EG UK Theory and Practice of Computer Graphics (2011)
Hamish Carr, Ian Grimstead (Editors)

Visual Saliency for Smell Impulses and Application to Selective Rendering

C. Harvey †[1], T.E.W.Bashford-Rogers[1], K. Debattista[1] and A. Chalmers[1]

[1]Visualisation Group, International Digital Laboratory, Warwick University, UK.

Abstract

A major challenge in generating high-fidelity virtual environments is to be able to provide interactive rates of realism. However this is very computationally demanding and only recently visual perception has been used in high-fidelity rendering to improve performance considerably by a series of novel exploitations; to render parts of the scene that are not currently being attended by the viewer at a much lower quality without the difference being perceived. This paper investigates the effect various smells have on the visual attention of the user when free viewing a set of engineered images. We verify the worth of investigating these saccade shifts (fast movements of the eyes) due to attention distraction to a congruent smell object. By analysing the gaze points, we identify time spent attending a particular area of a scene. We also present a technique from measured data to remodulate traditional saliency maps of image features to account for the observed results. We show that smell provides an impulse on attention to affect perception in such a way that this can be used to guide selective rendering of scenes through use of the remodulated saliency maps.

Categories and Subject Descriptors (according to ACM CCS): Computer Graphics [I.3.3]: Picture/Image Generation - Viewing Algorithms—Image Processing and Computer Vision [I.4.8]: Scene Analysis - Object Recognition—Image Processing and Computer Vision [I.4.8]: Scene Analysis - Tracking—

1. Introduction

In order to have *realism in real-time* for use in virtual reality (VR) applications now, it is necessary to exploit knowledge of the Human Visual System (HVS) to significantly reduce computation without loss in perceptual quality. The visually important features in a scene can be computed from Saliency Maps [IKN98, YPG01]. However, another key feature of human perception is cross-modality. Typical saliency models account for visual perception only. This paper considers the sense of smell and investigates how it affects eye saccade shifts and attention and whether this can be exploited to reduce computation without any perceptual loss in visual quality.

Our experiments aim to elicit quantitive proof of the possibility that olfaction, the sense of smell, guides visual attention in a bottom-up fashion. We present participants with visual cues under different smell conditions and monitor gaze patterns and durations as a quantity to human attention. We find that olfactory sensory stimuli guides visual attention to congruent objects and overrides traditional saliency models of visual feature prediction. These results are used to drive the generation of saliency maps to guide a sampling heuristic for selective rendering.

The paper is organised as follows. Section 2 describes previous work in the field of visual attention with specific reference to cross-modal interactions and covers prevalent work on the sense of smell. Section 3 describes the experimental setup and techniques and concludes with results. Section 4 discusses our new technique to remodulate visual saliency to account for the observed results. Finally, Section 5 draws conclusions and suggests future work in the area.

2. Related Work

Previous work has looked into other modalities and how they affect visual attention, notably sound. Mastoropoulou [MDCT05b, Mas06] has shown that participants, viewing an

† Carlo.Harvey@warwick.ac.uk

animation whilst an ambient sound was playing, were statistically significantly more likely to make incorrect choices as to the current frame rate when an acoustic stimulus was introduced together with a visual representation of that stimulus. Mastoropoulou *et al.* [MDCT05a] also showed that during an animation an ambient sound presented in the presence of a visual cue can be used to exploit perception to guide selective rendering. Furthermore Harvey *et al.* [HWBR*10] showed this phenomena further extended to spatial sound impacting gaze directions to attend a directional sound source when present. In additions, Ramic-Brkic *et al.* explored the cross-modal effects between smell and graphics when rendering grass in real-time [BCB*09], while Ellis *et al.* considered selective rendering in the presence of motion [EC06].

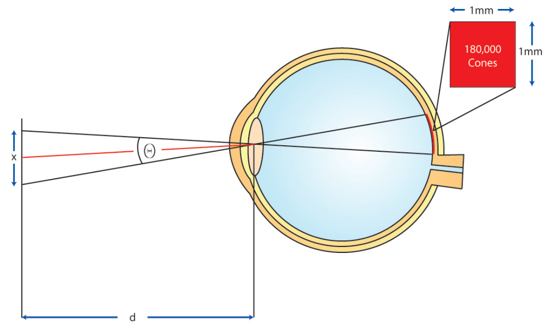

Figure 1: *The eyes' foveal angle. Due to the high number of cones (coloured light receptors) located in the fovea this area has the highest impact on visual perception of a colour image.*

2.1. Visual Attention

Coded into the retina [Dow87] is the sequential selection process the HVS uses to determine the hierarchy of visual cues used to deterministically select which objects in any given image are most important. This is necessary because there exists far too much information in any one image to remember it all with a single glance.

