Proceedings of Mirage 2005, INRIA Rocquencourt, France, March, 1-2 2005

RELIGHTING, RETINEX THEORY, AND PERCEIVED GRADIENTS

Todor Georgiev

tgeorgie@adobe.com
Adobe Systems, 345 Park Ave, W10-124, San Jose, CA 95110, USA

ABSTRACT

This paper provides a Mathematical model of relighting and adaptation of human Vision based on the approach of von Kries. It extends Retinex theory into a new image processing formalism using the mathematical concepts of Bundles and Connections, that is well suited for scene relighting applications.

Our model generates considerably improved results in the case of scratch/wire/wrinkle removal and has been implemented in Photoshop. The result is valid in arbitrary color space, and invariant under simultaneous lighting and adaptation transforms type von Kries.

This mathematical model also provides an exact solution to the problem of High Dynamic Range compression, and can be applied to a wide range of image processing algorithms.

Figure 1: *Semantic adaptation: The sun appears brighter than it is.*

1. INTRODUCTION

We have great flexibility in adapting to different lighting conditions. For example, entering a dark movie theatre, we can not see, but within a few seconds we are able to see well enough to find a seat. This is global adaptation.

Adaptation could also be local. It is a well known fact that the image as perceived is different from the image as measured by a camera. Perceived brightness, also called *lightness* in Retinex theory [6], depends on the surroundings of each pixel, and on the image as a whole. Experiments have shown that local adaptation is fast. Also the process is not limited to the retina only. Generating lightness is a complex process taking place in the cortex.

The process of addition, or generating lightness, is influenced by semantic information. For example, the sun on Figure 1 appears much brighter than it would be if measured by pixel value. A potential explanation for this could be the halo around the sun and our understanding that the sun must be very bright.

Some of the effects of local adaptation to lighting conditions are well understood in the general framework of Land's Retinex theory [6], [7]. However, this theory still does not provide a solution for actually calculating lightness in all cases. For example, it does not provide a solution in the case of objects of smoothly changing brightness in the image (as opposed to 'Mondrian' pictures consisting of patches of constant color where it works well). Even in cases where no solution is available, the main idea of Retinex theory - that we see pixel values as transformed by our visual system/interpretation - still holds true.

A big number of applications treat color additively, as a linear combination of three basis colors. This is the more widely accepted approach. In other cases color is treated multiplicatively.

Multiplicative representation of color is appropriate for description of light reflection/absorbtion. In sensor space this representation is equivalent to von Kries-type transforms of reflectance. Related approaches are widely used not just in Retinex theory, but also in image based rendering and lighting - for editing movies and a wide range applications of relighting [8], [9], [10].

We would like to consider these lighting/adaptation transforms together with Retinex theory in relation to image processing. As J. Koenderink notes [11], we should do image processing 'right' in terms of how images are perceived, taking into account the invariance of the image we see under a group of transformations of the physical image that can be interpreted as relighting.

In this paper we present our new mathematical model of relighting and adaptation. We propose a geometry of image space as a fibred manifold. Every image can be described as a section in a fibred space, and in this model we can clearly and accurately describe the essence of Retinex and other local adaptation theories, perform tone mapping by simulated relighting, and other image processing algorithms.

In the traditional approach grayscale images can be represented as surfaces in R^3. Pixels are defined by their coordinates x, y in the image plane, and their corresponding values in z. It is customary to assume that brightness of different pixels can be compared by this value of z. However, there are illusions that provide examples of same pixel values appearing different. Retinex and other adaptation theories are also based on considering the difference between pixel values and what we see. Pixels with higher z do not always appear brighter, and in certain cases the theory actually provides an algorithm to calculate visible brightness (lightness) based on raw pixel values.

The traditional method of comparison of pixels, and subsequently the concept of derivative, needs to be replaced with a new method of comparison that more accurately represents our visual perception and adaptation. This replacement of derivatives with modified (adapted) derivatives will alter our approach to image processing. Further, it will modify essentially any image processing algorithm, making it more closely related to human perception.

2. FIBRED SPACE APPROACH TO IMAGE PROCESSING

2.1. Our New Model of Image Space

The traditional model of image space is a Cartesian product of the image plane and the positive real line of pixel values R^+. Grayscale images are represented as surfaces in R^3. Pixels are defined by their coordinates x, y in the image plane, and their corresponding grayscale values in the z axis.

This mathematical structure contains two natural projections: For any point in image space we can immediately know which pixel it is, and what the pixel value is – according to the two components of the Cartesian product $R^2 \times R^+$. This makes possible the simple view of the image as a function $z = f(x, y)$.

