Universität des Saarlandes

Fachrichtung 6.1 – Mathematik

Preprint Nr. 260

Electrostatic Halftoning

Christian Schmaltz, Pascal Gwosdek,
Andrés Bruhn and Joachim Weickert

Saarbrücken 2010

Fachrichtung 6.1 – Mathematik
Universität des Saarlandes

Preprint No. 260
submitted: February 24, 2010

Electrostatic Halftoning

Christian Schmaltz
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science
Saarland University, Campus E1.1
66041 Saarbrücken
Germany
schmaltz@mia.uni-saarland.de

Pascal Gwosdek
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science
Saarland University, Campus E1.1
66041 Saarbrücken
Germany
gwosdek@mia.uni-saarland.de

Andrés Bruhn
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science
Saarland University, Campus E1.1
66041 Saarbrücken
Germany
bruhn@mia.uni-saarland.de

Joachim Weickert
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science
Saarland University, Campus E1.1
66041 Saarbrücken
Germany
weickert@mia.uni-saarland.de

Edited by
FR 6.1 – Mathematik
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken
Germany

Fax: + 49 681 302 4443
e-Mail: preprint@math.uni-sb.de
WWW: http://www.math.uni-sb.de/

Electrostatic Halftoning

Christian Schmaltz Pascal Gwosdek Andrés Bruhn
Joachim Weickert

February 24, 2010

Abstract

We introduce a new global approach for image dithering, stippling, screening, and sampling. It is inspired by the physical principles of electrostatics. Repelling forces between equally charged particles create a homogeneous distribution in flat areas, while attracting forces from the image brightness values ensure a high approximation quality. Our model is transparent and uses only two intuitive parameters: One steers the granularity of our halftoning approach, and the other its regularity. We evaluate two versions of our algorithm: A discrete version for dithering that ties points to grid positions, as well as a continuous one which does not have this restriction, and can thus be used for stippling or sampling density functions. Our methods create very few visual artefacts, reveal favourable blue-noise behaviour in the frequency domain, and have a lower approximation error under Gaussian convolution than state-of-the-art methods.

1 Introduction

Image halftoning describes a class of techniques to virtually increase the colour depth of printing or display devices. Though early examples go back to times before letterpress printing, the development of algorithms for digital image halftoning has recently enjoyed additional focus. Many devices are still unable to produce a high colour depth due to technical reasons. Printers and fax machines, for example, can often create only purely black and white images without intermediate grey values. Offset printers overlay few monochrome layers to create the impression of arbitrary colours. Even for digital image processing and storage, so-called *dithering* methods are frequently used to find a best representation of full-colour images in indexed colour palettes. However, today's accurate algorithms are no

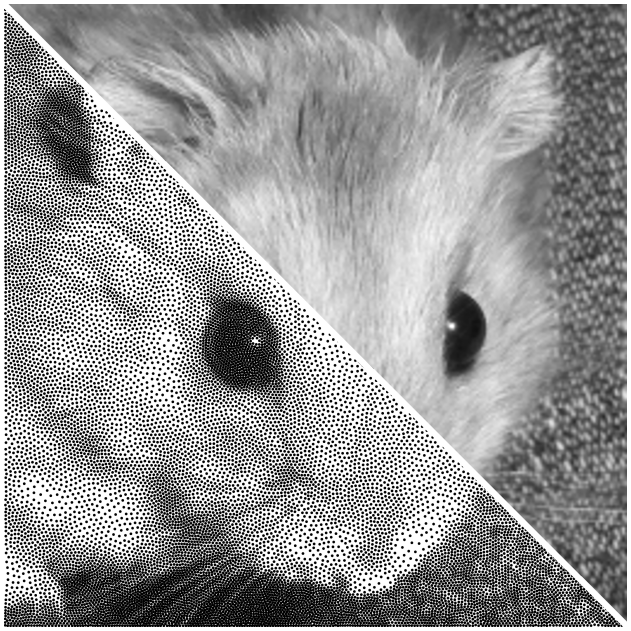

Figure 1: Comparison between halftoned result and original: Structure and tone are both well preserved.

longer restricted to preprint purposes only, but can additionally be applied to sampling problems occurring in rendering, re-lighting, or object placement, as well as for artistic non-photorealistic image visualisation.

One of the first dithering techniques was proposed by Goodall [Goo51]. In this work, images are intentionally perturbed by additive Gaussian noise prior to quantisation. Though the image might become less accurate using this method, quantisation boundaries almost disappear to the human observer. Another early class is the so-called ordered dithering [Bay73], which approximates colour values locally by a predefined pattern. This technique is often applied in modern laser printers. In 1994, Purgathofer *et al.* introduced an approach which replaces the optimal regular pattern used in [Bay73] by a more random arrangement [PTG94]. Although this yields less accurate results, there are also noticeably less artefacts with this method.

Further classic approaches frequently used today are those based on the so-called *error diffusion*. This concept considers a local image similarity measure and distributes the local error into the neighbourhood. Such algorithms often suffer from global artefacts, but remain popular thanks to their run-time efficiency. A prominent member of this class is the method proposed by Floyd and Steinberg [FS76]. Many methods are based upon this idea and extend it by a different neighbourhood treatment or grid structure [JJN76, Stu81, SA85], or by additional edge enhance-

ment [JJN76, JR76]. A variant of error diffusion, which is called dot diffusion, has been proposed by Knuth [Knu87].
More sophisticated approaches improved these early methods even further: Ostromoukhov extended the error diffusion idea by proposing a stencil design which depends on the grey values being processed [Ost01]. This reduces some artefacts created by the Floyd-Steinberg approach. Further improvements were introduced by Zhou and Fang [ZF03], who used threshold modulation [Kno89] and stencils optimised for this situation. Recently, Chang *et al.* [CAO09] proposed to use thresholds and error diffusion stencils depending on the local frequency, orientation, and contrast in the image. This is done by using five lookup tables created manually in a calibration step. A related technique which locally optimises space distance measures based on a physical model has been proposed and patented by Ilbery [Ilb00].
Another technique to reduce local errors is given by the optimisation of frequencies occurring within a local neighbourhood. Two equivalent models approaching this aim with neural networks and Markov simulations, respectively, have been proposed by Geist *et al.* [GRS93]. A more recent method of Pang *et al.* optimises a structure-preserving energy functional with related techniques, thereby obtaining sharp and detailed visual results [PQW*08]. For many practical applications, this method is still suboptimal since it often overemphasises edges noticeably. Furthermore, it relies on a good initialisation whose artefacts are often preserved. An algorithmically less demanding approach has recently been proposed by Vanderhaeghe and Ostromoukhov, whose idea is based on a polyomino tiling of the image [VO08] which helps to avoid regular patterns. For artistic purposes, there are also modern techniques which intentionally alienate images to simulate traditional drawing or printing techniques [OH95]. In this context, there has also been work focused on results in a continuous domain, like halftoning based on the Eikonal equation [PB96] which uses lines instead of points to visualise the image. A related class of algorithms that obtains continuous instead of discrete results has drawn attention in the field of non-photorealistic rendering: So-called *stippling* aims at an imitation of dot-based halftoning known from pointillism or technical drawings. While early approaches like [Sec02] focus on creating visually pleasing results, recent techniques also consider more objective criteria for evaluation [BSD09]. In both approaches, a weighted Voronoi tessellation are used. Note that such continuous methods are also required in the context of importance sampling, e.g. for Quasi-Monte Carlo processes, or to estimate integrals.
Many modern algorithms used in the previously mentioned fields today target application-dependent features such as structure-enhancement, and introduce these characteristics as an integral and mandatory part of their model. Thus, they are typically very inflexible and cannot adapt to related applications or changing external conditions. In contrast to these approaches, we propose a new continuous

model motivated by physics which finds numerous applications in the fields of stippling (cf. Figure 1), dithering, screening, or sampling. To this end, we design a very general model, and propose intuitive and flexible extensions to provide particular attributes. Consequently, our algorithm is edge enhancing to a freely adjustable degree, can easily be adapted to any printer resolution, works equally well on colour and grey images, and can be fine-tuned to yield either energetically or visually optimal results. This last property has become important in the last decade [MF92, BSD09].
Still, we like to stress that all of these extensions are entirely optional. In fact, a constant parameter set already yields highly accurate results of unsurpassed quality. We discuss these observations in more detail in Section 6.
Our paper is structured as follows: In Section 2 we give an overview of the particle model we use. After discussing extensions to this model in Section 3, we sketch our algorithm in Section 4. In Section 5 we then describe a GPU-based implementation. Finally, we evaluate the quality of our methods in Section 6, and conclude with a summary in Section 7.