The first psychological study of human task-orientated saccades was undertaken by Yarbus who noted that under task-based scenarios the eyes jump in saccades to new points of interest in the scene [Yar67]. However, once the object of interest has been found and lies within the foveal region the eye tracks the object in a smooth manner. More recent psychological research [NBH*08] shows that concurrent audio stimuli increases the visual system's ability to distinguish brief visual events.

2.1.1. Peripheral Vision

Spatial acuity is maximised around the fovea as shown in Figure 1. In [ML97,LM99,LMYM01] the authors had a gaze contingent multi resolutional display producing only high visual resolutions within the area attended by the fovea. Using pre-stored exhaustive combinations of possible image combinations they showed an update was required before 5 milliseconds elapsed after a fixation of the fovea on a region following a saccade. This was necessary to maintain the integrity of the illusion without disturbing the cognitive process. Loshky *et al.* [LMYM01] ran a series of tests in order to adjudge that a 4.1° representation for the foveal coverage of the multi resolutional display produced results which indicated the difference between it and a completely high resolution display was statistically indistinguishable.

2.2. Saliency Maps

Itti *et al.* [IKN98] proposed a model to decorrelate an image into local image intensity, colour and orientation feature maps. A combination of these maps produces a saliency map which is a 2D predictor of likelihood a region will draw attention of a viewer. Inattentional blindness is a failing of the human eye to see items accurately lying outside of the foveal region [MR98]. Saccade shifts for the human eye are very much driven by this phenomena. Itti and Koch's [IK00] ideas of top-down (user driven) and bottom-up (scene stimulus driven) phenomena directing attention are prevalent to inattentional blindness. This means user driven task-orientated viewing can cause inattenional blindness to scene stimuli and vice versa. Aleph maps introduced by Yee *et al.* [YPG01] are a combination of saliency maps and error tolerance which were used in dynamic environments to reduce rendering computation time.

Cater *et al.* [CCL02, CCW03] looked at how users perceived a selective quality animation versus a high quality animation under different task-based scenarios. When asked to count pencils during watching the animations, the participants were unable to notice the difference between high quality and selective quality animations where only the area surrounding the interest region was rendered in high quality via task maps. A combination of task maps and saliency maps created an importance map in the work by Sundstedt *et al.* [SDC05]. Debattista *et al.* [DC05] joined importance maps and component based rendering to direct a selective guidance system in a rendering application.

2.3. Smell and Attention

The sense of smell, olfaction, is a major sense of humans. Smell has been linked with influencing mood, emotion, memory, social behaviour and even partner choice [Jac07]. Despite this, smell has largely been omitted from virtual environments seemingly waiving the vast role it plays in human behaviour.

It has been well established that visual cues have a marked improved affect on olfactory performance. Zellner [ZBE91] demonstrated that odours matched appropiately with a colour were rated more pleasant than inappropriately matched odours. Sakai *et al.* [SIS*05] showed that watch-

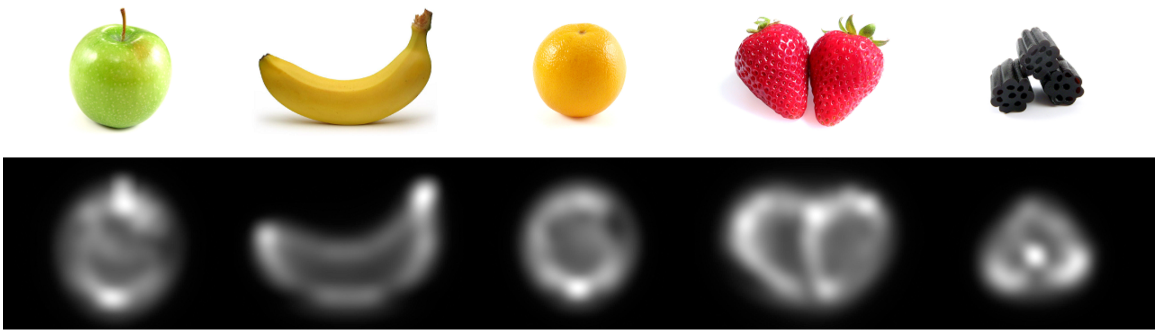

Figure 2: *All images used in the slides and respective individual saliency maps.*

ing congruent pictures had the effect of increasing the pleasantness and also odour intensity as opposed to incongruent picture matching. Seignuric *et al.* [SDJ*10] explored the influence of prior learned associations between an odour and a visual stimulus naturally associated with that odour on eye saccades and fixtations. They showed that the odour-related visual cue was explored faster and for a shorter time in the presence of a congruent odour.