Lightness depends on many factors, including local comparisons and semantic information. The complex nature of adaptation suggests that the second projection (measuring lightness) is unpredictable in computer vision, and hence does not exist a priori. Human visual system does not compare pixels by their luminance. We need a model of image space in which pixel values, even if well defined, are not comparable a priori.

We propose a model that replaces the Cartesian product structure of Image Space with a Fibred Space structure (see also [11]). The new structure is 'weaker' because it 'forgets' (in the mathematical sense) one of the projections. In this paper we are showing on the basis of two examples how fibred space structure can be more useful.

2.2. Definition of Fibred Spaces (Bundles)

We start with an introduction to fibred spaces in general, then we show how they can be used to model images and transformations.

By definition [12], a Fibred Space (E, π, B) consists of two spaces: *total space* E and *base space* B, and a mapping π, called *projection*, of the total space onto the base. Space B has lower dimension than E, so many E points map to the same B point, as shown in Figure 2.

In our model of grayscale images the total space is R^3, the base is the image plane, and π gives us the location of each pixel in the image plane. There is no mapping that would give us the grayscale value of lightness for a pixel.

For each point $p \in B$ there is the so-called *fibre* (F_p in Figure 2) in E, consisting of all points that are sent to p by π (definition of fibre). We cannot compare the lightness of two points from different fibres because there is no mapping that a priori would produce that lightness. Each fibre has its luminance coordinate, but luminances in different fibres are not related perceptually. This corresponds to the fact that π has no inverse. In other words, there is no distinguished mapping of B into E.

By definition, a section in a Fibred Space is a mapping f that sends points in B to E, and has the property $\pi(f(p)) = p$ for any $p \in B$. See Figure 3.

A section selects just one of the many points in each fibre. It defines one manifold (connected set of points) in total space E, with one point in E for each point in B. Intuitively it is "the closest we can get to the concept of function without defining a function".

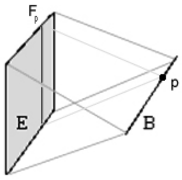

Figure 2: *Fibred space* $(E,\ \pi,\ B)$.

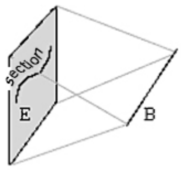

Figure 3: *Section in fibred space.*

A grayscale image is a section in a fibred image space $(R^3,\ \pi,\ R^2)$. Since there is no projection onto z, there is no comparison between different pixels. As a result, change in image lightness and directional derivative at a point in image space is defined not by a displacement vector in the image plane (x, y), as it is with functions, but in relation to a vector in the total space (x, y, z), which is called "the horizontal lift" of the vector in (x, y). (Of course, there is a luminance value z for each pixel, but it is a perceptually meaningless coordinate.)

Before giving rigorous definitions, let's look at one imaginary practical example that will help build some intuition.

2.3. Intuition

Imagine a scenario, in which three mathematicians are building houses on a hill. Figure 4 compares the three "architectures". The first mathematician doesn't know anything about the fibred structure of 3D due to gravity. In other words, he ignores the projection that singles out one direction as "vertical". This is obviously wrong.
Lesson: If our problem has fibred structure, we have to take it into consideration and cannot ignore the projection.

The second mathematician correctly takes into account the projection, but he uses a wrong measure of horizontality. He assumes that horizontal is the local surface of the earth. He calculates derivatives relative to the hill slope.
Lesson: The concept of 'horizontal' is not defined a priori in a fibred space, and we need to choose it appropriately for each particular application.

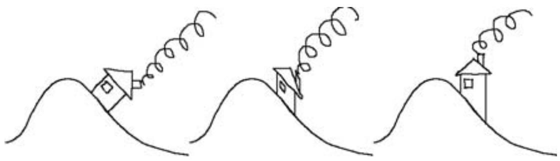

Figure 4: *Three houses on a hill.*

Finally, the third mathematician uses the correct measure of horizontality. If his coordinate system has x along the surface of the earth, he calculates the slope of the floor as $d/dx + a$, where a compensates for the slope of the hill. This way of calculating derivative, relative to a distribution of planes that are defining what "horizontal" means at each point, is called *Covariant Derivative* or *Connection*. Next we will give a more rigorous definition.

2.4. Connections

In fibred spaces, also known as Bundles, changes in the section (slopes of the section) are measured by the so called connection, or covariant derivative (instead of derivative). As the name suggests, connections show how fibres are connected or glued together. Connections are used like derivatives to compare or transfer pixel values from one fibre to another. In Gauge theories [13] the simplest example of such a field is the vector potential in Electrodynamics.

