# Generalizing a Closed-Form Correlation Model of Oriented Bandpass Natural Images

Zeina Sinno and Alan C. Bovik

Laboratory for Image and Video Engineering
Department of Electrical and Computer Engineering
The University of Texas at Austin, Austin, TX, USA

*Abstract*—**Building natural scene statistic models is a potentially transformative development for a wide variety of visual applications, ranging from the design of faithful image and video quality models to the development of perceptually optimized image enhancing techniques. Most predominant statistical models of natural images only characterize the univariate distributions of divisively normalized bandpass image responses. Previous efforts towards modeling bandpass natural responses have not focused on finding closed-form quantative models of bivariate natural statistics. Towards filling this gap, Su *et al.* [1] recently modeled spatially adjacent bandpass image responses over multiple scales; however, they did not consider the effects of spatial distance between the bandpass samples. Here we build on Su *et al.*'s model and extend their closed-form correlation model to non-adjacent distant bandpass image responses over multiple spatial orientations and scales.**

*Index Terms*—**Natural Scene Statistics; Bivariate Correlation Models; Bandpass Natural Images.**

## I. INTRODUCTION

Models of the function of Natural Scene Statistics (NSS) of perceived images have become essential building blocks in many reliable image and video processing algorithms. Such algorithms span a wide range of applications from image/video quality assessment models [2], [3], [4] to state of the art image enhancement techniques including image denoising [5], image defocus [6], and image super-resolution [7]. Tremendous effort has been made to fathom the relationships between NSS and visual perception and how these relationships might be exploited to produce perceptually relevant image processing models.

Spatial processing in primary visual cortex is often modeled in image analysis algorithms by a Gabor filter bank [8], [9], which decomposes and decorrelates the received signal over multiple scales and orientations, followed by nonlinear adaptive gain control (ACG) [10]. The resulting visual signal after decorrelation is well-modeled by certain probability distributions. Ruderman [11], operating on naturalistic images, showed that a local mean subtraction operation followed by division by a measure of local image energy, such as local variance (divisive normalization) reproduces this decorrelation effect. The resulting signal following these two operations is strongly Gaussianized [11]. This concept has been deeply exploited in the establishment of first order statistical models of bandpass natural images.

Among previous attempts to characterize the bivariate behavior of natural images, little attention has been applied to finding closed form models. In an attempt to fill this gap, Su *et al.* [1] proposed a closed form correlation model of horizontally adjacent oriented bandpass natural image responses across multiple pixels and multiple sub-band bandpass orientations and scales. The authors demonstrated that this model is useful for a wide variety of image processing applications including stereoscopic image quality prediction [12]. This predictor outperforms state-of-the-art full- and no-reference 3D IQA algorithms on both symmetrically and asymmetrically distorted stereoscopic image pairs. They also applied the new correlation model to design a depth estimator based on single luminance images [13]. This suggests that generalizing the model could be very propitious. Here, we extend Su *et al.*'s work [1] and present a closed form correlation model of oriented bandpass natural images covering a spatial distance of up to ten pixels, encompassing all discrete spatial angles, over four scales.

## II. RELEVANT OBSERVATIONS AND MODELS

Simoncelli *et al.* [14] observed that the coefficients of orthonormal wavelet decompositions of natural images are decorrelated but not independent. Liu *et al.* [15] noted the presence of inter and intra-scale dependencies between wavelet coefficients. Sendur *et al.* [16] used a circularly symmetric bivariate distribution to model the dependencies between image wavelet coefficients and their parents (at coarser scale locations).

