Predicting 3D Visual Discomfort using Natural Scene Statistics and a Binocular Model

Zeina Sinno[a] and Alan C. Bovik[b]

[a,b]Laboratory for Image and Video Engineering, The University of Texas at Austin, Austin, Texas, 78712

ABSTRACT

When humans observe stereoscopic images, visual discomfort may be experienced in the form of physiological symptoms such as eyestrain, a feeling of pressure in the eyes, headaches, neck pain, and more. These sensations can arise in cortical mechanisms related to early visual processing. For example, vergence eye movements and lens accommodation can provide conflicting information to the brain if the stereo images are distorted or presented on a flat display. Over the past decade, significant effort has been applied to understanding and characterizing how discomfort arises, towards being able to design safer and more comfortable 3D displays and to provide better guidelines on how to design, align, and capture 3D images and videos. Part of solving this problem consists of objectively predicting the visual discomfort that may arise from viewing a given pair of stereo images that are distorted. Researchers have built several models based primarily on cortical mechanisms that yield good predictions of visual discomfort. Here we study the role of natural scene statistics (NSS) of the disparity maps of stereoscopic images and their relationship to 3D visual discomfort. In particular, we focus on bivariate NSS models. We also build a new prediction model that combines information from binocular vision and the NSS models of disparity maps to accurately predict 3D visual discomfort, and we demonstrate that an algorithm that realizes the prediction outperforms other existing predictors.

Keywords: Binocular Vision, 3D Visual Discomfort, Vergence-Accommodation Conflict, Natural Scene Statistics

1. INTRODUCTION

The positioning of the two eyes in the front of the head, horizontally aligned but separated, allows them to obtain slightly different retinal images. The brain combines the left and the right images to obtain a fused, 'cyclopean' image. Based on the distance between corresponding points in the left and right images, disparity information is extracted and depths are computed. This ability facilitates a variety of exteroceptive and visuomotor tasks,[1] especially in the comprehension of complex visual presentations and those requiring hand-eye coordination.[2]

The binocular processing centers of the brain capture differences between the left and the right images obtained from the eyes. For humans to perceive depth correctly, the two images need to align closely. If for some reason they do not, then visual discomfort may be experienced.[3] 3D visual discomfort can take several different symptoms, including eye strain, nausea, fatigue and headaches.[4] There are several explanations of experienced visual discomfort when viewing stereo displays, including the eyewear required to present images to the two eyes, ghosting or cross-talk between the images, misalignment of the images, inappropriate head orientation, vergence-accommodation conflicts, visibility of flicker or motion artifacts, and visual-vestibular conflicts.[3] The vergence-accommodation conflict has often been identified as the primary culprit causing visual fatigue.[5,6]

Vergence describes the mechanism of binocular eye-movement that directs the eyes towards an object. When fixation on an object moves closer or farther, the eyes converge or diverge, respectively. Accommodation describes the mechanism of adjusting the focal power of the crystalline eye lens to acquire a clear and sharp retinal image of an object. As the object moves closer or farther, the focal power increases or decreases respectively.[7] In a natural environment, both mechanisms are coupled and occur in parallel. More importantly, the amount of

Further contact information:
E-mails: zeina@utexas.edu - bovik@ece.utexas.edu.

accommodation required to put an object into focus is proportional to the amount of of vergence needed to fixate on the object.[8] The human visual system (HVS) has evolved towards associating these processes neurologically; triggering vergence stimulates accommodation, and vice versa. Stereoscopic displays stimulate accommodation and vergence in an unnatural way, resulting in vergence-accommodation conflicts.

Over the past decade, safety and health issues related to stereo images and videos have been well studied. A significant aim for 3D camera acquisition makers and display manufacturers is to characterize the visual discomfort of stereo images accurately in an attempt to reduce it or eliminate it. Several efforts have been made in the literature to create such models. Early on, Nojiri *et al.*[9,10] found a close correlation between the range of parallax distribution and the degree of visual discomfort. In particular, they found that the reconstructed scene should be positioned behind the screen to deliver a more comfortable viewing experience. Yano *et al.*[11] measured the degree of visual fatigue from the change of accommodation response before and after viewing stereoscopic images. Choi *et al.*[12] used a Principal Component Analysis (PCA) approach to understand factors contributing to visual fatigue in stereoscopic videos including spatial complexity, depth position, temporal complexity, scene movement, depth gradient, crosstalk, and brightness. Kim *et al.*'s model[13] characterized horizontal and vertical disparities. In Ref. 14, Park *et al.* described a predictor which combines features from a neural population coding model with the statistics of horizontal disparity maps. Kim *et al.* in Ref. 15 described a more advanced second-order system model that forms a transfer functions integrating information about the optical nerve, the accommodation and vergence neural pathways, the oculomotor plant, and visual area MT. In Ref. 16, Oh *et al.* constructed different maps and extracted their features to create their model. The considered maps are the degree of out-of-focus map starting from the focal distance, the Panum's fusional area map representing how well the 3D object is fused, the stereoscopic map representing the output responses induced by processes of accommodation and vergence, and the conflict response map which accounts for the disagreement between the response when viewing stereo images on a flat screen vs in a natural environment. In Ref. 17, Park *et al.* proposed a model which accounts for both accommodation and vergence to predict quality, by making use of the physiological optics of binocular vision and foveation.

The models described above are all perception based. Deep convolutional neural networks (CNN) have also recently been applied to the problem. Very recently the authors in Ref. 18 used a CNN which is fed a disparity map to predict visual discomfort. Using Natural Scene Statistics (NSS) as a tool to assess visual discomfort has not been exploited yet. NSS has proved to be a powerful strategy for gaining insight into the HVS by measuring and analyzing the physical regularities of the natural environment, as the HVS has evolved based on the natural environment that we perceive. The power of this approach is that by characterizing the physical regularities in the visual environment, then we can gain insight into how those regularities could be exploited to perform visual tasks. In this paper, we first look at the statistics of multiple pixels, or the bivariate NSS of disparity maps, in attempt to understand how those statistics correlate with visual discomfort. Our hypothesis is that the statistics of natural disparity maps would likely not cause visual discomfort. We will extract NSS features of disparity maps, and combine them with a subset of the features of the binocular model in Ref. 17 to create a new 3D visual discomfort predictor.