Let us consider a vector field (a vector at each point) in the image plane. In tensor notations we use coordinates x^μ, $\mu = 1, 2$ in the plane, instead of the more traditional x and y. A vector field describes at each point the directional derivative $\mathbf{X}$ with components $X^\mu = X^\mu(x)$, and tells us how to differentiate functions in a given direction. A common notation for vector fields is $\mathbf{X} = X^\mu \partial/\partial x^\mu$, summation over μ. In the traditional treatment of images as functions on the plane, a vector field (directional derivative) is the way to calculate rate of change at a given point, in a given direction $\mathbf{X}$.

If the image is defined as a section in a fibred space, the above definition of change does not work because in fibred spaces there is no concept of derivative. Perceptually, we do not have a sense of horizontality because there is no natural projection on lightness. This sense of horizontality needs to be defined as an additional structure. This structure is called connection.

A *connection* on a bundle (E, π, B) is a way for any vector field $\mathbf{X}$ defined on B to differentiate a section σ on E and produce a new section. If F and S are functions on

B, the derivative of the product FS in direction $\mathbf{X}$ would be $\mathbf{X}(FS) = (\mathbf{X}F)S + F\mathbf{X}S$, which is known as the Leibniz rule for derivative of a product. The concept of connection is a generalization of the above Leibniz rule to the case of sections, which replace functions. By definition, if ∇ is a connection, $\nabla_{\mathbf{X}}(F\sigma) = (\mathbf{X}F)\sigma + F\nabla_{\mathbf{X}}\sigma$. Note that the derivative $\mathbf{X}$ acts on a function, while the "derivative" acting on the section is $\nabla_{\mathbf{X}}$.

In our image processing applications, a color picture is a section in a vector bundle, where each three dimensional fibre is a copy of the vector space of colors. We call it the *color bundle*. A connection is "adapted directional derivative of color", as perceived by the observer. In other words, it shows how the human visual system perceives directional change of color in a given state of adaptation.

Any section can be represented as a linear combination of a set of basis sections σ_i. In other words, $\sigma = F_i\sigma_i$. Summation is assumed and the coefficients F_i are functions. These functions are referred to as color channels (Photoshop terminology). Some more notations: Everywhere in this paper $i = 1, 2, 3$ enumerates color channels. We always assume summation over repeated indexes; upper and lower indexes are treated the same assuming Euclidean metric. We use greek indexes, $\mu = 1, 2$, to enumerate image coordinates x, y. We write ∂_μ for $\partial/\partial x^\mu$, and ∇_μ for $\nabla_{\partial/\partial x^\mu}$.

By the above definition of connection, ∇_μ would act on a section $\sigma = F_i\sigma_i$ in the following way:

$$\nabla_\mu\sigma = \nabla_\mu(F_i\sigma_i) = (\partial_\mu F_i)\sigma_i + F_i\nabla_\mu\sigma_i \qquad (1)$$

This expression simply extends the Leibniz rule for the action of a derivative on functions to a Leibniz rule for sections. We don't know what the action on the basis section is, but we know that it must be again a section, representable by the basis. So, it is $\nabla_\mu\sigma_i = -A^j{}_{i\mu}\sigma_j$ where $A^j{}_{i\mu}$ are functions.

$$\nabla_\mu(F_i\sigma_i) = (\partial_\mu F_i)\sigma_i - F_iA^j{}_{i\mu}\sigma_j \qquad (2)$$

As a matter of notation, often the basis σ_i is dropped, and we talk of the section as represented in terms of F_i. Then the action of the connection on F_i is:

$$\nabla_\mu F_i = \partial_\mu F_i - A^j{}_{i\mu}F_j. \qquad (3)$$

This expression for the connection, as a replacement of the derivative, will be our main tool throughout this paper. The rule of thumb is that connection ∇_μ replaces the gradient ∂_μ according to the so called "minimal substitution"

$$\partial_\mu \to \nabla_\mu = \partial_\mu - A_\mu. \qquad (4)$$

We call $\partial_\mu - A_\mu$ the *covariant derivative*, or *perceptual gradient*.

When representing images as a sections in the color bundle, in given RGB space coordinates, A_μ will be 3×3 matrices, as shown in the next section.

