Experiences on Attention Direction through Manipulation of Salient Features

Erick Mendez*
Graz University of Technology

Dieter Schmalstieg†
Graz University of Technology

Steven Feiner‡
Columbia University

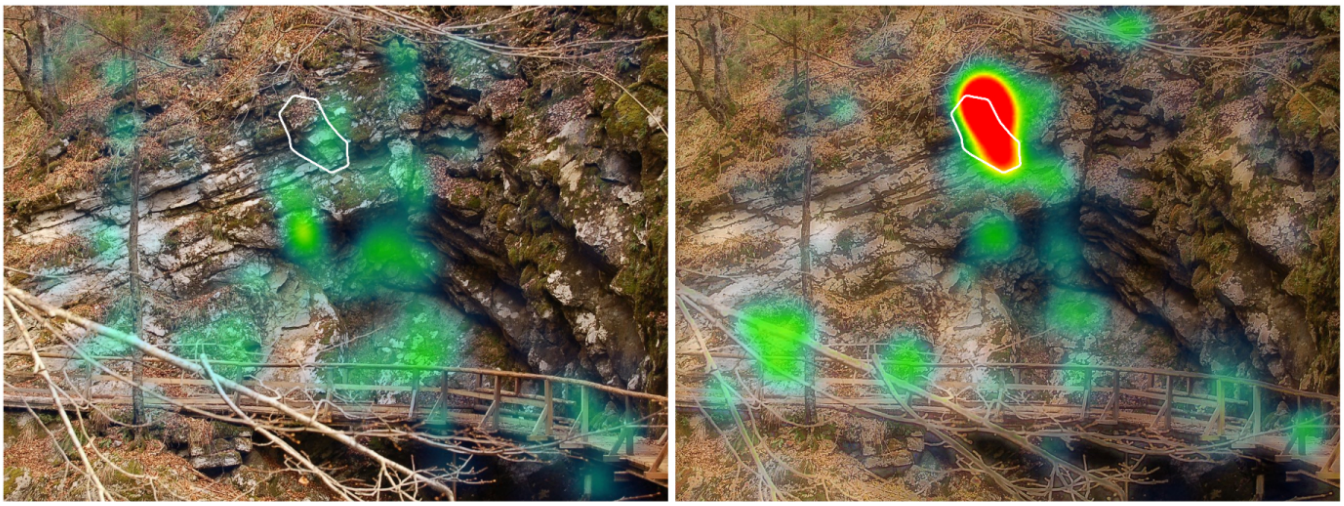

Figure 1: Heatmap of user fixations. These figure shows two example images used for our study, the first is the original unmodified source and the second is the result after our modulation procedure. The left image shows the heatmap generated by the fixations on the unmodified image. The right image shows the heatmap after our modulation technique. Notice how a large portion of the fixations land inside the defined focus region (white outline). Figure 2 shows the image before and after modulation without the eye gaze heatmaps.

Abstract

In this paper we present a study on the impact on user's fixations by localized modifications of the color properties of an image. We outline the technique used to influence the fixations of the participants, where it succeeded and where it failed. The question we address is whether pixel-wise modifications are sufficient to influence the order of fixations of the users and how strong this effect is. We developed a technique that performs localized adjustment in *value*, *saturation* and *hue* and conducted a user study with an eye tracker to asses its effectiveness. The results of our study imply that one can effectively influence the order and the duration of fixations of users based on localized adjustments.

1 Introduction and Related Work

This work addresses the efforts of directing the attention of users by modulating bottom up salient features of the image. We are interested in finding how the users observe an image and in which order fixations take place. Once having a measure of which locations are visually salient, we intend to direct the user's attention towards a predefined location. We explore how this can be done effectively and what the limitations are. Figure 1 illustrates the effectiveness of our technique with an attention heatmap. This map shows how a large portion of the fixations and their duration was spent on the focus region (white outline) after our modulation procedure. This modulation was done by considering how visually salient each pixel in the image is.

