Focus and Context in Mixed Reality by Modulating First Order Salient Features

Erick Mendez, Steven Feiner, and Dieter Schmalstieg

Graz University of Tehcnology
{mendez,schmalstieg}@icg.tugraz.at
Columbia University
feiner@cs.columbia.edu

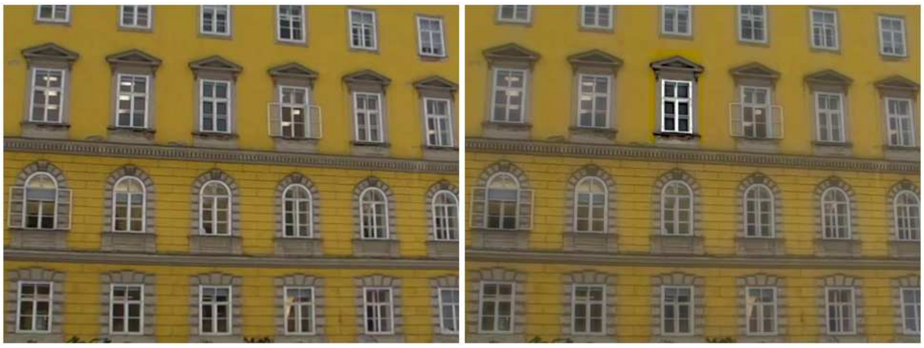

Fig. 1. Attention direction with our technique. (Left) Original image. (Right) Result of our modulation technique. The saliency of pixels is automatically decreased in the context area and increased in the focus area. Notice how the reflections of windows in the context area are slightly diminished yet not suppressed entirely. Pixel values of the modulated image differ on average by 1.28% from their counterparts in the original image.

Abstract. We present a technique for dynamically directing a viewer's attention to a focus object by analyzing and modulating bottom-up salient features of a video feed. Rather than applying a static modulation strategy, we inspect the original image's saliency map, and modify the image automatically to favor the focus object. Image fragments are adaptively darkened, lightened and manipulated in hue according to local contrast information rather than global parameters. The goal is to suggest rather than force the attention of the user towards a specific location. The technique's goal is to apply only minimal changes to an image, while achieving a desired difference of saliency between focus and context regions of the image. Our technique exhibits temporal and spatial coherence and runs at interactive frame rates using GPU shaders. We present several application examples from the field of Mixed Reality, or more precisely Mediated Reality.

1 Introduction

Focus and context (F+C) describes the concept of visually discriminating interesting objects (focus) from nearby related objects (context). In mixed reality, or more precisely mediated reality, F+C can be used to draw the attention of a user to certain objects of a scene, be it to indicate danger, to supplement detailed information, or to guide the user to a destination. There exist several strategies to achieve this, for example, by changing the color, overlaying augmenting artifacts, or distorting the area of attention [13].

Not all of the F+C strategies are universally effective, and the choice of technique depends heavily on the focus and context objects themselves. For example, suppose we were to draw the attention of the user to a particular region by drawing a circle around it. The effectiveness of this technique will depend on parameters such as the color or size of the circle and whether they offered sufficient contrast with the rest of the image.

Consequently, an adaptive discrimination of scene objects is needed, i. e., the F+C strategy has to be constantly adjusted. Specifically in mixed reality applications based on live video, one cannot easily impose constraints on visible objects or camera movements. The technique presented in this article therefore analyzes the video image in real time and computes the saliency for every fragment. In an image, an object is said to be visually salient if it stands out more than its surrounding neighborhood [14]. Our technique modifies the image by changing lightness and color contrast in order to have the highest attention salient inside the desired focus region. This is done in such as way that the applied changes are minimal and spatial and temporal coherence are respected. Consequently, the legibility of the context region is affected as little as possible.

All computations are carried out with GPU shaders in real time. We present several application examples from the field of mixed reality, including directing the attention of the use to an object in a search task, and highlighting a possibly dangerous object during car maintenance.

2 Background

There exists an extensive amount of work on trying to model the visual saliency of an image. The different techniques try including contextual information [24], non-parametric approaches [10], face detection [3] or using trained samples over large datasets [8].

The saliency is usually defined as a measure of how contrasting a particular location is from its surrounding in dimensions such as color, orientation, motion, and depth . Treisman and Gelade use *dimension* to refer to the range of variations, and *feature* to refer to values in a dimension (e.g., orientation and lightness are dimensions, while horizontal and dark are features) [25]. The *conspicuities* of a location are measures that represent how contrasting this location is to its surroundings in each of said dimensions. 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.

