Colour Constancy Beyond the Classical Receptive Field

Arash Akbarinia, *Member, IEEE,* and C. Alejandro Parraga, *Member, IEEE*

Abstract—The problem of removing illuminant variations to preserve the colours of objects (*colour constancy*) has already been solved by the human brain using mechanisms that rely largely on centre-surround computations of local contrast. In this paper we adopt some of these biological solutions described by long known physiological findings into a simple, fully automatic, functional model (termed Adaptive Surround Modulation or ASM). In ASM, the size of a visual neuron's receptive field (RF) as well as the relationship with its surround varies according to the local contrast within the stimulus, which in turn determines the nature of the centre-surround normalisation of cortical neurons higher up in the processing chain. We modelled colour constancy by means of two overlapping asymmetric Gaussian kernels whose sizes are adapted based on the contrast of the surround pixels, resembling the change of RF size. We simulated the contrast-dependent surround modulation by weighting the contribution of each Gaussian according to the centre-surround contrast. In the end, we obtained an estimation of the illuminant from the set of the most activated RFs' outputs. Our results on three single-illuminant and one multi-illuminant benchmark datasets show that ASM is highly competitive against the state-of-the-art and it even outperforms learning-based algorithms in one case. Moreover, the robustness of our model is more tangible if we consider that our results were obtained using the same parameters for all datasets, that is, mimicking how the human visual system operates. These results suggest a dynamical adaptation mechanisms contribute to achieving higher accuracy in computational colour constancy.

Index Terms—colour constancy, illuminant estimation, classical receptive field, surround modulation, centre-surround contrast.

1 Introduction

Colour is an essential property of our visual world. Apart from its aesthetic and emotional value, it provides valuable information about the environment by breaking the luminance pattern of cast shadows, facilitating the segmentation of objects from each other and the background [1]. To our brain the colour of an object appears largely the same throughout the day, despite dramatic changes in the spectral composition of the light reflected from a scene (e.g. the gamut of physical colours at sunset almost doubles in comparison to the "flattish" midday illumination [2]). This ability (termed *colour constancy*), is more impressive if we consider that mathematically, the problem of separating illumination form reflectance is *ill-posed*, i.e. there are infinite possible solutions.

Although there is no agreement on the precise mechanisms and brain areas responsible for colour constancy, most researchers group them according to the neural level where they likely operate [3]:

1) *Sensory level*: modelled by simple linear transformations of the photoreceptor responses, e.g. scaling responses by their mean activities over the image [4], [5].
2) *Perceptual level*: modelled considering various perceptual "cues" such as specular highlights [6], mutual reflections [7], achromaticity of edges [8], etc. segmenting the image into distinct components (reflections, edges and surfaces) to estimate the illuminant.
3) *Cognitive level*: modelled considering colour memory and/or the identification of objects to be able to compensate for the effects introduced by familiar objects [9].

The relative contributions of each of these processing levels is still a matter for debate. However, most researchers acknowledge that cognitive contributions are likely to be small since the phenomenon can be largely explained by low level mechanisms present in the retina and areas V1 and V4 of the visual cortex [10]. The significance of colour constancy to both human vision and computer vision communities is manifested by the many studies in object detection, tracking, feature extraction, etc. [11], [12], [13] which approach the subject from visual perception and computer vision perspectives, which have historically had different objectives. Most visual perception and neuroscience studies [10], [14], [15] aim at understanding the phenomenon while most computer vision studies [16], [17], [18] aim at predicting the effects of colour constancy. Despite this apparent divergence, one can assume that there might be computational advantages in incorporating the knowledge acquired by the brain's neural machinery after millions of years of evolution. To this end, this finely-tuned combination of low-level (mostly hard-wired) and high-level (learnt) mechanisms might be understood in terms of the *bias/variance* trade-off common in machine learning [19]. The choice of the best bias will depend on the nature of the training data (e.g. how much is known in advance about the problem) and the system's noise. Biological systems face similar choices. A simple organism living in a fix

• *A. Akbarinia and C. A. Parraga are with the Centre de Visió per Computador (CVC) – Universitat Autònoma de Barcelona (UAB), 08193, Spain.*
E-mail: {arash.akbarinia, alejandro.parraga}@cvc.uab.es

Manuscript received August 2, 2016.

environment does not need a strong bias and all individuals can safely share the same neural configuration. Complex organisms such as primates face variable environments and need to dedicate part of their brains to learning during their lifetime while leaving large scale neural structures like the sensory cortex genetically specified. This particular combination of bias/variance in complex organisms allows them to adapt to different environments while still keeping crucial survival skills. In the case of colour constancy, most of the brain computations are arguably done at the sensory level [10] indicating that "bias" may perhaps plays a larger role than "variance" (i.e. more of a *normalisation* problem than a *learning* problem). This is perhaps the reason why current learning-based solutions have trouble to replicate their results in new (non-learned) datasets [10], [20], using dataset-dependent parameters. Additionally, the majority of methods are constrained to consider only one source of illumination, which in effect hinders their applicability in real scenes [18].

1.1 Computational Solutions

From a mathematical point of view, retrieving the colour of a surface illuminated by light of unknown spectral distribution is underdetermined, and to computationally rectify biased images (in the same way colour constancy does) it is common to impose several assumptions regarding the scene illuminant, the statistical distribution of colours or edges, etc. [18]. In general, these algorithms can be divided into two categories: (i) learning-based approaches and (ii) low-level features-driven methods.

Learning-based approaches, e.g. [21], [22], [23], [24], train machine learning techniques on some relevant image features. One group of learning-based algorithms is "gamut mapping", which originated from the influential work of Forsyth [16], and was extended by others [25], [26], [27], [28], [29], following the assumption that only a finite set of colours is observable in real world images. Another large group of algorithms considers reflectance as the random variable of a normal distribution under a Bayesian framework [30], [31], [32]. Although learning-based approaches can obtain accurate results, they rely heavily on their training (i.e. their overall performance depends on the quality of their training data), which is likely to be cumbersome and slow [18].

The majority of low-level features-driven methods can be summarised by the Minkowski framework [8], [33]

$$L_c(p) = \left(\int f_c^p(x) dx \right)^{\frac{1}{p}} = k e_c, \tag{1}$$

where $f(x)$ is the image value at the spatial coordinate x; c is one of the three $\{R, G, B\}$ channels; p is the Minkowski norm; and k is a multiplicative constant chosen such that the illuminant colour, e, is a unit vector.

Substituting $p = 1$ in Eq. 1 reproduces the well known Grey-World assumption, in which the illuminant is estimated by presuming that all colours in the scene average to grey [34]. Setting $p \to \infty$ replicates the White-Patch algorithm, which assumes that the brightest patch in the image corresponds to a specular reflection containing all necessary information about the illuminant [14]. In general, it is challenging to automatically tune p for every image and at the same time inaccurate p values may corrupt the results noticeably [18].

Incorporating high-order image statistics into the Minkowski framework was proposed by van de Weijer et al. [8], under the assumption that the edges carry important information about the source of light, thus their algorithm is called "Grey-Edge". The Minkowski framework can be generalised further by replacing the $f(x)$ in Eq. 1 with its derivative

$$\left| \frac{\vartheta^n f_\sigma(x)}{\vartheta x^n} \right|, \tag{2}$$

where $|.|$ is the Frobenius norm; n is the order of the derivative; and σ is the scale of the Gaussian derivative filters convolved with the original image [35].

It has been noted [36], [37], [38], [39] that high-order derivatives have correspondences with the centre-surround mechanism as modelled in visual perception research. This mechanism is activated when localised sensory regions of the retina are stimulated by light. These sensory regions (also called "receptive fields" or RF) are characterised in terms of their contribution to cortical neurons' stimulation as "centre" and "surround" [40]. The interplay between centre and surround in receptive fields is typically modelled by a Difference-of-Gaussians (DoG) [41], [42], [43], [44]. Since the second order image derivative can be approximated by DoG, they can be a good tool for modelling the cortical and sub-cortical mechanisms involved in colour constancy. This simple model of the low-level properties of the mammalian visual system has a long history starting with Enroth-Cugell and Robson in 1966 [45], continuing with Marr in 1980 [46] and more recently applied to colour constancy by Gao et al. [43]. However, the efficiency of DoG in estimating the illuminant depends on finding an adequate width for the Gaussian kernel, σ, and the optimal weight of the broader Gaussian function, which are difficult to tune automatically. A solution to this problem has already been found by the human visual system (HVS) in the form of dynamic, contrast-based, centre-surround cortical interactions [47], [48] (see below), which are not present in the classical formulations. Although the ultimate purpose of these non-linear interactions is not known, we speculate here that they might play a role in colour constancy and accordingly, we propose a colour constancy model that overcomes the need for ad-hoc or dataset-dependent parameters (and in this sense it is *fully automatic*). In our model we incorporate three well known properties of cortical (area V1) neurons:

1) The size of the minimum RF varies according to the local contrast of the stimuli, i.e. enlarged when exposed to low-contrast [47];
2) The influence of the surround on the centre varies depending on the local contrast of both centre and surround, with greater inhibition for higher contrast stimuli [48];
3) Cortical RFs increase their diameters systematically by approximately a factor of three from lower to higher areas [49], as they pool signals over a large neighbourhood from the levels below.

