

# Linear Subspaces of the Appearance Space

Mylo, Marlon  
Fraunhofer FKIE  
Fraunhoferstraße 20  
53343 Wachtberg,  
Germany  
mylo@cs.uni-bonn.de

Klein, Reinhard  
University of Bonn  
Institute of Computer Science II  
Regina-Pacis-Weg 3  
53113 Bonn, Germany  
rk@cs.uni-bonn.de

Manipulating digital optical material representations is still a difficult problem because arbitrary manipulations lead almost certainly to an unrealistic impression of the material. In this paper we present an approach to material editing based on a digital model of the V1-area of the visual cortex. The V1-model is used to define the appearance space as the space of weighted sums of the cortical-model filter responses. We will show that it is possible to transform several optical material manipulation schemes into our editing scheme. As those optical material manipulation schemes may also be physical phenomena, we may introduce a new material edit. Our argumentation will be supported by comparing editing-examples.

## Keywords

Material Editing, BTF, Striate Cortex

## 1 INTRODUCTION

Editing digital representations of the measured reflectance-properties of material surfaces is an intensely studied but still difficult problem. Renderings of 3D-scenes, which give the impression as if they were real are of high significance e.g. in advertisement, film-productions and historical reconstruction projects. Most approaches target at manipulating the underlying physics whereas we present material editing as a matter of influencing the visual perception. Namely we will transfer several approaches to material editing into a computational model of the simple cells of the **primary visual cortex (V1)**. Using models of the visual cortex has a long tradition in computer vision for pattern recognition tasks and for the description of perceptual image-metrics but it is not yet an integral component of computer graphics. We will introduce the term **appearance space** which has mostly been used implicitly [15, 24]. We argue, that the set of all possibly occurring neuronal states in the visual cortex may be seen as this appearance space. So given a computational cortex transform model, we may define a *computational* appearance space as part of it. Seeing computer graphics from the perspective of human physiology is fruitful: Bayer filter in digital image sensors follow the cone distribution in the retina, retinal

displays mimic the cone-density and photo-sensors filter and cumulate the incoming energy comparable to photo receptor cells. Our approach is a kind of frequency editing. Frequency editing is a very old technique. Blurring e.g. has already been used hundred years ago in silent film. But it has mostly not been seen as an operation in the visual cortex but merely as a given visual effect.

Our point is that optical operations in the physical world are mapped to operations in the appearance space. Recent perceptual studies [11] suggest, that some of those cortical operations are linear. We will show that those operations may be applied by scaling of cortical filter responses. In other words we simulate a physiological linearisation of a physical phenomenon. This is in accordance with results from behavioral and brain science, due to which appearances may be seen as representation of optical phenomena, relevant for the human evolutionary adjustment to our environment [43]. Following this idea, we may state that our visual system is the simplest known representation of optics which still allows all possible perceptual manipulations.

After outlining the relevant related literature we will introduce our model of the V1-cortex. In the third section we will give a formal overview, the fourth section will be dedicated to the description and parametrization of the Gabor-base functions underlying our V1-model. In the fifth section we describe how to transfer optical material manipulation schemes into our model.

The presented results (section 6) will support the conclusion that perceptually motivated frequency manipulations may be seen as promising approach to the generation of new virtual materials (section 7).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.



Figure 1: Rendering of an edited wool-BTF. The left part of the image shows the result of a combination of the edge aware operator and the thickening operator (sections 6.2 and 6.3), the right image shows the result of the corresponding band-pass filter, according to [29] and in the middle, we show the original material.

We contribute to the field of material editing by presenting a system to transfer image manipulations into a model of the visual cortex which in many cases brings better results than the original editing scheme and we will provide a novel realistic material manipulation, namely frequency based moving of a light source.

## 2 RELATED WORK

Because of its outstanding role in visualisation, in advertisement and in filmmaking, editing of realistic material-surfaces is a vivid field of research. In the first paragraph we will portray the development in the field of digital cortex-modelling. The second paragraph will be dedicated to literature on comparable image processing schemes. The related work for the manipulation operators will be presented in section 5.

Our understanding of the structure and the modes of action of the animals visual cortex goes back to the work of Hubel and Wiesel during the late 50. and 60. of the last century [16–18]. Twenty years later Daugman fitted Gaussian and Gabor-filters to the cortical responses measured by Hubel and Wiesel [4, 5]. It is noticeable that neural networks develop Gabor-filter like structures by their own, when trained with random input [41]. Olshausen and Field found that optimizing a vector base for sparse linear coding of images leads to a set of Gabor-like base vectors which is in spatial frequency and orientation coverage comparable to the filtering system in the visual cortex [33, 34]. A publication which concentrates on mathematical aspects of the Gabor-filter-systems compatible with the neural responses of the V1-cortex is the work of Lee [23]. Lee gives explicit parameters for his filter systems and calculates the tightness of the Gabor-frames. A good overview over publications on cortical parameter measurements may be derived from [26, Table 1]. Recently Huth et al. published a python-toolbox for simulations of early vision [19].

The presented approach stands in the tradition of the pyramid-based texture analysis and synthesis published

by Heeger and Bergen in 1995 [15]. Heeger and Bergen use steerable pyramids to model the behaviour of the visual cortex. Gutman and Hyvärinen derive a probabilistic model of image statistics by modelling two cortical layers of simple and complex cells [13]. This publication may also be consulted for further references to Bayesian perception. In her dissertation Diana Turcsány [45] uses a convolutional neural network to model the deeper levels of the visual cortex for image editing.

## 3 OVERVIEW AND DEFINITIONS

The insight that light is not coloured but that the energy in a light beam provokes a sensation of colour goes back to Newton [32]. Heeger and Bergen used the word *appearance* to bridge the gap between the sensation of a texture and the physical phenomena on the surface of the texture [15]. We can locate the term *appearance* between the sole occurrence of physical phenomena and the set of sensations by identifying the **appearance space** with the set of all neural response-states in the visual cortex. In the ventral stream of the human visual system, the primary visual cortex follows after the **lateral geniculate nucleus (LGN)**. As receptive fields have directly been measured while exposing the macaque retina to visual stimuli, the influence of the LGN is an implicit part of the model but does not have to be modelled explicitly.

