Bachelor Thesis

Comparison of two multiscale spatial filtering models

Sebastian Keil

TU Berlin, Computer Engineering B.Sc.
Supervisor: Dr. Joris Vincent

TU Berlin, Computational Psychology

December 13, 2024

Summary

1 Introduction **1**
1.1 Light and the Visual System . . . 1
1.2 Low-Level Vision . . . 3
1.3 Modelling Human Vision . . . 5
1.4 ODOG and BIWaM . . . 6

2 Methodology **8**
2.1 Structure of the thesis . . . 8
2.2 Stimuli . . . 9
2.3 Model responses . . . 10
2.4 Preparations and Tools . . . 10

3 Robustness **11**

4 Filter parameters **12**

5 Aligned models **13**

6 CSF weighting **14**

7 Spatial interaction **15**

8 Discussion **16**

1 Introduction

1.1 Light and the Visual System

In the environment, light is emitted by a source of illumination, such as the sun. Any surface on which this light falls will reflect a portion of it as *luminance*. This portion is called the *reflectance* of the surface. The luminance is therefore the result of *illuminance* and reflectance, as shown in Figure 1 a). A lightmeter can measure the amount of luminance reflected by a surface, but it cannot tell what the reflectance of this surface is, because the luminance could be the product of any combination of illuminance and reflectance.

$$L = I \cdot R \tag{1}$$

The formula 1 shows the problem from a mathematical perspective. L represents luminance, I and R illuminance and reflectance, respectively. It is simply impossible to solve for I and R, when only L is been measured, since for every R there is an I to produce the measured L (Adelson, 2000).
However, the human visual system can solve this problem. It is also processing only luminance, yet it is able to generate the perception of reflectance, which is referred to as lightness. Figure 1 b) shows an abstraction of this process.

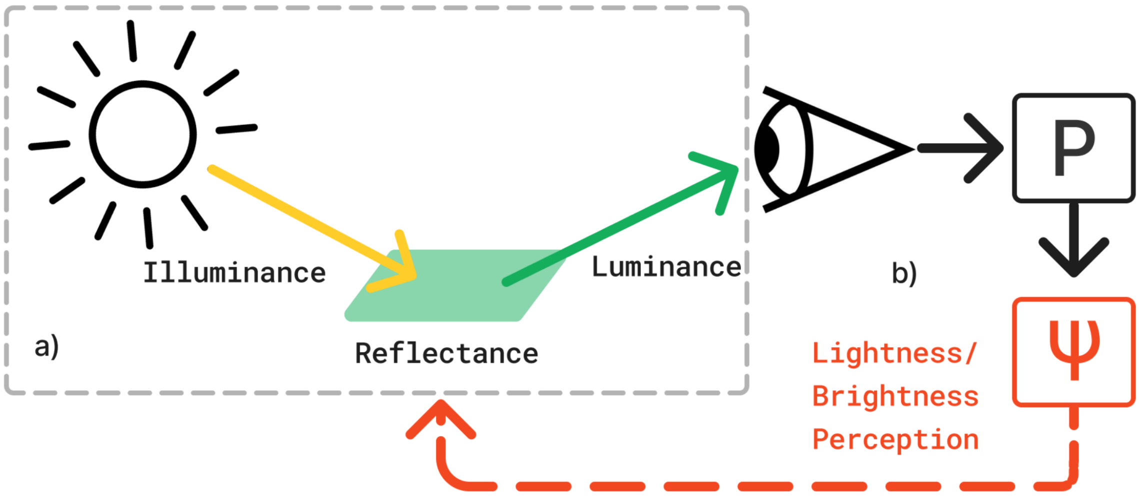


Figure 1: a) Relationship between illuminance, reflectance and luminance. Luminance is the result of illuminance and reflectance. b) The human visual system processes luminance through an unidentified mechanism, represented by P which is determining lightness and brightness perception ψ of the surface.

Next to lightness, humans also perceive brightness, which is the perception of luminance, seen in Figure 1 b). Unlike reflectance, luminance is directly available at the retinal image and brightness could in principle be derived from the retinal measurement. However, the visual system does not only consider the corresponding luminances when evaluating the perceived lightness and brightness of image areas, it also takes into account the luminances of the surrounding regions (Kingdom, 1997). As a result the perception can differ from the actual retinal information. The mechanisms through which the visual system accomplishes these tasks are the subject of current research and will be discussed further in the following sections.

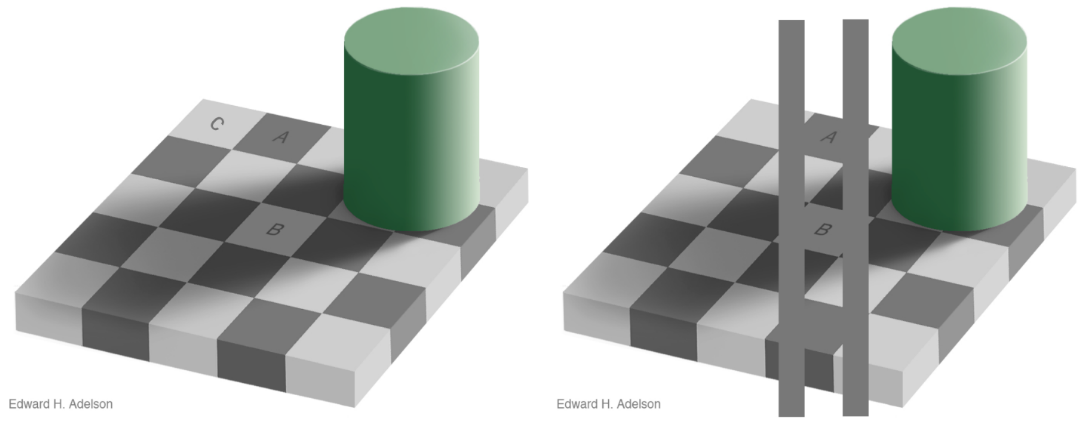
Figure 2: Lightness and brightness are distinguishable when illumination is visible. *"The walls of the house appear uniformly white — a lightness judgment — yet are brighter in some places than others — a brightness judgment"*. Quote and picture from Kingdom (2014).

The distinction between lightness and brightness becomes apparent when information about illumination is visible, as seen in Figure 2. *"The walls of the house appear uniformly white — a lightness judgment — yet are brighter in some places than others — a brightness judgment"* (Kingdom, 2014). This distinction is important because brightness is about the relationship between the object and its environment. In other words, it reveals how the object is exposed to illumination. Lightness, on the other hand, represents the intrinsic properties of the object, such as color, regardless of the environment.