The above formulation presents major differences with current DoG-based models like that of Gao et al. [43], where

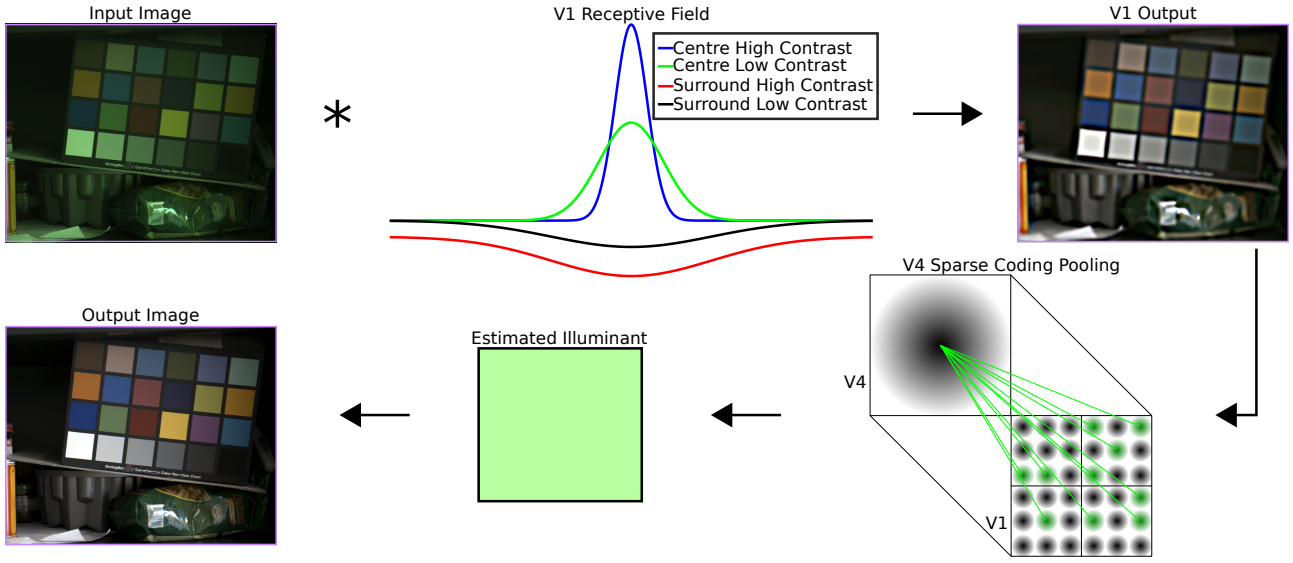

Fig. 1. The flowchart of our model. The input image is convolved by a centre-surround contrast-dependent asymmetric difference-of-Gaussian envelope (inspired by V1 neurons whose receptive fields are large at low contrast and suppressed by high contrast surround). The output of V1 is pooled by V4 neurons according to the sparse-coding principle considering global contrast of image.

the centre size is always constant and the contributions of both centre and surround to the receptive field responses are fixed. Also, the final estimation of luminance was previously based on a simple operation (e.g. selecting the maximum value), whereas we model hypothetical neurons from a higher area (area V4) whose receptive fields are substantially larger than those of V1 neurons, pooling signals from area V1 according to the contrast of the corresponding stimulus. In other words, previous models adopt the *classical* receptive field approach while we go beyond, including the latest physiological findings.

Fig. 1 shows a flowchart of our model. Although a step forward in terms of plausibility, our functional approach still entails an oversimplification of the much more complex (and less well known) interactions between the different neural layers and cortical feedback from higher regions. Following the Occam's razor principle we aimed for the most parsimonious solution that can produce competitive results. We would also like to highlight that we are not strictly interested in out-competing learning-based solutions in each of the testing datasets. Instead we want to produce an algorithm that works like the HVS does, i.e. produces the best possible results in all of the datasets at the same time and *with the same set of parameters*. Equally, we want our solution to be computationally efficient, that is, to incorporate the evolutionary knowledge accumulated by the primate brain in an algorithm potentially implemented in small portable devices. A broader objective of this work is to further understand the role of dynamically-sensitive visual cortical neurons.

In summary, the main contributions of this paper are: (i) the modelling of colour constancy based on physiological findings, i.e. two overlapping asymmetric Gaussian functions whose kernels and weights adapt according to centre-surround contrast, (ii) the estimation of the chromaticity of the light source by modelling higher visual cortical areas (i.e. neurons with large RFs pooling signals from lower areas) according to their local contrast, and (iii) the dynamic generalisation of the colour constancy problem by using the same parameters to predict results in different datasets with no need to "recalibrate", mimicking what the HVS does. Throughout this article we will refer to our model as *Adaptive Surround Modulation (ASM)*

2 Beyond the Classical Receptive Field

In this section we review important physiological findings regarding surround modulation in the visual cortex and describe how we modelled these properties.

2.1 Surround modulation in area V1

The concept of non-classical receptive field (RF) became established by the work of Allman et al. [50] and today numerous studies show that most V1 cells in cat and macaque are suppressed by stimuli extending beyond a critical distance (for a full review refer to [48]).

Quantitative results suggest that RFs in cortical area V1 of macaque change their responses when measured at low contrast [47]. Fig. 2 illustrates the responses of a typical macaque neuron when its RFs are stimulated by a vertically-oriented sinusoidal grating of constant spatial frequency and varying size [48]. The dashed line at the bottom shows the mean spontaneous firing rate of the neuron (no stimulation). The black curve represents the neuron's excitation when stimulated by a high (70%) contrast grating of increasing size (increasing grating radius). As the grating's size increases, more of the neuron's receptive field becomes stimulated producing an increase in the neuron's output, a process known as "facilitation". Its maximum output happens when the grating reaches a radius equal to sRF_{high}. After that, increasing the size of the grating only decreases the neuron's output, i.e. neighbouring neurons

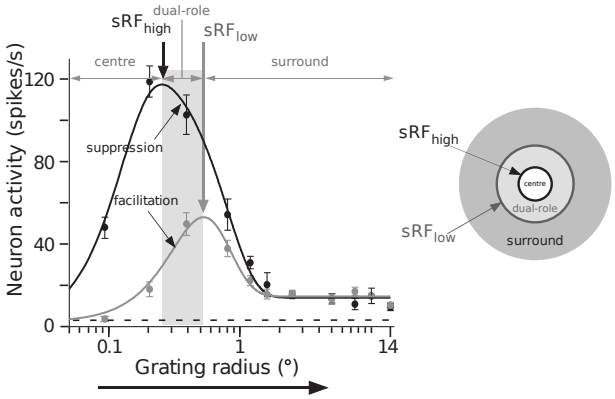

Fig. 2. Size tuning curve of a cell in macaque V1, adapted from [48]. Black and grey curves show responses to a grating of high and low contrast, respectively. The dual-role area is suppressive for high contrast stimuli, whereas it acts as a facilitator in the case of low contrast. The schematic on the right represents the RFs of a V1 neuron. Arrow heads point to radii that determine $sRF_{high}(0.26°)$ and sRF_{low} $(0.54°)$.

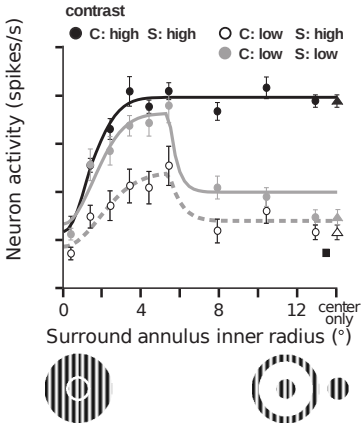

Fig. 3. The influence of surround on the centre, adapted from [48]. Response of a V1 cell in an anaesthetised macaque as a function of the inner radius of the surround annular grating. The triangles are responses to centre-only stimulation. The square indicates response to a surround stimulus of the smallest inner radius presented alone.

start to "suppress" the neuron's activity until it becomes close to zero. Correspondingly, the grey curve in Fig. 2 represents the same neuron's activity as a function of grating size when stimulated by a low (12%) contrast grating. The peak of the grey curve (maximum stimulation radius or sRF_{low}) has now shifted to the right of the plot. The area between the two peaks (shaded in the plot) defines a "dual-role" region, i.e. gratings of radii between these two values can either suppress or stimulate the neuron according to its contrast. The existence of this region implies a fundamental change in the way these visual cortex neurons operate, and we have incorporated it at the core of our model. Now the receptive field of the neuron can be separated in three regions, a "centre" with radius up to sRF_{high}, a "surround" with radius larger than sRF_{low} and a dual-role area in between which operates like the surround (i.e. suppression) when contrast is high and like the centre (i.e. facilitation) when the contrast is low (see right insert in Fig. 2).

Physiological recordings [47] have shown that the radius of the surround in V1 can be about five to six times larger than the value of sRF_{high} and its effects on the centre are significantly more complex than those described above. Fig. 3 illustrates changes in a typical V1 neuron's activity when the stimulation of the centre is fixed and the surround is stimulated by an annuli that becomes increasingly thinner. The plot shows results for three different cases (a) high-contrast is applied to both the centre and the surround; (b) low-contrast is applied to the centre and high-contrast to the surround and (c) low contrast is applied to both the centre and the surround. In all cases, centre-only stimulation (right side of the plot) produces higher neural activity than when both centre and surround are stimulated (left side of the plot). However, suppression is larger for high contrast stimuli (black curve reaches zero when the whole of the surround is stimulated) and is minimal when both centre and surround are stimulated by low contrast gratings (solid grey curve) [51]. In all cases, suppression is strongest when the orientation of centre stimuli is parallel to that of the surround, an effect known as iso-orientation suppression. Although we have only discussed the isotropic (circular RF) case, cortical RFs have preferences for particular spatial orientations. If this is the case, the suppression effect can turn into facilitation as the orientations of the stimuli applied to centre and surround move towards perpendicular directions and the contrast is low. In general, facilitation happens when centre and surround have different characteristics (e.g. different spatial frequency, phase or orientation) and increases when these differences increase.

Cortical RFs systematically increase their diameters by a factor of three along the ventral stream, i.e. the visual pathway specialised in object identification [49]. This is due to the pooling mechanism, which combines signals from the central region as well as neighbouring spatial locations from preceding areas. The whole structure suggests that local visual stimuli is processed by the lower cortical areas and the scope becomes increasingly global as the signal progresses throughout the visual pathway.

2.2 A model of contrast-dependent colour constancy

Surround modulation has been incorporated to biologically-inspired computer vision models with encouraging results, e.g. visual attention [52], saliency [53], tone mapping [54], and boundary detection [55]. However, in the field of computational colour constancy this important physiological finding seems to have been largely overlooked. In this section we investigate the implications of contrast-dependent centre-surround modulation on illuminant estimation by incorporating them into a simple deterministic model.

We recreated a typical RF and its surround using two overlapping asymmetric Gaussian functions which have been reported to adequately fit neuronal responses, e.g. [39], [41], [56]. These functions, referred in our modelling context as the spatially "narrower" and "broader" Gaussians, represent the centre and surround respectively. The width of the narrower Gaussian varies between $[\sigma, 2\sigma]$ and is inversely proportional to the centre contrast. This mimics the changes in size that occur when the centre is exposed to high or low contrast and is similar to incorporating the dual-role region of Fig. 2. Therefore, prior to convolving an image I with a

Gaussian kernel, we compute local contrast C at every pixel through the local standard deviation of I as

$$C_{c,d}(x,y;\sigma) = \sqrt{\left(I_c(x,y) - I_c(x,y) * \mu_d(\sigma)\right)^2 * \mu_d(\sigma)}, \quad (3)$$

where c indexes each colour channel $\{R,G,B\}$; d is the spatial orientation $\{h,v,i\}$ (horizontal, vertical, and isotropic) over which contrast is measured; (x,y) are the spatial coordinates of a pixel; μ is the average kernel with size σ in the direction d and $*$ is the convolution operator. In the case of horizontal contrast, μ is a column vector; in the case of vertical contrast, μ is a row vector; and in the case of isotropic contrast, μ is a square matrix.

