Effect of tone mapping operators on visual attention deployment

Manish Narwaria, Matthieu Perreira da Silva, Patrick Le Callet, Romuald
Pépion

HAL Id: hal-00724931
https://hal.science/hal-00724931

Submitted on 23 Aug 2012

Effect of tone mapping operators on visual attention deployment

Manish Narwaria, Matthieu Perreira Da Silva[*], Patrick Le Callet, Romuald Pepion

LUNAM University - IRCCyN CNRS UMR 6597, 44306, Nantes, France

ABSTRACT

High Dynamic Range (HDR) images/videos require the use of a tone mapping operator (TMO) when visualized on Low Dynamic Range (LDR) displays. From an artistic intention point of view, TMOs are not necessarily transparent and might induce different behavior to view the content. In this paper, we investigate and quantify how TMOs modify visual attention (VA). To that end both objective and subjective tests in the form of eye-tracking experiments have been conducted on several still image content that have been processed by 11 different TMOs. Our studies confirm that TMOs can indeed modify human attention and fixation behavior significantly. Therefore our studies suggest that VA needs consideration for evaluating the overall perceptual impact of TMOs on HDR content. Since the existing studies so far have only considered the *quality* or *aesthetic appeal* angle, this study brings in a new perspective regarding the importance of VA in HDR content processing for visualization on LDR displays.

Keywords: High Dynamic Range (HDR), visual attention, eye-tracking experiment

1. INTRODUCTION

High Dynamic Range imaging (HDRI) has been steadily gaining popularity in both academia and industry. The reason being that HDR contents are visually more appealing and realistic as they can represent the dynamic range of the visual stimuli present in the real world. In other words, HDR faithfully depicts the dynamic range of the real world luminance (typically varying from 10^{-1} cd/m^2 to 10^5 cd/m^2) by storing them as floating point values. As a result, an HDR image can capture very high contrasts which in turn enables it to incorporate maximum details that the human eye can discern. Not surprisingly HDRI can arguably be considered a major step in display technologies since the transition from black and white to color displays[1,23]. However, the cost of HDR display technologies is currently quite high and yet to reach consumer levels. In such scenario, the only alternative is to display HDR contents directly on commonly available devices such as CRT, LCD monitors, printers etc. which have a significantly low dynamic range (LDR). It follows that these cannot provide the necessary luminance range (usually their range lies between 1 to 300 cd/m^2) for a true HDR experience. Additionally, their contrast ratio is not good enough for displaying HDR contents. For example, even a good In-Plane Switching (IPS) LCD panel can achieve a contrast ratio of only about 1000:1 while the required contrast ratio of typical HDR scenes can be more than 10^6:1. Therefore an important issue in HDRI is to reduce the dynamic range of the HDR content. This problem has been commonly addressed by employing the so called tone mapping operators (TMOs). Tone mapping refers to the reduction in dynamic range so as to properly display the HDR content onto LDR devices. To that end, several TMOs[2-12] have been devised in literature. Some are simple and based on operations such as linear scaling and clipping while the more sophisticated ones exploit several properties of the Human Visual System (HVS) with the aim of preserving the details. But more often than not, TMOs lead to information loss which can reduce the perceptual quality of the tone mapped contents. This is expected since dynamic range compression invariably tends to destroy important details and textures and can introduce additional artifacts related to changes in contrast and brightness.

TMOs can be broadly classified into 2 categories namely local operators and global operators. As the name implies, local operators employ a spatially varying mapping which depends on the local image content. As opposed to this, global operators use the same mapping function for the whole image. Chiu et al. introduced[2] one of the first local TMOs by employing a local intensity function based on a low-pass filter to scale the local pixel values. The method[3] proposed by Fattal et al. is based on compressing the magnitudes of large gradients and solves the Poisson equation on the modified gradient field to obtain tone mapped images. Durand et al. presented[4] a TMO based on the assumption that an HDR image can be decomposed into a base image and a detail image. The contrast of the base layer is reduced using an edge-

---

[*] Corresponding author.
  Email: matthieu.perreiradasilva@univ-nantes.fr

preserving filter (known as the bilateral filter). The tone mapped image is obtained as a result of multiplication of the contrast reduced base layer with the detail image. Drago et al. adopted[5] logarithmic compression of the luminance values for dynamic range reduction in HDR images. They use adaptively varying logarithmic bases in order to preserve local details and contrast. The TMO proposed[6] by Ashikimin first estimates the local adaptation luminance at each point which is then compressed using a simple mapping function. In the second stage, the details lost in the first stage are re-introduced to obtain the final tone mapped image. Reinhard et al. applied[7] the dodging and burning technique (traditionally used in photography) for dynamic range compression. A TMO based on a perceptual framework for contrast processing in HDR images was introduced[8] by Mantiuk et al. This operator involves the transformation of an image from luminance to a pyramid of low-pass contrast images and then to the visual response space. It was claimed that in this framework, dynamic range reduction can be achieved by a simple scaling of the input. Mantiuk et al. also proposed[9] another TMO by formulating tone mapping as an optimization problem: one that minimizes the visible contrast distortions between a human visual system and the display. Another TMO known as iCAM06[10] has also been developed. It is based on the sophisticated image color appearance model (iCAM) and incorporates the spatial processing models in the HVS for contrast enhancement, photoreceptor light adaptation functions that enhance local details in highlights and shadows. With regards to global TMOs, the simplest one is the linear operation in which the maximum input luminance is mapped to the maximum output value (the maximum luminance mapping) or the average luminance mapping (i.e. mapping average input luminance to the average output value). Another global TMO is the one proposed[11] by Ward which focuses on the preservation of perceived contrast. In this method, the scaling factor is derived from a psychophysical contrast sensitivity model. Tumblin et al. have reported[12] a TMO based on the assumption that a real-world observer should be the same as a display observer. These are some of the existing TMOs and the list is by no means exhaustive. The interested reader is referred to survey papers[17] for a more complete and detailed study of TMOs.

The reader may have noticed that local TMOs seem to have received more attention than the global ones. This is partly due to the fact that as a result of their design local TMOs perform well in preserving the local details (but are less effective in reproducing the overall brightness and contrast). On the other hand, although global TMOs preserve the overall contrast they usually lead to loss of local details. But as an important advantage, global operators are generally computationally more efficient than the local ones. So while local and global TMOs have their own advantages and disadvantages, there does seem to be a general consensus: TMOs inevitably degrade the visual experience of the HDR image. For that reason, there have been subjective studies reported in literature all of which aim to analyze the impact of TMOs with regards to the perceptual visual quality.