In comparison, less research has investigated the influence of olfactory cues on visual attention. A study presented by Millot *et al.* has shown smell can impact participant behaviour towards visuals [MBM02], in that ambient smell produced a faster response to a visual stimuli in a sensory-motor task than a condition with no ambient smell. Knasko [Kna95] showed participants looked longer at slides in the presence of a pleasant odour such as baby powder and chocolate than with no odour present. Seo *et al.* [SRMN10] extended this by showing a distinct effect of congruency upon viewing time and in addition explored where eyes fixated, given congruent and incongruent images as a visual stimulus.

3. Psychophysical Experimental Layout, Procedure and Results

Our experiments aim to elicit quantitive proof of the possibility that olfaction, the sense of smell, guides visual attention in a bottom-up fashion.

3.1. Methodology

3.1.1. Variables

For the experiment we used two image sets. One set had an image with higher visual saliency and one set had an image with a lower saliency; strawberry and liquorice respectively. The other images for the remaining three quadrants were constant between sets; apple, banana, orange. All images can be seen in Figure 2. One of the four images from each image set was displayed on the screen in random quadrants. One possible combination for a higher saliency image set slide is shown in Figure 3.

The images were specifically chosen to have similar saliency besides the variable factor between the two sets which was the strawberry or liquorice image. The saliency map of the slide highlighted in Figure 3 is shown in Figure 4. As one can see; the strawberry comes out on top in a winner takes all combination of the local intensity, colour and orientation feature maps.

3.1.2. Experiment

Each participant was randomly assigned to one image set and were presented four one minute slides and four 30 second buffer slides for nose desensitisation, for a total of a six minute experiment. The buffer slides displayed a small red circle in the centre of the screen to minimise discrepancies in gaze quadrant analysis and to create a fixation in a free area so free-view conditions are started without bias when the next main slide starts. The four main slides cycled randomly through the subset of variables; i.e. which smell to emit alongside the display of the slide. Smells used in this experiment were apple, banana, orange and a control: no smell. Smells were delivered to the participant for the full duration of the slide. We eye tracked participants free viewing the slides under these conditions.

3.2. Equipment

The eye tracker used was as uninvasive as possible. It is a passive measuring device with no extraneous materials connected to the participant; making free viewing an image as natural as possible. faceLAB^{TM}provides a system which records real time blink, saccade and fixation estimates.

The smell delivery system used was a Vortex Activ produced by Dale Air. This device has four fan emission chambers which are programmable via a micro controller to guide chamber impulse onset, duration and which chamber in the device to fire. PTFE tubes were used to minimise the effect of adsorption of smell molecules onto the surface of the inner tubing which could bias subsequent data collection as participants went through the experiment. Tubes were blown clear before and after each participant, as well as an empty chamber pumping air when a smell chamber was not firing.

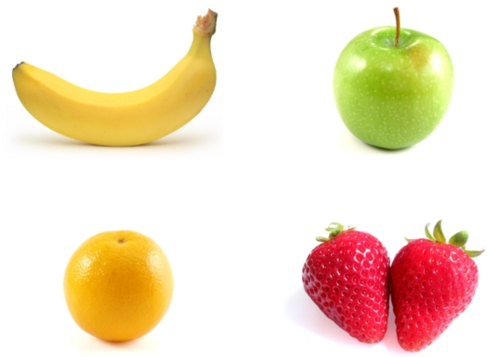

Figure 3: *A possible combination of images for a slide presented to a participant. Slide from the salient image set.*

Figure 4: *Saliency map for the slide shown in Figure 3.*

In addition to this, during experiment down time, room fans and a room neutral deodourant were used in order to minimise discrepancies between subsequent datasets.

3.3. Setup

Experimental setup and placement is shown in Figure 5. The participant was sat on a chair, with the backrest of the chair 115cm from the display. One metre long PTFE tubes were drawn from the device chambers and clipped to the top of a participants collar. The monitor used for display was a 37" 1280x1024 resolution display. The distance between equipment was standardised and controlled.