2 Electrostatic Model

2.1 Repulsion

Our halftoning approach is motivated by the intuitive assumption that, for regions of constant density, e.g. regions of constant grey value in an input image, black points in the output image shall be equally distributed. We obtain this solution by modelling points as charged particles in a global particle system. A similar idea was already mentioned in [Han05], but not pursued any further – in particular since this model only allows uniform densities as input. In contrast, we now introduce an easy and versatile new model which we extend to arbitrary images in the next section.
To formalise this idea, let us consider a continuous greyscale input image $f : \Omega \to [0,1]$. The output image g of the dithering or sampling process will admit only the values 0 and 1. We regard black pixels as small equally-charged particles in an environment without frictional forces. Due to repulsion, the steady state of this evolution will always be an equilibrium maximising the pairwise distances between all particles. Thus, one obtains a globally optimal result.
To achieve this goal, we first sample f at a regular grid, and obtain a discrete image $u : \Gamma \to [0,1]$ with

$$\Gamma := \left\{ (i,j)^\top \;\middle|\; i \in \{1,\dots,n_x\}, j \in \{1,\dots,n_y\} \right\}. \tag{1}$$

Based on this sampling, we then compute the number of black particles in our target image g as

$$|\mathcal{P}| := \text{round}\left(\sum_{\boldsymbol{x}\in\Gamma}(1-u(\boldsymbol{x}))\right), \tag{2}$$

where round(x) denotes the nearest integer value of x. With this choice, the average grey value of the image is approximately preserved. Let some particle $n \in \mathcal{P} := \{1,2,\ldots,|\mathcal{P}|\}$ be characterised by a position $\boldsymbol{p}_n$ and a charge q_n. We will now compute an energetically ideal particle distribution within a system of electrostatic repulsive forces.
Consider two point charges q_1, q_2 at positions $\boldsymbol{p}_1, \boldsymbol{p}_2$, as well as the unit vector $\boldsymbol{e}_{1,2}$ from $\boldsymbol{p}_1$ to $\boldsymbol{p}_2$:

$$\boldsymbol{e}_{1,2} := \frac{\boldsymbol{p}_2 - \boldsymbol{p}_1}{\|\boldsymbol{p}_2 - \boldsymbol{p}_1\|}, \tag{3}$$

where $\|\cdot\|$ denotes the Euclidean norm. Let us first compute the repulsive force $\boldsymbol{F}_{r,1,2}$ acting on the first particle. Different to classical physics, however, we do not consider a 3-D world here, but deduce this interaction in a pure 2-D model. For this purpose, we first obtain the electric flux Φ through a circular curve of radius r by

$$\Phi = 2\pi r|\boldsymbol{E}|, \tag{4}$$

where $\boldsymbol{E}$ is the electrical field [Mes06]. By the theorem of Gauß-Ostrogradski, this electric flux also equals

$$\Phi = \frac{q_2}{\varepsilon_0}, \tag{5}$$

where ε_0 is the electric constant. Combining (4) and (5) results in

$$|\boldsymbol{E}| = \frac{q_2}{2\pi\varepsilon_0 r}, \tag{6}$$

which is now used to compute the influence on the other particle:

$$|\boldsymbol{F}_{r,1,2}| = q_1|\boldsymbol{E}| = \frac{q_1 q_2}{2\pi\varepsilon_0 r} = \frac{k q_1 q_2}{\|\boldsymbol{p}_2 - \boldsymbol{p}_1\|}, \tag{7}$$

where $k := \frac{1}{2\pi\varepsilon_0}$, i.e. k equals twice Coulomb's constant. Since this force acts in direction $-\boldsymbol{e}_{1,2}$, we obtain

$$\boldsymbol{F}_{r,1,2} = -\frac{k \cdot q_1 \cdot q_2}{\|\boldsymbol{p}_2 - \boldsymbol{p}_1\|}\boldsymbol{e}_{1,2}. \tag{8}$$

The negative sign stresses the repulsive character of $\boldsymbol{F}_{r,1,2}$, and will be used later to separate it from positive attractive forces introduced in the next section. Note

that, in contrast to the 3-D world, there is a decay of forces of the form $\frac{1}{r}$ rather than $\frac{1}{r^2}$. This finding might seem surprising at first, but can also be derived from the 3-D world by considering infinitely long, thin, charged cylinders which are aligned parallel to each other. In this scenario, one plane perpendicular to the cylinders exactly represents our 2-D system, and the forces involved in this case are again given by (8) (cf. [Mes06]).
Next, we compute the accumulated force on one particle $n \in \mathcal{P}$ by summing up its interactions with all other particles:

$$\boldsymbol{F}_{r,n} = -\sum_{\substack{m \in \mathcal{P} \\ m \neq n}} \frac{k \cdot q_n \cdot q_m}{\|\boldsymbol{p}_m - \boldsymbol{p}_n\|} \boldsymbol{e}_{n,m}. \tag{9}$$

This is the first force that will be used in our approach. If we restrict the admissible positions of the particles to the finite, two-dimensional image domain, the forces will be zero if the particles are uniformly distributed. We will see later that the restriction on the particle positions is not necessary in our final model.

2.2 Attraction

So far, we have only considered particles with a uniform distribution over the image domain. However, this simple model is only justified for constant images having only a single grey value. To extend our approach to arbitrary images consisting of multiple grey values, we introduce additional attractive forces that pull particles towards dark image regions. This is done by regarding each image grid point $\boldsymbol{x} \in \Gamma$ with grey value $u(\boldsymbol{x})$ as a particle with charge $(1 - u(\boldsymbol{x})) \cdot q$, where q is the charge of a 'black', i.e. most attractive, grid point. The attracting force of the image onto a particle n is then given by:

$$\boldsymbol{F}_{a,n} = \sum_{\substack{\boldsymbol{x} \in \Gamma \\ \boldsymbol{x} \neq \boldsymbol{p}_n}} \frac{k \cdot q_n \cdot (1 - u(\boldsymbol{x})) \cdot q}{\|\boldsymbol{x} - \boldsymbol{p}_n\|} \boldsymbol{e}_{n,x}. \tag{10}$$

In a similar style as before, $\boldsymbol{e}_{n,x}$ denotes the unit vector from the particle position $\boldsymbol{p}_n$ to the image point $\boldsymbol{x}$.
To ensure that the particles will distribute over the complete image domain, the attractive image forces must equal the repulsive particle forces, i.e. the system is required to be electrically neutral. If we neglect rounding inaccuracies introduced by (2), this holds automatically when the charge of all particle are equal, e. g.

$$\forall n \in \mathcal{P}: \quad q_n := q := 1. \tag{11}$$

Combining (9), (10) and (11), we obtain the force

$$\boldsymbol{F}_n = k \cdot q^2 \cdot \left(\sum_{\substack{\boldsymbol{x} \in \Gamma \\ \boldsymbol{x} \neq \boldsymbol{p}_n}} \frac{1-u(\boldsymbol{x})}{\|\boldsymbol{x}-\boldsymbol{p}_n\|} \boldsymbol{e}_{n,x} - \sum_{\substack{m \in \mathcal{P} \\ m \neq n}} \frac{1}{\|\boldsymbol{p}_m - \boldsymbol{p}_n\|} \boldsymbol{e}_{n,m} \right). \tag{12}$$

Within regions with constant grey values, attractive forces will be homogeneously distributed. This means that energetically ideal positions within such regions are still given at locations which maximise the distance between the particles. The model fulfils exactly the properties imposed in Section 2.1. In textured regions or at image boundaries, however, attractive forces predominate their repulsive counterparts and tie particles onto locations with darker grey values.
Moreover, we can observe the effect that, whenever a region contains the amount of particles it should do according to its average grey value, it behaves neutral with respect to 'external' particles. This means, it will neither attract nor repulse any other particles. However, if one particle enters this region because of external force, another one will leave it at some less restrictive location to restore the neutral state. Note in particular that the whole image represents a neutral region now, which automatically binds particle locations to the image domain. Our previous restriction to certain admissible locations introduced at the end of Section 2.1 thus becomes redundant.

2.3 Algorithmic Solution

Our approach computes the halftoned image as steady-state of the described particle system. Since we are only interested in the steady-state, but not in the evolution, we do not study accelerated particles in the electric field. Instead, we propose an artificial time evolution which is numerically more stable and also leads to the desired equilibrium of forces: The translation of one particle performed at a time step is given by the weighted force vector currently acting on it. From an optimisation viewpoint this can be regarded as decreasing the overall imbalance of forces every time step by repositioning the particles in accordance with the force field. The corresponding time-discrete, space-continuous update equation is given by

$$\boldsymbol{p}_n^{k+1} = \boldsymbol{p}_n^k + \tau \left(\sum_{\substack{\boldsymbol{x} \in \Gamma \\ \boldsymbol{x} \neq \boldsymbol{p}_n}} \frac{1-u(\boldsymbol{x})}{\|\boldsymbol{x}-\boldsymbol{p}_n\|} \boldsymbol{e}_{n,x} - \sum_{\substack{m \in \mathcal{P} \\ m \neq n}} \frac{1}{\|\boldsymbol{p}_m - \boldsymbol{p}_n\|} \boldsymbol{e}_{n,m} \right) \tag{13}$$

where $\boldsymbol{p}_n^k$ is the position of particle n at time step k, and where the parameter τ controls the artificial time step size. In our experiments we use $\tau = 0.1$. Note that τ incorporates some constants such as the point charges and Coulomb's constant.

3 Extensions

3.1 Discrete Particle Locations

The model developed so far is perfectly suited to find optimal point locations for applications situated in the continuous domain. For dithering, however, the set of admissible locations is finite: It is given by the rectangular grid imposed by the discrete image domain.

In order to obtain an optimal solution, it is not advisable to simply discretise the continuous result: Hexagonal structures, which are optimal in the continuous setting, cannot be adequately represented on the rectangular grid. As a consequence, unpleasant artefacts and multiple particles being assigned to one grid position yield intolerable visual results and alter the image brightness.