2. BACKGROUND AND MOTIVATION

We first briefly describe some of the existing studies related to subjective evaluation of TMOs. The psychophysical experiments carried out[13] by Drago et al. aimed to evaluate six TMOs with regard to similarity and preference. Three perceptual attributes namely apparent image contrast, apparent level of detail (visibility of scene features), and apparent naturalness (the degree to which the image resembled a realistic scene) were investigated. It was found that naturalness and details are important attributes for perceptual evaluation of TMOs. The study[14] by Kuang et al. performed a series of three experiments. The first one aimed to test the performance of TMOs with regard to image preference. For this experiment, 12 HDR images were tone mapped using six different TMOs and evaluation was done using the paired comparison methodology. The second experiment dealt with the criteria (or attributes) observers used to scale image preference. The attributes that were investigated included highlight details, shadow details, overall contrast, sharpness, colorfulness, and the appearance of artifacts. The subsequent regression analysis showed that the rating scale of a single image appearance attribute is often capable of predicting the overall preference. The third experiment was designed to evaluate HDR rendering algorithms for their perceptual accuracy of reproducing the appearance of real-world scenes. To that end, a direct comparison between three HDR real-world scenes and their corresponding rendered images displayed on a low dynamic-range LCD monitor was employed. Yoshida et al. conducted[15] psychophysical experiments which involved the comparison between two real-world scenes and their corresponding tone mapped images (obtained by applying 7 different TMOs to the HDR images of those scenes). Similar to other studies, this one was also aimed at assessing the differences in how tone mapped images are perceived by human observers and was based on four attributes: image naturalness, overall contrast, overall brightness, and detail reproduction in dark and bright image regions. In the experiments conducted[16] by Ledda et al., the subjects were presented three images at a time: the reference
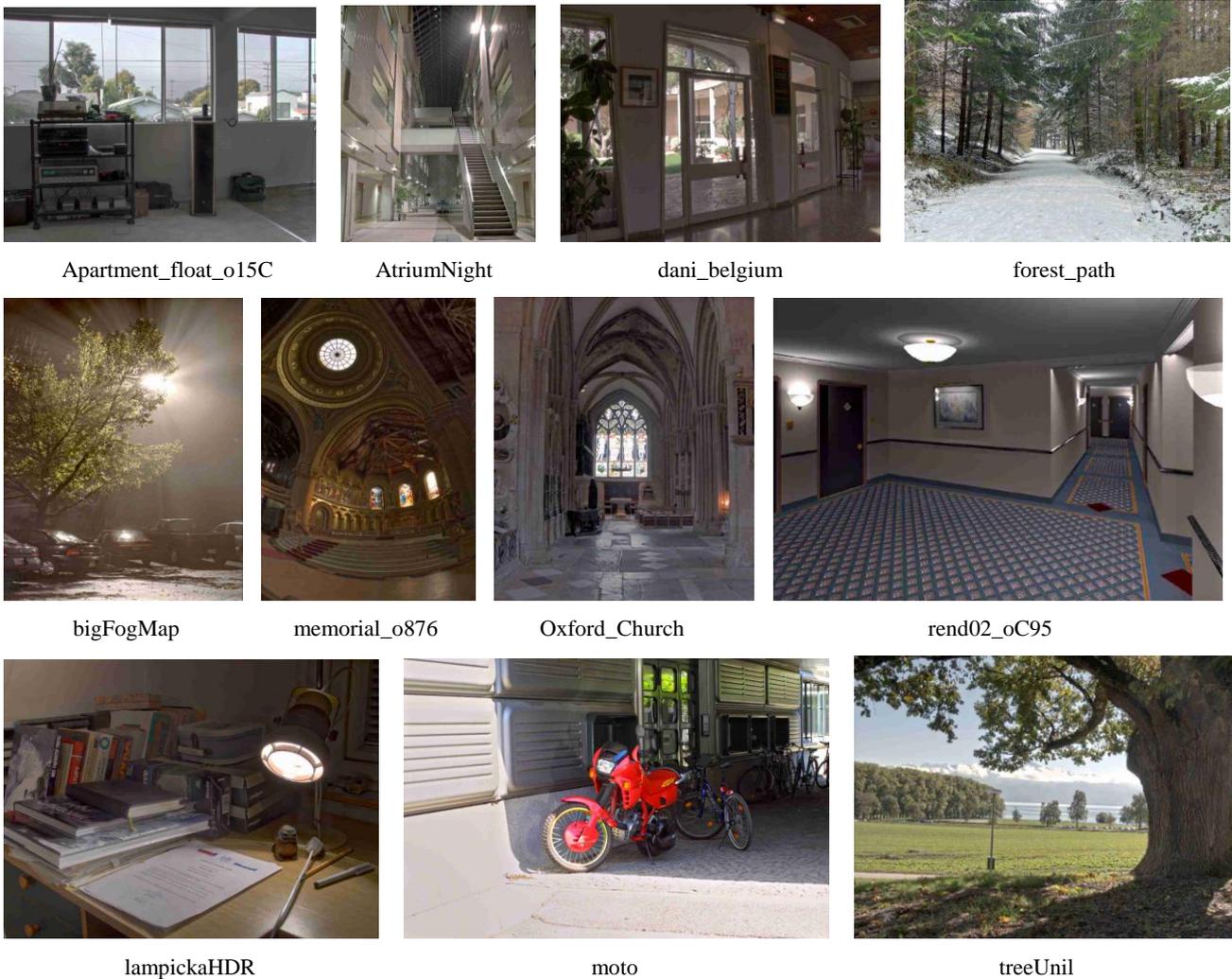
Figure 1. Images used for the eye-tracking experiments.

HDR image displayed on an HDR display and two tone mapped images viewed on LCD monitors. They had to choose the image closest to the reference. Because an HDR display was used, factors such as controlling screen resolution, dimensions, colorimetry, viewing distance and ambient lighting could be controlled. This is in contrast to using a real-world scene as a reference which might introduce uncontrolled variables. The authors have also reported the statistical analysis of the subjective data with respect to the overall quality and to the reproduction of features and details. Different from the mentioned studies, Cadik et al. adopted[17] both a direct rating (with reference) comparison of the tone mapped images to the real scenes, and a subjective ranking of tone mapped images without a real reference. They further derived an overall image quality estimate by defining a relationship (based on multivariate linear regression) between the attributes: reproduction of brightness, color, contrast, detail and visibility of artifacts. The analysis further revealed that contrast, color and artifacts are the major contributing factors in the overall judgment of the perceptual quality. However, it was also argued that the effect of attributes such as brightness is indirectly incorporated through other attributes. Another conclusion from this study was that there was agreement between the ranking (of two tone mapped images) and rating (with respect to a real scene) experiments. In contrast to this last observation, Ashikimin et al. found[18] that there were significant differences in subjective opinions depending on whether a real scene is used as a reference or not. The experiments conducted by Akyuz et al. focused only on contrast perception in tone mapped images. For this purpose they used artificially created HDR images in order to control other complex factors which are present in a real HDR image.

Table 1. Tone mapping operators evaluated in the subjective study

TMO (Abbreviation)	Description	Reference	Global/Local (G/L)
Ashikimin (ash)	A tone mapping algorithm for high contrast images	[6]	L
Chiu (chi)	Spatially nonuniform scaling functions for high contrast images	[2]	L
Drago (dra)	Adaptive logarithmic mapping for displaying high contrast scenes	[5]	G
Durand (dur)	Fast bilateral filtering for the display of HDR images	[4]	L
Fattal (fat)	Gradient domain high dynamic range compression	[3]	L
iCAM06 (ica)	A refined image appearance model for HDR image rendering	[10]	L
Linear (lin)	Simple linear scaling	-	G
Mantiuk (man)	Display Adaptive Tone Mapping	[9]	L
Reinhard (rei)	Photographic tone reproduction for digital images	[7]	L
Tumblin (tum)	Two methods for display of high contrast images	[12]	G
Wardhist (war)	A contrast-based scalefactor for luminance display	[11]	G

This is a significant divergence from the other mentioned studies which either used a real scene or an HDR image for comparing TMOs. As mentioned above, the current effort in subjective evaluation has been mainly directed towards assessing the impact of TMOs from quality and aesthetic appeal point of view. From these we may be able to study and analyze people's preference regarding visual appeal of the tone mapped content. However, we believe that quality is just one of the several aspects that need to be considered to make conclusions on how TMOs affect the overall quality of experience (QoE). One such issue is that of visual attention (VA) which has been well recognized as a crucial aspect in perceptual visual signal processing. It is well known that human eyes tend to focus more on certain areas in an image/video than others. Stated differently, some regions attract more eye attention and these are termed as salient regions. VA is therefore the ability of the HVS to find and focus on relevant information quickly and efficiently[19]. This has several applications since the more important signal information can be extracted and processed accordingly. For example in image/video coding, the visually salient parts can be assigned more bits in order to achieve higher efficiency and better visual quality. VA has also been employed in image quality assessment[25] where VA maps can be used as weighting factor for the extracted features. Further from an artistic viewpoint, TMOs could possibly change the way a scene is perceived by human eyes. This may lead to changes in the feelings and emotions conveyed by the image. Thus the intention of the artist /content author may not be "represented" correctly to the viewer. For instance, intricate details (like very fine texture) in some part of an image which attract viewer attention might be lost due to tone mapping and the photographer's intention of producing a compelling picture is jeopardized. . It is thus clear that VA plays an important in human perception and therefore the impact of TMOs should also be analyzed from this view point as well. A survey of literature however reveals that this issue has not been addressed in any of the existing works. This paper therefore seeks to address this by way of eye-tracking experiments on several still images processed by 11 TMOs.