In an image, an object is said to be visually salient if it stands out more than its surrounding neighborhood [9]. The saliency is a measure that states what is the likelihood of this location to call our attention. The *conspicuities* of a location are measures that represent how contrasting this location is to its surroundings in dimensions such as color, orientation, motion, or depth separately [6] [7] . Treisman and Gelade [12] use the term 'dimension' to refer to the range of variations, and 'feature' to refer to values in a dimension. The visual saliency of a location is the combination of all its *conspicuities*. A scene's saliency map is a map of the saliency values on each location in the image. Two types of factors influence saliency: bottom-up and top-down. Bottom-up factors depend on instantaneous sensory input, such as brightness and size, while top-down factors consider the internal state of the user, such as experiences or goals.

In the work presented here we focus on bottom-up *conspicuities* based on the model suggested by Koch and Ullman [7]. We adopt bottom-up factors that lend themselves to pixel-wise manipulation, namely *value* (brightness), *saturation* and $color-opponency$ (hue). Color opponency is based on the opponent process theory [4], stating that we perceive colors by processing the differences between opponents, red-green and blue-yellow [6]. There is much evidence that there is a correlation between our visual attention and the saliency map. Ouerhani et al. [10] and similarly Santella et al. [11] used an eye tracker to confirm that there exists a relationship between the saliency map and human visual attention. Lee et al.

*e-mail: mendez@icg.tugraz.at
†e-mail: schmalstieg@icg.tugraz.at
‡e-mail:feiner@cs.columbia.edu

[9] went one step further by using the saliency map to track objects being attended by the user.

The most relevant work to our research was carried out by Bailey et al. [2] on subtle gaze direction. Bailey et al. apply first-order modulations to the focus, only when the user is not looking there, as determined by an eye tracker. The modulations are not, however, dynamic. In contrast to that technique, our technique works at interactive frame rates without requiring an eye tracker and can thus support mediated reality applications with arbitrary scenes.

In this paper we share our experiences on trying to influence the user's attention, our methodology and an assessment of its success. We present the failure and success criteria and we outline the next steps for attention direction.

2 Modulation technique

The techniques used for directing the user's attention are often referred as Focus and Context (F+C) techniques [8]. Focus is the term used for the portion of the image towards which we wish to draw the attention. Context is the portion of the image we want to distract the user from.

In order to direct the attention of the viewer to a particular location, one can either increase the saliency of that location (focus), decrease the saliency of its surroundings (context), or do both. Because the saliency of a location is a combination of several *conspicuities*, the final goal is to modulate the appropriate *conspicuities* by location. Our modulation procedure is straightforward: first we quantify for each location how conspicuous it is, or in other words, what its likelihood to attract our attention is. Then, given a classification of what is focus and what is context, we compute the average $conspicuity$ values of the focus region. Subsequently, we compute the difference between each location's *conspicuities* and the average *conspicuities* of the focus. And finally, we use this difference to modify the saliency of the locations of the context. This implies that all modulations are not applied equally to all fragments. For example, some might need a strong $saturation$ modulation, but no *value* changes. It is important to note that in the present technique we concentrate on simply modulating the context region and we leave the focus region un-modulated.

Analysis. The calculation of the saliency of a location and hence its *conspicuities* has been widely published in the past. We use the method described by Itti et al. [6] with the modifications suggested by Lee et al. [9]. However, one difference to both descriptions is that during the *conspicuities* computation step we do not use the absolute value for the difference across pyramid levels. Instead we keep the sign in order to know the type of $conspicuity$ we are dealing with. This determines whether a location is highly conspicuous because it is too bright or because it is too dark. As mentioned before, we also need the average of the $conspicuity$ values in the focus region. This will give us a measure of whether a location needs modulation at all or not.

We now have the components necessary for the modulation step: the *conspicuities* at every location, and the average *conspicuities* in the focus area. We proceed to modify the video image only in the context region, while the focus remains unchanged. We modulate every $conspicuity$ in the same dimension space where it was measured: *value* (brightness), $saturation$, and $color-opponency$ (hue).

Modulation. A naïve modulation method would heavily decrease all the *conspicuities* of the context area regardless of what is present in the scene (for example, turn all of the context to black). This increases the emphasis of the focus object, but at the same time drastically reduces the contribution of the context, since all pixels are modified whether the change was necessary or not. Better discrimination can be achieved by choosing an appropriate dimension ($saturation$ or *value*, for example) on a per pixel basis in which to modify the image.

