

## NON-LOCAL RETINEX – A UNIFYING FRAMEWORK AND BEYOND \*

Dominique Zosso<sup>†</sup>, Giang Tran , and Stanley J. Osher

**Abstract.** In this paper, we provide a short review of Retinex and then present a unifying framework. The fundamental assumption of all Retinex models is that the observed image is a multiplication between the illumination and the true underlying reflectance of the object. Starting from Morel’s 2010 PDE model, where illumination is supposed to vary smoothly and where the reflectance is thus recovered from a hard-thresholded Laplacian of the observed image in a Poisson equation, we define our unifying Retinex model in similar but more general two steps.

We reinterpret the gradient thresholding model as variational models with sparsity constraints. First, we look for a filtered gradient that is the solution of an optimization problem consisting of two terms: a sparsity prior of the reflectance and a fidelity prior of the reflectance gradient to the observed image gradient. Second, since this filtered gradient almost certainly is not a consistent image gradient, we then fit an actual reflectance gradient to it, subject to further sparsity and fidelity priors. This generalized formulation allows making connections with other, variational or kernel-based Retinex implementations.

We provide simple algorithms for the optimization problems resulting from our framework. In particular, in the quadratic case, we can link our model to a plausible neural mechanism through Wilson-Cowan equations. Beyond unifying existing models, we derive entirely novel Retinex flavors by using more interesting non-local versions for the sparsity and fidelity prior. Eventually, we define within a single framework new Retinex applications to shadow detection and removal, nonuniformity correction, cartoon-texture decomposition, as well as color and hyperspectral image enhancement.

**Key words.** Retinex, non-local operators, reflectance, illumination, image decomposition, contrast enhancement, shadow detection, cartoon-texture decomposition, thresholding.

**AMS subject classifications.** 68U10, 65D19, 68T45, 49N45, 35Q93.

**DOI.**

---

This is a significantly expanded and overhauled version of work initially presented in much shorter and preliminary form in [83].

**1. Introduction.** Retinex is a theory on the human visual perception [38, 42, 39]. It was an attempt at explaining how a combination of processes supposedly taking place both in the retina and the cortex is capable of adaptively coping with illumination that varies spatially. The fundamental observation is the insensitivity of human visual perception with respect to a slowly varying illumination on a Mondrian-like scene, see figure 1.1.

Amongst other models for enhancement and noise removal—such as [75, 43, 65, 66], for an overview see [1]—Retinex has received particular attention because it is deeply connected with psycho-visual empirical evidence of human perception. However, depending on the application,

---

\*The authors are with the University of California, Los Angeles (UCLA), Department of Mathematics, 520 Portola Plaza, Box 951555, Los Angeles, CA 90095-1555. eMail: {zosso,giangtran,sjo}@math.ucla.edu.

DZ is supported by the Swiss National Science Foundation (SNF) under grants PBELP2-137727 and P300P2-147778. GT is supported by the UC Lab Fees Award ID# 12-LR-23660 and ONR N00014-10-10221. SO is supported by grants NSF DMS-118971, NSF DMS-0914561 and ONR N00014-12-10838.

<sup>†</sup>To whom correspondence should be addressed

the various Retinex assumptions are given different importance, and resulting implementations vary significantly; to the point where “opposing” schools of thought openly accuse each other of doing it wrongly. Retinex research is a minefield and here we explicitly want to stay non-partisan.

The outline of this paper is as follows: we start by providing a short review of some Retinex flavors, in section 2. We tentatively classify models and algorithms in 5 main classes: original Retinex formulation, threshold-based PDE-models, reset-based random walk and kernel-based methods, reset- and threshold-free center-surround models, variational Retinex. While fundamental connections and equivalences have been shown between random-walk, threshold, and the original Retinex formulation [30, 53], a divide still exists between recent kernel-based Retinex such as [4] and variational models like [36].

To fill this gap, we will then in section 3 recall some definitions and notions from non-local differential operators, and introduce a few new concepts. Based on these non-local differential operators we will be able to make a formal connection between existing kernel-based, center-surround, and variational Retinex models. Indeed, in section 4 we provide a short overview of relevant methods for constructing weight kernels, some of which are well known and others are suggested more unexpectedly. Then we will show in section 5 how threshold-based, kernel-based, and center-surround Retinex can be expressed as variational models, the latter two by making use of these non-local generalizations of differential operators using particular kernels.

In a second step, we will propose a tentative unifying, non-local framework for Retinex in section 6. Our proposed model takes the shape of a *generalized fidelity to thresholded-gradient problem*. We provide a quick outline of how we numerically solve our model for proof of concept, and we show important connections with Wilson-Cowan IDE equations of neural-networks. We will then show how the proposed unifying model can reproduce results of other state-of-the-art Retinex models in section 7. Beyond simply reproducing existing results, we will explore the new degrees of freedom of the proposed unifying framework, as shown in section 8. Especially, the flexibility in choosing the weights for the sparsity and fidelity terms, separately, brings about the possibility of adaptive or conditional thresholding, in section 9, inspired by the idea of interacting color channels and hue-sensitivity, which allows far-reaching connections with state-of-the-art shadow-removal algorithms far from traditional Retinex. Finally, we will conclude on our framework in section 10.

## 2. A short review of Retinex implementations.

**2.1. Original Retinex algorithm.** Land formalized the reflectance ratios, by summing thresholded log-ratios over continuous paths between two pixels [40]. He defines the relative reflectance of pixel  $i$  to  $j$  as (see figure 2.1):

$$(2.1) \quad R(i, j) = \sum_k \delta_\tau \log \frac{I_{k+1}}{I_k} \quad \text{with} \quad \delta_\tau(\cdot) := \begin{cases} \cdot & \text{if } |\cdot| \geq \tau \\ 0 & \text{otherwise} \end{cases}$$

where  $\delta_\tau$  denotes hard thresholding. The thresholding makes sure, that only sharp intensity transitions are included in the sum, and the slow drift due to smoothly varying illumination is eliminated. Provided that the thresholding yields perfect distinction between illumination gradients and actual feature edges, the relative reflectance of  $i$  to  $j$  is invariant to the path

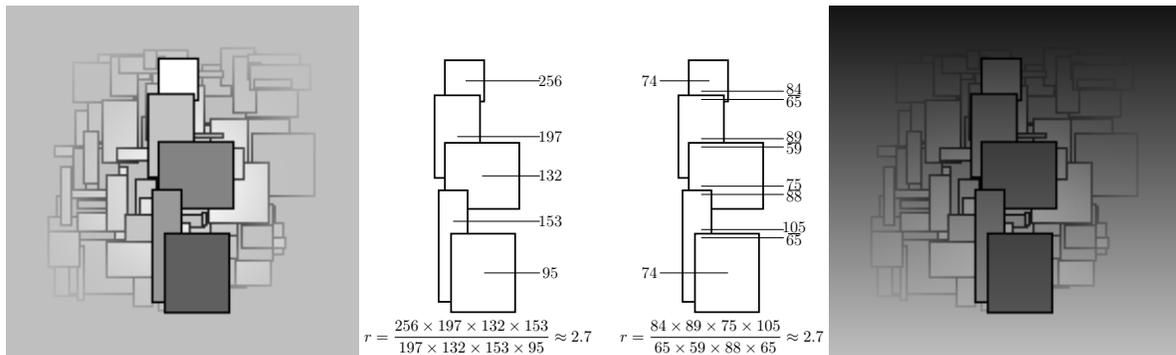


Figure 1.1: **Mondrian illusion.** **Left:** Under uniform illumination, the relative intensity ratio between the top-most and bottom-most patch in this synthetic “Mondrian” image is obvious. **Right:** Albeit these patches have the same absolute gray-level under smoothly varying illumination, their underlying relative reflectance ratio, 2.7, is still perceived by humans, and it can be recovered by multiplying the local intensity ratios at the discontinuities along any path joining the two patches, effectively discarding gradients due to lighting.

connecting them. The average relative reflectance at  $i$  is then estimated as

$$(2.2) \quad \bar{R}(i) = \mathbb{E}_j [R(i, j)] = \frac{1}{N} \sum_{j=1}^N R(i, j)$$

However, “the ultimate purpose is to describe any area by relating its reflectance to a single, standard, high reflectance somewhere in the Mondrian or to several equally high reflectances” [42]. Instead of localizing the highest reflectance in a preprocessing step, which seemed biologically unplausible, it was proposed to estimate the maximum reflectance directly while performing the sequential sum along each path. Indeed, whenever the intermediate sequential sum from  $j$  up to  $I_{k+1}$ , i.e. the relative reflectance of  $I_{k+1}$  to  $j$ , becomes positive—equivalent to a sequential product bigger than 1—, one has reached a new maximum reflectance, and the sequential sum is reset, with  $I_{k+1}$  as new reference. Due to the presence of the thresholding operator, the final reference pixel does not necessarily coincide with the brightest pixel along the path. For a mathematical definition and analysis of this reset mechanism, see [60].

There has been quite some debate about the respective role and importance of both threshold and reset in the Retinex, including by McCann himself [51]. The criterion will serve us dividing the many Retinex implementations in two broad classes: threshold-based versus reset-based. A third class of implementations is based on an alternative technique proposed by Land, which determines lightness as ratio of the local intensity compared to the average intensity of its immediate (circular) surroundings, without neither thresholding, nor reset [41]. A fourth class, finally, extracts the reflectance and illumination information variationally, by optimizing different energy functionals.

**2.2. Threshold-based Retinex implementations (PDE).** In 1974, Horn proposed a mathematical alternative to the Retinex algorithm that differs substantially in form [29]. He es-

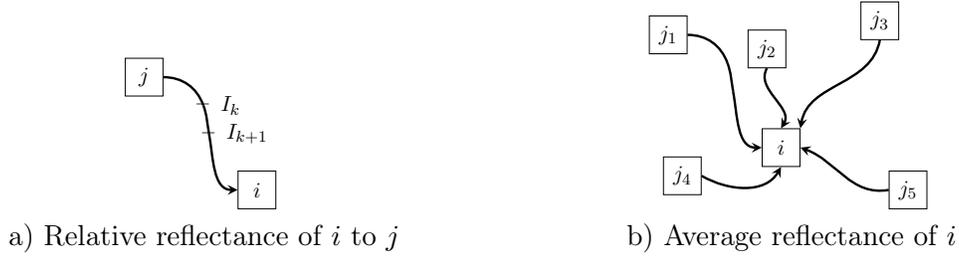


Figure 2.1: **Relative and average reflectance.** **a)** The relative reflectance of pixel  $i$  to pixel  $j$  is computed by accumulating all *important* local intensity gradients along any path  $\gamma$  joining these pixels. **b)** The average reflectance of  $i$  is then obtained as the average relative reflectance of  $i$  to different  $j$ .

essentially stripped the Retinex algorithm down to a smoothness prior on the illumination field, and thus to a thresholding on intensity derivatives. He poses the problem of recovering the underlying reflectance  $R$ , which multiplied by the illumination  $B$  resulted in the observed  $I$ :

$$(2.3) \quad I(x, y) = B(x, y)R(x, y)$$

By taking the logarithm,  $i := \log(I)$  etc., we obtain an additive impact of illumination:

$$(2.4) \quad i(x, y) = b(x, y) + r(x, y)$$

Since the illumination  $b$  is supposed to be varying smoothly, the spatial derivatives of the observed intensity are mostly due to edges in the reflectance  $r$ . However, he realized that first order derivatives are directional in the two-dimensional case of images, and that the lowest order isotropic derivatives are found in the scalar Laplacian operator:  $\Delta b$  will be finite everywhere, while  $\Delta r$  will be zero except at each edge separating regions [29]. Therefore, discarding the finite parts of the observed intensity Laplacian is supposed to yield the Laplacian of the reflectance (Poisson equation):

$$(2.5) \quad \Delta r = \delta_\tau \Delta i$$

A tight mathematical connection between Land's and Horn's computations, on the basis of Green's formula, has been shown in work by Hurlbert [30]. A fully discrete alternative to Horn's convolution and inversion scheme was proposed by [49]. There, an equivalent two-stage algorithm was introduced, which achieves inversion at very low computational cost, feasible in terms of neural networks. The steps are first local contrast computation and thresholding ( $\delta_\tau$ ), then inversion, as follows:

$$(2.6) \quad c(x) = \delta_\tau \left( i(x) - \sum_{y \neq x} w(x, y) i(y) \right)$$

$$(2.7) \quad r(x) = c(x) + \sum_{y \neq x} w(x, y) r(y)$$

While the first step is straightforward, the second step realizes inversion through feedback in terms of neural networks:  $r(y)$  is Retinex output *computed in parallel*, and the system (hopefully) converges to a steady state solution.

Horn’s model has been strongly backed up by a much more recent paper by Morel [53], where the authors show a very tight connection between Horn’s Laplacian thresholding and Land’s original, resetless Retinex algorithm. Indeed, “*if the Retinex paths are interpreted as symmetric random walks, then Retinex is equivalent to a Neumann problem for a linear Poisson equation*” [53]. The main difference between Horn and Morel concerns the argument of the hard thresholding operator: while Horn thresholds the scalar Laplacian, Morel thresholds the components of the gradient prior to computing the divergence. *De facto*, Morel thus effectively solves an  $L^2$ -gradient fitting problem:

$$(2.8) \quad \hat{r} = \arg \min_r \{ \|\nabla r - \delta_\tau \nabla i\|_2^2 \}$$

We refer to this basic model as  $L^2$ -Retinex. Note that reconstruction from thresholded gradient has earlier been proposed by Blake [8, 9, 11]. More recently, the  $L^1$ -equivalent thresholded-gradient fidelity Retinex has been proposed: The  $L^1$ -Retinex minimizes the isotropic  $L^1$ -distance [45]:

$$(2.9) \quad \hat{r} = \arg \min_r \{ \|\nabla r - \delta_\tau \nabla i\|_1 \}$$

**2.3. Reset-based Retinex implementations (Random walk).** Moving away from thresholding and relying purely on the reset mechanism, Frackle and McCann have patented their Retinex algorithm [20]. The Frackle-McCann algorithm replaces sequential products along paths by pairwise pixel ratios sampled along discrete spirals. Long-distance interactions are computed first, then the sampling progressively approaches the center pixel while decreasing the spacing. At each step, the lightness estimate is updated with a ratio-product-reset-average operation [21]. More recent variants of the algorithm mainly involve multiresolution image pyramids [50, 21], different sampling patterns [80, 31], or ratio modifiers [70].

Two of the main drawbacks of the Frackle-McCann algorithm are the strong dependence on the path length of the spiral pattern (represented as number of iterations), and the appearance of asymmetric halos due to the anisotropic sampling pattern. In order to avoid these issues, Provenzi et al. replace the path-based sampling pattern by a repeated sampling through random sprays [61]. Indeed, if the threshold is removed from the Retinex formulation, then the reset reduces the relative reflectance, computed using a specific path  $\gamma_k$ , to the ratio of central pixel  $I(i)$  and brightest pixel  $I(x)$  along that same path  $\gamma_k$  [60]:

$$(2.10) \quad R(i, j) = \log \frac{I(i)}{I(x)}$$

$$(2.11) \quad R(i) = \frac{1}{N} \sum_{k=1}^N \log \frac{I(i)}{I(x_k)}$$

where  $x_k$  is the brightest pixel encountered along  $\gamma_k$ . Therefore, many paths become redundant, and the maxima  $I(x_k)$  can be sampled alternatively. It was suggested that averaging

repeated random spray sampling of  $x_k$  directly, with radially decreasing sampling density, can substitute the path-based filtering efficiently. On the downside, such sampling of extrema exhibits high variance, and the recovered reflectances are typically noisy. More robust estimators, such as high-percentage quantiles rather than the most extreme sample, results in smoother illumination estimation and therefore lower noise.

Beyond, the (white-patch) random spray Retinex was combined with a (gray-world) model used for automatic color equalization (ACE) [65, 62]. Eventually, the random spray sampling was replaced by a kernel, representing the sampling density of the random spray in the limit case [4, 59]:

$$(2.12) \quad R(i) = \sum_{j:I(j)\geq I(i)} w(i,j) f\left(\frac{I(i)}{I(j)}\right) + \sum_{j:I(j)<I(i)} w(i,j)$$

where  $w(i,j)$  is the kernel, representing the probability density of picking a pixel  $j$  in the neighborhood of  $i$  [4]. Note that, here again, we find the ratio modifier  $f$  previously introduced by Sobol [70].