As frame for our (computational) appearance space we will use a **cortex transform model** [46] which we will derive from empirical data (section 4).

Our formal scaffold consists of a **model of the space of retinal responses**, a **model of the neural responses of simple cells in V1**, a **model of the visual stream** from the retina to the neural response and an **interpretation model for the retinal responses**.

**The space of retinal responses** describes the entrance of pictorial data into the visual system. We will use RGB-images with an edge-length of 256 pixels. Decorrelating the color space as in [15, Sec. 3.5], lead to

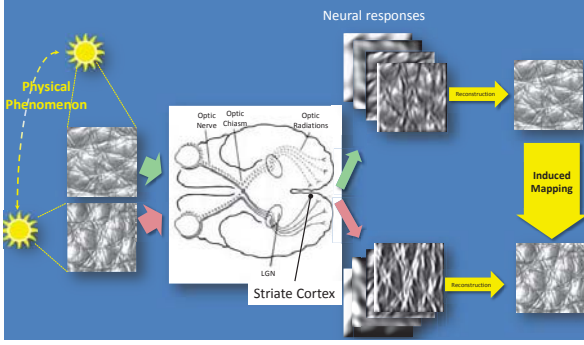


Figure 2: Visual path of a material patch, seen under different optical conditions. The physical phenomenon induces a mapping in the space of cortical responses. Brain drawing taken from <http://universe-review.ca/I10-85-opticpath.jpg>.

strong artefacts. Confining the manipulations to the value-channel of the HSV-color-space brought good results. So we define  $\mathbb{I} := [0, 1]^{256 \times 256}$  as retina model. The images, we use for testing, correspond to real-world patches with an edge length of approximately 5 cm. If a patch of this size has a distance of 57 cm from the observer, its retinal image approximately covers the fovea.

**Our model of the visual stream** is limited to the early ventral stream up to V1. While there have been suggested different filters for modelling V1-receptive fields [25], we use Gabor-filters [4], (see section 4). Our whole V1-model consists of a filter bank of 517 filters ( $\{\Gamma_\psi\}_{\psi \in \Psi}$ , see section 4.2.1).

**The space of neural responses** will be modelled as a stack of matrices  $\mathbb{G} := \mathbb{R}^{256 \times 256 \times 517}$ . We do not limit the amplitude of neural responses. It is not self-telling, that the spatial dimension of the neural responses ( $256 \times 256$ ) equals the dimension of  $\mathbb{I}$  (see paragraph 4.2.2) but it enables a direct comparison between the input and the result of the V1-transform.

**The interpretation space** is a set of mappings  $\mathbb{G} : \mathbb{G} \rightarrow \mathbb{I}$  with  $\mathbb{G} := \sum_{\psi} a_\psi \{\Gamma_\psi \star \mathcal{T}\}$ .

Now we define the **appearance space**  $\mathbb{A}$  as the image of the interpretation space. This leads to the following diagram, modelling the relations, depicted in figure 2:

$$\begin{array}{ccc}
 \mathcal{T} \in \mathbb{I} & \xrightarrow{\{\Gamma_\psi\}} & \{\Gamma_\psi \star \mathcal{T}\} \in \mathbb{G} & \xrightarrow{\mathbb{G}_I} & \mathbb{G}_I \mathcal{T} \in \mathbb{A} \\
 \mathfrak{P} \downarrow & & & \searrow \mathbb{G}_{\mathfrak{P}} & \downarrow \mathbb{E}_{\mathfrak{P}} \\
 \mathfrak{P}(\mathcal{T}) \in \mathbb{I} & \xrightarrow{\{\Gamma_\psi\}} & \{\Gamma_\psi \star \mathfrak{P}(\mathcal{T})\} \in \mathbb{G} & \xrightarrow{\mathbb{G}_I} & \mathbb{G}_I \mathfrak{P}(\mathcal{T}) \in \mathbb{A}
 \end{array} \quad (1)$$

The filter bank  $\{\Gamma_\psi\}$  maps the texture  $\mathcal{T}$  to the neural response space  $\mathbb{G}$ . By the definition of  $\mathbb{G}$  and  $\mathbb{A}$ , we may identify  $\mathbb{I}$  and  $\mathbb{A}$ . Neural responses are recombined to a texture  $\mathcal{T}_\chi := \mathbb{G}_X \mathcal{T}$  in the appearance space. A

physical phenomenon  $\mathfrak{P}$  induces a mapping from the appearance space  $\mathbb{E}_{\mathfrak{P}} : \mathbb{A} \rightarrow \mathbb{A}$  to itself (compare with figure 2). If we identify  $\mathbb{G}_X$  and  $\mathbb{G}_X \circ \{\Gamma_\psi\}$ , the operator  $\mathbb{G}_{\mathfrak{P}}$  may be constructed as linear approximation of  $\mathbb{E}_{\mathfrak{P}}$ .

$$\mathbb{G}_{\mathfrak{P}} \approx \mathbb{E}_{\mathfrak{P}} \quad (2)$$

Note that  $\mathbb{G}_I \approx \mathbb{I}$  is an approximation of the identity on  $\mathbb{A}$ .  $\mathbb{G}_{\mathfrak{P}}$  is the operator, we want to learn. For a full clarification of the symbols, see the following section.

## 4 THE COMPUTATIONAL MODEL OF THE EARLY VISION

In this section we will introduce the V1-model. The concept that the neural response of a simple cell in V1 cortex is linear in the intensity of the incoming optical stimulus is essential not only for the model of the visual pathway [1, 4] but also for all measuring methods of the receptive fields like subspace reverse correlation [39]. The function describing the weighted contribution from each position of the receptive field to the response of this cell is called **weighting function** and may be modelled by a linear filter [46].

