A center-surround framework for spatial image processing

Vassilios Vonikakis and Stefan Winkler; Advanced Digital Sciences Center (ADSC), Singapore

Abstract

This paper presents a computational framework inspired by the center-surround antagonistic receptive fields of the human visual system. It demonstrates that, starting from the actual pixel value (center) and a low-pass estimation of the pixel's neighborhood (surround) and using a mapping function inspired by the shunting inhibition mechanism, some widely used spatial image processing techniques can be implemented, including adaptive tone-mapping, local contrast enhancement, text binarization and local feature detection. As a result, it highlights the relations of these seemingly different applications with the early stages of the human visual system and draws insights about their characteristics.

Introduction

Center-surround antagonistic Receptive Fields (RFs) are abundant in the Human Visual System (HVS). They have been found in many areas, such as the retina, the Lateral Geniculate Nucleus, V1 or in higher visual areas. It seems that this is a typical strategy that the HVS employs for local signal comparisons, not only in vision but in other sensory areas as well.

The RFs of center-surround cells comprise two separate concentric regions sampling the photoreceptor mosaic (namely the center and the surround) that act *antagonistically* on the final output of the cell. ON center-surround cells exhibit increased output with higher photoreceptor activity on their center and decreased output with increased activity on their surround. Conversely, for OFF center-surround cells, higher photoreceptor activity on the center has a negative impact on their output, whereas, increased photoreceptor activity on the surround increases their output. The size of the two regions defines the spatial frequency of sampling: smaller RF sizes sample finer details from the photoreceptor mosaic, while larger sizes encode coarser scales of the same signal.

Center-surround cells are essentially a biological implementation of spatial filtering. Spatial filtering is a very broad term, encompassing any kind of filtering operations that depend on the local content of the signal and are not globally constant. Almost all existing image processing and computational photography techniques include some kind of spatial image processing. Modern denoising, local contrast enhancement, scale decomposition, exposure fusion, HDR tone mapping are some of them. Most of these methods have some common grounds with the basic computational models of the early stages of the HVS. However these similarities are not always so evident.

In this paper, we start from the computational model of the first stages of HVS, developed by Grossberg [24], and we *adapt* it for image processing operations. Explicitly modeling HVS is out of the scope of this paper. We rather draw inspiration from it in order to address real-world imaging problems. More specifically, we define a framework, inspired by Grossberg's theory, that describes center-surround signal interactions. We show that such a framework can give rise to existing spatial image processing techniques, as many of them are special cases of it. This gives a more *unified view* between image processing and biological vision models, highlighting their common ground and showing other potential applications that can be developed.

Modeling Center-Surround RFs

Traditionally, center-surround RFs have been modeled as Difference of Gaussians (DoG) [13]. This *linear* operator essentially approximates the Laplacian operator, by subtracting two Gaussians of different sigmas, centered in the same position. DoG is at the heart of many computer vision and image processing algorithms, such as edge detection [12], scale-space construction [1] and local feature detectors [11].

Contrary to the linear response of the DoG operator though, the center-surround cells of the HVS exhibit *non-linear* response in regards to their inputs. Interestingly, their nonlinear response is thought to contribute to illumination invariance and contrast enhancement [24]. According to the standard retinal model [6, 21], the output V_{ij} of an ON-center OFF-surround cell at grid position (i, j), obeying the membrane equations of physiology is given by

$$\frac{dV_{ij}(t)}{dt} = g_{leak}\left(V_{rest} - V_{ij}\right) + C_{ij}\left(E_{ex} - V_{ij}\right) + S_{ij}\left(E_{inh} - V_{ij}\right) \tag{1}$$

with

$$C_{ij} = \sum_{p,q} I_{pq} G_{\sigma_C}(i-p, j-q) \;,\; S_{ij} = \sum_{p,q} I_{pq} G_{\sigma_S}(i-p, j-q)$$

g_{leak} is a decay constant, I is a luminance distribution (i.e. the image formed in the photoreceptor mosaic), V_{rest} (the cells resting potential), while E_{ex}, the excitatory reversal potential, and E_{inh}, the inhibitory reversal potential, are constants related to the neurophysiology of the cell. G_{σ_C} and G_{σ_S} are Gaussians representing the center and the surround of the cells receptive field, respectively, normalized in order to integrate to unity and with $\sigma_C < \sigma_S$. The steady-state solution of equation (1), assuming $V_{rest} = 0$, is given by:

$$V_{ij,\infty} = \frac{C_{ij}E_{ex} + S_{ij}E_{inh}}{g_{leak} + C_{ij} + S_{ij}} = \frac{C_{ij} - S_{ij}}{g_{leak} + C_{ij} + S_{ij}} \tag{2}$$

For ON-center OFF-surround cells $E_{ex} = 1$ and $E_{inh} = -1$. Thus, the operator of equation (2) is closer to a DoG divided by a Sum of Gaussians (SoG), augmented by the decay constant g_{leak}, rather than a linear DoG response. This mechanism is formally known as *shunting inhibition* (or *divisive inhibition*) and has been shown to contribute to the illumination invariant and contrast enhancement characteristics of center-surround cells [24]. More specifically, in steady state, center-surround cells compute a *ratio*,