Figure 2: Modulation of bottom-up salient features of context. (Top) Original image. (Bottom) Result of our modulation technique. The saliency of pixels outside the focus region is automatically decreased. Pixel values of the modulated image differ on average by 2.25% from their counterparts in the original image.

Modulation is performed in three sequential steps: *value* (brightness) modulation, $saturation$ modulation, and $color-opponency$ modulation. All three modulations take place in the HSV color space. *Value* and $saturation$ are very similar as they involve one subtraction. Modulation of *value* is performed by changing the fragment's color space from RGB to HSV space. Once this conversion is done, we compute the difference between the fragments' $value\ conspicuity$ and the average $value\ conspicuity$ of the focus. We increase or decrease the *value* by this difference. Modulation of $saturation$ is identical; this time, however, the operation is done in the $saturation$ channel.

Finally, the modulation of $color-opponency$ takes place in the hue channel; however, its computation is a bit more complex. The $color-opponency$ theory states that the color channel pairs Red-Green and Blue-Yellow are each mutually opposing. The HSV space arranges colors in a cone form where the Hue, is encoded by the angular position around the cone [3]. The red color is at 0 degrees, yellow at 60, green at 120, cyan at 180, blue at 240 and magenta at 300 degrees. However, the fragment's hue does not tell us anything about its surroundings. Such information is encoded in the *conspicuities*. For example, assume that the fragment has a stronger

Blue-Yellow than Red-Green $color-opponency\ conspicuity$. This means that blue and yellow give a high contrast to the fragment and we should move its Hue away from either of them. We calculate then the distance of the fragments's hue to both red and green. For example, if the distance to red is shorter we may move the hue towards it. In this manner we can decrease the contrast created by the blue and yellow opponents.

3 User Study

To test the effect on the order and length of fixations of the images, we carried out a formal user study with an eye tracker. The goal of the study was to test whether we could influence the visual attention of the user to the regions of the images that we designated as focus, regardless of the information in the original image. Our hypotheses were:

- **H1.** The time before the first fixation on the focus region will be smaller for the images modulated than for the original unmodified images.
- **H2.** The total fixation time (i.e., sum of durations of all fixations on the focus) will be higher for the modified images than for the original unmodified images.
- **H3.** The percentage of participants that have at least one fixation on the focus region will be higher for the images modified than for the original unmodified images.

3.1 Experiment description

The experiment was composed of two phases: Artificial and Natural. The artificial set was created in order to verify the effectiveness of each of the three dimension modulations (*value*, *saturation* and $color-opponency$) separately. The Natural set of images involved images from real scenarios and was the main target of our experiment.

3.1.1 Artificial images

The artificial images were created with a popular graphics package, Adobe Photoshop. The images included an arrangement of stimuli in either of the three dimensions for a total of three image sets: artificial-value, artificial-saturation, Artificial-Color-Opponency. Each image was created with a single focus area and was subsequently modulated with a single step of our technique (*value* modulation, *saturation* modulation, or $color-opponency$ modulation) to change the focus to a different location. Figure 3, for example, shows an unmodified image of the artificial-value set (left), where the focus is the object with lower *value*, and the same image after our modulation technique (right). Subsequently, figure 4 shows a pair of original and modified images on the *saturation* dimension.

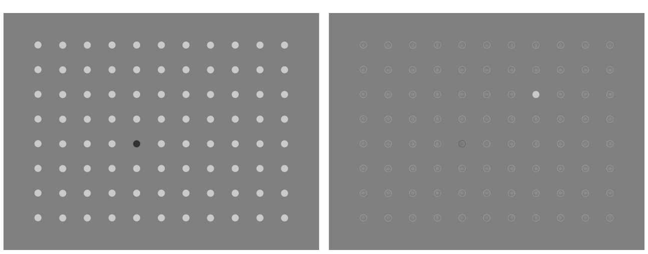

Figure 3: Example images from the artificial-value set. (Left) Image created artificially, where one of the circles has lower *value* than the rest. (Right) The same image after our modulation process tries to move the attention of the subject to a different region of the image.