**2.4. Center-surround Retinex implementations.** A simple alternative to threshold/reset based Retinex algorithms was proposed by Land based on findings of lateral inhibition [41]. The alternative consists in determining the local lightness (reflectance) as the ratio between local intensity and an average of its close surroundings. Land realized that this much simpler model could reproduce all Retinex behavior modeled so far, and had the additional competence of generating empirically perceived Mach bands [64, 17]. The fundamental idea is again that the low-frequency components are due to illumination, while the high-frequency details are features in the reflectance.

10 years later, only, the idea was picked up and formulated as single- and multi-scale center-surround Retinex [35, 34, 63]. The single-scale Retinex is given by

$$(2.13) \quad R(i) = \log I(i) - \log [F * I](i)$$

where  $F$  is a Gaussian kernel. The multi-scale Retinex is then simply the combination of different single-scale retineces:

$$(2.14) \quad R(i) = \log I(i) - \sum_n w_n \log [F_n * I](i)$$

where  $w_n : \sum_n w_n = 1$  are the weights of each scale, and  $F_n$  are Gaussian kernels of different scale.

Changing the order of log and Gaussian convolution in the single scale Retinex amounts to homomorphic filtering

$$(2.15) \quad R(i) = \log I(i) - [F * \log I](i)$$

which in turn can be identified as a special case of (resetless) kernel-Retinex, with the kernel  $w(i,j) \equiv F$  and ratio modifier  $f \equiv \log$ :

$$(2.16) \quad R(i) = \sum_j w(i,j) \log\left(\frac{I(i)}{I(j)}\right) = \log I(i) - \sum_j w(i,j) \log I(j)$$

**2.5. Variational Retinex.** A whole family of variational Retinex models handles the regularity priors on the reflectance and illumination parts of the Retinex decomposition in a more explicit way. First of its kind, the variational framework by Kimmel introduces competing  $H^1$  smoothness priors on both the illumination and reflectance fields, as well as a quadratic fidelity prior between illumination and observed intensity [36]. In addition, illumination is constrained to be bigger or equal to observed intensity, i.e. the reflectance is limited by an upper bound:

$$(2.17) \quad \min_b \left\{ \int_{\Omega} |\nabla b|^2 + \alpha(b - i)^2 + \beta |\nabla b - \nabla i|^2 dx dy \right\} \quad \text{s.t.} \quad b \geq i, \quad \langle \nabla b, \vec{n} \rangle = 0 \text{ on } \partial\Omega.$$

Kimmel solves this quadratic programming problem using projected normalized steepest descent at multiple resolutions [36]. Here, we rewrite the problem slightly, optimizing for the reflectance rather than the illumination, by substituting according to the coherence condition  $i = b + r$ :

$$(2.18) \quad \min_r \left\{ \|\nabla r - \nabla i\|_2^2 + \alpha \|r\|_2^2 + \beta \|\nabla r\|_2^2 \right\} \quad \text{s.t.} \quad r \leq 0, \quad \langle \nabla r, \vec{n} \rangle = 0 \text{ on } \partial\Omega.$$

This form makes clear that variational Retinex is an optimization between reflectance gradient fidelity and some sparsity penalties.

Subsequently, variations of this variational Retinex model have been proposed, mainly involving different norms for the fidelity and sparsity terms, and dropping the asymmetry constraint  $r \leq 0$ . First, Ma and Osher have dropped a few terms and replace  $H^1$  smoothness of the reflectance by a TV-prior [46]:

$$(2.19) \quad \min_r \left\{ \|\nabla r - \nabla i\|_2^2 + 2\lambda \|\nabla r\|_1 \right\}$$

As a complication, instead of the local TV prior, they also make use of non-local total variation. Further, Ng and Wang introduce an  $L^2$ -fidelity prior between reflectance and intensity [54]:

$$(2.20) \quad \min_r \left\{ \|\nabla r - \nabla i\|_2^2 + \alpha \|r - i\|_2^2 + 2\lambda \|\nabla r\|_1 \right\}$$

Chen et al. have used a TV- $L^1$ -based variational Retinex approach, which they call logarithmic total variation (LTV), for illumination normalized face detection [15]:

$$(2.21) \quad \min_r \left\{ \|\nabla r - \nabla i\|_1 + \alpha \|r\|_1 \right\}$$

**Remark 2.1.** *At this point it is worthwhile noting, that both the  $L^2$ - and  $L^1$ -Retinex [53, 45] have a threshold-free variational equivalent.*

Indeed, the hard threshold on the intensity gradient can be seen as a contraction of an  $L^0$ -sparsity prior on the gradients of the reflectance:

$$(2.22) \quad \min_r \left\{ \|\nabla r - \delta_\tau \nabla i\|_2^2 \right\} = \min_r \left\{ \left\| \nabla r - \arg \min_{\vec{q}} \left\{ \|\vec{q} - \nabla i\|_2^2 + \tau^2 \|\vec{q}\|_0 \right\} \right\|_2^2 \right\}$$

which is a relaxed version of the more complicated problem

$$(2.23) \quad \min_r \left\{ \|\vec{q} - \nabla i\|_2^2 + \tau^2 \|\vec{q}\|_0 \right\} \quad \text{s.t.} \quad \nabla r = \vec{q}$$

This connection between basic threshold and a related variational problem makes it intuitive, that other variational models, such as the TV-Retinex [46], can be retro-fit into a threshold based Poisson-problem, as well. *This is the fundamental insight leading to the non-local unification proposed in this paper.*

**3. Non-local differential operators (basic definitions).** In this section, we recall and give a few definitions of non-local differential operators, which we need in order to cast existing kernel-based Retinex methods into a variational framework, and based on which we will propose our unifying Retinex framework. To this end, we recall a few definitions of non-local operators [22]. Here, we employ the continuous definitions, but their translation into the discrete case is relatively straightforward.

**3.1. Products and norms.** First, we give the definitions of particular products and norms of scalars and non-local vectors. To begin with, we require appropriate inner products.

**Definition 3.1.** For scalars  $i : \Omega \rightarrow \mathbb{R}$ , we choose:

$$(3.1) \quad \langle i, j \rangle := \int_{\Omega} i(x)j(x)dx,$$

which is the common  $L^2$  inner product.

**Definition 3.2.** Accordingly, we introduce the following inner product for vectors  $\vec{v} : \Omega \rightarrow \Omega \times \Omega$ :

$$(3.2) \quad \langle \vec{u}, \vec{v} \rangle := \int_{\Omega \times \Omega} u(x, y)v(x, y)dxdy.$$

**Definition 3.3.** The associated  $L^2$  norms are respectively for scalars  $i : \Omega \rightarrow \mathbb{R}$ :

$$(3.3) \quad \|i\|_2 := \sqrt{\langle i, i \rangle} = \sqrt{\int_{\Omega} i(x)^2 dx},$$

and for vectors  $\vec{v} : \Omega \rightarrow \Omega \times \Omega$ :

$$(3.4) \quad \|\vec{v}\|_2 := \sqrt{\langle \vec{v}, \vec{v} \rangle} = \sqrt{\int_{\Omega \times \Omega} v(x, y)^2 dxdy}.$$

**Definition 3.4.** Similarly, the  $L^1$ -norm of the vector  $\vec{v}$ ,  $\|\vec{v}\|_1 : \Omega \times \Omega \rightarrow \mathbb{R}$ , is defined as

$$(3.5) \quad \|\vec{v}\|_1 := \int_{\Omega \times \Omega} |v(x, y)| dxdy.$$

**Definition 3.5.** Let  $w$  be a non-negative weighting function and  $\vec{v}$  a vector. The weighted  $L^0$ -“norm” of the vector  $\vec{v}$ ,  $\|\vec{v}\|_{0,w} : \Omega \times \Omega \rightarrow \mathbb{R}$ , is defined as

$$(3.6) \quad \|\vec{v}\|_{0,w} := \int_{\Omega \times \Omega} w(x, y)(1 - \delta(v(x, y))) dxdy$$

where  $\delta$  is the Dirac distribution. This functional emphasizes the  $(L^0)$  sparsity of important vector components.

**Definition 3.6.** Further, pointwise multiplication is written for scalars  $i$  and  $j$  as

$$(3.7) \quad (i \cdot j)(x) := i(x)j(x), \quad x \in \Omega$$

and for vectors  $\vec{u}$  and  $\vec{v}$  as:

$$(3.8) \quad (\vec{u} \cdot \vec{v})(x, y) := u(x, y)v(x, y), \quad x, y \in \Omega$$

**3.2. Differential operators.** We can now extend the gradient operator to the non-local case:

**Definition 3.7.** Let  $\Omega \in \mathbb{R}^n$ ,  $x \in \Omega$ ,  $i(x)$  be a real function  $i : \Omega \rightarrow \mathbb{R}$ . We define the non-local gradient of this function as the vector of all partial derivatives,  $\nabla_w i : \Omega \rightarrow \Omega \times \Omega$ :

$$(3.9) \quad (\nabla_w i)(x, y) := \sqrt{w(x, y)}(i(y) - i(x)), \quad x, y \in \Omega$$

for some non-negative weights  $w(x, y)$ .

**Remark 3.8.** Note that this definition of non-local gradients amounts to a pointwise multiplication between the vector of all finite differences,  $\vec{d}_i$ , and the weight-vector  $\vec{\omega}$ :

$$(3.10) \quad (\nabla_w i) \equiv \vec{\omega} \cdot \vec{d} \quad \begin{cases} \vec{\omega} : \omega(x, y) = \sqrt{w(x, y)} & x, y \in \Omega \\ \vec{d}_i : d_i(x, y) = i(y) - i(x) & x, y \in \Omega \end{cases}$$

Thus we will call any such pointwise product between a weighting function and another vector, say,  $\vec{q}$ , a quasi-gradient, insofar as it shares the form of an actual gradient up to the fact that the vector  $\vec{q}$  does not actually stem from non-local finite differences.

**Definition 3.9.** The associated divergence of a vector  $\vec{v} \in \Omega \times \Omega$ , namely  $\text{div}_w \vec{v} : \Omega \times \Omega \rightarrow \Omega$ , is then defined as the negative adjoint under the above inner products:

$$(3.11) \quad \langle \nabla_w i, \vec{v} \rangle = \langle i, -\text{div}_w \vec{v} \rangle,$$

The expression for the divergence is easily found as

$$(3.12) \quad (\text{div}_w \vec{v})(x) := \int_{\Omega} \sqrt{w(x, y)}v(x, y) - \sqrt{w(y, x)}v(y, x)dy.$$

**Definition 3.10.** The non-local Laplacian,  $\Delta_w i : \Omega \rightarrow \Omega$  is defined as the composition of non-local divergence and non-local gradient:

$$(3.13) \quad (\Delta_w i)(x) := (\text{div}_w(\nabla_w i))(x) = \int_{\Omega} (w(x, y) + w(y, x))(i(y) - i(x))dy.$$

**Lemma 3.11.** *Let  $w_s(x, y)$  be a symmetric weighting function, i.e.  $\forall x, y \in \Omega : w_s(x, y) = w_s(y, x)$ . This restriction simplifies the expressions of both the divergence and associated Laplacian:*

$$(3.14) \quad (\operatorname{div}_{w_s} \vec{v})(x) = \int_{\Omega} \sqrt{w_s(x, y)} (v(x, y) - v(y, x)) dy,$$

and

$$(3.15) \quad (\Delta_{w_s} i)(x) = (\operatorname{div}_{w_s} (\nabla_{w_s} i))(x) = 2 \int_{\Omega} w_s(x, y) (i(y) - i(x)) dy,$$

where the Laplacian now differs from the regular graph Laplacian by a factor 2.

*Proof.* The symmetry of weights,  $w_s(x, y) = w_s(y, x)$  in particular leads to:

$$(3.16) \quad \sqrt{w_s(x, y)} = \sqrt{w_s(y, x)}, \text{ and}$$

$$(3.17) \quad w_s(x, y) + w_s(y, x) = 2w_s(x, y).$$

From which the simplifications immediately follow. ■

**3.3. Filtered gradients.** Based on the non-local differential operators defined above, we now introduce *filtered gradients*, by making use of a filter function  $f$  acting on the scalar differences.

**Definition 3.12.** *Be  $f : \mathbb{R} \rightarrow \mathbb{R}$  a real-valued distortion function applied to the finite differences. We define filtered non-local gradients,  $\nabla_{w, f} i : \Omega \rightarrow \Omega \times \Omega$ , as the quasi-gradients obtained as follows:*

$$(3.18) \quad (\nabla_{w, f} i)(x, y) := \sqrt{w(x, y)} f(i(y) - i(x)), \quad x, y \in \Omega$$

**Definition 3.13.** *We call  $\Delta_{w, f}$  the filtered non-local Laplacian obtained by applying the (regular) divergence to filtered gradients*

$$(3.19) \quad (\Delta_{w, f} i)(x) := (\operatorname{div}_w (\nabla_{w, f} i))(x) = \int_{\Omega} w(x, y) f(i(y) - i(x)) - w(y, x) f(i(x) - i(y)) dy.$$

**Lemma 3.14.** *Let  $f_s$  be a symmetric real-valued function, i.e.  $f_s(z) = f_s(-z)$  and choose the weights  $w_s(x, y) = w_s(y, x)$  symmetrically. The associated filtered non-local Laplacian  $\Delta_{w_s, f_s}$  is always zero:*

$$(3.20) \quad (\Delta_{w_s, f_s})(x) = 0.$$

*Proof.* The symmetrically filtered non-local Laplacian with symmetric weights is written

$$(3.21) \quad (\Delta_{w_s, f_s})(x) = \int_{\Omega} w_s(x, y) f_s(i(y) - i(x)) - w_s(y, x) f_s(i(x) - i(y)) dy.$$

Due to symmetry, one has both  $w_s(x, y) = w_s(y, x)$  and  $f_s(i(x) - i(y)) = f_s(i(y) - i(x))$  and therefore the integrand vanishes. ■

**Lemma 3.15.** *Let  $f_a$  be an anti-symmetric real-valued function, i.e.  $f_a(z) = -f_a(-z)$  and choose the weights  $w_s(x, y) = w_s(y, x)$  symmetrically. The associated filtered non-local Laplacian  $\Delta_{w_s, f_a}$  is given by:*

$$(3.22) \quad (\Delta_{w_s, f_a})(x) = 2 \int_{\Omega} w_s(x, y) f_a(i(y) - i(x)) dy.$$

*Proof.* The anti-symmetrically filtered non-local Laplacian with symmetric weights reads

$$(3.23) \quad (\Delta_{w_s, f_a})(x) = \int_{\Omega} w_s(x, y) f_a(i(y) - i(x)) - w_s(y, x) f_a(i(x) - i(y)) dy.$$

Due to (anti-)symmetry, one has both  $w_s(x, y) = w_s(y, x)$  and  $f_a(i(x) - i(y)) = -f_a(i(y) - i(x))$ , from which the lemma is directly obtained. ■

**3.4. Non-stationary filtering.** In the above definitions, the filter function  $f$  was stationary, i.e. independent of the location. Here, we now introduce a non-stationary filtering function.

**Definition 3.16.** *Be  $f^{ns} : \mathbb{R} \times \Omega \times \Omega \rightarrow \mathbb{R}$  a non-stationary real-valued distortion function applied to the finite differences. We define non-stationary filtered non-local gradients,  $\nabla_{w, f^{ns}i} : \Omega \rightarrow \Omega \times \Omega$ , as follows:*

$$(3.24) \quad (\nabla_{w, f^{ns}i})(x, y) := \sqrt{w(x, y)} f^{ns}(i(y) - i(x), x, y), \quad x, y \in \Omega$$

**Definition 3.17.** *We call  $\Delta_{w, f^{ns}}$  the non-stationary filtered non-local Laplacian obtained by applying the (regular) divergence to non-stationary filtered gradients*

$$(3.25) \quad (\Delta_{w, f^{ns}})(x) := (\operatorname{div}_w(\nabla_{w, f^{ns}i}))(x) \\ = \int_{\Omega} w(x, y) f^{ns}(i(y) - i(x), x, y) - w(y, x) f^{ns}(i(x) - i(y), y, x) dy.$$

**Lemma 3.18.** *Let  $f$  be a cross-symmetric non-stationary real-valued filter function, i.e. let  $f(z, x, y) = -f(-z, y, x)$ , and choose the weights  $w_s(x, y) = w_s(y, x)$  symmetrically. The associated non-stationary filtered non-local Laplacian  $\Delta_{w_s, f}$  is given by:*

$$(3.26) \quad (\Delta_{w_s, f})(x) = 2 \int_{\Omega} w_s(x, y) f(i(y) - i(x), x, y) dy.$$

*Proof.* The lemma is again an immediate result of the (anti-)symmetry properties of the weights and the filter-function. ■

**3.5.  $L^0$  sparse quasi-gradients.** Based on the preceding definitions of non-stationary filtered gradients, we can now make a formal connection between particular types of sparse gradient approximations and related thresholding filter functions. We first consider the  $L^0$ -gradient sparsity.