3. THE GROUP OF VON KRIES LIGHTING / ADAPTATION TRANSFORMS.

3.1. Additive color vs. Multiplicative color

In the so-called cone space coordinates [14], color is represented as a vector $\vec{F} = (L, M, S)$. Here L, M, and S stand for Long, Medium, and Short wavelengths of light, and represent the three types of cone sensors in the human retina. Rod sensors, which play a role in low lighting conditions, are not considered in this model. In the case of a camera, L, M, and S are the three camera sensors.

Sometimes, Log-space representation of color is used in the context of cone sensors output, $\vec{f} = (l, m, s)$, where $l = \ln L$, $m = \ln M$, $s = \ln S$. In a simplified toy-model of human vision these Log signals are being sent from the eye to the cortex. This model will be of help for understanding our approach.

Usually in the literature color is treated additively, as in RGB representation. That is, each color as a point in a vector space is a sum or linear combination of basis colors.

There is also a multiplicative approach to color. In line with von Kries [14, 15] and Retinex [6] models, relighting the scene (or casting a shadow) is expressed as a multiplication by a matrix diagonal in LMS space. The same type of transformation also describes internal adaptation of the visual system to different lighting conditions, at least in the von Kries model. Starting from some fixed reference point, any color can be represented by such a transform.

The set of all those lighting/adaptation transformations forms a Lie group[1]. We will refer to it as the *von Kries*

[1] A Lie group is a very intuitive mathematical construction, which generalizes everyday examples of continuous transforms, like the plurality of all rotations in space or the plurality of all translations in space. A group is a set of transforms, $\phi_1, \phi_2, \ldots$, any two of which can be multiplied and the result $\phi_1\phi_2$ belongs to the group. Also, for any ϕ the inverse ϕ^{-1} belongs to the group, so that $\phi\phi^{-1} = 1$. Here 1 is the identity transform of the group. The elements of a Lie group can be expressed as finctions of a number of parameters, for example $\phi = \phi(\varphi_1, \varphi_2, \varphi_3)$, and satisfy natural smoothness properties.

group. The action of this group on a color RGB-vector $\vec{F}$ has been described first by von Kries back in 1902 [15], and in our notations it is:

$$\vec{F} \to \phi \vec{F} \tag{5}$$

where the matrix ϕ is diagonal in cone space coordinates, L, M, S. (Note however, that we can work in any RGB-space, and still the same equation (5) holds, with the appropriate matrix.) The von Kries group transform (5) can always be expressed as:

$$\vec{F} \to e^{K_i \varphi_i} \vec{F} \tag{6}$$

where K_i are the so called generators[2] of the group. In LMS-space the generators have their simplest form:

$$K_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$K_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$K_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

They are general matrices in other color space coordinates. Formula (6) will be our only use of group theory in this paper.

From the above it is clear that we are dealing with the group of all 3×3 matrices that are diagonal in cone space representation, where φ_i are logarithms of the diagonal elements of ϕ. In Log-space this is the Abelian group of translations $\vec{f} \to \vec{f} + \vec{\varphi}$. It is obvious that the group should be Abelian ($\phi_1 \phi_2 = \phi_2 \phi_1$) regardless of the representation: Consider lighting the scene with a red light, and then adding a green filter in front of the light source. This is the same as lighting the scene with a green light, and then adding a red filter.

3.2. Relighting Gradients

Changing the lighting conditions or *relighting* can be different at different points in the scene/image. The same must be true for adaptation. Let's see how the perceptual "covariant" derivative acts on the group transformations in the general adaptation case where group parameters are functions of image coordinates, $\varphi_i = \varphi_i(x, y)$. One can verify that

$$(\partial_\mu - A_\mu) e^{K_i \varphi_i} \vec{F} = e^{K_i \varphi_i} (K_i \partial_\mu \varphi_i + \partial_\mu - A_\mu) \vec{F} \tag{7}$$

Considering (7), there is a difference between the behavior of pixels and image gradients during adaptation to relighting: Pixels are only multiplied by $e^{K_i \varphi_i}$, while for gradients also there is an additive term proportional to $K_i \partial_\mu \varphi_i$. The natural expectation would be that gradients behave like pixels, just as it is when relighting and adaptation are constant in space.

This suggests the following idea: If we assume that the usual von Kries adaptation to relighting transform

$$\vec{F} \to e^{K_i \varphi_i} \vec{F} \tag{8}$$

corresponds in the visual system to perceptual gradient adaptation transform

$$A_\mu \to A_\mu + K_i \partial_\mu \varphi_i, \tag{9}$$

then the perceived gradient would transform during adaptation multiplicatively by $e^{K_i \varphi_i}$, same as the pixels. It is natural to assume that changing lighting would change the state of adaptation. Equations (8) and (9) give us the exact expressions for that process of adaptation, consistent with the model. This is a new result, which can be considered a contribution to Retinex theory of adaptation.