The rest of the paper is organized as follows. In Sec. 2 we characterize the bivariate NSS of disparity maps. Then in Sec. 3, we combine features bivariate NSS of disparity maps with perceptually based features to create a new predictor to evaluate visual discomfort. In Sec. 4 we evaluate its performance and conclude the paper in Sec. 5.

2. BIVARIATE NATURAL SCENE STATISTICS OF DISPARITY MAPS

For the case of 2D images, NSS models allowed us to understand how the HVS can efficiently process gigantic amounts of visual data.[19] Recent efforts has been directed towards understanding the joint statistics of multiple pixels.[20,21] It has been established in Ref. 22 that such statistics can be captured in closed-form for luminance 2D images, as well as for the case of chromatic images.[23] It was also shown that such models can be used to derive new methods for predicting the quality of images.[24] Next we describe how the model in Ref. 22 can be applied to 3D images. The IEEE-SA database[25] was used as a basis for developing our model. The database contains 160 scenes, captured with different disparity ranges, resulting in 800 S3D images pairs, each associated with a Mean Opinion Score (MOS). The resolution of all the images is 1920×1080. The content of this database

is diverse, containing a wide variety of objects. The scenes were captured indoors and outdoors. A flow chart of the model is presented in Fig. 1.

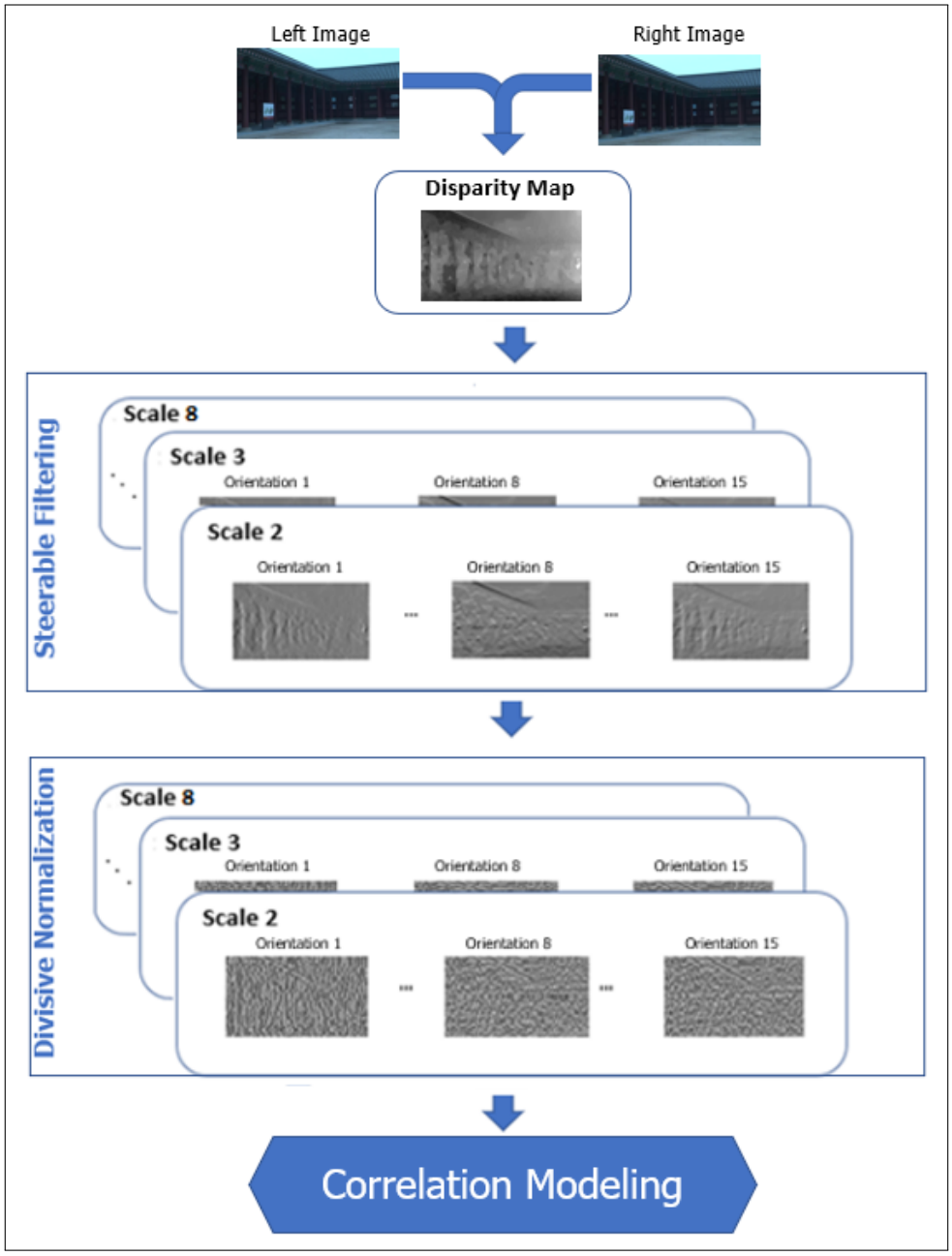


Figure 1: Flow chart of the NSS model.

2.1 Disparity Maps Extraction

The input to our system are left and right stereo pairs. The HVS infers depth information from the observed left and right images. To extract this information, we computed disparity maps. We used a classical technique described in Ref. 26, 27 to obtain a single disparity map for every image pair.

2.2 Steerable Filtering

Simple cells in the primary visual cortex act as bandpass filters. To model this process, we used steerable filters[28] in our simulations, owing to their simple, easily manipulated form, their invariance to content translations, and their good fit as a frequently used model of bandpass simple cells in primary visual cortex. A steerable filter at a given frequency tuning orientation θ_1 is defined by:

$$F_{\theta_1}(\boldsymbol{x}) = cos(\theta_1)F_x(\boldsymbol{x}) + sin(\theta_1)F_y(\boldsymbol{x}), \tag{1}$$

where $\boldsymbol{x} = (x, y)$, and F_x and F_y are the gradient components of the two-dimensional unit-energy bivariate isotropic gaussian function:

$$G(\boldsymbol{x}) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}}, \tag{2}$$

where σ is the scale.