3.4. Participants

A total of 30 participants took part in this experiment, 21 males and 9 females. 15 were assigned to the salient image set and 15 to the non-salient image set. Participants had an average age of 26.7 and had normal or corrected to normal vision at the time of testing. In addition participants reported

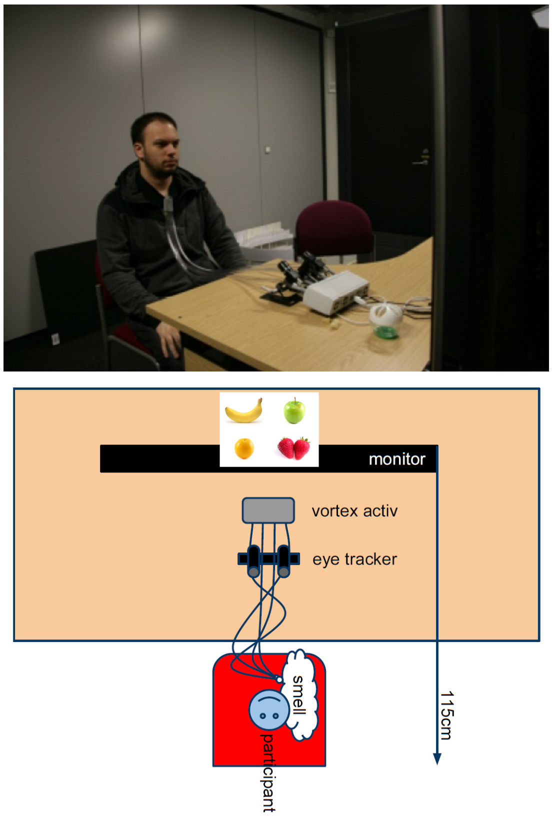

Figure 5: *Real(top) and figurative(bottom) experimental setup schemas.*

no anosmia (inability to smell), which has also been reported to be a factor of age.

3.5. Results

The average time spent in a quad region for each smell condition is shown in Tables 1 and 2 for saliency and no saliency conditions respectively. In addition the complete breakdown of the study on a participant by participant basis is provided in the appendix in Figure 9. Not every column total will sum to 60, the duration of a slide, due to lost tracking data or time spent gazing outside slide screen boundaries. But as can be seen time spent within the salient region decreases when smell condition is presented asynchronously with a congruent object. In addition time spent viewing the congruent object for the smell conditions is increased drastically under both image sets. When no smell condition is present there is no smell congreuncy; as such standard visual saliency predictive metrics prevail and the strawberry is most viewed. As can be seen in the control column of Table 2, the average time is balanced between quadrants, as their visual saliencies are all similar.

Quadrant Attended	Smell			No Smell
	Apple	Banana	Orange	Control
Apple	**21.13**	6.04	6.34	9.37
Banana	6.26	**26.34**	5.71	6.69
Orange	6.88	6.96	**23.6**	6.89
Strawberry	13.31	11.91	11.99	**23.91**

Table 1: *Average gaze time(s) within each quadrant for the salient image set under each odour condition. Largest values per smell condition emboldened.*

Quadrant Attended	Smell			No Smell
	Apple	Banana	Orange	Control
Apple	**20.38**	9.29	11.9	10.22
Banana	11.45	**23.2**	9.11	**14.18**
Orange	11.24	9.24	**20.81**	12.10
Liquorice	10.9	9.90	9.92	13.60

Table 2: *Average gaze time(s) within each quadrant for the non-salient image set under each odour condition. Largest values per smell condition emboldened.*

4. Gaze Point Analysis

4.1. Density Estimation

Gaze points within the area of interest for the pertaining smell condition appear, after density estimation on the set of gaze points, to be modulated by a standard visual saliency model; as seen in Figure 6. However, on top of being modulated by visual saliency it also overrides the standard visual saliency predictor that stipulated the strawberry would have been the region of most interest; which was the case under the control condition.

As such a new model is needed for saliency when we have smell impulses in a virtual environment. The recorded gaze points fits a normal distribution, as shown in Figure 7. Using a gaussian profile, with a standard deviation from the measured data points as the sigma on a smell material matte blur, gives us a predictor of the area of influence of the smell attention deviation: GaussSmellBlur in Equation 1 and Figure 8. Convolving this with the original saliency map for the scene and subsequently normalising gives us a reattenuation factor for the orginal saliency for the scene: Smell Modulation Map in Figure 8. Added to the original map guided via a weighting function, gives us our new model for visual saliency when a smell emitting object is present.

