Contrast-Limited Adaptive Histogram Equalization: Speed and Effectiveness

Stephen M. Pizer, R. Eugene Johnston, James P. Ericksen,
Bonnie C. Yankaskas, Keith E. Muller
Medical Image Display Research Group
University of North Carolina
Chapel Hill, NC 27599-3175, USA

ABSTRACT

The Contrast-Limited Adaptive Histogram Equalization (CLAHE) method for assigning displayed intensity levels in medical images, is supported by anecdotal evidence and evidence from detection experiments. Despite that, the method requires clinical evaluation and implementation achieving few-second application before it can be clinically adopted. Experiments attempting to produce this evaluation and a machine providing the required performance are described.

Introduction

Contrast-Limited Adaptive Histogram Equalization (CLAHE) is a method that has shown itself to be useful in assigning displayed intensity levels in medical images. The method is designed to allow the observer to easily see, in a single image, all contrast of clinical or research interest [Pizer, 1987]. The method examines a histogram of intensities in a contextual region centered at each pixel and sets the displayed intensity at the pixel as the rank of that pixel's intensity in its histogram. That histogram is a modified form of the ordinary histogram in which the contrast enhancement induced by the method at each intensity level is limited to a user-selectable maximum. Anecdotal evidence has shown CLAHE to be useful in viewing a wide variety of medical images (see figure 1, for example).

In the first section of this paper we report on an experiment intended to evaluate the clinical application of CLAHE to chest CT images. In the second section we report on MAHEM, a machine to compute CLAHE in a few seconds.

Effectiveness of CLAHE in Diagnostic CT

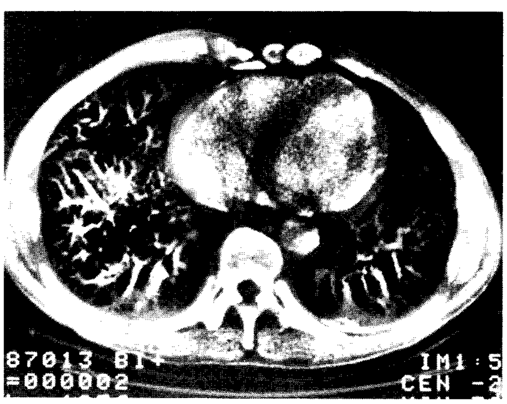
In various observer studies CLAHE has been shown to allow the detection of contrast changes as effectively as interactive intensity windowing. More precisely, Zimmerman [1989] showed by ROC studies on CT chest studies that CLAHE with moderate limitation of contrast enhancement allows as effective detection of simulated Gaussian lesions as CLAHE with no such limitation (AHE). In [Zimmerman, 1988] he reported that AHE allows as effective detection of these lesions in the lung as interactive intensity windowing and almost as effective detection in the mediastinum. Because CLAHE handles the mediastinum better with limitation of contrast enhancement than without and preset intensity windowing is less capable than interactive intensity windowing, it is reasonable to conclude that CLAHE is at least as good as preset intensity windowing for detection of small intensity increases.

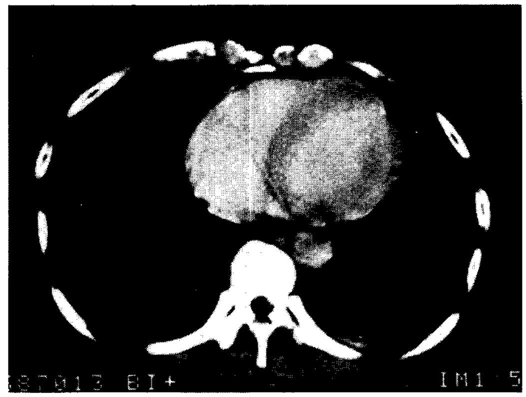
The remaining question is whether in clinical diagnosis, where not simply detection, but shape determinations, comparisons among regions, localizations, and other judgments are required, CLAHE communicates the information in the recorded intensities as effectively as intensity windowing. We chose to ask this question with regard to chest CT, because CLAHE would have real advantages in allowing a single image to replace the two normally used in clinical diagnosis, one with a lung window and one with a mediastinum window.