Simple cells in primary visual cortex act as a multi-scale bank of bandpass filters. Varying σ of the bivariate gaussian derivative functions (F_x and F_y) allows us to obtain a multi-scale bandpass image decomposition broadly resembling the responses of populations of simple cells in cortical area V1. We varied $\sigma \in \{2, 3, ..., 7\}$ and over 15 frequency tuning orientations $\theta_1 \in [0, \pi/15, 2\pi/15, ..., \pi]$, resulting in 90 bandpass responses for every disparity map.

2.3 Divisive Normalization

The output of each steerable filter is then subjected to divisive normalization. This step models the nonlinear adaptive gain control of V1 neuronal responses in the visual cortex.[29] This operation normalizes the energy of the local signal, and gaussianizes and decorrelates the image data as shown in Ref. 30, 31. The divisive normalization model used here is:

$$u_j(\boldsymbol{x}) = \frac{w_j(\boldsymbol{x})}{\sqrt{t + \sum_{\boldsymbol{y}} g(j(\boldsymbol{y}), w_j(\boldsymbol{y}))^2}}, \tag{3}$$

where w_j are the steerable filter responses for filters indexed by j, u are the coefficients obtained after divisive normalization, and $t = 10^{-4}$ is a stabilizing saturation constant. The weighted sum in the denominator is computed over a spatial neighborhood of pixels from the same sub-band, where $g(x_i, y_i)$ is a circularly symmetric Gaussian function having unit volume. The variance of $g(x_i, y_i)$ is also increased linearly as a function of the scale σ.

2.4 Correlation Modeling

Next, we define several important quantities. Consider each steerable filtered and divisive normalized image and define in each a window at a fixed position (Window 1) and another sliding window of the same dimensions (Window 2). The two windows were separated by horizontal and vertical separations δ_x and δ_y resulting in a Euclidean distance $d = \sqrt{\delta_x^2 + \delta_y^2}$, separating the two centers. δ_x and δ_y were varied over the integer range from 0 and 19, hence d took values between 0 and $\sqrt{19^2 + 19^2} = \sqrt{722} = 26.87$. The centers of the two windows are separated by a spatial orientation $\theta_2 = \arctan(\frac{\delta_y}{\delta_x})$ (relative to the horizontal axis), as illustrated in Fig. 2. We limited the range of θ_2 to $[0, \pi[$ since the quantities being measured are symmetrically defined and are π periodic. We considered the 8 most frequent values of θ_2 : $[0, 0.785, 1.570, 2.356]$ occurring 25 times in our considered window and $[0.436, 1.107, 2.034, 2.677]$ occurring 12 times there. The model could also extended to include other less frequent values of θ_2, but we found that including those does not add too much value. Next, define the relative angle $\theta_2 - \theta_1$, where θ_1 is the sub-band tuning of the bandpass filter orientation relative to the horizontal axis. The tuning orientation θ_1 is the frequency tuning orientation of the steerable filter. We used a discrete set of 15 sub-band orientations $\{0, \frac{\pi}{15}, \frac{2\pi}{15}, ..., \frac{14\pi}{15}\}$ to build our model. Hence for every fixed value of θ_2, 15 relative angle values $\theta_2 - \theta_1$ are obtained.

For each scale σ, distance d, and spatial angle θ_2, we computed the Pearson correlation function between the two windows as a function of the relative angle $\theta_2 - \theta_1$. Fig. 3 presents sample plots.

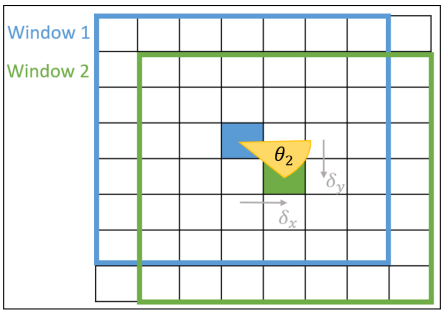
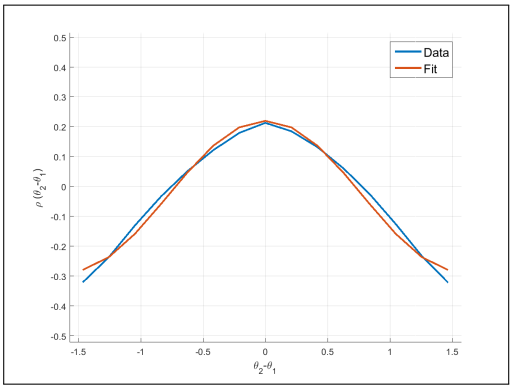
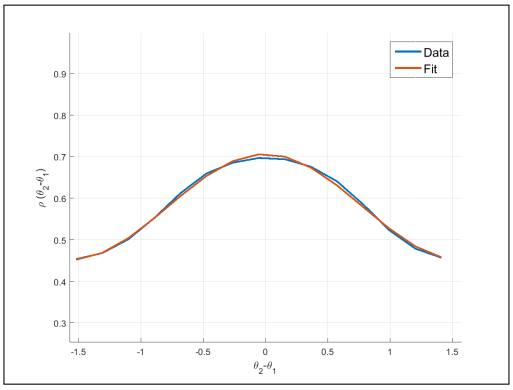


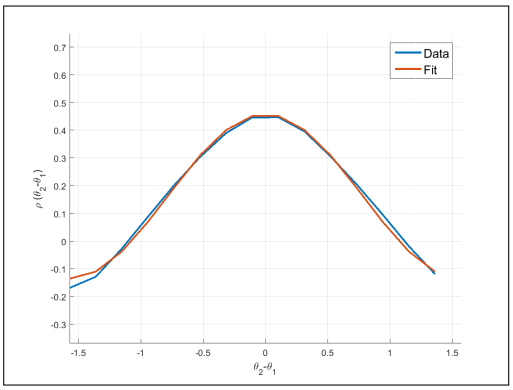
Figure 2: An illustration of an image after the divisive normalization and steerable filtering (of fixed σ and θ_1 values) are applied, with the two sliding windows, and how θ_2 is computed.