As a remedy, we model the rectangular grid with an additional constraint. In a first step, we thus introduce a force term to concentrate particles in the vicinity of grid points. This results in a much better solution when particles are mapped to the nearest discrete grid point after convergence. More precisely, we use the force

$$\boldsymbol{F}_{g,n} = \alpha \cdot \frac{1}{1 + \frac{\|\boldsymbol{d}_{n\boldsymbol{x}}\|^8}{\lambda^8}} \cdot \frac{\boldsymbol{d}_{n\boldsymbol{x}}}{\|\boldsymbol{d}_{n\boldsymbol{x}}\|}, \tag{14}$$

where $\boldsymbol{d}_{n\boldsymbol{x}}$ is the vector from the particle n to the nearest grid point $\boldsymbol{x} \in \Gamma$, the contrast weight $\lambda := \frac{1}{\sqrt{10}}$ decides by how much a particle is pulled back depending on the distance to this grid point, and the regularisation weight $\alpha := 3.5$ steers the influence of this force on the system. Algorithmically, this modification is realised by adding a term $\tau \boldsymbol{F}_{g,n}$ to (13).

In a second step, we account for the problem that there are still local minima in which particles can get stuck between the grid lines. When the converged system is eventually sampled to obtain a discrete dithering result, this fact can lead to ambiguous and thus sub-optimal mappings. Thus, on each time step, we project all particles back onto the closest horizontal or vertical line connecting the grid points. This operation forces particles to align in a rectangular grid structure instead of a hexagonal arrangement.

A special treatment is applied to white regions. Those should not contain any particles, and yet this can happen in particular if particles undergo large displacements per time step. While this does not constitute a problem for the continuous algorithm, our extensions for the discrete case dominate attractive forces of coloured image parts, and particles cannot return. As a remedy, we limit the maximal motion to one pixel per time step, and switch off our additional force term and projection in white regions. While the first aspect prevents a particle to move too far from desirable locations, the latter one ensures that the remaining forces suffice to let the particle return.

3.2 Edge Enhancement

To enhance the perceptual quality of dithering algorithms such as Floyd-Steinberg, an edge enhancing preprocessing of the initial image has been proposed in the literature [JJN76,JR76]. These algorithms modify the effect of a dithering method in such a way that it no longer approximates the original image as good as possible from a mathematical viewpoint, but creates a result that is visually more pleasant for human observers.
As it turns out, these filters are very similar to the well-known unsharp masking method [GW08]. Our model can easily employ this technique by applying a preprocessing step to the initial image f. This can either be a classic unsharp masking involving convolution with a Gaussian, or a discrete variant like it can be applied for the Floyd-Steinberg filter [JJN76]. Note that during edge enhancement, image values might leave the range $[0, 1]$. Though values greater than 1 even repel particles rather than attracting them, our algorithm works well under these conditions.

3.3 Different Pixel Sizes

Up to now, we assumed that a pixel in the image has the same size than a blob of the output device. However, this is not always the case: One way to deal with this problem in classic dithering schemes is to sparsify point clouds recursively, as is described in [KCDL06].
In contrast to such post-processing steps, the proposed algorithm can already adapt to arbitrary printer resolutions in a fully-automatic manner: Let η be the ratio between the area of a pixel and one blob. In that case, the number of particles is divided by η, while the charge of each particle is multiplied by η. Technically, this comes down to scaling the pixels in the input image by a factor $\frac{1}{\eta}$.

3.4 Multi-Channel Images

For colour images, we suggest a separable treatment using either the RGB or the CMY colour model. In both cases, the colour channels are independent from each other and can hence be computed separately. Although we did not optimise our algorithm for colour images, it turns out to perform well under these conditions: Both structure and colours are very accurately represented, as is shown in Section 6.1.
We are convinced that our model can be extended to 'coupled' colour models like CMYK. In such cases, the channels are no longer independent, but need to be simultaneously optimised with respect to the joint ('K') channel. This non-trivial extension is part of our ongoing work.

3.5 Artefact Prevention ('Jittering')

If there are large uniform regions in an image, our algorithm tends to arrange dots in energetically optimal, hexagonal structures. However, such regular patterns often attract the attention of the observer and are thus undesired in some applications. To deal with this problem, we now introduce a 'jittered' variant of our algorithm. Since we are convinced that the avoidance of regular patterns is a purely application-specific aim, we do not understand this extension as an integral part of the algorithm. In contrast to other methods proposed in the literature, we can control the degree of 'jittering' in our framework, and can thus find an optimal tradeoff between visual and analytic quality.

During the initialisation phase, we set up a dense, high-frequent turbulence field T, which we evaluate for all interactions with the underlying image. To this extent, we create a random vector

$$\begin{pmatrix} w_x \cdot R \cdot \sin(\varphi) \\ w_y \cdot R \cdot \cos(\varphi) \end{pmatrix} \tag{15}$$

for each grid point $\boldsymbol{x} \in \Gamma$, where $R(\boldsymbol{x})$ is a uniformly distributed random number in the interval $[0, 0.1]$, and $\varphi(\boldsymbol{x})$ is a random direction in the interval $[0°, 360°)$. The weights $w_i(f, \boldsymbol{x}) := 1 - |D_i f(\boldsymbol{x})|$ account for the fact that jittering is predominantly required in homogeneous regions, i.e. where the finite difference approximation $D_i f$ of the gradient $\partial_i f$ is close to zero. The turbulence field T is then obtained by bilinear interpolation between the grid points. Thus, we add $\tau T(\boldsymbol{p}_n^k)$ to (13) in order to obtain the new particle time step. Note that T is constant during this evolution, such that this process is convergent.

Due to the introduced turbulence field, the regular structures in the solution are broken apart. Consequently, the obtained results look much more pleasing to the eye. However, this comes at the cost of a slightly higher error.

4 Algorithm

In our algorithm, we implemented the particle evolution given by Equation (13). Since the attraction term $\boldsymbol{F}_{a,n}$ only depends on the current position, and not of the position of other particles, it is constant over the course of the evolution, and can hence be precomputed. Note that if the 'jittering' extension is desired for the current application, this precomputation also helps in this case to significantly reduce the computation time: The turbulence field described in Section 3.5 can be added to the image-based force field $\boldsymbol{F}_{a,n}$ once at initialisation time and will no longer be performed over the run of the evolution.

In order to obtain a smooth representation of the attractive force field, we sample the respective forces for all grid points and perform bilinear interpolation.

Our algorithm consists of three major steps:

1. Initialise the positions of the particles.
2. Precompute image-based force field $\boldsymbol{F}_{a,n}$.
3. Process the particle evolution until converged (or maximal number of iterations is reached).

For the first step, any initialisation with the correct amount of black pixels (see Equation (2)) can be used. Possible choices are, for example, the result from another dithering or sampling method, or a image in which the positions of the particles have been chosen randomly. In all our experiments, we used random positions whereas the probability that a position $\boldsymbol{x}$ is chosen is proportional to $1-u(\boldsymbol{x})$. The remaining two stages can be implemented straightforwardly by using Equations (10) and (13), respectively.
By applying Equation (13), the minimisation procedure is likely to end in a local minimum, thus yielding globally suboptimal results. In order to diminish this problem, we also included a procedure we called 'shaking': Every ten iterations, each particle is moved into a random direction with a random magnitude. Thereby, the maximal displacement decreases with the iteration number. More specifically, we set the maximal displacement in the a-th iteration to

$$c_1 \cdot \exp\left(\frac{-a}{1000}\right), \quad c_1 := \max\left(0, \frac{\log_2(NOI)-6}{10}\right), \tag{16}$$

where NOI is the total number of iterations to be done. From a physical standpoint, this can be regarded as Brownian motion, whereas the temperature of the particles decreases during the iterations. Such an approach is also known as *simulated annealing* which is a widely used optimisation strategy.

5 Implementation on a GPU

Since the proposed algorithm is perfectly suited to exploit the potential of Graphics Processing Units (GPUs), we implemented our method on an NVidia GeForce GTX 285 card using CUDA. Particle systems are well-known in the literature: Depending on the application, we find electrostatic or gravitational forces between the different bodies. A fast parallelisation technique for such an N-body particle system has been proposed by Nyland *et al.* [Ngu07]. We use the same idea as a basis for our algorithm. However, our model requires certain extensions and modifications to this approach, since it also relies on interaction with an input image.

As is described in the previous section, we split our algorithm in three stages, whereof the first stage, the initialisation phase, is computed on the CPU. The attractive image force field computed in step two is stored in texture memory. During the particle evolution, it can thus be read out with hardware-supported bilinear interpolation and a better caching strategy.

5.1 Creation of the Force Field

As a first GPU kernel, we compute the image-driven force field. It turns out that even this operation is structurally similar to a particle system: For a single test charge that is moved over all $i \cdot j = M$ locations, we compute the impact of all other $M-1$ point charges on grid points other than the one we are currently observing. All those forces can be arranged in an $M \times M$ force matrix with a zero main diagonal. The overall force $\boldsymbol{F}_m$ acting on any location m is then given by an integration of the m-th row:

$$\boldsymbol{F}_m = \sum_{k=1}^{M} A_{k,m}. \tag{17}$$

Without going too much into detail, we refer the reader to the well-written paper by Nyland *et al.* [Ngu07] and only sketch the differences:
Every pixel is determined by its position, which can be computed in-line, and by its grey value that can be retrieved from a texture fetch. To this extent, all threads in a block first compute the distance of 'their' coordinate to a certain other pixel, and fetch the grey value at this point. Then, they compute the resulting force increments, and add them to their force vector. Note that during this process, at most $M \cdot$ #Blocks many loads, and M writes from and to GPU memory are needed.