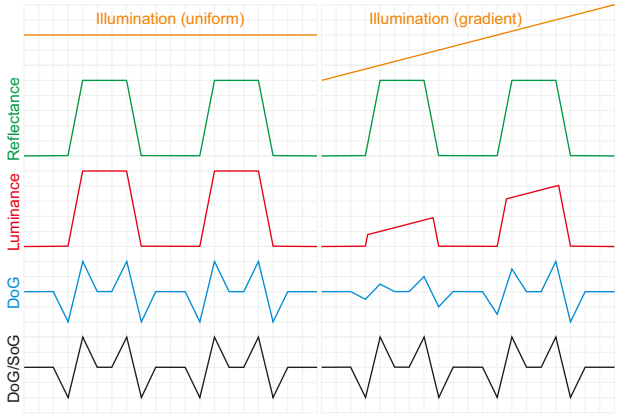

Figure 1: Simple 1-dimensional example depicting how an operator similar to equation (2) can be invariant to illumination gradients, contrary to the classic DoG (adapted from [24]).

normalizing locally the output of the linear DoG operator (numerator) with a local low-pass version of the input signal (denominator). In practice, this means that visual center-surround cells can increase their response under low illumination levels, due to the low activity of center and surround in the denominator of equation (2), exhibiting illumination invariant characteristics.

In the rest of the paper we will use the notation of center C and surround S (devoid of subscripts), to denote the fine and coarse scale of an input signal, respectively, at a given spatial position. It should be noted that, although in neurophysiology C and S are modeled by Gaussians, this should not necessarily be the case in other practical applications. A more intuitive way of thinking about this, is that S represents an averaging of the signal in the neighborhood of C. As it will be shown later, choosing what constitutes the neighborhood and how this averaging is calculated, can be done in many different ways and can have a great impact on the final output. Based on this notation, equation (2) can be transformed in the following way:

$$V = \frac{C-S}{g_{leak}+C+S} = \frac{C-S}{g_{leak}+2S+C-S} = \frac{x}{g_{leak}+2S+x} = \frac{x}{A+x}$$

where $x = C-S$ is a quantity describing the local contrast differences. It is clear that the cell's output V exhibits a non-linear response to x, adjusted by parameter A. This function is a form of the Naka-Rushton function [18] (also known as Michaelis-Menten in biochemistry), which has been identified in many vision-related cell types and has been associated with the enhancement of contrast sensitivity in the HVS [2]. The general form of the Naka-Rushton function is given by:

$$r(x) = \frac{B \cdot x}{A+x} \text{ with } x \in \mathbb{R}_0^+ \text{ and } A, B \in \mathbb{R}^+ \tag{3}$$

where x is the input signal, B is the maximum possible value of the output and A is a positive number that controls the non-linear degree of mapping. As Fig. 2a indicates, the Naka-Rushton function maps x from $[0,\infty) \rightarrow [0,B)$ in a non-linear way defined by parameter A. This type of mapping function is important in biology, because it can map an input x of arbitrary magnitude to an output bounded by B, since $\lim_{x\to\infty} r(x) = B$. Consequently, it can prevent the saturation of a cell's response. For the same reason,

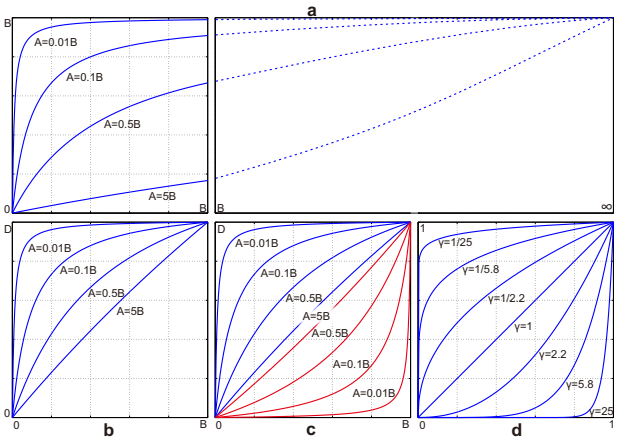

Figure 2: A. The Naka-Rushton mapping function $r(x)$. Different modulation factors A result into different non-linear mappings. Notice that all of them converge towards infinity, while in the interval $[0,B]$ the maximum output is not constant. B. Function $g(x,A)$. Notice that it has similar behavior to $r(x)$, but with fixed maximum output in the interval $[0,B]$. C. Function $g(x,A)$ (blue) and its inverse $h(x,A)$ (red). D. The gamma mapping functions.

function (3) is also used in tone-mapping HDR inputs to LDR outputs (with $A = 1$) [22]. In digital systems however, signals are always bounded within a specific range, i.e. $[0,B]$. As such, one important drawback of the Naka-Rushton function is that it does not exhibit a constant maximum output, within the $[0,B]$ interval, for different values of A. This creates limitations for many real-life applications, since the range of outputs is not independent of the non-linearity parameter A.