Agreement Experiment

Because CLAHE has advantages other than diagnostic quality, we simply wanted to demonstrate whether equivalent information could be obtained from CLAHEd images, as compared to images displayed according to the present clinical standard (see figure 1 for a single image from a single case displayed using CLAHE and each of the intensity windows). As a result an experimental paradigm was designed whereby knowledge of the correct diagnosis was unnecessary. Rather we aimed to measure whether a radiologist reading CLAHEd images (new) would agree equally well on diagnostic findings with a radiologist reading the intensity windowed (standard) pair as two radiologists would agree with each other if they were both reading an intensity windowed pair.

Agreement equality is defined in terms of what would be an acceptable difference of agreements. The first agreement compares performances using the new method to that with the standard method.





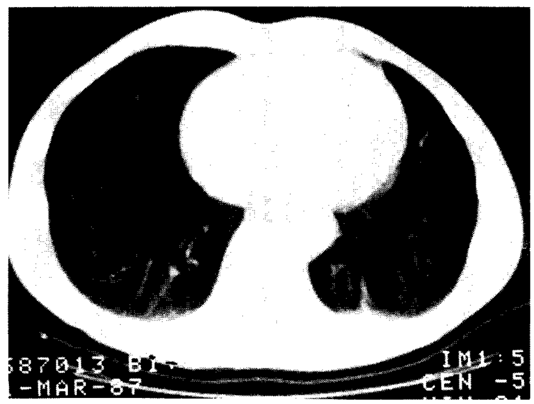


Figure 1. A single chest CT image from our study, first in CLAHE'd form and then with the two intensity windows.

The second agreement compares performances of two observers both using the standard method. This method involves two radiologist observers, both well trained to use the new and standard display approach so that the observers are operating with the same value system on both the new and standard display.

The complete set of study material (here chest CT cases) is divided into three groups, one containing one-half of the patient studies, which become designated as the *baseline*. Those images are read by both radiologists using the standard display approach. The resulting agreement measure serves as a baseline, measuring of the expected agreement of the observers when operating under the same conditions and viewing the same material. The remaining half forms the *comparison set*. The images in the comparison set are divided into two parts. One part is read by observer A with images using the new approach, and the same studies are read by observer B with images using the standard approach. For the other part the type of images given to each observer is opposite from that in the first part. This strategy avoids bias toward the special skill of an observer for a particular display approach and having observers reading an image twice. The data from all of these studies are then pooled to provide a measure of (new/standard) agreement. A more complete description of the agreement method is presented in [Johnston, 1990].

We conducted an agreement experiment as described above with a baseline of 40 chest CT cases and a comparison set of 40 cases. In the baseline set each slice appeared according to both preset intensity windows. The studies in the comparison set appeared in that form to one radiologist but as CLAHEd to the other. Each case consisted of the full set of 25-40 slice images making up a particular clinical CT chest study. In each case the radiologist filled out a clinical findings form listing 25 different "findings" for which the radiologist must assign a score. A 1 through 5 rating scale was used, where 1 meant "normal", 2 meant "probably normal", 3 meant "possibly abnormal", 4 meant "probably abnormal", and 5 meant "abnormal". Figure 2 shows a subset of the findings list.

Aware that new treatments of images required the radiologist to be well trained on the appearance of the CLAHEd images relative to the images they were accustomed to reading, we carried out the following training before the experiment commenced. Six cases, not part of the study set, were chosen that represented normal and a range of abnormal categories of findings. The radiologists carefully examined the CLAHEd images and scored their findings lists. Then they reviewed the same images with the standard intensity windows and compared their reading with the CLAHEd reading. Finally they compared both readings to a findings list previously generated from the intensity windowed images. After applying this process to the six CT cases, both radiologists indicated that they understood the CLAHEd images and were ready to undertake the study.

sternum	________
thoracic soft tissue	________
vertebrae	________
pleural surfaces	________
pleural space	________
mediastinum	________
aorta	________
lung parenchyma	________
RUL	________
RML	________
RLL	________
LUL	________
LLL	________
pericardium	________

Figure 2. Example of clinical findings categories for chest CT.

The cases in the study were read over a period of one or two sessions. For Dr. D. this period fell 2-3 weeks after the training session, and for Dr. P. this period fell 4-5 weeks after the training session. 3 weeks after his reading was complete, Dr. P. reread the CLAHE'd images that he had read in that form from the comparison set.

