

Prediction of the Perceived Quality of Streak Distortions in Offset-Printing with a Psychophysically Motivated Multi-channel Model

Konrad Gadzicki, Christoph Zetsche

Universität Bremen

Kognitive Neuroinformatik

Enrique-Schmidt-Str. 5, 28359 Bremen

eMail: konny@informatik.uni-bremen.de,

zetsche@informatik.uni-bremen.de

URL:http://www.informatik.uni-bremen.de/cog_neuroinf/

Abstract The evaluation of printing machines poses the problem of how distortions like streaks caused by the machine can be detected and assessed automatically. Although luminance variations in prints can be measured quite precisely, the measured functions bear little relevance for the lightness of streaks and other distortions of prints as perceived by human observers. First, the measurements sometimes indicate changes of luminance in regions which are perceived as homogeneous by humans. Second, the measured strength of a distortion correlates often weakly with its perceived strength, which is influenced by a variety of factors, like the shape of a streak's luminance profile and the distribution of luminance variations in its spatial surround. We have used a model of human perception, based on fundamental neurophysiological and psychophysical properties of the visual system, in order to predict the strength of streak distortions as perceived by human observers from a measured luminance signal. For the evaluation of the model, tests with naive and expert observers have been conducted. They showed that the model has a good correlation (> 0.8) to the assessments of human observers and is therefore suited for use in an automatic evaluation system.

1 Introduction

Modern offset printing machines are able to produce prints at high speed and quality. Nevertheless certain distortions, like streaks, are generated almost inevitably due to vibrations of the machine or an inadequate configuration. Streak distortions run orthogonally to the printing direction and result from slight shifts of the ink. All variations in the printing pattern can be captured metrologically very precisely, resulting in a signal in which the slightest changes of the density of a print are recorded. Unfortunately, the perception of humans deviates in several respects from the recorded signal. First, areas containing only small random fluctuations are perceived as homogeneous by humans. It is clear that such fluctuations, though measurable, have little practical relevance for the judgment of the quality of the printing process. Second, the perceived lightness of a streak differs substantially from its recorded intensity profile, both in dependence of the shape of the profile (as opposed to its mere height or contrast), and in dependence of the spatial surround of the streak (for example neighboring streaks or paper borders).

The evaluation of a printing machine thus requires a human expert in order to assess the prints with regard to the severity of streak distortions. Though procedures exist for metrological evaluation, the current methods either locate only severe distortions which are not arguable, while ignoring disputable distortions below their threshold which can still be visible to humans. Or, with the threshold lowered, they report relevant deviations in areas perceived as homogeneous by humans. In addition, they cannot take the surround of a streak into account for the evaluation of its strength. An automatic system for evaluation according to human perception essentially requires the incorporation of a model of the human visual system.

Several such models based on the properties of the human visual system have been used to assess image quality in the past [1][2][3][4]. These models commonly incorporate major properties like luminance invariance, sensitivity to frequency and orientations, and masking effects. Early models used a point-wise non-linearity or explicit gain modifications by a masking signal to account for masking effects while later vision models used divisive inhibition pooling acting over both spatial positions and frequency-selective channels [5]. The model suggested here takes these developments into account.

2 System and Model

The model is intended as part of a future system for the automatic evaluation of prints produced by SID¹. It is designed to work with scanned images of printings, but can also be adapted to other input sources (e.g. densitometers). Hardware and software for scanning and preprocessing of the scans in the current study was provided by SID. The scans were generated with 5000 dpi which provides sufficient resolution for discrimination of closely neighboring streaks.

2.1 Model

On the top level, the model processes two inputs in parallel. As shown in Figure 1, signal+background and background alone are processed in the same manner by a system

¹Sächsisches Institut für die Druckindustrie

explained in detail below. Streaks and other distortions are seen as the signal, while the background contributes to masking effects on this signal, e.g. by high-contrast luminance edges at the print borders. The local vector norm between the multi-channel outputs of the two pathways determines the final model output and thus the perceived strength of local distortion.