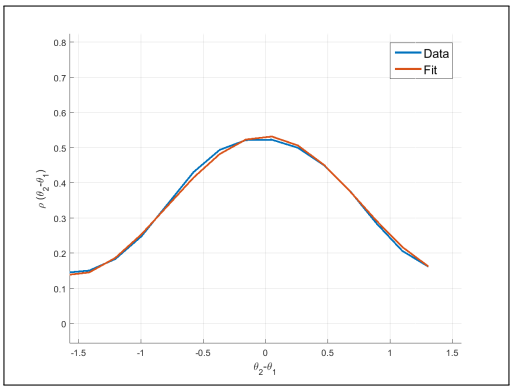
(a) $d = 7$, $\theta_2 = 0$ and $\sigma = 2$.



(b) $d = 4.24$, $\theta_2 = 0.785$ and $\sigma = 3$.



(c) $d = 12$, $\theta_2 = 1.507$ and $\sigma = 5$.



(d) $d = 12.72$, $\theta_2 = 2.352$ and $\sigma = 7$.

Figure 3: Sample correlation plots obtained from the data and their corresponding fits, obtained from the stereo pair ISS1-0-L.png and ISS1-0-R.png from the IEEE-SA database.[25] The sample correlation plots span various d, θ_2 and σ values.

The correlation function model expresses a periodic behavior in the relative angle $\theta_2 - \theta_1$, which is symmetrical around 0, as observed in Fig. 3 so we found that a cosine function yields a good fit:

$$\rho(d, \sigma, \theta_2) = A(d, \sigma, \theta_2) cos(2(\theta_2 - \theta_1)) + c(d, \sigma, \theta_2) \tag{4}$$

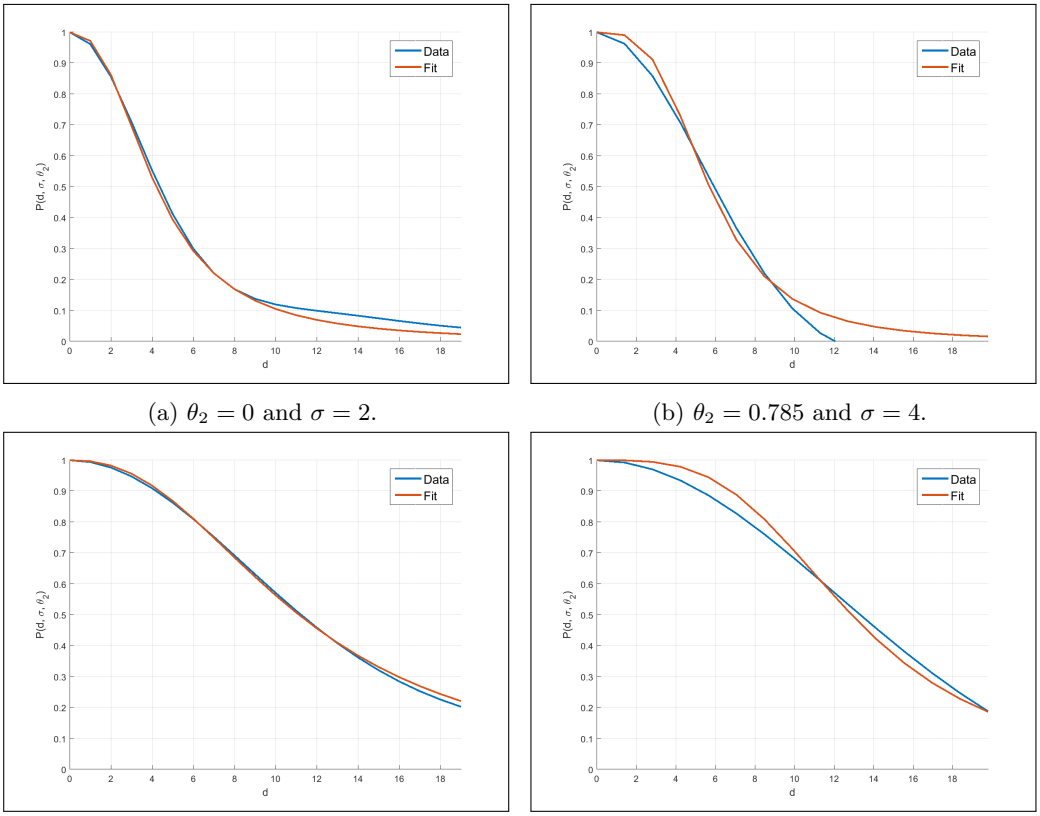
where $A(d, \sigma, \theta_2)$ is the amplitude, $c(d, \sigma, \theta_2)$ is an offset, d is the spatial separation between the target pixels, σ is the steerable filter scale parameter, and θ_2 is the spatial orientation. Looking at Fig. 3, we notice that the empirical data and its corresponding fits align very closely, validating (4).

To complete the model of the correlation function ρ in (4), we also model the amplitude and offset functions A and c. Define the peak of the correlation function as:

$$P = \max(\rho) = A + c. \tag{5}$$

wherein we may rewrite (4) as:

$$\rho(d, \sigma, \theta_2) = A(d, \sigma, \theta_2) cos(2(\theta_2 - \theta_1)) + [P(d, \sigma, \theta_2) - A(d, \sigma, \theta_2)] \tag{6}$$



(a) $\theta_2 = 0$ and $\sigma = 2$.

(b) $\theta_2 = 0.785$ and $\sigma = 4$.

(c) $\theta_2 = 1.507$ and $\sigma = 3$.