In this paper, we consider bottom-up saliency, which only relies on instantaneous sensory input and not on higher level factors such as intentions. Itti et al. provided a computational model for analysis of several bottom-up stimuli. From this work, we adopt those that lend themselves to pixel-wise manipulation, namely lightness, red-green color opponency and blue-yellow color opponency. This can be seen as a form of in-place F+C [13]. Highly conspicuous objects in the lightness dimensions are either dark objects in light surroundings or vice versa. Color opponency is based on the opponent process theory [5], stating that we perceive colors by processing the differences between opponents, red-green and blue-yellow [7]. This means that, for example, if two objects, one blue and one green, were placed on a yellow canvas, the blue will be more conspicuous due to its color opponency to yellow.

There is much evidence that there is a correlation between our visual attention and the saliency map. Ouerhani et al. [17] and similarly Santella et al. [19] used an eye tracker to confirm that there exists a relationship between the saliency map and human visual attention. Lee et al. [14] went one step further by using the saliency map to track objects being attended by the user.

Practically any change done to the image will modify its saliency map. Blurring, (de-) saturating, harmonizing and distorting are typical operations that implicitly change the saliency of the image. During the last few years there has been an increasing interest in directing the attention of the user through saliency manipulation for volume rendering [12], non-photorealistic stylization [19] and geometry [11]. However, previous work concentrates on creating salient features in focus regions rather than applying subtle modifications to existing images. For example, the work of Kim et al. [12] presents a visual-saliency-based operator to enhance selected regions of a volume. This operator is a part of the visualization pipeline and is applied before the image is generated, in contrast, our work receives an existing image as input and pursues the manipulation of its existing salient regions.

Closest to our intentions is the work by Su et al. [23] on de-emphasizing distracting image regions and by Bailey et al. [2] on subtle gaze direction. Su et al. focused on so-called second-order saliency by modulating variations in texture to redirect the user's attention to specific locations. Bailey et al. [2] apply first-order modulations to the focus, only when the user is not looking there, as determined by an eye tracker. In contrast to these techniques, our technique works with dynamic live video and can thus support mediated reality applications with arbitrary scenes and without requiring an eye tracker.

3 Conspicuities Analysis

We modulate the image on a frame-by-frame basis, in order to reflect the latest information available in the case of a video feed. Achieving this demands two general steps depending on the input before composing the final image:

- **Conspicuities analysis.** During this step, we compute the conspicuities of the whole image to have a measure of the naturally salient objects in the scene.
- **Conspicuities modulation.** (Section 4) Once we have quantified the image's conspicuities, we select and apply the appropriate modulations to the input image. The modulations are done sequentially for each of the conspicuities at multiple levels of coarseness, and ultimately produce an image whose highest salient is in the focus area.

The saliency of a location is given by a combination of its conspicuities, the final goal is then to modulate said conspicuities. We now present how the saliency of the image is analyzed so that modulation can take place. For simplicity, the term lightness is simply referred as L, red-green color opponency as RG and blue-yellow color opponency as BY.

Figure 2 illustrates the calculated conspicuities for L, RG and BY. For illustration purposes we show positive values in green and negative values in red, for example, dark objects near light objects have a negative conspicuity and it is shown in red. The right-most image shows the arithmetical average of the conspicuities representing the total saliency of the image.

In order to compute the conspicuity map of an image one must follow three steps: a) feature extraction, b) conspicuity computation and c) normalization. A feature is the value on a given dimension in a given location while conspicuity is the difference of the feature value of said location with its surroundings. Finally, the saliency is a combination of the conspicuity values.

Feature extraction. We use a slightly modified version of the conspicuity computation provided by Itti et al. [7]. That work computed the saliency of a location in the lightness, red-green color opponent, blue-yellow color opponent and orientation dimensions. We only compute the first three dimensions by converting the image from the RGB to the CIEL*a*b* space which already encodes the lightness and opponent colors dimensions similar to the work of Achanta et al. [1] (the initial RGB values are given in simplified sRGB with a Gamma of 2.2; we assume the observer at 2°, and use the D65 illuminant).