### 4.1 An empirically based model of the visual cortex

There exist many publications on the frequency distribution in Macaque V1-area [9, 42]. We used empirical data, measured and fitted by De Valois et al. [6]. We use two dimensional Gabor-base functions for spatial frequency filtering [4]. It is convenient, to introduce the Gabor-filtering system by starting with a transformation of the euclidean plane:

$$R_\Theta \circ T_{\mathbf{p}}(x, y) = \begin{pmatrix} \cos \Theta & \sin \Theta \\ -\sin \Theta & \cos \Theta \end{pmatrix} \begin{pmatrix} x - p_x \\ y - p_y \end{pmatrix} \quad (3)$$

With the point  $\mathbf{p} := \begin{pmatrix} p_x \\ p_y \end{pmatrix}$  and the rotation angle  $\Theta$ . The Gabor base function is the product of a wave-function, called **carrier** ( $\cos$ ), and an Gaussian **envelope** ( $\exp$ ):

$$a\gamma_{\omega, \sigma_\xi, \sigma_\eta, \phi}(\xi, \eta) = ae^{-\frac{1}{2}(\frac{\xi}{\sigma_\xi})^2 - \frac{1}{2}(\frac{\eta}{\sigma_\eta})^2} \cos(\omega\xi + \phi) \quad (4)$$

Here we use  $\xi$  and  $\eta$  for the position to emphasize that it refers to the local coordinate system. The preimage of the directional standard deviation of the Gaussian envelope forms an ellipse. The semi-minor axis, here the  $\xi$ -axis, of this ellipse is according to [23] and [20, Fig. 8A] parallel to the wave-vector of the carrier. We confine to a real plane-wave (see 4.2.1). So, with  $\psi := \{\omega, \sigma_\xi, \sigma_\eta, \phi, \Theta\}$ , we may define:

$$a\Gamma_{\psi, \mathbf{p}} := a\gamma_{\omega, \sigma_\xi, \sigma_\eta, \phi} \circ R_\Theta \circ T_{\mathbf{p}}(x, y) \quad (5)$$

## 4.2 Parameters

To compose the Gabor filter bank, we have to specify the parameters. We distinguish between the parameters, which we set up according to given publications in the field of neuro-science (the parameter set  $\psi$ , paragraph 4.2.1), the position of the filter center  $\mathbf{p}$  (paragraph 4.2.2) and the amplitude  $a$  (paragraph 4.2.3), which we will use for the definition of the editing operator  $\mathbf{G}$ .

### 4.2.1 The parameter-set $\psi$

The parameter-set  $\psi$  contains all parameters which have to be distributed according to measurements in the macaques or in the cats striate cortex.

#### The spatial frequency $\omega$

In the visual cortex, frequency sensitivity occurs not in exact but in rough steps of 0.3 to 0.5 octaves. As we drew the spatial frequency according to [6, Fig. 6.], we were limited to the bin width in this figure, which is 0.5 octaves.

**Differences between human and macaque visual system** The monkey visual system as model for the human visual system has been validated under several different aspects [37]. While the human visual system is from an anatomical and physiological perspective extremely similar to the macaque visual system, it has a slightly higher retinal magnification factor (about  $0.291/0.223$ ), which hints to a higher angular resolution [28]. Therefore we add another frequency bin at  $20.8c/^\circ$  and so we have to extrapolate to a plausible histogram of the human frequency distribution.

**The distribution given by de Valois** In [6, Fig. 6.] De Valois et al. describe their measurements of the spatial frequency distribution of the receptive fields of simple cells in macaques primary visual cortex. They distinguish between the cells with receptive fields in the fovea and in the parafovea region of the retina. We assume that our texture covers a visual angle of  $5^\circ$ . As we cannot expect observers to concentrate on a texture without any eye-movement, we merged the distributions for the fovea and the parafovea by normalized summation. The blue (including green-blue) bars in Figure 3 belong to the merged histogram from [6]. To extend this histogram to the slightly bigger frequency range of the human vision, we fitted a gaussian by an iterative Least Mean Square algorithm, moved the mean of the gaussian to the logarithmic middle of the new frequency distribution range and stretched the standard deviation proportional to the ratio of logarithmic ranges.

The following table shows the number of filters we have in every frequency bin. The absolute number of 517 filters has been chosen in order to have a good fitting to

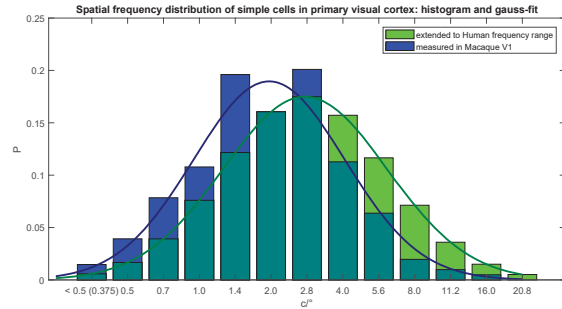


Figure 3: The blue part of the bars shows the histogram given in [6]. The green part describes the extrapolation results and has been added to account for the slightly wider frequency range of human vision [28].

the histogram and still stay comparable with the reconstruction scheme for tight frames (section 6.1).

$c/^\circ$	0.4	0.5	0.7	1.0	1.4	2.0	2.8
# filters	3	8	20	39	64	84	92
$c/^\circ$	4.0	5.6	8.0	11.2	16.0	20.8	
# filters	82	61	37	18	7	2	

#### The standard deviation in direction of the wave-vector $\sigma_\xi$

$\sigma_\xi$  and  $\omega$  are connected via the bandwidth. As Gaussian kernels have infinite support, the bandwidth is defined as **half amplitude bandwidth**. Bandwidths have been drawn on base of [6, Fig. 7]. In this diagram, De Valois et al. visualized the bandwidth with standard deviation as a function of the spatial frequency. As spatial frequencies were known, bandwidth-samples could be drawn under the assumption of normal-distribution within the same frequency range. Given the bandwidth  $B$  and the spatial frequency  $\omega$ , we may calculate:

$$\sigma_\xi = \frac{\sqrt{2 \ln 2} ((2^B + 1)/(2^B - 1))}{\omega} \quad (6)$$

#### The standard deviation orthogonal to the wave-vector $\sigma_\eta$

According to [38, FIG. 4.], there is a relation between  $\omega\sigma_\xi$  and  $\omega\sigma_\eta$ . This relation may be interpreted as functional graph with a small deviation. To make use of this relation, we fitted a cubic spline to the data and used this spline as function graph.

