Saliency as compact regions for local image enhancement

Clément Fredembach
Canon Information Systems Research Australia pty ltd.
{clement.fredembach@cisra.canon.com.au}

Abstract

Professional photographers compose and process an image to emphasise the image's subject. Images with high salience, where a region is highly distinct from its background, are perceived to be of much greater quality in panel tests. Because of technical and expertise considerations, "average" camera users often capture images that have a lesser salience, thereby decreasing the image's appeal.

The standard workflow to increase the perceived salience of an image's main subject consists in identifying the region of interest, and processing that region according to a set of rules. The level of analysis and processing can greatly vary, from increasing saturation or sharpness to identifying semantic concepts, e.g., faces, and employ a complex, tailored, modification.

This is a delicate problem to approach: saliency prediction algorithms are currently not precise enough, and region classification is necessarily limited to a few specific classes. Furthermore, the variety of content often precludes the usage of a fixed set of rules in the enhancement step.

Rather than attempting to predict saliency in images, we propose that important regions are somewhat distinct from their surroundings and can be identified by features that are spatially compact, in addition to standard compositional cues. Having identified the region of interest, we provide an enhanced image by increasing the values of its compact feature(s), i.e., increasing the perceived saliency of the region of interest. Preference studies indicate our modified images are significantly preferred to the original ones.

Introduction

An image, digital or hardcopy, is only an imperfect representation of a scene, produced by a device whose accuracy is limited by its technical (e.g., optical, mechanical or electronic) capabilities. Should one be able to create an ideal, perfect imaging device, the images it would produce would still not be optimal in a preferred, subjective sense. A device able to reproduce physical reality perfectly does not take into account the major processing centre that is the human brain, in particular the visual cortex. In addition, humans compare images to their perception or memory of an original scene. Memory is imperfect and affected by preference [13]. As a result, it is rare that users require or even desire that an image be a perfect replica of reality. Instead, a preferred reproduction is being sought; every image can be improved, in a preferred sense at least.

Improving the perceived or subjective quality of an image is as old as pictorial art itself, and photography in particular. There are indeed few images that cannot be improved by altering their contrast, saturation or colour balance. Improvements, enhancements, or image modifications with the intent of providing a (more) preferred representation of the scene have traditionally followed two distinct paths: global and local image modification.

Global methods affect all parts of an image equally and have been employed for a long time as a means to alter the reality of a scene, usually related to the nature of the light impinging on the scene or light-sensitive imaging elements. Examples of global methods include the use of yellow or red filters to modify image contrast in black and white photography, unsharp masking for sharpness, or white balancing. While capable of significant image improvement, the usefulness of global methods is generally limited either by the need for manual intervention, or the range of scenes to which they can be applied. Indeed, when a scene's statistics do not comply with the method's underlying assumptions, a frequent occurrence, global modifications can significantly decrease the perceived image quality [9].

Local modification methods affect regions differently, depending on the regions' characteristics. In particular, local methods can be better tailored to a specific image or region, thereby avoiding many of the pitfalls of global enhancement techniques. This increased precision allows small numerical modifications to result in significant perceptual changes. Because the image is subject to less changes, local enhancement methods are potentially more robust than global ones. However, increased precision is generally gained at the expense of versatility: local methods are highly specific and apply a single correction, or are targeted towards a particular use-case.

This paper proposes a method to improved subjective image quality by increasing the perceived salience of selected image regions. The difficulty in doing so traditionally stems from the lack of prior knowledge one has about what is important in a given image, see Section 2. Rather than building a saliency model or attempting to classify image regions into perceptually relevant classes, we hypothesise that an input image already exhibit some degree of contrast in its salient region. Specifically, we decompose the image into a series of simple Key Attributes (KA) such as opponent hues, luminance, sharpness, contrast and calculate their spatial compactness. The underlying idea is that KA that have a compact distribution are better indicators of content contrast and saliency than widely spread ones. We subsequently weigh the most compact attributes with information related to compositional rules, e.g., important salient regions are more likely to be located close to the centre of the image [7, 10] and region size, which reduces the chances of enhancing a distractor such as noise. Finally, we modify the values of the relevant KA to emphasise the region of interest. An illustration of our method is shown in Fig. 1. We validate our approach by performing a forced choice preference experiment that show our enhanced images are preferred over the original ones by a 9:1 ratio.