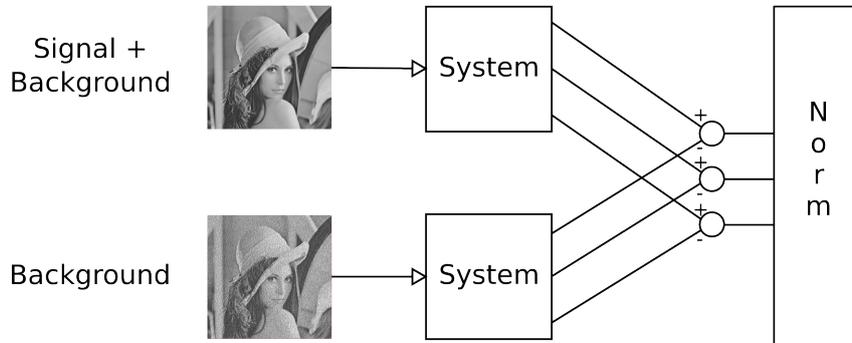


Figure 1: Model overview. The system (shown in detail in Figure 2) generates a multi-channel representation of both signal and signal+background. The distance between the two representations is assumed to be the perceived strength of the signal, and is computed by the difference norm shown at the right.

The system used for both pathways in Figure 1 is shown in detail in Figure 2. Its first stage is a luminance adaptivity stage, where local contrast is computed by a nonlinear multi-scale decomposition. In the next stage frequency- and orientation-selective linear band-pass filters are applied, leading to a further decomposition of each scale into its orientations. Finally, the filter outputs are normalized by local gain control mechanisms which pool over space, scales and orientations. These stages are now described in detail.

2.1.1 Luminance Invariance

The first step in the model is to transform the absolute luminance intensities. According to Weber's Law, the crucial variable for discrimination of luminance variations is contrast and not the absolute difference of luminance. The model uses the Ratio of Gaussian (ROG) operator [1] in order to calculate the contrast values. As the name suggests, the ROG is divisive operation of two low-pass inputs with different cut-off frequencies resulting in a non-linear band-pass output. Figure 3 illustrates the response of the operator.

For a computationally efficient implementation a pyramid scheme, similar to that in the Laplacian pyramid [6], is used to build the nonlinear representation of contrast, decomposed into several spatial scales.

2.1.2 Frequency- and Orientation Selective Filters

The majority of cells in early visual cortex are so-called Simple and Complex Cells which are responsive to patterns with a specific orientation and size. Orientation selectivity was the major finding of Hubel and Wiesel [7][8][9], and spatial-frequency selectivity was measured later in cats [10] and primates [11].

Mathematically the receptive field of such cells can be best described by Gabor functions [12], which offer the optimal resolution and selectivity in order to represent the response of