It is worth noting that colour processing by post-receptoral visual mechanisms is mostly done in a colour-opponent space [40]. Throughout this work we model $\{R,G,B\}$ colour channels for convenience, since most digital image datasets are in this colour space and we cannot accurately transform to a colour-opponent space without knowing about each manufacturer's camera sensors.

The receptive field's centre response CR is computed by convolution of the original image I at every channel c with the narrower Gaussian as follows:

$$CR_c(x,y) = I_c(x,y) * g_c(x,y;s_{c,h}(x,y),s_{c,v}(x,y)). \quad (4)$$

In Eq. 4, g is the two-dimensional Gaussian kernel:

$$g(x,y;\sigma_h,\sigma_v) = \frac{1}{2\pi\sigma_h\sigma_v}\exp\left(-0.5\left(\frac{x^2}{\sigma_h^2}+\frac{y^2}{\sigma_v^2}\right)\right), \quad (5)$$

where σ_d is the size of the Gaussian kernel in the direction d. The values of $s_{c,h}(x,y)$ and $s_{c,v}(x,y)$ in Eq. 4 represent the vertical and horizontal dimensions of the Gaussian kernel respectively. Since in our formulation the size of the RF's centre is inversely proportional to its local contrast (see Fig. 2), we compute it from the values obtained in Eq. 3:

$$s_{c,d}(x,y) \propto C^{-1}_{c,d}(x,y;\sigma), \quad (6)$$

inversely linking the size of the RF's central kernel to its contrast. In theory, $s_{c,d}$ can be calculated for each individual pixel, however, in practice convolving an image with a unique Gaussian kernel at every pixel is extremely expensive from a computational point of view. For this reason, we approximated $s_{c,d}$ through its uniform quantisation to l different levels, effectively limiting the number of convolutions to l. We computed this uniform quantisation by finding the range of local contrasts through the difference between the two extrema of $s_{c,d}$ and dividing it into an arbitrary number of contrast levels. For example, let's assume that local contrasts are in the range $[0,1]$ and the arbitrary number of contrast levels is 4: pixels with local contrast between $[0.00,0.25]$ are convolved with a Gaussian of 2σ; pixels in the range $(0.25,0.50]$ with a Gaussian of 1.66σ; pixels in the range $(0.50,0.75]$ with a Gaussian of 1.33σ; and pixels in the range $(0.75,1.00]$ with a Gaussian of σ.

To summarise, we calculated the centre response CR by convolving low contrast image pixels with large Gaussians and high contrast image pixels with small Gaussians. It is worth noting that σ_h and σ_v in Eq. 5 are not identical (a common assumption in computer vision) due to the fact that the local interactions in V1 are not always organised in a symmetric fashion [57].

The surround response, SR, was computed by convolution of the original image in every $\{R,G,B\}$ channel with the broader symmetric Gaussian kernel

$$SR_c(x,y) = I_c(x,y) * g_c(x,y;5\sigma,5\sigma), \quad (7)$$

where kernel size is constant in both directions regardless of local contrast. The decision to keep the size of the SR kernel fixed was made after considering the much smaller variations that occur in the surround RFs of neurons under different contrast levels [47].

The final RF response RR, was computed by combining centre and surround modulations as follows:

$$RR_c(x,y) = \lambda_c(x,y)CR_c(x,y) + \kappa_c(x,y)SR_c(x,y), \quad (8)$$

where λ and κ are the weights of centre and surround in each spatial location. These parameters model the fact that the strength of centre response and surround suppression depend of the contrast and relative orientations of the centre and surround stimuli (see Fig. 3 and the work of Shushruth et al. [47]). We modelled λ and κ as inversely proportional to the oriented contrast of centre and surround respectively, which was computed as

$$\begin{aligned} \lambda_c(x,y) &\propto C^{-1}_{c,i}(x,y;\sigma); \\ \kappa_c(x,y) &\propto C^{-1}_{c,i}(x,y;5\sigma), \end{aligned} \quad (9)$$

where i denotes the spatial direction. We modelled the fact that suppression can turn into facilitation when the centre is exposed to low contrast or when centre and surround stimuli are orthogonal from each other [48]. This can be done by allowing the sign of κ to change from minus (suppressive surround) to the occasional plus (facilitatory surround) transforming our model from a DoG to Sum-of-Gaussians (SoG). Although the model allows the possibility of a positive κ, we should note that the boundary between suppression and facilitation is cell specific and there is no universal contrast level or surround stimulus size that triggers facilitation across the entire cell population [48]. Due to this, and the fact that numerical surround suppression figures in macaque V1 neurons were reported to be all negative [47], the results we present in this paper were all obtained with a negative κ value.

Up to this point we implemented a model of RR based on well known properties of V1 neurons. In the next processing stage, the visual signal is pooled and sent to higher cortical areas whose exact location is unknown. Many authors [38], [58] have proposed area V4 as the most likely candidate for a colour constancy site. We hypothesised the existence of V4 neurons that perform operations on the outputs of those in V1. From the physiology, we know that cortical RFs increase their diameter systematically by approximately a factor of three from lower to higher areas [49]. This means that V4 RFs are about nine times larger than those in V1 (which is $0.26°$, see Fig. 2) thus, the centre and surround of a typical V4 RF subtend about $2.3°$ and $11.7°$ of visual angle respectively, which are equivalent to 117 and 585 pixels on a standard monitor viewed from a $100cm$ distance.

The exact pooling mechanism applied to these V1 signals is unknown, however "winner-takes-all" and "sparse coding" kurtotical behaviour are common to large groups

of neurons all over the visual cortex [59], [60] and it is not infeasible to assume that a small group of neurons with the largest activation dominate most of the process. We approximated this hypothetical behaviour of V4 neurons by selecting a small percentage of "winner neurons" whose RFs are highly activated. To simulate contrast adaptation behaviour in our hypothetical V4 neurons similar to those in V1, we inversely linked the percentage of pooled signals to the variability of the signal collected by their receptive field. In other words, when the "contrast" applied to V4 RF is high, a smaller percentage of signals from V1 is pooled and vice versa. As before, contrast was calculated as the local standard deviation of the input. Fig. 1 summarises the whole feedforward process in a flowchart. The first stage of the model simulates the operation of the typical V1 neuron with contrast-dependent RFs and the second stage simulates the V4 sparse-coding pooling of a small percentage of highly activated V1 neurons.

In practice RR (V1 output) is an image composed by three chromatic channels RR_c. We implemented the "winner-takes-all" behaviour via a histogram-based clipping mechanism [17], [61] as follows. Let H_c be the histogram of RR_c values obtained by applying Eq. 8 to colour channel c of the input image. In this histogram, the neural response of the cells contained in an individual bin b is represented by $RR_c(b_c)$. We estimate the scene illuminant by computing

$$L_c = RR_c(b_c), \tag{10}$$

with b_c chosen so that only the most activated ("winner") units contribute to the pooling (sum). To calculate b_c we start by estimating the average local contrast of all inputs to V4 in a given colour channel c using

$$p_c = \frac{1}{n}\sum_{x,y} F_c(x,y), \tag{11}$$

where F is the standard deviation of the pixels of RR_c computed using the average V4 neuron receptive field (nine times larger than that of a V1 neuron), i.e. $F_c(x,y) \approx C_{c,i}(x,y;9\sigma)$. Bear in mind that "contras" is just a fraction in the range $[0,1]$. Instead of choosing a fix percentage of neurons with the largest activation for each colour channel (as in [61]), we chose an adaptive activation level such that all neurons with activations higher than the one chosen account for fraction p_c of the total number of pixels. In other words, we computed b_c as the threshold activation level that defines a number of highly activated neurons equal to the contrast value calculated in Eq. 11 as follows:

$$p_c n \leq \sum_{k=b_c}^{n_b} H_c(k) \quad \text{and} \quad p_c n \geq \sum_{k=b_c+1}^{n_b} H_c(k), \tag{12}$$

where n is the total number of RR_c response units and n_b is the total number of bins in histogram H_c. This effectively links the number of highly activated neurons we use to compute the scene illuminant to the average contrast of the input to area V4.

We illustrated this mechanism of V4 pooling in Fig. 4, where RR_c is represented by the red, green and blue signals corresponding to each chromatic channel. Dashed vertical lines show b_c, i.e. cells (bins) on the right side of these lines are pooled by V4 and their sum for each colour channel is the estimated illuminant. In this example contrast is higher for the red signal and therefore a smaller percentage of cells are pooled in the red channel.

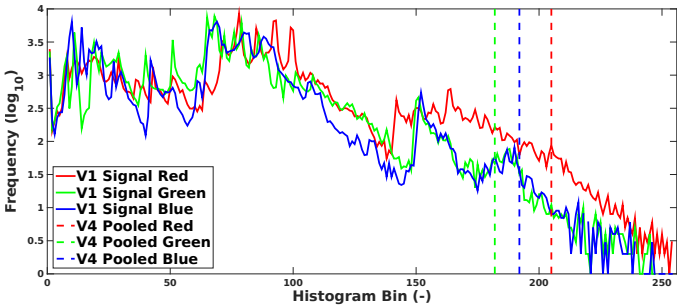

Fig. 4. V4 "winner-takes-all" mechanism. Each colour depicts its chromatic channel. Straight lines show which portion of V1 signals is pooled into V4. The ordinates are shown as log-axes due to the large variations in the counts of the different bins.

Mathematically, there is a direct relation between the fraction of "winner" pixels, p in Eq. 11, and value of the Minkowski norm in Eq. 1. When the fraction of "winner" pixels is equal to unity (i.e. 100% pooling) computations of Eq. 10 includes the responses of all V1 neurons, resembling the Grey-World assumption in which the exponential term of the Minkowski sum in Eq. 1 is equal to unity. Correspondingly, when the percentage of "winner" pixels tends to zero, only the most activated V1 response is pooled, resembling the White-Patch algorithm.