5.2 Evolution of the Particles

Each particle evolution step depends on three variables: The particle's own location, the image force at this point, and the other particles' locations. Since attractive image forces are encoded in a texture information that can immediately be evaluated once per update, we only need to find a suitable parallelisation strategy for the repulsive forces. Here we apply a similar technique we used for the computation of the force field (cf. Section 5.1): This time, however, we regard N particles each of which interacts with $N-1$ other ones. In this way, we obtain an $N \times N$ matrix of forces with a zero diagonal, which we integrate in one direction. For this purpose, we use similar implementation parameters as before.
In order to check the convergence after every iteration step, each thread computes the norm of its particle displacement and compares this value to a constant

threshold. Since GPU-CPU-bandwidth is rare, a per-block reduction operation is applied to these flags prior to downloading this information to the CPU. As long as one particle does not meet the convergence requirements, we proceed with our algorithm. Moreover, after a certain number of iterations, we usually want to apply 'shaking'. Since no random generator is available on the graphics card, we download the entire point cloud in this case, apply shaking on the CPU, and upload the new position list back to the device.

6 Experiments

6.1 Visual Evaluation

As a first quality criterion, we investigate the perceptual quality of our algorithm. To this extent, we compare both the normal and the 'jittered' variant of our space-continuous version against the results by Balzer *et al.* (with 1024 points per site and a Euclidean metric) [BSD09]. Moreover, the results of our space-discrete dithering approach are compared to those of the popular Floyd-Steinberg method [FS76], the techniques of Ostromoukhov [Ost01], Zhou *et al.* [ZF03], and Purgathofer *et al.* [PTG94] (using a matrix spanning the whole image), as well as to the neural network based method of Geist *et al.* [GRS93], and the structure-aware halftoning approach of Pang *et al.* (which is always initialised with Ostromoukhov's method) [PQW*08]. While it only required small modifications to integrate the freely available source code of Ostromoukhov and Balzer *et al.* into our evaluation framework, the remaining methods had to be implemented from scratch. In addition, we also evaluated against a modern commercial *stochastic screening* technique, for which sample images have generously been provided by a company that preferred to remain anonymous. In contrast to all other observed methods, this technique is optimised for printing purposes.

In our first experiment, we evaluate the behaviour of the proposed method for different grey values. To this end, we created a synthetic grey ramp covering 256 grey values, and halftoned it with different methods. As one can see in Figure 2, our pure continuous approach offers a better tone preservation than the method of Balzer *et al.*, while the latter technique suffers less from regularity artefacts. Our 'jittered' approach combines the advantages of both methods, and thus outperforms the other techniques. Among the discrete methods, the proposed approach performs best as well: Our method neither creates clustered artefacts in areas close to 50%, as is visible for the methods of Floyd-Steinberg, Ostromoukhov, and Purgathofer *et al.*, nor does it circumvent this problem at the cost of a higher graniness, as is done by the method of Zhou *et al.* or stochastic screening. Also the techniques of Pang *et al.* and Geist *et al.* do not reach the quality of our discrete

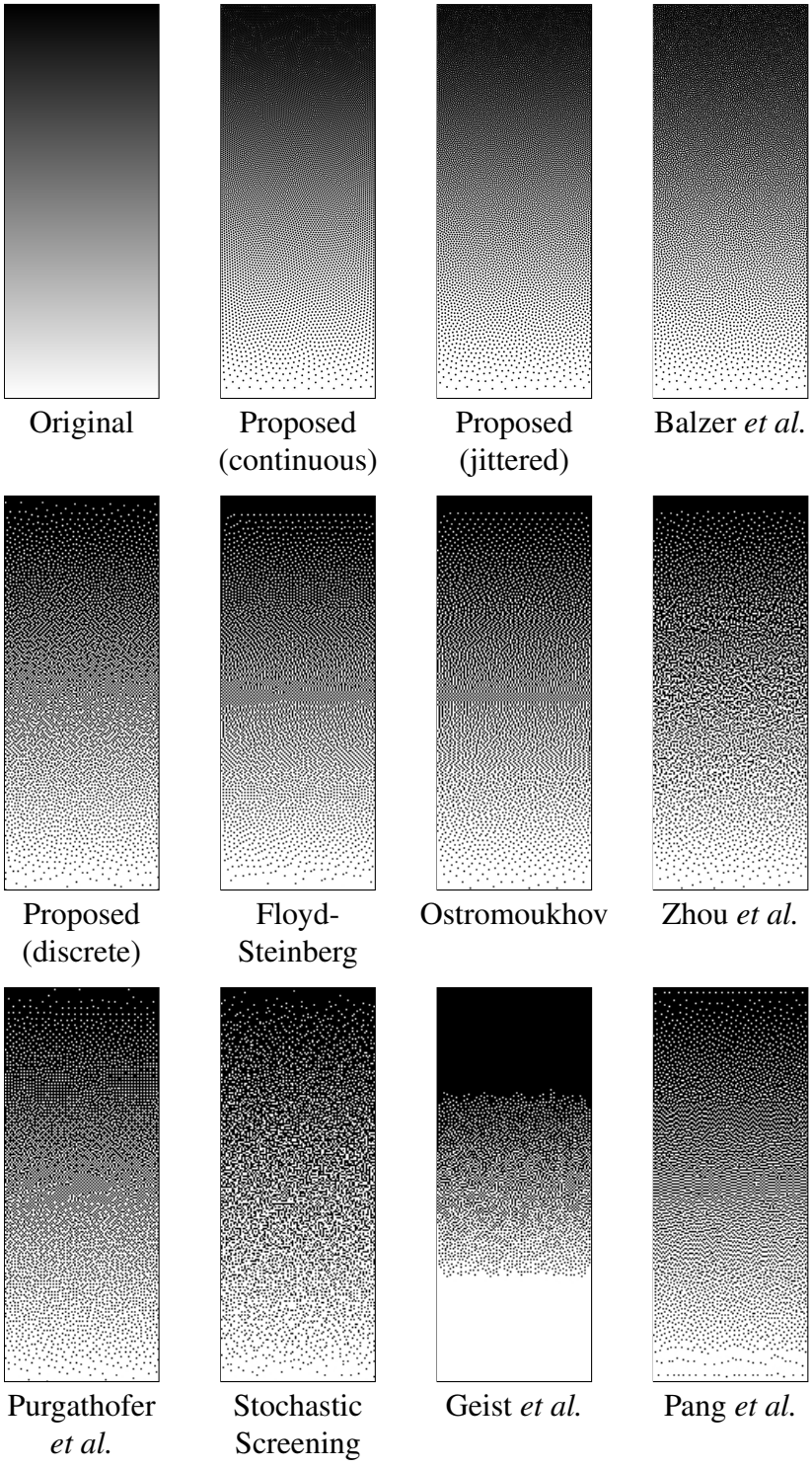

Figure 2: Visual quality of halftoning results for *grey ramp* (100×256).

approach. While the first one suffers from line artefacts instead of problems with clustering, the second one fails to represent bright and dark regions sufficiently well, since it considers only a limited neighbourhood.
Next, we compare halftoning results for the real-world colour image *statue* shown in Figure 3 using the CMY colour model. All observed continuous methods approximate the original very well, and are qualitatively similar to each other. However, in the comparison of discrete approaches, our method has less artefacts than the other algorithms. This can be seen particularly well at the top of the image, and at the chest of the statue. In contrast, the method of Pang *et al.* creates artificially-looking results by pronouncing edges too strongly, while the method of Geist *et al.* cannot reproduce the colour of the sky, as two channels are close to zero. Finally, one can observe that the stochastic screening results appear a little more grainy than others and the colours are represented differently. However, this characteristic might be intended as an optimisation to real printing devices.
As a third experiment, we evaluated a colour image containing large white areas, skin tones and bright, saturated colours, as it is depicted in Figure 4. This image is quite challenging due to two reasons: Human observers are very good at recognising errors in the representation of skin colour, as well as spotting single points in uniform areas. For the continuous methods, the visual quality is very similar at first glance. However, the method of Balzer *et al.* suffers from two minor problems visible when zooming into the image. Dark colours are not represented accurately and stray dots are visible in white regions. The latter problem occurs when colour information is distributed along the boundary of a Voronoi cell but represented by a dot in the centre. Such artefacts can for instance be observed between the middle and the ring finger of the left hand. Our continuous methods, in contrast, are not affected by these problems: They provide an accurate colour reproduction and artefact-free results. These nice properties carry over to the discrete case: Our approach outperforms the other dithering methods by a good preservation of tone, as well as by an artefact-free rendering of both coloured and empty regions. In the results of Ostromoukhov and Zhang *et al.*, one finds many stray particles in the white regions. Stochastic screening fills the entire white region with cyan dots, but we are again this property might be intentional.
The effect of our optional edge enhancement step (cf. Section 3.2) can be seen in the fourth experiment, which is depicted in Figure 5. Here we used the unsharp masking implemented in the GNU Image Manipulation Program (GIMP, version 2.2) [Pec06] as a preprocessing step. Note that small structures, like braids or curls, are much more emphasised and clearly visible. In some applications, such results are preferable to the outcome without preprocessing.
In our fifth experiment, we investigate the influence of the chosen dot size. As one can see in Figure 6, even for a dot radius of 2.0 and accordingly few blobs, lines are still perfectly preserved. Note that in all other experiments, the radius