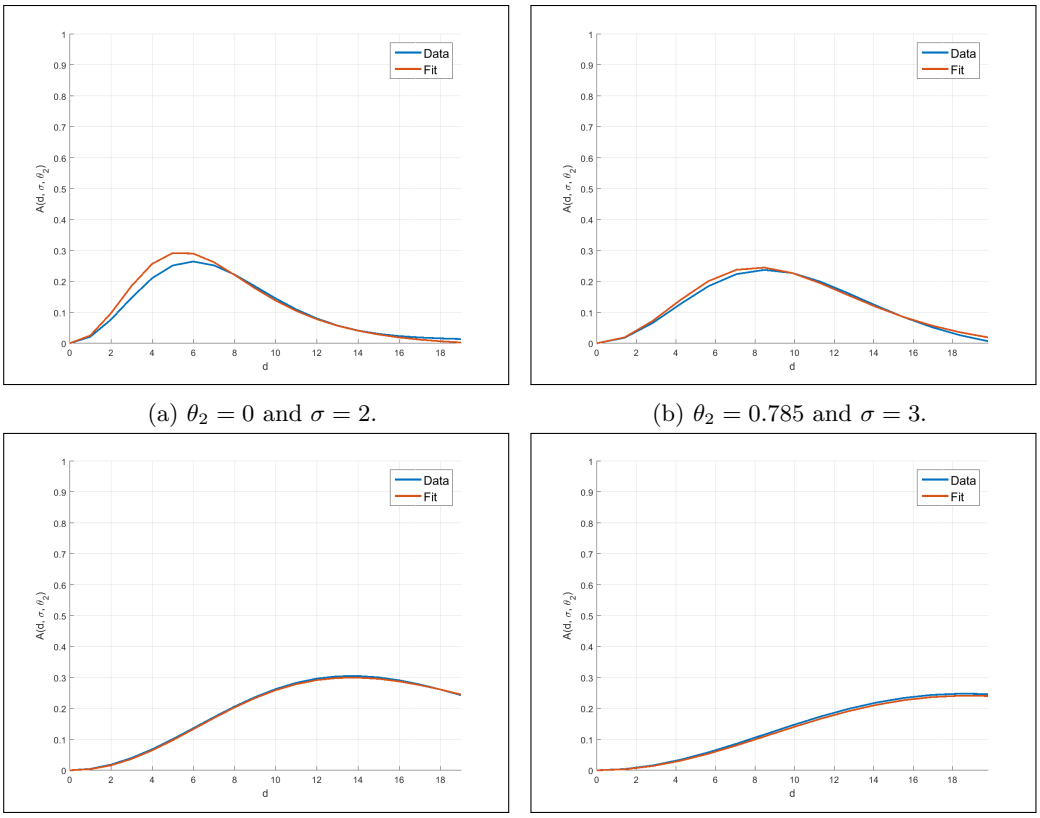
(d) $\theta_2 = 2.352$ and $\sigma = 7$.

Figure 4: Sample empirical P and their fits, obtained from the stereo pair ISS1-0-L.png and ISS1-0-R.png from the IEEE-SA database.[25] The sample P plots span various θ_2 and σ values.

Fig. 4 plots the empirical peak correlation function P against the sample separation d for a few values of σ and θ_2. The measured correlations decrease rapidly from a peak value of 1 as the spatial separation d is increased; this is expected since the correlations between pixels decreases as the spatial separation increases. P seems to take the shape of a $1/f$ process,[32] taking a reciprocal form as a function of the distance $\frac{1}{|d|^\beta}$. We modeled the peak correlation function as having a general version of the form $\frac{1}{|d|^\beta + 1}$, which allows P to take the value 1 when $d = 0$:

$$\hat{P}(d, \sigma, \theta_2) = \frac{1}{(\frac{d}{\alpha_0(\theta_2) * \sigma})^{\beta_0} + 1} \tag{7}$$

where $\{\alpha_0, \beta_0\}$ are parameters that control the shape and fall-off of the peak correlation function, and which depend on the spatial orientation θ_2.



(a) $\theta_2 = 0$ and $\sigma = 2$.

(b) $\theta_2 = 0.785$ and $\sigma = 3$.

(c) $\theta_2 = 1.507$ and $\sigma = 5$.

(d) $\theta_2 = 2.352$ and $\sigma = 7$.

Figure 5: Sample empirical A and their fits, obtained from the stereo pair ISS1-0-L.png and ISS1-0-R.png from the IEEE-SA database.[25] The sample A plots span various θ_2 and σ values.

Fig. 5 plots the amplitude function $A(d, \sigma, \theta_2)$ against d for few scales σ and spatial orientations θ_2. The graph of A rises from the value 0 at $d = 0$, then decreases as the spatial separation d increases. Given the similarity of the graph of A to the difference of two functions of the same form but different scales, and the close relationship between A and P, we model A as the difference of two functions of the form (7):

$$\hat{A}(d, \sigma, \theta_2) = \frac{1}{(\frac{d}{\alpha_1(\theta_2) * \sigma})^{\beta_1(\theta_2)} + 1} - \frac{1}{(\frac{d}{\alpha_2(\theta_2) * \sigma})^{\beta_2(\theta_2)} + 1} \tag{8}$$

where $\{\alpha_1, \beta_1, \alpha_2, \beta_2\}$ are parameters that are functions of θ_2 that control the shape of A.