3 Experiments and Results

The issue of observer's performance evaluation in colour constancy tasks using naturalistic stimuli is still an open problem [10], [72]. In the case of algorithms, popular measures consist of some kind of angular distance in chromatic space between the estimated illuminant and that of the ground truth. Although intuitively simple, psychophysical experiments have shown that these error measures do not always correspond to observer preferences [73]. Despite this, angular errors are still the most convenient way to compare algorithms and their use in the literature is widespread, being perhaps the most common the *recovery angular error* defined as

$$\epsilon^{\circ}_{recovery}\left(e_e, e_t\right) = cos^{-1}\left(\frac{e_e.e_t}{\|e_e\|.\|e_t\|}\right), \tag{13}$$

where $e_e.e_t$ is the dot product of the estimated illuminant e_e and the ground truth e_t, and $\|.\|$ represents the Euclidean norm of a vector. This simple measure has recently been the subject of criticism from Finlayson et al. [74] since it arguably produces different recovery errors for identical scenes viewed under two different coloured illuminants. For this reason, they proposed an improved version (termed *reproduction angular error*):

$$\epsilon^{\circ}_{reproduction}\left(e_e, e_t\right) = cos^{-1}\left(\frac{(e_t/e_e)}{\|e_t/e_e\|}.w\right), \tag{14}$$

where $w = \frac{e_t/e_t}{\sqrt{3}}$ is the true colour of the white reference.

To compare our results with those of state-of-the-art algorithms, we present the mean, median and trimean of both recovery and reproduction angular errors. The later two measures are considered to be more appropriate to assess the performance of colour constancy algorithms, because of their robustness to outliers [75], [76].

We evaluated our method on four benchmark datasets[1] without adjusting free parameters since ASM is automatic (i.e. dataset-independent) in contrast to most other algorithms whose results were acquired after adjusting their parameters to the optimum value for each dataset. Additionally, in order to better understand the contribution of the different components of our model, we conducted three extra experiments, which are explained later in this section.

3.1 Single-illuminant scenes

We tested our model on three single-illuminant benchmark datasets, (i) SFU Lab [62], (ii) Colour Checker [77], and (iii) Grey Ball [78]. Our results for single-illuminant scenes were obtained under four contrast levels, $l = 4$, with $\sigma = 1.5$. This σ is equivalent to 13 pixels or 0.26° of visual angle when viewed from $100cm$ in a standard monitor, which is also the size of sRF_{high} (see Fig. 2). We set the range of surround suppression to $\kappa = -[0.67, 0.77]$, considering the surround suppression index of macaque V1 neurons reported at [47]. The centre weight was retrieved directly from the contrast of pixels, $\lambda_c(x, y) = 1 + C^{-1}_{c,i}(x, y; \sigma)$.

3.1.1 SFU Lab

The SFU Lab dataset [62] consists of 321 images of size 637×468 captured in a controlled environment under eleven different sources of light. The scenes are partitioned into four categories: (a) minimal specularities, (b) non-negligible dielectric specularities, (c) metallic specularities, and (d) at least one fluorescent surface. We report the results of our method and several others on this dataset in Table 1. Our model's results show a clear improvement in the median and trimean angular errors (both reproduction and recovery) compared to state-of-the-art for the SFU Lab dataset.

3.1.2 Colour Checker

The Colour Checker dataset [32], [77] consists of 568 indoor and outdoor images of size 2041×1359. Each image contains a MacBeth colour-checker as a reference to retrieve the chromaticity of the actual source of light. We followed the best practices and guidelines of this dataset by masking out all colour-checker boards prior to processing an image with our model. The original images are non-linear due to gamma and tone curve correction. Shi and Funt [77] reprocessed the raw data and generated 12-bit images. Our results (see Table 2) show that our model is in par with the state-of-the-art for this dataset.

3.1.3 Grey Ball

The Grey Ball dataset [78] consists of 11346 non-linear images of size 360×240 extracted from two hours of video recorded under a large variety of conditions in both indoor and outdoor environments. In every image there is a grey sphere at the bottom right corner from which the ambient illuminant can be estimated. We also masked out all grey spheres prior to processing these images with our model. Our results (see Table 3) suggest that our model is in par with the learning-based state-of-the-art for this dataset, while it outperforms all other low-level features-driven methods.

3.2 Testing the role of each model component

We studied the contribution of each component (i.e., adaptive centre, dynamic surround and p estimation) by con-

1. All source code and experimental materials are available under this link https://goo.gl/nQUenN.

TABLE 1
Angular error of several methods on SFU Lab [62] benchmark dataset. Lower figures indicate better performance.

			Recovery Error			Reproduction Error		
	Method		Mean	Median	Trimean	Mean	Median	Trimean
	Do Nothing		17.3	15.6	16.9	17.3	15.6	16.9
Low-level features	Inverse-Intensity Chromaticity Space	[63]	15.5	8.2	10.7	15.1	9.3	11.5
	Grey-World	[34]	9.8	7.0	7.6	10.1	7.5	8.3
	White-Patch	[14]	9.1	6.5	7.5	9.7	7.4	8.2
	Shades of Grey	[33]	6.4	3.7	4.6	6.9	3.9	4.8
	General Grey-World	[33]	5.4	3.3	3.8	6.0	3.9	4.3
	First-order Grey-Edge	[8]	5.6	3.2	3.7	6.3	3.6	4.2
	Second-order Grey-Edge	[8]	5.2	2.7	3.3	5.8	3.0	3.8
	Local Surface Reflectance Statistics	[64]	5.7	2.4	-	-	-	-
	Random Sample Consensus	[65]	-	-	-	-	-	-
	Edge-based Grey Pixel	[66]	5.3	2.3	-	-	-	-
	Double-Opponency	[43]	4.8	2.4	3.5	-	-	-
Learning-based	Pixel-based Gamut Mapping	[16]	3.7	2.3	2.5	4.2	2.8	3.0
	Edge-based Gamut Mapping	[29]	3.9	2.3	2.7	4.5	2.7	3.2
	Spectral Statistics	[67]	5.6	3.5	4.3	-	-	-
	Weighted Grey-Edge	[68]	5.6	2.4	2.9	6.1	3.6	4.3
	Regression	[22]	-	2.2	-	-	-	-
	Thin-plate Spline Interpolation	[24]	-	2.4	-	-	-	-
	Bayesian	[32]	-	-	-	-	-	-
	Natural Image Statistics	[18]	-	-	-	-	-	-
	Exemplar-based method	[69]	-	-	-	-	-	-
	CNN Fine Tuned	[70]	-	-	-	-	-	-
	Deep Learning Colour Constancy	[71]	-	-	-	-	-	-
	ASM		**4.7**	**1.8**	**2.3**	**5.2**	**2.3**	**2.7**

TABLE 2
Angular error of several methods on Colour Checker [77] benchmark dataset. Lower figures indicate better performance.

			Recovery Error			Reproduction Error		
	Method		Mean	Median	Trimean	Mean	Median	Trimean
	Do Nothing		13.7	13.6	13.5	13.7	13.6	13.5
Low-level features	Inverse-Intensity Chromaticity Space	[63]	13.6	13.6	13.5	14.3	13.6	13.6
	Grey-World	[34]	6.4	6.3	6.3	7.0	6.8	6.9
	White-Patch	[14]	7.5	5.7	6.4	8.1	6.5	7.1
	Shades of Grey	[33]	4.9	4.0	4.2	5.8	4.4	4.9
	General Grey-World	[33]	4.7	3.5	3.8	5.3	4.0	4.4
	First-order Grey-Edge	[8]	5.3	4.5	4.7	6.4	4.9	5.3
	Second-order Grey-Edge	[8]	5.1	4.4	4.6	6.0	4.8	5.2
	Local Surface Reflectance Statistics	[64]	3.4	2.6	-	-	-	-
	Random Sample Consensus	[65]	3.2	2.3	-	-	-	-
	Edge-based Grey Pixel	[66]	4.6	3.1	-	-	-	-
	Double-Opponency	[43]	4.0	2.6	-	-	-	-
Learning-based	Pixel-based Gamut Mapping	[16]	4.2	2.3	2.9	4.8	2.7	3.4
	Edge-based Gamut Mapping	[29]	6.5	5.0	5.4	8.0	5.9	6.6
	Spectral Statistics	[67]	3.7	3.0	3.1	-	-	-
	Weighted Grey-Edge	[68]	-	-	-	-	-	-
	Regression	[22]	8.1	6.7	7.2	8.8	7.4	7.9
	Thin-plate Spline Interpolation	[24]	-	2.8	-	-	-	-
	Bayesian	[32]	4.8	3.5	3.9	5.6	3.9	4.4
	Natural Image Statistics	[18]	4.2	3.1	3.5	4.8	3.5	3.9
	Exemplar-based method	[69]	2.9	2.3	2.4	3.4	2.6	2.9
	CNN Fine Tuned	[70]	2.6	2.0	-	-	-	-
	Deep Learning Colour Constancy	[71]	3.1	2.3	-	-	-	-
	ASM		**3.8**	**2.4**	**2.7**	**4.9**	**3.0**	**3.4**

TABLE 3
Angular error of several methods on Grey Ball [78] benchmark dataset. Lower figures indicate better performance.

			Recovery Error			Reproduction Error		
	Method		Mean	Median	Trimean	Mean	Median	Trimean
	Do Nothing		8.3	6.7	7.2	8.3	6.7	7.2
Low-level features	Inverse-Intensity Chromaticity Space	[63]	6.6	5.6	5.8	7.0	6.0	6.2
	Grey-World	[34]	7.9	7.0	7.1	8.7	7.6	7.9
	White-Patch	[14]	6.8	5.3	5.8	7.1	5.5	6.0
	Shades of Grey	[33]	6.1	5.3	5.5	6.5	5.6	5.8
	General Grey-World	[33]	6.1	5.3	5.5	7.1	6.2	6.4
	First-order Grey-Edge	[8]	5.9	4.7	5.1	6.3	4.8	5.4
	Second-order Grey-Edge	[8]	6.1	4.8	5.3	6.5	5.0	5.6
	Local Surface Reflectance Statistics	[64]	6.0	5.1	-	-	-	-
	Random Sample Consensus	[65]	-	-	-	-	-	-
	Edge-based Grey Pixel	[66]	6.1	4.6	-	-	-	-
	Double-Opponency	[43]	-	-	-	-	-	-
Learning-based	Pixel-based Gamut Mapping	[16]	7.1	5.8	6.1	7.5	5.9	6.3
	Edge-based Gamut Mapping	[29]	6.8	5.8	6.0	7.3	5.8	6.3
	Spectral Statistics	[67]	10.3	8.9	9.1	-	-	-
	Weighted Grey-Edge	[68]	-	-	-	-	-	-
	Regression	[22]	-	-	-	-	-	-
	Thin-plate Spline Interpolation	[24]	-	-	-	-	-	-
	Bayesian	[32]	-	-	-	-	-	-
	Natural Image Statistics	[18]	5.2	3.9	4.3	5.5	4.3	4.7
	Exemplar-based method	[69]	4.4	3.4	3.7	4.8	3.7	4.0
	CNN Fine Tuned	[70]	-	-	-	-	-	-
	Deep Learning Colour Constancy	[71]	4.8	3.7	-	-	-	-
	ASM		**4.7**	**3.8**	**4.0**	**5.0**	**4.1**	**4.3**

ducting three experiments and analysing their results in terms of median and trimean angular errors, which were proposed by [75] and [76] as robust measures to evaluate colour constancy algorithms.

