Adaptive Contrast Enhancement by Entropy Maximization with a 1-K-1 Constrained Network[1]

I. King
king@cs.cuhk.hk

L. Xu
lxu@cs.cuhk.hk

Department of Computer Science
The Chinese University of Hong Kong
Shatin, New Territories, Hong Kong

Abstract

This paper uses the Maximum Entropy Principle to construct a 1-K-1 constrained sigmoidal neural network which adaptively adjusts its gain parameters to control the transfer function in order to maximize the entropy measure at the output for image contrast enhancement. We demonstrate how the model works with the standard `lena` image.

1 Introduction

In recent years, information theoretic approaches have been successful in characterizing the functionality of unsupervised sensory processing in neural systems, e.g., Infomax (Linsker, 1988), redundancy reduction (Barlow, 1989; Atick, 1992), mutual information maximization among output units with spatial coherence (Becker and Hinton, 1992), mutual information maximization between inputs and outputs (Linsker, 1989), etc.

In the context of image processing, a transfer function similar to the *Cumulative Distribution Function* (CDF) of the input image yields the maximization of the entropy at the output. This transformation acquires a uniformed output histogram distribution which is histogram equalization used in image enhancement (Gonzalez and Woods, 1992).

To perform histogram equalization on a set of images, one needs to calculate the CDF for every image which is computationally intensive. An *unsupervised* and *adaptive* method which approximates the CDF function can generalize to other on-line images quickly without recomputing the CDF each time while maintaining maximized output entropy.

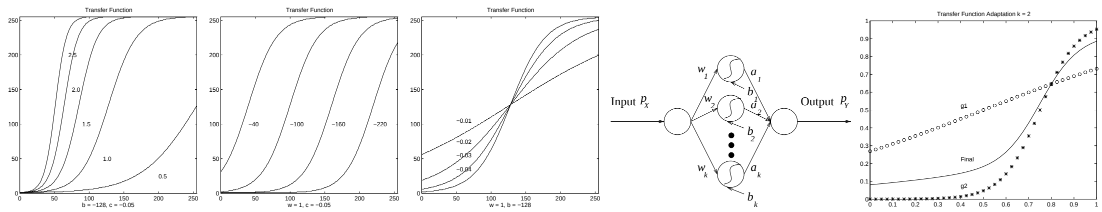
Moreover, entropy maximization of the output image achieves: (1) maximum contrast with the maximum variance, (2) utilization of the full dynamic range of the sensory and output circuit, and (3) maximum information being preserved.

Motivated by the information theoretic approaches, we use the Maximum Entropy Principle (MEP) to guide us in constructing a 1-K-1 constrained neural network. It approximates the CDF for image contrast enhancement by adaptively modifying its parameters to obtain output entropy maximization through a linear combination of monotonically increasing transfer functions.

2 Maximum Entropy Principle

The MEP simply maximizes the entropy measurement subject to satisfying a set of given constraints. Let p_X and p_Y be the *Probability Density Function* (PDF) of continuous random variables X and Y respectively with natural constraints: $\int dx\ p_X = \int dy\ p_Y = 1$ and $0 \leq p_X(x), p_Y(y) \leq 1$. Suppose $g : X \mapsto Y$, then $p_X dx = p_Y dy$ or $p_X = p_Y g'(x, \theta)$. Here $g(\cdot)$ depends on a set of parameters, $\theta = \{\theta_1, \theta_2, \cdots, \theta_m\}$, e.g., $g(Y|X, \theta_1, \cdots, \theta_m)$. We will assume that X denotes the input intensity information and Y denotes the transformed output.

[1]This work was supported in parts by the Hong Kong Research Grant Council, NO: 220500720 and 220500910.