Inspired by Geman *et al.*'s work, [17] deploys a Markov random field model to implement image restoration at low signal-to-noise ratios, Portilla *et al.* [18] targeted the problem of natural image texture modeling. They used a set of parametric constraints on pairs of complex wavelet coefficients at adjacent spatial locations, orientations and scales in conjunction with a non-Gaussian Markov Random Field. The major issue with this method is in the choice of statistical constraints, which were obtained by applying a form of reverse-engineering of the early Human Visual System (HVS). The selection process of the parameters was achieved by observing failures to synthesize particular types of texture rather than seeking the optimality of the solution. Given the under-determined nature of the problem, additional constraints would be useful. Po *et al.* [19] modeled natural images using a hidden Markov tree, a Gaussian mixture

model, and two dimensional contourlets to capture interlocation, interscale, and interdirection dependencies. Mumford *et al.* [20] proposed an infinitely divisible model of generic image statistics. This model assumes that the environment may be subdivided into objects cast against an ergodic field, containing regions with very little information (e.g. blue sky). This model was used to fit parts of images but falls short of capturing the 2D dependencies between (bandpass) image luminances.

None of the above-mentioned models provides a closed form bivariate correlation model of natural images. Such a model, of "perceptually transformed" bandpass and normalized images, could supply powerful priors on numerous visual processing problems. Su *et. al* [1] attempted to bridge this gap by modeling the responses of adjacent oriented bandpass natural image pixels. Here we broaden their model to encompass non-adjacent distances, while also exploring simplifications of the model.

## III. BASIC IMAGE MODEL

First we describe the preprocessing steps applied to the images before modeling them. The images used are high quality pristine images from the LIVE IQA database [3] and from the Berkeley image segmentation database [21].

### A. Color Space Transformation

Each natural image was first transformed to the CIELAB color space. This color space relates to human color perception [1]. Only the L component was used in our current development of the correlation model.

### B. Steerable Filters

The steerable filters [22] were applied as a simple model of the bandpass characteristic of simple cells in primary visual cortex. A steerable filter at a given frequency tuning orientation $\theta_1$ is defined by:

$$F(\theta_1) = cos(\theta_1)F_x + sin(\theta_1)F_y \qquad (1)$$

where $F_x$ and $F_y$ are the gradients of the two dimensional bivariate gaussian function with respect to the horizontal and vertical axes respectively. $F_x$ and $F_y$ are normalized to get unit energy. Image decompositions using steerable filters yield decorrelated representations over scale and orientation which resemble spatial cortical responses. Similarly to [23], altering the variance $\sigma$ of the bivariate gaussian function (differentiated to obtain $F_x$ and $F_y$) allows to account for the multi-scale decomposition computed by simple cells in area V1. For this purpose, we set the variance $\sigma$ to 1, 2 and 4 as in [23] in addition to 1.5, 2.5, 3 and 3.5. The (half-peak) octave bandwidth of the steerable filter (1) is about 2.6 octaves. We computed responses on all images over 15 frequency tuning orientations $\theta_1$ ranging over $[0, \pi/15, 2\pi/15, ..., \pi]$.

### C. Divisive Normalization

Divisive normalization was applied on all the steerable filter responses. This step models the non linear adaptive gain control of V1 neuronal responses in the visual cortex [24]. This process further decorrelates and gaussianizes the image data [11], [14]. The divisive normalization model used here is:

$$u(x_i, y_i) = \frac{w(x_i, y_i)}{\sqrt{s + \boldsymbol{w_g^T w_g}}} = \frac{w(x_i, y_i)}{\sqrt{s + \sum_j g(x_j, y_j)w(x_j, y_j)^2}} \qquad (2)$$

where $(x_i, y_i)$ are spatial coordinates, $w$ are the wavelet coefficients, $u$ are the coefficients obtained after divisive normalization, and $s = 10^{-4}$ is a semi saturation constant. The weighted sum is computed over a spatial neighborhood of pixels in the same sub-band index by $j$ (assuming a window of dimensions $3 \times 3$ hence $j = 9$). The Gaussian weighting function, $g(x_i, y_i)$, is circularly symmetric and unit volume.

### D. Modeling the Bivariate Joint Distribution

The bivariate joint distribution model aims to characterize pairs of pixels having a relative distance between 1 and 10, and covering all possible discrete angles. This extends Su *et al.*'s work [1], where the bivariate joint distribution was computed from adjacent pixels only.