4.2. Weighting Function

$$\begin{aligned} &SmellSaliency = \\ &\quad (1-w)(Saliency - GaussSmellBlur) + \\ &\quad (w)\frac{GaussSmellBlur \otimes Saliency}{|GaussSmellBlur \otimes Saliency|} \end{aligned} \tag{1}$$

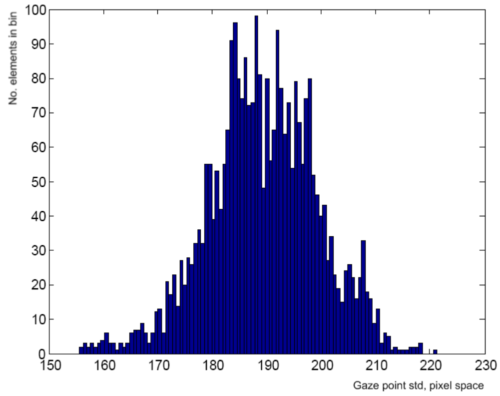

Figure 7: *Histogram of gaze point standard deviation within a smell congruency quadrant.*

where the weight (w) in Equation 1 is the ratio of the time spent looking at the smell congruency, versus the total time spent looking at the rest of the image. This, when averaged over all participants, gave a value for w of 0.621.

4.3. Measurements in Practice in Selective Rendering

Figure 8 highlights how our metric from measured data could be used in an interactive rendering pipeline. As there are insufficient resources to compute full global illumination effects in real time this pipeline allows resources to be redirected to the areas which are most relevant when a congruent smell impulse is delivered in the virtual environment.

5. Conclusions and Future Work

This paper has presented a method which exploits the HVS's bottom-up approach. The fact that the HVS is guided by other modal impulses allows us to selectively render the regions attended to in higher quality and the remainder of the image in a lower quality capitulating on Inattentional Blindness. The results extend previous work of smell and graphics, such as [BCB*09] and confirm the impact the inclusion of smell into a high-fidelity virtual environments has for selective rendering. Smell is a key human sense. Including it into virtual environments, not only improves their perceived realism, but also enables only those parts of the environments being attended to, to be rendered at the highest quality while the remainder of the scene can be computed at a much reduced quality for a significantly lower cost.

Future work will investigate the variability of the weighting function w across variable smell sets. In evolutionary terms, certain smells are more prevalent to certain people, and in fact, some demographs are anosmic or less responsive to some smells. The hedonic tone of a smell is highly

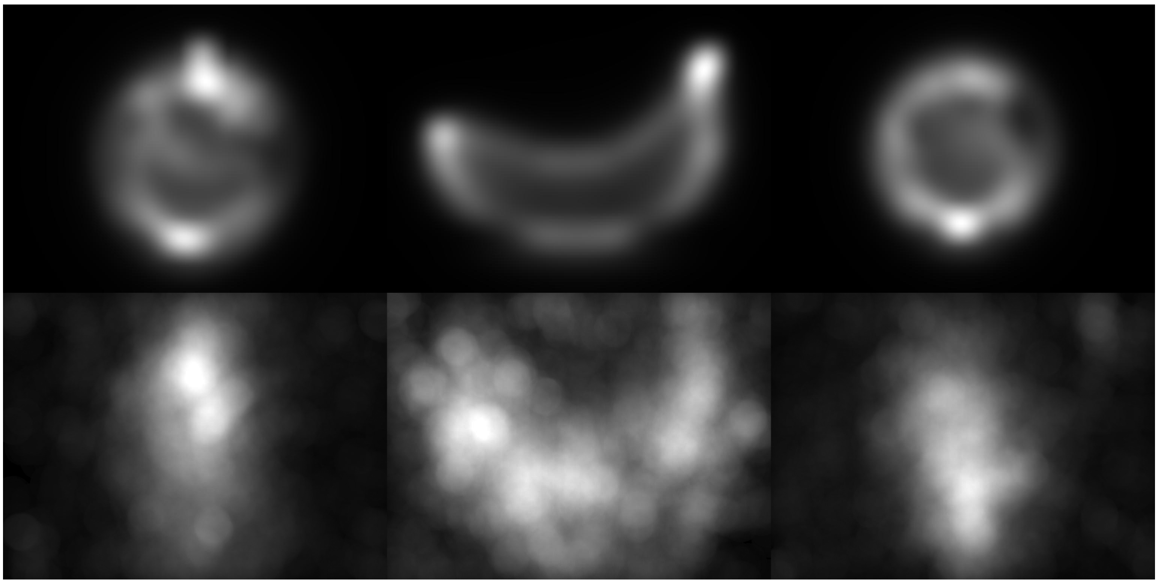

Figure 6: *Top Row: Saliency maps for smell emitting images used in the slides: (l) Apple, (m) Banana, (r) Orange. Bottom Row: The corresponding density estimation of the gaze points collected on the smell emitting visual congruency.*

subjective and this makes it hard to categorise smells across the board into pleasant or unpleasant but we wish to investigate whether the effect described here prevails when we have participants with a degree of anosmia to a certain smell and also when participants express variations of hedonic tone towards the exposure molcules. In addition we would like to investigate the effect multiple smell impulses have on visual attention. A similar avenue of research is to study the intensity of the smell impulse at which the effect on visual attention comes into play. Also we would like to look into the plausibility of the application to selective rendering. We also aim to investigate if smell saliency works as well as the previous saliency models where applied to selective rendering and answer if the rendering quality threshold is noticed.

