Image Halftoning by Mean Field Annealing

Thomas K. Miller

David E. Van den Bout

Center for Communications and Signal Processing
North Carolina State University

October 27, 1988

CCSP-TR-88/27

The image halftoning problem is to find the best "black and white" representation of a grey-level image as viewed by a human observer. We can formulate the problem mathematically as follows: Let G represent the pixels of an $N \times N$ grey level image such that $g_{ij} \in [0,1]\ \forall i,j$, and let H represent the pixels of the best corresponding halftone image such that $h_{ij} \in \{0,1\}\ \forall i,j$. Now we must find the halftone image which minimizes the objective function $E = f(H)$. One reasonable objective function is the sum squared error between the pixels of the grey-level image and the halftone image, filtered by the human visual response. This can be written in the spatial frequency domain as

$$E_{sse} = \sum_i \sum_j \mathcal{V}_{ij}\mathcal{V}_{ij}^*(\mathcal{H}_{ij} - \mathcal{G}_{ij})(\mathcal{H}_{ij} - \mathcal{G}_{ij})^* \tag{1}$$

where $\mathcal{G}_{ij}$ and $\mathcal{H}_{ij}$ are the pixels of the 2D-DFT of the original image and the halftone image, respectively, and $\mathcal{V}$ is the spatial frequency domain representation of the human visual response.

In mean field annealing, the discrete degrees of freedom, or *spins*, are allowed to vary continuously between zero and one. In this case, the spins are the pixels of the halftone image, H. If those pixels are allowed to vary continuously between zero and one, then clearly E_{sse} is minimized when $H = G$. To enforce the constraint that $h_{ij} \in \{0,1\}$ we can add a *penalty term* which goes to zero when all pixels of H are either zero or one. This yields a new objective function,

$$E = \sum_i \sum_j \mathcal{V}_{ij}\mathcal{V}_{ij}^*(\mathcal{H}_{ij} - \mathcal{G}_{ij})(\mathcal{H}_{ij} - \mathcal{G}_{ij})^* \ + \ \gamma \sum_i \sum_j h_{ij}(1 - h_{ij})\,. \tag{2}$$

Since the DFT is a linear operator, we can write E as a quadratic function of H, which gives us the form of an Ising Hamiltonian. To achieve this, we rewrite the equation as follows:

$$E \ = \ (\mathbf{V}(\mathbf{Fh} - \mathbf{Fg}))^H \mathbf{V}(\mathbf{Fh} - \mathbf{Fg}) \ + \ \gamma(\mathbf{Ih})^T \mathbf{I}(\mathbf{j} - \mathbf{h}) \tag{3}$$

where

$$\begin{aligned}
\mathbf{V} &= \begin{bmatrix} \mathcal{V}_{00} & 0 & \cdots & 0 \\ 0 & \mathcal{V}_{01} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \mathcal{V}_{N-1,N-1} \end{bmatrix} \\
f_{ij} &= \frac{1}{N^2}\exp\left(\frac{2\pi \jmath (ij + \lfloor \frac{i}{N} \rfloor \lfloor \frac{j}{N} \rfloor)}{N}\right) \\
\mathbf{g} &= [g_{00}\, g_{01}\ \cdots\ g_{N-1,N-1}]^T \\
\mathbf{h} &= [h_{00}\, h_{01}\ \cdots\ h_{N-1,N-1}]^T \\
\mathbf{j} &= [1\ 1\ \cdots\ 1]^T\,.
\end{aligned}$$

Simplifying equation 3 yields

$$E \ = \ \mathbf{h}^T \Omega \mathbf{h} \ + \ \mathbf{h}^T \beta \ + \ C \tag{4}$$

where

$$\Omega \ = \ \mathbf{Q} \ - \ \gamma \mathbf{I}$$

$$\begin{aligned} \beta &= \gamma\mathbf{j} - 2\mathbf{Q}\mathbf{g} \\ C &= \mathbf{g}^T\mathbf{Q}\mathbf{g} \\ \mathbf{Q} &= \mathrm{Real}\left\{\mathbf{F}^H\mathbf{V}^H\mathbf{V}\mathbf{F}\right\} . \end{aligned}$$

(Note that $\mathbf{F}$ is the $N^2 \times N^2$ matrix which implements the 2D-DFT. Note also that the imaginary part of $\mathbf{Q}$ is dropped because $\mathbf{Q}$ is Hermitian, and the complex conjugates exactly cancel the imaginary component of $\mathbf{Q}$ in equation 3.)