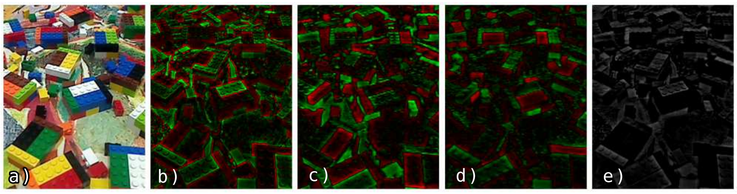

Fig. 2. Illustration of conspicuities. These images illustrate the conspicuities of the different dimensions used in this paper, green is used for positive values, while red is used for negative. a) Original image. b) Lightness conspicuity. c) Red-Green Color Opponency conspicuity. d) Blue-Yellow Color Opponency conspicuity. e) Saliency map.

Conspicuities computation. The next step computes the conspicuities for each separate dimension. This is done by calculating the difference between a location and its neighborhood by the center-surround technique. This technique calculates the relation of a location to its surroundings by checking the difference across fine and coarse levels. Accessing finer and coarser levels of the image is done by using the built-in hardware mip-mapping as suggested by Lee et al. [14]. The center-surround technique as described by Itti et al. [7] is:

Let k_i be the fragment's feature k on pyramid level i. The conspicuity c_k is then defined as:

$$c_k = \frac{\sum_{n=0}^{n=2} \sum_{m=n+3}^{m=n+4} k_n - k_{n+m}}{p} \tag{1}$$

Where k_i is the conspicuity value $k \in \{L, RG, BY\}$ at mipmap level i and $p = 6$. The value of p states the number of levels of coarseness being considered. An important difference between our work and others is that we do not use the absolute value of the conspicuities before adding them up. This allows us to keep the sign of the conspicuity, e.g. if the current location (fragment) has a negative lightness conspicuity then it is a dark location on light surroundings.

Normalization. We use a normalization that considers the global conspicuity maxima as described by Lee et al. [14]. This has the effect of reducing non-contributing high-frequency artifacts on each dimension. The normalized conspicuity is then defined as follows:

Let $max(c_k)$ be the maximum conspicuity value of the feature k of the whole image. The normalized conspicuity at every location n_k is then

$$n_k = \frac{c_k}{max(c_k)} \text{ where } k \in \{L, RG, BY\} \tag{2}$$

The calculation of the normalization weights is a computationally demanding task. We allow the computation of these weights to be done every few frames. The number of frames is determined by the current framerate of the system in order to maintain at least 15fps.

Saliency computation The saliency of a location is the arithmetical average of its normalized conspicuities. The computation of the saliency s at a given location is:

$$s = \frac{\sum n_k}{d} \text{ where } k \in \{L, RG, BY\} \text{ and } d = 3 \tag{3}$$

The value of d states the number of dimensions being considered.

4 Conspicuities Modulation

Modulating a location means either reducing its conspicuity (in the case of context) or increase it (in the case of focus). To modulate the conspicuity of a location we must, for example, lighten or darken it, reduce or increase its "redness" or "greenness". A brute force method of attention direction would heavily modify the image to highlight the object of interest, for example, by turning all the context area to black. Although effective, such a technique also eliminates the information of the context, since all fragments are suppressed whether it was necessary or not. The purpose of our algorithm is to apply the appropriate amount of change to the image such that the information of the context is not lost. This means that the modulations are not all applied to every fragment equally. For example, some fragments might need a strong red-green modulation, but no blue-yellow modulation. Modulation is performed in three sequential steps: first, lightness is modulated, then red-green opponency and finally blue-yellow opponency. The order of this sequence is given by the sensitivity of the human visual system in each dimension; we are more sensitive to lightness than to chromatic information, and we are more sensitive to red-green stimulus than to blue-yellow [22][21].

Classification. Before we can modulate the image, we need a classification of the objects in the scene. This classification tells us whether we want to direct attention towards a given object (focus) or away from it (context). In this paper, we assume that this classification is given through a-priori knowledge of the scene or user interaction.

Modulation thresholds. The set of conspicuities encodes the difference of every location with its surroundings. However, in order to modulate conspicuities adaptively, we need a threshold for every dimension to compare a locations's conspicuities to. For automatic determination of thresholds, we empirically found the average conspicuity values of the focus object to be a good value. Throughout the rest of this paper, the threshold values will be referred as t_k where k is the given dimension.