**Lemma 3.19.** *Let  $w_1$  and  $w_2$  be two non-negative weighting functions. We look for a vector  $\vec{q}_{L^0}$  which is  $L^0$  sparse as weighted by  $w_1$ , while the quasi-gradient  $\sqrt{w_2} \cdot \vec{q}_{L^0}$  remains close to the observed gradients  $\nabla_{w_2} i$ . This is the solution of the following optimization problem*

$$(3.27) \quad \vec{q}_{L^0} = \arg \min_{\vec{q}} \{ \lambda^2 \|\vec{q}\|_{0, w_1} + \|\sqrt{w_2} \cdot \vec{q} - \nabla_{w_2} i\|_2^2 \}$$

which has a closed form as a component-wise hard-thresholding applied to the non-local finite differences:

$$(3.28) \quad q_{L^0}(x, y) = S_\tau^h(i(y) - i(x)), \quad \text{where } \tau = \lambda \sqrt{\frac{w_1(x, y)}{w_2(x, y)}} \quad \text{and} \quad S_\tau^h(z) = \begin{cases} 0 & |z| \leq \tau \\ z & \text{otherwise} \end{cases}$$

*Proof.* We first proceed by rewriting the optimization problem component-wise, and explicitly using the definition of the  $L^0$ -cost function:

$$(3.29) \quad q_{L^0}(x, y) = \arg \min_{q \in \mathbb{R}} \left\{ \lambda^2 w_1(x, y) (1 - \delta(q)) + w_2(x, y) (q - (i(y) - i(x)))^2 \right\},$$

$$\forall (x, y) \in \Omega \times \Omega.$$

It is easy to see that the sub-differential of the expression to be minimized contains 0 in at most two points, i.e. there are at most two local minima (just one if they coincide), namely:

$$(3.30) \quad q_{L^0}(x, y) \in \{0, i(y) - i(x)\}.$$

Since the expression goes to  $+\infty$  for  $q \rightarrow \pm\infty$ , the global minimum is determined by comparing the cost associated with just these two candidates:

$$(3.31) \quad q_{L^0}(x, y) = \begin{cases} 0 & w_2(x, y)(i(y) - i(x))^2 \leq \lambda^2 w_1(x, y) \\ i(y) - i(x) & \text{otherwise,} \end{cases} \quad \forall (x, y) \in \Omega \times \Omega.$$

Identifying this expression with hard-thresholding completes the proof.  $\blacksquare$

**Remark 3.20.** *Let  $w_3$  be another weighting function, based on  $w_1$  and  $w_2$  as*

$$(3.32) \quad w_3(x, y) = \max(w_1(x, y), w_2(x, y)) \quad x, y \in \Omega$$

*The quasi-gradient  $\sqrt{w_3} \cdot \vec{q}_{L^0}$  is an instance of non-stationary filtered non-local gradient:*

$$(3.33) \quad (\sqrt{w_3} \cdot \vec{q}_{L^0})(x, y) = (\nabla_{w_3, f^{ns}} i)(x, y) = \sqrt{w_3(x, y)} f^{ns}(i(y) - i(x), x, y)$$

with

$$(3.34) \quad f^{ns} = S_\tau^h \quad \text{and} \quad \tau(x, y) = \lambda \sqrt{\frac{w_1(x, y)}{w_2(x, y)}}$$

**3.6.  $L^1$ /TV sparse quasi-gradients.** After making the relation between hard-thresholded gradients and  $L^0$ -sparsity, we now highlight the similar connection between soft-thresholding (shrinkage) and  $L^1$ /TV-sparsity.

**Lemma 3.21.** *Let  $w_1$  and  $w_2$  be two non-negative weighting functions. We look for a vector  $\vec{q}_{TV}$  such that the quasi-gradient  $\sqrt{w_1} \cdot \vec{q}_{TV}$  has small  $L^1$ -norm, while the quasi-gradient  $\sqrt{w_2} \cdot \vec{q}_{TV}$  remains close to the observed gradients  $\nabla_{w_2} i$ . More precisely, we are interested in the solution of the following convex optimization problem:*

$$(3.35) \quad \vec{q}_{TV} = \arg \min_{\vec{q}} \{2\lambda \|\sqrt{w_1} \cdot \vec{q}\|_1 + \|\sqrt{w_2} \cdot \vec{q} - \nabla_{w_2} i\|_2^2\}$$

It is found as component-wise soft-thresholding applied to the non-local finite differences:

$$(3.36) \quad q_{TV}(x, y) = S_\tau^s(i(y) - i(x)), \quad \text{where } \tau = \lambda \frac{\sqrt{w_1(x, y)}}{w_2(x, y)} \quad \text{and} \quad S_\tau^s(z) = \begin{cases} z + \tau & z < -\tau \\ 0 & |z| \leq \tau \\ z - \tau & z > \tau \end{cases}$$

*Proof.* For shorter notation, let us write  $\tau := \lambda \frac{\sqrt{w_1(x, y)}}{w_2(x, y)} \geq 0$ , and  $z := i(y) - i(x)$ . Again we rewrite the problem component-wise:

$$(3.37) \quad q_{TV}(x, y) = \arg \min_q \{f(q) := 2\tau|q| + (q - z)^2\}, \quad \forall (x, y) \in \Omega \times \Omega.$$

The cost functional  $f(q)$  is non-differentiable at  $q = 0$ , and it grows to infinity for  $q \rightarrow \pm\infty$ . As a consequence, it is minimized for a finite  $q^*$  such that the functional's sub-gradient contains 0 at that point,  $\partial_q f(q^*) \ni 0$ . The sub-differential is easily computed as

$$(3.38) \quad \partial_q f(q) = \begin{cases} 2(\tau + q - z) & q > 0 \\ [-\tau - z, \tau - z] & q = 0 \\ 2(-\tau + q - z) & q < 0. \end{cases}$$

There are, thus, three cases to consider:

**Case 1a:**  $z > \tau$ . Clearly, in this case 0 is not in the sub-differential at  $q = 0$ . Also, the differential does not vanish for any negative  $q$ . Instead,  $q = z - \tau$  is the only minimizer.

**Case 1b:**  $z < -\tau$ . Similar to case 1a, except that the only minimizer is found for  $q = z + \tau$ .

**Case 2:**  $z \in [-\tau, \tau]$ . In this case,  $0 \in \partial_q f(q = 0)$ , while no  $q \neq 0$  leads to a vanishing derivative. The minimizer is thus given by  $q = 0$ . In addition, we have:

$$(3.39) \quad f(0) = z^2$$

$$(3.40) \quad f(q > 0) - f(0) = 2\tau q + (z - q)^2 - z^2 = 2(\tau - z)q + q^2 > 0$$

$$(3.41) \quad f(q < 0) - f(0) = -2\tau q + (z - q)^2 - z^2 = 2(-\tau - z)q + q^2 > 0,$$

which confirms  $q = 0$  as only minimizer in this case. Identifying the three cases and their minimizers with soft-thresholding (shrinkage) completes the proof. ■

**Remark 3.22.** Let again  $w_3$  be the weighting function such that  $w_3 = \max(w_1, w_2)$ . The quasi-gradient  $\sqrt{w_3} \cdot \vec{q}_{TV}$  is an instance of non-stationary filtered non-local gradient:

$$(3.42) \quad (\sqrt{w_3} \cdot \vec{q}_{TV})(x, y) = (\nabla_{w_3, f^{ns}} i)(x, y) = \sqrt{w_3(x, y)} f^{ns}(i(y) - i(x), x, y)$$

with

$$(3.43) \quad f^{ns} = S_\tau^s \quad \text{and} \quad \tau(x, y) = \lambda \frac{\sqrt{w_1(x, y)}}{w_2(x, y)}$$

**3.7.  $L^2/H^1$  sparse quasi-gradients.** The next step is to consider  $L^2$  gradient “sparsity”:

**Lemma 3.23.** Let  $w_1$  and  $w_2$  be two non-negative weighting functions. We look for a vector  $\vec{q}_{H^1}$  such that the quasi-gradient  $\sqrt{w_1} \cdot \vec{q}_{H^1}$  has small  $L^2$ -norm, while the quasi-gradient  $\sqrt{w_2} \cdot \vec{q}_{H^1}$  remains close to the observed gradients  $\nabla_{w_2} i$ . This corresponds to the following quadratic, convex optimization problem:

$$(3.44) \quad \vec{q}_{H^1} = \arg \min_{\vec{q}} \{ \lambda \|\sqrt{w_1} \cdot \vec{q}\|_2^2 + \|\sqrt{w_2} \cdot \vec{q} - \nabla_{w_2} i\|_2^2 \}$$

Its solution is found as scaling applied to the non-local differences:

$$(3.45) \quad q_{H^1}(x, y) = S_\tau^u(i(y) - i(x)), \quad \text{where} \quad \tau = \lambda \frac{w_1(x, y)}{w_2(x, y)} \quad \text{and} \quad S_\tau^u(z) = \frac{z}{1 + \tau}$$

*Proof.* The cost function is fully differentiable and can be solved pointwise:

$$(3.46) \quad q_{H^1}(x, y) = \arg \min_q \{ \lambda w_1 q^2 + w_2 (q - (i(y) - i(x)))^2 \}, \quad \forall (x, y) \in \Omega \times \Omega.$$

Since the cost functional approaches  $+\infty$  for  $q \rightarrow \pm\infty$ , the minimum is achieved for vanishing first variation. Optimality thus requires from a minimizer  $q^*$ :

$$(3.47) \quad \left( \lambda \frac{w_1(x, y)}{w_2(x, y)} + 1 \right) q^* = i(y) - i(x),$$

from which the lemma is directly obtained. ■

**Remark 3.24.** Let again  $w_3$  be the weighting function such that  $w_3 = \max(w_1, w_2)$ . The quasi-gradient  $\sqrt{w_3} \cdot \vec{q}_{H^1}$  is an instance of non-stationary filtered non-local gradient:

$$(3.48) \quad (\sqrt{w_3} \cdot \vec{q}_{H^1})(x, y) = (\nabla_{w_3, f^{ns}} i)(x, y) = \sqrt{w_3(x, y)} f^{ns}(i(y) - i(x), x, y)$$

with

$$(3.49) \quad f^{ns} = S_\tau^u \quad \text{and} \quad \tau(x, y) = \lambda \frac{w_1(x, y)}{w_2(x, y)}$$

**Remark 3.25.** For the particular choice of identical weights  $w = w_1 \equiv w_2$ , we have  $w_3 = w$ , and the thresholdings in both  $L^0$  and  $L^2/H^1$  sparse quasi-gradients become a constant in space, that is  $\tau(x, y) = \lambda$  uniformly, which makes the filtering function stationary and anti-symmetric. If the weights are binary, then this extends to the  $L^1/TV$  case, as well. On the other hand, if the weights controlling the sparsity and the fidelity are different, then those thresholding functions  $\tau$  vary spatially, which results in adaptive thresholding. We will further discuss this in section 9.

**3.8. TV augmented quasi-gradients.** The last case deals with a gradient enhancement:

**Lemma 3.26.** *Let  $w_1$  and  $w_2$  be two non-negative weighting functions. We look for a vector  $\vec{q}^{TV}$  such that the quasi-gradient  $\sqrt{w_1} \cdot \vec{q}^{TV}$  has increased  $L^1$ -norm, while the quasi-gradient  $\sqrt{w_2} \cdot \vec{q}^{TV}$  remains close to the observed gradients  $\nabla_{w_2} i$ . It's the solution of the following optimization problem:*

$$(3.50) \quad \vec{q}^{TV} = \arg \min_{\vec{q}} \{-2\lambda \|\sqrt{w_1} \cdot \vec{q}\|_1 + \|\sqrt{w_2} \cdot \vec{q} - \nabla_{w_2} i\|_2^2\}$$

and is found as unshrinkage applied to the non-local differences:

$$(3.51) \quad q^{TV}(x, y) = S_{-\tau}^s(i(y) - i(x)), \quad \text{where } \tau = \lambda \frac{\sqrt{w_1(x, y)}}{w_2(x, y)} \quad \text{and} \quad S_{-\tau}^s(z) = \begin{cases} z + \tau & z > 0 \\ 0 & z = 0 \\ z - \tau & z < 0 \end{cases}$$

*Proof.* In this case again, we proceed by rewriting the problem component-wise, and substituting  $\tau := \lambda \frac{\sqrt{w_1(x, y)}}{w_2(x, y)}$  and  $z := i(y) - i(x)$ :

$$(3.52) \quad q^{TV}(x, y) = \arg \min_q \{f(q) := -2\tau|q| + (q - z)^2\}, \quad \forall (x, y) \in \Omega \times \Omega.$$

The cost function  $f(q)$  is non-differentiable at  $q = 0$ , and  $f(q) \rightarrow \infty$  as  $q \rightarrow \pm\infty$ . The subgradient of  $f$  is:

$$(3.53) \quad \partial_q f(q) = \begin{cases} 2(-\tau + q - z) & q > 0 \\ \emptyset & q = 0 \\ 2(\tau + q - z) & q < 0. \end{cases}$$

There are now 5 cases to be considered:

**Case 1a:**  $z \geq \tau$ . The only vanishing subgradient is found for  $q = z + \tau$ .

**Case 1b:**  $z \in (0, \tau)$ . There are two locations with vanishing subgradient, namely  $q = z \pm \tau$ , corresponding to two local minima, one of which is the minimizer. The global minimum is found by evaluating the cost function at these two locations:

$$(3.54) \quad f(z + \tau) = -2\tau(z + \tau) + \tau^2 = -2\tau z - \tau^2$$

$$(3.55) \quad f(z - \tau) = 2\tau(z - \tau) + (-\tau)^2 = +2\tau z - \tau^2,$$

therefore  $f(z + \tau) < f(z - \tau)$ , and the global minimizer is found as  $q = z + \tau$ .

**Case 2a:**  $z \leq -\tau$ . The only vanishing subgradient is found for  $q = z - \tau$ .

**Case 2b:**  $z \in (-\tau, 0)$ . As in case 1b, there are two local minima, at  $q = z \pm \tau$ . This time,  $f(z - \tau) < f(z + \tau)$  and  $q = z - \tau$  is the unique minimizer.

**Case 3:  $z = 0$ .** In this case, there are two equal minimizers,  $q = \pm\tau$ , since  $f(\tau) = f(-\tau) = -\tau^2$ , here.

The cases 1a through 2b combine into the gradient enhancement of the lemma. For the single point  $z = 0$  we are left with an ambiguity between two minimizers. *In order to resolve this ambiguity, we deliberately set the solution to  $q = 0$  for  $z = 0$ , albeit this is clearly not a minimizer, but a local maximum instead.* This choice is primarily motivated by the desire to obtain an odd filtering function, and since no image gradients are to be created out of nowhere.

■

**Remark 3.27.** Let again  $w_3$  be the weighting function such that  $w_3 = \max(w_1, w_2)$ . The quasi-gradient  $\sqrt{w_3} \cdot \bar{q}^{TV}$  is an instance of non-stationary filtered non-local gradient:

$$(3.56) \quad (\sqrt{w_3} \cdot \bar{q}^{TV})(x, y) = (\nabla_{w_3, f^{ns}i})(x, y) = \sqrt{w_3(x, y)} f^{ns}(i(y) - i(x), x, y)$$

with

$$(3.57) \quad f^{ns} = S_{-\tau}^s \quad \text{and} \quad \tau(x, y) = \lambda \frac{\sqrt{w_1(x, y)}}{w_2(x, y)}$$

**4. Computing the weights.** At this point, it is worthwhile spending some time on different choices for the weight vectors used in the non-local differential operators. We will discuss common choices such as local weights, patch-based non-local weights, semi-local Gaussian kernels, and finally we suggest the use of cosine-based distances.

**4.1. Local weights.** It is possible to construct a local weight vector that reproduces the finite difference scheme of standard local differential operators.

**Definition 4.1.** Indeed, we construct two-dimensional local weights between two points  $x$  and  $y$  as

$$(4.1) \quad w_\epsilon(x, y) := \frac{1}{\epsilon} \left[ \delta \left( x - y + \begin{pmatrix} \epsilon \\ 0 \end{pmatrix} \right) + \delta \left( x - y + \begin{pmatrix} 0 \\ \epsilon \end{pmatrix} \right) \right],$$

where  $\delta$  is the Dirac distribution. Then, in the limit  $\epsilon \rightarrow 0$ , the non-local derivative approaches the standard gradient:

$$(4.2) \quad \lim_{\epsilon \rightarrow 0} (\nabla_{w_\epsilon} u)(x) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \begin{pmatrix} u \left( x + \begin{pmatrix} \epsilon \\ 0 \end{pmatrix} \right) - u(x) \\ u \left( x + \begin{pmatrix} 0 \\ \epsilon \end{pmatrix} \right) - u(x) \end{pmatrix} = (\nabla u)(x)$$

**4.2. Semi-local Gaussian kernel.** The Gaussian kernel corresponds to a symmetrical extension and mollification of local weights.