References

[BCB*09] Brkic B., Chalmers A., Boulanger K., Pattanaik S., Covington J.: Cross-modal affects of smell on real-time rendering of grass. In *In SCCG'09: Spring Conference on Computer Graphics* (2009), ACM SIGGRAPH Press, pp. 175–179. 2, 5

[CCL02] Cater K., Chalmers A., Ledda P.: Selective quality rendering by exploiting human inattentional blindness: looking but not seeing. In *VRST '02: Proceedings of the ACM symposium on Virtual reality software and technology* (New York, NY, USA, 2002), ACM, pp. 17–24. 2

[CCW03] Cater K., Chalmers A., Ward G.: Detail to attention: Exploiting visual tasks for selective rendering. In *Eurographics Symposium on Rendering 2003* (June 2003), ACM, pp. 270–280. 2

[DC05] Debattista K., Chalmers A.: Component-based adaptive sampling. In *SIBGRAPI 2005* (October 2005), IEEE Computer Society Press, pp. 375–382. 2

[Dow87] Dowling J.: *The retina: An approachable part of the brain*. Cambridge: Belknap, 1987. 2

[EC06] Ellis G., Chalmers A.: The effect of translational ego-motion on the perception of high fidelity animations. In *In Spring Conference on Computer Graphics* (2006), ACM SIGGRAPH Press. 2

[HWBR*10] Harvey C., Walker S., Bashford-Rogers T., Debattista K., Chalmers A.: The Effect of Discretised and Fully Converged Spatialised Sound on Directional Attention and Distraction. In *Theory and Practice of Computer Graphics* (Sheffield, United Kingdom, 2010), Collomosse J., Grimstead I., (Eds.), Eurographics Association, pp. 191–198. 2

[IK00] Itti L., Koch C.: A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research 40*, 10-12 (May 2000), 1489–1506. 2

[IKN98] Itti L., Koch C., Niebur E.: A model of saliency-based visual attention for rapid scene analysis, 1998. 1, 2

[Jac07] Jacob T.: A tutorial on the sense of smell. World Wide Web electronic publication, http://www.cardiff.ac.uk/biosi/staffinfo/jacob/teaching/sensory/olfact1.html, 2007. 2

[Kna95] Knasko S. C.: Pleasant odors and congruency: Effects on approach behavior. *Chem Senses 20*, 5 (1995), 479 – 487. 3

[LM99] Loschky L., McConkie G.: Gaze contingent displays: Maximizing display bandwidth efficiency. In *ARL Federated Laboratory Advanced Displays and Interactive Displays Consortium* (1999), Advanced Displays and Interactive Displays Third Annual Symposium, pp. 79–83. 2

[LMYM01] Loschky L., McConkie G., Yang J., Miller M.: Perceptual effects of a gaze-contingent multi-resolution display based on a model of visual sensitivity. In *ARL Federated Laboratory Advanced Displays and Interactive Displays Consortium* (2001), Advanced Displays and Interactive Displays Fifth Annual Symposium, pp. 53–58. 2

[Mas06] MASTOROPOULOU G.: The effect of audio on the visual perception of high-fidelity animated 3d computer graphics. PhD Thesis, University of Bristol, 2006. 1

[MBM02] MILLOT J. L., BRAND G., MORAND N.: Effects of ambient odors on reaction time in humans. *Neurosci Lett. 322*, 2 (Apr 2002), 79 – 82. 3

[MDCT05a] MASTOROPOULOU G., DEBATTISTA K., CHALMERS A., TROSCIANCO T.: Auditory bias of visual attention for perceptually-guided selective rendering of animations. In *GRAPHITE '05: Proceedings of the 3rd international conference on Computer graphics and interactive techniques in Australasia and South East Asia* (2005), ACM Press, pp. 363–369. 2