Solving analytically for $\mathbf{Q}$, we find that

$$q_{ij} = \frac{1}{N^4}\sum_{k=0}^{N^2-1} v_{kk}v_{kk}^* \cos\left(\frac{2\pi\left(k(i-j) + \lfloor\frac{k}{N}\rfloor\left(\lfloor\frac{i}{N}\rfloor - \lfloor\frac{j}{N}\rfloor\right)\right)}{N}\right) . \tag{5}$$

¿From this equation we see that the diagonal elements of Ω can be annhiliated by choosing

$$\gamma = \frac{1}{N^4}\sum_{k=0}^{N^2-1} v_{kk}v_{kk}^* .$$

For the Ising Hamiltonian, the "mean field" experienced by the i^{th} pixel in $\mathbf{h}$ is

$$\Phi_i = \frac{\partial E}{\partial h_i} = 2\sum_{j=0}^{N^2-1} \omega_{ij}h_j + \beta_i . \tag{6}$$

We can now write down an algorithm for finding a halftone image which seeks to minimize the objective function E:

Initialize each spin average to its high temperature limit and calculate the initial energy of the configuration:

$$\begin{aligned} h_i &= \frac{1}{2}\,(i = 0, \ldots, N^2 - 1) \\ E &= \mathbf{h}^T\Omega\mathbf{h} + \mathbf{h}^T\beta + C \end{aligned}$$

Initialize T to the "critical temperature," T_c.

Repeat until "frozen":

Repeat until in equilibrium at T:

Update all spins in *random* order:

$$\Phi_i = 2\sum_{j=0}^{N^2-1} \omega_{ij}h_j + \beta_i$$

$$
\begin{aligned}
h_i' &= \frac{1}{1+\exp\left(\Phi_i/T\right)} \\
E &= E + \Phi_i(h_i' - h_i) \\
h_i &= h_i'
\end{aligned}
$$

End Repeat.

Reduce temperature by α: $T = \alpha T\,(\alpha < 1)$

End Repeat.

A couple of properties of the spin interaction matrix Ω are worthy of note. First, Ω is independent of the input image; it depends only on the human visual response, $\mathbf{V}$. Second, we note that the i^{th} row in Ω tells us how the i^{th} pixel in H is affected by all other pixels in H. We find that each row in Ω is a permutation of the first row, such that a symmetric kernel is formed around the corresponding pixel. To illustrate this, consider the example of an 8×8 image. The coefficients of the human visual response filter (*these are fake!*) are given in the following table.

1.000	1.000	0.700	0.300	0.200	0.300	0.700	1.000
1.000	1.000	0.700	0.300	0.200	0.300	0.700	1.000
0.700	0.700	0.490	0.210	0.140	0.210	0.490	0.700
0.300	0.300	0.210	0.090	0.060	0.090	0.210	0.300
0.200	0.200	0.140	0.060	0.040	0.060	0.140	0.200
0.300	0.300	0.210	0.090	0.060	0.090	0.210	0.300
0.700	0.700	0.490	0.210	0.140	0.210	0.490	0.700
1.000	1.000	0.700	0.300	0.200	0.300	0.700	1.000

Examining the row of Ω corresponding to $h_{3,3}$ in "pixel-morphic" fashion, we see the symmetric interaction terms around the pixel:

0.107	-0.020	-0.735	-1.373	-0.735	-0.020	0.107	0.052
-0.020	0.004	0.135	0.252	0.135	0.004	-0.020	-0.010
-0.735	0.135	5.049	9.437	5.049	0.135	-0.735	-0.360
-1.373	0.252	9.437	0.000	9.437	0.252	-1.373	-0.672
-0.735	0.135	5.049	9.437	5.049	0.135	-0.735	-0.360
-0.020	0.004	0.135	0.252	0.135	0.004	-0.020	-0.010
0.107	-0.020	-0.735	-1.373	-0.735	-0.020	0.107	0.052
0.052	-0.010	-0.360	-0.672	-0.360	-0.010	0.052	0.026

Note that the large terms in the matrix are clustered in the neighborhood of the pixel. If we compute the "power" of the matrix as the sum of squares of its coefficients (*probably not very meaningful*) we find that

97% of the "power" in this case is in a 3×3 kernel around each pixel. (If we look at the "magnitude" as the sum of the absolute values, then 77% of the "magnitude" is in the 3×3 kernel.) This modification would reduce the complexity of the MFA halftoning algorithm from $O(M^2)$ to $O(M)$, where M is the number of pixels in the image. This also suggests the possibility of constructing a neural net for image halftoning with a neuron for each pixel, where each neuron is connected to its eight nearest neighbors. (This idea needs to be investigated further using the *real* human visual response!)