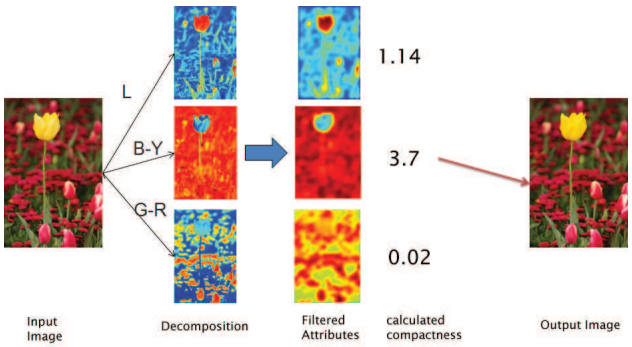

Figure 1. The different steps of our proposed image enhancement method. The input image is decomposed in a series of simple Key Attributes: here the L^*, a^*, and b^* channels of the CIELab colourspace; the channels are filtered using a centre bias filter and a size filter; the compactness of the KA is calculated and the attribute whose compactness is the highest (here: b^*). is selected for modification, resulting in the output, enhanced, image.

Salient regions as compact regions

Putting a strong emphasis on the main subject of an image, i.e., making it highly salient, can be done with compositional (e.g., selecting a background that contrasts with the foreground object) or physical (e.g., bokeh obtained by shooting with a large aperture) techniques. For general images, however, inferring what part of the image is important to the photographer or viewer from the pixel values is a very arduous task. Eye tracking is a reliable way to acquire that information [12], but it is highly unpractical in a general context.

Predicting regions of interest in images is generally performed by region classification over pre-determined classes, e.g., memory colours [4, 11, 13] or faces [3], or by determining salient regions in images from a viewer's perspective [1, 6]. The variety of natural images, however, diminishes the relevance of region classification, while salient region prediction methods do not currently correlate well enough with actual salience [5].

To avoid the difficulties associated with saliency prediction and image classification, we assume that regions of interest are (somewhat) distinct from the rest of the image in terms of simple low-level features. Instead of considering the magnitude of that distinction directly in terms of feature values, we calculate the spatial compactness of the distribution of these features across the entire image. The compactness is calculated as the kurtosis of the feature distribution, i.e., the "peakier" the distribution, the higher the kurtosis. In this work, we will use seven Key Attributes as features: Luminance; 0°, 45°, 90°, and 135° hue lines in the a^*b^* plane; contrast; sharpness. Contrast at a pixel is measured as the Michelson contrast of a 15×15 pixel window centred at that pixel. Sharpness at a pixel is taken to be the spectrally weighted magnitude of the Fourier transform taken over a 21×21 pixel window centred at that pixel. Luminance is the L^* channel of CIELab, while the hue features are defined, from a^* and b^* to be:

$$\begin{aligned} H_0 &= a^* \\ H_{45} &= \text{sign}(a^*+b^*)|\sqrt{a^{*\,2}+b^{*\,2}}\cos(\frac{\pi}{4}-\text{atan}(\frac{b^*}{a^*}))| \\ H_{90} &= b^* \\ H_{135} &= \text{sign}(b^*-a^*)|\sqrt{a^{*\,2}+b^{*\,2}}\cos(\frac{3\pi}{4}-\text{atan}(\frac{b^*}{a^*}))| \end{aligned}$$