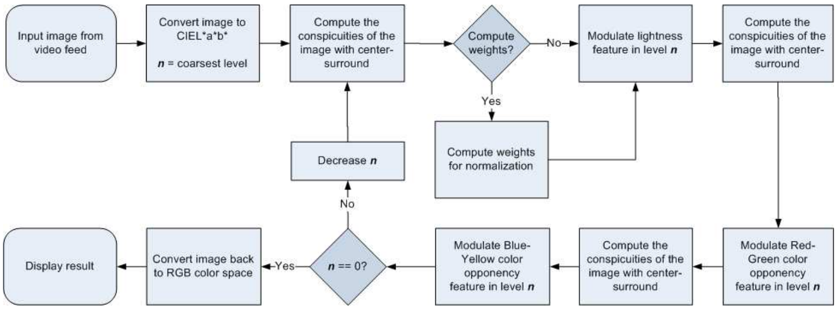
Input image from video feed
Convert image to CIEL*a*b*
n = coarsest level
Compute the conspicuities of the image with center-surround
Compute weights?
No
Modulate lightness feature in level n
Compute the conspicuities of the image with center-surround
Yes
Compute weights for normalization
Decrease n
No
Display result
Convert image back to RGB color space
Yes
n == 0?
Modulate Blue-Yellow color opponency feature in level n
Compute the conspicuities of the image with center-surround
Modulate Red-Green color opponency feature in level n

Fig. 3. Flow chart of our technique. This flow chart illustrates the iterative process of our modulation technique.

Modulation steps. The modulation procedure is a series of analyses and adjustments. The analysis step generates information such that the adjustment step can verify further changes in a given location are necessary. The adjustment step modifies the location in each of its dimensions separately in order to reduce its conspicuity (or increase it depending on the classification). These steps are done multiple times from coarse to fine levels of the image pyramid by using the built-in mip-map capability of the graphics unit. This allows changes affecting a large region to occur early in the process, while later steps progressively refine the result. An implicit benefit of starting with coarse resolutions is that modulation of lower frequencies introduces less noticeable artifacts. Each analysis and adjustment carried out in a fragment shader that executes in a render pass on the current image. The total number of passes necessary for modulation can be expressed as: $2 + 6 * n$ where n is the number of pyramid passes. Two passes are necessary to convert the image to and from CIEL*a*b* space and six passes are necessary for the adjustments in the three considered dimensions and their respective analyses. The analysis step computes the conspicuity values of the input image as described in the section 3. Figure 3 shows a flowchart with a detailed description of all the necessary steps.

Compute the modulation values to be applied. To know the modulation value to be applied to the current location (such as making it more blue or less yellow), we first verify that the location's absolute conspicuity value exceeds the given threshold. The conspicuity value is provided by the analysis step. If it does not exceed it, then no modulation is necessary and we leave the feature value unmodified. Otherwise, we compute the difference between the threshold and the current conspicuity value. This difference is then used as modulation value to be applied to the current feature. At this point, it is important to remember the roles of conspicuity and feature. A conspicuity is the difference between a location and its surroundings. We cannot modify a conspicuity directly, instead, it is modified indirectly by changing the feature values of the location.

Let m_k be the modulation to be applied to the location, c_k the conspicuity of the location and t_k the threshold of the conspicuity where k is the given dimension.

$$m_k = \begin{cases} 0 & c_k < t_k \\ c_k - t_k & \text{otherwise} \end{cases} \tag{4}$$

Before applying this modulation, a few checks must be performed in order to avoid unpleasing visual artifacts. For example, in the chromatic channels, we must also take care that the modulation should not flip the hue completely, i.e. blue never becomes yellow, and red becomes green (or vice versa). This is done by preventing a flip on the sign of the feature value, a positive value (e.g. red) cannot become negative (e.g. green).

Coherence. The modulation process seeks to reduce the amount of changes in the original image. A naïve implementation only considers the appropriate values for each location without regard to the global coherence of the image. As a result, noise artifacts can occur, typically chromatic, on the final image as illustrated in Figure 4. Such artifacts happen when two spatially close locations are matched to different modulation values, when the conspicuity of a location is increased (focus) and the original chromatic values are close to zero. To resolve this problem, we compute the average between the modulation computed at the previous pyramid level and the current level. A side effect of this filtering is that