Note that this process automatically takes place for transformations of lighting conditions that are constant across the image, and are discounted by the visual system. The above condition (9) ensures that this same mechanism would also work for the gradient when the relighting transform is not constant across the image. In other words, (8) and (9) lead to the right adaptation transform for the perceived gradient

$$(\partial_\mu - A_\mu) \vec{F} \to (\partial_\mu - A_\mu - K_i \partial_\mu \varphi_i) e^{K_i \varphi_i} \vec{F} = e^{K_i \varphi_i} (\partial_\mu - A_\mu) \vec{F}. \tag{10}$$

After adaptation (to a given relighting), the expression for the perceptual gradient acting on the image is transformed into adapted perceptual gradient acting on the adapted image, which is equal to von Kries-type adaptation of the original perceived gradient. Adaptation to the relighting is such that the perceived gradient is always transformed by the same von Kries matrix as the pixels themselves. In this way the visual system discounts not only for uniform relighting, but also for shading/shadows and other lighting conditions changing across the image.

[2] Generators of a Lie Group are matrices defining infinitesimal group transforms in the sense that they are the linear terms in a Taylor series expansion around the identity group element. From the generators the whole group can be reconstructed. We will simply use the final result (6) for our particular case.

This result extends the Retinex theory model's description of relighting objects of piecewise constant reflectance to the more general case of relighting images of arbitrary changing reflectance or perceived gradient.

Note also that in our model A_μ is not restricted to such values that can be represented by simple adaptation to a given image $K_i \partial_\mu \varphi_i$. We believe that A_μ can be more general: Human vision adapts to the perceived gradient in the image, not to the true gradient. Perceived gradients can be modified from their true values by different factors, including nonlinearity and semantic adaptation. In other words, we can potentially observe in the real world gradients that can not be integrated into an image, and this could lead to a state of adaptation with unrestricted A_μ.

4. SCRATCH REMOVAL

This section will describe one specific application of the above lighting/adaptation theory.

Relighting an image differently at different pixels is a powerful operation, because any change to an image can be described in terms of relighting. We can model changes of lighting that are constant throughout the image, smoothly graduated, or widely different between one pixel and its neighbors. Relighting is described by a group: Given an arbitrary transformation, we can always model a transformation to undo it. In this section, we present one application of the lighting/adaptation theory above.

4.1. Harmonic Reconstruction

Often in images there are unwanted elements/defects like wires, scratches, or wrinkles (in human faces). Removing them, also known as Inpainting [3], is a desirable feature for many applications.

Consider the noisy image similar to the one in Figure 5, courtessy of Russell Williams. If we zoom in the small rectangle at the top, we see a lot of detail, including a scratch (Figure 6). One intuitive approach to fixing the scratch is to reconstruct the image surface for each channel in the defective area as a harmonic function $F(x, y)$, i.e. solution of the Laplace equation, with Dirichlet boundary conditions.

$$\triangle F = 0 \tag{11}$$

In this paper $\triangle$ denotes the Laplacian on the Euclidean plane. This solution is continuous everywhere, including at the boundary.

Figure 5: Noisy picture (basilica San Marco, Venice).

A simple way to solve the Laplace equation in a given area with Dirichlet boundary conditions is to iterate with the following kernel (divided by 4):

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Each iteration reads from a 3×3 pixel area, adds up the pixel values as described by the kernel to calculate the average value, then writes it to the central pixel. We need to write only in the selected area, while sampling from outside the boundary when part of the kernel covers outside pixels. See [5] for faster numerical methods.

Further extensions of this approach to higher order Laplace-type PDEs are possible if we use the appropriate kernel. They can be designed to have both continuous solution and continuous derivatives at the boundary, thus fitting better to the original image [2].

4.2. The Problem with Reconstruction

Figure 6 shows the result of the above reconstruction of the three individual color channels on a fragment of a noisy image. Film grain and other artifacts in the image are not appropriately matched in the reconstructed area.

Figure 7 presents the same image fixed with the Photoshop Healing Brush. Intuitively, the problem solved is – reconstruction of the image surface by a Laplace-type equation which has embedded in it specific structure, so that the new surface is no longer smooth, but matches the surroundings in appearance. Jumping ahead, this structure is the connection (covariant derivative) extracted from another area of the image.