Of course, determining saliency in such a fashion potentially introduces errors when distractors (small, highly distinct objects drawing attention away from the image's main subject) or noise are present. To maximise the chances of detecting actual regions of interest, we filter the feature distributions with a compositional filter and a size filter. The compositional filter is based on observations by several saliency prediction experiments which concluded that salient regions were, in everyday photographs, overwhelmingly located near the centre due to average photographer's compositional bias [7, 8]. Our compositional filter is a 2D gaussian centred in the middle of the image, with a standard deviation of a quarter of the image width. Recent work [2] has shown that large online photo-collections increasingly followed the compositional "rule of thirds". The primary use of the compositional filter is to prevent otherwise salient distractors or artefacts to be detected. The simple central gaussian filter we employed performed satisfactorily, but one should keep in mind that for different images/databases, the compositional filter can be modified to better reflect the photographers' skill or bias, e.g., the rules of thirds is likely to be more prevalent in a semi-professional image repository, while centre bias occurs with a greater frequency in casual snapshots.

The size filter gives more weight to regions of a given size. To be adequate, it has to decrease the perceived importance of small potential salient elements such as noise and image artefacts that generally subtend less than one degree of visual angle, while emphasising perceptually relevant regions. Several studies have suggested that human eye fixations may be correlated with the size of the underlying region [7, 10] . Our own experiments indicate that most salient objects are comprised between 3 and 5 degrees of subtended visual angle. The results presented in this paper were obtained using a 4° box filter as our size filter.

Let I be the input image to the process and f_i, $i = 1, \ldots, 7$ be the seven Key Attributes used to analyse the image. We write the decomposition of I in key attributes as:

$$F_i = f_i(I) \tag{1}$$

the "equivalent probability distribution" of the decomposed image is:

$$P_i = \frac{F_i}{\sum_x \sum_y F_i} \tag{2}$$

where x and y are the row and column indexes of the image, i.e., the operation is done on a per-pixel basis.

Prior to calculating the compactness, we filter the image with the compositional filter C and the size filter S, as defined above.

$$Pc_i = P_i * C \tag{3}$$

$$Pcs_i = Pc_i * S \tag{4}$$

where $*$ is the convolution operator.

The relevant KA for defining saliency and enhancing the image is k_{opt}:

$$k_{\text{opt}} = \max_i(k_i) \tag{5}$$

Where k_i is the kurtosis of Pcs_i, a measure of the concentration (or "peakedness") of the two dimensional distribution associated with every key attribute.

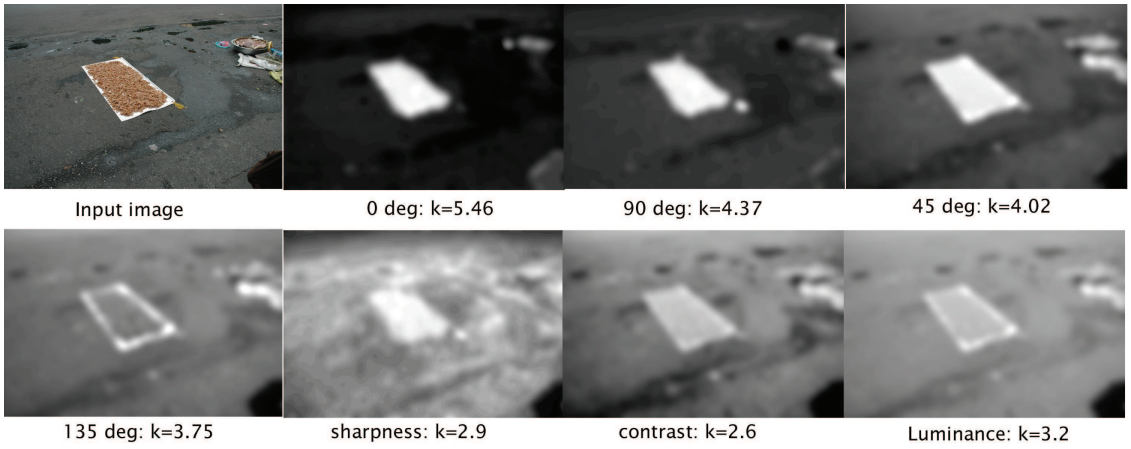

Figure 2. Original image and its decomposition into the seven proposed KAs: Luminance, 0°, 45°, 90°, and 135° hue lines, contrast, and sharpness. The calculated compactness value is shown underneath each image. For space reasons, only Pcs_i is shown.