Results and Conclusions

Our findings are that the CLAHEing led to a change in the radiologists' use of the five categories of abnormality and normality. The data are given in figure 3. Because table entries are based on 20-40 cases, the standard error of such a proportion is approximately .08-.11. Hence it is unlikely that differences smaller than .2 are reliable. Nonindependence of observations complicates many comparisons in the figure. It can be observed that for intensity windowing readings, the two subjects used the categories with about the same frequencies, both for the cases that they read in common and for all the intensity windowed cases read by each. On the other hand, CLAHEing caused Dr. P. to decrease by 10% the number of cases that he called normal or probably normal and to increase those called abnormal or probably abnormal by 3%. In reverse, CLAHEing caused Dr. D. to *in*crease by 10% the number of cases that he called normal or probably normal and to *de*crease those called abnormal or probably abnormal by 16%. Furthermore, the confidence of both observers in their calls on CLAHEd images decreased as compared to their calls on intensity windowed images: CLAHE caused the use of the categories normal and abnormal by both subjects under CLAHE to be clearly decreased relative to the use of probably normal and probably abnormal.

Relative Response Frequency of Rating

	Baseline(IW)		All IW		CLAHE		Repeat CLAHE
	Dr. P	Dr. D	Dr. P	Dr.D	Dr.P	Dr.D	Dr.P
Rating							
1	0.46	0.42	0.46	0.44	0.24	0.35	0.14
2	0.31	0.32	0.27	0.26	0.43	0.34	0.47
3	0.04	0.03	0.04	0.03	0.10	0.08	0.14
4	0.02	0.07	0.04	0.08	0.02	0.07	0.12
5	0.17	0.16	0.20	0.20	0.15	0.05	0.13
N*	0.77	0.74	0.73	0.70	0.67	0.84	0.61
A*	0.19	0.23	0.24	0.28	0.22	0.07	0.25
# of Cases	40	40	20	20	20	20	20

*N is the sum of rating 1 and 2, A is the sum of rating 4 and 5

Figure 3. Use of rating categories by subjects.

The result of these changes in the use of the categories is that the agreement statistics for this experiment cannot answer whether the information gleaned via CLAHE is the same as that gleaned from intensity windowing. In order to obtain this information, we shall need to repeat the experiment with an additional training session with feedback to the observers until they learn to use the categories in the same proportions as they do with intensity windowing. A subsequent run of the agreement experiment will then yield an answer to our original question.

This experience confirms the likelihood of image processing to change viewers' reading behavior, independent of any change in information available in the images in the processed form. Experimental paradigms that depend on no change must assure recalibration of the observers before the experiment is carried out.

Interviews of our subjects revealed two additional properties that need experimental verification. First, it appears that CLAHEd images convey at least as much clinically relevant structural information as the intensity-windowed images. Second, it appears that some clinically relevant information given by absolute CT number levels is conveyed in the intensity-windowed images but lost in CLAHEing.

An Engine for Fast CLAHE

As a display algorithm, CLAHE needs to operate in a few seconds. However, on a computer of only a few MIPS the method can require 1-2 hours unless approximations based on spatial sampling and interpolation of the mapping are used. These approximations need to be avoided because of the artifacts they can produce.

We will describe the structure of a Multiprocessor AHE Machine (MAHEM) that we have designed and built to provide speedy CLAHE. We will also present the algorithm on which it is based. MAHEM can apply CLAHE to a 512 x 512 image in four seconds (and significantly longer on larger images) and allows user control of both the contextual region size and the limit of maximum contrast enhancement. MAHEM is built from off-the-shelf components costing $16,000, including printed circuit board fabrication. It is based on the simultaneous calculation of the effect of any input image pixel on n output pixels by the use of n simple processors in parallel. For the present machine, n=64.

MAHEM Description

MAHEM is a self-contained image processing machine, built on a triple-height VME bus chassis. Images and commands are sent to and from a host microVAX computer through a parallel interface with a bandwidth of approximately 500K bytes/second. MAHEM is built from readily available parts and technologies: the pixel processors are implemented with standard TTL logic, and a microprocessor handles distribution of pixels to individual image processors and performs pre- and post-processing of contrast-limited pixel values.

A complete MAHEM system includes one controller board and several pixel processor boards. The controller board contains the microprocessor, the interface to the host computer, and memory for input, intermediate, and output images. Each pixel processor board contains 16 individual pixel processors. In MAHEM's version of the CLAHE algorithm the only operations required of the processors are comparison and addition. A minimum MAHEM system (and the prototype) features four such boards, for a total of 64 pixel processors. The system design allows expandability and upgradability; the addition of more memory and additional pixel processors allows larger images to be processed more rapidly.