**Definition 4.2.** The isotropic 2-dimensional normalized Gaussian kernel is given by:

$$(4.3) \quad w_g(x, y) := \frac{1}{2\pi\sigma^2} e^{-\frac{d(x, y)^2}{2\sigma^2}},$$

where  $d(x, y)$  denotes the Euclidean distance between points  $x$  and  $y$ , and  $\sigma$  is a scale factor.

**4.3. Patch-based non-local weights.** The non-local weights commonly used in imaging have been introduced by Buades, Coll and Morel in [13].

**Definition 4.3.** *Based on an image  $i(x)$ , we define non-local weights as*

$$(4.4) \quad w_{nl}(x, y) := e^{-\frac{d(I(x), I(y))^2}{h^2}} \quad \text{with} \quad d(I(x), I(y))^2 := \int_{\Omega} G_{\sigma}(t)(i(y+t) - i(x+t))^2 dt,$$

where  $h > 0$  is a scale parameter and  $G_{\sigma}$  is a Gaussian window with standard deviation  $\sigma$ . This weight is close to zero if the regions (patches) around  $x$  and  $y$ ,  $I(x)$  and  $I(y)$ , have an important Gaussian-weighted  $L^2$ -distance  $d(I(x), I(y))$ . Practically, for each pixel  $x$  we only calculate  $w_{nl}(x, y)$  for  $y$  in a small search window centered at  $x$ , keep only a few, large coefficients and discard the rest [13, 22, 46]. These weights are generally used today, and little attention is typically paid to where they come from.

However, an important connection has been shown by [2], where the exponential weights have been connected with maximum-entropy distributions known from statistical mechanics [32, 25]. Indeed, the Gaussian weights are maximum-entropy weights minimizing the non-local  $H^1$ -energy based on patch-distances  $d(I(x), I(y))$ . This argument suggests, that different weights are optimal when functionals other than non-local  $H^1$  are considered, such as non-local TV or non-local  $L^0$ , corresponding to differing choices of distances  $d$ .

**4.4. Cosine-distances.** Indeed, it must appear wrong to use (illumination-sensitive) non-local weights based on the input image, to measure the non-local  $H^1$  of the illumination invariant reflectance (or its fidelity to filtered gradients). Instead of computing and updating the weights on the current estimate of the reflectance, we suggest using illumination-robust weights stemming from the input image as a proxy, by normalizing patches by their  $L^2$ -norm before computing distances. Such normalized  $L^2$ -distance can be shown to be equivalent to a cosine distance:

$$(4.5) \quad d(I(x)/\|I(x)\|_2, I(y)/\|I(y)\|_2) = 1 - \langle I(x)/\|I(x)\|_2, I(y)/\|I(y)\|_2 \rangle = 1 - \cos \alpha_{xy},$$

where  $\alpha_{xy} := \angle(I(x), I(y))$  is the angle between two patches.

Further, exponentials of such cosine distances have the same low order terms in their series expansion as powers of the cosine:

$$(4.6) \quad e^{-\lambda(1-\cos \alpha_{xy})} \approx \cos^{\lambda} \alpha_{xy} \quad \alpha_{xy} \ll 1$$

**4.5. Color distances.** In a natural image, objects of similar material may have similar texture and thus small patch-angle. However, more likely pixels of the same material will have a very similar hue. Here, cosine distances come in very handy. Indeed, to measure the similarity in material between pixels, we use the cosine distance in RGB space. This comes from the observation that two pixels in RGB space have the same hue if the angle between the two corresponding RGB vectors is small—irrespective of their strength of illumination, encoded as magnitude. Given any two pixels  $x, y$  with the corresponding RGB values  $(I_1(x), I_2(x), I_3(x))$  and  $(I_1(y), I_2(y), I_3(y))$ , the angle between those two vectors,  $\alpha_{xy}$ , is computed via

$$(4.7) \quad \cos(\alpha_{xy}) := \frac{\langle I(x), I(y) \rangle}{\|I(x)\| \|I(y)\|} = \frac{I_1(x)I_1(y) + I_2(x)I_2(y) + I_3(x)I_3(y)}{\sqrt{I_1^2(x) + I_2^2(x) + I_3^2(x)} \sqrt{I_1^2(y) + I_2^2(y) + I_3^2(y)}}$$

**Definition 4.4.** Then we define the color/hue-based weight as follows:

$$(4.8) \quad w_c(x, y) = \begin{cases} \cos(\alpha_{xy}) & \text{if } \cos(\alpha_{xy}) \geq c \\ 0 & \text{otherwise,} \end{cases}$$

where  $0 < c < 1$  is a parameter close to one. Note that this definition of color weight in RGB space can be directly generalized to weights based on cosine-distances in hyperspectral images [3].

**5. Closing the gap between kernel and variational Retinex.** The link between PDE-Retinex (thresholded Poisson equations) and variational models has been shown before. The definitions of non-local differential operators will now allow us to make a formal connection between semi-local kernel-based and center-surround Retinex models on the one hand, and local variational as well as threshold-based models on the other hand. This will eventually provide a big umbrella under which all Retinex methods can be classified.

**5.1. Homomorphic filtering as variational problem.** We have already mentioned that the homomorphic filtering Retinex can be rewritten as a Gaussian-kernel  $w_g(x, y)$  based computation of the following form:

$$(5.1) \quad r(x) = i(x) - \sum_y w_g(x, y)i(y) = - \sum_y w_g(x, y) (i(y) - i(x)),$$

provided that the Gaussian kernel is normalized, i.e.  $\sum_y w_g(x, y) = 1$ . The second sum now clearly identifies with our definition of non-local Laplacian (see Definition 3.10 and (3.15)), and we may thus also write:

$$(5.2) \quad r(x) + \frac{1}{2} \Delta_{w_g} i(x) = 0$$

This in turn is the Euler-Lagrange equation corresponding to the following convex optimization problem:

$$(5.3) \quad \min_r \left\{ \underbrace{\|\nabla_{w_g} r - \nabla_{w_g} i\|_2^2}_A - \underbrace{\|\nabla_{w_g} r\|_2^2}_B + 2 \underbrace{\|r\|_2^2}_C \right\}$$

The interpretation of this variational model is as follows: We look for a reflectance  $r$  whose gradients are similar to those of  $i$  (A), but enhanced (B), while having minimal energy (C).

**5.2. Perceptual contrast enhancement and non-local derivatives.** In their award-winning model, Bertalmio and colleagues have used their kernel-based lightness estimate together with a grey-world prior and a fidelity constraint to build a ‘‘perceptually inspired variational framework’’ for image enhancement [4, 59]. Their anti-symmetrized kernel-based Retinex has a variational formulation, which is very close to the ACE model [5], namely:

$$(5.4) \quad \min_R \left\{ \int_{\Omega} \left[ \alpha (R(x) - 1/2)^2 + \beta (R(x) - I(x))^2 \right] dx + C \frac{\min}{\sqrt{w}}(R) \right\}$$

where  $C_{\sqrt{w}}^{\frac{\min}{\max}}(R)$  is a contrast function. For particular, but reasonable choices of the contrast functions, the contrast term can be written as

$$(5.5) \quad C_{\sqrt{w}}^{\frac{\min}{\max}}(R) \equiv - \int_{\Omega \times \Omega} \sqrt{w(x,y)} |f_a(r(y) - r(x))| dx dy = -\|\nabla_{w,f_a} r\|_1$$

In particular,  $f_a$  may be the identity. Thus, we rewrite the the perceptual contrast enhancement in terms of non-local derivatives as follows:

$$(5.6) \quad \min_{r=\log(R)} \left\{ \alpha \|R - 1/2\|_2^2 + \beta \|R - I\|_2^2 - \|\nabla_w r\|_1 \right\}$$

where the first term represents the gray-world prior, the second is a fidelity term with respect to the observed intensity, and the contrast term increases non-local TV of the reflectance.

**6. Non-local Retinex (proposed model).** So far we have seen that all Retinex models have a variational cousin, potentially through the use of non-local differential operators. Even more, these variational counterparts all share a very similar structure: the energy typically comprises one or two fidelity terms (image and/or its gradient), as well as sparsity priors or alternatively, through negation, enhancement terms.

Also, we have shown that this type of variational problem can be retrofitted into a Horn/Morel-style gradient-fidelity problem (PDE-Retinex), potentially adding further terms. In particular, we have shown in the previous section, how different gradient sparsity and fidelity terms translate into different associated thresholding functions.

Here, we want to formulate this retrofitted PDE-Retinex model as a general recipe. We tackle the Retinex problem in a two step approach:

*1) Gradient filtering.* We realize that the reflectance obeys both to some gradient sparsity priors and some gradient fidelity priors. In a first step, we thus look for an optimal quasi-gradient that best satisfies those two constraints. This quasi-gradient is obtained as filtered gradient of the observed image  $\nabla_{w,f}i$ . Here, we write  $w$  as short for  $w_3$ , since both  $w_1$  and  $w_2$  are entirely hidden within the generic gradient filter  $f$ , inspired by the threshold employed by Horn and Morel [29, 53].

*2) Gradient fitting.* However, the resulting quasi-gradient almost certainly is not a valid gradient by itself, and we wish to fit a reflectance, whose gradient comes closest to the quasi-gradient determined in the first step, while possibly respecting some additional constraints:

$$(6.1) \quad \hat{r} = \arg \min_r \left\{ \|\nabla_w r - \nabla_{w,f}i\|_p^p + \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 \right\}$$

The sparsity and gradient fidelity terms of the first step will determine the exact filter function  $f$  to be used, while the sparsity/smoothness priors on the illumination will essentially govern the gradient fidelity norm  $p$  of the second step.

The interest of such a two step procedure is manifold: First, each step, i.e. thresholding the input gradient, followed by a gradient fitting is relatively simple to compute, compared to the non-compacted variational model. Further, the computational tools required to solve the gradient reconstruction step become independent of the gradient sparsity imposed.

Finally, this two step procedure is fully compatible with Marr’s theory of lightness computation in the (primate) retina, given in (2.6) and (2.7) [49]. Beyond, it particularly neatly fits into Wilson-Cowan equations, modeling large-scale activity in cortical neural populations [76, 77], as will be shown below.

**6.1. Numerical optimization.** In the following, we present basic solvers used for proof of concept. Note that solvers with better efficiency could easily be devised, but doing so is beyond the scope of this manuscript.

**6.1.1.  $L^2$  gradient fidelity.** The  $L^2$ -based problem is differentiable and we propose to solve its Euler-Lagrange equations. The energy of the  $L^2$  gradient-fidelity non-local Retinex is

$$J(r) = \|\nabla_w r - \nabla_{w,f} i\|_2^2 + \alpha \|r\|_2^2 + \beta \|r - i\|_2^2$$

The corresponding Euler-Lagrange equations are

$$(6.2) \quad 0 = 2(-\Delta_w \hat{r} + \Delta_{w,f} i + \alpha \hat{r} + \beta(\hat{r} - i)),$$

and we recover an estimate of the reflectance,  $\hat{r}$ , as

$$(6.3) \quad \hat{r} = ((\alpha + \beta)I - L)^{-1} (\beta i - \Delta_{w,f} i),$$

where  $I$  is the identity matrix, and  $L$  is the Laplacian matrix derived from the weights  $w(x, y)$ :

$$(6.4) \quad L_{xy} = \begin{cases} w(x, y) + w(y, x) & x \neq y \\ -\sum_z w(x, z) + w(z, x) & x = y. \end{cases}$$

**Remark 6.1.** *Since the graph Laplacian  $L$  is negative semi-definite, the operator  $(\alpha + \beta)I - L$  is diagonally dominant. Therefore, we can solve for  $\hat{r}$  either by a Gauss-Seidel algorithm or conjugate gradient method with a few iterations.*

**Remark 6.2.** *Note that the problem at hand strongly resembles the  $L^2$  statistical ranking problem [33, 28, 57], from which alternative optimization strategies could be inspired.*

**6.1.2. Wilson-Cowan equations.** Originally, the Wilson-Cowan equations [76, 77] provided a description of the (temporal) evolution of the coarse-scale, mean activity of a population of both inhibitory and excitatory neurons in the cortex [6]. The equations have later been generalized in order to model the spatio-temporal distribution and patterns of excitation in the visual cortex (V1) [12]. A first connection between the Wilson-Cowan equations and Retinex was proposed by Cowan and Bressloff, where it was shown that Marr’s Retinex model could actually be written in terms of Wilson-Cowan equations, and thus be implemented by corresponding neuron populations [49, 16].

**Definition 6.3.** *The relevant descriptor of mean activity,  $a(x, \phi, t)$ , at cortical coordinates  $x$ , orientation preference  $\phi$  is given by the following integro-differential equation (IDE) [12, 6]:*

$$(6.5) \quad \frac{\partial a(x, \phi, t)}{\partial t} = -\nu a(x, \phi, t) + \mu \int_o^\pi \int_\Omega w(x, \phi, y, \varphi) \sigma [a(y, \varphi, t)] dyd\varphi + \lambda h(x, \phi, t)$$

where  $h(x, \phi, t)$  is the external stimulus,  $w$  is a kernel that decays with the distances  $|x - y|$  and  $|\phi - \varphi|$ , and  $\sigma$  is a sigmoid function.

**Definition 6.4.** Like previous authors, we ignore the orientation  $\phi$  and assume the external stimuli to be constant in time:

$$(6.6) \quad \frac{\partial a(x, t)}{\partial t} = -\nu a(x, t) + \mu \int_{\Omega} w(x, y) \sigma [a(y, t)] dy + \lambda h(x)$$

**Lemma 6.5.** The proposed  $L^2$ -based non-local Retinex model is a steady-state of the above orientation-insensitive Wilson-Cowan equations.

*Proof.* First, let us write the gradient-descent equation associated with the first variation

$$(6.7) \quad \begin{aligned} \frac{\partial r(x, t)}{\partial t} = & 2 \int_{\Omega} w(x, y) (r(y, t) - r(x, t)) dy \\ & - 2 \int_{\Omega} w(x, y) f(i(y) - i(x)) dy - 2\alpha r(x, t) - 2\beta (r(x, t) - i(x)) \end{aligned}$$

After rearrangement,

$$(6.8) \quad \begin{aligned} \frac{\partial r(x, t)}{\partial t} = & -2 \left( \alpha + \beta + \int_{\Omega} w(x, y) dy \right) r(x, t) + 2 \int_{\Omega} w(x, y) r(y, t) dy \\ & + \beta i(x) - 2 \int_{\Omega} w(x, y) f(i(y) - i(x)) dy, \end{aligned}$$

the individual terms can easily be identified with the elements of the Wilson-Cowan equation:

$$(6.9) \quad \nu = 2\alpha + 2\beta + 2 \int_{\Omega} w(x, y) dy$$

$$(6.10) \quad \mu = 2$$

$$(6.11) \quad \sigma[r] = r$$

$$(6.12) \quad \lambda h(x) = \beta i(x) - 2 \int_{\Omega} w(x, y) f(i(y) - i(x)) dy$$

where the first coefficient can easily be made constant by imposing kernel normality, i.e.  $\forall x : \int_{\Omega} w(x, y) dy = 1$ . Further, the central term clearly identifies as lateral excitation, whereas the input stimulus  $i(x)$  is affected by lateral inhibition through the (thresholded) gradients  $\int_{\Omega} w(x, y) f(i(y) - i(x)) dy$ .

As usual, the solution of the initial minimization problem is associated with the steady state of its gradient descent, and since the gradient descent corresponds to a Wilson-Cowan IDE, therefore, the  $L^2$  non-local gradient Retinex is a steady state of the corresponding Wilson-Cowan IDE. ■

**Remark 6.6.** It is to note that a similar claim was made in [6], where the perceptually inspired variational Retinex model is associated with a Wilson-Cowan IDE. However, there, the claim involves the sigmoid function  $\sigma$  to be both spatially varying and depending on  $r(x, t)$ ; a complication which is not required in the Retinex IDE proposed here.

**6.1.3.  $L^1$  gradient fidelity.** Let us now consider the optimization of the proposed gradient fitting after filtering in the  $p = 1$  case. We would like to point out that this problem is formally equivalent to the  $L^1$  statistical ranking problem, for which efficient graph-cuts-based solvers have been developed [58]. Here, as simpler yet less efficient alternative for illustrative purposes we might as well explore a more intuitively accessible split Bregman/augmented Lagrangian based approach, as outlined in the following paragraphs [7, 23, 55].