Center-surround framework

In order to overcome this limitation, we modify the Naka-Rushton function in order to exhibit a constant maximum output in the range of $[0,D]$, independent of A.

$$g(x,A) = \frac{(B+A)x}{A+x} \cdot \frac{D}{B} \tag{4}$$

$$h(x,A) = g^{-1}(x,A) = \frac{A \cdot x}{B+A-x} \cdot \frac{D}{B} \tag{5}$$

with $x \in [0,B]$ and $A,B \in (0,\infty)$. As Fig. 2b indicates, $g(x,A)$ maps x from $[0,B] \rightarrow [0,D]$ with a similar behavior to the Naka-Rushton function. However, it exhibits a constant maximum output for different values of A. The function in equation (4), and its inverse (5), define a *family of mapping curves* which can be a helpful tool for signal manipulation (Fig. 2c). Once B and D are defined according to the range of input and output data, respectively, varying A accordingly can act as a *modulation*, resulting in different non-linear mappings between input and output, ranging from a steep non-linearity ($A \approx 10^{-2}B$) to an approximately linear mapping ($A \geq 5B$). This also highlights the reason why A is selected to be used as an input variable (and not a parameter) in equations (4) and (5); the non-linearity of the mapping can be variable and not necessarily constant.

In comparison to other widely used families of curves, like the gamma curves (Fig. 2d), exhibits two main advantages. First and most important, curvature is more evenly distributed across the input and output axis. This is even more evident when $D = B$,

Table 1: Possible options for combining C, S, local differences ($C-S$ or $S-C$) and a constant α with the *signal* x and the *modulation* A of mapping functions (4),(5).

	x	A	**Model**
1	C	α	Tone mapping
2	C	$f(C)$,$f(S)$	Adaptive tone mapping
3	C	$f(C-S)$	Local contrast enhancement
4	S	α	Surround mapping
5	S	$f(C)$,$f(S)$	
6	S	$f(C-S)$	
7	$C-S$	α	Contrast mapping
8	$C-S$	$f(C)$,$f(S)$	Text binarization Illumination invariant scale-space
9	$C-S$	$f(C-S)$	

which is the case for Fig. 2c and 2d. In such cases, the mapping curves are *symmetric* over the line $y = B - x$. On the contrary, gamma curves are *asymmetric* and, for steep non-linear mappings, many input values are mapped to a single output value, resulting in loss of information. For example, when $\gamma = 25$, approximately 80% of the input values are mapped to zero. This behavior is much less evident for the g and h functions with a similarly steep non-linearity ($A \approx 10^{-2}B$). This difference is highlighted also in Fig. 3b and 3d. Second, the calculation of the curve consists only of simple operations, without the need to compute powers.

This family of mapping functions can be employed in order to create a computational framework that can address a variety of image processing problems. Following the retinal architecture, each pixel can be considered to be processed *independently* by a processing element, resembling a center-surround cell. As such, for each pixel 3 pieces of information are available: its actual value (C - center), a low-pass estimation of the values in its neighborhood (S - surround) and their difference ($C-S$ or $S-C$), which corresponds to the local details of the signal. The final value of the processing element is determined by the mapping functions g and/or h of equations (4),(5), depending on how the signal x and the modulation A are selected among C, S and $C-S$. All the possible combinations of options are listed in Table 1, along with the resulting models. Note that this approach is not only limited to 2D signals (e.g. images), but could also be applied to 1D cases (e.g. sound), such as in Fig. 1, or higher dimensional data.

This relatively simple framework, inspired by the local divisive normalization mechanism of equation (2), can be used in many different scenarios, for a variety of image processing and computer vision applications, giving rise to some seemingly unrelated methods, ranging from image tone manipulation, local contrast enhancement, binarization of text or detection of local feature.

Applications

Here we discuss some existing approaches and applications, arising as special cases of the aforementioned framework, namely, tone-mapping, local contrast enhancement, text binarization, and local feature detection. In all cases, p represents the new pixel value in position (i, j), C (center) the old pixel value, S (surround) the surround value in position (i, j), B and D the maximum value of the input and output signal (e.g. 255 for 8-bit images), respectively, and $f(x)$ a function that associates the modulation A with some other quantity (either C, S, $C-S$ or a constant α). Choosing different types of functions f and different input signals x can give rise to different spatial and global image processing techniques, as Table 1 indicates.

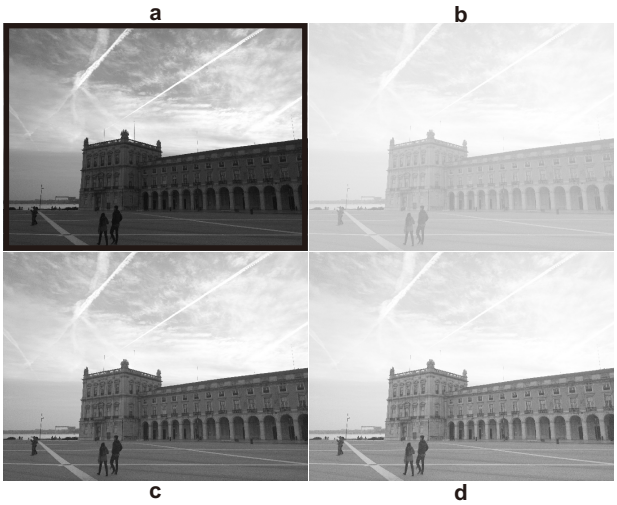

Figure 3: Tone-mapping using the function (4). a. Original image. b. Tone-mapping using a gamma of $1/6$. c. Tone-mapping using $g(C, \alpha = 0.2B)$. d. Tone-mapping using $g(C, \alpha = 0.1B)$.