#### The Phase angle $\phi$

To draw the phase parameter  $\phi$ , we used the histograms given in [38, FIG. 7A/B].

## The orientation $\Theta$

By definition of  $\xi = (x - p_x) \cos \Theta + (y - p_y) \sin \Theta$ ,  $\Theta$  is the angle between the  $\xi$  and the  $x$ -axis. We drew the orientation equally distributed from  $\{i\frac{\pi}{8}\}_{i \in \{1, \dots, 16\}}$ . Where possible, directions have been drawn in orthogonal pairs.

All random experiments have been done in several passes and brought comparable results. The set of all parameter-sets  $\psi$  in the Gabor-filter bank, will be denoted by  $\Psi$ .

### 4.2.2 The position $\mathbf{p}$

Every neural measurement provides us just one sample of the domain of neural responses. Be  $\Gamma$  the Gabor filter, best fitting the receptive field of a given neuron with filter center  $\mathbf{p}$ : now the neural response is modelled as  $a \langle \Gamma_{\mathbf{p}}, \mathcal{T} \rangle$ .

$\langle \rangle$  is the standard inner product in the image domain.  $\Gamma$  has to be appropriately sized and evaluated on the spatial grid of the image and  $\mathcal{T}$  has to be zero-padded, where necessary. The filter-centers are often chosen to be elements of the spatial image grid ( $\mathbf{p} \in \{1, \dots, 256\}^2$ ) [5], sometimes with the constant stride ( $h := p_{z_{i+1}} - p_{z_i}$ ) between consecutive grid points increasing with an increasing wavelength and/or starting with a stride smaller than one (e.g. [23]). In order to make use of the convolution theorem and to avoid a resampling step we will assume the parameter set  $\psi$  to be constant over the whole grid and set the stride  $h = 1$  to one and keep the image-grid. Nevertheless we have to emphasize that our approach might distract the statistics: as the statistics of DeValois et al. [6] are based on the measurements of individual cells, a higher spatial resolution goes to the cost of the angular resolution and the variety of the phase values. Particularly in the case of low frequencies, the spatial domain is probably oversampled. The results of the undulation experiment 6.3.2 might indicate this problem (see figure 11).

To locate the neural response, we multiply it by the canonical base matrix  $\mathbf{e}_{\mathbf{p}} \in \mathbb{R}^{256 \times 256}$  at position  $\mathbf{p}$  and sum those matrices up  $\sum_{\mathbf{p}} \langle \Gamma_{\mathbf{p}}, \mathcal{T} \rangle \mathbf{e}_{\mathbf{p}}$ . As we confined to real-valued Gabor-base functions (equation 4), meaning  $\Gamma^* = \Gamma$ , we may write that summation-formula as cross-correlation  $\star$ :

$$a \sum_{\mathbf{p}} \langle \Gamma_{\mathbf{p}}, \mathcal{T} \rangle \mathbf{e}_{\mathbf{p}} = a \Gamma \star \mathcal{T} \quad (7)$$

Where we appoint the amplitude  $a$  to be fixed for a changing position  $\mathbf{p}$ .

### 4.2.3 The amplitude $a$

We use the amplitude to combine filter responses to operators. Be  $\mathbf{E} : \mathbb{A} \rightarrow \mathbb{A}$  an operator, than we want to find  $a_{\psi}^{\mathbf{E}}$  to approximate  $\mathbf{E}$  (see section 5.1):

$$\mathbf{E} \approx \sum_{\psi \in \Psi} a_{\psi}^{\mathbf{E}} \Gamma_{\psi} \star \quad (8)$$

This mapping operates via cross-correlation, it may be visualized by applying it to the discrete dirac  $\delta \in \mathbb{A}$ .

## 5 TRANSFERRING EDITS TO THE MODEL OF THE VISUAL CORTEX

Now, that we have introduced our model of the visual cortex, we want to introduce the operators. First we will discuss the editing scheme itself and how to transform into it. Than we will present the editing paradigms to transfer.

### 5.1 Learning an operator

To transfer a given edit, we take a collection of test-textures  $\mathcal{T}_{i \in \{1, \dots, m\}}$  and solve

$$\mathbf{a}_i^{\mathbf{E}} = \min_{\mathbf{c} \in \mathbb{R}^n} \arg \|\mathbf{E} \mathcal{T}_i - \sum_{\psi \in \Psi} c_{\psi} \Gamma_{\psi} \star \mathcal{T}_i\|_2 \quad (9)$$

for each texture  $\mathcal{T}_i$ . We could define  $\mathbf{a}^{\mathbf{E}} := \min_{\mathbf{c} \in \mathbb{R}^n} \sum_i \|\mathbf{E} \mathcal{T}_i - \sum_{\psi} c_{\psi} \Gamma_{\psi} \star \mathcal{T}_i\|_2$  but this definition lead to undesired activities in higher frequency-bands. Instead, we apply a singular value decomposition to  $\mathbf{A} := a_{i\psi}$ :

$$\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}' \quad (10)$$

and use the base vector  $\mathbf{a}^{\mathbf{E}} = (a_{\psi}^{\mathbf{E}})_{\psi \in \Psi} = (V_{\psi 1})_{\psi \in \Psi}$ . So we may declare our new editing operator

$$\mathbf{G}_{\mathbf{E}} := \sum_{\psi \in \Psi} a_{\psi}^{\mathbf{E}} \Gamma_{\psi} \star \quad (11)$$

### 5.2 The operators

We will explore four different operators: the identity, linear edge enhancement, bandpass filters and spotlight moving.