[MDCT05b] MASTOROPOULOU G., DEBATTISTA K., CHALMERS A., TROSCIANCO T.: The influence of sound effects on the perceived smoothness of rendered animations. In *APGV 2005: Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization* (August 2005), ACM SIGGRAPH, pp. 9–15. 1

[ML97] MCCONKIE G., LOSCHKY L.: Human performance with a gaze-linked multi-resolutional display. In *ARL Federated Laboratory Advanced Displays and Interactive Displays Consortium* (1997), Advanced Displays and Interactive Displays First Annual Symposium, pp. 25–34. 2

[MR98] MACK A., ROCK I.: Inattentional blindness. *Massachusetts Institute of Technology Press* (1998). 2

[NBH*08] NOESSELT T., BERGMANN D., HAKE M., HEINZE H.-J., FENDRICH R.: Sound increases the saliency of visual events. *Brain Research 1220* (2008), 157 – 163. Active Listening. 2

[SDC05] SUNDSTEDT V., DEBATTISTA K., CHALMERS A.: Perceived aliasing thresholds in high-fidelity rendering. In *Proceedings of the 2nd symposium on Applied perception in graphics and visualization* (New York, NY, USA, 2005), APGV '05, ACM, pp. 166–166. 2

[SDJ*10] SEIGNEURIC A., DURAND K., JIANG T., BAUDOUIN J., SCHAAL B.: The nose tells it to the eyes: crossmodal associations between olfaction and vision. *Perception 39*, 10 (2010), 1541 – 1554. 3

[SIS*05] SAKAI N., IMADA S., SAITO S., KOBAYAKAWA T., DEGUCHI Y.: The effect of visual images on perception of odors. *Chem Senses 30*, 1 (2005), 244 – 245. 2

[SRMN10] SEO H.-S., ROIDL E., MÜLLER F., NEGOIAS S.: Odors enhance visual attention to congruent objects. *Appetite 54*, 3 (Jun 2010), 544 – 549. Epub. 3

[Yar67] YARBUS A.: Eye movements during perception of complex objects. *In L. A. Riggs, Ed., Eye Movements and Vision 7* (1967), 171–196. 2

[YPG01] YEE H., PATTANAIK S., GREENBERG D. P.: Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Trans. Graph. 20*, 1 (2001), 39–65. 1, 2

[ZBE91] ZELLNER D. A., BARTOLI A. M., ECKARD R.: Influence of color on odor identification and liking ratings. *The American Journal of Psychology 104*, 4 (1991), 547 – 561. 2

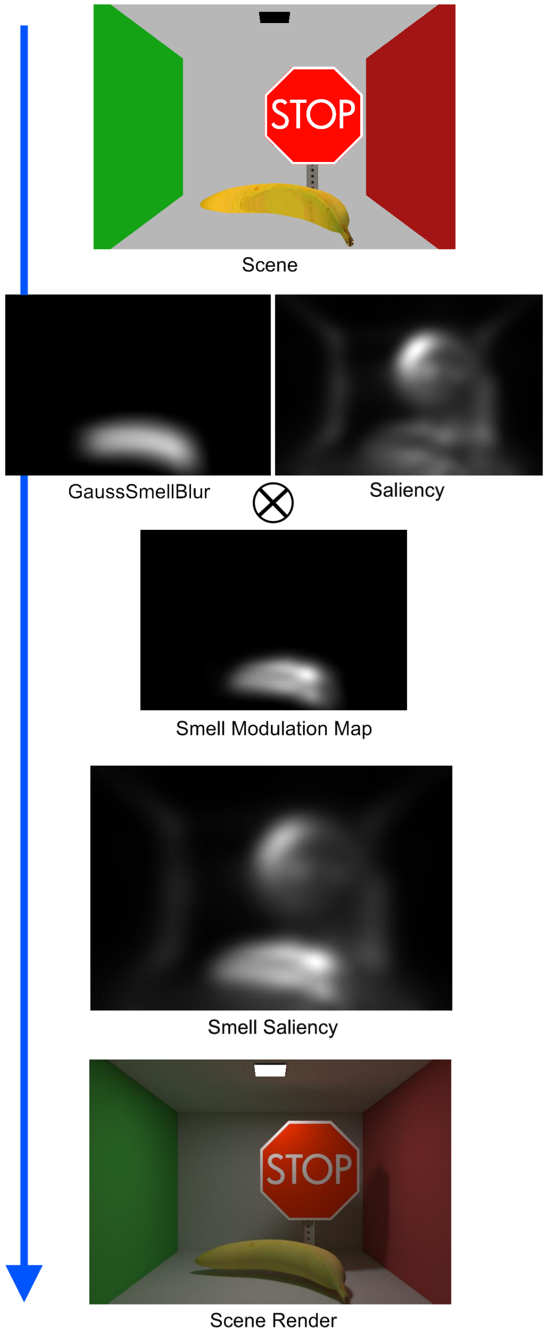

Figure 8: *Selective rendering saliency remodulation pipeline for a virtual scene; highlighting the procedure in Equation 1.*