Tone-mapping

Tone mapping (Table 1, line 1) is a simple operation that adjusts the values of the pixel according to a mapping function, in order to change the appearance of the image for aesthetic purposes or to compensate for overall under/overexposure. The new pixel value depends solely on its previous value and the value of the constant α. Consequently, the operation is *global* across the whole image. One of the simplest tone-mapping operators is the gamma function in Fig. 2d. Based on the center-surround framework, tone-mapping can be achieved by using the center C as the input signal and a constant α for the modulation.

$$p = g(C, \alpha) \text{ or } p = h(C, \alpha) \tag{6}$$

Equation (6) makes sure that all pixels will be processed with the same transformation, either increasing their values (through function $g(x,A)$), or decreasing them (through function $h(x,A)$). Fig. 3 depicts the results of a tone-mapping function implementing equation (6). Note the differences between Fig. 3b and 3d. Although they have similar impact on the sky region, the contrast on the building region is higher for Fig. 3d. Selecting the values of D and B adjusts the operation according to different use cases. When $D < B$ compression to a lower range medium is achieved, such as in HDR imaging. When $D = B$ range remains constant, such as when changing the tones of a single image.

Adaptive Tone-mapping

Correcting image tones in a local way is another important image processing task, aiming to compensate for the under/overexposed image regions caused by non-uniform illumination conditions. Based on the proposed framework, adaptive tone-mapping (Table 1, line 2) (also known as local tone-mapping) can be achieved by using the center C as the input signal and the sur-

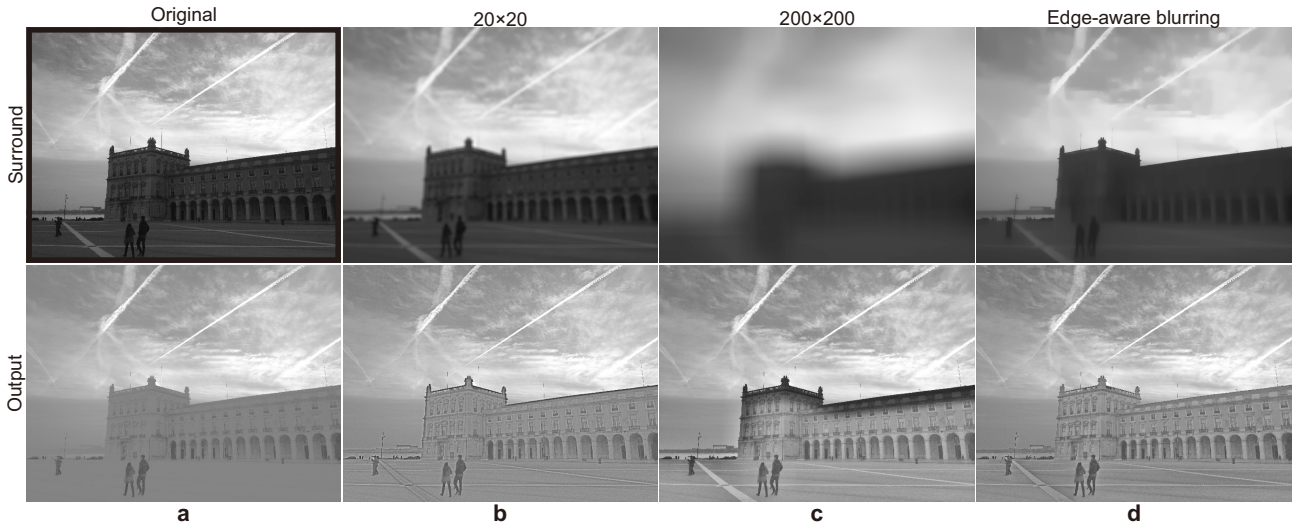

Figure 4: The impact of different types and sizes of surround S on the same enhancement algorithm [27]. a. Using the original image as a surround, i.e. $S = C$. b. A simple averaging filter of 20×20 pixels. c. An averaging filter of 200×200 pixels. d. Edge-aware blurring.

round S as the modulation factor.

$$p = \begin{cases} g(C, f(S)) & S \leq \frac{B}{2} \\ h(C, f(B-S)) & S > \frac{B}{2} \end{cases} \tag{7}$$

where f is a polynomial function that associates the modulation A with the surround S. Equation (7) makes sure that pixels located in dark image regions will increase their values (through function $g(x,A)$), whereas pixels in bright areas will decrease their values (through function $h(x,A)$), resulting in a more balanced overall image exposure. This approach has been used in [27], where f was selected as a quadratic function, along with other parameters that controlled the coefficients of the polynomial. A very similar approach was also reported in [15, 22], but without the lower part of equation (7). As such, pixels can only increase their values in the underexposed regions, whereas the overexposed regions remain untouched. Fig. 4 depicts the results of an algorithm implementing equation (7) for different sizes and types of surround S. Similarly to global tone-mapping, the relation between D and B defines the use case; compression from HDR to LDR or enhancing a single image.