#### 5.2.1 The identity

The first operator maps the image to itself. This is a reconstruction. There is no canonical reconstruction scheme for Gabor-Wavelets as they are overcomplete. There has been many efforts to produce models of the visual cortex which had good mathematical properties [23, 40]. Lee introduces a reconstruction scheme which relies on the tightness of the frame [7, 23]. There are many approaches, to adjust the filter responses of a Gabor-filter bank to a partition of unity in the frequency domain [46].

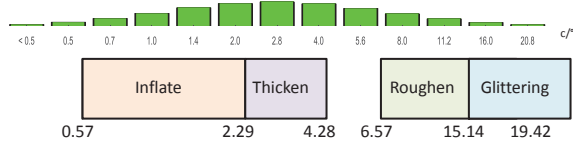


Figure 4: Under the frequency distribution, we show the range of the edits of the editing scheme.

### 5.2.2 Edge aware imaging

Edge aware imaging has been subject of intensive study during the last years. Bilateral filters [36,44] are among the most popular tools for edge-aware image processing. One recent approach gives a linear approximation of a bilateral filter [30]. He et al. suggest to improve the edge-preservation property of filters by the use of a guidance map [14]. Paris et al. use Laplacian Pyramids for strengthening or weakening edges in images: they argue, that edges are a jump in not only one level in the laplacian pyramid but merely in all levels [35]. Laparra et al. use those insides to build a system for perceptually optimized image rendering [22]. Fattal [8] detects edges by the use of second generation wavelets. Using Gabor-filters for edge detection has a long tradition, e.g. [27].

### 5.2.3 Bandpass filtering

**Affordance** is a concept from psychology, introduced by Gibson [10], and describes the possibilities of actions which may be done on a given object. By user studies, Giesel and Zaidi found a relation between certain affordances or material properties and spatial frequency bands in material-images [11]:

0.57-2.29	$c/^\circ$	Inflated and deflated
2.29-4.28	$c/^\circ$	Deep and flat
6.57-15.14	$c/^\circ$	Soft and rough
15.14-19.42	$c/^\circ$	Sparkling and dull

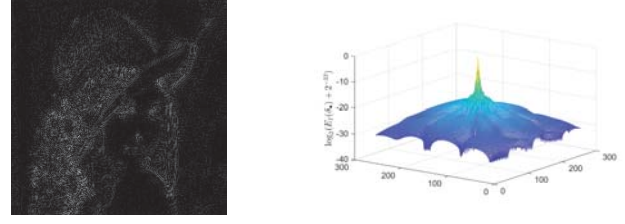
The connection between affordance and spatial-frequencies gives rise to a semantically founded editing scheme by simply enhancing or weakening particular frequency bands [12, 29]. Because the underlying physical effects are too complicated, those effects may not be seen as the result of inverse optics [12] and are therefore examples for complex physical operations with a linear representation in the visual cortex. It is striking, that those manipulations cover nearly the whole frequency range of the visual cortex (figure 4).

### 5.2.4 Moving spotlight

Given a directionally illuminated texture patch. We will show, that it is possible in our model to learn and reproduce small movements of the light source.

## 6 RESULTS

In this section we want to show and discuss some results. The presented results have been calculated on



(a)  $500 \cdot |\mathcal{L} - \mathbf{E}_I(\mathcal{L})|$

(b)  $\log(\mathbf{E}_I(\delta) + 2^{-32})$

Figure 5: 500 times amplified reconstruction error of Lena image. Boundary cut off in a distance of 10 pixel. (a). To get a better impression of the delta-spike, we added  $2^{-32}$  and applied the binary logarithm (b).

colour or reflectance maps. The colour maps had a dynamic of 48 dB and the reflectance maps had a dynamic of 96 dB. All operations on the HDR-images had been performed in log-space. As training samples, we used textures from the USC-SIPI Image Database from the University of Southern California and the describable texture dataset [2].

## 6.1 Identity

Figure 5b visualizes the learned identity operator. The difference between the reconstructed and the original image is with bare eyes intractable. The maximum pixel intensity difference between the Lena image ( $\mathcal{L}$ ) and the reconstruction of it was  $\max_{ij} |\mathcal{L}_{ij} - \mathbf{G}_I(\mathcal{L})_{ij}| = 6.9 \cdot 10^{-3}$  which corresponds to 2 steps in an 8 bit grey-scale image. While such a small deviation will not stand-out when affecting the intensity channel, sensible people might perceive colour aberrations if the operator was applied channel-wise to an RGB-image.

As it is not possible, to reconstruct an Gabor-filtered image perfectly, we will compare against the approximative reconstruction scheme, presented by Lee [23]: a frame  $\{\Gamma_\beta\}$  (for a definition and constraints on the parameter set  $\beta$ , see [7, 23]) is tight when the following equation holds for a given constant  $c$  and a small positive number  $\varepsilon$ :

$$\forall \mathcal{T}: \quad c \|\mathcal{T}\|^2 \leq \sum_{\beta} |\langle \Gamma_\beta, \mathcal{T} \rangle| \leq (c + \varepsilon) \|\mathcal{T}\|^2 \quad (12)$$

Lee investigated for which parameter sets  $\beta$  this frame becomes a tight frame ( $\varepsilon \searrow 0$ ). Note that Lee uses complex-valued Gabor-base functions, which does not make sense in our setting as we do not apply filters to filtered values and have therefore no complex multiplications. In his definition of the Gabor-base functions, the amplitude  $a$  is part of the definition of  $\Gamma$  and the position  $\mathbf{p}$  is an element of the parameter set  $\beta$  and he uses a pyramid sampling scheme. For a tight frame the following reconstruction formula may be applied:

$$\mathcal{T} \approx \frac{2}{2c + \varepsilon} \sum_{\beta} \langle \mathcal{T}, \Gamma_\beta \rangle \Gamma_\beta \quad (13)$$

To compare against [23], we sample over 16 directions  $\theta$ , made three steps per octave and set the stride  $h$  to 0.5. This yields a value for  $\varepsilon$  of approximately 0.0001, the number of base-vectors was 864.

Figure 6 visualizes that even with this very tight frame the quality of this reconstruction scheme is not high enough to allow for applications in computer graphics.