3. DETAILS OF EYE-TRACKING EXPERIMENTS

3.1 Test scenes

It is necessary that the results and conclusions out of a subjective study should be independent of the image contents. In other words, the resultant analysis should be general and not be influenced by the specific image/video contents used. Generally, test images should include a large variety of image content, such as landscape, architecture, human portraiture, and images with differing light-source sizes. Accordingly we used 11 different HDR images (shown in Fig. 1) which are representative of a wide range of content such as indoor and outdoor scenes with varying illumination conditions. These include outdoor only scenes ('treeUnil', 'bigforgmap' and 'forest_path'), both indoor and
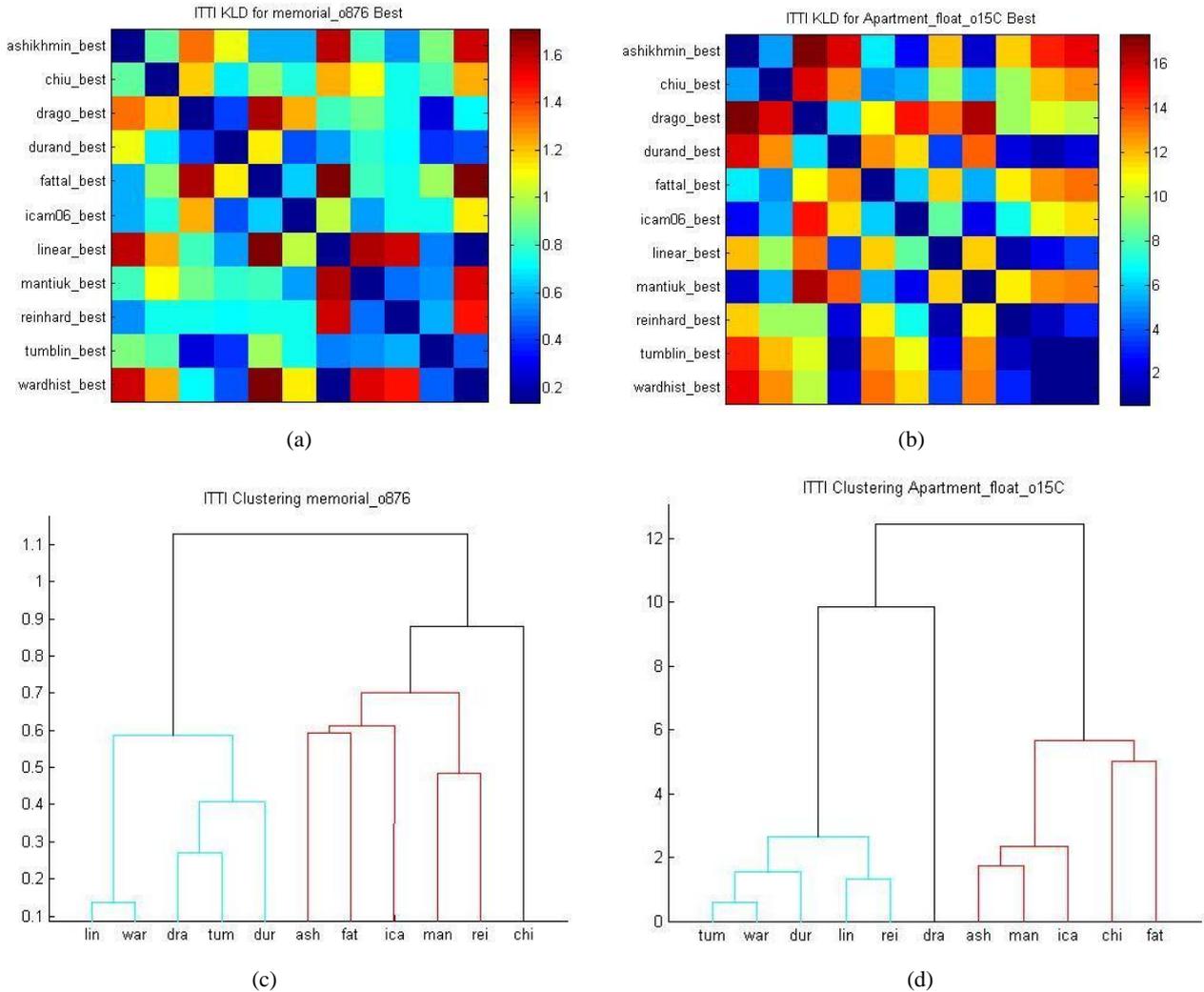
Figure 2. Objective analysis to reduce the number of images for the eye-tracking experiments. (a) KLD based distance matrix for 'memorial_o876' image, (b) KLD based distance matrix for 'Apartment_float_o15C' image, Cluster tree diagram generated for distance the matrix in (a) and (d) Cluster tree diagram generated for the distance matrix in (b)

outdoor scenes ('apartment', 'dani_belgium' and 'moto'), indoor only scenes ('lamp', 'church', 'memorial' and 'rend02').

3.2 Tone Mapping Operators

In order to arrive at reliable conclusions, it is necessary to use a large number of TMOs. Using only a few TMOs can affect the analysis and results of the subjective study. Therefore, our study utilized images processed from 11 different TMOs. Some of these TMOs are classical ones while others are more recent. In addition, a few are global operators while others belong to the local operator category. The 11 TMOs used in this paper are given in Table 1.

We used the HDR toolbox[23] which provides Matlab implementations of several TMOs for generating the tone mapped images from all but Mantiuk TMO. For Mantiuk TMO, we used the implementation which is freely available online[28]. An important issue to be addressed here is that each TMO has one or more parameters which are normally user specified. Unfortunately, for many TMOs, the use of the default parameter set yielded tone mapped images with 'poor' visual quality. The issue of 'best' parameter selection is further complicated since it can be image content specific. That is, a set of parameters which is suitable for one image content may not be optimal for the other. Note that in our

ITTI average Clustering Mean of the KLD on all images

ITTI complete Clustering Mean of the KLD on all images

(a)

(b)

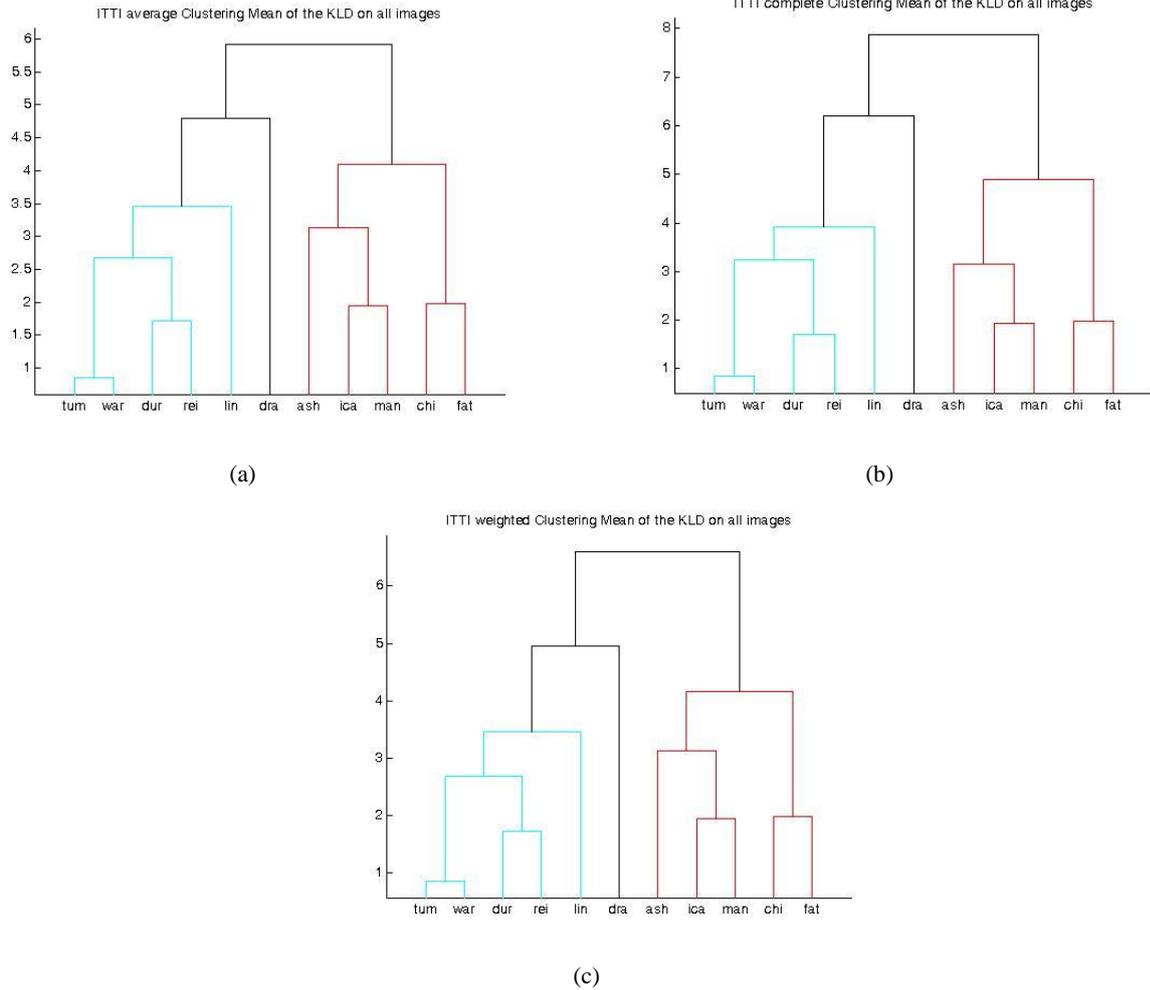ITTI weighted Clustering Mean of the KLD on all images

(c)

Figure 3. Objective analysis to reduce the number of images for the eye-tracking experiments. (a) Cluster tree based on average clustering, (b) Cluster tree based on complete clustering, (c) Cluster tree based on weighted clustering. These cluster trees are for the data from Itti's model. Refer to Table 1 for abbreviations used in the figure.

discussion 'best' refers to the parameter set which yields 'good' quality tone mapped images. Given the subjective nature of defining 'good' quality, we conducted a small pilot study to choose the 'best' parameters for each TMO. To that end, we experimented with a large number of combinations of different parameter values before narrowing down to a set which yielded images with 'good' quality. The list of 'best' TMO parameters for each HDR image that we chose based on our study can be made available to the interested reader by contacting the corresponding author.