Local contrast enhancement

Local contrast enhancement (also known as detail enhancement) (Table 1, line 3) is another important operation in image processing. It can add *definition* in images, or compensate for low contrast imaging conditions, such as hazy scenes or scenes affected by strong glare. According to the center-surround framework, the local contrast of images can be enhanced by using the center C as the input signal and $C-S$ or $S-C$ as the modulation factor A.

$$p = \begin{cases} g(C, f(C-S)) & C \geq S \\ h(C, f(S-C)) & C < S \end{cases} \tag{8}$$

where f is a function of the form $k/(l+x)$, with k and l constants, associating the modulation A with the local contrast surround $C-S$ or $S-C$. As a result, if the center C has a value greater than its surround S, the new pixel value p is increased. Conversely, if the center C has a lower value compared to its surround S, the new pixel value p is decreased. This increases local pixel differences with a variable degree, based on an estimation of the local contrast [26]. Fig. 5 depicts an implementation of equation (8), for two different scenes.

Figure 5: Implementation of the local contrast enhancement function described by equation (8), for two different scenes.

Surround mapping

Many existing image enhancement and adaptive tone-mapping techniques [3], as well as some image appearance models [10], rely on decomposing the image into 2 layers: a coarse layer (or base layer), which is a low pass version of the image and can be seen as an estimation of the illumination, and a layer of details, which can be seen as an estimation of reflectance. Depending on the final objective, the two layers are processed separately and recombined in order to form the output image. The coarse layer is usually tone-mapped in order to compensate for underexposed regions, using a gamma function. This operation can be implemented by the proposed framework (Table 1, line 4) by mapping the surround S values in a similar way to the tradi-

tional tone-mapping.

$$p = g(S,\alpha) \text{ or } p = h(S,\alpha) \tag{9}$$

Depending on the required processing, function g may be used for compensating for underexposure or function h for overexposure.

Contrast mapping

Extraction of local features is the first stage in many computer vision applications. One of the most popular local feature detectors is the Scale-Invariant-Feature-Transform (SIFT), which utilizes a scale space for the detection of the most important keypoints. Although the local minima and maxima of the scale space are invariant to illumination, in practice, a global threshold is used on the gradient amplitude in order to filter out keypoints corresponding to noise. However, uneven illumination can dramatically reduce the amplitude of the gradient, resulting in filtering out many valid keypoints located in dark image regions. One possible solution is to change the magnitude of the gradient in such a way that meaningful keypoints are above the global threshold [29]. This approach essentially applies a global mapping to the magnitude of the gradient and can be implemented by the proposed framework by using $S-C$ as the signal x, and a constant α as the modulation A, while $D=B$.

$$c = g(S-C,\alpha) = \frac{(B+\alpha)(S-C)}{\alpha+|S-C|} \tag{10}$$

where c is the new value of the gradient $S-C$. Notice that in the denominator, the absolute value $|S-C|$ is used rather than $S-C$. This departure from the typical form of function g in equation (4) is necessary in order to retain the polarity of the local contrast.

Illumination invariant scale-space

A more elaborate solution to the problem of low gradient magnitude caused by non-uniform illumination, is to adjust it in a *local* way, rather than global, according to an estimation of the illumination. In such a case, the magnitude of gradient will be increased in the underexposed regions, whereas it will remain the same in the well-exposed ones. The proposed framework can be adapted in order to implement this approach and form an Illumination Invariant DoG (iiDoG) operator [30] (Table 1, line 8), by using $S-C$ as the signal x, $S+C$ as the modulation A and $D=B=2-A$ in function g of equation (4).

$$\text{iiDoG} = \frac{(B+A)(S-C)}{A+|S-C|} = \frac{2(S-C)}{S+C+|S-C|} \tag{11}$$

Similarly to the previous section, the absolute value $|S-C|$ is used in the denominator in order to retain the polarity of the local contrast. The operator of equation (11) can be used to construct an illumination invariant scale-space. In this scale-space, a *single* global threshold can be used in order to filter out noise-related local extrema, while still detecting keypoints in the dark image regions. Fig. 6 depicts a comparison of the typical DoG-based SIFT detector and one based on the iiDoG operator.

Text binarization

Text binarization (Table 1, line 8) is an important part of automatic document analysis and of outdoor vision systems (e.g. license plate recognition). Uneven illumination though, can make the problem challenging. A black character in a highly illuminated image region can have a higher pixel value compared to a white region in a shadow. As such, no global threshold can achieve a reliable separation between characters and background. On the other hand, humans are excellent in distinguishing black letters on white background, irrelevant of the ambient illumination. In HVS, there are two separate neural pathways associated with encoding luminance increments and decrements, namely, the ON and OFF center-surround pathways. Text, which comprises dark characters over a brighter background (i.e. light decrements), is conveyed by the OFF pathway. Consequently, using the proposed center-surround framework one can implement a text binarization technique by using $S-C$ as the signal x and the surround S as the modulation factor A.