3.2.1 Experiment 1 – constant vs. adaptive centre size

In order to measure the contribution of the adaptive size of the narrower Gaussian, we kept all other parameters fixed (i.e. the centre-surround influence, $\lambda = 1.00; \kappa = -0.77$, and the percentage of pooled signal, $p = \infty$). We tested two scenarios: (a) all pixels were convolved with a constant Gaussian of width σ (essentially the Double-Opponency algorithm [43]) whereas, in (b) this width was varied within the range $[\sigma, 2\sigma]$ and computed for each pixel. These two conditions were called "Constant Gaussian Width" (CGW) and "Adaptive Gaussian Width" (AGW). Additionally, since the Grey-Edge hypothesis captures high-order image features similar to the DoG, we tested whether this centre adaptation can improve the first and second order Grey-Edge algorithm with a Minkowski norm $p = \infty$.

The results of experiment 1 (see Fig. 5) show that both measures of median and trimean errors are always smaller

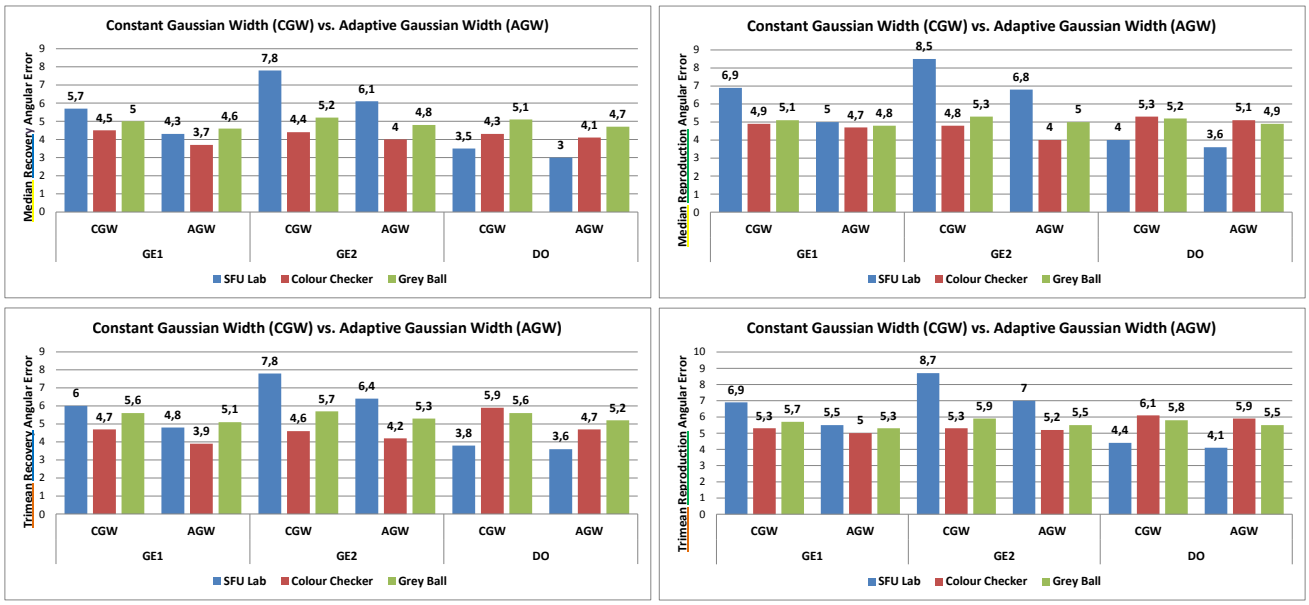

Fig. 5. Influence of contrast-dependent RF size on illuminant estimation.

in the adaptive case (AGW) than in the constant case (CGW). This is true for both recovery and reproduction angular errors. The largest and smallest improvements are achieved in the SFU Lab (about 19% on average) and Grey Ball (about 6% in average) datasets, respectively.

3.2.2 *Experiment 2 – constant vs. adaptive surround*

In order to measure contribution of the adaptive surround modulation, we kept all other parameters fixed (i.e. the centre adaptation, $l = 1$, and the percentage of pooled signal, $p = \infty$). We tested three scenarios, the first and second were computed under a constant surround influence, $\kappa = -0.67$ and $\kappa = -0.77$, respectively (both extrema of our adaptive κ), as well as constant centre weight, $\lambda = 1.00$. In the third scenario, the centre-surround influence was adaptive, $\lambda = 1 + C_{c,i}^{-1}(x, y; \sigma)$ and $\kappa = -[0.67, 0.77]$, under four contrast levels $l = 4$.

Fig. 6 shows the results for Experiment 2, where the median and trimean errors (both recovery and reproduction) obtained with a dynamic surround suppression, $\kappa = -[0.67, 0.77]$, are always lower in comparison to the constant κ. The gain across datasets appear to be similar (around 3% for both error measures).

3.2.3 *Experiment 3 – constant vs. adaptive p estimation*

In order to measure the contribution of the adaptive clipping, we examined five different scenarios. In the first four, histograms (see Eq. 12) were clipped with constant percentages, $p = \{5, 1, 0.5, 0.1\}\%$, i.e. a fixed set of V1 cells were

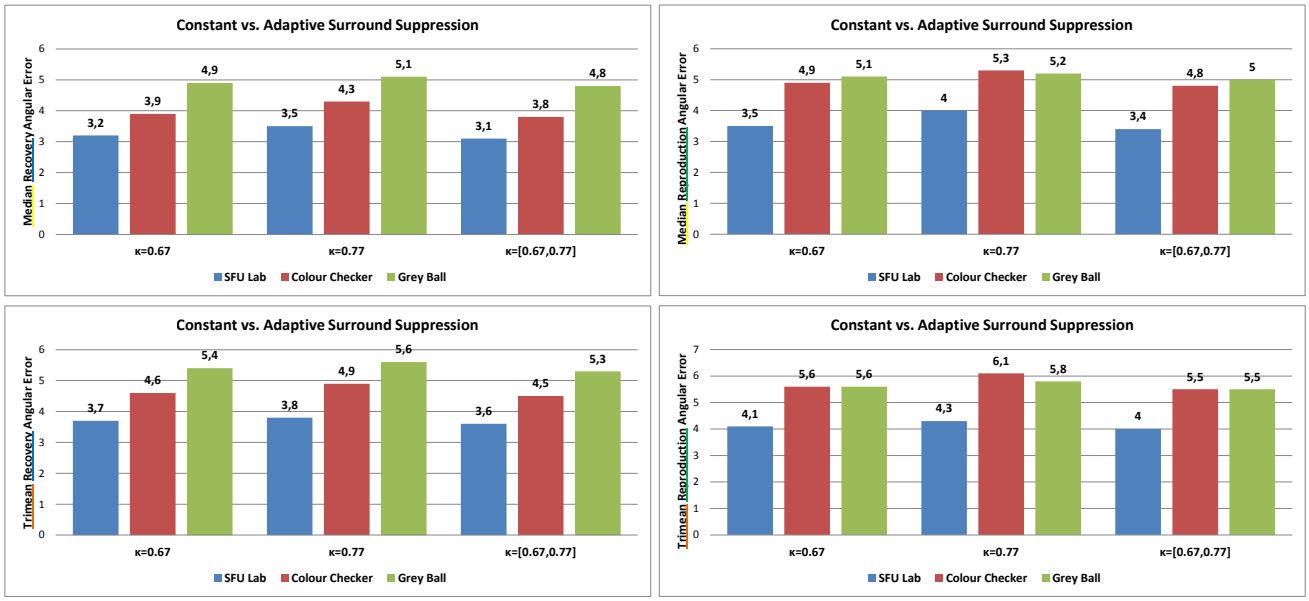

Fig. 6. Influence of contrast-dependent surround suppression on illuminant estimation.

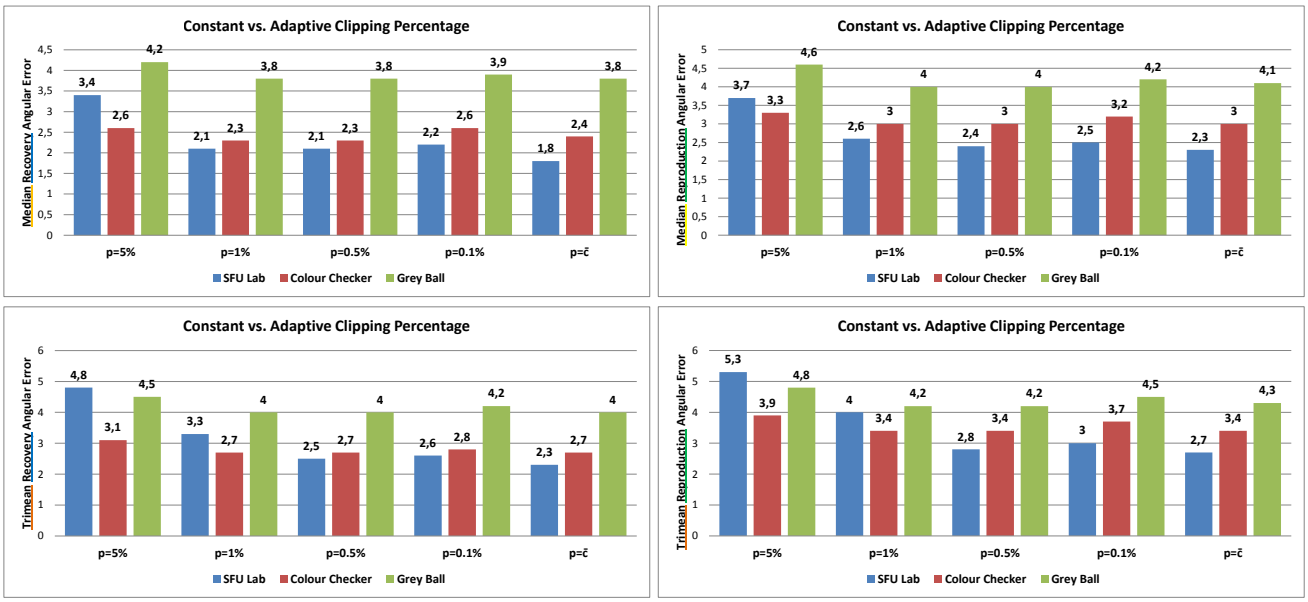

Fig. 7. Influence of "winners" percentage p on illuminant estimation.

pooled into V4. In the fifth case, value of p was adaptive and computed as the average contrast of RR (see Eq. 11).