Since reflectance is only implicitly perceivable, it can lead to uncertain situations. For instance, a shadow can dim an area so that a white surface within the shadow reflects the same amount of light as a black surface in full illumination next to it. Despite this, human observers can usually distinguish between the white and the black surface (Arend, 1993).

This phenomenon is illustrated by Edward H. Edelson's *checkerboard shadow illusion*, shown in Figure 3. The two patches A and B on the checkerboard in Figure 3a appear to have different colors, even though they are emitting the same light, as shown in Figure 3b.

The cylinder seems to cast a shadow, even though there is no real light source, since the image is just a two-dimensional representation of the scene. However, the visual system is designed to process images coming from three-dimensional scenes with illumination and shadows and so it processes the checkerboard shadow illusion with all the available information about depth and illumination. To logically follow the processing, one can say, that patch B reflects the same amount of light as patch A (seen in Figure 3b), but is located in the shadow of the cylinder and therefore must have a higher reflectance. In other words, the visual system needs to react to differences in illumination and compensate for them in order to estimate the reflectance. This behavior ensures that the perception of a scene is closely related to the reflectance of its surfaces and is largely unaffected by the illumination.



(a) (b)

Figure 3: (a) The Checkerboard shadow illusion image; (b) proof image (Adelson, 1995).

In everyday life, illusions in brightness and lightness perception are rare, likely due to the vast amount of information that the visual system can make use of. Shading, shadowing, and spatial depth can guide the perception of brightness and lightness, like in the checkerboard shadow illusion. However, there are illusions with less information available for the visual system, which need a different explanation.

1.2 Low-Level Vision

One very simple illusion is the classic *simultaneous contrast illusion*, as seen in the upper part of Figure 4. Two identical grey squares appear to be different in brightness, depending on their surroundings. The lack of information about illumination and spatial depth is crucial in comparison with the checkerboard shadow illusion. The parts of the visual system, which are processing illumination or spatial depth, will have no information available in the simultaneous contrast illusion. Here the idea of detecting illumination and compensating for it will fail, hence a different explanation is needed.

An explanation for the simultaneous contrast illusion is offered by neural units in the retina. Hering (1834 —1918) was the first describing their *center surround fields*, which compare luminance areas with their surrounding areas. With that they could account for the simultaneous contrast illusion. The lower part in Figure 4 illustrates the principle. The blocks under the illusion represent neurons responding to areas of the illusion image. The surrounding blocks subtract and the center block adds their responses to the fourth block underneath. When the surrounding blocks are sensing the darker surrounding of the left patch, their response is small and so the subtraction is small. As a result the left summing block receives a higher response, correlating with a human observer experiencing the left patch to be brighter. On the right patch the surrounding blocks are sensing bright surroundings and so the subtraction is higher and the summing block receives a lower response, also correlating with a human observer experiencing the right patch to be darker.

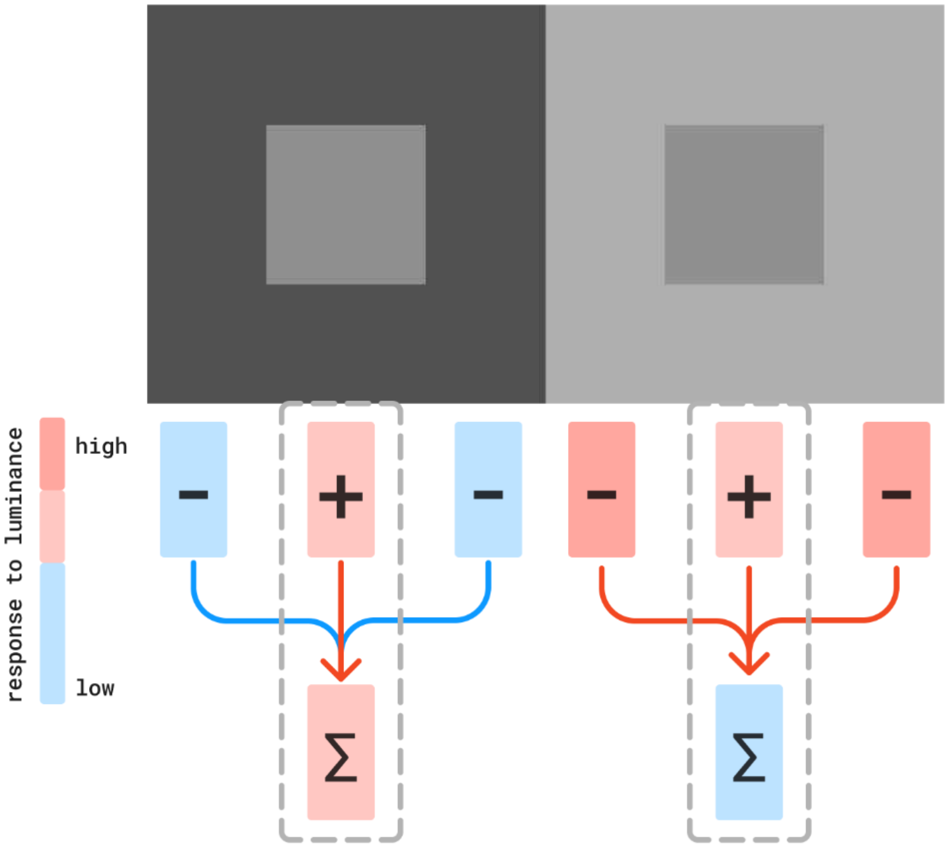


Figure 4: Above: The simultaneous contrast effect. Two identical grey patches appear to be different in brightness, depending on their surrounding. The left patch appears brighter than the right patch.
Below: The principle of center surround fields. The blocks on each side represent neurons, where the surrounding blocks subtract and the center block adds their responses to the fourth block underneath. On the left side the center response is medium (light red), corresponding to the left patch in the illusion, while the surround response is low, corresponding to the surrounding dark gray in the illusion. The summation results in a medium response. On the right side the center is the same, but the surround response is higher (dark red) and so the summation in the fourth block is lower. The responses of the patches in the illusion depends on their surrounding.

Simple mechanisms like the center surround fields could be responsible for human brightness perception[1]. They exist in different sizes and their outputs are also interacting with each other. This complex neural processing results in so called *sensory channels*, where each channel is selectively sensitive to different sizes of contrast areas, also referred to as spatial frequencies (Sachs et al., 1971). Large center surround fields will respond to low spatial frequency information like large objects and gradual changes across the image. Small center surround fields will respond to high spatial frequency information like fine details and edges.
The organization of center surround fields can isolate specific image properties and the interaction of their outputs provides a mechanism to process the retinal information in a much more meaningful way. This has inspired researchers to investigate the computational modeling of center-surround fields and their interactions.