(a) (b) (c) (d) (e)

Figure 1: (a) Adaptive sigmoidal transfer function, $b = -128, c = -0.05, w = \{0.5, 1.5, 2.0, 2.5\}, X = \{0, \cdots, 255\}$, (b) Adaptive sigmoidal transfer function. $w = 1, c = -0.05, b = \{-40, -100, -160, -220\}$. (c) Adaptive sigmoidal transfer function, $w = 1, b = -128, c = \{-0.1, -0.2, -0.3, -0.4\}$. (d) A 1-$K$-1 sigmoidal approximation neural network. (e) Summation (solid line) of two monotonically increasing functions ('o' and '*').

Our goal is then to maximize the entropy, $\mathcal{H}(Y|X, \theta)$, of the output variable Y with respect to the parameters, θ, as

$$\begin{aligned}
\max_{\theta_1, \cdots, \theta_m} \mathcal{H}(Y) &= \max_{\theta_1, \cdots, \theta_m} \left(-\int dy\ p_Y \log p_Y \right) = \max_{\theta_1, \cdots, \theta_m} \left(-\int dx\ p_X \log \frac{p_X}{g'(x, \theta)} \right) \\
&= \max_{\theta_1, \cdots, \theta_m} \left(-\int dx\ p_X \log p_X + \int dx\ p_X \log g'(x, \theta) \right) \\
&= \mathcal{H}(X) + \max_{\theta_1, \cdots, \theta_m} \int dx\ p_X \log g'(x, \theta) \qquad (1)
\end{aligned}$$

Let $J = \int dx\ p_X \log g'(x, \theta)$. We may reformulate Eq. (1) as $\max_{\theta_1, \cdots, \theta_m} \mathcal{H}(Y) \equiv \max_{\theta_1, \cdots, \theta_m} J$ since $\mathcal{H}(X)$ is independent of θ.

2.1 K-piecewise Sigmoidal Approximation Network

The transfer function, $g(x, \theta)$, can be modeled in general by a constrained network consists of a linear combination of K-piecewise bounded, monotonic, and invertible logistic functions as follows,

$$g(x, \theta) = \sum_{k=1}^{K} \frac{a_k}{1 + \exp(w_k x + b_k)}, \text{ with } a_k = \frac{\exp(\alpha_k)}{\sum_{l=1}^{K} \exp(\alpha_l)}, \theta = \{w_k, \alpha_k, b_k; k = 1, \cdots, K\} \qquad (2)$$

which ensures that $a_k > 0, k = 1, 2, \cdots, K$ with $\sum_{k=1}^{K} a_k = 1$ for the function to remain monotonic and bounded since the summation of monotonic functions yields a monotonic function. Moreover, these constraints make our model different from the typical feedforward neural networks. By having the linear combination of these K-piecewise monotonic functions, the network has a greater degree of freedom to approximate various CDF functions for entropy maximization of different images.

The network from Eq. (2) has a 1-K-1 architecture with a single input, K hidden neurons in the second layer, and a single output as shown in Fig. 1 (d). Fig. 1 (e) demonstrates the linear combination of logistic functions for $K = 2$.

If $g(x, \theta) = \sum_{k=1}^{K} a_k / (1 + \exp(c_k(w_k x + b_k)))$ then c_k can be viewed as a scaling function which controls the "window" of operation that essentially defines the slope of the transfer function in each of the intervals defined by a_k as shown in Fig. 1 (c). However, it is a redundant variable since by varying w_k and b_k we in essence obtain the same result; hence, we do not use c_k in the