The  $L^1$ -based problem writes:

$$(6.13) \quad \min_r \left\{ \|\nabla_w r - \nabla_{w,f} i\|_1 + \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 \right\}$$

which we may split into the following linearly constraint minimization problem over two variables:

$$(6.14) \quad \min_{r,e} \left\{ \|e - \nabla_{w,f} i\|_1 + \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 \right\} \quad \text{s.t.} \quad e = \nabla_w r$$

We address the constraint by introducing the following augmented Lagrangian, which includes the constraint as both a quadratic penalty and Lagrangian multiplier term:

$$(6.15) \quad AL(r, e, \mu) = \|e - \nabla_{w,f} i\|_1 + \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 + \rho \|\nabla_w r - e\|_2^2 + 2 \langle \mu, \nabla_w r - e \rangle$$

The  $L^1$  minimization problem can now be solved by iteratively finding a saddle point to this augmented Lagrangian, iteratively along the different problem dimensions (ADMM). This process essentially involves iteration of three steps: solving the  $L^2$  minimization in  $r$ , shrinkage of  $e$ , updating the Lagrangian multiplier  $\mu$ :

$$(6.16) \quad \left\{ \begin{array}{l} r^{k+1} = \arg \min_r \left\{ \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 + \rho \left\| \nabla_w r - e^k + \frac{\mu^k}{\rho} \right\|_2^2 \right\} \\ \quad = ((\alpha + \beta)I - \rho L)^{-1} \left( \beta i - \rho \operatorname{div}_w \left( e^k - \frac{\mu^k}{\rho} \right) \right) \quad (L^2\text{-Retinex}) \\ e^{k+1} = \arg \min_r \left\{ \|e - \nabla_{w,f} i\|_1 + \rho \left\| e - \nabla_w r^{k+1} - \frac{\mu^k}{\rho} \right\|_2^2 \right\} \\ \quad = S_{1/2\rho}^s \left( \nabla_w r^{k+1} - \nabla_{w,f} i + \frac{\mu^k}{\rho} \right) + \nabla_{w,f} i \quad (\text{Shrinkage}) \\ \mu^{k+1} = \mu^k + \rho \left( \nabla_w r^{k+1} - e^{k+1} \right) \quad (\text{Dual ascent}) \end{array} \right.$$

**Remark 6.7.** *The most time consuming part is the update  $r^{k+1}$ . To speed up, we use a fixed  $\rho$  for every iteration as suggested by the split Bregman method [24], resulting in repeated inversion using the same system. Also, since the system matrix  $((\alpha + \beta)I - \rho L)$  is sparse and strictly positive, we can use the Gauss-Seidel or conjugate gradient methods to solve  $r^{k+1}$  with a few iterations.*

**6.1.4.  $L^0$  gradient fidelity.** Finally, let us look at the non-convex  $L^0$ -based gradient-fidelity optimization problem. This  $L^0$ -based problem writes:

$$(6.17) \quad \min_r \left\{ \|\nabla_w r - \nabla_{w,f} i\|_0 + \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 \right\}$$

This problem is hard to solve. We propose to treat it very similarly to the above  $L^1$ -problem and first split it into the following linearly constraint minimization problem over two variables:

$$(6.18) \quad \min_{r,e} \left\{ \|e - \nabla_{w,f} i\|_0 + \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 \right\} \quad \text{s.t.} \quad e = \nabla_w r$$

The resulting sub-optimization problems are now easy to solve, but we have no guarantee for overall convergence. Indeed, we address this problem in analogy to the  $L^1$  gradient fidelity, introducing both quadratic penalty and Lagrangian multiplier, and solve iteratively along each direction, where the only change occurs in the  $e$ -step:

$$(6.19) \quad \left\{ \begin{array}{l} r^{k+1} = \arg \min_r \left\{ \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 + \rho^k \left\| \nabla_w r - e^k + \frac{\mu^k}{\rho^k} \right\|_2^2 \right\} \\ \quad = \left( (\alpha + \beta)I - \rho^k L \right)^{-1} \left( \beta i - \rho^k \operatorname{div}_w \left( e - \frac{\mu^k}{\rho^k} \right) \right) \quad (L^2\text{-Retinex}) \\ e^{k+1} = \arg \min_r \left\{ \|e - \nabla_{w,f} i\|_0 + \rho^k \left\| e - \nabla_w r^{k+1} - \frac{\mu^k}{\rho^k} \right\|_2^2 \right\} \\ \quad = S_{1/\sqrt{\rho^k}}^h \left( \nabla_w r^{k+1} - \nabla_{w,f} i + \frac{\mu^k}{\rho^k} \right) + \nabla_{w,f} i \quad (\text{Thresholding}) \\ \mu^{k+1} = \mu^k + \rho^k \left( \nabla_w r^{k+1} - e^{k+1} \right) \quad (\text{Dual ascent}) \\ \rho^{k+1} = \rho^k \cdot s \quad (\text{Step reduction}) \end{array} \right.$$

where  $s > 1$  is a constant that essentially reduces the step size at each iteration and causes the iterative process to stabilize eventually.

**Remark 6.8.** *Note that the  $L^0$ -optimization problem is not convex. In order to enforce convergence, we add one more step to the iterative process, that is decreasing the effective step-size of the variable-split at every iteration. As a ramification, however, the matrix  $(\alpha + \beta)I - \rho L$  is not constant, and the  $r^{k+1}$ -update is less efficient. Moreover, said system-matrix becomes singular at some point. Numerically, we will converge sooner and stop before that. The proof of the convergence of this scheme is still an open question. However, we can see later that despite rigorous theory, this model offers interesting new applications of Retinex in shadow detection with interesting, stable results.*

**7. Results I: Relations to existing models.** The first results section is dedicated to demonstrate the unifying power of the proposed non-local two-step Retinex model.

**7.1. Model correspondences.** In the following paragraphs, we want to show how the existing Retinex implementations can be reproduced in our proposed, fundamental non-local Retinex model. In all these models, we can restrict ourselves to identical weights  $w = w_1 = w_2 = w_3$ . The different correspondences are summarized in table 7.1.

Table 7.1: Filtered non-local gradient-fidelity based approximations to existing Retinex models. Both Poisson PDE [53] and  $L^1$ -Retinex [45] employ gradient filtering natively. For the other methods, the filtered gradient reproduces a gradient sparsity term.

Model	Norm $p$	Weights $w$	Filter $f_a$	Additional terms
Poisson PDE [53]	$L^2$	local	$S_\lambda^h$	—
$L^1$ -Retinex [45]	$L^1$	local	$S_\lambda^h$	—
TV-Retinex [46]	$L^2$	local	$S_\lambda^s$	—
Variational Retinex [36]	$L^2$	local	$S_\lambda^u$	$\alpha \ r\ _2^2 \quad (r \leq 0)$
TV-Retinex [54]	$L^2$	local	$S_\lambda^s$	$\beta \ r - i\ _2^2$
TV- $L^1$ [15]	$L^1$	local	—	$\alpha \ r\ _2^2 \quad (\alpha \ r\ _1)$
Random walk/Kernel [59]	$L^2$	Gaussian	$S_{-\lambda}^s$	$\alpha \ r\ _2^2 + \beta \ r - i\ _2^2$

**7.1.1. Poisson.** The PDE version of Retinex [53] can be derived exactly from the  $L^2$ -version,  $p = 2$ , of the proposed Retinex model, under local weights  $w_l$  and gradient thresholding  $f_a = S_\lambda^h$ . Indeed:

$$(7.1) \quad \hat{r} = \arg \min_r \{ \|\nabla_{w_l} r - \nabla_{w_l, f_a} i\|_2^2 \} \quad (L^2\text{-Retinex})$$

implies the Euler-Lagrange equations

$$(7.2) \quad (\Delta \hat{r})(x, y) = (\Delta_{w_l, f_a} i)(x, y) \quad x, y \in \Omega. \quad (\text{Poisson PDE Retinex})$$

**7.1.2.  $L^1$ -Retinex.** The next close relative of the proposed non-local Retinex model is its local  $L^1$  predecessor,  $L^1$ -Retinex [45]. The closest match to  $L^1$ -Retinex in the proposed framework is obtained if we choose the weights  $w_l(x, y)$  such as to reproduce the well-known local finite differences differential operators, gradient filtering  $f_a = S_\lambda^h$ , and with  $p = 1$ :

$$(7.3) \quad \hat{r} = \arg \min_r \{ \|\nabla_{w_l} r - \nabla_{w_l, f_a} i\|_1 \} = \arg \min_r \left\{ \int_\Omega \sum_{k=1}^n \left| \nabla_k r(x) - S_\lambda^h(\nabla_k i(x)) \right| dx \right\}$$

which is the anisotropic  $L^1$ -distance for local gradient fidelity.

**7.1.3. TV regularized Retinex.** In [46], the authors propose to solve directly for an image, whose gradient is close to the observed gradient in  $L^2$ , while minimizing isotropic TV:

$$(7.4) \quad \hat{r} = \arg \min_r \{ \|\nabla r - \nabla i\|_2^2 + 2\lambda \|\nabla r\|_2 \}_1, \quad (\text{TV Retinex})$$

This is the constraint  $L^1$ -relaxed gradient sparsity assumption, solved through Bregman iterations:

$$(7.5) \quad \hat{r} = \arg \min_r \{ \|q - \nabla i\|_2^2 + 2\lambda \|q\|_2 \}_1 \quad \text{s.t.} \quad q = \nabla r \quad (\text{Bregman TV Retinex})$$

A similar model can be obtained through the proposed general Retinex model by employing soft-shrinkage gradient filtering,  $f_a = S_\lambda^s$ , to which the according potential is recovered:

$$(7.6) \quad \hat{r} = \arg \min_r \left\{ \|\nabla_{w_l} r - \nabla_{w_l, f_a} i\|_2^2 \right\}$$

Again, the main difference is the use of anisotropic TV through gradient filtering in the proposed framework.

**7.1.4.  $H^1+L^2$  regularized.** The variational Retinex model in [36] can be rewritten exactly as

$$(7.7) \quad \hat{r} = \arg \min_r \left\{ \|\nabla r - \nabla i\|_2^2 + \alpha \|r\|_2^2 + \lambda \|\nabla r\|_2^2 \right\} \quad \text{s.t.} \quad r \leq 0 \quad (H^1/L^2 \text{ Retinex})$$

In [36], the authors motivate the  $L^2$  term mainly as “a regularization of the problem that makes it better conditioned”, and they state that “in practice this penalty term should be weak [...] and  $\alpha$  should therefore be very small.” The constraint  $r < 0$  corresponds to the reset in the original Retinex theory. The constraint and  $L^2$  norm together push the reflectance close to white.

We may find a similar problem within the proposed framework, where we choose uniform gradient scaling  $f_a = S_\lambda^u$  and omit the clipping constraint:

$$(7.8) \quad \hat{r} = \arg \min_r \left\{ \|\nabla_{w_l} r - \nabla_{w_l, f_a} i\|_2^2 + \alpha \|r\|_2^2 \right\}$$

**7.1.5. TV+ $L^2$  regularized.** Recently, a mixture of TV regularized and Kimmel’s variational approach was proposed [54]. This model essentially boils down to:

$$(7.9) \quad \hat{r} = \arg \min_r \left\{ \|\nabla r - \nabla i\|_2^2 + \beta \|r - i\|_2^2 + 2\lambda \|\nabla r\|_1 \right\} \quad (\text{TV}/L^2 \text{ Retinex})$$

Again, we may approximate this model with a similar energy based on similarity to filtered gradients, with  $f_a = S_\lambda^s$ :

$$(7.10) \quad \hat{r} = \arg \min_r \left\{ \|\nabla_{w_l} r - \nabla_{w_l, f_a} i\|_2^2 + \beta \|r - i\|_2^2 \right\}$$

The main difference is the use of anisotropic TV through gradient filtering in the proposed framework.

**7.1.6. TV- $L^1$ .** The “logarithmic total variation” (LTV) model was suggested for extraction of illumination invariant features for face recognition [15]. It is defined as an TV- $L^1$  based on the logarithmic input image and its logarithmic illumination:

$$(7.11) \quad \hat{r} = \arg \min_r \left\{ \|\nabla r - \nabla i\|_1 + \alpha \|r\|_1 \right\} \quad (\text{TV-}L^1)$$

Its equivalent in the proposed framework is found by relaxing the second term to an  $L^2$ -norm, i.e. TV- $L^2$  Retinex.

**7.1.7. Bertalmío.** To approximate the perceptually inspired Retinex model through our proposed general framework, we replace or complete the intensity fidelity by a gradient fidelity:

$$(7.12) \quad \hat{r} = \arg \min_r \left\{ \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 + \|\nabla_w r - \nabla_w i\|_2^2 - 2\lambda \|\nabla_w r\|_1 \right\} \quad (-\text{TV}+L^2 \text{ Retinex})$$

This is essentially homomorphic filtering with TV in place of  $H^1$ . Again, we may now substitute by incorporating the TV-enhancement term as an input-gradient filter  $f_a = S_{-\lambda}^s$ :

$$(7.13) \quad \hat{r} = \arg \min_r \left\{ \|\nabla_w r - \nabla_{w, f_a} i\|_2^2 + \alpha \|r\|_2^2 + \beta \|r - i\|_2^2 \right\}$$

**7.2. Examples.** In the following we provide a range of example results obtained using the proposed non-local Retinex model to illustrate the broad spectrum of Retinex flavors it includes.

**7.2.1. The Logvinenko illusion.** We have applied the whole range of Retinex “modes” retrofitted above to existing Retinex implementations to a single common test image extracted from the Logvinenko illusion pattern [44]. The test image is shown in Fig. 7.1a). The illusion consists of the following: due to the suggested smoothly varying lighting, the oblique grey diamonds of the upper row appear darker than the diamonds of the lower row. However, as shown in the adjacent Fig. 7.1b), their actual intensity is exactly equal. In this example, the Retinex model is expected to separate the almost smooth shading from the rough checkerboard-like reflectance, thereby truly rendering the two rows of diamonds at different reflectances.

The first model,  $L^2$ -Retinex equivalent to [53], produces the standard result in Fig. 7.1c). It can be clearly seen that in particular the lower row of diamonds is not recovered completely flat, since the illumination is not smooth everywhere. The related  $L^1$ -Retinex in Fig. 7.1d), inspired by [45], suffers from very similar artifacts. In Fig. 7.1e) we show the results of our model with parameters set to correspond to TV-regularized Retinex [46], resulting in less artifacts. Adding an  $L^2$  fidelity-constraint ( $\beta > 0$ ), as in [54], injects more of the initial shading into the estimated reflectance, see Fig. 7.1f).

The TV- $L^1$ -inspired model [15] is in our case an TV- $L^2$  model for illumination recovery, where the TV-sparsity of the extracted illumination is tuned by the parameter  $\alpha$ . It is clearly appreciated in Fig. 7.1g–j) that the impact of the parameter is quite severe, with higher values corresponding to the output desired for illumination invariant feature extraction. The choice of parameters inspired by Kimmel’s Retinex formulation yields the output shown in Fig. 7.1k–m), which corresponds well to the behavior expected from [36]. The parameter  $\alpha$  controls the degree of dynamic range compression applied, i.e. the dominance of local contrast enhancement. Finally, in Fig. 7.1n–p) we provide the output produced by model parameters mimicking Bertalmío’s perceptually inspired Retinex [4]. Here, the unshrinking of the gradients has the unpleasant effect of amplifying pixel noise.

**7.2.2. Other applications.** To show the “unifying power” of the proposed Retinex model, we provide a few more illustrating examples in Fig. 7.2.