[1]The terms brightness and lightness become synonymous without information about illumination and will be used interchangeably in the following sections, as we will discuss only such illusions

1.3 Modelling Human Vision

The basic idea of center surround fields is to compare luminances with their surround luminances. Since computers handle image data as discrete pixel values, it is mathematically straight forward to model this comparison with algorithms. A common approach is to design a convolution filter, representing the center surround field and convolve it with the image pixel values. In Figure 5 the principle of a convolution on a grayscale image is shown. The filter values are applied on the input pixels by an element-wise multiplication. In the example of Figure 5 the filter is comparing every pixel with its direct neighboring pixels.

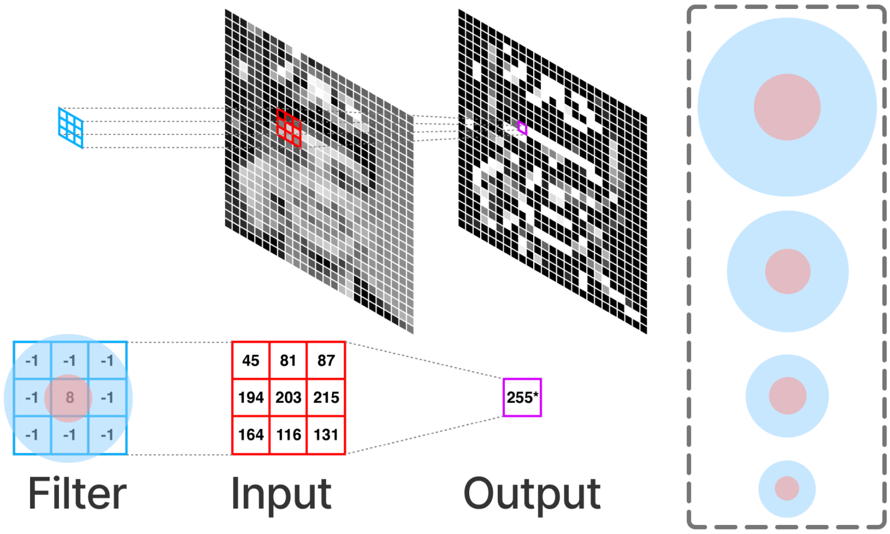


Figure 5: Applying a convolution on an image is a common approach to model center surround fields. Every pixel of the input image is multiplied with the center value of the filter (8) and then summed up with every surround pixel multiplied with the corresponding value in the filter (-1). The resulting sum is the new pixel value for the output image at the position of the original pixel. the left Figure is inspired by Gundersen (2017). The right side shows a filterbank, existing of multiple filters in different sizes.

To address the benefits of different sized center surround fields of the retina, it is sufficient to create multiple filters of varying sizes. On the right in Figure 5 a so called filterbank is shown. Each of the different sized filters will be applied on the input image and generate its own output, similar to the sensory channels of the visual system. Larger filters will compare more pixels and will respond stronger to low spatial frequencies. Smaller filters will respond to high spatial frequencies. The outputs from all the filters of a filterbank are representing the image information decomposed by frequencies. In order to reconstruct the image, it is sufficient to sum the filter outputs. The reconstructed image is identical to the original image, since all information is kept during the process.

To replicate human perception the DOG (Difference Of Gaussian) model from Blakeslee and McCourt does a reweighting of the filter outputs before reconstruction (Blakeslee & Mccourt, 1997). Inspired by the contrast sensitivity function of the human visual system, it attenuates lower frequencies. As a consequence the output image is no longer identical to the input image. In fact for the simultaneous contrast illusion the left patch is brighter and the right patch is darker, aligning with the human perception of the illusion.

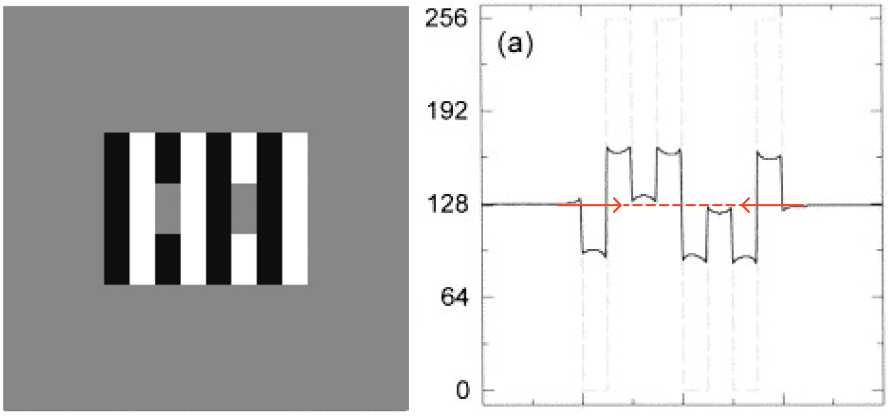


Figure 6: In White's Effect (left) the shift in perceived brightness is in the opposite direction compared to brightness contrast. Both patches are identical, but the left patch on the black bar appears to be brighter than the patch on the white bar, even if it shares most of its edges with white surfaces (White, 1979). The diagram on the right shows the processed White's Effect illusion by the ODOG model. The dashed line refers to the luminance profile across the horizontal center of the illusion. The solid line represents the models output along the same line. The red markers indicate that the models output is in accord with the human perception (Blakeslee & Mccourt, 1999).

The reweighting between decomposition and reconstruction appears to be a key mechanism of the DOG model making it capable of replicating human perception of the simultaneous contrast illusion. However, the DOG model cannot account for all brightness phenomena. For instance, the White's effect, illustrated on the left side of Figure 6, cannot be explained by Blakeslee and McCourt's initial model. One year later, in 1999, they developed an extended version of the model, called ODOG (Oriented Difference Of Gaussian), which can account for White's effect shown on the right side of Figure 6 (Blakeslee & Mccourt, 1999). Two changes are responsible for its new capabilities. They extended the filterbank by six different orientations for each filter, therefore they also changed the isotropic filters of the DOG model to anisotropic filters in order to be able to rotate them. The second change is a normalization step before reconstruction. The ODOG model will be discussed in more detail at a later point in the thesis.
The ODOG model became very famous and inspired a whole family of so-called spatial filtering models. Among them is a model with a different approach, the BIWaM (Brightness Induction Wavelet Model) model from Otazu. It doesn't use a filterbank to decompose the image, but rather convolves its filters within a wavelet transformation (Otazu et al., 2008).