Figure 2 shows an image decomposed by the seven KA and their measured compactness.

Image Enhancement

Because a strong salient region of interest is preponderant for perceptual image quality, most images can be enhanced by increasing the saliency of the region of interest. One however has to exercise caution when increasing salience, because introducing defects or artefacts in a highly visible region is especially detrimental to image quality [5].

We propose to noticeably but conservatively enhance the image by *modifying what is already there*. The "optimal" key attribute is the one whose distribution over the image is the most compact, i.e., it is the most discriminative KA between the region of interest and its background. Because this "optimal" KA already crystallises the saliency of the region of interest, a small modification of its value will have a visible effect. Because the modification is small, no artefacts are expected to be introduced, hence the robustness of the method.

In practice, we employ an s-shaped curve that will increase the high values of F_{opt} and decrease its low values. An s-shaped curve is ideal to prevent clipping of the KA values. We calculate F_{out}, the modified value of F_{opt} as:

$$F_{out} = h(F_{opt}) \tag{6}$$

where

$$h(x) = \frac{1}{1 + e^{n((x*\alpha)-\beta)}} \tag{7}$$

and n, α, and β are chosen to be 0.55, 20, and 10, respectively. An illustration of the type of tonal curves induced by $h(x)$ is shown in Fig.3.

While working on F_{opt} may modify image values outside of the actual region of interest, it is preferable than employing Pcs_{opt} for the modification step because the modified edges introduced by the successive filtering operations can produce halo-type artefacts.

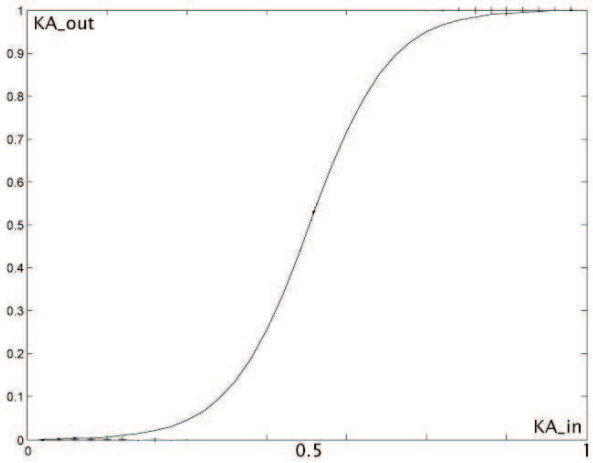

Figure 3. The tone curve $h(X)$ used to modify the values of the selected KA.

Results

Figure 4 displays the results obtained with the algorithm described in this paper. We show the input image, output image and indicate which one of the seven key attributes was selected for modification. All the input images are JPGs captured by a point and shoot camera with no additional processing.

The results show that our enhancement parameters, while conservative, are noticeable[1] and do not generate unwanted artefacts. This qualitative assessment is corroborated by the results of a panel test comprising 14 observers and 12 images. Both the original and enhanced image, printed on high quality glossy photo paper, were presented together and the observers were asked to indicate which one they preferred. In 91% of the cases, the enhanced image was preferred. The preference is significant ($p < 0.01$) with respect to both observers and images.

Colour features were responsible for most of the compact KA in the images shown in Fig. 4, as well as the other images of our tests. This appears to be in contradiction with most existing

[1]The results are better appreciated when viewed on screen.