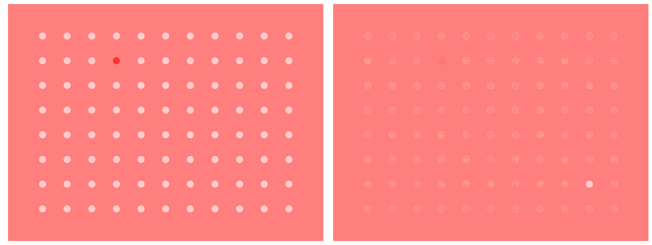

Figure 4: Example images from the artificial-saturation set. (Left) Image created artificially where one of the circles has higher *saturation* than the rest. (Right) The same image after our modulation process tries to draw the attention of the participant to a different region of the image.

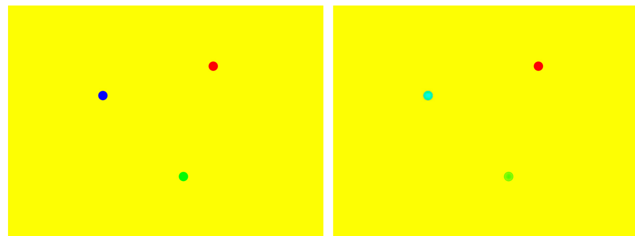

Figure 5: Example images from the artificial-color-opponency set. (Left) Image created artificially with a yellow background and three circles in red, green and blue. According to the opponent process theory, the blue circle is the highest salient. (Right) The same image after our modulation process tries to move the attention of the subject to the red circle. Notice how the hue of the blue has been moved towards cyan.

The effect of the modulation on $color-opponency$ of the four colors, red, green, blue and yellow was tested by, for example, an image with one of these colors as the background and three circles with the remaining colors. Figure 5 shows an example image from the artificial-color-opponency set for yellow. According to the color opponent process theory, the blue circle is the most salient of the three circles. The right image shows the result after saliency modulation which directs the attention of the subject is to the red circle. Notice how the blue circle is moved towards the cyan hue.

3.1.2 Natural images

The natural image set was a collection of photographs of outdoor environments. These images were selected to include gardens, streets, people gatherings, city landscapes and so on. They included many other dimensions of attention such as shape, orientation or texture detail, which are not considered for modulation by our technique. The designated focus regions (i.e., regions we want the participants to look at) were scattered around, including and excluding human faces, perspective vanishing points on the horizon, etc. These focus regions were always placed away from the region of the original image with the highest salient. Every image was modulated with all three dimensions sequentially: *value*, *saturation*, and $color-opponency$.

Figure 2 and Figure 6 each show an image before and after our modulation procedure. Figure 2 shows a bridge over a gorge with a wall of rocks in back. The modulated image tries to direct the attention of the user towards a rock on the upper part of the wall. Figure 6 shows a busy street with multiple colors, perspective lines, faces and so on. The modulated image tries to direct the attention towards the second farthest lamp on the right hand side.

The order in which the two sets of images were presented was randomized, as was the order of presentation of images within each

