*, 3157–3163.

Rowe, G., Hirsh, J., and Anderson, A. 2007. Positive affect increases the breadth of attentional selection. In *Proc. Nati. Acad. Sci.* 383–388.

Schmitz, T. W., De Rosa, E., and Anderson, A. K. 2009. Opposing influences of affective state valence on visual cortical encoding. *J. Neurosci. 29*, 22, 7199–7207.

Stirk, J. A. and Underwood, G. 2007. Low-Level visual saliency does not predict change detection in natural scenes. *J. Vis. 7*, 10, 1–10.

Torralba, A. 2003. Modeling global scene factors in attention. *J. Opt. Soc. Amer. 20*.

Underwood, G., Templeman, E., Lamming, L., and Foulsham, T. 2007. Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. In *Consciousness and Cognition*.

Received May 2012; accepted June 2012