1.4 ODOG and BIWaM

ODOG and BIWaM are oriented spatial-filtering models, highlighting the importance of low-level vision. Both models process an input image using filters, which are inspired by the center surround fields. As shown in Figure 7, their processing begins with the decomposition of the input image (Step 1), using spatially scaled and oriented filters. Each filter results in an channel that captures different orientations and spatial frequencies of the image. The channels are then processed (Step 2) by reweighting and normalization, each model has it own strategy. In the final step (Step 3) all outputs are merged to reconstruct the output image.

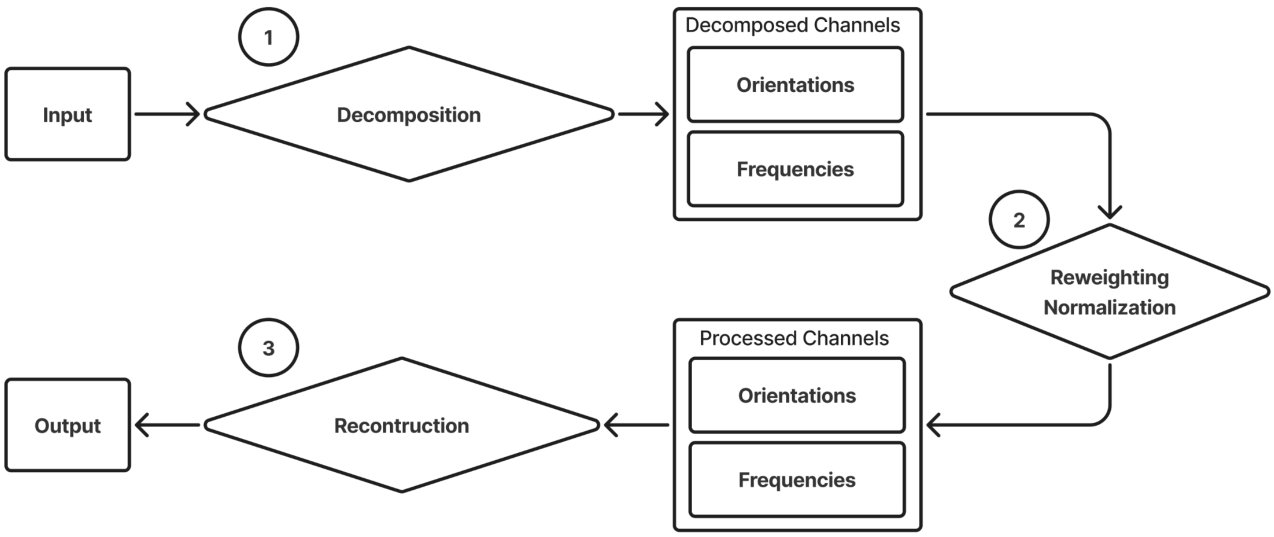


Figure 7: Structural overview of ODOG and BIWaM models, three steps to analyze

The ODOG model uses a filterbank consisting of 42 filters in seven different scales and six different orientations. These filters are applied on the input image by a convolution for every filter, to create the frequency specific and oriented channels. The outputs of the filters within the same orientation are summed, with weights that are determined by the spatial frequency. Lower frequencies receive smaller weights than higher frequencies, similar to the contrast sensitivity of the visual system. These responses are normalized by their root-mean-square energy, which is computed across all pixels and summed to yield the model output. As a consequence of the response normalization, orientations with little energy in the input image will have a proportionally larger influence on the model output (Betz et al., 2015).

For the decomposition in the BIWaM model, a wavelet transformation is used instead of a filterbank. This wavelet transformation also performs convolutions with different filters. Only three orientations are used, but a wider range of scales from 1 cpd (cycles per degree) to the Nyquist frequency. It iterates through the decomposition recursively, downsampling the image at each iteration but using the same filter. Therefore, it generates channels for different filter-to-image scales in each iteration, similar to the ODOG model. After decomposition, all channels are reweighted by a function that is also inspired by the contrast sensitivity function of the visual system. For normalization, they use factors generated for specific areas of the image, resulting in a local normalization. This could be a main difference between the models (Betz et al., 2015). In the final step, the BIWaM model merges the filter outputs to generate the output image.

ODOG and BIWAM may seem different at first glance, because of their distinct processing techniques. But some processes are similar to those in the other model. What exactly is making a difference? Is the decomposition and reconstruction interchangeably in both models, does the processing on the channels make the difference? These questions arise when dealing with both models and are the focus of this thesis. The aim is to explore whether they process identical information as a subset of the other, or if they address distinct concepts.

2 Methodology

2.1 Structure of the thesis

Spatial filtering models are providing insights in the field of vision research and come with advantages and disadvantages. On some images the choice of the model can make a significant difference(TODO: Example). But these differences in model behavior are not trivial. To better understand the models algorithmic and the resulting effects, i plan to study two the models ODOG and BIWaM, which are known for... (TODO: why these models?). Due to the complexity of the models, my first goal is to catalog parameters and link them to specific behaviors for each of the models, then discuss similarities and differences. The methodological approach to the goal is to define meaningful parameters in the source codes and find their effects on model response by modifying them. After that i should be able to align the models and narrow down the search for differences. As there are a lot of test cases i will cover on several stages of both models, the thesis will not have a classic methods and results structure, instead there will be five sections which built up on each other. Each section will have some methods and ends in some results. In the Discussion chapter i will link results between the sections and compare the models more comprehensively. To provide a red line through these sections, there will be an overview at the beginning of each section like Figure 8.