Figure 6: The upper right of the image shows the Lena image reconstructed by the formula 13. The wavelet family forms a frame with  $\delta = 0.0001$ . The lower left shows our reconstruction.

## 6.2 Edge aware imaging

For this edit, we learned randomly linear edge aware filtering kernels: we used Gabor and Sobel-filters  $((1, 2, 1)' \otimes (1, 0, -1))$ , we will write:  $\mathbf{SX}, \mathbf{SY}$ . Note that the parameters of the filter kernels and the intensity of the filters had been drawn randomly and so they were in general not in the set  $\Gamma_\Psi$ . Intensities were always enhanced. We used 1000 editing samples of varying photos for learning. The resulting filter (figure 8) may be seen as the average of all projected filter-kernels. It is a good approximation of the sign-inverted discrete Laplace operator with weights on the diagonals. In comparison with other state of the art edge aware imaging operators (figure 7), it is noteworthy, that the learned operator enhances very fine structure and the material still looks realistic. A physical effect, bound to this appearance, is a higher fibrousness.

## 6.3 Affordance editing

In this section we will compare our results against pure frequency band scaling. While there is evidence, that the frequency-bands are subject to a recognition step [12] and consecutively to a scaling step in the visual cortex, according to the original perceptual studies, an edge length of a material patch should cover a viewing angle of  $3.5^\circ$ , this corresponds to an observer-distance of approximately 82 cm. We will confine to the roughen and the undulation operation. The thicken and the glitter-operator will be compared on bidirectional texture functions (section 6.5).

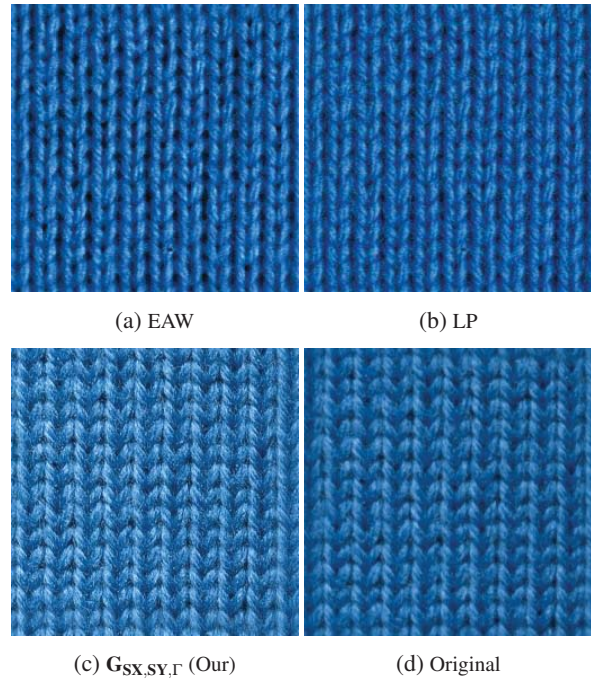


Figure 7: In the top row you can see as comparison the results of two non-linear edge filters: the edge avoiding wavelets of Fattal et al. (7a, [8]) with an exponent of 1.15 (slightly enhancing fine details, see publication) and the local Laplacian filters of Paris et al. (7b, [35]) with  $\sigma_{\text{publ}} = 0.2$  and  $\alpha_{\text{publ}} = 0.2$ . The bottom row shows the result of our algorithm (7c) and the original material (7d).

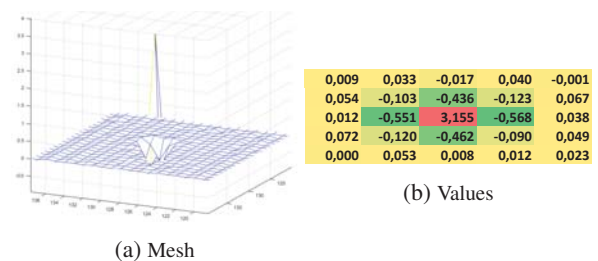


Figure 8: The learned filter.

### The comparison edit

The influence of the absolute value for the strength of the edit is not directly comparable. The learned edits were mostly weaker than the originals. To compensate for that, we made a relaxation step based on the HDR VDP 2.2-metric as published by Mantiuk et al. in 2015 [31], by scaling the edit with a positive number  $s$  with

$$s := \operatorname{argmin}_{r>0} |d(\mathcal{T}, \mathbf{E}\mathcal{T})_{VDP} - d(\mathcal{T}, r\mathbf{G}_E\mathcal{T})_{VDP}| \quad (14)$$

to minimize the visual difference between  $\mathbf{E}$  and  $\mathbf{G}_E$ . Of course  $r\mathbf{G}_E := r(\mathbf{G}_E - \mathbf{G}_I) + \mathbf{G}_I$

#### 6.3.1 Roughening

Roughening seems to work comparably good in the Fourier-domain (operator  $\mathbf{F}$ ) and in the Cortex-filter-

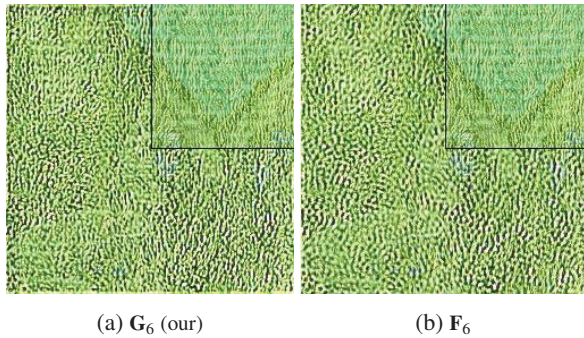


Figure 9: Comparison of the roughening filter. The original material in the top right corner.

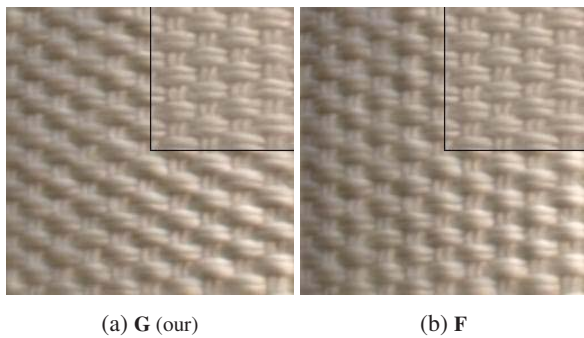


Figure 10: Inflating a material. The original material in the top right corner.