Fig. 7 shows results of this experiment suggesting that using a contrast-dependent pooling mechanism reduces the recovery/reproduction angular errors in all cases considered in the SFU Lab dataset (blue bar with $p = \bar{c}$ is smaller than all the others). In the Colour Checker and Grey Ball datasets (red and green bars respectively), estimating p adaptively yields angular errors very close to the best constant p values. Among the constant clipping percentages $p = 0.5\%$ performs best: moving towards a Grey-World pooling deteriorates the results ($p = 5\%$ obtain the highest angular errors) and moving towards a White-Patch solution also worsens angular errors ($p = 0.5\%$ always performs better than $p = 0.1\%$). This suggests that the optimal pooling mechanism is close to our proposal of pooling a set of highly activated cells. A comparison between the best fixed p ($= 0.5\%$) and adaptive p ($= \bar{c}$) shows a 4% improvement of median and trimean errors (average of all three datasets) in the case of adaptive p.

3.3 Multi-illuminant scenes

We tested our model on a multi-illuminant benchmark dataset [79] consisting of 78 images, where each image was captured under the illumination of two different light sources. The dataset contains two sets of images: (a) laboratory (58 images of size 452×260) and (b) real-world images (20 images of size 452×302).

The extension of our model to multi-illuminant scenes is done straightforward by modelling each region or pixel with a similar contrast-dependent pooling mechanism (Eq. 10, 11, 12 will be region or pixel dependent). This solution is biologically-plausible as different V4 neurons pool signals from different V1 neurons. For this multi-illuminant dataset we used the exact same parameters as single-illuminant datasets (refer to Section 3.1). Here we defined four simple image regions (by halving the image in both horizontal and vertical directions) and computed the source of light in each region accordingly. These results are reported alongside several others in Table 4. Since other methods have not reported their respective trimean and reproduction angular errors in this dataset, we only report the mean and median recovery angular error. Our results are competitive with the state-of-the-art.

TABLE 4
Recovery angular error of several methods on Multi-illuminant [79] benchmark dataset. Lower figures indicate better performance.

		Laboratory		Real-world	
Method		Mean	Median	Mean	Median
Do Nothing		10.6	10.5	8.9	8.8
Grey-World	[34]	3.2	2.9	5.2	4.2
White-Patch	[14]	7.8	7.6	6.8	5.6
First-order Grey-Edge	[8]	3.1	2.8	5.3	3.9
Second-order Grey-Edge	[8]	3.2	2.9	6.0	4.7
Gijsenij et al.	[80]	4.8	4.2	4.2	3.8
Double-Opponency	[43]	4.6	4.4	7.8	4.9
STD-based Grey Pixel	[66]	2.9	2.2	5.7	3.5
MI Random Field	[79]	2.6	2.6	4.1	3.3
ASM		**2.7**	**2.5**	**5.1**	**3.5**

4 Discussion

Fig. 8 illustrates the results of our *Adaptive Surround Modulation (ASM)* model alongside three other algorithms on four exemplary images that were captured under different illuminations: "synthetic indoor", "natural daylight", "dim evening", and "multi-illuminant" (one from each of the benchmark datasets). This qualitative analysis show that ASM can efficiently estimate the source of light present in synthetic and natural images, bright and dark environments, and in both single- and multi-illuminant scenes. This agrees with the quantitative results presented in Section 3. ASM surpasses all other state-of-the-art algorithms in the SFU Lab dataset (Table 1). In the Grey Ball dataset ASM performs the best amid methods driven by low-level features and obtains comparable results to the learning-based

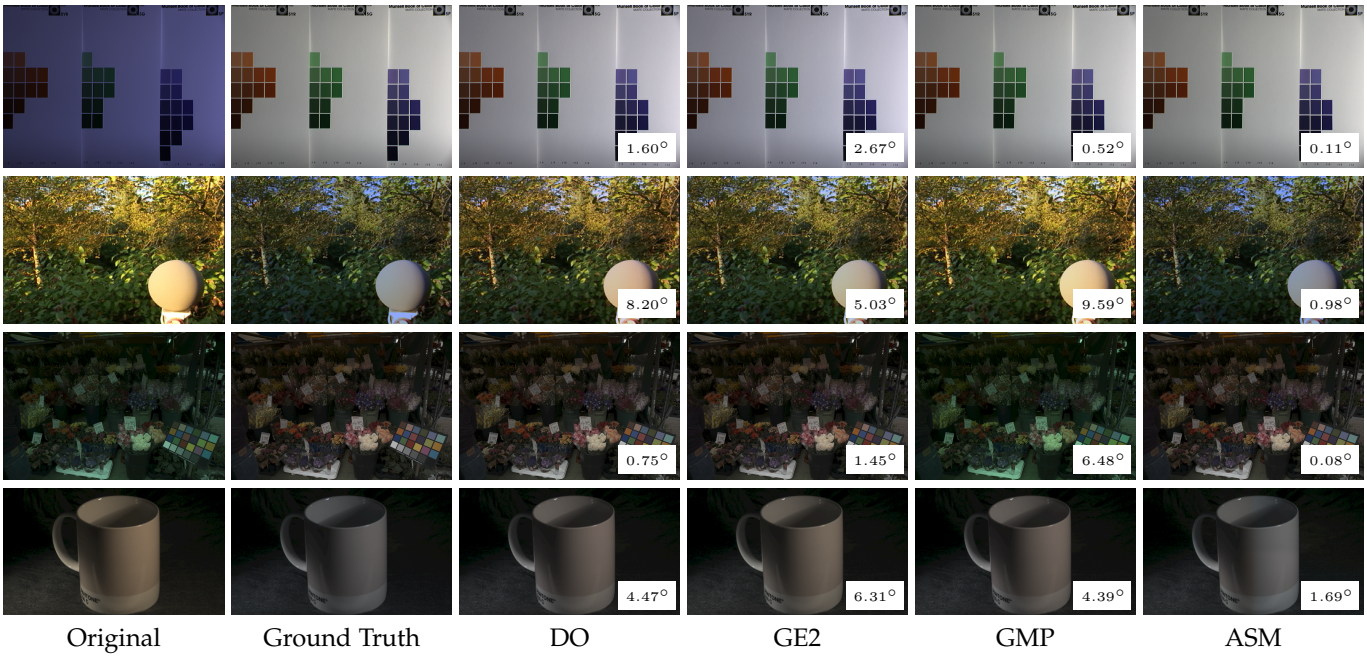

Fig. 8. Colour constancy results of several methods. The recovery angular error is indicated on the right bottom corner. The first row shows results for a picture from the SFU Lab dataset, the second row from the Grey Ball dataset, the third row from the Colour Checker dataset, and the last row from the Multi-illuminant dataset.

techniques (Table 3). Similarly, in the Colour Checker and Multi-illuminant datasets the angular errors obtained by our model are highly competitive with the best of the state-of-the-art (Tables 2 and 4, respectively). This is despite using a fix set of parameters for all four datasets, unlike the other methods.

A comparison among Tables 1–3 and Fig. 5 suggests that methods driven by higher-order image statistics (e.g. Grey-Edge and Double-Opponency) are highly sensitive to their choice of parameters. For example, in the SFU Lab dataset, the median recovery angular error of the second order Grey-Edge (GE2) escalates from 2.7° (Table 1) to 7.8° (Fig. 5) under the optimum ($p = 7, \sigma = 4$) and non-optimum parameters ($p = 1, \sigma = 1$), respectively. This could be because the role of contrast is not taken fully into consideration by those algorithms. Indeed a number of physiological and psychophysical studies have shown that the processing of contrast and colour are thoroughly intertwined [3], [10], [58]. In particular, it has been shown that colour constancy is significantly influenced by the variance of surrounding regions [81], a concept solidly incorporated into ASM where the model's output is dependent on both, the variance and the average colour of the surround. This insight is also supported by results of our experiments.

The results of experiment 1 (see Fig. 5) show that the performance of colour constancy methods driven by high-order image statistics (e.g. Grey-Edge and Double-Opponency) can be improved, as much as 21%, by adapting their Gaussian width σ based on local contrast at pixel level. This improvement, which comes from the *Adaptive Gaussian Width* (AGW), appears to be larger for the Grey-Edge (about 13% in average) than for the Double-Opponency (about 7% in average). In the Grey-Edge, the centre-surround mechanism is modelled in one operation, therefore the AGW implicitly influences both centre and surround. However, in the AGW version of Double-Opponency neither the surround size nor its contribution to the centre changes according to the contrast level.

Experiment 2 (see Fig. 6) shows that contrast-dependent surround modulation can improve the angular errors up to 15%, however the average improvement is a more modest figure of about 3%. This is explained by the fact that surround modulation depends on a number of other parameters in addition to the local contrast of stimuli, such as spatial frequency and orientation. In this work, we restricted our studies to the role of contrast on surround modulation and therefore the range of surround suppression we could explore was rather limited to $\kappa = -[0.67, 0.77]$. However, we believe that our results can be improved even further by taking into account the orientation selectivity of surround suppression and consequently allowing a larger range of κ values. In this way ASM can oscillate between DoG and SoG to account for both surround inhibition and facilitation. This can be achieved for example by wavelet decomposition, which we propose as future work. Such pyramids of wavelets have been successfully used to model the operation of neurons in the visual cortex in the case of contrast induction [82] and saliency [53].

The results of experiment 3 (see Fig. 7) indicate that our "winner-takes-all" hypothesis appears to be correct as well. The lowest angular errors are obtained when only a small percentage of V1 signals are pooled into V4 and when this percentage is high (5%) the results deteriorate significantly. However, there is no unique p to minimise the angular errors across different datasets for both measures of median and trimean. Determining the "winners" according to the average contrast of V1 RFs ($p = \bar{c}$) produces the lowest angular errors across datasets. This is in line with findings of [83] that reported an inverse relationship between the contrast of stimulus and the percentage of the pooled signal

in cat visual cortex. Conceptually, in a low contrast image a few bright pixels can hint of the source of light, whereas in a high contrast image (i.e. with high variation of pixel values) more samples are required to determine the scene illuminant. This is in line with the results of Joze et al. [84], which indicate that bright pixels play a vital role in illuminant estimation. A better estimation of p might be obtained by a more thorough modelling of V4 neurons (for example by calculating p in different image regions, rather than the entire population of V1 neuron).