Inspired by the fact that the univariate generalized Gaussian distribution successfully models univariate natural scenes, we deploy a multivariate generalized Gaussian distribution (MGGD) to model the bivariate joint histogram of two target pixels at two different spatial locations in the bandpass normalized image. This choice is also justified by the fact that the MGGD is an accurate model of multi-dimensional image histograms [25]. The probability density function of the MGGD is:

$$p(\boldsymbol{x}; \boldsymbol{M}, \alpha, \beta) = \frac{1}{|\boldsymbol{M}|^{\frac{1}{2}}} g_{\alpha,\beta}(\boldsymbol{x^T M^{-1} x}) \qquad (3)$$

Where $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{M}$ is an $N \times N$ scatter matrix, $\alpha$ and $\beta$ are scale and shape parameters respectively, and $g_{\alpha,\beta}(.)$ is the density generator:

$$g_{\alpha,\beta}(y) = \frac{\beta\Gamma(\frac{N}{2})}{(2^{\frac{1}{\beta}}\pi\alpha)^{\frac{N}{2}}\Gamma(\frac{N}{2\beta})} e^{-\frac{1}{2}(\frac{y}{\alpha})^\beta} \qquad (4)$$

where $\Gamma$ is the digamma function and $y \in \mathbb{R}^+$. Note that if $\beta = 0.5$ then (3) is a multivariate Laplacian distribution, and when $\beta = 1$, (3) is a multivariate Gaussian distribution.

The bivariate empirical histograms of the sub-band coefficients of natural images are thus modeled using a bivariate generalized Gaussian distribution (BGGD), by setting $N = 2$. This also presumes that the images have not been distorted, which may change their statistics. Similar to [1] the parameters of the BGDD were estimated using an efficient maximum likelihood estimation method [26].

## IV. THE GENERALIZED MODEL

The BGGD model captures the way the shape, height, and dispersion of the bivariate distributions vary with the spatial locations and the tuning orientation of the sub-band responses. Su *et al.* [1] found that when the spatial orientation $\theta_2$ between the bandpass samples matches the frequency tuning orientation $\theta_1$, the joint distribution becomes peaky and extremely elliptical, implying that the bandpass responses are highly correlated. Conversely, when the spatial relationship and the sub-band tuning orientation becomes orthogonal, the joint distribution approaches a circular Gaussian suggesting nearly independent sub-band responses.When the steerable filters (1) are used, the best-fitting BGGD distribution never approaches circularity, which may be because the steerable filters have a fairly wide bandwidth. In their study of the correlation behavior of spatially adjacent subband responses as a function of the relative orientation (difference between spatial and tuning orientations), the authors of [1] observed a periodic behavior from which they deduced an exponentiated cosine of the correlation coefficients:

$$\rho = A cos(\theta_2 - \theta_1)^{2\gamma} + c \qquad (5)$$

as a function of relative orientation, where $\theta_1$ and $\theta_2$ are the sub-band and spatial tuning orientations respectively, $A > 0$ is the amplitude, $\gamma$ is a shape exponent and $c$ is an offset. The relative orientation is represented by $\theta_2 - \theta_1$. The authors obtained a closed form model by fitting the correlation coefficients between horizontally adjacent bandpass responses as the function (4) of the sub-band tuning orientations using all the high quality images in the LIVE IQA database [3]. Of course, the periodicity of the model is unsurprising; however the good fits obtained using a simple parametric functional model is both unexpected and useful.

Here we generalize Su *et al.*'s model to non-adjacent distances between the bandpass samples. For consistency with the followed convention for $\theta_1$ and $\theta_2$, the relative angle $\theta_2 - \theta_1$ increases in a counterclockwise direction as in [1].

The spatial orientation $\theta_2 = \arctan(\frac{\delta_y}{\delta_x})$ where $\delta_x$ and $\delta_y$ are the relative row and column differences between coordinates of the responses after divisive normalization. The tuning orientation $\theta_1$ is defined as the normal to a sinusoidal wave front.

We used the midpoint circle algorithm [27] to generate digital circles of integer radius varying between 1 to 10 in the image space. Points on each circle defined the spatial pairs $(\delta_x, \delta_y)$.