A simple way to reconstruct the area of a scratch in a color image (Figure 6) is to replace defective pixels with a solution of the Laplace equation. This is equivalent to minimizing the energy expression

$$\int \partial_\mu F_i \partial_\mu F_i dxdy \tag{12}$$

in the selected area, where we replace the 'defective' pixels with pixels representing smallest sum of gradients squared.

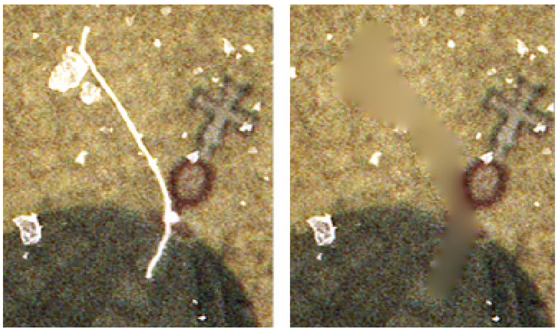

Figure 6: Harmonic reconstruction of image on the left.

This result is not perfect. Obviously, we want not just a smooth fill in, but a fill in with variations similar to those in the surrounding image areas. Reconstuction not having the right texture is a problem.

Let's also notice another problem: (12) implies the wrong image model – color valued function. The image has to be a section in a fubre bundle, not a function. The approach proposed next fixes both problems at the same time.

4.3. Deriving the equation

As a rule, the surrounding image is not smooth, measured by pixel values. It has certain texture. However, this texture changes smoothly in the sense that each small area is consistent with the other areas. We would like to think of a model visual system, perfectly adapted to all image variations (noise, texture) in some representative for this image area. For example it could be film grain or variable shading due to lighting a rough surface like paper or a human face.

We would like to consider such variations and a visual system adapted to them. For such a visual system the image is perfectly smooth: Adapted gradient (covariant derivative) is zero because the adapted visual system discounts for the changes.

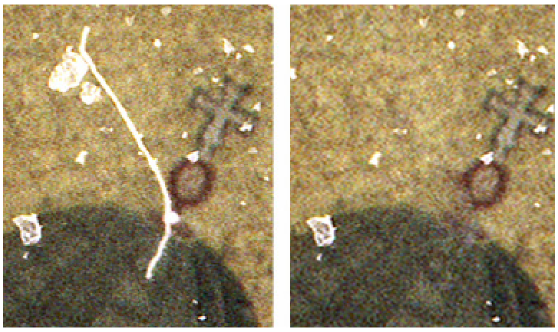

Figure 7: Healing Brush reconstruction of image on the left.

Now we take that adapted visual system and direct it to the defective area that needs 'healing'. Again we want to replace the pixels with a solution that has minimal sum of the adapted gradients squared (12). In other words, we want to minimize the energy

$$\int (\delta^j{}_i \partial_\mu - A^j{}_{i\mu}) F_j (\delta^k{}_i \partial_\mu - A^k{}_{i\mu}) F_k dxdy \tag{13}$$

where A_μ represents adaptation from the previous area. Note that (13) implies that the image is a section in the color bundle, and adapted gradients are connections. The Euler-Lagrange equation minimizing this energy is

$$\triangle F_i = (\partial_\mu A^j{}_{i\mu} + A^k{}_{i\mu} A^j{}_{k\mu}) F_j. \tag{14}$$

If we choose A_μ to be "pure adaptation to an image", corresponding to group parameters φ_n (compare with (9)),

$$A^j{}_{i\mu} = K^j{}_{in}\partial_\mu\varphi_n. \tag{15}$$

Using the expression (which can be verified by simple differentiation)

$$\triangle e^{K_n\varphi_n} = (K_n\triangle\varphi_n + K_n\partial_\mu\varphi_n K_m\partial_\mu\varphi_m)e^{K_i\varphi_i}, \tag{16}$$

(14), (15) can be simplified into

$$\triangle\vec{F} = (\triangle e^{K_n\varphi_n})e^{-K_m\varphi_m}\vec{F} \tag{17}$$

which can be written in a more compact form

$$\triangle\vec{F} = (\triangle\phi)\phi^{-1}\vec{F} \tag{18}$$

using the expression for the von Kries group element $\phi = e^{K_n\varphi_n}$. That's the most general reconstruction or Healing Brush equation.