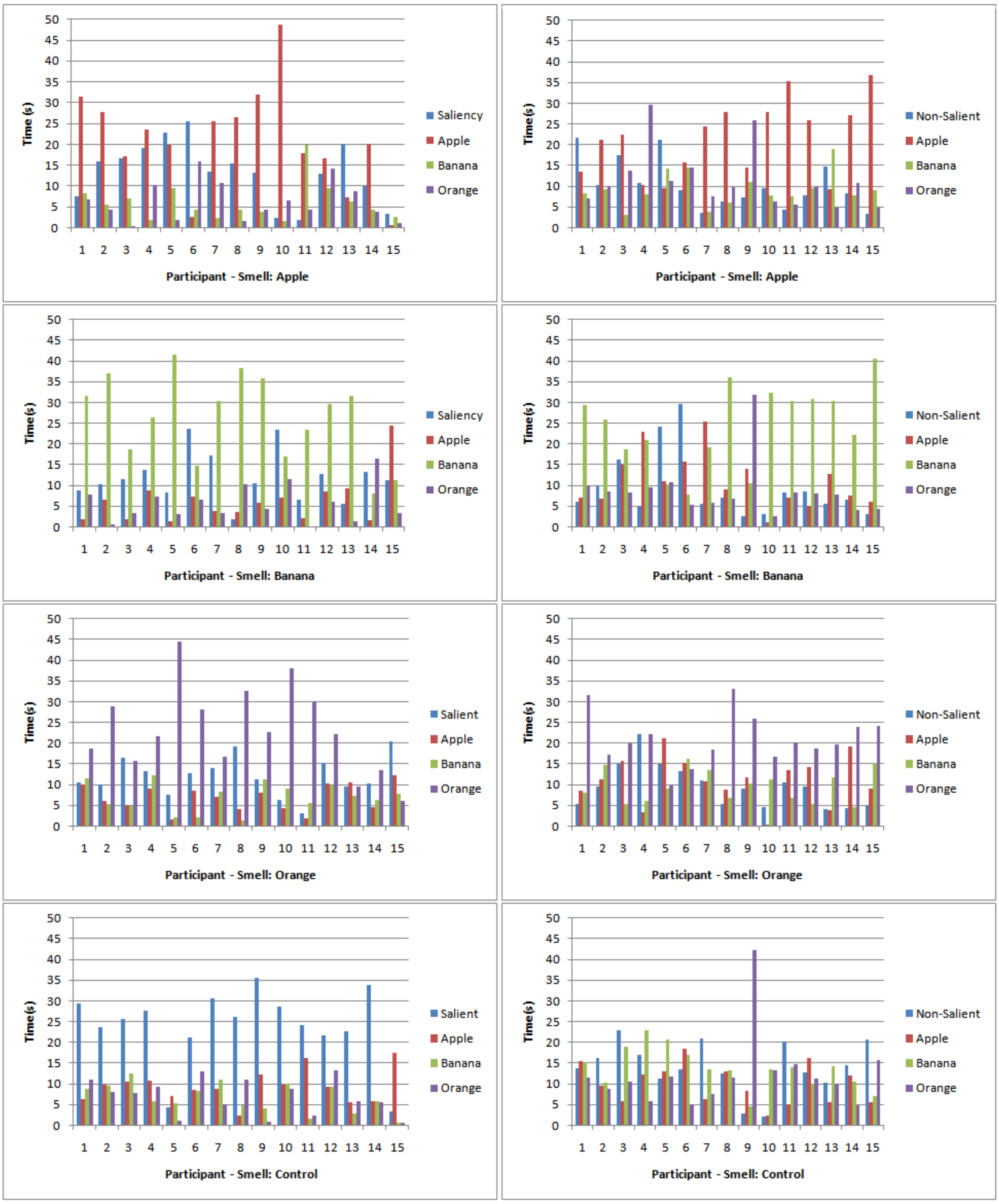

Figure 9: *Graphs of participants gaze durations in slide quadrants; Salient set (l), Non-Salient set (r). Smell emitted: Apple (top), Banana (second top), Orange (second bottom), Control/ No Smell (bottom).*