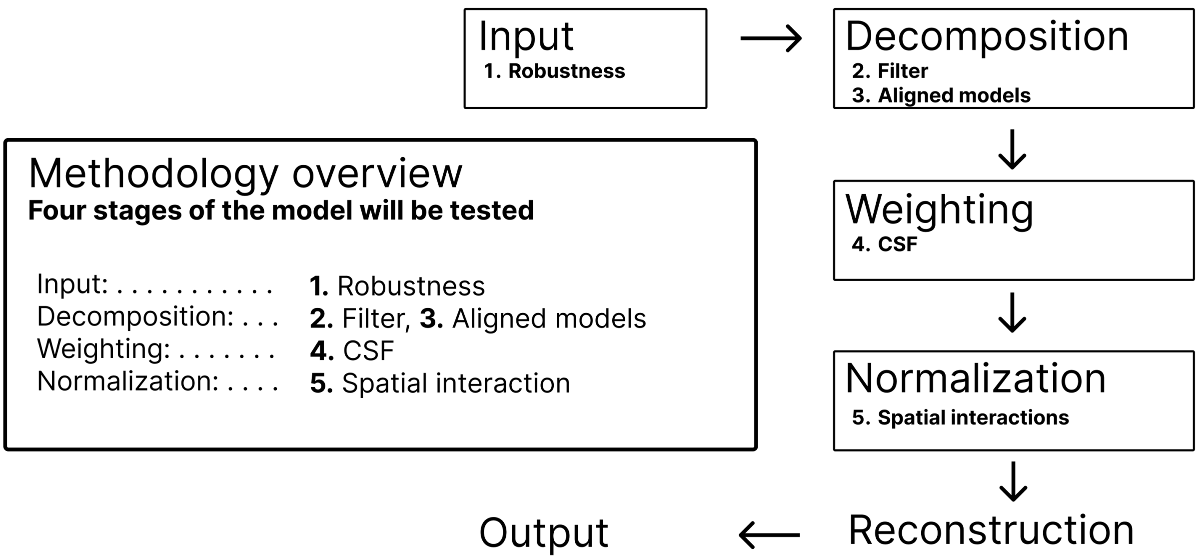


Figure 8: Methodology overview of analysis

Each of the five sections (1. - 5.) in Figure 8, are correlating with a processing stage of the models. For each stage there are specific parameters which have an impact on the models response. These parameters will be studied and prepared for the next section, so the sections built up on each others.

2.2 Stimuli

Both models are tested on a set of 7 different stimuli. The variety in stimuli lowers the risk of drawing conclusions based on specific features of a single stimulus. The set includes two types of brightness perception phenomena: Assimilation and contrast phenomena. The stimuli used are: Two versions of Whites effect, two versions of Simultaneous Brightness Contrast (SBC), Checkerboard illusion, Circular Whites effect and the Dungeon illusion. The last one is specifically chosen, because the authors of the BIWaM model mentioned, that their model will account for it, while it was previously unexplained (Otazu et al., 2008). A List of all stimuli is shown in Figure 9.

(TODO: tell more about the stimuli, why these?)

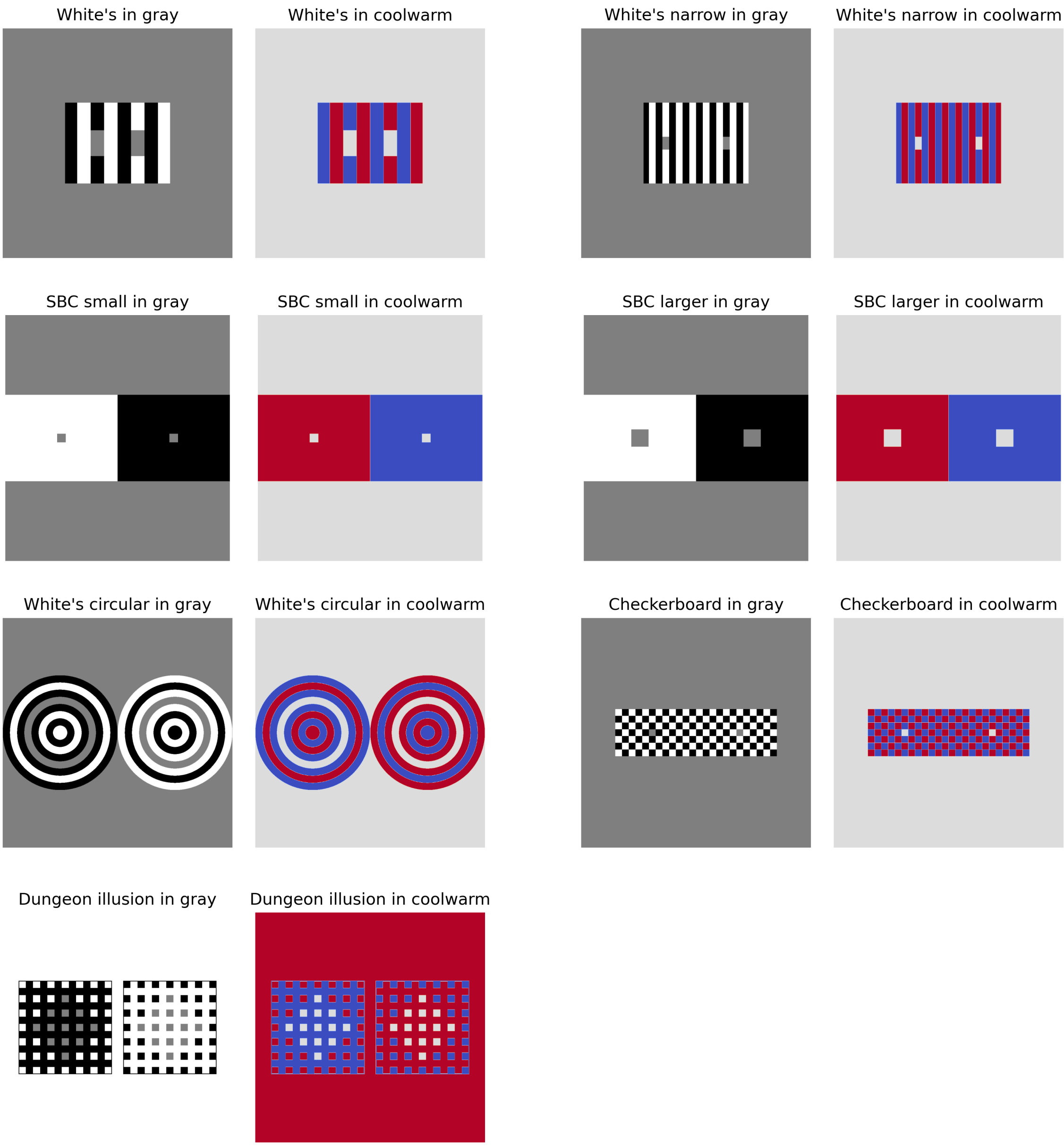


Figure 9: 7 different stimuli are used to test the models. Each stimuli is shown in greyscale and in coolwarm, to avoid the illusion take effect when being observed.

2.3 Model responses

TODO: explain ways to compare model responses with the center line plot, a mean patch difference and a difference image (?)