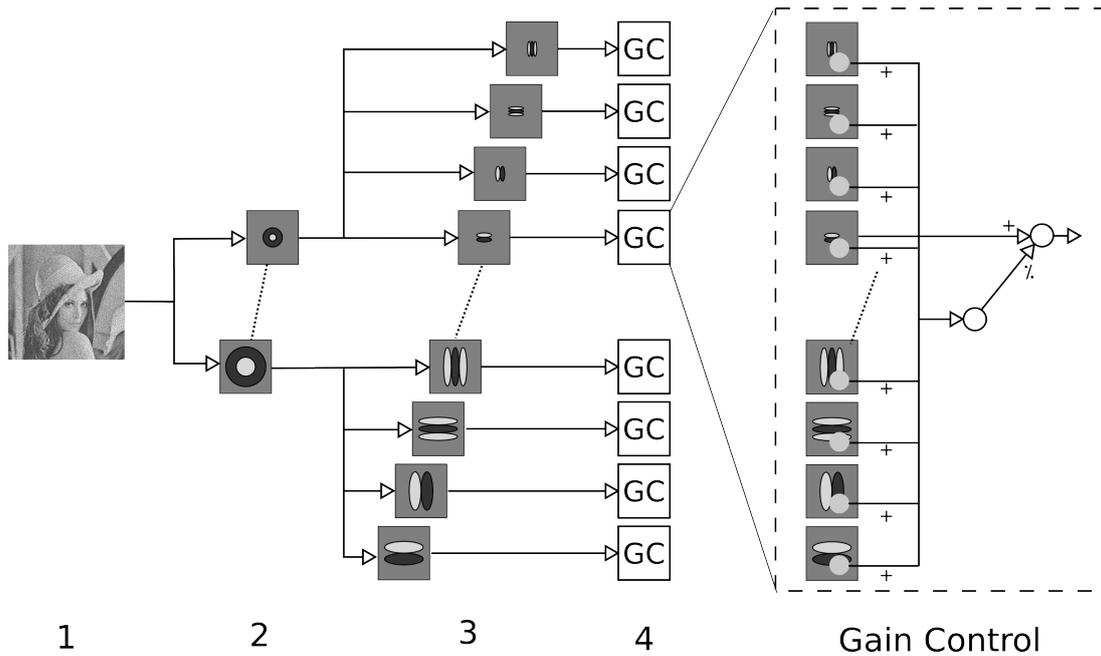


Figure 2: Overview over the system (as used for both pathways in Figure 1). From the input (1), the contrast is computed by non-linear ROG filters (2) and passed to a set of frequency- and orientation-selective linear filters (3). The outputs are then passed through gain control mechanisms (4). One of these is shown in detail on the right hand side. Hence each channel is normalized by spatial pooling over the other channels.

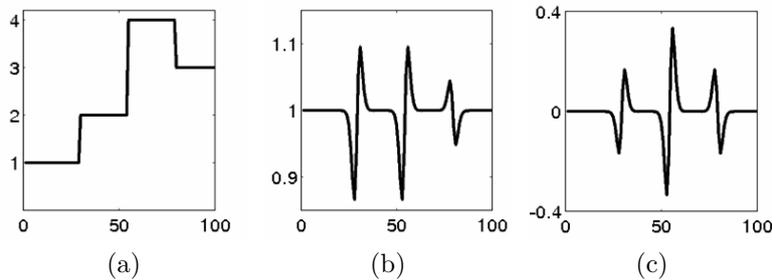


Figure 3: Response of a Ratio of Gaussian operator to luminance step edges. (a) input. (b) the ROG output represents luminance contrast. (c) linear DOG response shown for comparison.

V1 cells [13][14]. Gabor functions can be separated into even- and odd-symmetric parts, which respond best to saddles or edges respectively.

Though Gabor filters fit the psychophysical data well, they can be problematic in practice due to the DC component (zero frequency) of the even-symmetric part which does not drop to zero for filters with typical bandwidth. A Log-Gabor function [15], essentially a Gabor on a logarithmic scale, can be used to avoid this problem.

2.1.3 Masking

In visual perception the term masking usually refers to the reduced detectability of a stimulus in the presence of another stimulus. Such contrast masking has been modeled by a non-linear transducer function in early models [16]. Psychophysical experiments investigating the contrast response function of neurons from cats and monkeys [17] introduced the Naka-Rushton function, a hyperbolic ratio, as the best fitting description for neural response to contrast. Further investigations of the response properties of neurons revealed a pooling effect [18][19][20]. Masking effects can thus be explained by a suppressive signal derived by pooling over neurons. This mechanism is commonly referred to as Cortical Gain Control. It is modeled by a divisive pooling over space and channels and was included into recent visual system models, e.g. by Watson [5].

3 Methods

For the purpose of data acquisition, several experiments have been run. The first experiment employed naive observers with an on-screen display of distorted patterns in order to establish a baseline for the model in a noise-free environment. The only source of noise in this setup was effectively perceptual noise of the observers.