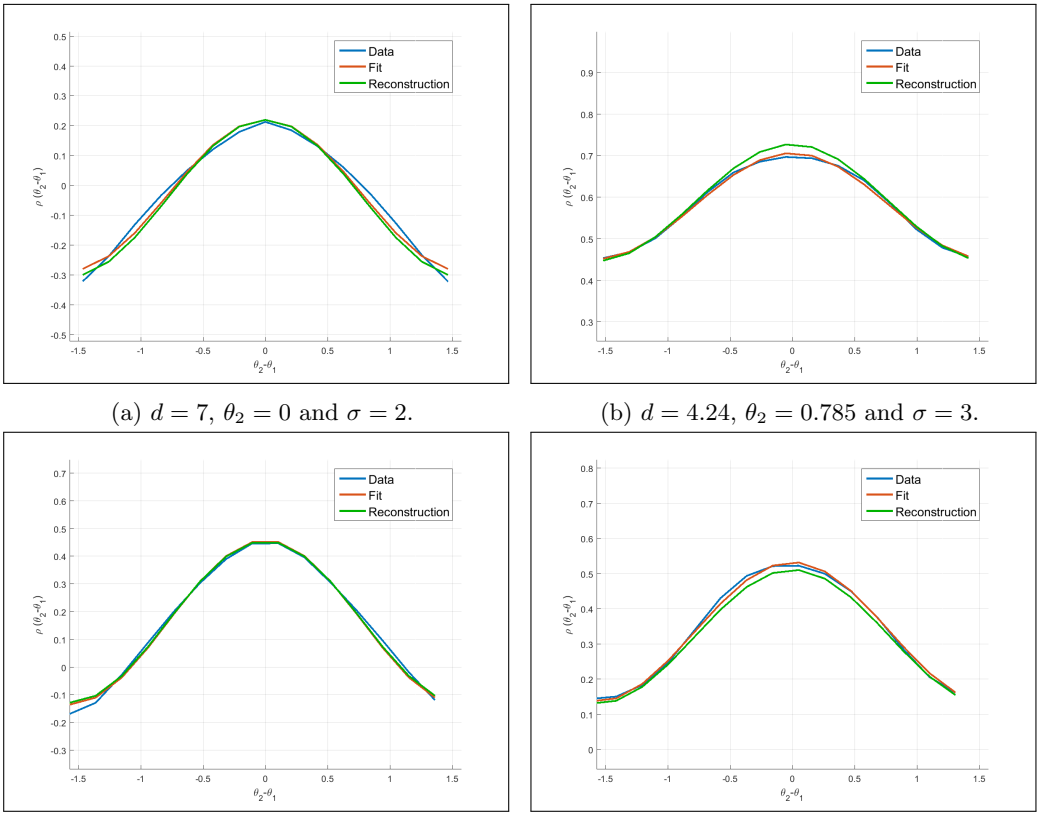
Finally, to complete the correlation model, we found the values of the parameters $\{\alpha_0, \beta_0\}$ that produce the best fit to (7), and the parameters $\{\alpha_1, \beta_1, \alpha_2, \beta_2\}$, that yield the best fit to (8), in the least mean squared error sense, for a fixed spatial orientation θ_2. To do so, we form two optimization systems for P and A that account for scale, to find the optimal values. It was shown in Ref. 22 that the developed correlation model is scale invariant, meaning that if P is plotted as a function of d/σ, the obtained plots at multiple scale values (σs) overlap. This property holds for A as well, so we use it as a core to our optimization systems. We applied unconstrained nonlinear regression using the quasi newton method.[33] The four functions $P(d, \sigma, \theta_2)$, $A(d, \sigma, \theta_2)$, $\hat{P}(d, \sigma, \theta_2)$, and $\hat{A}(d, \sigma, \theta_2)$ form vectors of size $m \times 1$, where m is the number of occurrences of θ_2 inside the span of interest. Denote by D the set of distances for a given spatial orientation θ_2. For the case $\theta_2 = 0$ or $\pi/2$, $D = \{0, 1, 2, 3, ..., 18, 19\}$. For the case $\theta_2 = \pi/4$ or $3\pi/4$, $D = \{0, \sqrt{2}, \sqrt{8}, \sqrt{18}, ..., \sqrt{648}, \sqrt{722}\}$. Our optimization systems are then expressed as:

$$\min_{\alpha_0, b_0} \quad \sum_{d \in D} \sum_{\sigma=2}^{7} (P(d, \sigma, \theta_2) - \hat{P}(d, \sigma, \theta_2))^2 \tag{9}$$

and

$$\min_{\alpha_1, b_1, \alpha_2, \beta_2, b_2} \quad \sum_{d \in D} \sum_{\sigma=2}^{7} (A(d, \sigma, \theta_2) - \hat{A}(d, \sigma, \theta_2))^2 \tag{10}$$

Using $\{\alpha_0, \beta_0\}$ to reconstruct P and $\{\alpha_1, \beta_1, \alpha_2, \beta_2\}$ to reconstruct A allows us to in turn reconstruct the correlation functions ρs with very small errors, at various d, σ and θ_2 values. In Fig. 6 we present some examples of the reconstructed ρ, at various d, θ_2, and σ values. The great overlap between the empirical correlation data, its fit and its reconstruction demonstrates the validity of our model. Similar observations can be made across all the images of the IEEE-SA database,[25] and across different d, θ_2 and σ values. This claim is also validated by computing the Mean Squared Error between the empirical data and its reconstruction.



(a) $d = 7$, $\theta_2 = 0$ and $\sigma = 2$.

(b) $d = 4.24$, $\theta_2 = 0.785$ and $\sigma = 3$.

(c) $d = 12$, $\theta_2 = 1.507$ and $\sigma = 5$.

(d) $d = 12.72$, $\theta_2 = 2.352$ and $\sigma = 7$.

Figure 6: Sample correlation plots obtained from the data and their corresponding fits and reconstructions, obtained from the stereo pair ISS1-0-L.png and ISS1-0-R.png from the IEEE-SA database.[25]

3. 3D VISUAL DISCOMFORT PREDICTOR

Motivated by the observation that stereoscopic displays stimulate accommodation and vergence in an unnatural way resulting in the vergence-accommodation conflict, we decided to combine binocular vergence and accommodation features computed based on the disparity maps along with the NSS features (α_0, β_0, α_1, β_1, α_2, β_2) along different spatial orientations θ_2 values to create a regression module that would map all those features to human ratings in the IEEE-SA[25] database.

3.1 Bivariate Natural Scene Statistics Depth Features

We study the correlation between the bivariate NSS depth features $\{\alpha_0,\ \beta_0,\ \alpha_1,\ \beta_1,\ \alpha_2,\ \beta_2\}$ at the 8 most frequently occurring spatial orientations $\theta_2 \in [0, 0.436, 0.785, 1.107, 1.570, 2.034, 2.356, 2.677]$. As a first test, we trained a Support Vector Regression (SVR) model[34] using a radial basis function and 80-20 split on the IEEE-SA database[25] using 46 features as input, obtaining Pearson's linear correlation coefficient (PLCC) and Spearmans rank ordered correlation coefficient (SROCC) both in the range of 0.71 and 0.63 respectively, when the experiment was repeated over multiple iterations. We found that taking a subset of those features and complementing them with other binocular model features helped improve performance. In particular, we observed that a combination of the bivariate features at $\theta_2 = 0$ and $\theta_2 = 2.677$ correlated the most with the MOS. We denote these features by F_1 - F_{12}

3.2 Binocular Model Features

We complemented our model with four other statistical based binocular model features. The authors in Ref. 17 proposed perceptual features to evaluate visual discomfort. We considered a subset of their proposed features, in particular the ones related to disparity maps which were reported to have a PLCC of 0.83 and an SROCC of 0.76 over multiple iterations.[17] We used the method described in Ref. 26, 27 to extract the disparity maps. The main motivation behind the use of the considered features relates to the close correlation between visual discomfort and the parallax distribution.[9,10] If the reconstructed scene is positioned behind the screen then the viewing experience is comfortable. In that case, the eyes diverge, and the disparity is positive. Otherwise, the eyes converge, as the disparity is negative, leading to visual discomfort. Thus the correlation between the sign of disparity and visual discomfort. The features that we included capture the sign of the disparity. The first feature represents the mean of positive disparities defined by:

$$F_{13} = \begin{cases} \frac{1}{N_{Pos}} \sum_{D(n)>0} D(n), & \text{if } N_{Pos} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

where $D(n)$ is the n^{th} smallest value in the disparity map and N_{Pos} is the number of positive elements in the map.