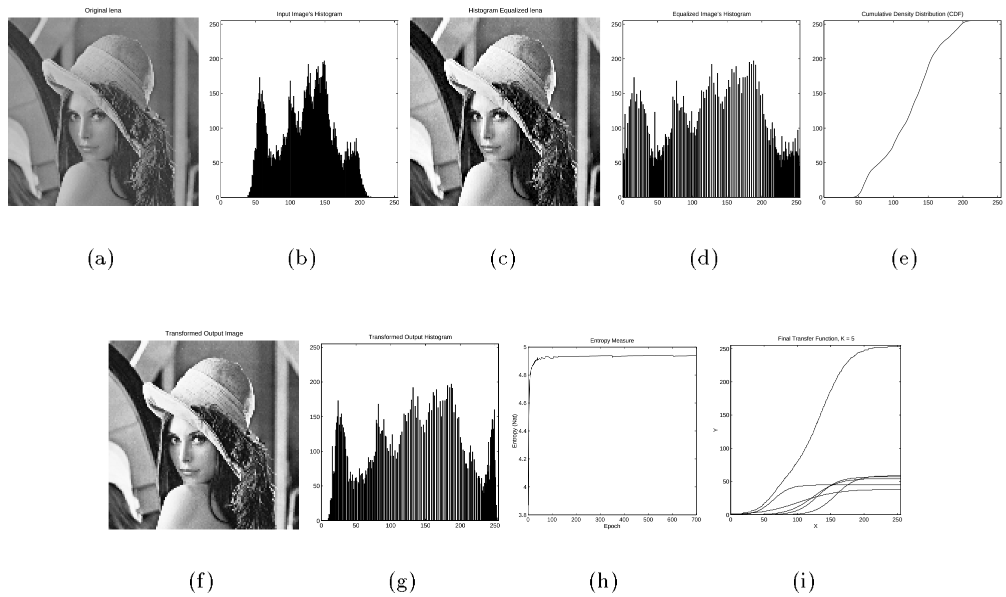


Figure 2: (a) original `lena`, (b) `lena`'s histogram, (c) `lena` histogram equalized, (d) histogram for equalized `lena`, (e) the Cumulative Distribution Function (CDF) function, (f) output image calculated from the 1-5-1 network, (g) entropy vs. learning steps, (h) output image's histogram (i) the transfer function formed by the 1-5-1 network, in comparison with the CDF function of (e).

equation. Both b_k and w_k are parameters which shift the transfer function along its input axis but each performs the shift differently. More specifically, parameter w_k controls the nonlinear shift whereas b_k is the bias which controls the linear shift of the operating range shown in Fig. 1 (a) and (b) respectively. The a_k, in turn, controls the output magnitude of each k'th unit.

With the substitution of $g(\cdot)$, J is written as

$$J = \int dx\ p_X \log g'(x,\theta) = \int dx\ p_X \log\left(-\sum \frac{a_k w_k \exp(w_k x + b_k)}{(1+\exp(w_k x + b_k))^2}\right). \tag{3}$$

We now write the maximization of J using the gradient ascend approach. With the Lebesque's Dominated Convergence Theorem, $\frac{d}{d\theta}\int g(x,\theta)dx = \int \frac{\partial}{\partial\theta} g(x,\theta)dx$, the differentiation outside of the integral can be moved inside,[2] which yields,

$$\frac{\partial J}{\partial \theta} = \int dx\ p_X\ \frac{\partial}{\partial \theta}\ \log\left(-\sum_{k=1}^{K} \frac{a_k w_k \exp(w_k x + b_k)}{(1+\exp(w_k x + b_k))^2}\right). \tag{4}$$

To simplify equations, we let $\Gamma = \frac{1}{\sum_{i=1}^{K} \frac{a_i w_i \exp(w_i x + b_i)}{(1+\exp(w_i x + b_i))^2}}$. From Eq. (4), we obtain

$$\begin{aligned}
\frac{\partial J}{\partial w_k} &= \int dx\ p_X \Gamma \left(\frac{a_k \exp(w_k x + b_k)}{(1+\exp(w_k x + b_k))^2} - \frac{2 a_k w_k x \exp(2 w_k x + 2 b_k)}{(1+\exp(w_k x + b_k))^3} + \frac{a_k w_k x \exp(w_k x + b_k)}{(1+\exp(w_k x + b_k))^2}\right) \quad (5)\\
\frac{\partial J}{\partial \alpha_k} &= \sum_{l=1}^{K} \frac{\partial J}{\partial a_l}\frac{\partial a_l}{\partial \alpha_k}
\end{aligned}$$

[2] The key condition is to show the existence of a dominating function $g'(x,\theta)$ with a finite integral, i.e., not wildly behaved. It can also be seen as a special case of the Leibnitz's rule of differentiation.