3.3 Statistical analysis for test material preparation

As mentioned, 11 HDR image content and 11 TMOs were chosen. Consequently there will be totally 121 tone mapped images for subjective eye-tracking experiments. However, since there are 11 image contents, it could lead to the situation where the same image (though processed by a different TMOs) is shown many times to the same subject. As a consequence subjects might start to 'memorize' the content unconsciously. This can lead to a situation where previous test scene could have an impact on the currently viewed scene which in turn might lead to strongly biased visual attention results. Clearly this is undesirable and the impact of memory must be reduced if not completely eliminated. However, this would mean that we must recruit a very large subject panel which can obviously create issues related to
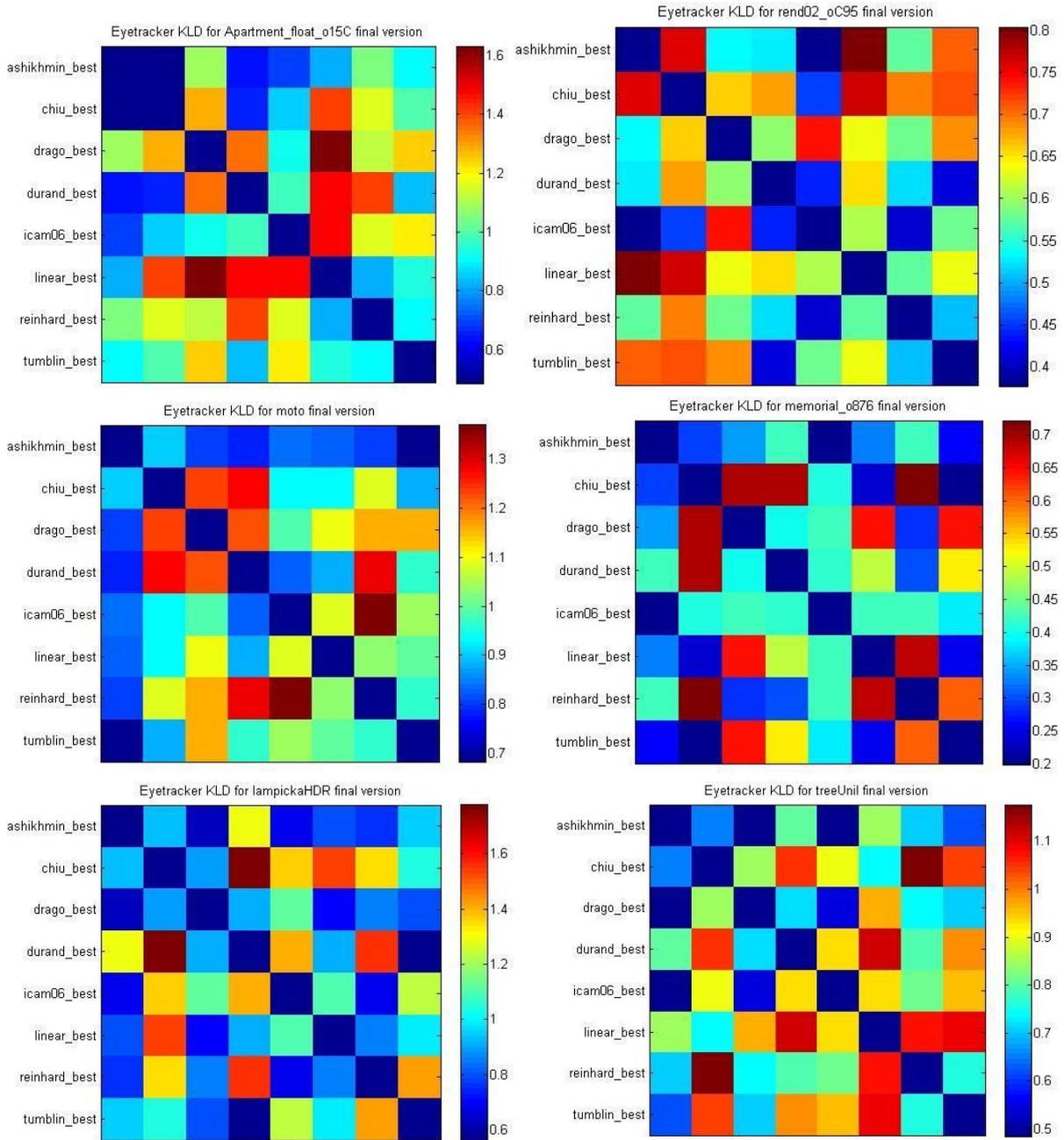
Figure 4. KLD based distance matrices for 6 HDR images. Each matrix denotes the distance between the VA maps (obtained from eye-tracking experiments) in terms of KLD.

logistics, more time and effort, all of which will result in a rather large and cumbersome experimental setup. Suppose we have i image contents (i.e. i HDR images) and to avoid the memory effect we assume that we should not show more than twice the same image content to each observer/subject. Further assume that each image needs to be viewed at least n times (obviously higher the value of n higher is the accuracy and reliability). Then in such a scenario we will need at least $i/2 * n$ subjects. Clearly this number increases with both i and n. Therefore, to reduce the impact of memory and the scale of the eye-tracking experiments, we adopted two measures: (a) we completely randomized the order of presentation of images to each observer, (b) statistical analysis was carried out to reduce the number of images used.

Pertaining to the latter measure, a clustering based analysis was done to eliminate TMOs which may have similar 'impact' in terms of VA. For this, we first obtained the VA maps objectively using the well known Itti's model[26]. Next, we used a Kullback Leibler divergence (KLD) based distance measure to obtain the distance matrices between the VA maps of images processed by different TMOs. KLD is measure of dissimilarity between two probability distributions and is defined as

$$D_{KL}(P \parallel Q) = \sum_{x \in X} P(x) \, \log \frac{P(x)}{Q(x)}$$

Here P and Q are the 2D probability density functions of VA maps and x represents the spatial coordinates of the pixel. When the two probability densities are strictly equal, the KLD value is zero. Since this KLD is an information based theoretical dissimilarity measure, it is not symmetric. We therefore used the mean of the KLD between the two distributions. That is

$$KLD(P \parallel Q) = \frac{D_{KL}(P \parallel Q) + D_{KL}(Q \parallel P)}{2}$$

The KLD defined above is obviously symmetric and thus represents 'distance' in a better sense. These KLD values were then used to derive distance matrices. Note that there will be one distance matrix for each of the 11 HDR images chosen. To illustrate this, example matrices for the 'memorial_o876' and 'Apartment_float_o15C' images are shown in Fig. 2 (a) and (b) respectively. Each entry in these matrices denotes the KLD between the VA maps of the corresponding TMOs. Note that the order of TMOs for the columns is the same as that of rows. So for example, the entry (3,5) (third row and fifth column) denotes the KLD between the VA maps of images processed by Drago and Fattal TMOs. Note that the notation 'TMO_best' in these matrices simply refers to the fact that the 'best' parameters in terms of visual quality (as elaborated in section 3.2) were used for each TMO to generate the tone mapped images. Obviously, the distance matrices would be symmetric i.e. element (i,j) is the same as element (j,i). Next, hierarchical clustering[30] was employed to analyze the data from the distance matrices.