The second set of experiments were evaluations of real printings conducted by experts from the printing industry. In this setup several sources contributed to the noise of the system, namely the printing process itself, the scanner system, and perceptual noise of the observers.

3.1 Scale

Participants rated distortions on a qualitative grade scale similar to the EBU scale used in television picture quality assessment [21]. The grades used here ranged from 0 to 6 in 0.5 steps with 6 meaning a severe distortion and 1 a barely visible one. 0 was reserved for undetected distortions, thus it was not assigned by the participants directly but rather during the evaluation in case that participants did not see particular distortions in trials. The scale used here is reversed in comparison to the EBU scale and is actually the one used historically by the printing industry.

3.2 Experiments with Naive Observers

The experiments with naive observers used patterns presented on-screen only, in order to establish full control of the presentation and eliminate any additional sources of noise which are inevitably introduced in the full process of printing and scanning of patterns. The patterns were presented on analog screens with analog input which allowed for a luminance resolution equivalent to roughly 10 bits.

The set of patterns consisted of 30 images with a total of 75 streak distortions. For each participant three sessions were conducted. The whole set was presented during each session. The participants marked the position of a distortion by clicking with the mouse. The level of distortion (grade) and the classification (edge or saddle) were entered with the keyboard.

The distortions for the whole set were generated in a random fashion. This randomized set was used for all participants and sessions. The generation parameters included the number, position and type (Gauss or sharp edge) of distortions, width of the distortion, width of the edge, and contrast.

The participants were students with no prior involvement with the printing industry except for being exposed to printed products like books, magazines etc. in their daily life. They were not trained to perform this specific evaluation of printings task that was required in this experiment.

In total 21 participants participated in the experiment. Six of them were excluded from the evaluation because they either did not finish all sessions or had a significantly higher deviation of responses from the average.

3.3 Experiments with Experts from the Printing Industry

This experiment involved evaluation of real printings and was done by experts from various companies (Heidelberger Druckmaschinen, Koenig & Bauer, Manroland) from the printing industry.

In total, 52 printings were specially produced for the purpose of this experiment. Among them, 36 contained artificially generated distortions while the remaining 16 printings contained realistic distortions produced by an imprecisely configured printing machine. In the case of the artificial patterns, contrast, position, width of the distortion, and width of edges were varied.

The evaluation process was conducted under standardized conditions according to [22]. Each participant completed three sessions. The positions for the realistic-distortion patterns were fixed by the experts before the actual assessment. The positions of artificial distortions were known beforehand, but in the case of additional distortions as a result of the printing process, the positions were mapped manually by experts as well. The level of distortion at a specific position was recorded by an assistant. Details on the experiment can be found in [23].

In contrast to former experiments with on-screen presentation, several sources of noise influenced the results of this experimental setup. The printing process itself introduced considerable noise so that in the case of artificial patterns, the signal was noisy, even though the parameters used for generation were known. With regard to the realistic-distortion patterns, the actual signal of the pattern was not known at all.

3.4 Estimation of Model Parameters

For evaluation, model parameters were fitted to the data. The first approach was to use a simplex search [24] which turned out to be error-prone. This approach was not able to cope with the non-linear nature of this optimization problem, and produced unstable results. Finally a global search using particle swarm optimization [25][26][27] was used to fit the model parameters.

3.5 Evaluation

A local maximum search was used for comparison between model output and assessments. For the on-screen experiments this was due to the fact that the non-expert participants

were often sloppy with regard to clicking at the exact location of the distortion, so that the recorded positions were often several pixel off. In the case of the expert experiments with printings the positions were given in millimeters and were very precise, but the scans of the printings mapped roughly five pixels to a millimeter. Due to this the position in millimeters only indicated an area in the model output.

In order to assess participants' performance, the standard deviation of the participants' responses was calculated. The deviations of the responses indicate how stable the observers' assessments were.

The performance of the model was measured by its correlation with the participants' average assessment for each distortion.