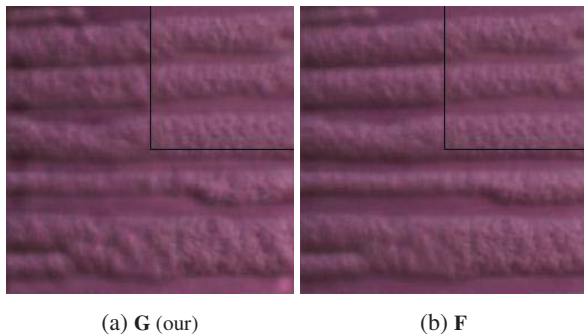


Figure 11: Inflating a material. The original material in the top right corner.

bank. For stronger edits (images 9a and 9b) our operator shows less artefacts.

### 6.3.2 Inflation

For relatively small structures, the undulation-operation works slightly better in the cortical filter bank (image 10a) than in the fourier domain (image 10b, operator **F**). For bigger structures, the manipulation in the cortical filter bank is not capable of reproducing the results of the bandpass filtering in the Fourier-domain (figure 11).

## 6.4 Spotlight moving

To make the moving spotlight experiment, we used the BTF-measurements of the UBO14-database of the university of Bonn [47]. For learning, we used the leather

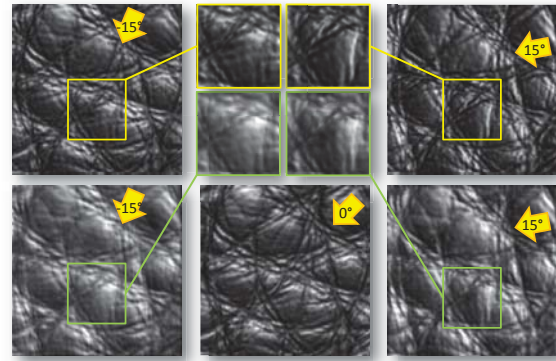


Figure 12: In the middle of the bottom row, you can see the original material test-patch. To compare against the real physical operation, we show in the top row a photograph of the same patch, illuminated under an azimuthal angle of  $-15^\circ$  (left) and illuminated under an azimuthal angle of  $15^\circ$  (right). We compare those results against the application of the spotlight moving-operator (bottom row, left ( $-15^\circ$ ) and right ( $15^\circ$ )). Here we show only the value channel of the material patch.

materials with the numbers 1-3 and 5-12. The testing results will be presented on the leather4 material. The camera position had been in the zenith above the material. Material-patches which were illuminated from a polar angle of approximately  $30^\circ$  against the zenith and from an azimuthal angle of  $0^\circ$  were considered as unedited material samples. We interpreted material patches, taken under the same conditions but illuminated from an azimuthal angle of  $15^\circ$  or  $-15^\circ$  respectively as the edited versions of the original material patch and used those patches for learning the motion of the spotlight. The results are presented in figure 12 and in a short movie in the additional material.

We can see that small moves of a spotlight can be represented and learned in the cortical domain.

## 6.5 Editing of high-dimensional material representations

Image based modelling of spatially varying and measured material reflectance properties has been introduced by Dana et al. [3] in form of bidirectional texture functions (**BTFs**). An approach to editing BTFs is, to deal with BTFs as with textures. In 2007, Kautz et al. showed, that applying operators from picture editing to the spatial or to the angular domain of a BTF may bring reasonable results [21] and introduced several different operators to BTF-editing. Our BTF editing approach is comparable to the frequency band scaling, published by Mylo et al. [29]. A detailed description of editing compressed BTF-data may be found there. In figure 1 we compare against the thicken and the glittering-operator from the band scaling approach ([29], see sec-



tion 5.2.3). Here the results of our V1-editing approach are clearly superior.

Energy preservation and other expressions of physical phenomena are lost after the editing step. Instead of using the suggested thicken or the undulation-operator, one may estimate the surface structure and operate on the new geometry. For higher frequencies, the inverse optics are highly complicated.

## 6.6 Time requirement

Experiments have been done on an *i7-4770 CPU @ 3.4 GHz* with 8 GB RAM. Learning took between 51' (spotlight moving, 6.4) and 90' (edge aware imaging, 6.2). The editing step took about 0.4' for a texture and 40' for a BTF.

## 7 CONCLUSION

In this work we have presented different linear editing operations based on a model of the V1-region of the visual cortex. We could show, that it is possible to reconstruct material patches in an appropriate quality by simple summation of the filter responses of the suggested Gabor-filter bank (section 6.1). We learned band-pass filtering which shows in many cases less artefacts when the corresponding band-pass filter in the fourier domain (section 6.3). This effect might be due to the immanent pre-filtering but we have to pronounce that prefiltering the bandpass in the fourier-domain is difficult because it varies between destroying the effect and producing strong sidelobes. Applying those learned operators to a BTF brought notably better results than the corresponding band-pass filters (section 6.5). Above that, we could also learn a physical effect (section 6.4).

An important subset of the appearance space is the set of all realistic appearances, meaning appearances which are inter-subjectively considered as pictorial representations of a real environment like e.g. photos. We have shown, that starting with a valid element of the space of realistic appearances, the presented operators define an affine linear subspace with limited diameter.