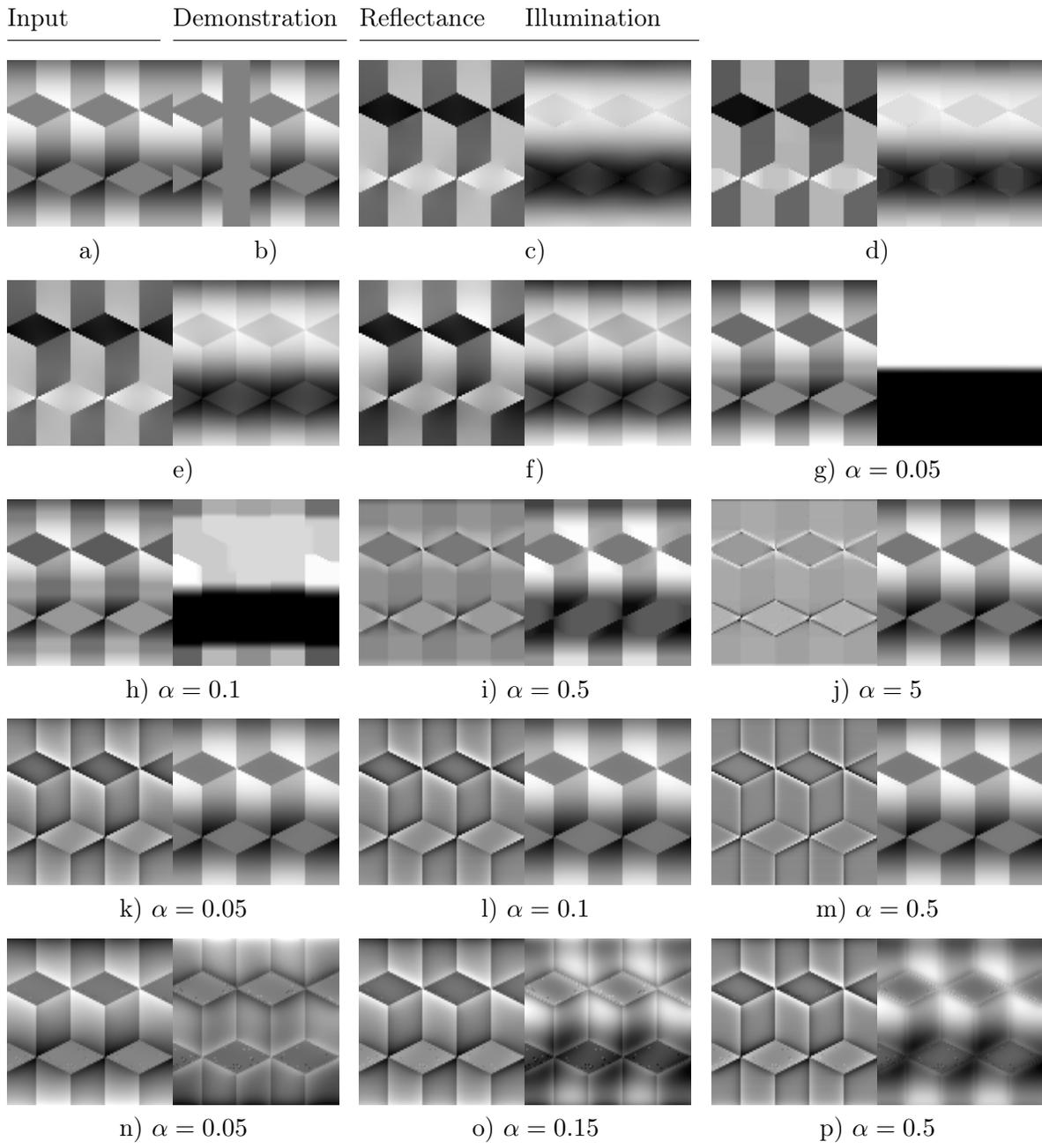


Figure 7.1: **Logvinenko illusion and different Retinex decompositions.** **a)** Input image. **b)** Demonstration of the illusion: despite the appearances, the “horizontal squares” actually have equal intensity. Retinex is supposed to reproduce this illusion of intensity difference. **c)** Reflectance and illumination recovered using the  $L^2$ -Retinex (hard thresholding,  $p = 2$ ,  $\alpha = \beta = 0$ ). **d)**  $L^1$ -Retinex (hard thresholding,  $p = 1$ ,  $\alpha = \beta = 0$ ). **e)** TV-regularized Retinex (soft thresholding,  $p = 2$ ,  $\alpha = \beta = 0$ ). **f)** Ng-Wang-like Retinex (soft thresholding,  $p = 2$ ,  $\alpha = 0$ ,  $\beta = 0.0015$ ). **g–j)** TV- $L^2$  Retinex (no thresholding,  $p = 1$ ,  $\alpha > 0$ ,  $\beta = 0$ ). **k–m)** Kimmel-like Retinex (gradient scaling,  $p = 2$ ,  $\alpha > 0$ ,  $\beta = 0$ ). **n–p)** Bertalmio-like Retinex (Gaussian kernel weights, gradient unshrinkage,  $p = 2$ ,  $\alpha > 0$ ,  $\beta = 0.002$ ). **g–p)** Where applicable,  $\alpha$  drastically tunes the amount of dynamic range compression.

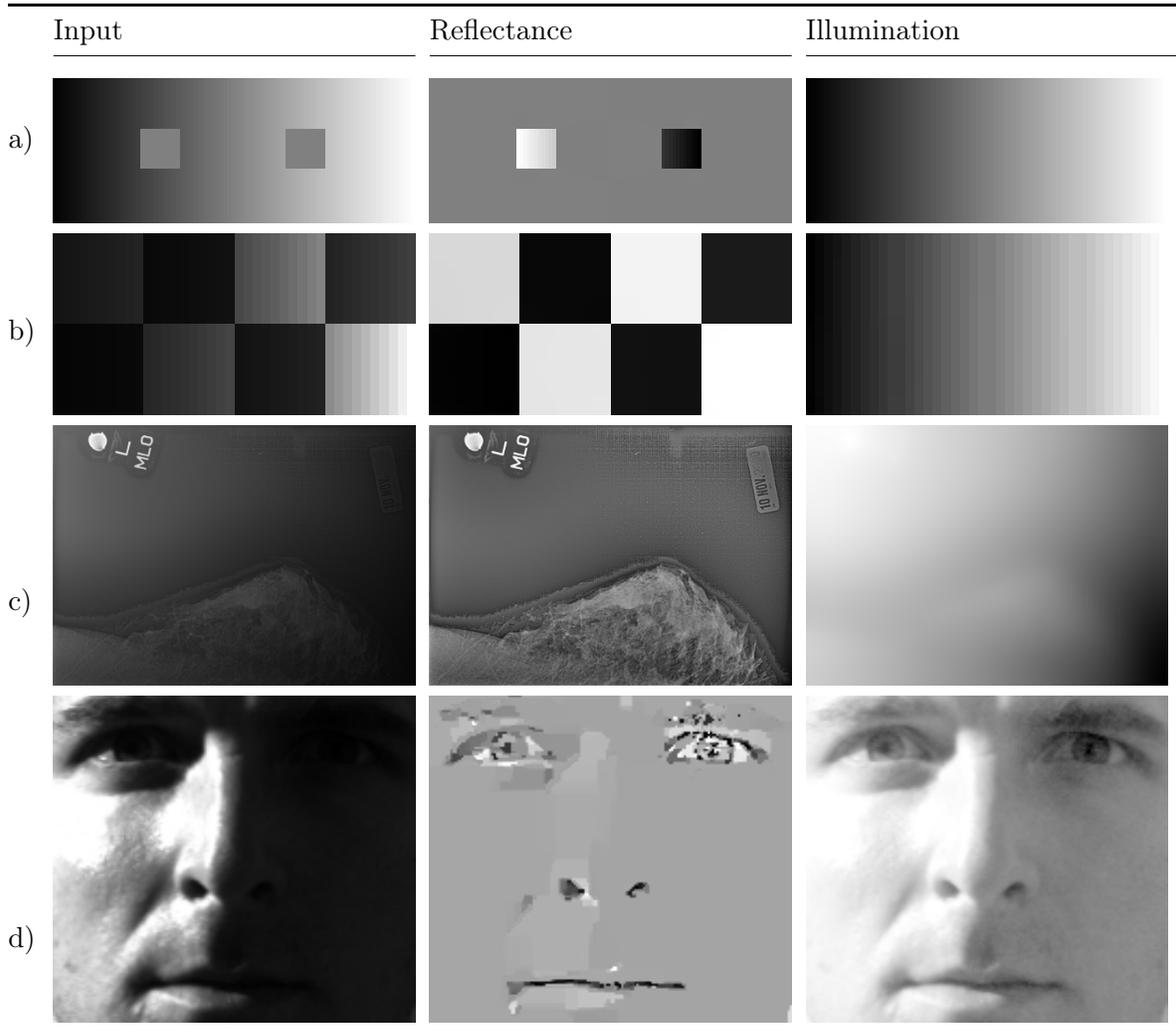


Figure 7.2: **Unifying non-local Retinex model.** a–b) Our unifying model decomposes an input image into underlying reflectance and estimated illumination, and successfully reproduces basic Retinex behavior (b/w squares and checkerboard, respectively). c) The same model allows dynamic range compression and local contrast enhancement (here: radiography), as well as d) illumination-invariant feature extraction, e.g. for face detection.

First, in Fig. 7.2a–b) two standard results are shown, that are based on  $L^2$ -gradient thresholding. We unsurprisingly succeed in separating the smooth gradient illumination from the sharp reflectance features.

The third example in Fig. 7.2c) is an (artificially) unevenly exposed radiography, where important features are masked due to the great dynamic range. We perform center-surround-like exposure correction and dynamic range compression simply by choosing  $p = 2$ , wide Gaussian kernel based weights, no thresholding, and  $\alpha = 0.01$ .



Figure 7.3: **HSV color Retinex.** Retinex is performed on the value-channel of a color image in HSV space. The final output is obtained by applying post-processing enhancement as in [36].

The fourth example is based on the challenge of illumination-invariant feature extraction for face recognition [15]. For this, we use  $p = 1$ , soft-thresholding  $\lambda = 0.2$ , and dynamic range compression  $\alpha = 5$ .

In Fig. 7.3 we provide an example inspired by Kimmel’s variational Retinex [36], where color images are involved. Here, Retinex is applied to the V-channel (lightness) of the color image (as opposed to separately on each RGB channel, as originally proposed), and we use matching parameters in our proposed framework. After application of their suggested post-processing steps, we obtain images very similar to Kimmel’s original results.

**8. Results II: New perspectives.** Beyond reproducing existing Retinex models, our proposed framework also has the potential to yield new results thanks to its generalizing power. In the next sections, we explore a few new possibilities offered by choosing new sets of parameters, in particular based on  $p = 0$  gradient fidelity, with applications to shadow detection and removal, and cartoon-texture decomposition.

**8.1.  $L^0$  gradient fidelity.** In Fig. 7.1 we have shown a series of decomposition results obtained with different model configurations. The best results in terms of piecewise constant reflectance versus illumination have been achieved with the basic hard thresholding models ( $L^2$ - and  $L^1$ -Retinex), as well as the soft-thresholding based TV-Retinex. However, all these models suffer from artifacts of illumination estimation at the edges and corners of the flat diamonds, where illumination smoothness is not a stringent enough prior. Therefore, we propose seeking for further illumination gradient penalty by choosing  $p = 0$ , corresponding to  $L^0$  gradient fidelity (as opposed to TV- or  $H^1$ -sparsity of the illumination). In Fig. 8.1 we show a few results where we make use of TV-regularization of the reflectance (soft thresholding). In particular in combination with Gaussian kernel weights, the decomposition exhibits less artifacts than previous results, see Fig. 8.1b).

**8.2. Shadows in artificial images: the Adelson checker illusion.** Another instance, where  $H^1$  smoothness ( $p = 2$ ) or TV-regularity ( $p = 1$ ) constraints are not sparse enough priors for the illumination field, is the Adelson checker illusion. We show a grayscale image of the

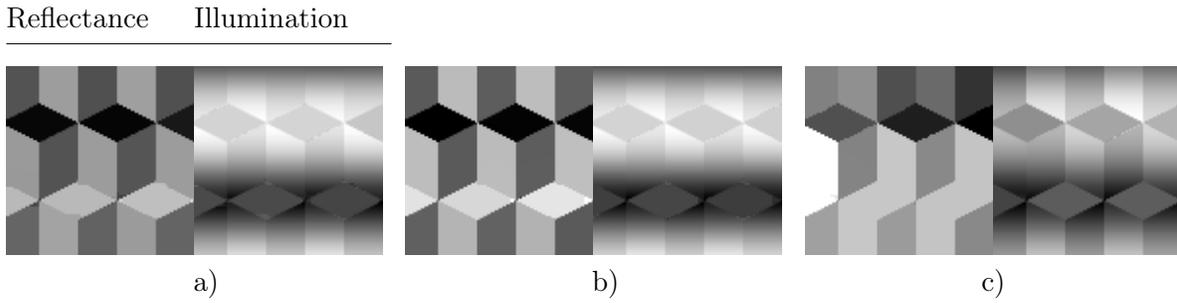


Figure 8.1: **Logvinenko illusion and new  $L^0$ -based Retinex decompositions.** Soft thresholding,  $p = 0$ ,  $\alpha = \beta = 0$ . **a)** Local weights. **b)** Narrow Gaussian kernel weights,  $\lambda = 0.33$ . **c)** Wider Gaussian kernel weights,  $\lambda = 0.8$ .

illusion image in Fig. 8.2a): the squares A and B appear to be of different intensity, for the human visual system corrects actual intensity by perceived shading. However, as shown in the adjacent demonstration, the diamonds truly have identical intensity. Here again, the role of Retinex is to separate the shading from the underlying reflectance. However, while the reflectance is expected to be piecewise constant, the illumination has both smooth and parts and sharp transitions. The sharp transitions are not sufficiently accounted for under simple  $L^2$  hard-thresholding, as shown in Fig. 8.2e–g); the estimated illumination always turns out too smooth. Even additional dynamic range compression cannot entirely fix the issues, as now parts of the checkerboard’s reflectance also appear in the illumination, see Fig. 8.2d). Thanks to the new possibilities of the proposed framework, however, the problem is rather nicely solved using a combination of  $L^0$  gradient fidelity, hard thresholding, and slight dynamic range compression, as shown in Fig. 8.2c).

**8.3. Shadow detection in natural images.** Shadow removal from a single (natural) image plays an important role in many computer vision algorithms. Most methods are based on a two-step procedure: first detect shadows, and then reconstruct shadow-free images. Shadow detection can be based on features such as intensity, gradients or texture, and even make use of supervision or training data [82, 27, 67]. Once the shadow regions have been reliably detected, several techniques aim at reconstructing a shadow-free images, through matting, inpainting, or Poisson editing [67, 78].

Here, we explore the applications of the proposed unified Retinex model for single step shadow detection and removal from a single image. We propose to use the  $L^0$  gradient fidelity criterion combined with dynamic range compression, without any gradient thresholding. The unfiltered  $L^0$  gradient fidelity is a strong prior on illumination gradient sparsity, while the dynamic range compression tends to take large intensity modulations out of the reflectance, and balance the mean intensities of inside- and outside-shadow regions. Our model can detect shadows in monochromatic and color images. We show a few example results for shadow detection in Fig. 8.3. We believe that the proposed model can largely compete with the recent state-of-the-art shadow detection scheme proposed Guo et al. [26, 27].

Moreover, our model also provides a shadow-free reflectance estimate at the same time. Re-

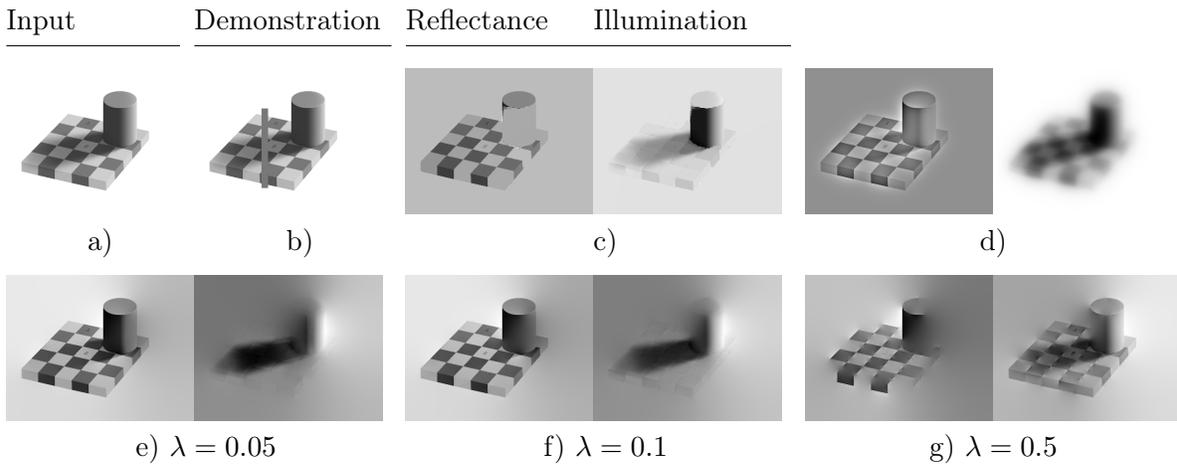


Figure 8.2: **Adelson checker illusion.** **a)** Input image. **b)** Demonstration of the illusion: despite the appearances, the squares A and B actually have equal intensity. Retinex is supposed to reproduce this illusion of intensity difference by removing the shadows. **c)** Reflectance and illumination recovered using the novel  $L^0$ -based Retinex. Hard thresholding  $\lambda = 0.15$ ,  $p = 0$ ,  $\alpha = 0.04$ . **d)**  $L^2$ -Retinex with dynamic range compression has too smooth illumination (hard thresholding  $\lambda = 0.025$ ,  $p = 2$ ,  $\alpha = 0.01$ ). **e–g)**  $L^2$ -Retinex without dynamic range compression. The drop shadow is removed nicely, but the cylinder shading is not.

reflectance output (i.e. after shadow removal) is illustrated in Fig. 8.4. However, in most natural scenes, the actual border between shaded and unshaded regions is rather smooth, called the penumbra, which is due to the spatial extent of the light source. Hence, the estimated shadow boundary in the proposed model is consistently overly sharp, and the estimated shadow-free reflectance image includes artifacts, see Fig. 8.4c). This problem can partially be tackled by smoothing the estimated illumination field in post-processing, as shown in Fig. 8.4d). A noticeable difference in texture is still visible, however, due to the missing specular highlights in the shadowed region, exclusively lit by ambient light.