We used the Euclidean distance as the metric to form the clusters. The other factor based on which clusters are separated is the linkage criteria which determines the distance between the sets of observations as a function of the pair wise distances between observations. In this paper, we experimented with three commonly used[31] criteria: complete, single and weighted. Based on the cluster analysis, it was found that out of the 11 TMOs, 3 could be removed since their impact in terms of objective VA can be accounted for in the remaining TMOs. To exemplify this, we show the cluster tree diagram generated from the two matrices shown in Fig. 2 (a) and (b). As can be seen from the cluster diagram in Fig. 2 (c) , for instance the linear and Wardhist TMOs can be clustered together. We also computed the mean cluster tree diagram (i.e. averaged over all the 11 images) for each linkage criteria. These are shown in Fig. 3. Based on our analysis over all the 11 images, we finally removed three TMOs namely Mantiuk, Wardhist and Fattal. In order to make sure that the conclusions drawn were reliable, we repeated the whole analysis by replacing Itti' method by the one proposed by Perreira Da Silva et al.[27] and Bruce et al.[29]. We found that the conclusions and analysis were largely the same as with Itti's model. Therefore, the number of TMOs for the subjective eye-tracking experiment was reduced from 11 to 8. We note that this method of reducing the number of TMOs is not ad-hoc and based on a statistical analysis.

3.4 Participants

The 88 tone mapped images (11 images processed by 8 TMOs) were viewed by a total 48 subjects (28 male and 20 female). They were all between 19 and 51 years old. Prior to the test, subjects were screened for visual acuity by using a Monoyer optometric table and for normal color vision by using Ishihara's tables. All subjects had normal or corrected to normal visual acuity and normal color perception. All were inexperienced observers (not expert in image or video processing) and naive for the purpose of this study. The 88 tone mapped images were divided into 4 image groups each with 22 images. Similarly, the subjects were divided into 4 observer groups each with 12 observers. Images in each image group were viewed by observers who came from different observer groups (i.e. the observers for each image group did not overlap). As already mentioned, the order of visualization was also randomized so no observer sees the images in the same order. All these ensured that the impact of memory effect was minimized.
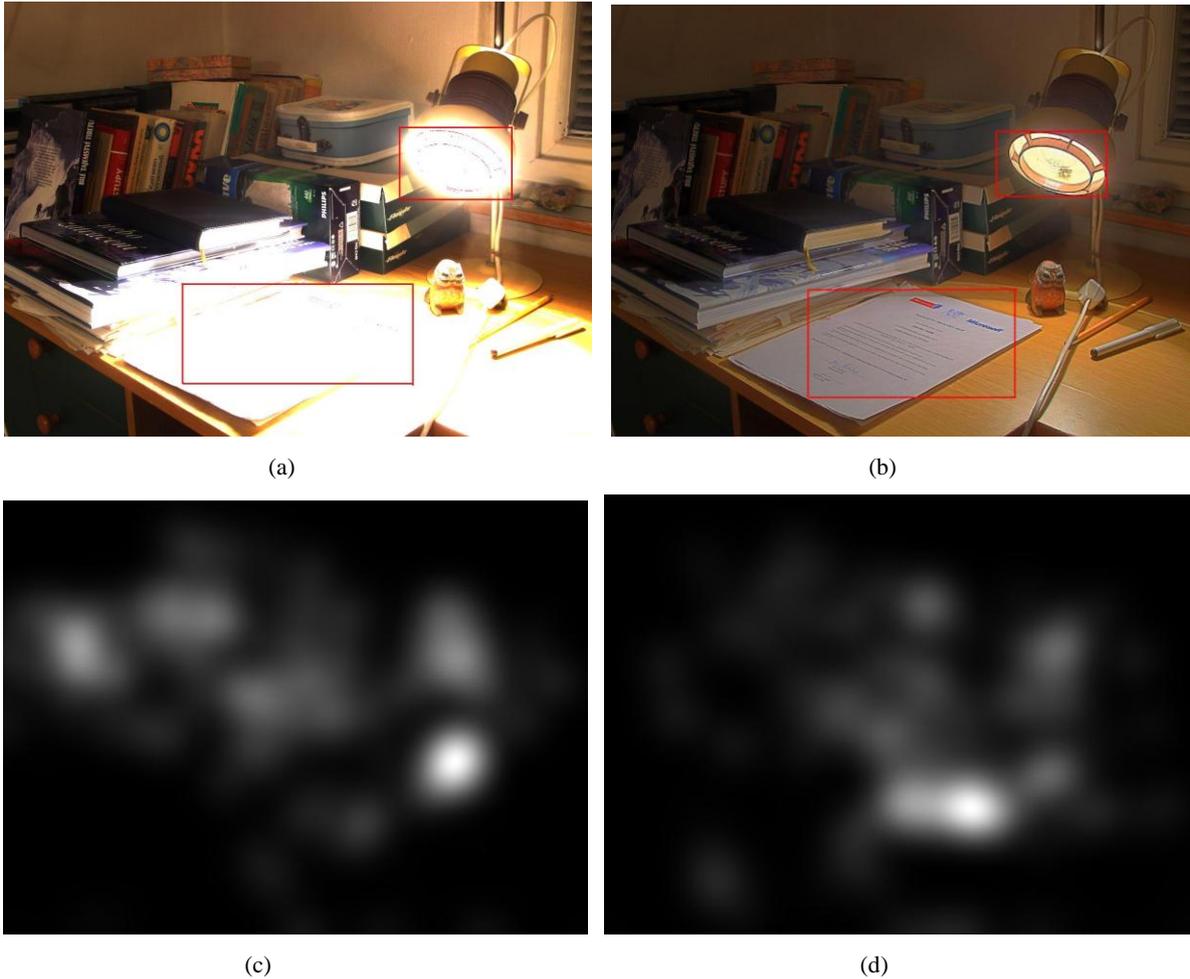
Figure 5. Illustration of the effect of global and local TMOs. (a) Image processed by Tumblin TMO (global), (b) Image processed by Ashikimin TMO (local), (c) VA map for (a) and (d) VA map for (b). The red boxes highlight two areas in the images where details are lost and preserved by global and local TMOs respectively.

3.5 Apparatus

Experiments were performed with a dual-Purkinje eye tracker, "the High Speed" from SensoMotoric Instrument (SMI) company. The eye tracker was mounted on a headrest that incorporates an infrared camera, an infrared semi transparent mirror and two infrared illumination sources. Before each trial, the subject's head was correctly positioned on a headrest so that their chin pressed on the chin-rest and their forehead lean against the head-strap. The heights of the chin-rest and head-strap system were adjusted so that the subject sat comfortable and their eye level was aligned with the center of the presentation display. The eye tracker is able to record the movement of both eyes.