The empirical correlations are defined as Pearson correlation coefficients. Each $(\delta_x, \delta_y)$ defines a spatial orientation $\theta_2$. The 15 sub-band orientations are drawn from the set $\{0, \frac{\pi}{15}, \frac{2\pi}{15}, ...,, \frac{14\pi}{15}\}$ rad at each scale of the steerable filters' decomposition. This process was performed on all the images in the LIVE IQA database [3] (29 images) and on some images from the Berkeley image segmentation database (71 images) [21] for each $\theta_1$ and $\theta_2$. Model fits are applied on
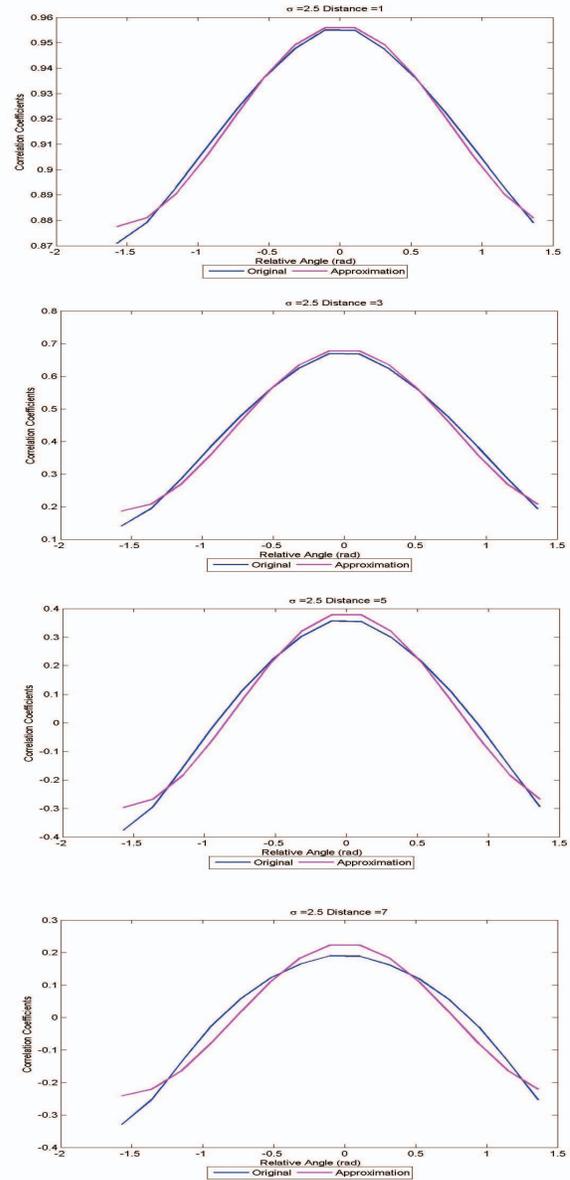


Fig. 1. Average correlation coefficients at scale 1, $\theta_2 = \pi/2$ and $\sigma = 1$ for separations of 1,3, 5 and 7 and their corresponding fits.

the average correlation values of the 100 images from the two databases.

By examining the correlation coefficients plots for small distances $\delta = \sqrt{\delta_x^2 + \delta_y^2}$, we observe that the maximal correlation is obtained when $\theta_2 - \theta_1$ is equal to 0. Generally, the maximal correlation drops as the relative distance between the origin and target increases. We also observe a similar trend across the seven different scales; the correlation functions exhibit a similar shape with a difference in the ranges attributed to passing different frequencies.