**8.4. Shadow removal in color images.** If the images are treated as color images, however, a few shortcomings of the simple shadow-removal model become obvious, beyond the penumbra-issue. In Fig. 8.4e) we show the output of Retinex being applied to the lightness channel in HSV-space only. Since the shadowed region was lit by (sky-blueish) ambient light only, compared to warmer direct sun light, the colorcast after intensity correction becomes really striking. If, in contrast, we perform Retinex on all three RGB channels independently, the colorcast can be successfully avoided, see Fig. 8.4f). However, since the three channels are not coupled, the respective shadow-boundaries differ slightly, creating local color-artifacts.

The observed issues are just a manifestation of a more fundamental Retinex problem when dealing with color images: namely the correct choice of color space and channels in which to perform Retinex. Traditionally, Retinex theories consider RGB-images and treat the color channels independently in order to achieve color constancy (color normalization) [42]. To some purists, this is the one and only right way. Other authors have suggested

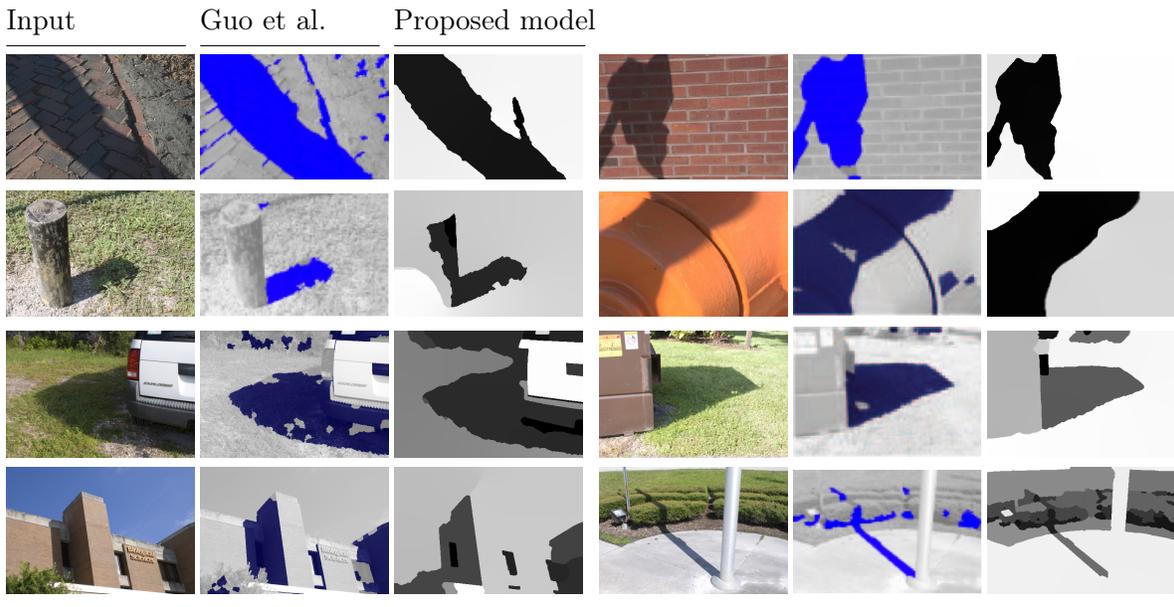


Figure 8.3: **Shadow detection results.** We compare the shadow detection results (illumination output) of our proposed model against the recently published results (blue mask) from Guo et al. [26, 27]. The results of the first row are very comparable, while we believe the examples of the second row are in favor of the proposed model. Indeed, our illumination output may be “multilevel” rather than just binary, and therefore better reflect the different nuances of shade in natural images (pole). On the other hand, our approach is less subject to local artifacts and produces more coherent shadow estimates.

to perform illumination correction by just working on the V-channel (lightness) of images in HSV color space—Retinex is expected to correct the amount of lighting but should conserve the general tone-trend in an image. Both approaches can yield unsatisfactory results in some situations. In particular, HSV-Retinex is unable to normalize differences between different lighting-temperatures (for example, blueish ambient lighting versus directly lit parts of a scene). It was alternatively proposed to perform color correction in CIELAB colorspace, where Retinex again works on the lightness channel, while a co-correction is performed on the chroma-channels based on the estimated change of illumination [73, 72].

An intermediate compromise between channel-wise RGB and lightness-only HSV Retinex could be devised as follows. The main advantage of RGB-Retinex is the capability of color-normalizing (“greying”) the respective shadow/light regions in a scene independently. Its main drawback is the lack of coupling between shadow boundaries in the three color channels. For HSV-retinex the situation is exactly opposite. A potential solution could perform channel-wise Retinex in RGB, but using a coupling term that encourages shadow boundaries to be collocated in all three color channels. Such a goal could be achieved by replacing the current channel-wise gradient fidelity by a grouped gradient fidelity derived from group sparsity [47, 48]. More advanced, another gradient fidelity functional involving stronger channel coupling could be

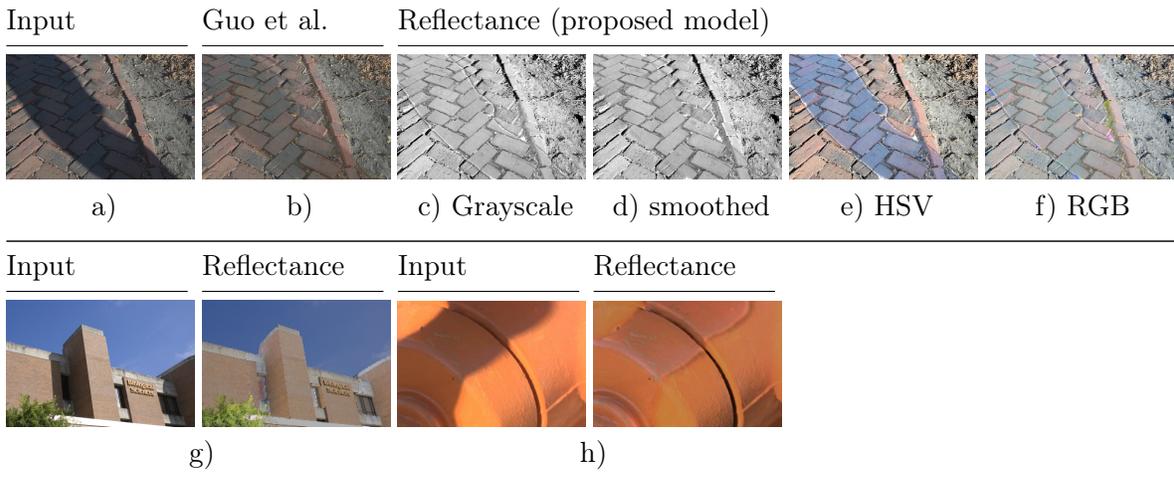


Figure 8.4: **Shadow removal results.** **a)** Input image. **b)** Recently published results from Guo et al. [26, 27]. **c)** Reflectance output of input image reduced to grayscale. The sharp boundary of the detected shadow region creates artifacts in the penumbra. **d)** The artifacts are almost entirely removed by smoothing the estimated illumination in post-processing. **e)** If the Retinex model is applied only to the V-channel of the color image in HSV-space, then strong colorcast becomes apparent, due to different lighting color for direct and ambient light. **f)** The colorcast is avoided by correcting all three RGB channels (colorbalancing). However, local artifacts appear due to inconsistent shadow region boundaries in the three individual channels.



Figure 8.5: **Cartoon-texture decomposition.** For important  $\alpha$ , the  $L^0$  model separates texture (reflectance) from cartoon (illumination).

employed, such as the Color Beltrami energy, that aligns gradients across color channels [37].

**8.5. Cartoon-texture decomposition.** The separation of an image into a piecewise regular component (cartoon) and its high-frequency parts (texture) is generally referred to as cartoon-texture decomposition [79, 56, 71]. If we give even more importance to dynamic range compression, then our proposed  $L^0$  gradient-fidelity based Retinex model can be used to this very same end. Indeed, the “reflectance” will only contain the texture of the image, whereas all larger scale intensity patches will be attributed to illumination (cartoon part). The scale of separation is determined by the weight of the dynamic range compression,  $\alpha$ . In Fig. 8.5 we show results of cartoon-texture decompositions of two natural images.

**9. Adaptive thresholding with texture and color.** In the preceding examples and comparisons, we always considered equal weights for gradient sparsity and fidelity,  $w = w_1 = w_2$ . However, doing so is not a requirement of the method, and choosing alternative weights for the two terms tremendously broadens the spectrum of possibilities. The fundamental principle is as follows: the reflectance of two pieces of same material is likely to be similar. If the non-local weights are constructed in a way that strongly connects same-material pixels, then the reflectance should have low non-local gradient magnitude. The first milestone is the definition of a suitable “material-distance” and the construction of the associated weights-graph. In the following, we quickly want to discuss two possible routes, namely the use texture and color-based weight graphs, as introduced in section 4, and their connections to existing methods.

**9.1. Texture-based non-local sparsity.** Here, the assumption is that pixels belonging to objects of the same material (and thus same reflectance) are characterized by local image structure (image patches) that are similar to (at least some) other, distant patches of the same material, and dissimilar to patches around pixels of different material.

This idea of texture sparsity in reflectance has been employed in non-local retinex already, to different extents [68, 74, 46].

In the proposed, unified model, this idea could be easily implemented as follows. The reflectance gradient is expected to be non-locally sparse, therefore we would pick  $w_1 = w_{nl}$ . The gradient fidelity, however is required locally, thus one chooses local weights  $w_2 = w_\epsilon$ , or mollified semilocal weights  $w_2 = w_g$ . This choice, together with an appropriate norm on both terms leads to one of the *non-stationary* gradient filters exposed in sections 3.5 through 3.7. In order to improve the material-similarity weights based on texture distances, illumination normalized cosine distances may be used in lieu of the classical non-local weights  $w_{nl}$ .

**9.2. Chroma-conditioned gradient thresholding.** The argument of texture-based material (dis-)similarity can easily be extended into a color-based (dis-)similarity metric, by using color-distances (see section 4.5). As pointed out earlier, classical Retinex treats the color channels independently and ignores any relations between them. In shadow-removal applications, however, this is clearly a shortcoming, since the presence of a shadow can be expected to similarly affect all color channels, and it is reasonable to exploit this interaction.

The role of threshold in the original Retinex theory is “to remove the effects of nonuniform illumination over the scene” [40]. However, those gradients are not necessarily small, for example the ones cross the shadow edges [18]. Explicitly, a large gradient belongs to reflectance if the material is different on both sides but to illumination if the material is the same. Such similar conditional thresholdings have been used in shadow removal [18, 19], and appeared in intrinsic image decomposition [10, 69, 52, 81, 14]. All these models aim at enforcing reflectance gradient sparsity between pixels that have similar hue, since these are believed to belong to objects of the same material and thus similar reflectance. Conversely, reflectance gradients are conserved only, if they are “motivated” by a hue gradient, suggesting a material boundary.

Such a method can easily be implemented within the non-local Retinex framework presented in this paper. Indeed, hue constancy requires sparsity of reflectance under illumination invariant but hue-sensitive weights. Thus one would choose  $w_1 = w_c$ . In contrast, the gradient fidelity weights can again be chosen locally,  $w_2 = w_\epsilon$ , or semilocally,  $w_2 = w_g$ . The resulting non-stationary filter function thresholds the observed gradients according to these criteria, and

then a reflectance gradient is fitted to this thresholded target.

**10. Conclusions.** In this paper, we have made an attempt at providing a fair overview of the broad spectrum of Retinex implementations existing in literature, and at unifying them within a single computational framework. The proposed unifying framework is a generalization of the well-known threshold-based two-step Retinex implementations. Indeed, we formulate the problem in the following two steps. First, we define an inner gradient sparsity and fidelity problem, that leads to a particular choice of gradient filtering function. Second, we look for a reflectance function that approximates this thresholded input gradient, while potentially optimizing additional sparsity or fidelity terms. The model traditional threshold model is thus generalized along four main directions: we use non-local differential operators in lieu of the local derivatives, we replace the hard-threshold by a more generic odd gradient filtering function, we extend the quadratic gradient fidelity to different fidelity norms, including  $L^1$  and  $L^0$ , and we include further terms in the reflectance reconstruction problem, such as reflectance fidelity or dynamic range compression.

We show, how various gradient-sparsity and -fidelity problems result in different gradient filtering functions, including soft- and hard-thresholding. Further, we discuss different general schemes to compute the weight functions used in the non-local differential operators, including local weights, semi-local Gaussian kernels, patch-based non-local weights, and finally cosine-distance-based color-distances.

Further, we briefly describe a possible numerical implementation to solve the proposed generalized non-local Retinex model in different configurations. In particular, we can show that the  $L^2$ -based non-local Retinex model can directly be implemented by Wilson-Cowan equations, which are used to model the mean activity of a population of both inhibitory and excitatory neurons in the cortex. This potential biological feasibility of at least the  $L^2$ -version of our framework conveys additional plausibility. Both  $L^1$  and  $L^0$ -based non-local Retinex models are proposed to be solved through variable splitting and the alternate direction method of multipliers.

Using the proposed generalized Retinex functional, we are able to expose relations with all major classes of existing Retinex implementations, such as kernel and variational Retinex, perceptual-contrast enhancement, and threshold-based PDE-Retinex. These relations are partly shown explicitly, partly made implicitly through well-established formal connections between some of these models. In particular, we provide an extensive list of equivalences between existing variational Retinex implementations and their counterparts within the proposed generalized framework. The capability of the proposed non-local Retinex model to reproduce these existing algorithms is illustrated on a few sample problems.

Beyond simply reproducing existing Retinex models, our proposed framework also offers potential for new forms of Retinex. In particular the  $L^0$ -based non-local Retinex model produces interesting results in terms of shadow detection and removal. While these results are convincing for gray-scale images, we realize that this simple model has some shortcomings with color-images, where ambient-lighting-color correction and color-channel-coherent shadow boundaries cannot currently be obtained simultaneously. However, we do point out how different forms of gradient-coupling across color channels (group sparsity, Beltrami functional) could be employed almost directly to address this issue. Pushing the dynamic range compression

in the reflectance component even further, we show how the very same Retinex model can be employed for cartoon-texture decomposition.

An important, yet largely unexplored, property of the proposed non-local Retinex model is the use of hybrid weights in the thresholding part; more specifically, different weight functions can be employed in the gradient sparsity and gradient fidelity component, leading to a gradient filtering function that is spatially varying (non-stationary, adaptive thresholding). While we do not provide practical examples of this capability in the present paper, we do outline the resulting parallels with so-called conditional-thresholding methods, for example known from shadow-removal.

While in the present paper we were able to introduce the unifying framework and outline many relations to existing Retinex models and shadow removal algorithms, future work will focus on providing experimental support. In particular, most of the sketched “new perspectives” (shadow removal, cartoon-texture decomposition, color-cosine-distance based Retinex, channel-coupling and group sparsity in multi-channel Retinex, and adaptive thresholding) are only superficially exposed here, but would certainly deserve more focused attention. Similarly, the numerical algorithms presented here are valid for proof-of-concept implementations only, but faster schemes are known to exist. Also, a detailed analysis of minimizers, in particular existence and uniqueness, and convergence of the algorithms is an important aspect of the model, but beyond the possible scope of the present paper as well.

**Acknowledgements.** The authors are grateful to Jean-Michel Morel for helpful advice and discussion at various stages of this work.