Observers were seated in a standardized room conforming to the International Telecommunication Union Recommendation (ITU-R) BT500-11 recommendations[24]. The display used for this experiment was the TvLogic LVM401. Calibration was performed with an Eye One Pro luminance meter: gamma correction was set to 2.2 and white point to 6500K for a luminance maximum of 180 cd/m^2 Room illumination was calibrated as well, and set to 25 cd/m^2 behind the screen, hence around 15% of the perceived screen brightness. Finally, viewing distance was set to three times the height of the screen (active part), that is 150 cm. Each image was displayed for ten seconds and a gray screen appeared for three seconds between each image display. In between, recalibration was done when it was necessary for some observers. We computed the VA maps at 0.5 seconds, between 0.5 and 4 seconds and finally 4 seconds of viewing
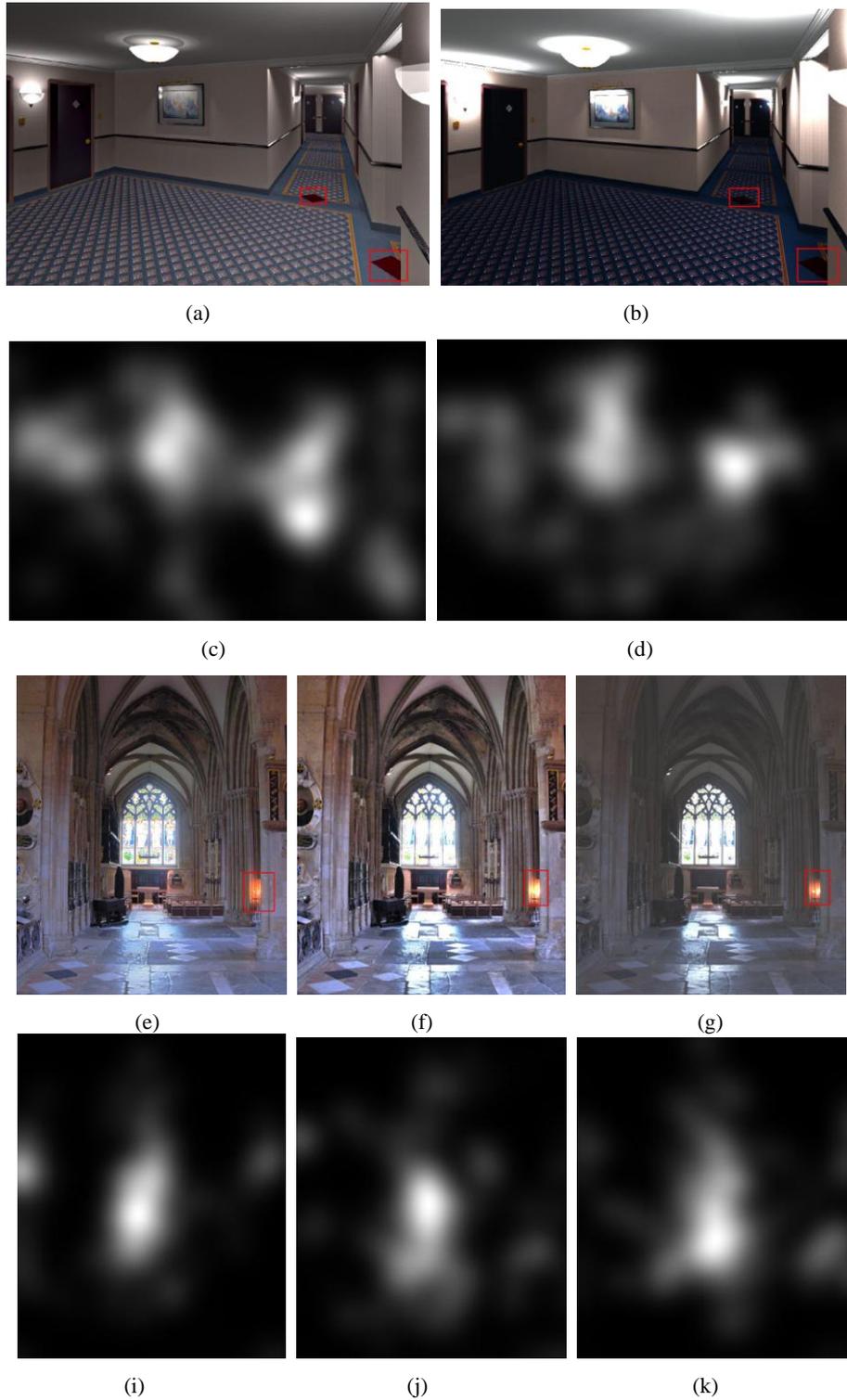
Figure 6. Effect of TMOs on VA. (a) 'rend02_oC95' image processed by iCAM06 TMO, (b) 'rend02_oC95' image processed by Drago TMO, (c) VA map for (a) and (d) VA map for (b), (e)-(g) 'Oxford_Church' image processed by Tumblin, iCAM06 and linear TMOs respectively, (i)-(k) VA maps for the images shown in (e)-(g) respectively. The red boxes highlight the area(s) in the images which become salient or non-salient depending on the overall impact of TMO.
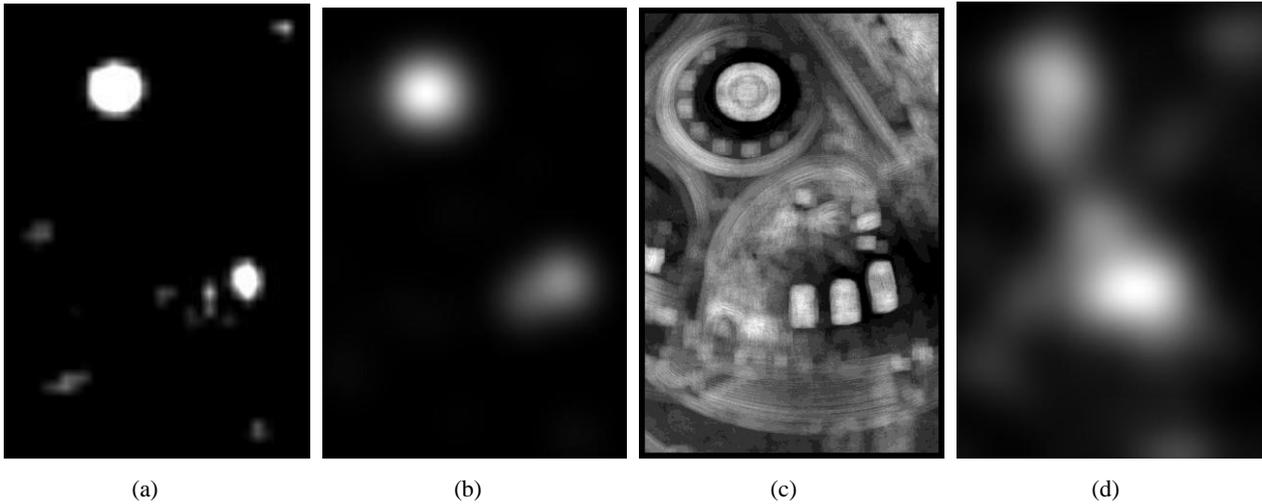
Figure 7. VA maps from different methods for the image 'memorial_o876' processed by Chiu TMO. (a) Itti et al. Model, (b) Perreira Da Silva et al. model, (c) Bruce et al. Model and (d) Eye-tracking model (subjective).

time as further elaborated in Section 4. Unless explicitly stated, all the VA maps from eye-tracking experiments shown in this paper were computed for 4 seconds of viewing time.

3.6 Human Priority Maps

The visual inspection of the visual field is studied through the eye movements. Analysis of the eye movement record was then carried out off-line after completion of the experiments. The raw eye data is segmented into saccades and fixations using iGaze software. Saccades are very rapid eye movements allowing the viewer to explore his visual field. Fixation is a residual movement of the eye when the eye is locked on a particular area of the visual field. The fixation occurs between two saccade periods. Visual fixation allows the viewer to lock the central part of the retina, the fovea, on a particular target. The fovea plays a critical role in sensing details since most of the visual sensory resources are concentrated on this central part. The start and end time of the fixation were extracted as well as its spatial coordinates. Human priority maps were then built by adding all fixations from all observers into a priority map and then blurring the resulting map with a two-dimensional Gaussian kernel whose size is proportional to viewing distance and eye-tracking accuracy. The rationale behind using the Gaussian kernel based filtering is two-fold: (a) observers do not gaze at a point of the visual field but rather an area having a surface close to the size of the fovea, (b) it is also used to reflect the limited accuracy of the eye tracking apparatus. There is one issue related to the use of fixation points. When we want VA maps for a very short viewing time (say 0.5 seconds), then there are not enough points to obtain statistically reliable VA maps. As detailed later in Section 4, our studies also used maps for the viewing duration of 0 to 0.5 seconds. Therefore, instead of using fixations, we used the gaze points to generate the saliency maps. It may also be pointed out that for larger viewing duration (4 seconds or more), the use of either fixation or gaze points yielded quite similar VA maps. In light of this, we used the gaze points to generate the maps in this paper.

4. ANALYSIS AND DISCUSSION

Similar to the objective analysis with the two computational VA models, we generated the distance matrices one for each of the 11 HDR image (processed by 8 TMOs since 3 TMOs were removed as explained in Section 3.3). These are shown in Fig. 4 (to save space matrices for only 6 HDR images are shown). The difference is that in this case the VA maps were obtained from the eye-tracking experiments (instead of computational models). As a result, these VA maps are 'ideal' and represent the actual human fixation behavior when viewing tone mapped images. It follows that the distance matrices (shown in Fig. 4) obtained from the real fixation maps also provide a more reliable data for drawing conclusions. We made several observations based on the subjective data obtained from the 48 subjects. These are described next.

The first observation is that for each image content, the TMOs modify the fixation behavior significantly. For instance, in the distance matrix for 'memorial_o876' image shown in Fig. 4, one can observe significant divergences between the fixation maps obtained for images processed by different TMOs. Of course the difference is more in some cases than others. But the consensual observation is that TMOs have a sufficiently perceivable (and different) impact on the VA behaviour. One can observe for instance from Fig. 5 (c) and (d) the VA maps for the two images shown in Fig. 5 (a) and (b).
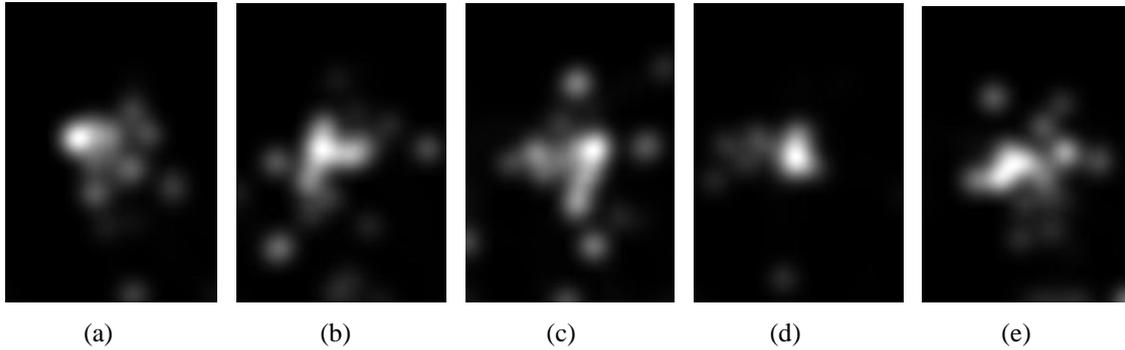

| (a) | (b) | (c) | (d) | (e) |

Figure 8. VA maps for the tone mapped versions of 'bigFogMap' image processed by, (a) Ashikimin TMO (b) Chiu TMO, (c) Durand TMO, (d) Linear TMO and (e) Reinhard TMO. These maps are obtained for a viewinf time of 0.5 seconds (i.e. from 0 to 0.5 seconds of viewing time).