## 8 REFERENCES

- [1] B. Andrews and D. Pollen. Relationship between spatial frequency selectivity and receptive field profile of simple cells. *The Journal of physiology*, 287(1):163–176, 1979.
- [2] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, , and A. Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014.
- [3] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real-world surfaces. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 151–157, 1997.
- [4] J. G. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision research*, 20(10):847–856, 1980.
- [5] J. G. Daugman. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(7):1169–1179, 1988.
- [6] R. L. De Valois, D. G. Albrecht, and L. G. Thorell. Spatial frequency selectivity of cells in macaque visual cortex. *Vision research*, 22(5):545–559, 1982.
- [7] R. J. Duffin and A. C. Schaeffer. A class of nonharmonic fourier series. *Transactions of the American Mathematical Society*, 72(2):341–366, 1952.
- [8] R. Fattal. Edge-avoiding wavelets and their applications. *ACM Transactions on Graphics (TOG)*, 28(3):22, 2009.
- [9] K. Foster, J. P. Gaska, M. Nagler, and D. Pollen. Spatial and temporal frequency selectivity of neurones in visual cortical areas v1 and v2 of the macaque monkey. *The Journal of physiology*, 365(1):331–363, 1985.
- [10] J. J. Gibson. Perceiving, acting, and knowing: Toward an ecological psychology. *The Theory of Affordances*, pages 67–82, 1977.
- [11] M. Giesel and Q. Zaidi. Adaptation reveals frequency band based inferences of material properties. volume (In press), 2012.
- [12] M. Giesel and Q. Zaidi. Frequency-based heuristics for material perception. *Journal of vision*, 13(14):7–7, 2013.
- [13] M. U. Gutmann and A. Hyvärinen. A three-layer model of natural image statistics. *Journal of Physiology-Paris*, 107(5):369–398, 2013.
- [14] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE transactions on pattern analysis and machine intelligence*, 35(6):1397–1409, 2013.
- [15] D. J. Heeger and J. R. Bergen. Pyramid-based texture analysis/synthesis. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 229–238. ACM, 1995.
- [16] D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574–591, 1959.
- [17] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–154, 1962.
- [18] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–243, 1968.
- [19] J. Huth, T. Masquelier, and A. Arleo. Convis: A toolbox to fit and simulate filter-based models of early visual processing. *bioRxiv*, page 169284, 2017.
- [20] J. P. Jones and L. A. Palmer. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of neurophysiology*, 58(6):1187–1211, 1987.

- [21] J. Kautz, S. Boulos, and F. Durand. Interactive editing and modeling of bidirectional texture functions. In *ACM Transactions on Graphics (TOG)*, volume 26, page 53. ACM, 2007.
- [22] V. Laparra, A. Berardino, J. Ballé, and E. P. Simoncelli. Perceptually optimized image rendering. *arXiv preprint arXiv:1701.06641*, 2017.
- [23] T. S. Lee. Image representation using 2d gabor wavelets. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18:959–971, 1996.
- [24] S. Lefebvre and H. Hoppe. Appearance-space texture synthesis. *ACM Transactions on Graphics (TOG)*, 25(3):541–548, 2006.
- [25] T. Lindeberg. A computational theory of visual receptive fields. *Biological cybernetics*, 107(6):589–635, 2013.
- [26] F. Mechler and D. L. Ringach. On the classification of simple and complex cells. *Vision research*, 42(8):1017–1033, 2002.
- [27] R. Mehrotra, K. R. Namuduri, and N. Ranganathan. Gabor filter-based edge detection. *Pattern recognition*, 25(12):1479–1494, 1992.
- [28] W. H. Merigan and L. M. Katz. Spatial resolution across the macaque retina. *Vision research*, 30(7):985–991, 1990.
- [29] M. Mylo, M. Giesel, Q. Zaidi, M. Hullin, and R. Klein. Appearance bending: A perceptual editing paradigm for data-driven material models. In *Vision, Modeling & Visualization*, 2017.
- [30] P. Nair, A. Popli, and K. N. Chaudhury. A fast approximation of the bilateral filter using the discrete fourier transform. *Image Processing On Line*, 7:115–130, 2017.
- [31] M. Narwaria, R. K. Mantiuk, M. P. Da Silva, and P. Le Callet. Hdr-vdp-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images. *Journal of Electronic Imaging*, 24(1):010501–010501, 2015.
- [32] I. Newton. Opticks: Or a treatise of the reflexions, refractions, inflexions and colours of light. 1704.
- [33] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607, 1996.
- [34] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research*, 37(23):3311–3325, 1997.
- [35] S. Paris, S. W. Hasinoff, and J. Kautz. Local laplacian filters: Edge-aware image processing with a laplacian pyramid. *ACM Trans. Graph.*, 30(4):68–1, 2011.
- [36] S. Paris, P. Kornprobst, J. Tumblin, F. Durand, et al. Bilateral filtering: Theory and applications. *Foundations and Trends® in Computer Graphics and Vision*, 4(1):1–73, 2009.
- [37] R. Rajalingham, K. Schmidt, and J. J. DiCarlo. Comparison of object recognition behavior in human and monkey. *Journal of Neuroscience*, 35(35):12127–12136, 2015.
- [38] D. L. Ringach. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of neurophysiology*, 88(1):455–463, 2002.
- [39] D. L. Ringach, G. Sapiro, and R. Shapley. A subspace reverse-correlation technique for the study of visual neurons. *Vision research*, 37(17):2455–2464, 1997.
- [40] E. Salinas and L. Abbott. Do simple cells in primary visual cortex form a tight frame? *Neural computation*, 12(2):313–335, 2000.
- [41] T. D. Sanger. Analysis of the two-dimensional receptive fields learned by the generalized hebbian algorithm in response to random input. *Biological cybernetics*, 63(3):221–228, 1990.
- [42] B. Selby. Development of an integrated model of primary visual cortex. Master’s thesis, University of Waterloo, 2016.
- [43] R. N. Shepard. Perceptual-cognitive universals as reflections of the world. *Behavioral and brain sciences*, 24(4):581–601, 2001.
- [44] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE, 1998.
- [45] D. Turcsány. *Deep learning models of biological visual information processing*. PhD thesis, University of Nottingham, 2016.
- [46] A. B. Watson et al. The cortex transform- rapid computation of simulated neural images. *Computer vision, graphics, and image processing*, 39(3):311–327, 1987.
- [47] M. Weinmann, J. Gall, and R. Klein. Material classification based on training data synthesized using a btf database. In *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part III*, pages 156–171. Springer International Publishing, 2014.