4.1 Why does ASM works?

The previous results show that ASM works because of a combination of all its parts: in some cases the contrast-dependent centre-surround mechanism provides most of the improvement, in others the contrast-dependent pooling and in most cases both. In general, implementing a contrast-dependent RF always reduces angular errors, even if this reduction is minimal. Conceptually, our operator is intuitive: on homogeneous regions a larger window is applied to represent surround variation, whereas on heterogeneous regions a small neighbourhood suffices. A second stage collects the brightest pixels also in a dynamical, contrast-dependent manner and use them to calculate the chromaticity of the illuminant. This is particularly advantageous since the most important cues about the illuminant are given by (spectrally flat) bright objects and specular highlights.

The operation of the model can be fully appreciated by examining its flowchart in Fig. 1, where the inset image labelled "V1 Output" shows the results of processing a typical scene with our contrast-dependent kernel. There, the coloured surfaces of the Macbeth chart have been smoothed, except near the edges, where they have been enhanced (a similar effect can be observed in Mondrians of the "Adaptive V1" image in Fig. 9). The contrast-dependent pooling stage of the model collects the largest pixel values (e.g. those near the edges of the white patch) and estimate the illuminant from them. Another source of illuminant information is the strong specular highlight situated below the Macbeth chart in Fig. 1. Pixels within this (high contrast) highlight are convolved with a small operator that enhances them while diminishing the effect of clipped values at the same time. This dual effect (smoothing over potentially uninformative and noisy areas while enhancing informative ones) is achieved by linking the size of the operator to local contrast.

Similar types of contrast-dependent surround modulation have been shown to boost edges while suppressing undesired textural information [85]. Theoretically, our variations of the Gaussian kernel width, σ, are resemblant of processing an image through a Gaussian pyramid (although not of fixed one-octave log increments in size like those found in the cortex). Correspondingly, the overall effect of our variation in the surround weight in the DoG operator, κ, resembles a Laplacian pyramid. Finally, our variation in the percentage of pooled signal, p, correlate with finding an optimum Minkowski norm for each image.

To mitigate the influence of higher-level visual cues, we tested our algorithm with the exact same parameters on 1000 randomly generated Mondrian images under random illuminants. Median reproduction angular error of adaptive surround modulation (our full model) was 2.3; the same measure for our model in its constant-size kernel form (i.e. no contrast-dependent V1 and V4 neurons, similar to Gao et al. [43]) was 3.8. In more than 77% of the images, ASM obtained better results compared to the constant-surround variant. We illustrated one example of this experiment in Fig. 9. If we compare "Constant V1" to "Adaptive V1", we can observe that constant-surround modulation indiscriminately blurs the entire image, whereas the adaptive-surround formulation allows for sharper edges. Similarly, in case of "Constant V4" the estimated illuminant is significantly greener than the actual illuminant and therefore the corrected image appears reddish.

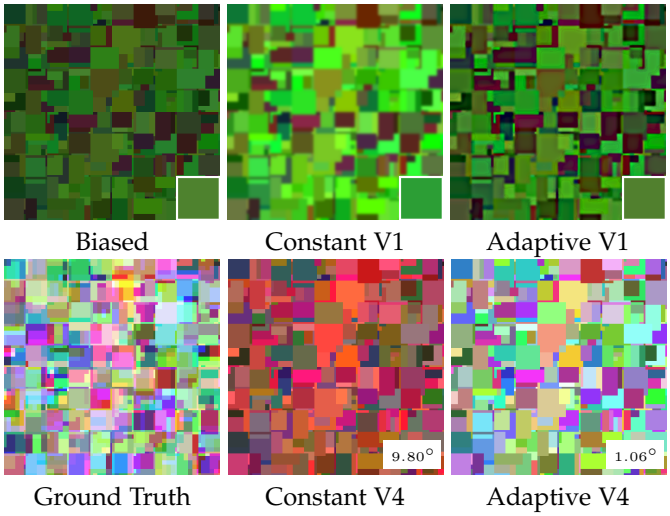

Fig. 9. Constant versus adaptive V1 and V4 modules. Colour patches on the right bottom corner of images in the first row depict the ground truth illuminant (in the case of the biased image) and estimated illuminants (for the constant and adaptive results).

Computationally, ASM is very efficient as no training is required. The complexity of our algorithm is l (number of contrast levels, 4 in this article) times more expensive than a simple DoG. However, each level is 100% independent and their convolutions can easily run in parallel, as it is implemented in our source code.

5 Conclusion

Local contrast greatly influences the appearance of colours in a scene [3], [10], [81]. In this paper, we show that adopting some of the computations that the human visual system forged after millions of years of evolution into a simple, functional model allows us to obtain results on par to much more complex computational learning approaches. The mechanisms in question are three: (i) adaptation of receptive field size depending on local contrast, (ii) influence of surround-on-centre also depending on local contrast, and (iii) computation of global contrast in higher visual areas to produce the final illuminant estimation. Their particular contributions were quantified by performing additional experiments. We compared our results to current state-of-the-art algorithms in four benchmark datasets showing a significant improvement regarding other low-level feature-driven methods, while still highly competitive with respect

to the best learning-based methods. The significance of this performance is emphasised considering that our model is (a) *automatic* and *parameter-free* (i.e. it does not require learning the properties of each dataset since all its initial variables are set at the beginning), (b) *parsimonious* (it follows the Occam's razor simplicity principle), and (c) *biologically-inspired* on well established findings within the neurophysiology and visual perception communities. These properties make it an excellent choice to be implemented in small image-gathering devices such as webcams and mobile phones. Furthermore, our *Adaptive Surround Modulation* model does not only provides a good solution to the *engineering* problem of removing the illuminant in images but, because of its close links to the properties of cortical neurons allows us to speculate on the *scientific* question regarding the evolutionary role of these properties of the visual system, something that other algorithms are unable to do.

As a final note, we would like to express our conviction that complex multidimensional problems such as colour constancy cannot be solved by one-fits-all solutions. In other words, the results of fully automatic solutions should not be interpreted the same as those of learning-based solutions. Our view is that these belong to different and sometimes orthogonal directions and should be considered according to their own particular merits.

ACKNOWLEDGMENTS

We would like to thank members of the NeuroBiT Group and anonymous reviewers of the manuscript for their thoughtful insights and comments. This work was funded by the Spanish Secretary of Research and Innovation (TIN2013-41751-P and TIN2013-49982-EXP).

REFERENCES

[1] D. H. Brainard and A. Radonjic, "Color constancy," in *The visual neurosciences*, vol. 1, 2004, pp. 948–961.
[2] P. M. Hubel, "The perception of color at dawn and dusk," *Journal of Imaging Science and Technology*, vol. 44, no. 4, pp. 371–375, 2000.
[3] A. Hurlbert and K. Wolf, "Color contrast: a contributory mechanism to color constancy," *Progress in brain research*, vol. 144, pp. 145–160, 2004.
[4] J. Von Kries, "Chromatic adaptation," *Festschrift der Albrecht-Ludwigs-Universität*, vol. 135, pp. 145–158, 1902.
[5] D. L. MacAdam, *Sources of color science*. Mit Press, 1970.
[6] H.-C. Lee, "Method for computing the scene-illuminant chromaticity from specular highlights," *JOSA A*, vol. 3, no. 10, pp. 1694–1699, 1986.
[7] B. V. Funt, M. S. Drew, and J. Ho, "Color constancy from mutual reflection," *International Journal of Computer Vision*, vol. 6, no. 1, pp. 5–24, 1991.
[8] J. Van De Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Transactions on image processing*, vol. 16, no. 9, pp. 2207–2214, 2007.
[9] T. Hansen, M. Olkkonen, S. Walter, and K. R. Gegenfurtner, "Memory modulates color appearance," *Nature neuroscience*, vol. 9, no. 11, pp. 1367–1368, 2006.
[10] D. H. Foster, "Color constancy," *Vision research*, vol. 51, no. 7, pp. 674–700, 2011.
[11] T. Gevers and A. W. Smeulders, "Color-based object recognition," *Pattern recognition*, vol. 32, no. 3, pp. 453–464, 1999.
[12] K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data," *Image Processing, IEEE Transactions on*, vol. 11, no. 9, pp. 972–984, 2002.
[13] T. Gevers and A. W. Smeulders, "Pictoseek: combining color and shape invariant features for image retrieval," *Image Processing, IEEE Transactions on*, vol. 9, no. 1, pp. 102–119, 2000.
[14] E. H. Land *et al.*, *The retinex theory of color vision*. Citeseer, 1977.
[15] L. T. Maloney and B. A. Wandell, "Color constancy: a method for recovering surface spectral reflectance," *JOSA A*, vol. 3, no. 1, pp. 29–33, 1986.
[16] D. A. Forsyth, "A novel algorithm for color constancy," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 5–35, 1990.
[17] M. Ebner, *Color constancy*. John Wiley & Sons, 2007, vol. 6.
[18] A. Gijsenij and T. Gevers, "Color constancy using natural image statistics and scene semantics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 4, pp. 687–698, 2011.
[19] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural computation*, vol. 4, no. 1, pp. 1–58, 1992.
[20] B. Funt, K. Barnard, and L. Martin, "Is machine colour constancy good enough?" in *ECCV*, 1998, pp. 445–459.
[21] V. C. Cardei, B. Funt, and K. Barnard, "Estimating the scene illumination chromaticity by using a neural network," *JOSA A*, vol. 19, no. 12, pp. 2374–2386, 2002.
[22] B. Funt and W. Xiong, "Estimating illumination chromaticity via support vector regression," in *Color and Imaging Conference*, vol. 2004, no. 1, 2004, pp. 47–52.
[23] V. Agarwal, A. V. Gribok, and M. A. Abidi, "Machine learning approach to color constancy," *Neural Networks*, vol. 20, no. 5, pp. 559–563, 2007.
[24] L. Shi, W. Xiong, and B. Funt, "Illumination estimation via thin-plate spline interpolation," *JOSA A*, vol. 28, no. 5, pp. 940–948, 2011.
[25] G. D. Finlayson, "Color in perspective," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 10, pp. 1034–1038, 1996.
[26] G. Finlayson and S. Hordley, "Improving gamut mapping color constancy," *Image Processing, IEEE Transactions on*, vol. 9, no. 10, pp. 1774–1783, 2000.
[27] K. Barnard, "Improvements to gamut mapping colour constancy algorithms," in *ECCV*, 2000, pp. 390–403.
[28] M. Mosny and B. Funt, "Cubical gamut mapping colour constancy," in *Conference on Colour in Graphics, Imaging, and Vision*, vol. 2010, no. 1, 2010, pp. 466–470.
[29] A. Gijsenij, T. Gevers, and J. Van De Weijer, "Generalized gamut mapping using image derivative structures for color constancy," *International Journal of Computer Vision*, vol. 86, no. 2-3, pp. 127–139, 2010.
[30] D. H. Brainard and W. T. Freeman, "Bayesian color constancy," *JOSA A*, vol. 14, no. 7, pp. 1393–1411, 1997.
[31] C. Rosenberg, A. Ladsariya, and T. Minka, "Bayesian color constancy with non-gaussian models," in *Advances in neural information processing systems*, 2003, p. None.
[32] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian color constancy revisited," in *CVPR*, 2008, pp. 1–8.
[33] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Color and Imaging Conference*, vol. 2004, no. 1, 2004, pp. 37–41.
[34] G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin institute*, vol. 310, no. 1, pp. 1–26, 1980.
[35] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 9, pp. 891–906, 1991.
[36] E. H. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proceedings of the national academy of sciences*, vol. 83, no. 10, pp. 3078–3080, 1986.
[37] H. Spitzer and Y. Barkan, "Computational adaptation model and its predictions for color induction of first and second orders," *Vision Research*, vol. 45, no. 27, pp. 3323–3342, 2005.
[38] B. R. Conway, S. Chatterjee, G. D. Field, G. D. Horwitz, E. N. Johnson, K. Koida, and K. Mancuso, "Advances in color science: from retina to behavior," *The Journal of Neuroscience*, vol. 30, no. 45, pp. 14 955–14 963, 2010.
[39] R. Shapley and M. J. Hawken, "Color in the cortex: single-and double-opponent cells," *Vision research*, vol. 51, no. 7, pp. 701–717, 2011.
[40] C. A. Parraga, "Color vision, computational methods for," *Encyclopedia of Computational Neuroscience, Ed. D. Jaeger and R. Jung, SpringerReference*, vol. 10, p. 58, 2013.
[41] J. R. Cavanaugh, W. Bair, and J. A. Movshon, "Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons," *Journal of neurophysiology*, vol. 88, no. 5, pp. 2530–2546, 2002.