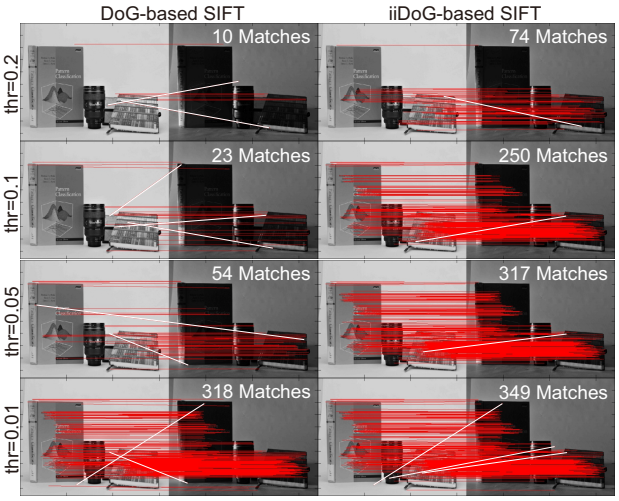

Figure 6: Matched features for a scene under 2 different types of illumination. Results are shown both for the classic DoG operator and the iiDoG operator of equation (11), for different values of the detector threshold.

$$p = \begin{cases} g(S-C, f(S)) & S-C > 0 \\ 0 & S-C \le 0 \end{cases} \tag{12}$$

where $f(x)=x$ is a simple identity function associating the modulation A with the surround S. Since the main objective is to detect text (darker signal over a brighter background), we are not interested in the cases where $C \ge S$ (brighter signal over a darker background) and thus, the lower branch of equation (12) is set to 0. This approach has been shown to be a good solution for *equalizing* light decrements across different illumination levels [28,31]. The final binary output can be acquired by using a simple global thresholding technique on the equalized decrement response, such as Otsu's thresholding [19]. Fig. 7 depicts the results of such an approach. Applying a binarization technique directly on the original image, results in a loss of information inside the dark image regions. However, if equation (12) is applied prior to binarization, information is maximized.

On choosing the surround

Selecting what constitutes the surround S and how it is computed can have a considerable impact on the final output. In gen-

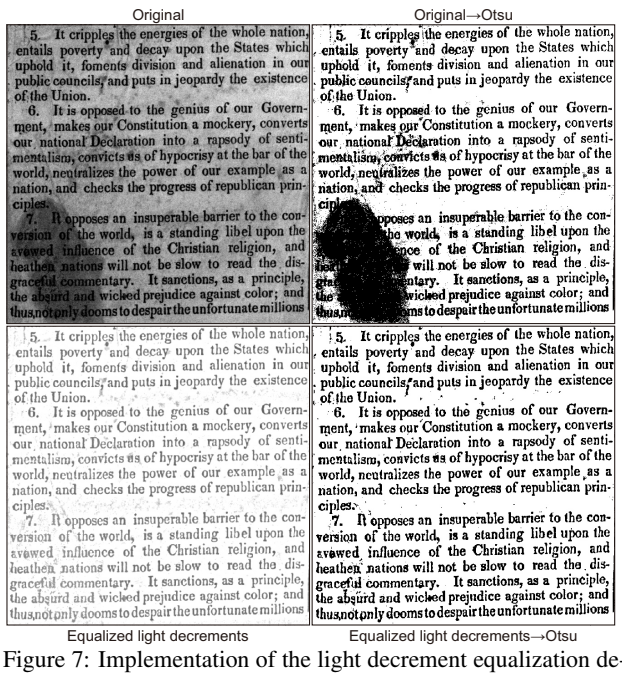

Figure 7: Implementation of the light decrement equalization described by equation (12).

eral, there is a direct relation between radius and degree of interaction between different spatial regions of the signal, level of local details, artifacts and computational complexity. There are many different approaches in computing S. All of them are characterized by the degree and type of low-pass filtering (blurring), the level of artifacts on the strong intensity transitions (halos and/or gradient reversal) and their computational burden. As a rule of thumb, stronger low-pass filtering leads to better estimation of local details (positive), but at the same time increases artifacts (Fig. 4) and, in some methods, the computational burden (negative).

For Gaussian blurring, which is considered one of the simplest ways of computing S, a larger radius and standard deviation leads to stronger blurring, better level of local details, larger halo artifacts and greater computational burden (Fig. 4c). Decreasing the size of spatial interaction, and thus the strength of blurring, is not an acceptable solution for dealing with halos, since it leads to low level of local details, resulting in a 'flat' and unnatural look on the final output (Fig. 4a, Fig. 4b). One of the first strategies to deal with the halo artifact problem was to use a collection of Gaussians with different radii [8, 9]. Soon however, it became evident that local filters that adjust to the local intensity of the signal are more useful [14].

This eventually led to the concept of Edge Aware Blurring (EAB) (or edge-preserving filtering), which comprises a big part of today's image processing research. EAB includes iterative methods such as Anisotropic Diffusion [20] and Robust Recursive Envelops [23], local filters such as Bilateral [25] and Guided filtering [7], scale decomposition, such as Edge Avoiding Wavelets [5] and Weighted Least Squares decomposition [4] and other global optimization techniques [17, 32]. All the above techniques, as well as many others, share many common elements. For an in-depth unified review please refer to [16]. Fig. 8 depicts the results of various EAB techniques.