2.4 Preparations and Tools

TODO: write out the bullet points

- To modify the models, it was necessary to have access to their source codes
- for ODOG i used the Multyscale library (version 0.2)
- for BIWaM i used a python reimplementation of the available Matlab code (which is tested to work the same, see appendix)
- I recreated the input stimuli using the Stimupy package (version 1.1.2).
- For the thesis i used Python 3.12.7, Jupyter Client 8.6.2, Matplotlib 3.9.1, Numpy 2.1.3, Scipy 1.14.0

3 Robustness

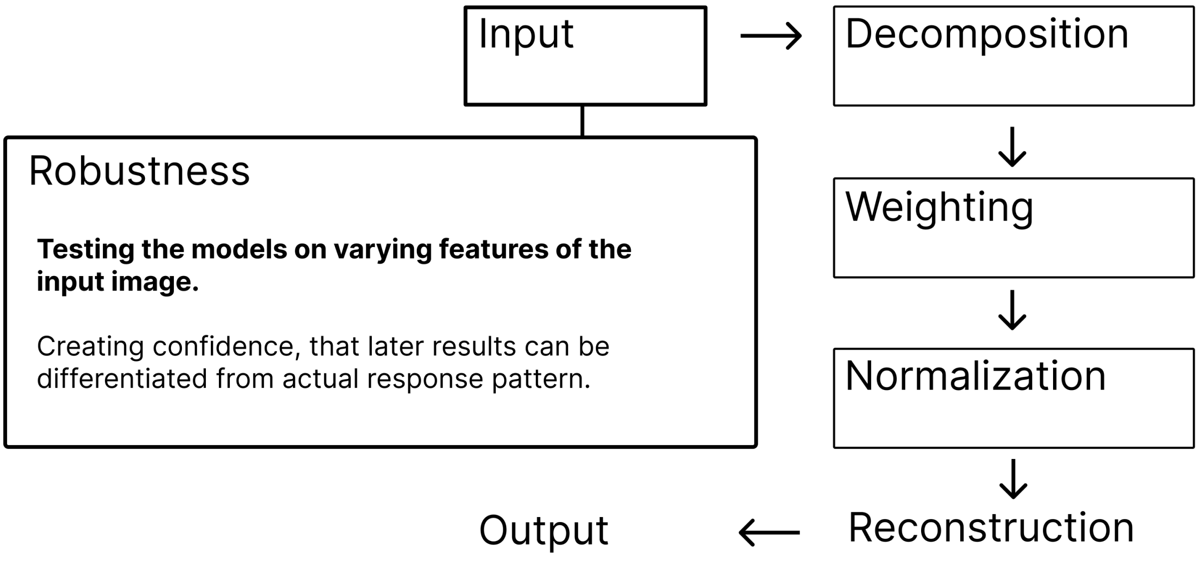


Figure 10: Methodology overview of robustness testing

Before changing parameters, I want to ensure that the patterns of the model responses on the stimuli are robust. This step ensures that any changes I observe later are more meaningful and can be attributed to parameter adjustments rather than the input stimuli.

By observing the output patterns under these controlled changes, I aimed to answer two key questions:

- What are the overall patterns of the models for the 7 test stimuli?
- Does the output changes in predictable ways?

To evaluate robustness, I applied both ODOG and BIWaM to a set of test stimuli that underwent controlled, minor modifications. These modifications included: Brightness of target patches, brightness of background, size of image content, size of target patches.

TODO: go through robustness test, explain what i did, show the plots and point out the predictable behavior.

The robustness observed in both models provides a solid foundation for parameter tuning. Since the outputs are consistent and predictable under minor changes to the input images, I am confident that future changes in the results can be attributed to parameter adjustments rather than the variability of the stimuli.

4 Filter parameters

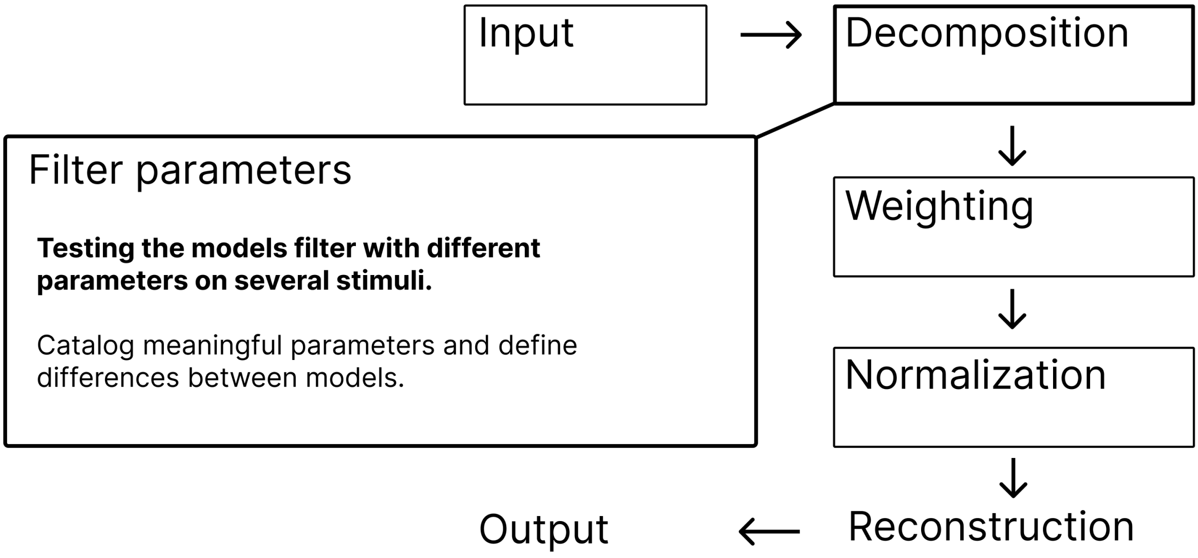


Figure 11: Methodology overview of filter testing

To further understand the behavior of ODOG and BIWaM, I analyzed the effects of the filter parameters during the decomposition stage. This evaluation aimed to catalog the parameters by identifying their impact on model responses.
Parameters tested
ODOG:

- Orientations: ODOG uses oriented filter spread across 0-180 degrees. The original model has 7 orientations.
- Scales: The filter bank's scales define the amount of extracted spatial frequencies. The original model has 6 scales.

BIWaM:

- Wavelet Levels: The depth of the wavelet decomposition defines the amount of extracted spatial frequencies.
- Residual Image: Whether to include the residual image in the wavelet transformation. This residual contains information not captured by the filters and is used during reconstruction.

TODO: go through parameter tests, show the effects and outline why the residual is making a big difference.