4 Results

4.1 Stability of Assessments

Table 1 shows the standard deviations of the responses of the naive and expert observers. The deviations were calculated intra-individually, e.g. average for each participant, and inter-individually, thus average for each distortion. The intra-individually averaged responses provide information on the consistency of an observer with regard to his or her own responses, i.e. the extent to what the participant can reproduce his or her own responses over subsequent sessions. On the other hand the inter-individually average responses show how the participants agree with each others assessments.

Table 1: Deviations of observers' assessments

	Naive	Expert
Intra-individually averaged	0.382	0.309
Inter-individually	0.707	0.454

One can see that both naive and expert observers were able to reproduce their own responses in a quite reliable way. Though the experts were slightly more consistent, the gap to the naive observers was not large and only differed by 0.073.

With regard to the deviation of responses over all participants for each distortion, the experts were much more consistent as a group. The experts were actually trained to assess distortions on an absolute scale due to their daily work which requires this particular skill. The naive observers on the other hand – even though consistent in their own opinion – have only received a brief explanation of the scale with examples for particular absolute grades. But with no further feedback on their own assessment, it is not surprising that the results of this group varied to a much larger degree.

4.2 Correlation between Model and Assessments

The correlation is shown for inter-individually averaged data where the average assessment for each distortion was mapped to the model output. Figure 4a shows the results for naive observers while figure 4b represents the expert observers.

Interestingly the naive observers achieved a higher correlation than the expert observers. With regard to the deviations of assessments (Table 1) this result is somewhat surprising.

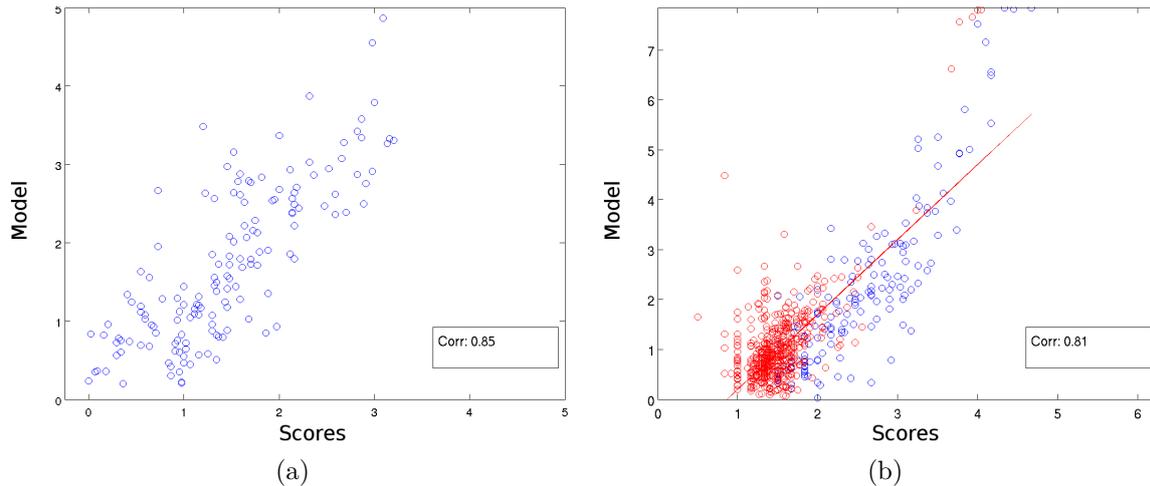


Figure 4: Correlation between model and inter-individually averaged data for (a) naive observers and (b) experts.

Though the processing pipeline for the experiment with printings introduced additional noise in comparison to the on-screen experiments, it is not clear whether this circumstance already accounts for the better results with naive observers.

A further difference between the two experiments is the respective set of test patterns. While in the on-screen experiments the strength of the streaks was approximately uniformly distributed over the grade scale, the expert experiment used a high percentage of realistic prints, which contain many barely visible streaks. Ratings performed close to the threshold of perception can be a problem, as shown below.