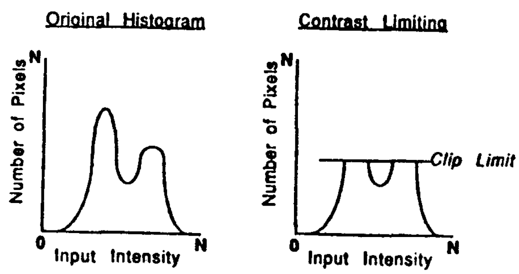
Loading a 512 x 512 16-bit image from the host computer to MAHEM takes about five seconds, in addition to the time for MAHEM to process the image. This processing time is 4 seconds for the first processing of an image and 2 seconds for each new value of the clip limit parameter. MAHEM's two parameters can be set and altered through commands sent to the host computer. The enhanced image may be displayed on a video monitor monitor connected to MAHEM. Additionally, users at remote sites can use ethernet to send images to and from MAHEM via the host computer, thus allowing MAHEM to serve as a high-speed CLAHE contrast enhancement server.

MAHEM Algorithm for CLAHE

All histogram equalization methods attempt to best use the range of available display levels by distributing pixels evenly among them, "flattening" the histogram. Thus, the degree of enhancement is sensitive to the distribution of recorded intensities, i.e., is proportional to the maximum height of the histogram. To assure that this sensitivity to the intensity distribution is limited to visually relevant intensities at any location of interest, the adaptive methods (AHE, CLAHE) use only the pixels within the contextual region of a pixel in calculating the rank of that pixel. For a 512 x 512 image, this contextual region is typically a 64 x 64 square with the pixel being processed (the *affected pixel*) at the center; all pixels in the contextual region are considered *affecting pixels*. For each position of the contextual region, only the center pixel's output value is computed.

In CLAHE the output value for a pixel is its rank in a histogram of pixel intensity values in the contextual region; this is the same as counting the number of pixels in the contextual region whose intensities are less than the affected pixel. The histogram is the actual histogram of recorded intensities centered at the pixel in question, but clipped at a particular height with the clipped pixels redistributed uniformly across all intensities in the range of the recorded image (see figure 4). The effect of the clipping is to lessen the enhancement of noise in relatively homogeneous areas of the image by varying the maximum possible level of contrast enhancement.

In MAHEM [Ericksen, 1990] the clipping effect is achieved by letting the contribution of each pixel to the rank of



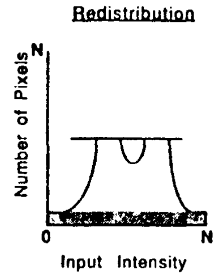


Figure 4. Histogram modification in CLAHE

another be unity if nothing is to be clipped at the contributing pixel's intensity and a fraction otherwise, with the fraction chosen so that the contribution of all the pixels at that intensity sums up to the clip level. The complement of each fraction is counted toward the pool of redistributed pixels. Moreover, the effect of redistribution can be shown to be taking the output image as a weight times the original image plus the image produced from a histogram with clipping but without redistribution. The weight is proportional to the total in the pool of redistributed pixels.

The fractional effect of an affecting pixel P is a function of the number of pixels with the same intensity as P. Normally, the bin height, and therefore the increment value for P, is dependent on the contextual region being examined. To avoid having to recompute each affecting pixel's increment value for each affected pixel, MAHEM considers only the affecting pixel's own contextual region in computing that pixel's increment value, and uses that increment value in computing the output value for every pixel which is affected by P.

MAHEM performs CLAHE in two distinct passes: the first "preprocessing" pass which counts the number of pixels in the contextual region which are equal to the center pixel (used to calculate the increment values for the contrast limitation), and the second pass in which the ranks are computed. The center, or affected, pixels in the first pass become the affecting pixels in the second pass.

```
/* CLAHE Algorithm (MAHEM variation) */
/* first pass: count equal pixels */
for each (x,y) in image do
{
   eqcount[x,y] = 0
    for each (i,j) in contextual region of (x,y) do
      if image[x,y] == image[i,j] then
          eqcount[x,y] = eqcount[x,y] + 1
}

  /* second pass: calculate partial rank, redistributed  area,
    and output values */
for each (x,y) in image do
{
    cliptotal = 0
    partialrank = 0
    for each (i,j) in contextual region of (x,y) do
   {
      if eqcount[i,j] > CLIPLIMIT then
          incr = CLIPLIMIT / eqcount[i,j]
      else
       incr = 1
      cliptotal = cliptotal + (1 - incr)
      if image[x,y] > image[i,j] then
           partialrank = partialrank + incr[i,j]
   }
    redistr = (cliptotal / CONTEXTAREA) * image[x,y]
    output[x,y] = partialrank + redistr
}
```