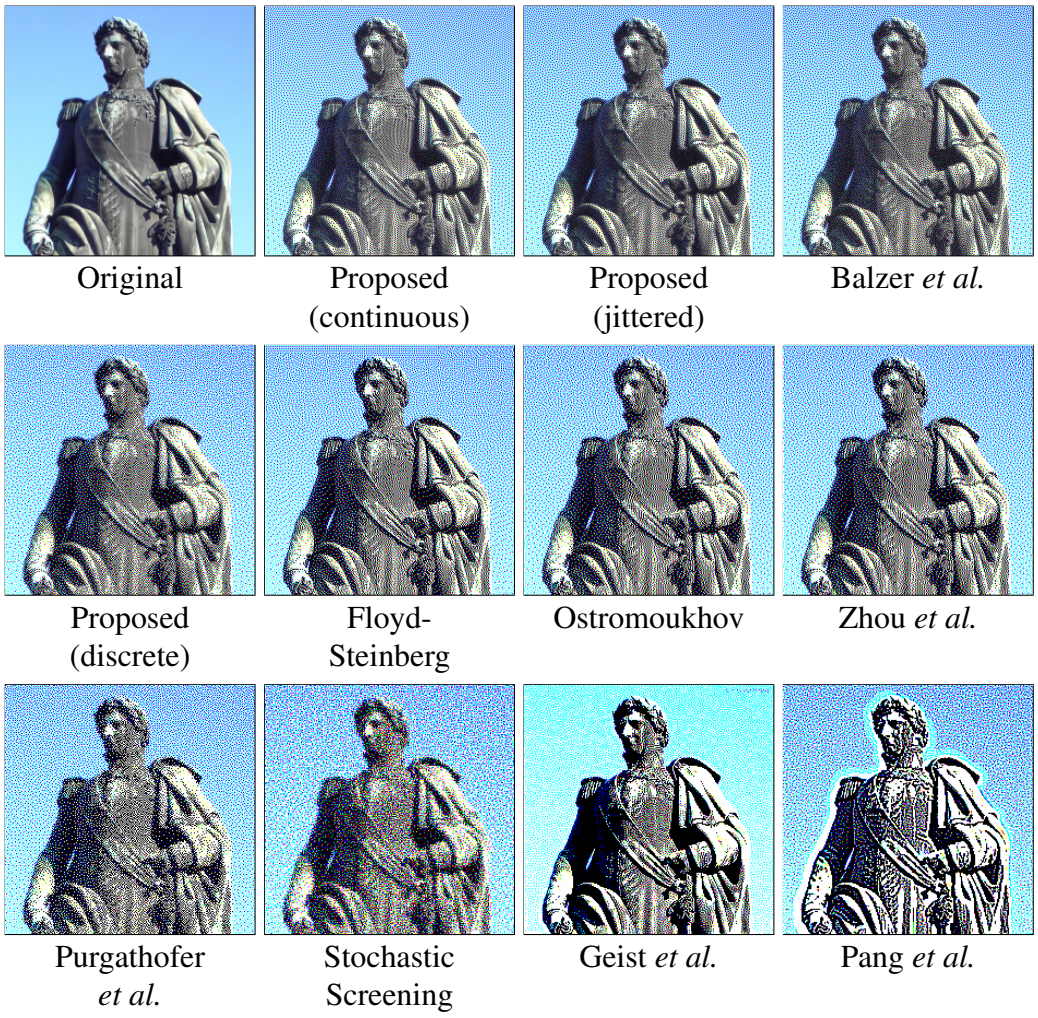

Figure 3: Visual quality of stippling and dithering results for a photograph of a statue of Charles XIV, in Norrköping (229×229). Original image licenced as CC-BY-SA 2.0 by User:Thuresson, published on Wikimedia Commons. The continuous results are visualised with dots, the remaining images have been upsampled.

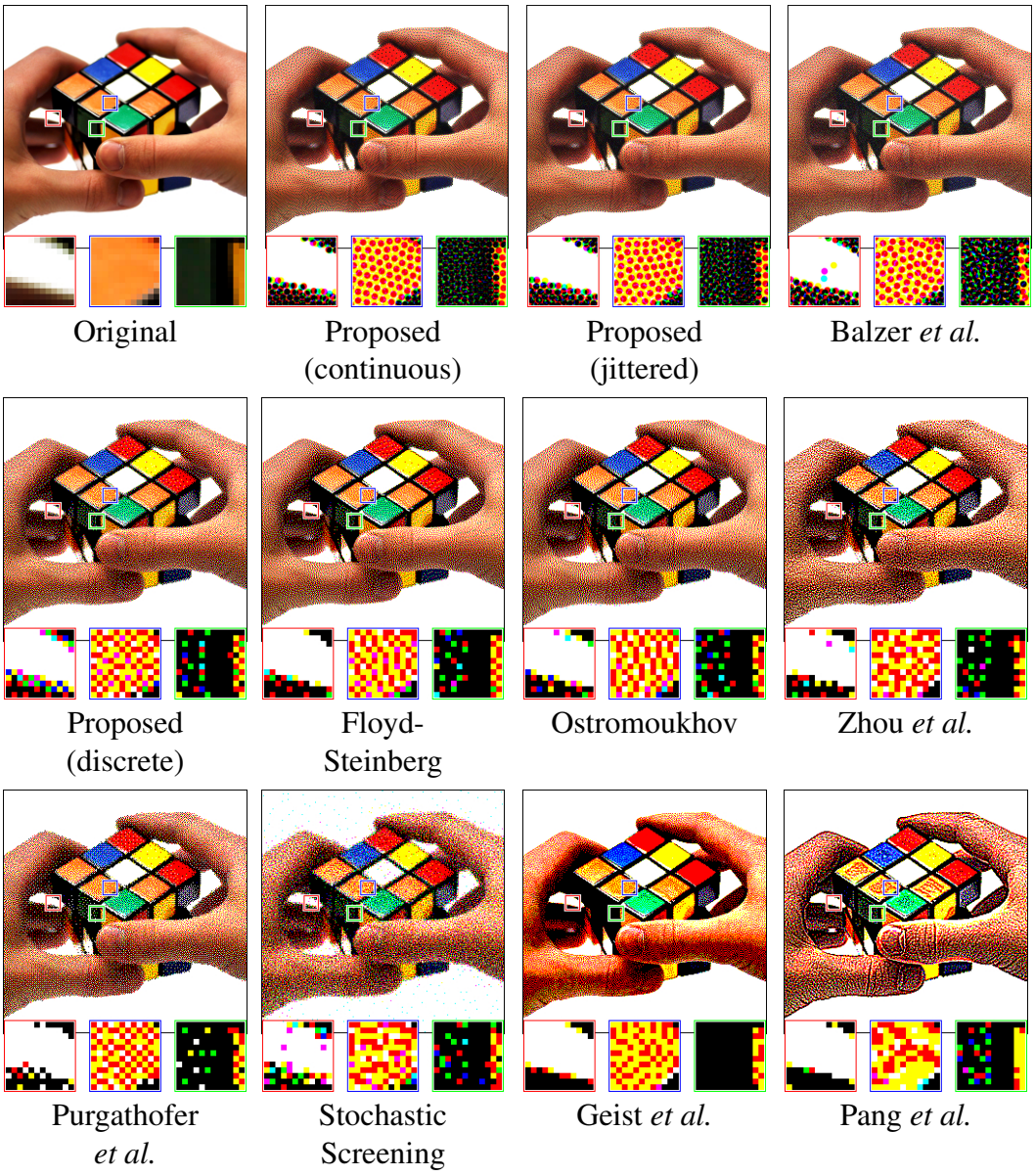

Figure 4: Visual quality of stippling and dithering results for the *Rubik's cube* colour test image (256×256). The continuous results are visualised with dots, the remaining images have been upsampled.

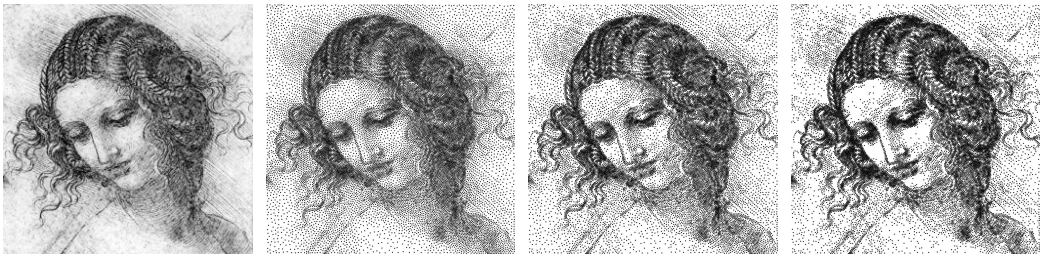

Figure 5: Leonardo da Vinci: *Study for the Head of Leda* (256×256). **From left to right:** Original, continuous result, ditto with little unsharp masking (radius 5, factor 0.5), and with more unsharp masking applied (radius 10, factor 1.0).