4.3 Patterns at Threshold of Perception

Printings are generally noisy with regard to the luminance signal. Therefore many weak streaks are difficult to detect even for trained observers. This is illustrated by the fact that naive observers who have performed the assessment task with realistic printings found approximately three streaks per print whereas the experts found up to 25. However, almost all additional streaks found by the expert group were rated with 1 (barely visible).

4.3.1 Illustration of the Problem

Figure 5a illustrates the distortions of the grades which can occur at the boundaries of the grade scale. First, in the lower part of the scale, near the threshold of perception, a range of different signal levels, from weakly visible to almost invisible signals, is mapped to grade 1. (Since the locations of the streaks are marked, they will typically receive a grade 1, even if critical testing without markings would presumably reveal that some participants are not able to see them.) Ideally there should be a proportionality between signal level and grade also in this regime, but since the grade 1 is used for even the least visible signal, the mapping runs into a plateau.

A similar problem can occur at the upper boundary of the grade scale as well, exaggerated by the fact that participants tend to avoid to use the worst grade, irrespective of how strong the streaks are. Hence the mapping of signal to grade scale can have a less steep

slope and may reach a plateau also in this range

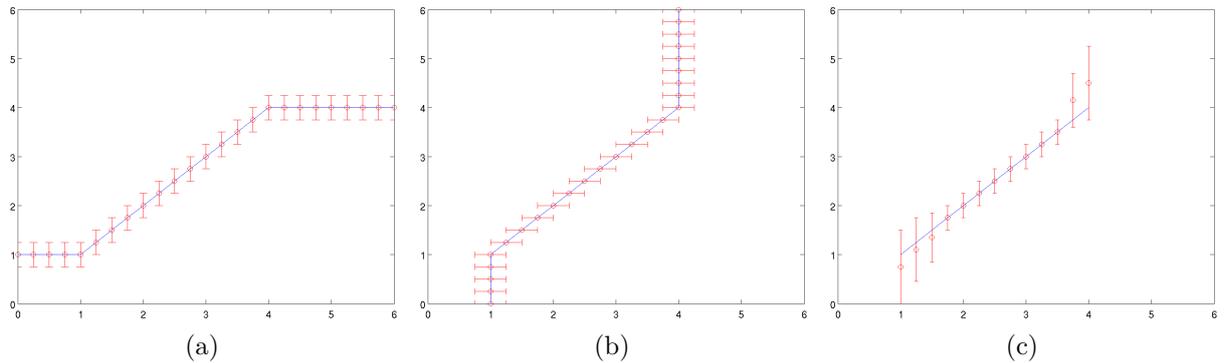


Figure 5: Visualization of distortions of the grade scale at the lower and upper boundary of the scale.

(a) hypothetical effective signal mapped to the grade scale (b) grades mapped to the ideal model. (c) apparent deviations of the model as caused by the grade scale distortion effects.

Figure 5b illustrates the problem from the point of view of an ideal model. For this the axes from figure 5a have to be switched. From the view of the model the minimum grade is assigned to range of values of the model output. At the upper end of the model outputs the same effect is visible, i.e. grades at the upper end will be mapped to a wider range of model outputs.

Figure 5c shows the expected effects with regard to the resulting apparent deviations of the ideal model from the participants’ responses. In the interior range of the grade scale the model output corresponds well to the responses of the participants. The deviations here reflect the “true” deviations of the model, while at the boundaries the apparent deviation of the model is artificially increased.

4.3.2 Correlation between Assessments and Model with Varying Percentage of Near-threshold Distortions

The correlations between model and participants’ assessments for the whole set, including the afore-mentioned distortions near the threshold of perception, have already been presented in figure 4b. We now analyze the influence of those near-threshold responses on the overall performance.

Subsets of the data with a varying percentage of near-threshold responses have been investigated with regard to their effect on the correlation of the model predictions. The criterion for assigning a particular streak to the near-threshold group was the percentage of minimum grades given to that streak. As the minimum grade was effectively given to streaks which were barely visible or not seen at all by some observers, this criterion offers a flexible way to determine the subset of near-threshold streaks. The hardest version of the criterion is to mark a streak as near-threshold if at least one minimum grade has been assigned to it by one of the participants. In subsequent plots assessments of distortions are plotted in red if they contain a minimum grade response.