In practice, for applications like adaptive tone-mapping and local contrast enhancement, a very strong edge-preserving blurring is required for high quality of results. Consequently, as it is evident from Fig. 8, Bilateral and Guided filters are not adequate for this task, since they exhibit inferior edge-preserving characteristics for the required level of blurring. On the other hand, scale decomposition techniques, as well as global optimization methods exhibit strong blurring with very high quality edge preservation. As such, they are better candidates for the calculation of the surround. Contrary to adaptive tone-mapping and local contrast enhancement, gradient manipulation applications and text binarization may not necessarily require EAB approaches. In these cases, simple Gaussian or averaging filters are found to be adequate for good quality of results.

Discussion

Although the transformations described in the previous section might initially look very different, they have one common ground; they all perform the same local signal comparisons in a neighborhood. What sets them apart is the type of signals used in these comparisons. As a result, the actual pixel value, a low-pass version its neighborhood, as well as their difference can be efficiently used in many different applications.

It should be noted that the transformations presented in the previous section should not necessarily be considered individually. They may also form *compound* transformations when they are applied as a cascade, using the same original estimations for center and surround. One example is local contrast enhancement and adaptive tone-mapping, which can be used together. Fig. 9 depicts this combination of transformations. It is evident that the final result incorporates the characteristics of both.

Conclusions

This paper introduced a computational framework inspired by the center-surround receptive fields of the HVS. It demonstrated that, starting from the actual pixel value (center) and a low-pass estimation of the pixel's neighborhood (surround) and using a mapping function inspired by the shunting inhibition mechanism, one can implement some widely used spatial image processing techniques. These include adaptive tone-mapping, local contrast enhancement, equalization of light decrements for text binarization and illumination invariant local feature detection. These seemingly different applications are all special cases of the same simple scheme, which can be traced back to the physiological structure of the early stages of the HVS and its center-surround RFs.

Acknowledgments

This study is supported by the research grant for the Human-Centered Cyber-physical Systems Programme at the Advanced Digital Sciences Center from Singapore's Agency for Science, Technology and Research (A*STAR).

References

[1] P.J. Burt and E.H. Adelson. The laplacian pyramid as a compact image code. *Communications, IEEE Transactions on*, 31(4):532–540, Apr 1983.

[2] Thang Duong and Ralph D. Freeman. Contrast sensitivity is enhanced by expansive nonlinear processing in the lateral geniculate nucleus. *Journal of Neurophysiology*, 99(1):367–372, 2008.

Figure 8: Results of various EAB techniques for strong blurring. From top left to right: original image, Bilateral filtering [25], Weighted Least Squares decomposition [4], L0 gradient minimization [32]. From bottom left to right: Robust Recursive Envelops [23], Guided filtering [7], Edge Avoiding Wavelets [5], Fast Global Image Smoothing [17].

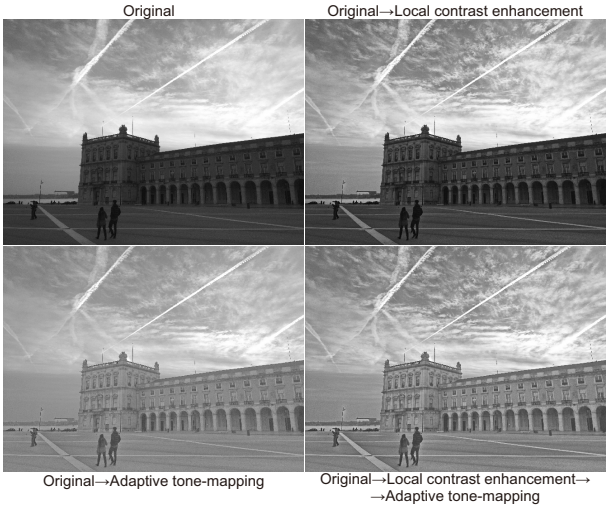

Figure 9: Combining local contrast enhancement and adaptive tone-mapping.

[3] Frédo Durand and Julie Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '02, pages 257–266, New York, NY, USA, 2002. ACM.

[4] Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM SIGGRAPH*, pages 67:1–67:10, New York, NY, USA, 2008. ACM.

[5] Raanan Fattal. Edge-avoiding wavelets and their applications. SIGGRAPH, pages 22:1–22:10, New York, NY, USA, 2009. ACM.

[6] Stephen Grossberg and Dejan Todorovic. Neural dynamics of 1-d and 2-d brightness perception: A unified model of classical and recent phenomena. *Perception and Psychophysics*, 43(3):241–277, 1988.

[7] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(6):1397–1409, June 2013.

[8] C. Iakovidou, V. Vonikakis, and I. Andreadis. Fpga implementation of a real-time biologically inspired image enhancement algorithm. *Journal of Real-Time Image Processing*, 3(4):269–287, 2008.

[9] D.J. Jobson, Z.-u. Rahman, and G.A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Image Processing, IEEE Transactions on*, 6(7):965–976, Jul 1997.