Multiprocessing In MAHEM

Rather than using multiple processors in parallel to compute a pixel's intensity (a typical local-region image processing approach), MAHEM has each processor computing one pixel's intensity at any given time; multiple processors are computing multiple pixels simultaneously. The central idea to making this work is that each affecting pixel affects (is in the contextual region of) many affected pixels; these affecting pixels are broadcast to all processors simultaneously, each of which is computing the value for one of the pixels affected by the broadcast (affecting) pixel.

Memory bandwidth limitations and economics make it impractical to build a machine in which an entire contextual region is processed in parallel. We have found it efficient and feasible to provide enough pixel processors to process one row, or one column, of the contextual region. The image buffers are

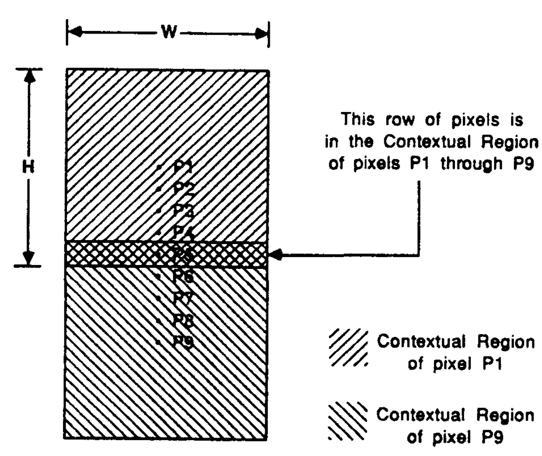


Figure 5. Pixel Processor Access to Pixels.

implemented with VRAM chips which allow rapid access to a horizontal row of sequential pixels. Suppose the contextual region is W pixels wide by H pixels high. Then there are H contextual regions that contain any given row of W pixels (see figure 5). The row of affecting pixels can therefore be broadcast to H pixel processors which simultaneously are computing the result value for H (vertically adjacent) pixels.

MAHEM can be programmed for any arbitrary contextual region size, up to 128 x 128. With 64 pixel processors, a contextual region size greater than 64 x 64 will require that the image be processed in multiple passes; however, additional pixel processors can easily be added to the machine.