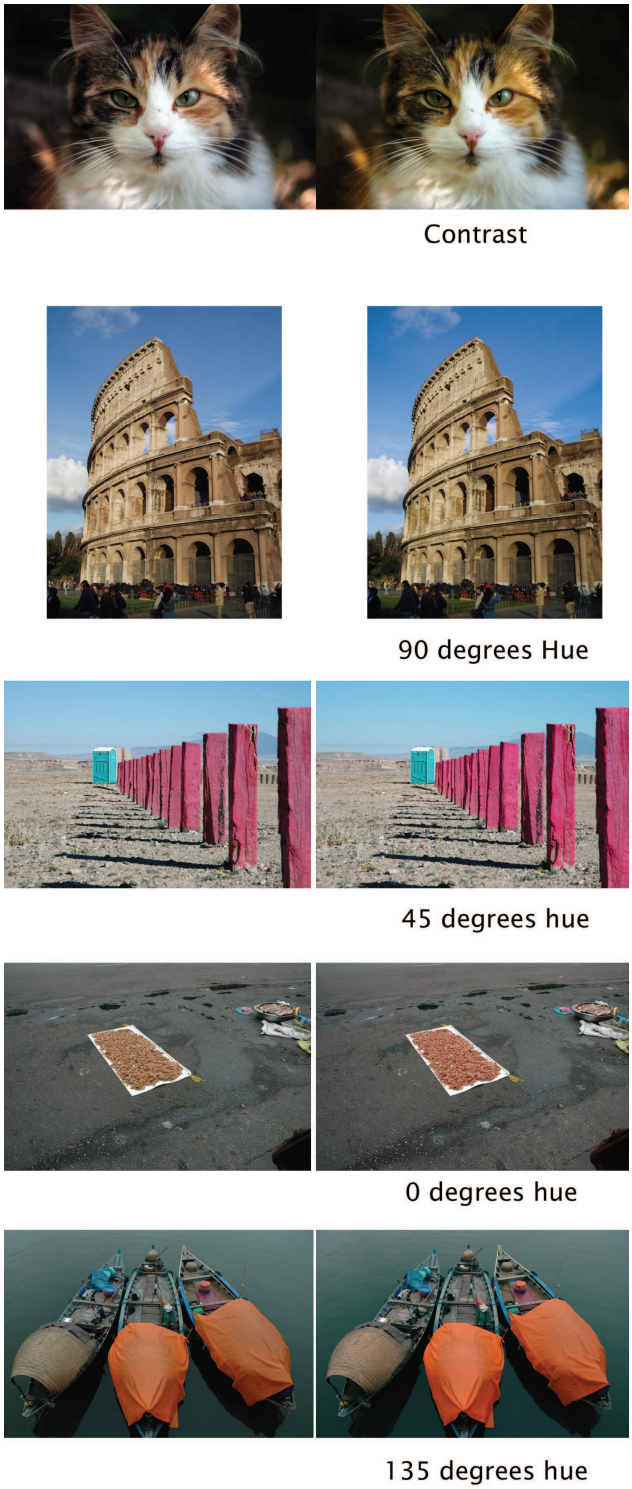

Figure 4. Original (left) and enhanced images with our method (right). For each image, the modified channel is also displayed.

saliency prediction algorithms, which weight colour very slightly. This phenomenon could be the result of our dataset (most compact camera have an almost infinite depth of field and contrast is sometimes equalised during the in-camera processing), coupled with the fact that people generally prefer more colourful images.

Applicability

An implicit assumption of our method is that the selected key attributes are sufficient to discriminate regions of interest from their background. While this assumption holds true for regions whose saliency is induced by low-level features, it is well known that a number of highly salient regions are identified through higher order cognitive processes, for example: faces and text. Some of these semantic regions can be detected with our method (e.g., text if sufficiently distinct from its background will be highly compact), but others normally require more sophisticated features for accurate detection. Similarly, if depth of field differences are not significant, one cannot increase the saliency of a face by enhancing its "faceness".