## REFERENCES

- [1] VIVEK AGARWAL, ANDREAS KOSCHAN, AND MONGI A ABIDI, *An Overview of Color Constancy Algorithms*, Journal of Pattern Recognition Research, 1 (2006), pp. 42–54.
- [2] PABLO ARIAS, GABRIELE FACCILOLO, VICENT CASELLES, AND GUILLERMO SAPIRO, *A Variational Framework for Exemplar-Based Image Inpainting*, International Journal of Computer Vision, 93 (2011), pp. 319–347.
- [3] EDWARD A. ASHTON, BRIAN D. WEMETT, ROBERT A. LEATHERS, AND TRIJNTJE V. DOWNES, *A Novel Method for Illumination Suppression in Hyperspectral Images*, in Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIV, Sylvia S. Shen and Paul E. Lewis, eds., vol. 6966, Apr. 2008, pp. 69660C–1–8.
- [4] MARCELO BERTALMIÓ, VICENT CASELLES, AND EDOARDO PROVENZI, *Issues About Retinex Theory and Contrast Enhancement*, International Journal of Computer Vision, 83 (2009), pp. 101–119.
- [5] MARCELO BERTALMIÓ, VICENT CASELLES, EDOARDO PROVENZI, AND ALESSANDRO RIZZI, *Perceptual Color Correction Through Variational Techniques*, IEEE Transactions on Image Processing, 16 (2007), pp. 1058–1072.
- [6] MARCELO BERTALMIÓ AND JACK D COWAN, *Implementing the Retinex algorithm with Wilson-Cowan equations*, Journal of Physiology - Paris, 103 (2009), pp. 69–72.
- [7] DIMITRI P. BERTSEKAS, *Constrained optimization and Lagrange Multiplier methods*, vol. -1, Academic Press, Boston, 1982.
- [8] ANDREW BLAKE, *Boundary conditions for lightness computation in Mondrian World*, Computer Vision, Graphics, and Image Processing, 32 (1985), pp. 314–327.
- [9] ———, *On lightness computation in Mondrian world*, in Central and Peripheral Mechanisms of Colour Vision, D. Ottoson and S. Zeki, eds., MacMillan, New York, 1985, pp. 45–49.
- [10] ADRIEN BOUSSEAU, SYLVAIN PARIS, AND FRÉDO DURAND, *User-assisted intrinsic images*, ACM Transactions on Graphics, 28 (2009), pp. 130:1–10.

- [11] GAVIN BRELSTAFF AND ANDREW BLAKE, *Computing lightness*, Pattern Recognition Letters, 5 (1987), pp. 129–138.
- [12] PAUL C. BRESSLOFF, JACK D. COWAN, MARTIN GOLUBITSKY, PETER J. THOMAS, AND MATTHEW C. WIENER, *Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex.*, Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 356 (2001), pp. 299–330.
- [13] ANTONI BUADES, BARTOMEU COLL, AND JEAN-MICHEL MOREL, *A Non-Local Algorithm for Image Denoising*, IEEE, 2005.
- [14] CHE-HAN CHANG, YU-TING CHENG, AND YUNG-YU CHUANG, *A Non-local Sparse Model for Intrinsic Images*, 2013.
- [15] TERRENCE CHEN, WOTAO YIN, XIANG SEAN ZHOU, DORIN COMANICIU, AND THOMAS S. HUANG, *Total variation models for variable lighting face recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28 (2006), pp. 1519–1524.
- [16] JACK D. COWAN AND PAUL C. BRESSLOFF, *Visual Cortex and the Retinex Algorithm*, in Electronic Imaging 2002, June 2002, pp. 278–285.
- [17] DAVID M. EAGLEMAN, *Visual illusions and neurobiology.*, Nature reviews. Neuroscience, 2 (2001), pp. 920–6.
- [18] GRAHAM D. FINLAYSON, S.D. HORDLEY, AND M.S. DREW, *Removing shadows from images using retinex*, 10th Color Imaging Conference, (2002).
- [19] GRAHAM D FINLAYSON, STEVEN D HORDLEY, CHENG LU, AND MARK S DREW, *On the removal of shadows from images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28 (2006), pp. 59–68.
- [20] JONATHAN A. FRANKLE AND JOHN J. MCCANN, *Method and apparatus for lightness imaging*, 1983.
- [21] BRIAN FUNT, FLORIAN CIUREA, AND JOHN MCCANN, *Retinex in MATLAB*, Journal of Electronic Imaging, 13 (2004), p. 48.
- [22] GUY GILBOA AND STANLEY OSHER, *Nonlocal Operators with Applications to Image Processing*, Multiscale Modeling & Simulation, 7 (2008), pp. 1005–1028.
- [23] ROLAND GLOWINSKI AND PATRICK LE TALLEC, *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1989.
- [24] TOM GOLDSTEIN AND STANLEY OSHER, *The Split Bregman Method for L1-Regularized Problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 323–343.
- [25] SILVIU GUIASU AND ABE SHENITZER, *The principle of maximum entropy*, The Mathematical Intelligencer, 7 (1985), pp. 42–48.
- [26] RUIQI GUO, QIEYUN DAI, AND DEREK HOIEM, *Single-image shadow detection and removal using paired regions*, in CVPR 2011, IEEE, June 2011, pp. 2033–2040.
- [27] ———, *Paired Regions for Shadow Detection and Removal.*, IEEE Transactions on Pattern Analysis and Machine Intelligence, (2012).
- [28] ANIL N. HIRANI, KAUSHIK KALYANARAMAN, AND SETH WATTS, *Least Squares Ranking on Graphs*, (2010).
- [29] BERTHOLD K.P. HORN, *Determining lightness from an image*, Computer Graphics and Image Processing, 3 (1974), pp. 277–299.
- [30] ANYA HURLBERT, *Formal connections between lightness algorithms*, Journal of the Optical Society of America A, 3 (1986), p. 1684.
- [31] DONG-GUK HWANG, WOO-RAM LEE, YOUNG-JIN OH, AND BYOUNG-MIN JUN, *Frankle-McCann Retinex by Shuffling*, in Convergence and Hybrid Information Technology, 2012, pp. 381–388.
- [32] EDWIN THOMPSON JAYNES, *Information Theory and Statistical Mechanics*, Physical Review, 106 (1957), pp. 620–630.
- [33] XIAOYE JIANG, LEK-HENG LIM, YUAN YAO, AND YINYU YE, *Statistical ranking and combinatorial Hodge theory*, Mathematical Programming, 127 (2010), pp. 203–244.
- [34] DANIEL J. JOBSON, ZIA-UR RAHMAN, AND GLENN A. WOODDELL, *A multiscale retinex for bridging the gap between color images and the human observation of scenes*, IEEE Transactions on Image Processing, 6 (1997), pp. 965–76.
- [35] ———, *Properties and performance of a center/surround retinex*, IEEE Transactions on Image Processing, 6 (1997), pp. 451–62.

- [36] RON KIMMEL, MICHAEL ELAD, DORON SHAKED, RENATO KESHET, AND IRWIN SOBEL, *A Variational Framework for Retinex*, International Journal of Computer Vision, 52 (2003), pp. 7–23.
- [37] RON KIMMEL, R MALLADI, AND NIR SOCHEN, *Images as Embedded Maps and Minimal Surfaces: Movies, Color, Texture, and Volumetric Medical Images*, Int. J. Comput. Vis., 39 (2000), pp. 111–129.
- [38] EDWIN H. LAND, *The Retinex*, American Scientist, 52 (1964), pp. 247–253, 255–264.
- [39] ———, *The Retinex Theory of Color Vision*, Scientific American, 237 (1977), pp. 108–128.
- [40] ———, *Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image*, Proceedings of the National Academy of Sciences of the United States of America, 80 (1983), pp. 5163–9.
- [41] ———, *An alternative technique for the computation of the designator in the retinex theory of color vision*, Proceedings of the National Academy of Sciences of the United States of America, 83 (1986), pp. 3078–80.
- [42] EDWIN H. LAND AND JOHN J. MCCANN, *Lightness and Retinex Theory*, Journal of the Optical Society of America, 61 (1971), pp. 1–11.
- [43] JONG-SEN LEE, *Digital image smoothing and the sigma filter*, Computer Vision, Graphics, and Image Processing, 24 (1983), pp. 255–269.
- [44] ALEXANDER D. LOGVINENKO, *Lightness induction revisited*, Perception, 28 (1999), pp. 803–816.
- [45] WENYE MA, JEAN-MICHEL MOREL, STANLEY OSHER, AND AICHI CHIEN, *An L1-based variational model for Retinex theory and its application to medical images*, in CVPR 2011, IEEE, June 2011, pp. 153–160.
- [46] WENYE MA AND STANLEY OSHER, *A TV Bregman iterative model of Retinex theory*. 2010.
- [47] ANGSUL MAJUMDAR AND RACHEL K. WARD, *Compressive color imaging with group-sparsity on analysis prior*, in International Conference on Image Processing, 2010, pp. 1337–1340.
- [48] ———, *Non-convex group sparsity: Application to color imaging*, in Acoustics Speech and Signal Processing, 2010, pp. 469–472.
- [49] DAVID MARR, *The computation of lightness by the primate retina*, Vision Research, 14 (1974), pp. 1377–1388.
- [50] JOHN J. MCCANN, *Lessons Learned from Mondrians Applied to Real Images and Color Gamuts*, in Seventh Color Imaging Conference, 1999, pp. 1–8.
- [51] ———, *Capturing a black cat in shade: past and present of Retinex color appearance models*, Journal of Electronic Imaging, 13 (2004), p. 36.
- [52] JEAN-MICHEL MOREL, *Personal communication*, 2012.
- [53] JEAN-MICHEL MOREL, ANA-BELEN PETRO, AND CATALINA SBERT, *A PDE Formalization of Retinex Theory*, IEEE Transactions on Image Processing, 19 (2010), pp. 2825–2837.
- [54] MICHAEL K. NG AND WEI WANG, *A Total Variation Model for Retinex*, SIAM Journal on Imaging Sciences, 4 (2011), pp. 345–365.
- [55] JORGE NOCEDAL AND STEPHEN J. WRIGHT, *Penalty and Augmented Lagrangian Methods*, in Numerical Optimization, 2006, ch. 17, pp. 497–528.
- [56] STANLEY J. OSHER, ANDRÉS SOLÉ, AND LUMINITA VESE, *Image decomposition and restoration using total variation minimization and the H-1 Norm*, Multiscale Modeling & Simulation, 1 (2003), pp. 349–370.
- [57] BRAXTON OSTING, CHRISTOPH BRUNE, AND STANLEY OSHER, *Optimal Data Collection for Improved Rankings Expose Well-Connected Graphs*, Journal of Machine Learning Research.
- [58] BRAXTON OSTING, JÉRÔME DARBON, AND STANLEY OSHER, *Statistical Ranking Using the l1-Norm on Graphs*, AIMS J. Inverse Problems and Imaging, 7 (2013), pp. 907–926.
- [59] RODRIGO PALMA-AMESTOY, EDOARDO PROVENZI, MARCELO BERTALMIÓ, AND VINCENT CASELLES, *A perceptually inspired variational framework for color enhancement.*, IEEE transactions on pattern analysis and machine intelligence, 31 (2009), pp. 458–74.
- [60] EDOARDO PROVENZI, LUCA DE CARLI, ALESSANDRO RIZZI, AND DANIELE MARINI, *Mathematical definition and analysis of the Retinex algorithm*, Journal of the Optical Society of America, 22 (2005), pp. 2613–2621.
- [61] EDOARDO PROVENZI, MASSIMO FIERRO, ALESSANDRO RIZZI, LUCA DE CARLI, DAVIDE GADIA, AND DANIELE MARINI, *Random Spray Retinex: A New Retinex Implementation to Investigate the Local Properties of the Model*, IEEE Transactions on Image Processing, 16 (2007), pp. 162–171.

- [62] EDOARDO PROVENZI, CARLO GATTA, MASSIMO FIERRO, AND ALESSANDRO RIZZI, *A spatially variant white-patch and gray-world method for color image enhancement driven by local contrast*, IEEE transactions on pattern analysis and machine intelligence, 30 (2008), pp. 1757–70.
- [63] ZIA-UR RAHMAN, DANIEL J. JOBSON, AND GLENN A. WOODSELL, *Retinex processing for automatic image enhancement*, Journal of Electronic Imaging, 13 (2004), p. 100.
- [64] FLOYD RATLIFF, *Mach bands: quantitative studies on neural networks in the retina*, Holden-Day, San Francisco, CA, USA, 1965.
- [65] ALESSANDRO RIZZI, CARLO GATTA, AND DANIELE MARINI, *A new algorithm for unsupervised global and local color correction*, Pattern Recognition Letters, 24 (2003), pp. 1663–1677.
- [66] ———, *From Retinex to Automatic Color Equalization: issues in developing a new algorithm for unsupervised color equalization*, Journal of Electronic Imaging, 13 (2004), p. 75.
- [67] ANDRES SANIN, CONRAD SANDERSON, AND BRIAN C. LOVELL, *Shadow detection: A survey and comparative evaluation of recent methods*, Pattern Recognition, 45 (2012), pp. 1684–1695.
- [68] LI SHEN, PING TAN, AND STEPHEN LIN, *Intrinsic image decomposition with non-local texture cues*, in 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, June 2008, pp. 1–7.
- [69] LI SHEN AND CHUOHAO YEO, *Intrinsic images decomposition using a local and global sparse representation of reflectance*, in CVPR 2011, IEEE, June 2011, pp. 697–704.
- [70] ROBERT SOBOL, *Improving the Retinex algorithm for rendering wide dynamic range photographs*, Journal of Electronic Imaging, 13 (2004), p. 65.
- [71] JEAN-LUC STARCK, MICHAEL ELAD, AND DAVID L. DONOHO, *Image decomposition via the combination of sparse representations and a variational approach.*, IEEE Transactions on Image Processing, 14 (2005), pp. 1570–1582.
- [72] MASATO TODA AND MASATO TSUKADA, *High dynamic range rendering method for YUV images with global luminance correction*, in 2011 IEEE International Conference on Consumer Electronics (ICCE), IEEE, Jan. 2011, pp. 255–256.
- [73] MASATO TODA, MASATO TSUKADA, AKIRA INOUE, AND TETSUAKI SUZUKI, *High dynamic range rendering for YUV images with a constraint on perceptual chroma preservation*, in 2009 16th IEEE International Conference on Image Processing (ICIP), IEEE, Nov. 2009, pp. 1817–1820.
- [74] VITOMIR ŠTRUC AND NIKOLA PAVEŠIĆ, *Illumination invariant face recognition by non-local smoothingn*, in Biometric ID management and multimodal communication, Springer Berlin / Heidelberg, 2009, pp. 1–8.
- [75] ROBERT H. WALLIS, *An approach to the space variant restoration and enhancement of images*, in Symposium on current mathematical problems in image science, 1976.
- [76] HUGH R. WILSON AND JACK D. COWAN, *Excitatory and inhibitory interactions in localized populations of model neurons.*, Biophysical journal, 12 (1972), pp. 1–24.
- [77] ———, *A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue*, Kybernetik, 13 (1973), pp. 55–80.
- [78] TAI-PANG WU, CHI-KEUNG TANG, MICHAEL S. BROWN, AND HEUNG-YEUNG SHUM, *Natural shadow matting*, ACM Transactions on Graphics, 26 (2007).
- [79] WOTAO YIN, DONALD GOLDFARB, AND STANLEY OSHER, *Image Cartoon-Texture Decomposition and Feature Selection Using the Total Variation Regularized L1 Functional*, in Variational, Geometric, and Level Set Methods in Computer Vision, 2005, pp. 73–84.
- [80] WEN-CHAO ZHANG, XIANG-JING AN, AND SHENG-DONG PAN, *An improved recursive retinex for rendering high dynamic range photographs*, in 2011 International Conference on Wavelet Analysis and Pattern Recognition, IEEE, July 2011, pp. 121–125.
- [81] QI ZHAO, PING TAN, QIANG DAI, LI SHEN, ENHUA WU, AND STEPHEN LIN, *A Closed-form Solution to Intrinsic Image Decomposition with Retinex and Non-local Texture Constraints.*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34 (2012).
- [82] JIEJIE ZHU, KEGAN G. G. SAMUEL, SYED Z. MASOOD, AND MARSHALL F. TAPPEN, *Learning to recognize shadows in monochromatic natural images*, 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2010), pp. 223–230.
- [83] DOMINIQUE ZOSSO, GIANG TRAN, AND STANLEY OSHER, *A unifying retinex model based on non-local differential operators*, in IS&T/SPIE Electronic Imaging, Charles A. Bouman, ed., Feb. 2013, pp. 865702–1–16.