Observe how the two maps are quite different even though the source HDR image ('lampickaHDR') is the same. The reason is that TMOs being non-transparent can destroy perceivable image information which can make non-salient regions into salient ones and vice-versa. To visually exemplify this, we have shown in Fig. 5, the tone mapped versions of the 'lampickaHDR' HDR image processed by Ashikimin and Tumblin TMOs. We can immediately make two observations from this Fig.. Firstly, the image processed by Ashikimin TMO has more details preserved in the regions highlighted by red boxes. As opposed to this, in the same regions of the image processed by Tumblin TMO, the details are clearly missing. Secondly, the overall contrast of the image in Figure 5 (a) is clearly better than the one in Figure 5 (b). To visualize and relate this to the impact of these TMOs on VA, we have shown the corresponding VA maps obtained from eye-tracking. The reader can notice that for the image processed by Tumblin, the attention regions are mainly the books in the background while the letter pad (in the foreground) is nearly unnoticed by the observers. We hypothesize that this happens because with Tumblin being a global TMO the overall image contrast is maintained and the details in the dark areas of the image (like the books in background) are well retained. Further the 'owl' below the 'lamp' is also clearly visible and is a salient region. However, as already mentioned, all this comes with the price of losing finer details mainly in the bright areas (like the lamp and the letter pad) as highlighted. As a result, the 'letter pad' is nearly non-salient since the useful information (the text inside) has been 'washed away' by the TMO. In contrast to this, the VA map of the image processed by Ashikimin TMO shows that the written text on the 'letter pad' is the most salient portion. Also, the darker background (mainly the books) seems to have become less 'eye-catching' since the contrast in that part is reduced. Another example to illustrate that TMOs can modify the attention regions is shown in Fig. 6. Here the images in the first and second rows are the tone mapped versions of 'rend02_oC95' image processed by Drago and iCAM06 TMOs and the corresponding human priority maps (VA maps) respectively. It can be seen that the two 'red mats' (highlighted by red boxes) are more clearly visible in the image processed by iCAM06 since there is high contrast preserved in and around that region. Consequently, one can see from the corresponding VA map that these indeed are salient regions for the human observers. On the other hand, in the image processed by Drago there is much lower contrast in the said regions. As a result of these attract much lesser eye attention as seen in the corresponding VA map. A second set of examples is shown in the third and fourth rows of Fig. 6. Here the third row shows three tone mapped versions of 'Oxford_Church' image (tone mapped by Tumblin, iCAM06 and Linear TMOs) while the corresponding VA maps are shown in the fourth row below each image. Again, one finds that the 'orange spot' (highlighted by red box) is a salient region only in case on linear TMO (see the VA map in Fig. 6 (k)) since this TMO destroys contrast on other regions which makes the 'orange spot' stand out and thus eye catching. As opposed to this, Tumblin and iCAM06 TMOs provide much better contrast in other parts of the image as well. So the 'orange spot' is nearly non-salient in these two images as the observers' attention is attracted to other parts. Therefore, based on our experiments and analysis of the VA maps obtained from eye-tracking experiments, we can say that contrast of the resultant tone mapped image plays an important role in VA behaviour. As exemplified by visual examples in Figs. 5 and

6, the areas that attract eye-attention can vary even within the same image depending on whether contrast is preserved or destroyed by the TMO. This therefore suggests that contrast indeed is a vital dimension in HDR content processing from VA view point. Our observations are also in agreement with those made in previous subjective studies[2-12] regarding the impact of TMOs on perceptual quality and visual appeal: among the several attributes involved, contrast has a significant impact[17] on visual quality judgement.

The reader will recall that we used 11 HDR image contents. It can be noted from Fig. 4 that the distance matrices corresponding to different content exhibit clear differences. This implies that the impact of TMOs on the fixation behavior is dependent on the image content. This can be explained since each image has different regions of interest (i.e. the portions that attract more attention) and that the effect of TMOs on these regions can be quite different as elaborated above. For instance, referring to the entry (6,3) in the matrix shown in Fig. 5 one can observe large difference between the VA maps of images processed by Ashikimin and Tumblin TMOs for 'Apartment_float_o15C' image.

Therefore, as analyzed and explained we note significant differences for both intra (i.e. for each image content) and inter (i.e. between different image contents) cases. A theoretical explanation for this is the manner in which TMOs operate. Most of them sacrifice one or the other type of visual information in order to reduce the dynamic range. In the process, additional artifacts (such as additional contours) might be introduced. The eventual result is that a non-attentional region in the HDR image become attentional one in the tone mapped version. The opposite case is that in which structural information is destroyed due tone mapping. In such cases, an attentional region in the HDR image becomes less important (or less eye catching) in the tone mapped image. For example, a contrast that was visible in the HDR image becomes invisible in the processed image (loss of visible contrast). We have already provided some visual examples to illustrate these points.

We have mentioned in Section 3.3 that apart from the subjective experiments, we also generated VA maps from 3 existing computational models i.e. the methods proposed by Itti[26], Perreira Da Silva et al.[27] and Bruce et al.[29]. The reader may naturally wonder how these objective methods compare with the subjective ground truth in identifying the visually salient regions of tone mapped images. Our preliminary analysis of the VA maps obtained from these methods reveals that none of them are accurate enough in predicting fixation points that can replace human priority maps. We observed that the methods by Itti et al. and Perreira Da Silva et al. tend to be 'conservative' in terms of highlighting the salient regions. That is the VA maps obtained from them usually have fewer salient regions as compared to the ground truth. For better comprehension of this point, in Fig. 7, we show the ground truth VA maps with the VA maps from these methods for the image 'memorial_o876' processed by Chiu TMO. Notice how the maps from Itti et al. and Perreira Da Silva et al. methods recognize very few salient points. We suspect this is due to the theoretical design of these methods wherein the more salient parts (peaks) are emphasized in the final map resulting in more 'concentrated' saliency maps. On the other hand, the method by Bruce et al. produces VA maps which highlight almost all the edges as salient portions. In that sense, this method appears to overestimate the salient parts of the image as opposed to the ones by Itti et al. and Perreira Da Silva et al. which tend to give 'underestimated' results.

The reader may note that for the eye-tracking experiments, we displayed each image for 10 seconds and a gray screen appeared for three seconds between each image display. In this paper we considered only still tone mapped images and VA behavior could change with the time allowed to view an image. This is in contrast to eye-tracking with video stimuli where each frame appears only for a pre-defined duration Therefore it will be interesting to analyze the observer responses for different viewing durations. To that end, we considered three time intervals: (a) between 0 to 0.5 seconds, (b) between 0.5 to 4 seconds, (c) between 0 to 4 seconds of viewing time. These were chosen based on the idea that as soon as we observe an image, the bottom-up mechanism of the HVS drives where we direct our attention in the image. After the initial duration, the top-down effect comes into picture and at steady state both the mechanisms compete and the human response is affected by both of them. So we computed the VA maps for the three chosen durations. Note that the second case was chosen to see how the VA maps change in between 0.5 and 4 seconds of viewing time. For the duration between 0 and 0.5 seconds, we found that there is larger difference (in terms of the KLD values as compared to the second and third cases case) in the VA maps of images processed by different TMOs. We can explain this because in such a small duration (of 0.5 seconds), although the viewer attention is mainly in and around the central part of the displayed image there are some points/regions which might be far from each other. Consequently while centrally biased VA maps were obtained for this case, the KLD values indicated larger differences. This is because there are fewer isolated points (due to very short viewing time) to compute the KLD. So there is larger distance between these isolated points/regions of saliency. To give a numerical example, in this case the KLD values could be as high as 3 while in the remaining two cases they were around 1.6 (or less). The lower KLD for the remaining cases is because with increasing

viewing time there would be more 'overlap' in terms of what the viewer might be attracted to leading to more uniform maps. An example of VA maps for the first case (viewing time 0 to 0.5 seconds) is shown in Fig. 8 (5 VA maps of 'bigFogMap' image processed by Ashikimin, Chiu, Durand, Linear and Reinhard TMOs) where we find that while there is a tendency of being centrally biased, these maps are still different from each other since there are more isolated points (which increases the KLD). We further found that the VA maps for the remaining two cases (i.e. second and third chosen time intervals) were quite similar to each other. Thus, we find that TMOs indeed change the human attention behaviour even if the image is displayed for a very short time interval. Obviously, with increasing viewing time, such differences become larger.