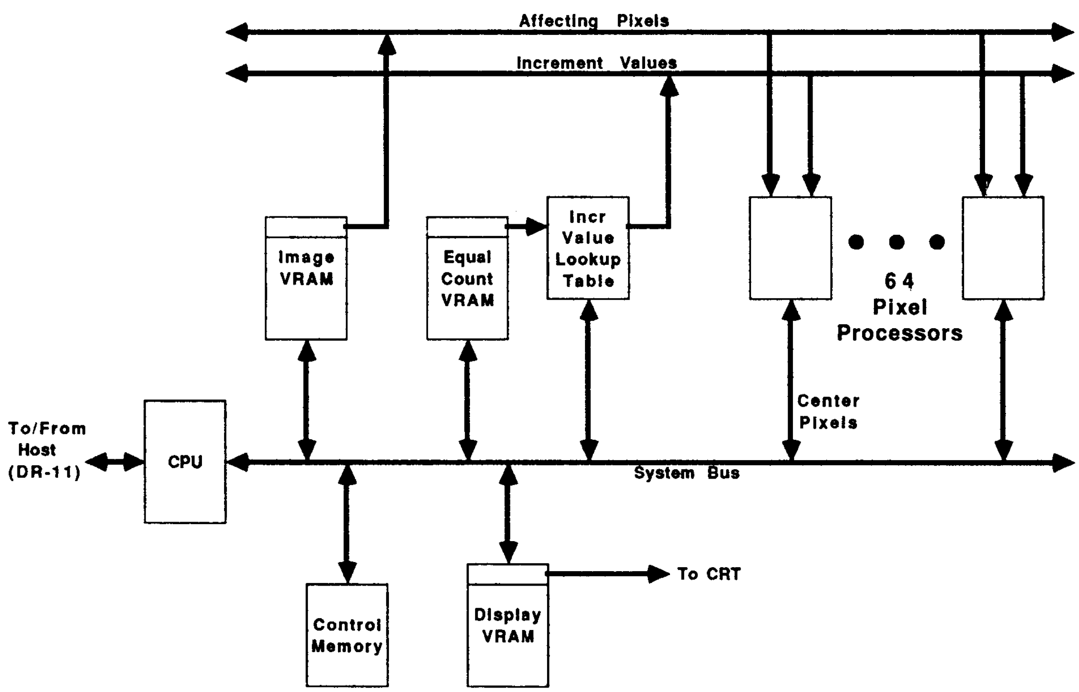


Figure 6. MAHEM system block diagram.

Implementation

The block diagram of MAHEM is given in figure 6. Details of this architecture and its realization are described in [Ericksen, 1990]. A summary of a few of the details of the implementation is given in the following.

A 16-bit original image buffer provides the affecting pixel values, which are broadcast to the pixel processors, The equal counts computed in pass 1 are then scaled to 8 bits and stored in a second buffer. The final results are stored in the 8-bit display memory buffer, which drives the video output and is also available for reading by the host computer.

The basic components of each pixel processor are a register for the center pixel (the pixel being processed), a comparator, and two "fractional accumulators" A and B, each consisting of an adder and a register.

The Increment Lookup Table is a high-speed RAM that translates equal count values to increment values while the affecting pixel data is shifted out. The increment is a 3-bit fixed-point fraction plus one overflow bit, giving a range of 0.0 to 1.0 in steps of 0.125.

Conclusions

CLAHE can be computed in 4 seconds after a 5-second loading time using a specially designed parallel engine made from a few thousand dollars of off-the-shelf components. The processing appears to be useful for a wide range of medical images, but the limitations of observer calibration have made it impossible to demonstrate such usefulness by agreement experiments. Concerns as to the loss of absolute intensity information by CLAHE may limit its usefulness for certain clinical problems.

Acknowledgements

We are most grateful to Drs. David Delany and Leonard Parker for their invaluable service as observers. We gratefully acknowledge the help of John Austin, who created the first MAHEM design; Philip Stancil, Vernon Chi, and the staff of the Microelectronics Systems Laboratory at UNC Computer Science for help with building MAHEM; John Zimmerman, whose work with AHE evaluation has been important; and Dr. Julian Rosenman for his support and advice. We thank Jeffrey Reese for data analysis, Bo Strain for photography, and Carolyn Din for help with manuscript preparation. This research has been partially supported by NIH grants numbers R01 CA44060 and R01 CA47982.

REFERENCES

[1] Ericksen, JP, SM Pizer, JD Austin, "MAHEM: a Multiprocessor Engine for Fast Contrast-Limited Adaptive Histogram Equalization", *Medical Imaging IV: Image Processing*, Volume 1233, SPIE, Bellingham, WA, 1990.

[2] Johnston, RE, B Yankaskas, JR Perry, SM Pizer, DL Delany, and LA Parker, "Agreement Experiments: A Method For Quantitatively Testing The Equivalence of New Medical Image Display Approaches," *Medical Imaging IV: Image Capture and Display*, Volume 1232, SPIE, Bellingham, WA, 1990.

[3] Pizer, SM, EP Amburn, JD Austin, R Cromartie, A Geselowitz, T Greer, BM ter Haar Romeny, JB Zimmerman, K Zuiderveld, "Adaptive Histogram Equalization and Its Variations," *Comp. Vis., Graphics, & Im. Proc.* **39**:355-368, 1987.

[4] Zimmerman, JB, SM Pizer, EV Staab, JR Perry, W McCartney, BC Brenton, "An Evaluation of the Effectiveness of Adaptive Histogram Equalization for Contrast Enhancement", *IEEE Trans. Med. Imaging*, **7**(4): 304-312, 1988.

[5] Zimmerman, JB, SB Cousins, ME Frisse, KM Hartzell, MG Kahn, "A Psychophysical Comparison of Two Methods for Adaptive Histogram Equalization", *J. Dig. Im.*, **2**(3), 1989.