[10] Jiangtao Kuang, Garrett M. Johnson, and Mark D. Fairchild. icam06: A refined image appearance model for hdr image rendering. *J. Vis. Comun. Image Represent.*, 18(5):406–414, October 2007.

[11] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[12] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London B: Biological Sciences*, 207(1167):187–217, 1980.

[13] P. Martin and U. Grunert. Ganglion cells in mammalian retinae. In L.M. Chalupa and J.S. Werner, editors, *The Visual Neurosciences*, pages 410–421. MIT Press, 2004.

[14] L. Meylan and S. Susstrunk. High dynamic range image rendering with a retinex-based adaptive filter. *Image Processing, IEEE Transactions on*, 15(9):2820–2830, Sept 2006.

[15] Laurence Meylan, David Alleysson, and Sabine Ssstrunk. A Model of Retinal Local Adaptation for the Tone Mapping of Color Filter Array Images. *Journal of the Optical Society of America A (JOSA A)*, 24(9):2807–2816, 2007.

[16] P. Milanfar. A tour of modern image filtering: New insights and methods, both practical and theoretical. *Signal Processing Magazine, IEEE*, 30(1):106–128, Jan 2013.

[17] Dongbo Min, Sunghwan Choi, Jiangbo Lu, Bumsub Ham, Kwanghoon Sohn, and M.N. Do. Fast global image smoothing based on weighted least squares. *Image Processing, IEEE Transactions on*, 23(12):5638–5653, Dec 2014.

[18] K. I. Naka and W. A. H. Rushton. S-potentials from colour units in the retina of fish (cyprinidae). *The Journal of Physiology*, 185(3):536–555, 1966.

[19] Nobuyuki Otsu. A Threshold Selection Method from Gray-level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, 1979.

[20] P. Perona and J. Malik. Scale-space and edge detection using

anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639, Jul 1990.

[21] Luiz Pessoa, Ennio Mingolla, and Heiko Neumann. A contrast- and luminance-driven multiscale network model of brightness perception. *Vision Research*, 35(15):2201 – 2223, 1995.

[22] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '02, pages 267–276, New York, NY, USA, 2002. ACM.

[23] Doron Shaked and Renato Keshet. Robust Recursive Envelope Operators for Fast Retinex. Technical report, HP Laboratories Israel, 03 2004.

[24] Grossberg Stephen. Visual boundaries and surfaces. In L.M. Chalupa and J.S. Werner, editors, *The Visual Neurosciences*, pages 1624–1639. MIT Press, 2004.

[25] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846, Jan 1998.

[26] V. Vonikakis and I. Andreadis. Multi-scale image contrast enhancement. In *10th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pages 856–861, 2008.

[27] V. Vonikakis, I. Andreadis, and A. Gasteratos. Fast centre-surround contrast modification. *Image Processing, IET*, 2(1):19–34, Feb 2008.

[28] V. Vonikakis, I. Andreadis, and N. Papamarkos. Robust document binarization with off center-surround cells. *Pattern Analysis and Applications*, 14(3):219–234, 2011.

[29] V. Vonikakis, D. Chrysostomou, R. Kouskouridas, and A. Gasteratos. Improving the robustness in feature detection by local contrast enhancement. In *Imaging Systems and Techniques (IST), 2012 IEEE International Conference on*, pages 158–163, July 2012.

[30] Vasillios Vonikakis, Dimitrios Chrysostomou, Rigas Kouskouridas, and Antonios Gasteratos. A biologically inspired scale-space for illumination invariant feature detection. *Measurement Science and Technology*, 24(7):074024, 2013.

[31] Vassilios Vonikakis, Ioannis Andreadis, Nikos Papamarkos, and Antonios Gasteratos. Adaptive document binarization - a human vision approach. In Alpesh Ranchordas, Helder Arajo, and Jordi Vitri, editors, *VISAPP (2)*, pages 104–109, 2007.

[32] Li Xu, Cewu Lu, Yi Xu, and Jiaya Jia. Image smoothing via l0 gradient minimization. *ACM Trans. Graph.*, 30(6):174:1–174:12, December 2011.

Author Biography

Dr. Vassilios Vonikakis received his degree in Electrical and Computer Engineering, as well as his Ph.D., from the Democritus University of Thrace, Greece. Between 2009 and 2012 he worked as a researcher for the University of Rome 'Sapienza' and the CRAS Institute of Aerospace Research, in the enhancement of High Dynamic Range (HDR) radiographic medical images. He currently works as a Research Scientist for the Advanced Digital Sciences Center (ADSC), in Singapore.

*Dr. Stefan Winkler got his degree in electrical engineering/communications engineering at the Technische Universitat Wien (Vienna), Austria and his Ph.D. at the Ecole Polytechnique Federale de Lausanne, Switzerland. Currently, he is Principal Scientist and Director of the Video & Analytics Program at the Advanced Digital Sciences Center (ADSC) in Singapore, a joint research center by the University of Illinois at Urbana-Champaign (UIUC) and the Agency for Science, Technology and Research (A*STAR), leading the Vision & Interaction Group.*