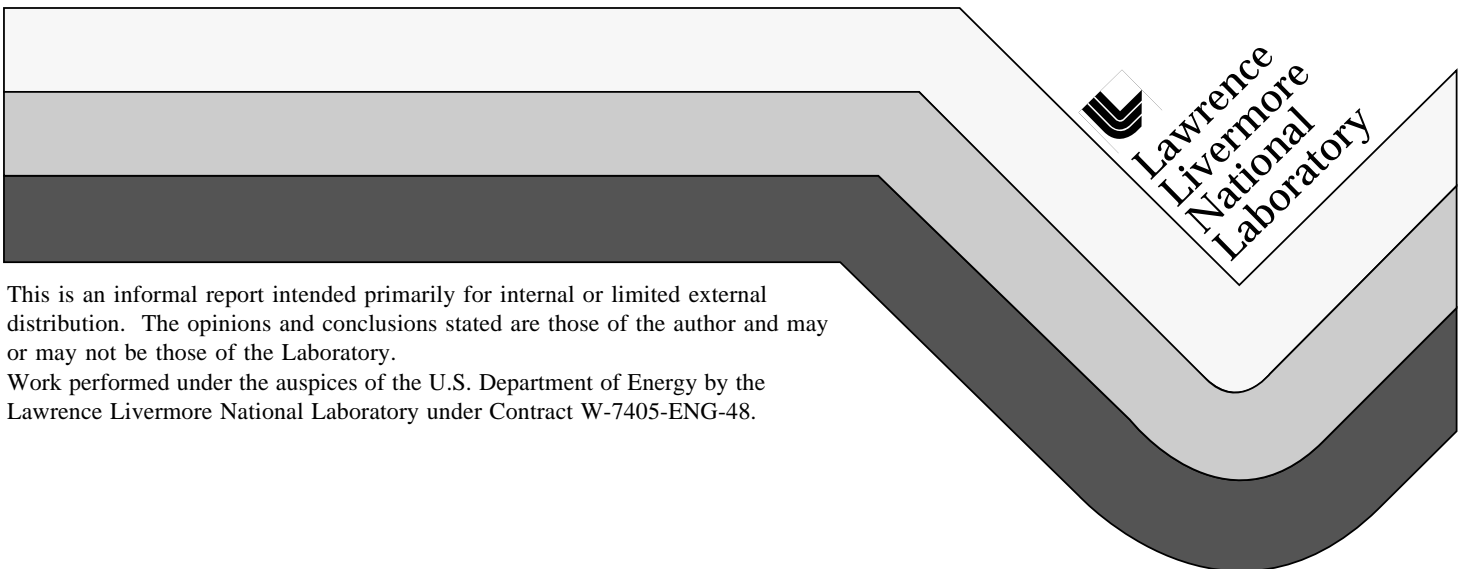


# Image and Video Compression/Decompression based on Human Visual Perception System and Transform Coding

Chi Yung Fu  
L. I. Petrich  
M. Lee

February 1997



This is an informal report intended primarily for internal or limited external distribution. The opinions and conclusions stated are those of the author and may or may not be those of the Laboratory.

Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under Contract W-7405-ENG-48.

## DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This report has been reproduced  
directly from the best available copy.

Available to DOE and DOE contractors from the  
Office of Scientific and Technical Information  
P.O. Box 62, Oak Ridge, TN 37831  
Prices available from (615) 576-8401, FTS 626-8401

Available to the public from the  
National Technical Information Service  
U.S. Department of Commerce  
5285 Port Royal Rd.,  
Springfield, VA 22161