Figure 6a shows the subset consisting of inter-individually averaged assessments, which received at least one minimum grade response. As one can see, the apparent variance

of the model output is quite high, which is in correspondence to the aforementioned hypothesis that in this regime different signal levels are mapped to the minimum grade. The omission of this subset, i.e. of streaks which have received at least one minimum grade from the evaluation, leads to an increase in the correlation for the averaged data from 0.81 to 0.88, as shown in Figure 6b. Note that although the overall dynamic range of the data is reduced, the correlation increases. This corroborates our hypothesis that the excluded subset is not well behaved.

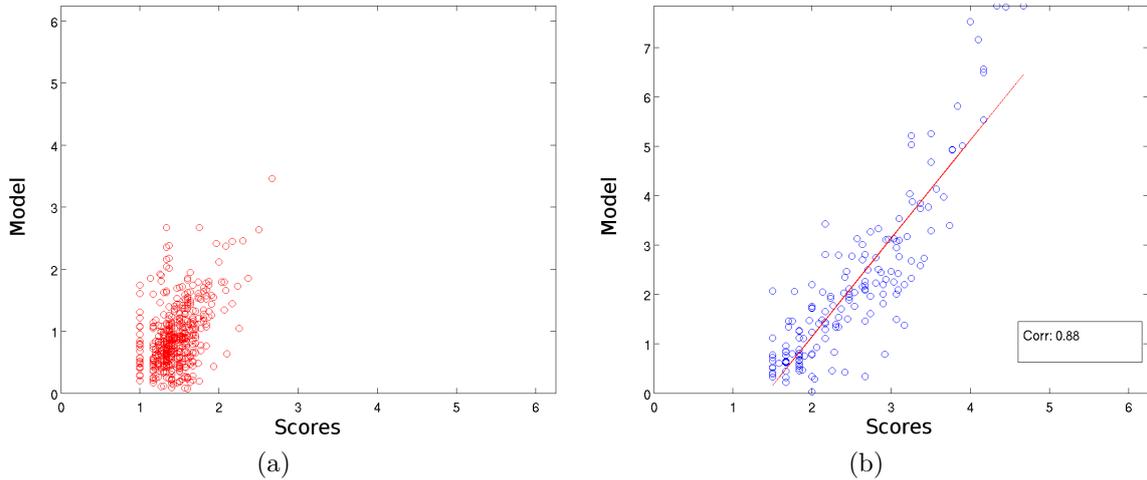


Figure 6: Correlation between model and inter-individually averaged assessments for ratings not containing the minimum grade (b) and for ratings containing at least one minimum grade response (a).

5 Conclusion

We have presented a model for the automatic prediction of the perceived strength of streak distortions in offset printing. The model is based on recent neurophysiological and psychophysical results. It can describe shape effects due to the shape of the luminance function of a streak and masking effects as caused by the configuration in its spatial surround.

In experiments with naive and expert observers we have collected assessments on a large number of streak patterns. We have then shown that the model shows a good correlation in predicting the perceived strength of these distortions.

The prediction of distortions close to the perceptual threshold proved to be problematic, resulting in a big deviation of model responses for barely visible distortions. The removal of distortions which were rated with the minimum grade increased the correlation of the model predictions, although the dynamic range of the data has been reduced by this removal.

This model of the human visual system seems thus to be suitable for the automatic evaluation of printings.

Acknowledgement

This work was supported by NCT Bremen, Sächsisches Institut für die Druckindustrie Leipzig and Forschungsgesellschaft Druckmaschinen e.V. Frankfurt am Main.