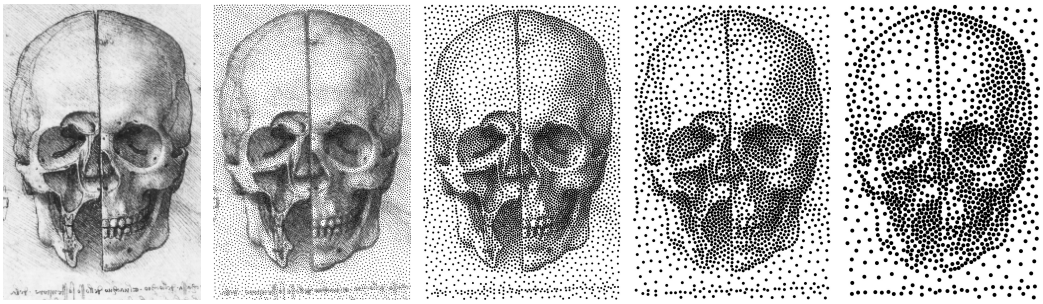

Figure 6: Leonardo da Vinci: *The skull bisected and sectioned* (200×300). Rendering of different blob sizes. **From left to right:** Original, continuous method with blob radius 0.5, 1.0, 1.5, 2.0.

has been chosen in such a way that the area of a dot is equal to the area of one pixel.
Taking these experiments together, we can conclude that our continuous methods yield accurate results of high quality. Also the discrete variant performs better than existing dithering approaches from the literature, which often suffer from 'worm'-like artefacts. In the following paragraphs, we will underline these qualitative findings by a quantitative evaluation based on different error measures.

6.2 Evaluation in Gaussian Scale Space

In this section, we evaluate the similarity of halftoning results to the original image. Motivated from the fact that human observers at a certain distance to the screen should not be able to distinguish an ideal result from the original image, we blur both the original and the halftoned image by a convolution with a Gaussian. This is meant to approximate the impulse response of the human visual system.

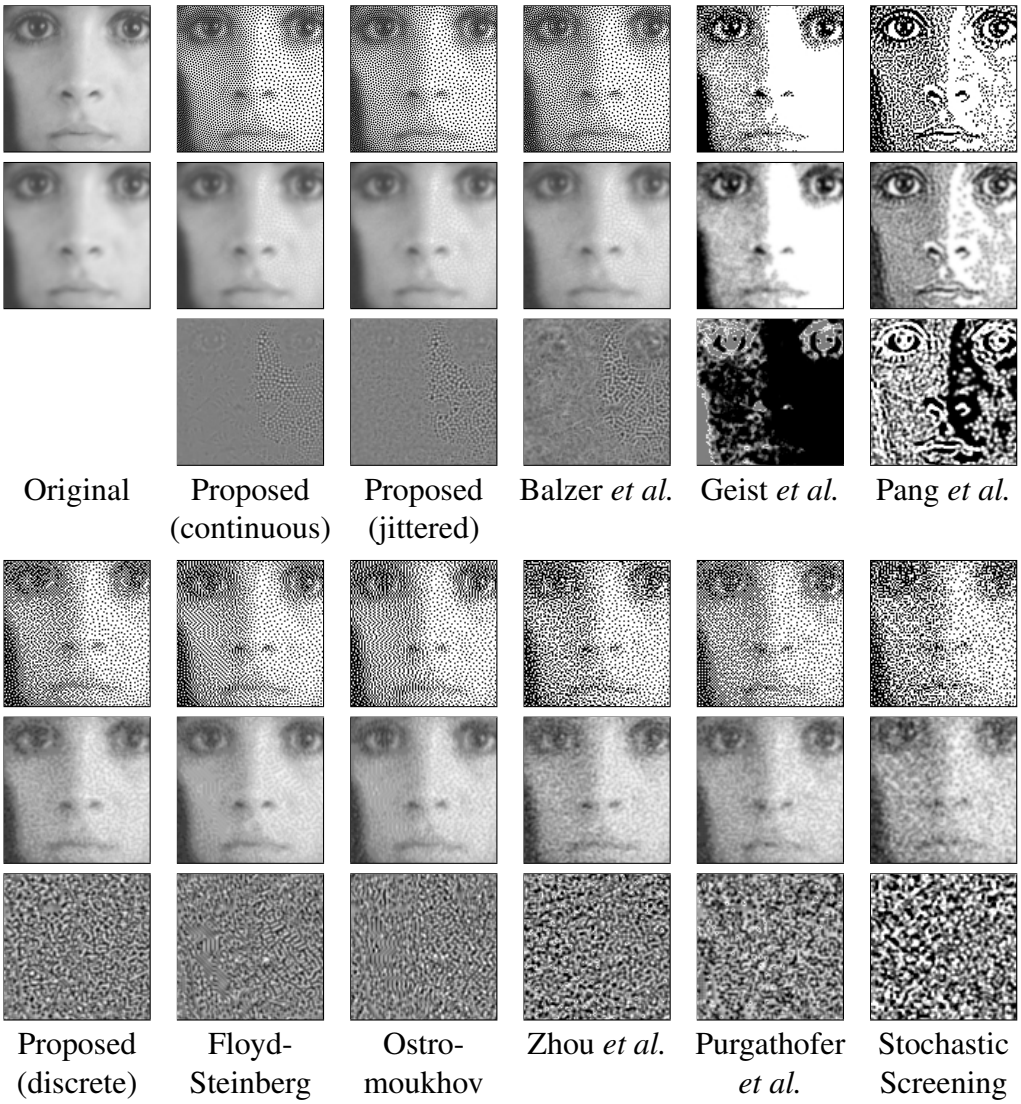

Figure 7: Zoom into the *trui* test image. From top to bottom: Dithering/Stippling result, ditto blurred with $\sigma = 1$, scaled difference image. Error-free regions appear as 50% grey in the difference image. The apparent different dot sizes in the proposed continuous result are an optical illusion that disappears when zooming into the image.

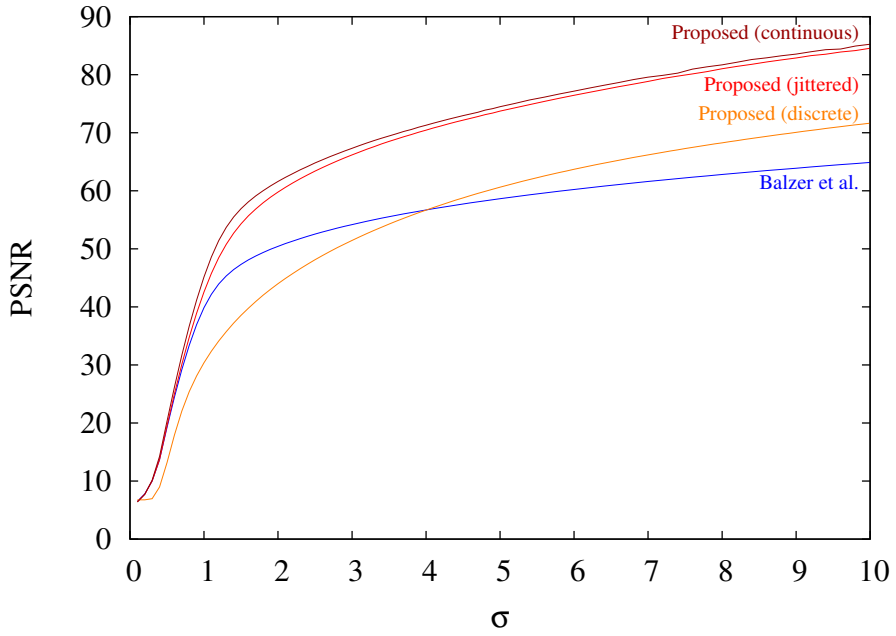

Figure 8: Peak Signal to Noise Ratio (PSNR) for stippling and screening methods under Gaussians of varying standard deviation σ. The proposed dithering method's plot is additionally given for comparison purposes.

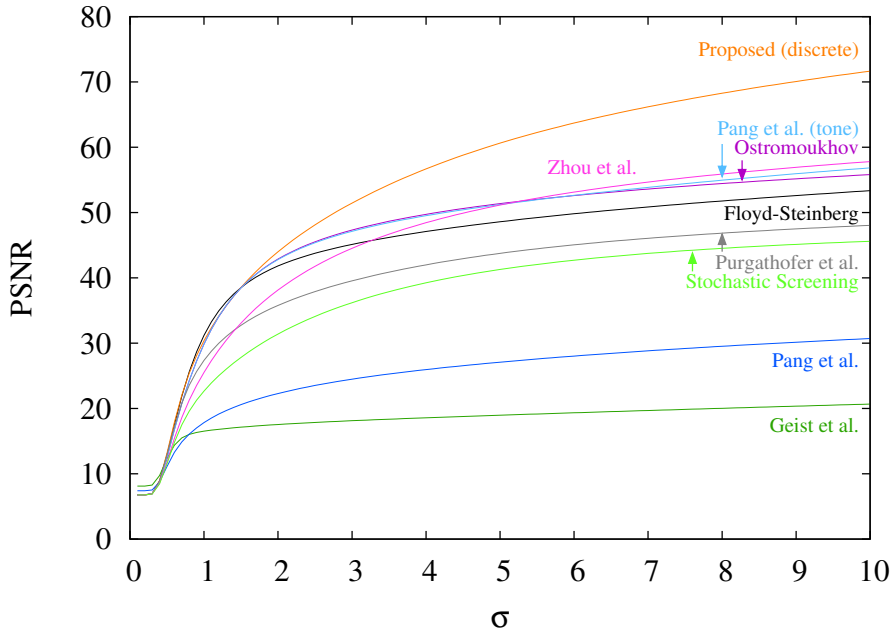

Figure 9: Peak Signal to Noise Ratio (PSNR) for dithering methods under Gaussians of varying standard deviation σ.