In a similar fashion, we let the mean of negative disparities be another feature defined by:

$$F_{14} = \begin{cases} \frac{1}{N_{Neg}} \sum_{D(n)\leq 0} D(n), & \text{if } N_{Neg} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

where N_{Neg} is the number of negative elements in the map.

We also include the mean of the lowest 5^{th} and highest 95^{th} percentiles as additional features, defined as:

$$F_{15} = \frac{1}{N_{P_{5^{th}}}} \sum_{D(n)\leq D(P_{5^{th}})} D(n) \tag{13}$$

where $N_{P_{5^{th}}}$ is the number of elements smaller than or equal to the 5^{th} percentiles in $D(n)$. And:

$$F_{16} = \frac{1}{N_{P_{95^{th}}}} \sum_{D(n) \geq D(P_{95^{th}})} D(n) \tag{14}$$

where $N_{P_{5^{th}}}$ is the number of elements greater than or equal to the 95^{th} percentiles in $D(n)$.

Taking the combination of all the 16 features results in a predictor that is summarized in Table 1.

Table 1: Summary of the features used in the predictor.

Feature	Description
F_1	α_0 at $\theta_2 = 0$ rad
F_2	β_0 at $\theta_2 = 0$ rad
F_3	α_1 at $\theta_2 = 0$ rad
F_4	β_1 at $\theta_2 = 0$ rad
F_5	α_2 at $\theta_2 = 0$ rad
F_6	β_2 at $\theta_2 = 0$ rad
F_7	α_0 at $\theta_2 = 2.677$ rad
F_8	β_0 at $\theta_2 = 2.677$ rad
F_9	α_1 at $\theta_2 = 2.677$ rad
F_{10}	β_1 at $\theta_2 = 2.677$ rad
F_{11}	α_2 at $\theta_2 = 2.677$ rad
F_{12}	β_2 at $\theta_2 = 2.677$ rad
F_{13}	mean of positive disparities
F_{14}	mean of negative disparities
F_{15}	mean of the smallest 5% of the values in the disparity map
F_{16}	mean of the largest 95% of the values in the disparity map

4. RESULTS

Using a regression module, we constructed a mapping from the feature space, (Table 1) to the MOS, resulting in a measure of 3D visual discomfort. We used a support vector regressor (SVR),[34] in particular the LIBSVM package[35] to implement the SVR with a radial basis function (RBF) kernel and to predict the MOS scores. We split the images randomly and used 80% of it for training and the rest for testing, then we normalized our features, and fed them into the SVR module to predict the MOS score. We repeated the process 50 times. We obtained a median Pearsons linear correlation coefficient (PLCC) of about 0.89 and a Spearmans rank ordered correlation coefficient (SROCC) of about 0.83 against MOS. Table 2 compares the performances of other various reported algorithms. The performance of our model was only approached by DeepVDP,[18] which uses a complex convolutional neural network.

5. CONCLUSION

In this paper, we studied the bivariate NSS of disparity maps, and modeled them in closed form. We showed that using 6 features per spatial orientation allows us to capture those statistics with very small error. We demonstrated a close relationship between those statistics and 3D visual discomfort. We combined those features

Table 2: Mean PLCC and SROCC and their standard deviations over the IEEE-SA database,[25] with 80-20% splits, over 50 iterations.

Model	PLCC	SROCC
Nojiri *et al.*[9]	0.6854 ± 0.0788	0.6108 ± 0.0732
Yano *et al.*[11]	0.3988 ± 0.0748	0.3363 ± 0.0798
Choi *et al.*[12]	0.6509 ± 0.0783	0.5851 ± 0.0798
Park *et al.*[14]	0.8310 ± 0.0526	0.7534 ± 0.0498
Kim *et al.*[15]	0.7018 ± 0.0771	0.6151 ± 0.0700
Oh *et al.*[16]	0.8590 ± 0.0452	0.7887 ± 0.0405
Park *et al.*[17]	0.8524 ± 0.0482	0.7785 ± 0.0451
Oh *et al.*[18]	0.8849 ± 0.0283	0.8164 ± 0.0254
Proposed Model	**0.8884** ± 0.0197	**0.8264** ± 0.0292

along with other simple statistics of disparity maps related to positive and negative disparities to create a powerful 3D visual discomfort predictor that outperforms state of the predictors, which are perceptually based and/or use deep convolutional networks. In the future, studies of the spatio-temporal dynamics of 3D NSS may prove quite useful, given that rapidly changing depths/disparities may contribute strongly to experienced visual discomfort.

REFERENCES

[1] Jones, R. K. and Lee, D. N., "Why two eyes are better than one: the two views of binocular vision.," *Journal of Experimental Psychology: Human Perception and Performance* **7**(1), 30 (1981).

[2] Fielder, A. R. and Moseley, M. J., "Does stereopsis matter in humans?," *Eye* **10**(2), 233 (1996).

[3] Kooi, F. L. and Lucassen, M., "Visual comfort of binocular and 3D displays," in [*Human Vision and Electronic Imaging VI*], **4299**, 586–593, International Society for Optics and Photonics (2001).

[4] Pastoor, S., "Human factors of 3d imaging: results of recent research at heinrich-hertz-institut berlin," in [*Proc. 2nd International Display Workshop IDW'95*], **3**, 69–72 (1995).