The problem of semantic classes is thus twofold: detecting these regions require complex features, which in turn cannot be readily used for enhancement. Addressing these issues is, however, not as hard as it appears, because the compactness hypothesis still holds. Indeed, if we include a face detector in our set of key attributes and look at its output pixel-wise to form equation (1), we can readily follow the rest of our method to determine whether faces are salient. If a single face is present in the image, the output of the detector will be highly compact and the face will be deemed salient. When many faces are present, the output of the detector will have a low compactness and other elements of the image will be regarded as more important from a saliency point of view.

The enhancement step for semantic classes would likely have to be modified, however, preferred rendering of high-level classes has been much studied and automatic steps exist, e.g., people prefer warmer tones in faces. Including such classes would necessitate an optimisation of the enhancement parameters based on observers' responses in a larger psychophysical experiment, departing from the generic character of the method presented here.

Conclusion

Local image enhancement and region of interest prediction are both hard problems. In this paper, we have proposed a simple, yet effective method to enhance an image by considering compact regions as salient, and increasing the features that induce this saliency, effectively bypassing the prediction task. Because the image is modified only with respect to attributes that already exhibit a difference between the detected region of interest and the rest of the image, small modifications produce visible effects without risking to introduce visible artefacts. Psychophysical experiments demonstrate that images processed with our method are significantly preferred over original one, by a 9:1 ratio.

References

[1] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk. Frequency-tuned salient region detection. In *IEEE conf. on Computer Vision and Pattern Recognition*, 2009.

[2] Sagnik Dhar, Vicente Ordonez, and Tamara L. Berg. High level describable attributes for predicting aesthetics and interestingness. In *IEEE conf. on Computer Vision and Pattern Recognition*, 2011.

[3] C. Dubout, M. Tsukada, R. Ishiyama, C. Funayama, and S. Süsstrunk. Face image enhancement using 3d and spectral

information. In *IEEE International Conference on Image Processing*, pages 697–700, 2009.
[4] C. Fredembach, M. Schroeder, and S. Süsstrunk. Region-based image classification for automatic color correction. In *IS&T/SID 11th Color Imaging Conference*, 2003.
[5] C. Fredembach, J. Wang, and G. Woolfe. Saliency, visual attention, and image quality. In *IS&T/SID 18th Color Imaging Conference*, 2010.
[6] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, pages 1489–1506, 2000.
[7] T. Judd, K. Ethinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *IEEE International Conference on Computer Vision*, 2009.
[8] D. Parkhurst, K. Law, and E. Niebur. Modeling the role of salience in the allocution of overt visual attention. *Vision Research*, pages 107–123, 2002.
[9] S. Pizer and et al. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, pages 355–368, 1987.
[10] S. Ramanathan, H. Katti, N. Sebe, M. Kankanhalli, and T.-S. Chua. An eye fixation database for saliency detection in images. In *European Conference on Computer Vision*, 2010.
[11] M. Schroeder and S. Moser. Automatic color correction based on generic content image analysis. In *IS&T/SID 9th Color Imaging Conference*, 2001.
[12] Alfred L. Yarbus. *Eye movements and Vision*. Plenum Press, 1967.
[13] S. Yendrikhovskij. Memory representation of object color. *Perception*, 1996.

Author Biography

Clément received an M.Sc. from EPFL, Switzerland in 2003 with internships in Fujifilm Japan and Gretag Imaging, and a Ph.D. from the University of East Anglia, Norwich, U.K. in 2007 on illuminant estimation, shadow detection and removal. From 2007 to 2009 he was a post-doc of the IVRG/LCAV at EPFL working on novel aspects of near-infrared imaging. He is now a senior research engineer with Canon Research (CiSRA) in Sydney, Australia working on saliency and perceptual image quality. He is a member of the IS&T.