Figure 7 shows excerpts from this experiment on the *trui* test image. The halftoned results presented in the top row are convolved with a Gaussian of standard deviation 1.0 to obtain the images shown in the second row. Finally, the difference image of the blurred original and this blurred result is scaled by a factor of ten to increase visibility. This error map is visualised in the bottom row. We expect a good halftoning method to approximate the initial image well. That is, its error should be globally small, and no image structures should occur in the error image. Among the continuous methods, our approach performs best. Thanks to a good localisation of halftoning points, most regions are already well reconstructed by convolution with this very small Gaussian. This is clearly visible in the error image. Note that the artefacts visible in sparsely sampled regions are theoretically justified, since larger Gaussians are required to fill in the area properly. A related problem is present for all discrete methods, where point distances induced by the fixed grid spacing can also generate high errors. As a sign of its superior quality, though, our discrete approach neither reveals structural information, nor any regular patterns such as stripe artefacts. In this context, one can also observe the poor performance of the halftoning method of Pang *et al.*, or the neural network based approach by Geist *et al.*. While their structure-enhancing properties are often considered as a strength, they fail to approximate the input image well. This is clearly visible in the error map.

The standard deviation σ of the Gaussian used for blurring immediately corresponds to the simulated distance of a viewer to the image or, equivalently, the printing resolution. Thus, we compare the image quality of the image *trui* for a large range of standard deviations to stippling and screening methods in Figure 8, as well as to dithering methods in Figure 9. Plotted are the standard deviation of the Gaussian against the *Peak-Signal-to-Noise-Ratio (PSNR)*.

In this experiment, our methods are again clearly superior to current state-of-the-art techniques. In particular, this holds for larger σ which attests a good approximation of the original image under arbitrary observation distances, or printing resolutions. This performance is mainly due to the global optimisation of all point locations within our particle system. If we perturb the force field to reduce artefacts, the quality of our methods still remains excellent. In Figure 8, we can observe that even our discrete dithering technique outperforms the method of Balzer *et al.* for sufficiently large σ, though continuous approaches should be able to achieve more accurate results than discrete ones. As a consequence, our discrete approach also outperforms the other dithering methods. Figure 9 further confirms the high errors for the structure-enhancing methods we observed before: While structure preservation is often considered an important property, this advantage is bought with a tradeoff in tone preservation. Indeed, if we only consider the tone similarity term of the model by Pang *et al.* as for the PSNR experiments in [PQW*08], this method goes head to head with many other dithering techniques. However, there

is no similar way to improve the method of Geist *et al.*.

6.3 Evaluation of Blue Noise Behaviour

Our third way to evaluate our results concentrates on the primary aim of our method: To distribute black pixels equally within regions of constant grey value. To check this property, we perform a frequency analysis in the Fourier domain. For this purpose, we halftone constant images, and compute the difference between result and original. This power spectrum of the difference image is then radially averaged to obtain statistical information about contained frequencies. Since high frequencies are in the nature of the problem, they cannot be eliminated. Low frequencies, however, should not be in the difference image at all. This so-called Blue Noise behaviour analysis has first been proposed by Ulichney [Uli88] and is widely used for evaluating dithering and screening methods.
Figure 10 depicts our measurements. In order to make them more robust against outliers and noise, we average the power spectra of ten randomly selected patches of the obtained results [Bar78]. Since error diffusion algorithms can suffer from boundary artefacts, we only consider patches with a sufficient distance to the image boundary.
Ulichney identifies a well-formed dithering method from a specific shape of the radially averaged power spectrum of difference images [Uli88]. Depending on the observed grey value, he first computes the *principal frequency* by

$$f_g = \sqrt{0.5 - |g - 0.5|} \tag{18}$$

where g is the grey value normalised to the range $[0,1]$. At this principal frequency, Ulichney expects a low-frequency cutoff, characterised by a sharp transition region below f_g and a potential overshoot. Above this frequency, he allows a flat blue noise behaviour much lower than the peak. A corresponding sketch is shown in Figure 10 (top left).
Note that all continuous solutions have an additional higher principal frequency depicted in their graphs. Here, the higher one refers to the principal frequency of a hypothetical hexagonal grid described by the energetically optimal hexagonal arrangement of points within uniform regions. However, since the points still depend on the underlying rectangular input image, we expect the characteristic peak to occur somewhere in between these two frequencies. This property is indeed confirmed by our experiments.
Observing the graphs in Figure 10, we see that our continuous approaches outperform the method of Balzer *et al.* due to their steeper transition region. In the discrete case, we achieve a similar performance than the method of Floyd-Steinberg, and a slightly better one than the method of Zhou *et al.*. Again, we observe a much

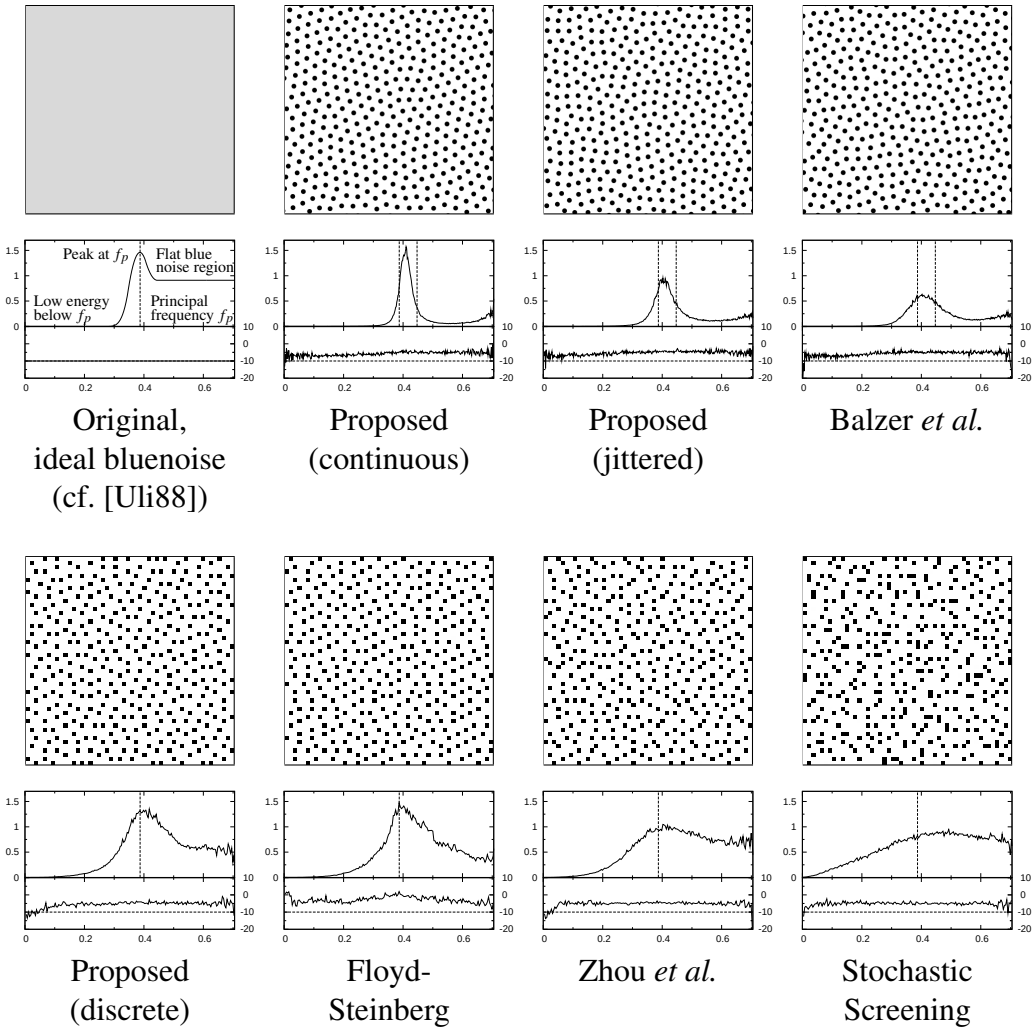

Figure 10: Evaluation of the blue noise behaviour for 85% grey. Every method is represented by a patch of the generated result, by its radially averaged power spectrum (upper graph), and the anisotropy of the power spectrum in decibels (lower graph). The diagrams are obtained from an average of 10 periodograms.