[5] Emoto, M., Niida, T., and Okano, F., "Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television," *Journal of display technology* **1**(2), 328 (2005).

[6] Hoffman, D. M., Girshick, A. R., Akeley, K., and Banks, M. S., "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of vision* **8**(3), 33–33 (2008).

[7] Shibata, T., Kim, J., Hoffman, D. M., and Banks, M. S., "Visual discomfort with stereo displays: effects of viewing distance and direction of vergence-accommodation conflict," in [*Stereoscopic Displays and Applications XXII*], **7863**, International Society for Optics and Photonics (2011).

[8] Lambooij, M. T., IJsselsteijn, W. A., and Heynderickx, I., "Visual discomfort in stereoscopic displays: a review," in [*Stereoscopic Displays and Virtual Reality Systems XIV*], **6490**, 64900I, International Society for Optics and Photonics (2007).

[9] Nojiri, Y., Yamanoue, H., Hanazato, A., and Okano, F., "Measurement of parallax distribution and its application to the analysis of visual comfort for stereoscopic HDTV," in [*Stereoscopic Displays and Virtual Reality Systems X*], **5006**, 195–206, International Society for Optics and Photonics (2003).

[10] Ide, S., Yamanoue, H., Okui, M., Okano, F., Bitou, M., and Terashima, N., "Parallax distribution for ease of viewing in stereoscopic hdtv," in [*Stereoscopic Displays and Virtual Reality Systems IX*], **4660**, 38–46, International Society for Optics and Photonics (2002).

[11] Yano, S., Ide, S., Mitsuhashi, T., and Thwaites, H., "A study of visual fatigue and visual comfort for 3D HDTV/HDTV images," *Displays* **23**(4), 191–201 (2002).

[12] Choi, J., Kim, D., Choi, S., and Sohn, K., "Visual fatigue modeling and analysis for stereoscopic video," *Optical Engineering* **51**(1), 017206 (2012).

[13] Kim, D. and Sohn, K., "Visual fatigue prediction for stereoscopic image," *IEEE Transactions on Circuits and Systems for Video Technology* **21**(2), 231–236 (2011).

[14] Park, J., Oh, H., Lee, S., and Bovik, A. C., "3D visual discomfort predictor: Analysis of disparity and neural activity statistics," *IEEE Transactions on Image Processing* **24**(3), 1101–1114 (2015).

[15] Kim, T., Lee, S., and Bovik, A. C., "Transfer function model of physiological mechanisms underlying temporal visual discomfort experienced when viewing stereoscopic 3D images," *IEEE Transactions on Image Processing* **24**(11), 4335–4347 (2015).

[16] Oh, H., Lee, S., and Bovik, A. C., "Stereoscopic 3D visual discomfort prediction: A dynamic accommodation and vergence interaction model," *IEEE Transactions on Image Processing* **25**(2), 615–629 (2016).

[17] Park, J., Lee, S., and Bovik, A. C., "3D visual discomfort prediction: Vergence, foveation, and the physiological optics of accommodation.," *Journal of Selected Topics in Signal Processing* **8**(3), 415–427 (2014).

[18] Oh, H., Ahn, S., Lee, S., and Bovik, A. C., "Deep visual discomfort predictor for stereoscopic 3d images," *IEEE Transactions on Image Processing* (2018).

[19] Field, D., "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America* **4**(12), 2379–2394 (1987).

[20] Sinno, Z. and Bovik, A. C., "Generalizing a closed-form correlation model of oriented bandpass natural images," in [*IEEE Global Conference on Signal and Information Processing (GlobalSIP)*], 373–377, IEEE (2015).

[21] Sinno, Z. and Bovik, A. C., "Relating spatial and spectral models of oriented bandpass natural images," in [*IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*], 89–92, IEEE (2016).

[22] Sinno, Z., Caramanis, C., and Bovik, A. C., "Towards a closed form second-order natural scene statistics model," *IEEE Transactions on Image Processing* **27**(7), 3194–3209 (2018).

[23] Sinno, Z. and Bovik, A. C., "On the natural statistics of chromatic images," *IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)* (Apr 2018).

[24] Sinno, Z., Caramanis, C., and Bovik, A. C., "Second order natural scene statistics model of blind image quality assessment," *IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (Apr. 2018).

[25] Li, S., "Stereoscopic (3-D imaging) Database." Standard for the Quality Assessment of Three Dimensional (3D) Displays, 3D Contents and 3D Devices based on Human Factors http://grouper.ieee.org/groups/3dhf/. (Accessed: 07 July 2018).

[26] Sun, D., Roth, S., and Black, M. J., "Secrets of optical flow estimation and their principles," in [*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 2432–2439, IEEE (2010).

[27] Sun, D., Roth, S., and Black, M. J., "A quantitative analysis of current practices in optical flow estimation and the principles behind them," *International Journal of Computer Vision* **106**(2), 115–137 (2014).

[28] Freeman, W. T. and Adelson, E. H., "The design and use of steerable filters," *IEEE Transactions on Patternn Analysis and Machine Intelligence* (9), 891–906 (1991).

[29] Carandini, M., Heeger, D. J., and Movshon, A. J., "Linearity and normalization in simple cells of the macaque primary visual cortex," *Journal of Neuroscience* **17**(21), 8621–8644 (1997).

[30] Ruderman, D. L. and Bialek, W., "Statistics of natural images: Scaling in the woods," in [*Advances in neural information processing systems*], 551–558 (1994).

[31] Simoncelli, E. P., "Modeling the joint statistics of images in the wavelet domain," in [*Wavelet Applications in Signal and Image Processing VII*], **3813**, 188–196, International Society for Optics and Photonics (1999).

[32] Keshner, M. S., "1/f noise," *Proceedings of the IEEE* **70**(3), 212–218 (1982).

[33] Gill, P. E. and Murray, W., "Quasi-newton methods for unconstrained optimization," *IMA Journal of Applied Mathematics* **9**(1), 91–108 (1972).

[34] Schölkopf, B., S. A. J. W. R. C. and Bartlett, P. L., "New support vector algorithms," *Neural Computation* **12**(5), 1207–1245 (2000).

[35] Chang, C.-C. and Lin, C.-J., "Libsvm: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)* **2**(3), 27 (2011).