To ensure a fair comparison between the models in the next section:

- ODOG and BIWaM will be tested using same number of orientations and scales.
- BIWaM will be evaluated without the residual image.

5 Aligned models

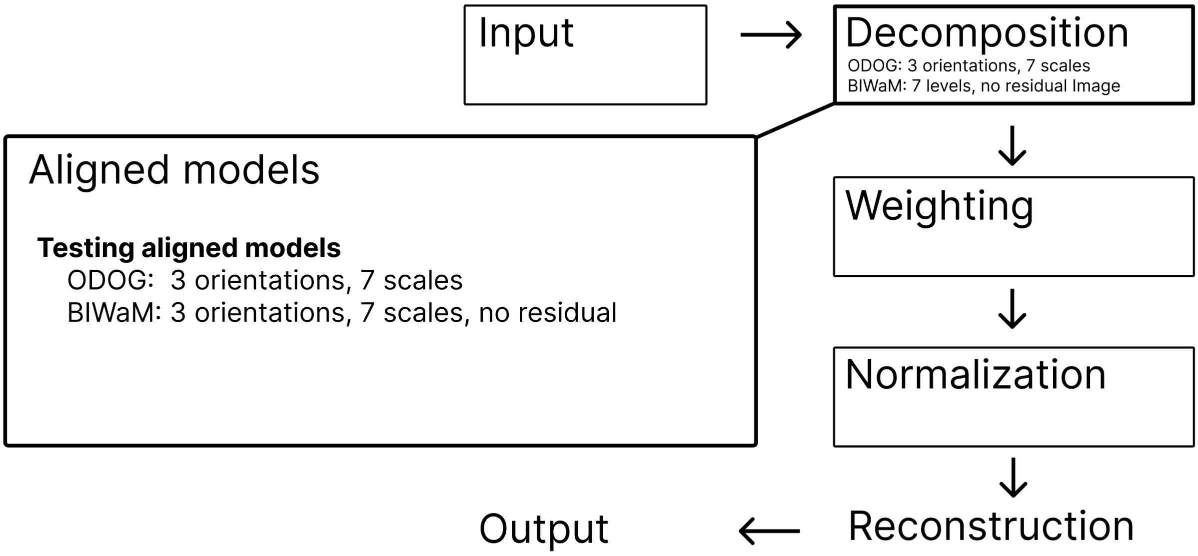


Figure 12: Methodology overview of testing aligned models

To compare ODOG and BIWaM in a fair setting, I evaluated both models on all seven test stimuli with aligned decomposition stage as established in the previous section. The parameters were set as follows:
ODOG:

- 7 scales
- 3 orientations

BIWaM:

- 7 wavelet levels
- No residual image

Each of the seven test stimulus was processed through both models with the aligned parameters. The outputs were analyzed for changes in the response pattern, sensitivity to features of the stimuli (e.g. edges, brightness) and the prediction of the illusion.

TODO: go through tests on aligned models, show the effects of the alignment and outline that the aligned models are getting closer.

This comparison demonstrates that the decomposition stage of the models, when aligned, exhibit similar overall behavior while retaining distinct characteristics based on the following stages (weighting and normalization). These findings establish a baseline for further alignments.

6 CSF weighting

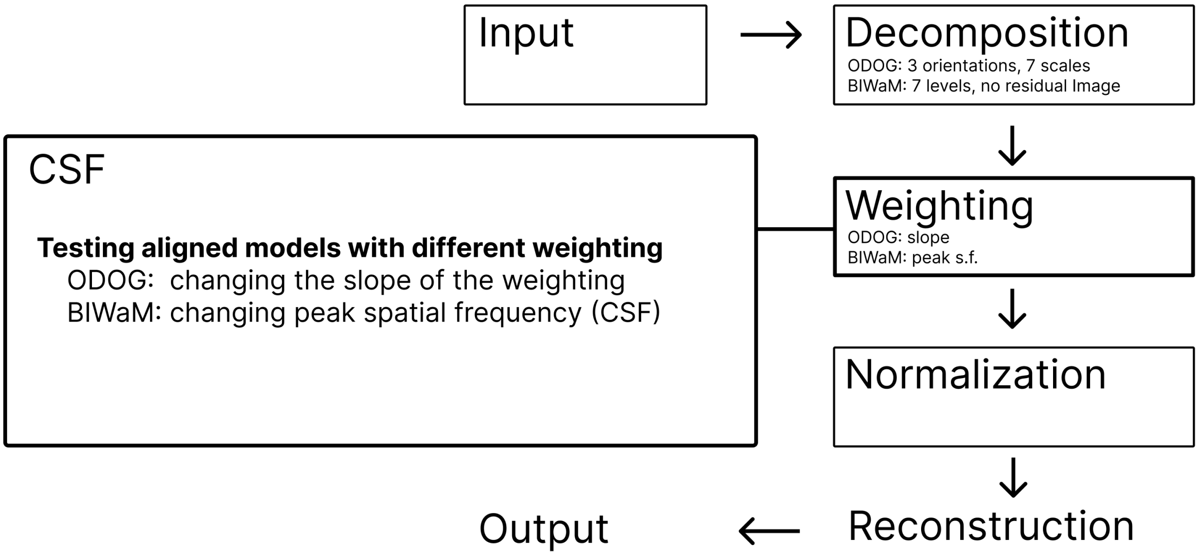


Figure 13: Methodology overview of CSF testing

With the models aligned at the decomposition stage, the next step involves analyzing their weighting mechanisms for frequency channels. Both models incorporate mechanisms to weight spatial frequencies, inspired by the characteristics of the Contrast Sensitivity Function (CSF) in human vision and can be accessed in the source code.
Parameters tested
ODOG:

- Slope in Weighting Step: This parameter defines the slope of a linear function around 1, where lower frequencies are attenuated (factor < 1), and higher frequencies are enhanced (factor > 1). The weighting follows a pattern similar to the CSF, since the scales are spaced logarithmically and therefore the linear function effectively is the positive side of a gaussian.

BIWaM:

- Peak_SF: This parameter shifts the peak of the CSF function along the x-axis. The CSF is modeled using two Gaussians, allowing the model to enhance lower or higher frequencies based on the value of Peak_SF.

TODO: go through both tests on aligned models, show the effects of the different weightings.