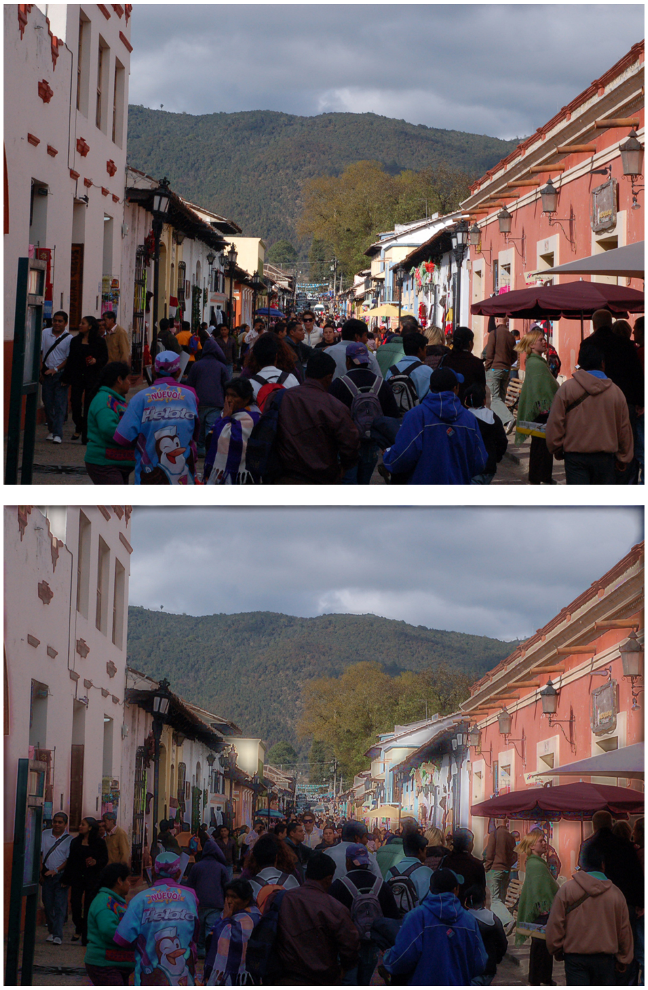

Figure 6: Picture of a busy street before and after modulation. Attention was directed towards the second farthest lamp on the wall on the right. The pixel values of the modified image differ on average by 1.84% from their counterparts on the original image.

set. There were a total of 24 images in the Artificial set (6 for artificial-value, 6 for artificial-saturation and 12 for artificial-color-opponency) and 27 in the Natural set. Each image was shown for 5000ms. In between images, a blank slide with a cross in the middle was shown for 2000ms in order to standardize the participant's initial gaze position before the image was presented.

3.1.3 Participants

The test was performed on 30 participants (6 female, 24 male) between 24 and 34 years old ($\bar{x}$=28). All participants had no known color sensitivity deficiencies; which was confirmed by an Ishihara color test. Participants were compensated with a gift certificate from a popular online shop.

3.1.4 Apparatus

The tracking device was an SMI desktop-mounted eye tracker, operating at 60 Hz. The stimuli were presented in the form of a slide show on a 19” monitor at a 70cm distance from the participant. The resolution of the images was 1280x960, and all were presented without resizing in order to avoid interpolation by the graphics card.

3.2 Analysis

Analysis was performed with paired-samples t-tests, α levels were adjusted (Bonferroni) to ensure a level of 5%. We discarded the data for the first 200ms after the image was shown, based on the assumption that this is the amount of time the brain needs to build the saliency map [5].

3.2.1 Analysis of the artificial images

Tables 1-9 show the results of the tests for the artificial images of all three tested dimensions separately. Standard deviations and statistical significances are stated. It is important to mention that there was no interaction among the sets since the images used for each set were different.

value	Original [ms]	Modulated [ms]	Speed up [ms]
H1	4069 (σ=766.9)	2141.4 (σ=1079.5)	-1927.6

Table 1: Results for hypothesis **H1** of the artificial-value set (average time before participants fixated on the focus). These results are statistically significant ($p < .00625$).

value	Original [ms]	Modulated [ms]	Increment [ms]
H2	102.61 (σ=92.5)	819.7 (σ=402.2)	717.1

Table 2: Results for hypothesis **H2** of the artificial-value set (sum of the total fixation time that participants spent on the focus). These results are statistically significant ($p < .00625$).

value	Original [%]	Modulated [%]	Difference [%]
500ms	0	33.3	33.3
1000ms	5.8	50	44.1
5000ms	20	79.1	59.1

Table 3: Results for hypothesis **H3** of the artificial-value set (percentage of participants that had at least one fixation on the focus region by the first 500ms, 1000ms, and 5000ms).

Artificial-value. Tables 1-3 show the results of the analysis of the artificial-value set. Participants had a first fixation on the focus significantly faster on the images modulated with our technique (2141.4ms) than on the original unmodified images (4069ms) (Table 1). Participants also spent significantly longer time fixated on the focus of the images modulated with our technique (819.73ms) than on the original unmodified images (102.61ms) (Table 2). By the first 500ms, 33.33% of the participants had had already at least one fixation on the focus of the modulated images while none of the participants had fixated on the focus of any of the original unmodified images. By the end of the stimulus, 59.16% more of the participants had at least one fixation on the focus on the modulated images than on the original unmodified images (Table 3).

Artificial-saturation. Tables 4-6 show the results of the analysis of the artificial-saturation set. Participants had a first fixation on the focus significantly faster on the images modulated with our technique (1736ms) than on the original unmodified images (4596.1ms) (Table 4). Participants also spent, significantly longer time fixated on the focus of the images modulated with our technique (742.7ms) than on the original unmodified images (35.48ms) (Table 5). By the first 500ms 38.33% of the participants had had already at least one fixation on the focus of the modulated image while none of the participants had a fixation on the focus of any of the original unmodified images. By the end of the stimulus 57.5% more of the participants had at least one fixation on the focus of a modulated image than on its original unmodified image (Table 6).