As pointed out in the introduction, to our knowledge, none of the existing subjective studies have evaluated the impact of TMOs on VA. It is also interesting to note that the conclusions from those studies and ours seem to agree. They have all found that impact of TMOs is significantly different and depends on the image content. Our study also leads to a similar conclusion albeit regarding VA and not visual quality. Their focus has been the perceptual quality or visual appeal of the tone mapped images while this has paper focused on VA. We therefore believe that VA is a complimentary and crucial aspect (in addition to visual quality) for assessing the overall perceptual impacts of TMOs. Not surprisingly, a few recently developed[20-21] TMOs employ VA for dynamic range reduction. The method[20] by Mei et al. uses the visual saliency map of the HDR and then use the saliency of the local regions to control the local tone mapping curve such that highly salient regions will have their details and contrast better protected so as to remain salient and attract visual attention in the tone mapped display. The TMO[21] proposed by Lin et al. also utilized attention maps to locally adjust the contrast of the HDR image according to attention and adaptation models found in psychophysics. It was found that their proposed TMO produced images with better quality in terms of preserving details and chromaticity of visual saliency. VA has also been used in selective tone mapping[22] for hardware applications where computational complexity is a critical issue. The base idea was to reduce the computational costs of existing local TMOs by focusing more on the salient portions in the image.

5. CONCLUSIONS

Most of the existing subjective studies were directed towards assessing the impact of TMOs in terms of perceptual quality. As a result, the important dimension of VA was neglected by these studies. To address that, eye movements of observers watching in free task the tone mapped images were recorded. A large number of HDR images and TMOs were employed in our tests to ensure better reliability of the conclusions being drawn. A systematic statistical analysis was also performed to reduce the number of images to be viewed by subjects. The results indicate that TMOs indeed modify fixation behavior greatly. This is an important result since it suggests that exploiting VA for HDR content processing will be advantageous. Further, an important observation was that contrast plays a big role in specifying the salient regions. It is therefore useful to point out that eye-tracking experiments on LDR images should take this into account. That is the contrast of the display devices used in say two different two eye-tracking studies must be the similar if we are to make any meaningful comparson and conclusions about the results obtained. As mentioned, already few TMOs based on VA have been developed and shown promising performance. Further, we strongly believe that the use of VA may not be just limited to developing TMOs but might also be beneficial for objective perceptual quality assessment as well. It will therefore be interesting to investigate both VA and quality of tone mapped images together and see if the two could provide better models for objective quality assessment.

REFERENCES

[1] Akyuz AO, Fleming R, Riecke BE, Reinhard E, Bulthoff HH. Do HDR displays support LDR content? A psychophysical evaluation. ACM Transactions on Graphics 2007;26(3).

[2] Chiu K, Herf M, Shirley P, Swamy S, Wang C, Zimmerman K. Spatially nonuniform scaling functions for high contrast images. In: Proceedings of graphics interface '93. 1993. p. 245‑53..

[3] Fattal R, Lischinski D, Werman M. Gradient domain high dynamic range compression. In: Proceedings of the 29th annual conference on computer graphics and interactive techniques. New York, NY, USA: ACM Press; 2002. p. 249‑56.

[4] Durand F, Dorsey J. Fast bilateral filtering for the display of high-dynamic range images. In: Proceedings of the 29th annual conference on computer graphics and interactive techniques. New York, NY, USA: ACM Press; 2002. p. 257‑66.

[5] Drago F, Myszkowski K, Annen T, Chiba N. Adaptive logarithmic mapping for displaying high contrast scenes. Computer Graphics Forum 2003;22(3).

[6] Ashikhmin M. A tone mapping algorithm for high contrast images. In: 13th eurographics workshop on rendering. Eurographics Association; 2002. p. 145‑56.

[7] Reinhard E, Stark M, Shirley P, Ferwerda J. Photographic tone reproduction for digital images. In: Proceedings of the 29th annual conference on computer graphics and interactive techniques. ACM Press; 2002. p. 267‑76.

[8] Mantiuk R, Myszkowski K, Seidel H-P. A perceptual framework for contrast processing of high dynamic range images. In: Proceedings of the 2nd symposium on applied perception in graphics and visualization, APGV'05, New York, NY, USA: ACM Press; 2005. p. 87‑94.

[9] Mantiuk R, Daly S and Kerofsky L. Display Adaptive Tone Mapping. In: ACM Transactions on Graphics 27 (3), 2008.

[10] Kuang, J., Johnson, G.M., Fairchild M.D. iCAM06: A refined image appearance model for HDR image rendering, . J. Visual Communication and Image Representation 18(5): 406-414 (2007).

[11] Ward LG. A contrast-based scalefactor for luminance display. Graphics Gems 1994;IV:415‑21.

[12] Tumblin J, Hodgins JK, Guenter BK. Two methods for display of high contrast images. ACM Transactions on Graphics 1999;18(1):56‑94.

[13] Drago F, Martens WL, Myszkowski K, Seidel, H-P. Perceptual evaluation of tone mapping operators. In: Proceedings of the SIGGRAPH 2003 conference on sketches & applications, GRAPH'03, New York, NY, USA: ACM Press; 2003. p.1.

[14] Kuang, J., Johnson, G., and Fairchild, M., "iCAM06: a refined image appearance model for HDR image rendering," J. Vis. Comun. Image Represent. 18(5), 406–414 (2007).

[15] Yoshida A, Blanz V, Myszkowski K, Seidel, H-P. Perceptual evaluation of tone mapping operators with real-world scenes. Human Vision & Electronic Imaging X, San Jose, CA, USA: SPIE;2005. p. 192‑203.

[16] Ledda P, Chalmers A, Troscianko T, Seetzen H. Evaluation of tone mapping operators using a high dynamic range display. In: Proceedings of the 32nd annual conference on computer graphics and interactive techniques, SIGGRAPH 05. ACM Press; 2005, p. 640‑8.

[17] Cadik M, Wimmer M, Neumann L, Artusi A. Evaluation of HDR tone mapping methods using essential perceptual attributes, Computers & Graphics 32 (2008) 330‑349

[18] Ashikhmin M, Goyal J. A reality check for tone-mapping operators. ACM Transactions on Applied Perception 2006;3(4):399‑411.

[19] Rensink R. "Visual attention" In Encyclopedia of Cognitive Science, London: Nature Publishing Group, 2002.

[20] Mei Y, Qiu G, Lam K. Saliency Modulated High Dynamic Range Image Tone Mapping. In: Proceedings of the Sixth International Conference on Image and Graphics, 2011, p. 22-27.

[21] Lin W, Yan Z. Attention-based high dynamic range imaging. Vis Comput (2011) 27: 717–727.

[22] Artusi A, Roch B, Chrysanthou Y, Michael D, Chalmers A. Selective Tone Mapper. University of Cyprus, Technical Report, TR-05-07.

[23] Banterle F, Artusi A, Debattista K, Chalmers A. Advanced High Dynamic Range Imaging: Theory and Practice. ISBN: 978-156881-719-4, AK Peters (CRC Press), Natrick, MA, USA.

[24] Recommendation ITU-R BT.500-11, "Methodology for the subjective assessment of the quality of television pictures", June 2002.

[25] Q. Ma and L. Zhang, "Image Quality Assessment with Visual Attention," In: Proceedings of International Conference on Pattern Recognition (ICPR), December 8-11, 2008,p.1-4.

[26] Itti L., Koch C., Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11), 1998, p.1254–59

[27] Perreira Da Silva., Courboulay V. Implementation and evaluation of a computational model of attention for computer vision. Developing and Applying Biologically-Inspired Vision Systems: Interdisciplinary Concepts, editors Mark Pomplun and Junichi Suzuki, IGI Global 2012, doi: 10.4018/978-1-4666-2539-6.

[28] http://www.mpi-inf.mpg.de/resources/hdr/datmo/

[29] Bruce N, Tsotsos J. Saliency, Attention, and Visual Search: An Information Theoretic Approach. Journal of Vision 9:3, p 1-24, 2009.

[30] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Prediction, Inference and Data Mining. Second Edition, Springer Verlag, 2009.