4.4. One single channel

It may be more intuitive to derive the equation in the simple case of one single channel or a grayscale image F. Also the result is not significantly different. Equation (18) becomes:

$$\triangle F = F\triangle G/G, \tag{19}$$

where G is the sampled image. (19) can be independently derived from the energy

$$I = \int(\partial_\mu - A_\mu)F(\partial_\mu - A_\mu)F dxdy, \tag{20}$$

the Euler-Lagrange equation for which is:

$$\triangle F = \mathbf{A}^2F + \partial_\mu A_\mu F. \tag{21}$$

Here A_μ is extracted from the adaptation area, in which we require that the section represented by G is horizontal, i.e. – complete adaptation to the image $(\partial_\mu - A_\mu)G = 0$ or:

$$A_\mu = \partial_\mu G/G. \tag{22}$$

In other words, the connection is extracted from the adaptation area and used for the reconstruction (healing) process. This gives us the grayscale healing equation (19).

Note:
Here it is easier to see the meaning of "general A_μ" discussed above: The vector field A_μ does not have to be gradient of a scalar field (describing image). In (22) $A_\mu = \partial_\mu \ln G$, but in general it could be any other vector field and then the equation to solve would be (21) instead of (19). A possibility of adaptation described by a non-gradient adaptation field has not been discussed in the literature before. This new idea involves grayscale images that can not be represented as functions, but are known only by their gradients. (Those images sometimes can be represented locally, but not globally.) We believe that the visual system sometimes actually perceives such images in the real world. A potential area of furure research would be how to simulate them with traditional imaging devices in order to achieve similar perception based on inducing appropriate state of adaptation.

4.5. Quality of Result and Healing Brush Algorithm

Assuming variations in the images are 'small' relative to the typical pixel value in the image, we can approximate F/G by a smooth function, or even a constant. This approximation is very close to the result in Poisson editing [4]. It means that we simply multiply the right hand side of Poisson equation by a constant which scales variations appropriately to achieve better fidelity.

This improvement in quality is especially clear when cloning from bright to dark, or from dark to bright areas. In other words, the main difference between our healing equation and Poisson editing is that we produce perceptually better results in cases of change in lighting conditions like shadows and so on.

Also, our theoretical approach reveals the deeper meaning of the 'guidance' field in [4], which is related to the connection (adapted gradient).

In this paper we work only with the Laplace equation but the idea of using connections (covariant derivatives) with the appropriate energy expression can be extended to any PDE, second order or higher.

For the sake of speed and simplicity, we would ignore the details of the rigorous result, and define an easy Healing algorithm, treating each channel separately:

(1) Use as input the difference function $H_{in} = F - G$.
(2) Reconstruct H_{in} (solving PDE) using difference boundary conditions. Result of reconstruction is R.
(3) Define healed image H as $H = R + G$.

The result H of this algorithm is a solution of $\triangle H = \triangle G$ with appropriate boundary conditions.

This algorithm is simpler and faster by solving Laplace instead of Poisson equation. Also, it can be immediately converted into similar biharmonic or higher order algorithm by using bi-Laplace (or other PDE) solver at step (2). As discussed in another paper[2], it has better smoothness properties at the boundary. We have implemented this biharmonic or "bi-Poisson" algorithm for the Healing Brush in Photoshop [1, 2].

In the general case (19) or (21) we need to solve Poisson or "bi-Poisson" equation.

5. HDR COMPRESSION AS RELIGHTING

This section describes another specific application of our lighting/adaptation theory.

A central problem in dealing with high dynamic range images (HDR) is how to display them on a low dynamic range device, like a monitor. Just like scratch removal, the problem of HDR compression can be expressed in terms of relighting. As an example of how our method works, we will reproduce the results of one of the best approaches [17] starting from first principles.

Here is a short review of the algorithm of [17]: Treat only the luminance, F. Calculate logarithm $f = \ln F$; find the gradient of it; attenuate big gradients to reduce dynamic range; then integrate back to get a real image in log space; and finally take the exponent to produce the output luminance.

The logarithm of luminance is used simply because human visual system is approximately logarithmic, and not based on theoretical reasons. Our approach will provide theoretical justification of the use of logarithm.