7 Spatial interaction

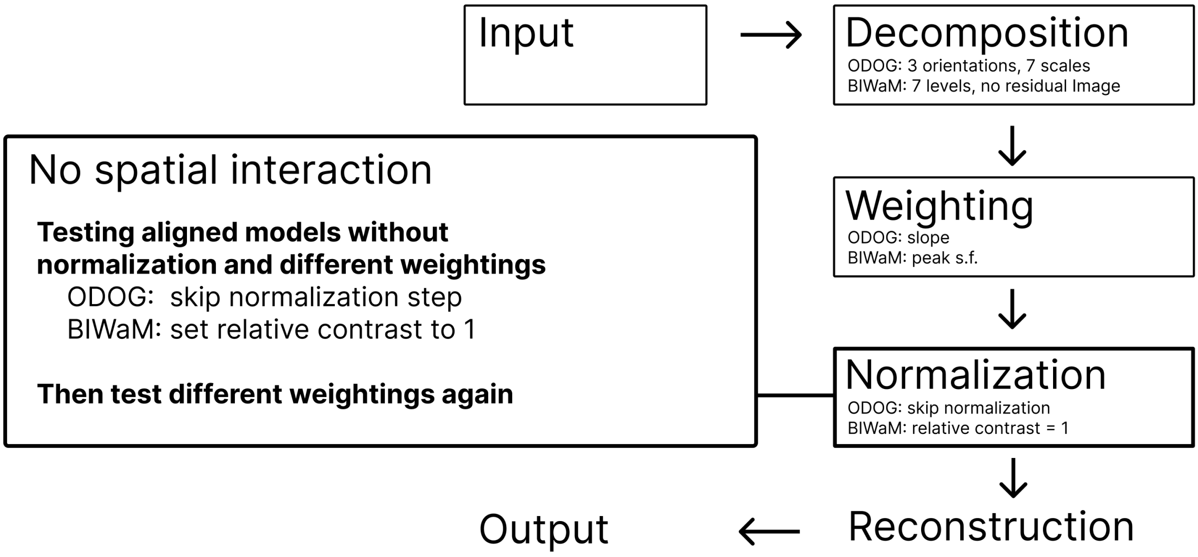


Figure 14: Methodology overview of testing models without spatial interaction

The final analysis focuses on the spatial interaction mechanisms of the models, implemented through normalization processes. To evaluate the importance of normalization, the models were tested with this feature disabled. To confirm the role of normalization, the models were also tested with different CSF weightings as used in the previous section. This helped determine whether CSF weighting alone could account for the models' behavior in the absence of normalization.

ODOG:

- The normalization step was skipped entirely.

BIWaM:

- The relative contrast was set to 1, effectively disabling normalization.

TODO: go through the tests on aligned models without normalization, show the effects of it and then go through CSF test without normalization again.

This analysis underscores the critical role of normalization in brightness prediction. It also highlights the limitations of frequency weighting alone, emphasizing the need for both (if previous section also shows need for CSF) mechanisms to achieve robust and realistic model behavior.

8 Discussion

1. **Linking differences in structure with results in behavior**
2. ODOG is literally a Fourier decomposition. It gets a base frequency components and then its harmonics as the filters are logarithmic spaced. The magnitude of every component gets modified by CSF(slope)
3. Resolution of image effects the models. link to hendrik
4. Over-/Undershooting, log-gabor filter as better filter for both models.
5. **Insights into decomposition**:
   - Information loss through downsampling is compensated due to residual image.
6. Speed of models

List of Figures

1 Relationship between illuminance, reflectance and luminance 1
2 Lightness and brightness are distinguishable 2
3 Checkerboard shadow illusion 3
4 The principle of center surround fields 4
5 Applying a convolution on an image 5
6 Models response to White's Effect 6
7 Structural overview of models 7
8 Methodology overview of analysis 8
9 Overview 7 different stimuli 9
10 Methodology overview of robustness testing 11
11 Methodology overview of filter testing 12
12 Methodology overview of testing aligned models 13
13 Methodology overview of CSF testing 14
14 Methodology overview of testing models without spatial interaction 15

References

Adelson, E. H., & Pentland, A. P. (1996). The perception of shading and reflectance.

Adelson, E. H. (1995). Checkershadow illusion. https://persci.mit.edu/publications

Adelson, E. H. (2000). Lightness perception and lightness illusions. *The New Cognitive Neurosciences, 2nd ed., M. Gazzaniga, ed. Cambridge, MA: MIT Press,* 339–351.

Arend, L. E. (1993). Lightness, brightness, and brightness contrast.

Betz, T., Shapley, R., Wichmann, F. A., & Maertens, M. (2015). Noise masking of white's illusion exposes the weakness of current spatial filtering models of lightness perception. *Journal of Vision*, *15*. https://doi.org/10.1167/15.14.1

Blakeslee, B., & Mccourt, M. E. (1997). Similar mechanisms underlie simultaneous brightness contrast and grating induction.

Blakeslee, B., & Mccourt, M. E. (1999). A multiscale spatial filtering account of the white effect, simultaneous brightness contrast and grating induction. www.elsevier.com

Brainard & Longere, K.
bibinitperiod. (2003). Colour constancy: Developing empirical tests of computational models. In R. Mausfeld & D. Heyer (Eds.), *Colour perception: Mind and the physical world* (pp. 308–326). Oxford University Press.

Gundersen, G. (2017). From convolution to neural network. https://gregorygundersen.com/blog/2017/02/24/cnns/

Hanson, A. R., & Riseman, E. M. (1977). *Recovering intrinsic scene characteristics from images*. Academic Press.

Kingdom, F. A. (1997). Simultaneous contrast: The legacies of hering and helmholtz [PMID: 9474338]. *Perception*, *26*(6), 673–677. https://doi.org/10.1068/p260673

Kingdom, F. A. (2014). Brightness and lightness. *MIT Press*, 499–509.

Murray, R. F. (2021). Lightness perception in complex scenes. https://doi.org/10.1146/annurev-vision-093019

Otazu, X., Vanrell, M., & Párraga, C. A. (2008). Multiresolution wavelet framework models brightness induction effects. *Vision Research*, *48*, 733–751. https://doi.org/10.1016/j.visres.2007.12.008

Robinson, A. E., Hammon, P. S., & de Sa, V. R. (2007). Explaining brightness illusions using spatial filtering and local response normalization. *Vision Research*, *47*, 1631–1644. https://doi.org/10.1016/j.visres.2007.02.017

Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-frequency channels in human vision. *J. Opt. Soc. Am.*, *61*(9), 1176–1186. https://doi.org/10.1364/JOSA.61.001176

White, M. (1979). A new effect of pattern on perceived lightness. *Perception, 8(4)*, 413–416. https://doi.org/10.1068/p080413