Table 1: Initial values used in simulating the 1-5-1 network.

	η_w 1.0e-07	η_α 1.0e-05	η_b 1.0e-05	w	a	b
Values	0.1526	0.1526	0.3052	0.05:0.0125:0.1	0.2	-3:-1.75:-10

$$= \int dx\ p_X \sum_{l=1}^{K} \left(\Gamma\ \frac{w_l \exp(w_l x + b_l)}{(1+\exp(w_l x + b_l)^2}\ a_l(\delta_{k,l} - a_k) \right),\ \text{where } \delta_{k,l} = \begin{cases} 1 & k = l \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$$\frac{\partial J}{\partial b_k} = \int dx\ p_X \Gamma \left(\frac{a_k w_k \exp(w_k x + b_k)(1 - \exp(w_k x + b_k))}{(1+\exp(w_k x + b_k))^3} \right). \tag{7}$$

From gradient equations Eq. (5)-(7), we obtain the following adaptive updating rules:

$$\Delta w_k = \eta_w \Gamma \left(\frac{a_k \exp(w_k x + b_k)}{(1+\exp(w_k x + b_k))^2} - \frac{2 a_k w_k x \exp(2 w_k x + 2 b_k)}{(1+\exp(w_k x + b_k))^3} + \frac{a_k w_k x \exp(w_k x + b_k)}{(1+\exp(w_k x + b_k))^2} \right) \tag{8}$$

$$\Delta \alpha_k = \eta_\alpha \sum_{l=1}^{K} \left(\Gamma\ \frac{w_l \exp(w_l x + b_l)}{(1+\exp(w_k x + b_k))^2}\ a_l(\delta_{k,l} - a_k) \right),\ a_k = \frac{\exp(\alpha_k)}{\sum_{i=1}^{K} \exp(\alpha_i)} \tag{9}$$

$$\Delta b_k = \eta_b \Gamma \left(\frac{a_k w_k \exp(w_k x + b_k)(1 - \exp(w_k x + b_k))}{(1+\exp(w_k x + b_k))^3} \right) \tag{10}$$

3 Simulation Result and Conclusion

The simulation used a 1-5-1 neural network. The iterative updates are calculated according to Eq. (8)-(10). The initial values for the network are listed in Table 1. Using digitized grayscale images, the continuous random variables X and Y are converted into discrete random variables with $n = 256$ states (graylevels). The image we used is 128×128 of `lena` shown in Fig. 2 (a). Its histogram, histogram equalized output, equalized histogram, and the CDF are illustrated in Fig. 2 (b)-(e) respectively. Results are shown in Fig. 2 (f)-(i). Fig. 2 (f) is the transformed output from the network. The quality is similar to (c) which is calculated by the standard histogram equalization technique. The transformed output's histogram in (g) which should be similar to (d) is more uniformly distributed than in (b). Furthermore, Fig. 2 (h) illustrates that the set of adaptive rules does increase the entropy at the output. Fig. 2 (i) displays the final transformation function formed by the 1-5-1 network. The larger diagonal slope from the bottom-left to the top-right is similar to the one calculated for `lena` shown in (e).

In this paper, we have derived a set of adaptive rules and constructed a 1-K-1 constrained network for histogram equalization in image enhancement using the Maximum Entropy Principle. The results demonstrate the network's ability to approximate the CDF function achieving contrast enhancement, good utilization of the dynamic range, and maximum information transfer.

References

Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? *Network: Computation in Neural Systems*, 3:213–251.

Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, 1(3):295–311.

Becker, S. and Hinton, G. E. (1992). A self-organising neural network that discovers surfaces in random-dot stereograms. *Nature*, 355:161–163.

Gonzalez, R. and Woods, R. E. (1992). *Digital Image Processing*. Addison-Wesley, Reading, MA, USA.

Linsker, R. (1988). Self-organization in a perceptual network. *IEEE Computer*, 21(3):105–117.

Linsker, R. (1989). How to generate ordered maps by maximizing the mutual information between input and output signals. *Neural Computation*, 1(3):402–411.