Artificial-color-opponency. Tables 7-9 show the analysis results for the artificial-color-opponency set. The results in this table for

saturation	Original [ms]	Modulated [ms]	Speed up [ms]
H1	4596.1 (σ=475.2)	1736 (σ=867.2)	-2860.1

Table 4: Results for hypothesis **H1** of the artificial-saturation set (average time before participants fixated on the focus). These results are statistically significant ($p < .00625$).

saturation	Original [ms]	Modulated [ms]	Increment [ms]
H2	35.4 (σ=43.0)	742.7 (σ=329.3)	707.2

Table 5: Results for hypothesis **H2** of the artificial-saturation set (sum of the total fixation time that participants spent on the focus). These results are statistically significant ($p < .00625$).

saturation	Original [%]	Modulated [%]	Difference [%]
500ms	0	38.3	38.3
1000ms	1.6	60	58.3
5000ms	17.5	75	57.5

Table 6: Results for hypothesis **H3** of the artificial-saturation set (percentage of participants that had at least one fixation on the focus region by the first 500ms, 1000ms, and 5000ms).

c.opp.	Original [ms]	Modulated [ms]	Speed up [ms]
H1	1896.6 (σ=713.4)	2077.3 (σ=632.6)	180.7

Table 7: Results for hypothesis **H1** of the artificial-color-opponency set (average time before participants fixated on the focus). These results are not statistically significant ($p > .00625$).

c.opp.	Original [ms]	Modulated [ms]	Increment [ms]
H2	848.8 (σ=355.2)	833.3 (σ=279.9)	-15.5

Table 8: Results for hypothesis **H2** of the artificial-color-opponency set (sum of the total fixation time that participants spent on the focus). These results are not statistically significant ($p > .00625$).

c.opp.	Original [%]	Modulated [%]	Difference [%]
500ms	17.5	22.9	5.4
1000ms	47	47.5	0.4
5000ms	82	79.58	-2.5

Table 9: Results for hypothesis **H3** of the artificial-color-opponency set (percentage of participants that had at least one fixation on the focus region by the first 500ms, 1000ms, and 5000ms).

hypotheses H1 and H2 were not found to be statistically significant. Participants tended to spend less time before fixating for the first time on the focus on the original unmodified images (1896.6ms), than on the modulated counterpart of the same image (2077.3ms) (Table 7). Participants also tended to spend more time fixated on the focus on the original unmodified images (848.85ms), than on the modulated counterpart of the same image (833.33ms) (Table 8). Interestingly, on the third hypothesis (H3), by the first 500ms the modulated images of the $color-opponency$ dimension called the attention of an extra 5.41% of the participants but by the end of the presentation of the stimulus, the modulation had a negative impact of 2.5% (Table 9).

3.2.2 Analysis of the natural images

Tables 10-12 show the results of the tests for the natural images. The results are placed in contrast with the hypotheses formulated earlier. Standard deviations and statistical significances are stated. It is important to note, that each image in this set was modulated across all three dimensions (unlike the images in the artificial sets, which were each modulated in only one dimension).

H1, duration before first fixation. Table 10 shows the average time that passed before the participants fixated on the focus region for the first time in both original and modulated conditions, as well as the speed up (difference) across all natural images. Participants had a first fixation on the focus significantly faster on the images modulated with our technique (3382.8ms) than on the original unmodified images (3842.3ms).

H2, total fixation time. Table 11 shows the average total time that the participants spent fixated on the focus region in the original and modulated conditions, as well as the difference. Participants spent significantly longer time fixated on the focus of the images modulated with our technique (338.31ms) than on the original unmodified images (176ms).

H3, percentage of participants with at least one fixation. Table 12 shows the percentage of the participants that had at least one fixation on the focus region by the first 500ms, 1000ms and by the end of the stimulus, 5000ms across all natural images. By the first 500ms, 8.33% had had already at least one fixation on the focus of the modulated image while only 2.96% of the participants had a fixation on the focus on the original unmodified counterpart, a factor of two. By the end of the stimulus an extra 14.63% of the participants had had at least one fixation on the focus on the modulated images than on the original unmodified images.