[42] J. Zhang, Y. Barhomi, and T. Serre, "A new biologically inspired color image descriptor," in *ECCV*, 2012, pp. 312–324.
[43] S.-B. Gao, K.-F. Yang, C.-Y. Li, and Y.-J. Li, "Color constancy using double-opponency," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 37, no. 10, pp. 1973–1985, 2015.
[44] C. A. Parraga and A. Akbarinia, "Colour constancy as a product of dynamic centre-surround adaptation." *Journal of Vision*, vol. 16, no. 12, pp. 214–214, 2016.
[45] C. Enroth-Cugell and J. G. Robson, "The contrast sensitivity of retinal ganglion cells of the cat," *The Journal of physiology*, vol. 187, no. 3, pp. 517–552, 1966.
[46] D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 207, no. 1167, pp. 187–217, 1980.
[47] S. Shushruth, J. M. Ichida, J. B. Levitt, and A. Angelucci, "Comparison of spatial summation properties of neurons in macaque v1 and v2," *J. Neurophysiology*, vol. 102, no. 4, pp. 2069–2083, 2009.
[48] A. Angelucci and S. Shushruth, "Beyond the classical receptive field: Surround modulation in primary visual cortex," *The new visual neurosciences*, pp. 425–444, 2013.
[49] H. Wilson and F. Wilkinson, "Configural pooling in the ventral pathway," *New visual neurosciences*, pp. 617–626, 2014.
[50] J. Allman, F. Miezin, and E. McGuinness, "Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons," *Annual review of neuroscience*, vol. 8, no. 1, pp. 407–430, 1985.
[51] M. K. Kapadia, M. Ito, C. D. Gilbert, and G. Westheimer, "Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in v1 of alert monkeys," *Neuron*, vol. 15, no. 4, pp. 843–856, 1995.
[52] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature reviews neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.
[53] N. Murray, M. Vanrell, X. Otazu, and C. A. Parraga, "Saliency estimation using a non-parametric low-level vision model," in *CVPR*, 2011, pp. 433–440.
[54] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," in *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3. ACM, 2002, pp. 267–276.
[55] A. Akbarinia, C. A. Parraga *et al.*, "Biologically-inspired edge detection through surround modulation," in *Proceedings of the British Machine Vision Conference*, 2016, pp. 1–13.
[56] J. M. Ichida, L. Schwabe, P. C. Bressloff, and A. Angelucci, "Response facilitation from the suppressive receptive field surround of macaque v1 neurons," *Journal of Neurophysiology*, vol. 98, no. 4, pp. 2168–2181, 2007.
[57] G. A. Walker, I. Ohzawa, and R. D. Freeman, "Asymmetric suppression outside the classical receptive field of the visual cortex," *The Journal of Neuroscience*, vol. 19, no. 23, pp. 10 536–10 553, 1999.
[58] K. R. Gegenfurtner, "Cortical mechanisms of colour vision," *Nature Reviews Neuroscience*, vol. 4, no. 7, pp. 563–572, 2003.
[59] M. Carandini and D. J. Heeger, "Normalization as a canonical neural computation," *Nature Reviews Neuroscience*, vol. 13, no. 1, pp. 51–62, 2012.
[60] B. A. Olshausen *et al.*, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.
[61] G. D. Finlayson, S. D. Hordley, and M. S. Drew, "Removing shadows from images," in *ECCV*, 2002, pp. 823–836.
[62] K. Barnard, L. Martin, B. Funt, and A. Coath, "A data set for color research," *Color Research & Application*, vol. 27, no. 3, pp. 147–151, 2002.
[63] R. T. Tan, K. Nishino, and K. Ikeuchi, "Color constancy through inverse-intensity chromaticity space," *JOSA A*, vol. 21, no. 3, pp. 321–334, 2004.
[64] S. Gao, W. Han, K. Yang, C. Li, and Y. Li, "Efficient color constancy with local surface reflectance statistics," in *ECCV*, 2014, pp. 158–173.
[65] B. Funt and M. Mosny, "Removing outliers in illumination estimation," in *Color and Imaging Conference*, vol. 2012, no. 1, 2012, pp. 105–110.
[66] K.-F. Yang, S.-B. Gao, and Y.-J. Li, "Efficient illuminant estimation for color constancy using grey pixels," in *CVPR*, 2015, pp. 2254–2263.
[67] A. Chakrabarti, K. Hirakawa, and T. Zickler, "Color constancy with spatio-spectral statistics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 8, pp. 1509–1519, 2012.
[68] A. Gijsenij, T. Gevers, and J. Van De Weijer, "Improving color constancy by photometric edge weighting," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 5, pp. 918–929, 2012.
[69] H. R. V. Joze and M. S. Drew, "Exemplar-based color constancy and multiple illumination," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 5, pp. 860–873, 2014.
[70] S. Bianco, C. Cusano, and R. Schettini, "Color constancy using cnns," in *CVPR Workshops*, 2015, pp. 81–89.
[71] Z. Lou, T. Gevers, N. Hu, and M. P. Lucassen, "Color constancy by deep learning," in *BMVC*, September 2015, pp. 1–12.
[72] J. Roca-Vila, C. A. Parraga, and M. Vanrell, "Chromatic settings and the structural color constancy index," *Journal of vision*, vol. 13, no. 4, pp. 3–3, 2013.
[73] J. Vazquez-Corral, C. Párraga, R. Baldrich, and M. Vanrell, "Color constancy algorithms: Psychophysical evaluation on a new dataset," *Journal of Imaging Science and Technology*, vol. 53, no. 3, pp. 31 105–1, 2009.
[74] G. D. Finlayson and R. Zakizadeh, "Reproduction angular error: An improved performance metric for illuminant estimation," *perception*, vol. 310, no. 1, pp. 1–26, 2014.
[75] S. D. Hordley and G. D. Finlayson, "Reevaluation of color constancy algorithm performance," *JOSA A*, vol. 23, no. 5, pp. 1008–1020, 2006.
[76] A. Gijsenij, T. Gevers, and M. P. Lucassen, "Perceptual analysis of distance measures for color constancy algorithms," *JOSA A*, vol. 26, no. 10, pp. 2243–2256, 2009.
[77] L. Shi and B. Funt, "Re-processed version of the gehler color constancy dataset of 568 images," http://www.cs.sfu.ca/~colour/data/.
[78] F. Ciurea and B. Funt, "A large image database for color constancy research," in *Color and Imaging Conference*, vol. 2003, no. 1, 2003, pp. 160–164.
[79] S. Beigpour, C. Riess, J. Van de Weijer, and E. Angelopoulou, "Multi-illuminant estimation with conditional random fields," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 83–96, 2014.
[80] A. Gijsenij, R. Lu, and T. Gevers, "Color constancy for multiple light sources," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 697–707, 2012.
[81] R. O. Brown and D. I. MacLeod, "Color appearance depends on the variance of surround colors," *Current Biology*, vol. 7, no. 11, pp. 844–849, 1997.
[82] X. Otazu, C. A. Parraga, and M. Vanrell, "Toward a unified chromatic induction model," *Journal of Vision*, vol. 10, no. 12, pp. 5–5, 2010.
[83] I. Lampl, D. Ferster, T. Poggio, and M. Riesenhuber, "Intracellular measurements of spatial integration and the max operation in complex cells of the cat primary visual cortex," *Journal of neurophysiology*, vol. 92, no. 5, pp. 2704–2713, 2004.
[84] H. R. V. Joze, M. S. Drew, G. D. Finlayson, and P. A. T. Rey, "The role of bright pixels in illumination estimation," in *Color and Imaging Conference*, vol. 2012, no. 1, 2012, pp. 41–46.
[85] A. Akbarinia and C. A. Parraga, "Feedback and surround modulated boundary detection," *International Journal of Computer Vision*, pp. 1–14, 2017.

Arash Akbarinia received his BSc. degree in Software Engineering from University of Gothenburg, Sweden. Was awarded MSc. in Computer Vision and Robotics (VIBOT) from an Erasmus Mundus programme coordinated by University of Burgundy, France, University of Girona, Spain, and Heriot-Watt University, Scotland. Currently pursuing his Ph.D. at Universitat Autònoma de Barcelona. His research interests include computational modelling of human visual system.

C. Alejandro Parraga received his first degree in Physics from the Univ. Nacional de Tucumán (Argentina) and his PhD in Visual Perception from the University of Bristol (UK). In 2008 was recipient of the "Ramón y Cajal" fellowship (Spain). Since 2012 works at the Comp. Vision Centre / Comp. Sci. Dept. of the Universitat Autònoma de Barcelona on biologically-plausible computational models of visual perception, colour vision and psychophysics.