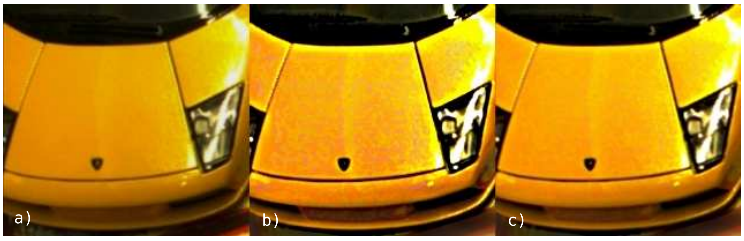

Fig. 4. Spatial coherence. This figures illustrates the problem arising from the emphasis of contrast in the focus area. A tungsten light on the metallic surface of this car model caused this particular artifacts. Notice the red or green dust in the middle image. a) Original image. b) Image affected by naïve conspicuity enhancement. c) Image after applying our spatial coherence technique.

the strength of the modulation is smoothed, leading to more visually pleasing results.

As stated before, the computation of the normalization weights is amortized over several frames, depending on the current framerate. If the amortization period is too long, the changes on normalization weights may be drastic. This can introduce temporal discontinuity artifacts between two adjacent frames. We therefore compute the weight and thresholds using a sliding average using a history of a few frames.

Once our modulation value has been computed and all checks necessary to ensure that no drastic changes occur (either spatially or temporally), we can apply this value to the location. To decrease the conspicuity of the context we merely subtract the change value we computed from the given location's feature. This has the effect of reducing the distance between the current conspicuity value and the threshold. To increase the conspicuity of a location, instead of subtracting, we simply add the change value we computed to the given location's feature. This has the effect of increasing the distance of the current conspicuity and the threshold.

Let f_k be the feature value of the location and $f\prime_k$ the modulated feature value and m_k the modulation to be applied where k is the given dimension.

$$f\prime_k = \begin{cases} f_k - m_k & \text{if the location is marked as context} \\ f_k + m_k & \text{if the location is marked as focus} \end{cases} \tag{5}$$

5 Results and Applications

We now present several examples that can make use of our technique. First we contrast naïve modification with our technique. Then we show applications that involve image modifications to achieve mediated reality effects on either the focus or the context of a scene. The first is a classical search task where the system tries to direct the attention of the user towards an item in the field of view. The second application tries to direct attention to an object that is not the main actor in the current task. All of the images shown in this paper were computed on a 3.0 GHz Intel Dual Core CPU with an NVIDIA GTX280 graphics card. We used GLSL for our fragment shaders and framebuffer objects for texture handling. The video feed used for our examples was at a 640x480 resolution. The lowest framerate we experienced was 15fps. Computing the modulation of an image on a single pyramid level was achieved in 1.023ms, computing it with 7 render passes was done in 36.47ms.

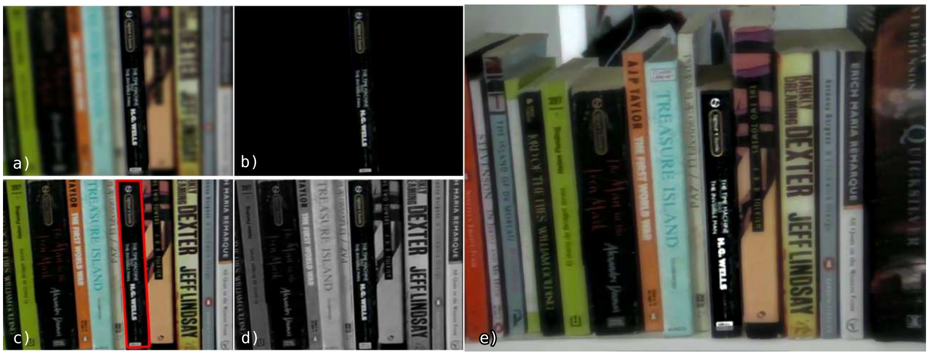

Fig. 5. Comparison between traditional techniques and our work. a) by Gaussian blur with a kernel size of 4. b) by total suppression of the context. c) by augmentation with a border of 4 pixels wide and a red color. d) by de-saturation. e) our approach.