4 Discussion

As can be seen, the artificial sets for *value* and *saturation* were successfully modulated, satisfying all three hypotheses. In average, *value* and *saturation* modulation presented a significant speed up of first fixation time and an significant increase on total fixation time as well as a higher percentage of participants having at least one fixation on the focus region. However, no statistically significant results were obtained for modulation of the $color-opponency$ dimension. We speculate that the reason why the $color-opponency$ failed to direct the attention of the user was the relative number of absolute elements in the scene. The *value* and *saturation* sets had an array with a total of 88 elements of which one would call the attention of the user. In contrast the $color-opponency$ set had only three elements (one for each hue) of which one would be the focus of attention.

As can be seen from the results, our modulation procedure could also effectively draw the attention of the participants to the focus region in the natural set of images. Figure 1 shows the heatmaps of the input image in its unmodified (left) and modified state (right), these heatmaps correspond to the images shown in figure 2. A white outline denotes the position of the focus region. It can be noted from this image that the user fixations spent a significant amount of time on the focus.

Figure 7 shows graphs of the saliency map of Figure 2. The top graph was obtained from the original image; a red overlay indicates the region we wish to set as focus. As it can be seen, the image has many high salient locations competing for the user's attention. The bottom graph shows the result after our modulation process and how our technique effectively keeps the object of interest highly salient in the scene. To find out the how much our technique changes the image, we computed the average pixel difference between the original image and after our modulation procedure. This was done by calculating the square root of the sum of squared differences in the RGB space divided by the number of pixels in the image. For example, the pixels of the right image of Figure 2 differ on average by 2.25% from their counterparts in the left image. The total average pixel difference across all images in the user study between modified and original images is 1.86%.

Natural	Original [ms]	Modulated [ms]	Speed up [ms]
H1	3842.3 (σ=412.3)	3382.8 (σ=463.1)	-459.5

Table 10: Results for hypothesis **H1** of the Natural set. This table shows the average time before participants fixated on the focus. These results are statistically significant ($p < .00625$).

Natural	Original [ms]	Modulated [ms]	Increment [ms]
H2	176 (σ=65.63)	338.3 (σ=115.6)	162.3

Table 11: Results for hypothesis **H2** of the Natural set. This table shows the sum of the total fixation time that participants spent on the focus. These results are statistically significant ($p < .00625$).

Natural	Original [%]	Modulated [%]	Difference [%]
500ms	2.9	8.3	5.3
1000ms	11.2	20.3	9.08
5000ms	40.9	55.5	14.6

Table 12: Results for hypothesis **H3** of the Natural set. This table shows the percentage of participants that had at least one fixation on the focus region by the first 500ms, 1000ms, and 5000ms.

Although numerically the difference between the original and the modified image is small, in practice the damage of the modified image is noticeable. The cause of this damage is twofold: First, the image modifications were done without considering spatial coherence, this means that spatially close fragments could be matched to different modulation procedures, such as one fragment being darkened to decrease its *conspicuity*, while a neighbor was lightened for the same purpose. Second, the modulation procedure is a single modulation pass, this means that there is no iterative process to try to solve overshot modulations. We have worked on a better approach to fix these problems and the manuscript describing it is currently under review process [1]. This new approach works by doing multiple modulation passes and calculating the *conspicuity* values before each modulation to verify whether changes are necessary at all. Moreover, the iterations are made on different levels of the image pyramid in order to carry large changes first and progressively refine them. A final consideration is to use a different color space for the modulations. The HSV space is quite convenient for handling *value* and *saturation* but does not support an intrinsic representation of $color-opponency$. Our new approach works on the CIEL*a*b* space, which is based on the opponent process [3].

We have presented a user study to asses the effects of localized modulations on real world images. The results of the study indicate that we can effectively manipulate the attention of the users by local modifications. After our modulation procedure, we could direct the user's attention to a target location significantly faster and retain it significantly longer than the original unmodified image. We showed that a higher percentage of the participants had a least one fixation in our designated focus region.