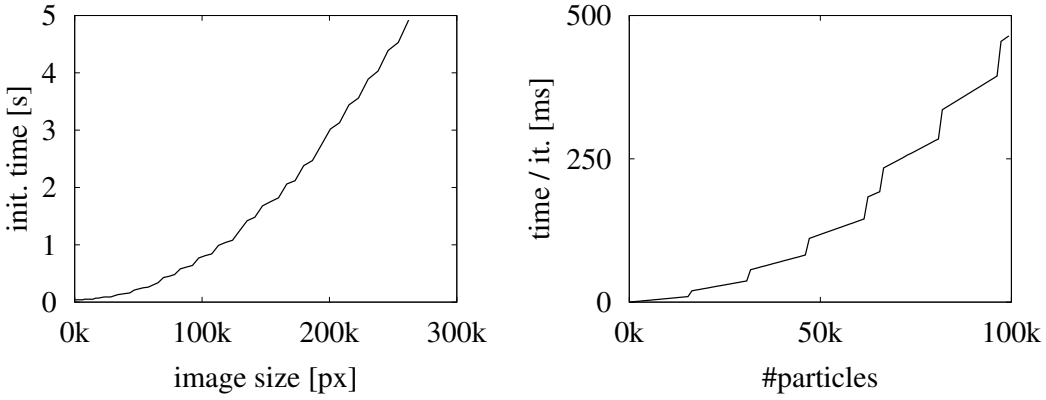

Figure 11: Time required by the proposed algorithm for initialisation (left), and for particle evolution (right).

higher graininess of stochastic screening compared to other methods, which is reflected in a less pronounced blue noise behaviour.
Additionally, we investigate the anisotropy of the (averaged) power spectra, i.e. the variance on concentric circles around the lowest frequency. From a perfect halftoning method, we expect rotational invariance, which expresses in a high isotropy of the power spectrum. Since we considered ten patches per method, we expect stochastic background noise at a level of -10dB. We denote this theoretical limit by a dashed line. Most observed approaches perform equally well with an average anisotropy of about -5dB. The only exception is the method of Floyd-Steinberg, which offered a good blue noise behaviour, but performs noticeably worse here. This confirms our observation that in contrast to early error diffusion algorithms, these methods do not reveal striking directional artefacts on any scale.

6.4 Runtime

Figure 11 illustrates the runtime required by the proposed (continuous) algorithm. Note that the initialisation of the force field only depends on the number of pixels, while the particle evolution depends on the number of particles. Thus, both initialisation and evolution times are shown. The jumps in both benchmarks are due to CUDA specifics, and can best be explained by means of the right graph: On the GTX 285, we use 128 blocks, each consisting of 512 threads. This explains the discontinuity for $128 \cdot 512 = 65536$ particles. All further discontinuities are due to the number of 30 streaming multiprocessors, which cause jumps at $n \cdot 65536 + m \cdot 30 \cdot 512$ particles whenever the number of blocks exceeds a multiple of 30. For the initialisation depicted in the left graph, the same behaviour can be observed with respect to the pixels.

This first implementation of our algorithm has a quadratic complexity in the number of particles, and a linear dependence on the number of iterations. A good initialisation like the randomised approach described above positively affects the runtime, and reasonable results can already be obtained after several hundred iterations: 300 iterations on 256×256 pixels with 16384 particles (25%) take about 6.5 seconds. Furthermore, we are currently investigating the quality of fast approximations to the proposed model, and are confident to obtain a linear complexity without a noticeable deterioration of the results in the near future.

7 Summary

We have proposed a novel and physically justified technique for image dithering, sampling, and stippling. Our flexible model can easily be adopted to many different fields of application, and outperforms the best algorithms known from the literature. It can be applied on and discretised to any scale depending on the printer resolution, allows arbitrary edge enhancement, and works equally well on greyscale and colour images. In applications in which visual artefacts must be avoided, a second parameter can be used as a tradeoff between quality and prevention of artefacts. Due to the global optimisation process involved in our approach, it yields excellent results, independent of the viewing distance or resolution. Future work includes the development of fast numerical schemes as well as specific models for other colour spaces, such as CMYK.

Acknowledgements

Our research was partly funded by the Deutsche Forschungsgemeinschaft under project We2605/5-2 and the Cluster of Excellence "Multimodal Computing and Interaction". This is gratefully acknowledged. The authors further thank the reviewers for their helpful comments, and the company that provided samples for stochastic screening.

References

[Bar78] BARTLETT M. S.: *An Introduction to Stochastic Processes with Special Reference to Methods and Applications*, third ed. Cambridge University Press, Cambridge, UK, 1978.

[Bay73] BAYER B.: An optimum method for two-level rendition of continuous-tone pictures. In *IEEE 1973 International Conference on Communications* (1973), vol. 1, pp. 11–15.

[BSD09] BALZER M., SCHLÖMER T., DEUSSEN O.: Capacity-constrained point distributions: A variant of Lloyd's method. *ACM Transactions on Graphics 28*, 3 (2009), Article No. 86.

[CAO09] CHANG J., ALAIN B., OSTROMOUKHOV V.: Structure-aware error diffusion. *ACM Transactions on Graphics 28*, 5 (2009), Article No. 162.

[FS76] FLOYD R. W., STEINBERG L.: An adaptive algorithm for spatial grey scale. In *Proc. Society of Information Display* (1976), vol. 17, pp. 75–77.

[Goo51] GOODALL W. M.: Television by pulse code modulation. *Bell System Technical Journal 30* (Jan. 1951), 33–49.

[GRS93] GEIST R., REYNOLDS R., SUGGS D.: A Markovian framework for digital halftoning. *ACM Transactions on Graphics 12*, 2 (1993), 136–159.

[GW08] GONZALEZ R. C., WOODS R. E.: *Digital Image Processing*, third ed. Prentice Hall, Upper Saddle River, 2008.

[Han05] HANSON K. M.: Halftoning and Quasi-Monte Carlo. In *Sensitivity Analysis of Model Output*, Hanson K. M., Hemez F. M., (Eds.). Los Alamos Research Library, 2005, pp. 430–442.

[Ilb00] ILBERY P. W. M.: U.S. patent No. 6,124,844, 2000.

[JJN76] JARVIS J. F., JUDICE C. N., NINKE W. H.: A survey of techniques for the display of continuous tone pictures on bilevel displays. *Computer Graphics and Image Processing 5* (1976), 13–40.

[JR76] JARVIS J. F., ROBERTS C. S.: A new technique for displaying continuous tone images on a bilevel display. *IEEE Transactions on Communications 24* (1976), 891–898.

[KCDL06] KOPF J., COHEN-OR D., DEUSSEN O., LISCHINSKI D.: Recursive Wang tiles for real-time blue noise. *ACM Transactions on Graphics 25*, 3 (2006), 509–518.

[Kno89] KNOX K. T.: Edge enhancement in error diffusion. In *Paper Summaries from The Society for Imaging Science and Technology, 42nd Annual Conference* (Boston, May 1989), pp. 310–313.

[Knu87] KNUTH D. E.: Digital halftones by dot diffusion. *ACM Transactions on Graphics 6*, 4 (1987), 245–273.

[Mes06] MESCHEDE D.: *Gerthsen Physik*, 23rd (revised) ed. Springer-Lehrbuch. Springer, Berlin/Heidelberg, 2006.

[MF92] MCCOOL M., FIUME E.: Hierarchical Poisson disk sampling distributions. In *Proceedings of the Conference on Graphics Interfaces '92* (San Francisco, CA, USA, 1992), Booth K. S., Fournier A., (Eds.), Morgan Kaufmann Publishers Inc., pp. 94–105.

[Ngu07] NGUYEN H. (Ed.): *GPU Gems 3*. Addison–Wesley Professional, Aug. 2007.

[OH95] OSTROMOUKHOV V., HERSCH R.: Artistic screening. In *Proc. SIGGRAPH '95* (New York, NY, Aug. 1995), pp. 219–228.

[Ost01] OSTROMOUKHOV V.: A simple and efficient error-diffusion algorithm. In *Proc. SIGGRAPH 2001* (Los Angeles, CA, Aug. 2001), Fiume E., (Ed.), Computer Graphics Proceedings, Annual Conference Series, pp. 567–572.

[PB96] PNUELI Y., BRUCKSTEIN A. M.: Gridless halftoning: A reincarnation of the old method. *CVGIP: Graphical Models and Image Processing 58*, 1 (Jan. 1996), 38–64.

[Pec06] PECK A.: *Beginning GIMP: From Novice to Professional*, first ed. Apress, 2006.

[PQW*08] PANG W.-M., QU Y., WONG T.-T., COHEN-OR D., HENG P.-A.: Structure-aware halftoning. *ACM Transactions on Graphics 27*, 3 (Aug. 2008), 89:1–89:8.

[PTG94] PURGATHOFER W., TOBLER R. F., GEILER M.: Forced random dithering: Improved threshold matrices for ordered dithering. In *Proceedings of 1st IEEE International Conference on Image Processing* (Austin, Texas, Nov. 1994), vol. 2, pp. 1032–1035. Also in Graphics Gems 5, p 297–301.

[SA85] STEVENSON R. L., ARCE G. R.: Binary display of hexagonally sampled continuous-tone images. *Journal of the Optical Society of America A 2*, 7 (Mar. 1985), 1009–1013.

[Sec02] SECORD A.: Weighted Voronoi stippling. In *Proceedings of the second international symposium on Non-photorealistic animation and rendering* (New York, NY, USA, 2002), ACM, pp. 37–43.

[Stu81] STUCKI P.: *MECCA – A multiple-error correcting computation algorithm for bilevel image hardcopy reproduction*. Tech. Rep. RZ1060, IBM Research Lab, Zurich, Switzerland, 1981.

[Uli88] ULICHNEY R. A.: Dithering with blue noise. *Proceedings of the IEEE 76*, 1 (Jan. 1988), 56–79.

[VO08] VANDERHAEGHE D., OSTROMOUKHOV V.: Polyomino-based digital halftoning. In *IADIS International Conference on Computer Graphics and Visualization 2008* (July 2008), Isaías P., (Ed.), pp. 11–18.

[ZF03] ZHOU B., FANG X.: Improving mid-tone quality of variable-coefficient error diffusion using threshold modulation. *ACM Transactions on Graphics 22*, 3 (2003), 437–444.