Fig. 1 illustrates the trend in the shapes of the correlation function as the separation $\sqrt{\delta_x^2 + \delta_y^2}$ increases. Fig. 2 illustrates the trend in the parameters versus the spatial separation
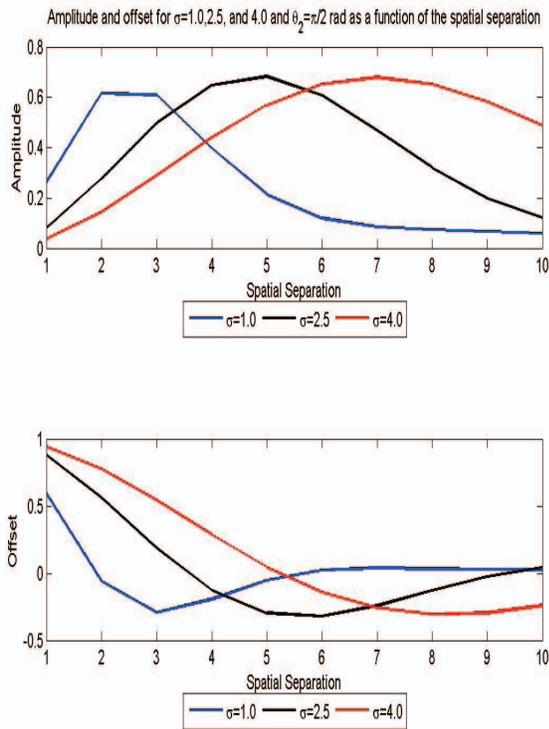
Fig. 2. The trend in the amplitude and offset in the correlation function fit (5) versus spatial separation for $\theta_2 = \pi/2$ and $\sigma = 1, 2.5$ and 4.

for $\theta_2 = \pi/2$. $A$ and $c$ are presented only as $\gamma = 0.5$.

## V. VALIDATION OF THE GENERALIZED MODEL

We computed the Mean Squared Error (MSE) and Pearson's $\chi^2$ test on the model against the empirical correlation coefficients over all of the pristine images of the LIVE IQA database [3] and the VCL@FER Image Quality Assessment Database [28] to validate our model. The $\chi^2$ test is computed as:

$$\chi^2 = \sum_{i=1}^{N} \sum_{j=1}^{S} \frac{(\rho_{i_j} - \rho'_j)^2}{\rho'_j} \qquad (6)$$

where $\{\rho'_j\} = \boldsymbol{\rho'} \in \mathbb{R}^D$ is the model, $\{\rho_{i_j}\} = \boldsymbol{\rho_i} \in \mathbb{R}^D$ are the correlation coefficients of the $i^{th}$ pristine image, $S = 15$ is the total number of sub-band angles, and $N$ is the number of pristine images. $N = 100$ for images from the LIVE IQA database combined with the images from the Berkeley image segmentation database, and $N = 23$ for the VCL@FER database. The MSE and $\chi^2$ test of the samples previously presented are shown in Table 1.

The MSE is low for all separations considered. The $\chi^2$ test has small values except for some points at high distances, which may be attributed to the small magnitude of the correlation coefficients magnifying any minor deviation from the model.

|  | MSE training | $\chi^2$ training | MSE testing | $\chi^2$ testing |
|---|---|---|---|---|
| Distance 1 | 0.004178 | 0.459442 | 0.001053 | 0.11582 |
| Distance 3 | 0.166615 | 48.77902 | 0.038803 | 11.43254 |
| Distance 5 | 0.364689 | -27.6250 | 0.070773 | -0.46758 |
| Distance 7 | 0.323636 | 290.3075 | 0.057003 | 47.40307 |

TABLE I

MSE AND $\chi^2$ OF THE GENERALIZED MODEL AGAINST EMPIRICAL CORRELATIONS AT SCALE 1 AND $\theta_2 = \pi/2$ FOR SEPARATIONS OF 1, 3, 5 AND 7.

## VI. CONCLUSION

We described a generalized closed-form correlation model of oriented bandpass natural images accounting for separation between the bandpass samples up to 10 pixels. In the future, we plan to make use of this correlation model as a building block in applications including non-reference image quality prediction, texture modeling and image interpolation.

## REFERENCES

[1] C. Su, L. Cormack, and A. Bovik, "Closed-form correlation model of oriented bandpass natural images," *Signal Processing Lett., IEEE*, vol. 22, no. 1, pp. 21–25, Jan 2015.

[2] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[3] H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3440–3451, Nov 2006.

[4] Z. Wang and A. Bovik, "A universal image quality index," *Signal Processing Lett., IEEE*, vol. 9, no. 3, pp. 81–84, Mar 2002.