Comparison of adaptive saliency modification and naïve image modification. There exist multiple techniques for attention direction. One may, for example, point at the object of interest with an augmentation. One may also de-saturate the context, blur it, or plainly suppress it entirely. However, the effectiveness of such attention direction techniques depends on the objects in the scene themselves. For example, color, shape and size of augmentations are heavily influenced by the objects in the field of view. An arrow pointing at an object will be only as effective as it is contrasting with its background, and text labels only be effective when displayed on adequate surfaces. In the same sense, traditional pixel-wise attention direction techniques rely on the assumption that the focus object has the necessary properties to stand out from the modified context. A de-saturation technique will be ineffective if the focus object has little saturation itself. Moreover, properties such as the strength of the modification are typically assigned a priori, for example, the kernel sized used for a Gaussian blur. This becomes critical in a mixed or mediated reality scenario where we do not have control over which objects are visible in the scene as the user is allowed free camera movements. Figure 5 compares different attention direction techniques side by side. The most effective of these techniques is image b) where the context is entirely suppressed. The least effective is d) due to the lack of saturation of the focus itself.

Reminding the user of dangerous objects in the proximity. The application we present now is that of giving maintenance to a car engine. The engine presents surfaces that are dangerous to the touch, such as being too hot, and the user should maintain a constant awareness of them. These objects, however, are not the main interest of the user. The user is engaged on a task which does not require the direction of the system. Our technique is well suited for this task by constantly reminding the user of the location of the dangerous surfaces (focus) while minimizing the obstruction of the main working area (context). Figure 6 shows an example of our technique before and after modulation. Bailey et al. [2] suggest that modulation does not need to be constantly applied. Instead, modulating the image during 1 second is already capable of directing the users gaze towards the focus area. This modulation can then be repeated every few seconds to keep attracting the gaze of the user towards the dangerous surfaces.

Finding objects task. Finding a particular object from a collection of similar objects is a common task presented in mixed reality. In this example we present a shelf where multiple books are visible to the user. The problem is to find a specific book in this shelf. Figure 5 illustrates this application which is reminiscent of that presented by Schwerdtfeger et al. [20]. The unmodulated image contains multiple items with colorful covers, all competing for the user's attention. Our system then subtly suggests the user where the target book is. Notice that the target book does not have any particular salient features such as a colorful book cover or large letters, yet the system is capable of accentuating its contrast and diminishing that of the context.

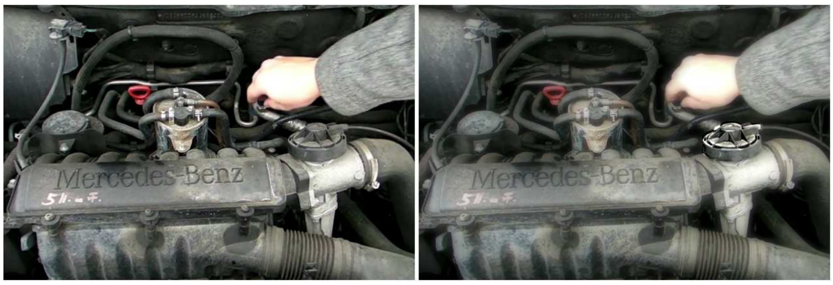

Fig. 6. One advantage of our technique is that it is not entirely detrimental to the Context region. This is helpful in situations where we want to highlight an object that is not the main actor on the current task of the user. For example, in this image we highlight an item that is not set for maintenance but may yet be dangerous and should be avoided.

6 Discussion and Conclusion

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 squared root of the sum of squared differences in the RGB space divided by the number of pixels in the image. The total average pixel difference across all images in this paper between modified and original versions is 1.34%.

We have presented a technique for the modulation of visual saliency for attention direction, specifically designed for interactive mediated reality application. Image saliency is analyzed in real time, and modulations are made in such a way that a desired distance of focus and context regions is achieved with only minimal changes. This technique can be seen as a way of reducing the contrast of the context area and of increasing it in the focus area. This contrast manipulation takes place on the lightness and color opponents dimensions. We have shown a number of application examples that indicate the usefulness of this approach.

However, it should be noted that the use of the technique presented in this article may not always be warranted or even possible. For some applications, the perception of the context may be unimportant. In other cases, there may be moving or blinking objects in the context, which may not be sufficiently suppressed with the presented pixel-wise techniques. Moreover, the modulations we have shown are incapable of directing the attention to an object that is not present in the image, nor can they show the direction on which this object may be found. In those cases, guiding attention direction through direct augmented overlays, such as described by Schwerdtfeger et al. [20] may be a more viable alternative.