Minimize the energy written in log-space

$$\int (\partial_\mu f - A_\mu)(\partial_\mu f - A_\mu) dxdy \tag{23}$$

to produce the Poisson equation

$$\triangle f = \partial_\mu A_\mu \tag{24}$$

for the logarthm of luminance, where $\mathbf{A}$ is the attenuated gradient of the log of the input. "Integrate back" in the above algorithm means "solve (24)". Without attenuation, (24) would produce seemless cloning from any image G if $A_\mu = \partial_\mu G/G$. We can also write $g = \ln G$ and then

$$\triangle f = \triangle g. \tag{25}$$

In our approach, the energy expression is written based on requirements for adaptation invariance. In other words, a multiplicative shadow/relighting G on the source image produces an additive to A_μ term in such a way that the new output image is multiplied by the same shadow/relighting. This simple requirement for energy invariance produces the result (24), (25), automatically placed in log-space. The transforms are (see also (8), (9)):

$$F \to GF \tag{26}$$

$$A_\mu \to A_\mu + \frac{\partial_\mu G}{G}. \tag{27}$$

The simplest energy expression that has the above invariance is

$$\int \frac{(\partial_\mu - A_\mu)F(\partial_\mu - A_\mu)F}{F^2} dxdy. \tag{28}$$

If we substitute $\mathbf{A}$ from (22), the Euler-Lagrange equation for this energy would be:

$$\triangle \ln F = \triangle \ln G, \tag{29}$$

which is exactly (25).

Because of the logarithm in our result, we produce exactly (24), the same as [17]. We did not depend on intuition to motivate this use of log space; instead, it comes directly from our mathematical model based on first principles. This can be seen as theoretical motivation for using log space in any visual system.

Note that $\mathbf{A}$ is adaptation vector field, and it can be more general than gradient of a function. We adapt to what we see, and not to the pixel values of energy illuminating the retina. Due to these adaptation effects, what we see is not always representable in pixels or as a picture.

6. CONCLUSION AND FUTURE WORK

The success of our approach in the case of the Healing Brush and HDR suggests the following theoretical idea: In many cases of image processing it is appropriate to treat the image as a section in a fibred space, rather than as a function. A connection can be chosen to represent adaptation or perceptual derivative. In this new formalism we need to use expressions for the energy based on connections (covariant derivatives) in relation to any PDE or other image processing algorithm. This approach is not limited just to the Laplace equation, scratch removal or HDR, but – applicable to any relighting or perception-based application.

7. REFERENCES

[1] Adobe Systems, "Photoshop 7.0 User Guide," Adobe Systems, San Jose, 2002.

[2] T.G. Georgiev, "Photoshop Healing Brush: a Tool for Seamless Cloning," in Proceedings of ECCV, ACV Workshop, pp 1-8, May 2004.

[3] M. Bertalmio, G. Sapiro, V. Castelles, C. Ballester,"Image Inpainting," in Proceedings of ACM SIGGRAPH 2000, pp. 417-424, ACM Press, 2000.

[4] P. Prez, M. Gangnet, A. Blake, "Poisson Image Editing," in Proceedings of ACM SIGGRAPH 2003, pp. 313-318, 2003.

[5] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, "Numerical Recipes in C," Cambridge University Press, 1992.

[6] E. Land, J. McCann, "Lightness and Retinex Theory," Journal of the Optical Society of America, v. 61, no. 1 pp. 1-7, 1971.

[7] W.H. Land, "Recent Advances in Retinex Theory," Vision Res. Vol. 26, No. 1, pp 7-21, 1986.

[8] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker,W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," in Proceedings of SIGGRAPH 2000, pp 145-156, July 2000.

[9] T. Hawkins, J. Cohen, and P. Debevec,"A photometric approach to digitizing cultural artifacts," In Proc. 2nd International Symposium on Virtual Reality, Archaeology, and Cultural Heritage (VAST 2001), pp 333-342, December 2001.

[10] A. Wenger, T. Hawkins, and P. Debevec. "Optimizing Color Matching in a Lighting Reproduction System for Complex Subject and Illuminant Spectra", Eurographics Symposium on Rendering, 2003.

[11] J. Koenderink, A. van Doorn, "Image Processing Done Right," in Proceedings of ECCV 2002, pp. 158-172. Springer 2002.

[12] D. Saunders, "The Geometry of Jet Bundles," Cambridge University Press, 1989.

[13] Y. Choquet-Bruhat, C. DeWitt-Morette, "Analysis, Manifolds and Physics," Amsterdam, North-Holland, 1982.

[14] M. Fairchild, "Color Appearance Models," Addison wesley, 1988.

[15] J. von Kries, "Chromatic Adaptation," Fribourg, 1902. [Translation: D. MacAdam, "Sources of Color Science," MIT press, 1970.]

[16] B. Horn, *Determining Lightness from an Image*, Computer Graphics and Image Processing, 3, pp. 277-299, 1974.

[17] R. Fattal, D. Lischinski, M. Werman, *Gradient Domain High Dynamic Range Compression*, Proceedings of ACM SIGGRAPH 2002, pp. 249-256, 2002.