[5] J. Perry and W. S. Geisler, "Natural scene statistics for image denoising," Center for Perceptual Systems, The University of Texas at Austin, Tech. Rep., 2013.

[6] J. Burge and W. S. Geisler, "Optimal defocus estimation in individual natural images," *Proc.of the National Acad. of Sci.*, vol. 108, no. 40, pp. 16 849–16 854, 2011. [Online]. Available: http://www.pnas.org/content/108/40/16849.abstract

[7] A. D. D'Antona, J. S. Perry, and W. S. Geisler, "Humans make efficient use of natural image statistics when performing spatial interpolation," *J. Vision*, vol. 13, no. 14, p. 11, 2013.

[8] M. Clark and A. C. Bovik, "Experiments in segmenting texton patterns using localized spatial filters," *Pattern Recognition*, vol. 22, no. 6, pp. 707–717, 1989.

[9] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Machine Intel.*, vol. 12, no. 1, pp. 55–73, 1990.

[10] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex," *J. of Neurophys.*, vol. 58, no. 6, pp. 1233–1258, 1987.

[11] D. L. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," *Phys. Rev. Lett.*, vol. 73, no. 6, p. 814, 1994.

[12] C. Su, L. Cormack, and A. Bovik, "Oriented correlation models of distorted natural images with application to natural stereopair quality evaluation," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1685–1699, May 2015.

[13] C. Su, L. Cormack, and A. C. Bovik, "Bayesian depth estimation from monocular natural images," *IEEE Trans. Pattern Anal. Machine Intell.*, Jan 2015.

[14] E. P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain," *SPIE Int'l Symp. on Opt. Sci., Eng., and Instrum.*, pp. 188–195, 1999.

[15] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. Image Process.*, vol. 10, no. 11, pp. 1647–1658, 2001.

[16] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2744–2756, 2002.

[17] S. Geman and C. Graffigne, "Markov random field image models and their applications to computer vision," *Proc. Int'l Congr. Math.*, vol. 1, p. 2, 1986.

[18] J. Portilla and E. P. Simoncelli, "Texture modeling and synthesis using joint statistics of complex wavelet coefficients," *IEEE workshop on stat. comp. th. vision*, vol. 12, 1999.

[19] D.-Y. Po and M. N. Do, "Directional multiscale modeling of images using the contourlet transform," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1610–1620, 2006.

[20] D. Mumford and B. Gidas, "Stochastic models for generic images," *Quarterly App. Math.*, vol. 59, no. 1, pp. 85–112, 2001.

[21] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Int. Conf. Comput. Vision*, vol. 2.   IEEE, 2001, pp. 416–423.

[22] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Machine Intell.*, no. 9, pp. 891–906, 1991.

[23] A. Chakrabarti, K. Hirakawa, and T. Zickler, "Color constancy with spatio-spectral statistics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 8, pp. 1509–1519, 2012.

[24] M. Carandini, D. J. Heeger, and A. J. Movshon, "Linearity and normalization in simple cells of the macaque primary visual cortex," *J. Neurosci.*, vol. 17, no. 21, pp. 8621–8644, 1997.

[25] F. Pascal, L. Bombrun, J.-Y. Tourneret, and Y. Berthoumieu, "Parameter estimation for multivariate generalized gaussian distributions," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5960–5971, 2013.

[26] C.-C. Su, L. K. Cormack, and A. C. Bovik, "Bivariate statistical modeling of color and range in natural scenes," in *Proc. SPIE, Human Vis. Electron. Imag. XIX*, vol. 9014, Feb. 2014.

[27] J. R. Van Aken, "An efficient ellipse-drawing algorithm," *IEEE Comput. Graph. App.*, vol. 4, no. 9, pp. 24–35, 1984.

[28] A. Zaric, N. Tatalovic, N. Brajkovic, H. Hlevnjak, M. Loncaric, E. Dumic, and S. Grgic, "Vcl@ fer image quality assessment database," *AUTOMATIKA*, vol. 53, no. 4, pp. 344–354, 2012.