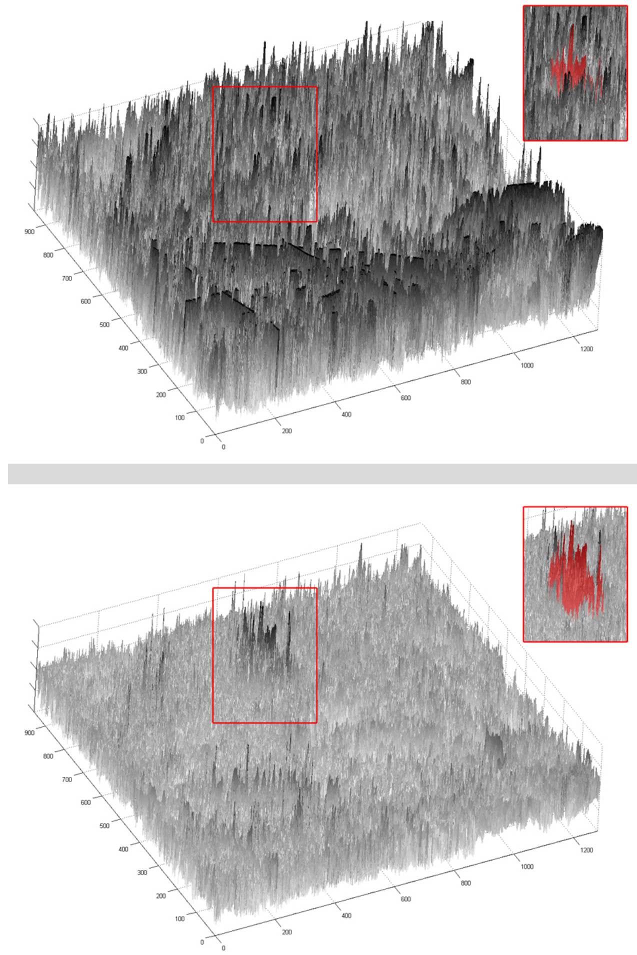

Figure 7: Comparison of methods. Graphs of the saliencies of Figure 2. (Top) Original unmodified image; many locations in the scene are competing for the user's attention. (Bottom) Modulated image. The modulation step eliminates competitors and clearly isolates the focus. Plots were created in an inverse scale with black as the highest value for better readability.

References

[1] Anonymous. Focus and context in mixed reality by modulating first order salient features. In *Under Review Process: Smartgraphics 2010*. Springer, 2010.

[2] R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm. Subtle gaze direction. 28, 2009.

[3] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. Achromatic and colored light - the hls color model. In *Computer Graphics: Principles and Practice*, pages 592–595. Massachusetts. 1996. 2nd ed.. 1174 p., 1996.

[4] E. Hering. Outlines of a theory of the light sense. pages 292–302, 1964.

[5] L. Itti. *Models of bottom-up and top-down visual attention*. PhD thesis, Pasadena, CA, USA, 2000. Adviser-Koch, Christof.

[6] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. In *IEEE Trans. Pattern Analysis and Machine Intelligence 20, 11*, pages 1254–1259, 1998.

[7] C. Koch and S. Ullman. Shifts in selective visual attention. 4:219–227, 1985.

[8] R. Kosara, H. Hauser, and D. Gresh. An interaction view on information visualization. In *EUROGRAPHICS*, pages 123–137, 2003.

[9] S. Lee, G. J. Kim, and S. Choi. Real-time tracking of visually attended objects in interactive virtual environments. In *VRST '07: Proceedings of the 2007 ACM symposium on Virtual reality software and technology*, pages 29–38, New York, NY, USA, 2007. ACM.

[10] N. Ouerhani, R. Von Wartburg, and H. Hugli. Empirical validation of the saliency-based model of visual attention. In *Electronic Letters on Computer Vision and Image Analysis*, volume 3, pages 13–24, 2007.

[11] A. Santella and D. DeCarlo. Visual interest and npr: an evaluation and manifesto. In *NPAR '04: Proceedings of the 3rd international symposium on Non-photorealistic animation and rendering*, pages 71–150, New York, NY, USA, 2004. ACM.

[12] A. M. Treisman and G. Gelade. A feature-integration theory of attention. In *Cognitive Psychology*, volume 12, pages 97–136, 1980.