We are considering multiple directions of work. Specifically, we are planning on incorporating additional saliency models. This would allow us to possibly modify the strength of our modulation depending on other conspicuity dimensions, for example texture variation as demonstrated by Su et al. [23]. An important future work direction is the validation of our research via user studies. This work is based on our experiences and preliminary user tests with an eye tracker [15]. In our previous work we found that modulation of bottom up stimuli can effectively direct the attention of users. The initial applications to mixed reality look promising on an informal level, but clearly a quantitative analysis is needed to fully understand the involved phenomena. To this aim, we are currently setting up a user study performed with an eye tracker to investigate which image modifications have a particular effect. In particular, future studies will evaluate not only the effectiveness of attention direction towards the focus, but also the circumstances of preservation of the contextual information.

Acknowledgments. This work was partially funded through the HYDROSYS project (EC-FP7 grant 224416).

References

1. Achanta, R., et al.: Salient Region Detection and Segmentation. In: ICVS 2008, pp. 66-75

2. Bailey, R., et al.: Subtle Gaze Direction. In: ACM TOG, Vol. 28, No. 4, 2009, pp. 1-14
3. Cerf, M., et al.: Predicting human gaze using low-level saliency combined with face detection. In: J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, NIPS. MIT Press, 2007
4. Engel, S., et al.: Colour tuning in human visual cortex measured with functional magnetic resonance imaging. In: Nature 388, 6637, pp. 68-71
5. Hering, E.: Outlines of a Theory of the Light Sense. In: Harvard University Press, Cambridge, Mass. 1964
6. Hurvich, L. M., Jameson, D.: An opponent-process theory of color vision. In: Psychological Review 64 (6, Part I), 1957, pp. 384-404
7. Itti, L., et al.: A model of saliency-based visual attention for rapid scene analysis. In: IEEE TPAMI 20, 11, 1998, pp. 1254-1259
8. Judd, T., et al.: Learning to predict where people look. In: ICCV 2009
9. Koch, C., et al.: Shifts in selective visual attention. In: Human Neurobiology 4, 1985, pp. 219-227
10. Kienzle, F., et al.: A nonparametric approach to bottom-up visual saliency. In: B. Scholkopf, J. C. Platt, and T. Hoffman, editors, NIPS, MIT Press, 2006, pp. 689-696
11. Kim, Y., Varshney, A.: Persuading Visual Attention through Geometry. In: IEEE TVCG, Vol. 14, No. 4, July 2008, pp 772-782
12. Kim, Y., Varshney, A.: Saliency-guided Enhancement for Volume Visualization. In: IEEE TVCG, vol.12, no.5, 2006, pp.925-932
13. Kosara, R., et al.: An interaction view on information visualization. In: EUROGRAPHICS, 2003, pp. 123-137
14. Lee, S., et al.: Real-Time Tracking of Visually Attended Objects in Interactive Virtual Environments. In: ACM VRST, 2007, pp. 29-38
15. Mendez, E. et al.: Experiences on Attention Direction through Manipulation of Salient Features. In: IEEE VR 2010 PIVE Workshop, pp. 4-9
16. Niebur, E.: Saliency Map. In: scholarpedia, Web address: http://www.scholarpedia.org/article/Saliencymap, Date: January 10, 2010
17. Ouerhani, N., et al.: Empirical validation of the saliency-based model of visual attention. In: Electronic Letters on Computer Vision and Image Analysis 3, 1, pp. 13-24
18. Rosenholtz, R.: A simple saliency model predicts a number of motion popout phenomena. In: Vision Research 39, 19, 1999, pp. 3157-3163
19. Santella, A., Decarlo, D.: Visual interest and NPR: an evaluation and manifesto. In: Proceedings of NPAR, 2004, pp. 71-150
20. Schwerdtfeger, B., Klinker, G.: Supporting Order Picking with Augmented Reality. In: IEEE and ACM ISMAR, 2008, pp. 91-94
21. Sangwie S.J., et al.: The colour image processing handbook, first edition: Published by Chapman and Hall, 1998
22. Spillman, L.: Visual Perception: The Neurophysiological Foundations. In: Academic Press, 1990
23. Su, S., et al.: De-Emphasis of Distracting Image Regions Using Texture Power Maps. In: Workshop on Texture Analysis and Synthesis at ICCV, 2005
24. Torralba, et al.: Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. In: Psychological Review. Vol 113(4), 2006, pp. 766-786
25. Treisman, A. M., Gelade, G.: A feature-integration theory of attention. In: Cognitive Psychology 12, 1980, pp. 97-136