References

- [1] Christoph Zetsche and G Hauske. Multiple channel model for the prediction of subjective image quality. In *Human Vision, Visual Processing and Digital Display*, volume 1077 of *Proc. SPIE*, pages 209–216, 1989.
- [2] Christoph Zetsche and G Hauske. Principal features of human vision in the context of image quality models. In *IEEE 3rd Int. Conf. Image Proc. and its Appl*, Proc. of IEEE, pages 102–106, 1989.
- [3] Scott Daly. The visible differences predictor: an algorithm for the assessment of image fidelity. In Andrew B. Watson, editor, *Digital images and human vision*, pages 179–206. MIT Press, Cambridge, MA, USA, 1993.
- [4] Jeffrey Lubin. The use of psychophysical data and models in the analysis of display system performance. In Andrew B Watson, editor, *Digital images and human vision*, pages 163–178. MIT Press, Cambridge, MA, USA, 1993.
- [5] A.B. Watson and J.A. Solomon. Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America.*, 14(9):2379–91, September 1997.
- [6] P. Burt and E. Adelson. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, 31(4):532–540, April 1983.
- [7] D.H. Hubel and TN Wiesel. Receptive fields of single neurones in the cat’s striate cortex. *The Journal of Physiology*, 148(3):574–591, 1959.
- [8] D.H. Hubel and TN Wiesel. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of Physiology*, 160(1):106–154, 1962.
- [9] D.H. Hubel and TN Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1):215–243, 1968.
- [10] F. W. Campbell, G. F. Cooper, and Christina Enroth-Cugell. The spatial selectivity of the visual cells of the cat. *The Journal of Physiology*, 203(1):223–235, 1969.
- [11] R.L. De Valois, K.K. De Valois, J. Ready, and H. Blanckensee. Spatial frequency tuning of macaque striate cortex cells. *Assoc. Res., Vision Ophthalmol*, 15, 1975.
- [12] D Gabor. Theory of communication. Part 1: The analysis of information. *Engineers-Part III: Radio and Communication*, 1946.
- [13] S Marcelja. Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America*, 70(11):1297–300, November 1980.

- [14] J G Daugman. Spatial visual channels in the Fourier plane. *Vision Research*, 24(9):891–910, January 1984.
- [15] David J Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America.*, 4(12):2379–94, December 1987.
- [16] G E Legge and J M Foley. Contrast masking in human vision. *Journal of the Optical Society of America*, 70(12):1458–71, December 1980.
- [17] Duane G. Albrecht and David B. Hamilton. Striate Cortex of Monkey and Cat: Function Contrast Response. *Journal of Neurophysiology*, 48(1), 1982.
- [18] GC DeAngelis, JG Robson, I Ohzawa, and RD Freeman. Organization of suppression in receptive fields of neurons in cat visual cortex. *Journal of Neurophysiology*, 68:144–163, 1992.
- [19] WS Geisler and DG Albrecht. Cortical neurons: isolation of contrast gain control. *Vision Research*, 32(8):1409–1410, 1992.
- [20] DJ Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–198, 1992.
- [21] ITU. Recommendation ITU-R BT.500-11, Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union, Geneva*, pages 1–48, 2002.
- [22] Norm ISO 3664:2009-04. Technical report, International Organization for Standardization, April 2009.
- [23] Versuchsdurchführung bei der Bewertung von Druckbogen. Technical report, FOGRA, 2010.
- [24] J C Lagarias, J A Reeds, M H Wright, and P E Wright. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9:112–147, 1998.
- [25] James Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948. IEEE, 1995.
- [26] Kennedy J, RC Eberhart, and YH Shi. *Swarm Intelligence*. Academic Press, 2001.
- [27] Said M. Mikki and Ahmed a. Kishk. *Particle Swarm Optimizaton: A Physics-Based Approach*, volume 3. January 2008.
- [28] R L De Valois, E W Yund, and N Hepler. The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, 22(5):531–544, 1982.
- [29] AB Watson. Detection and recognition of simple spatial forms. In O. J. Braddick and A. C. Sleight, editors, *Physical and Biological Processing of Images*, 1983.