# Image and Video Compression/Decompression based on Human Visual Perception System and Transform Coding

*Chi Yung Fu, L. I. Petrich, and M. Lee*

Lawrence Livermore National Laboratory  
(510) 423-1175  
fu1@llnl.gov

## ABSTRACT

The quantity of information has been growing exponentially, and the form and mix of information have been shifting into the image and video areas. However, neither the storage media nor the available bandwidth can accommodate the vastly expanding requirements for image information. A 24-bit color picture measuring 8.5 in. by 11 in. at 300 dots per inch will generate 25 Mbytes of data. Just 40 of these images will fill up a 1-Gbyte hard disk. Storage is not the only problem here; such huge files can take a long time to load into a program or save onto a disk. Moving this amount of information over a network or a broadcast channel is an enormous task. Transmission of these images over commonly used transmission channels will be extremely costly, especially for video, because of the large amount of data involved. Even with further network advances, it is expected that growing needs will outstrip the expected increase in bandwidth, and thus information congestion is still expected.

A vital, enabling technology here is compression/decompression. The expected payoffs are in the storage and transmission of information; they will affect projects ranging from automatic target recognition, such as the U.S. Army's Takeall system, to visualization of scientific and engineering data, such as those derived from Advanced Design and Production Technology (ADaPT), the National Ignition Facility (NIF), and the Accelerated Strategic Computing Initiative (ASCI).

Our compression work is based on a combination of feature-based algorithms inspired by the human visual-perception system (HVS), and some transform-based algorithms, (such as our enhanced discrete cosine transforms, wavelet transforms,) vector quantization, and neural networks. All our work was done on desktop workstations using the C++ programming language and commercially available software.

During FY 1996, we explored and implemented an enhanced feature-based algorithms, vector quantization, and neural-network-based compression technologies. For example, we improved the feature compression for our feature-based algorithms by a factor of two to ten, a substantial improvement. We also found some promising results when using neural networks and applying them to some video sequences. In addition, we also investigated

objective measures to characterize compression results, because traditional means such as the peak signal-to-noise ratio (PSNR) are not adequate to fully characterize the results, since such measures do not take into account the details of human visual perception.

We have successfully used our one-year LDRD funding as seed money to explore new research ideas and concepts, the results of this work have led us to obtain external funding from the DoD. At this point, we are seeking matching funds from DOE to match the DoD funding so that we can bring such technologies into fruition.

## INTRODUCTION

Information has been growing exponentially, and the form and mix of information has been shifting into the image and video areas. But neither the storage media nor the available bandwidth can accommodate the vastly expanding image information requirements. A 24-bit color picture measuring 8.5" by 11" at 300 dots per inch will generate 25 megabytes of data. Just forty of these images will fill up a 1-gigabyte hard disk, a storage capacity that is a minimal for scientific and engineering computation but exceeds most of those used in home and corporate computers. Storage is not the only problem here; such huge files can take a long time to load into a program or save into a disk. Moving this amount of information over a network or sending it through a broadcast channel is an enormous task. Transmission of these images over commonly-used transmission channels will be extremely costly, especially for video, because of the large amount of data involved. Even with future network advances, it is expected that growing needs will outstrip the expected increase in bandwidth, thus creating information congestion; the recent breakdowns in America Online's system may only be the beginning. A vital enabling technology is compression/decompression. As an example of a very common bandwidth limitation, let us consider a modem connection. The current maximum bandwidth of analog modems for telephone lines is 28.8 kilobaud (28,800 bits/s); a value greater than what many users now have. At this data rate, it takes 4 minutes to transfer just *one* uncompressed relatively low resolution 640\*480 24-bit-color image! However if the image was compressed 100:1, the transmission time goes down to a tiny 2.6 seconds. This narrowness of bandwidth is one of the major barriers in the digitalization of the battlefield and other such DoD imaging problems. The expected payoffs are for both the storage and transmission of information, impacting projects ranging from automatic target recognition, such as the U.S. Army's Takeall system, to scientific and engineering data visualization such as those derived from the Advanced Design and Production Technology (ADaPT), the National Ignition Facility (NIF), and the Accelerated Strategic Computing Initiative (ASCI) programs.

## ALGORITHM DEVELOPMENT and RESULTS

### Feature Representation, Vector Quantization and Huffman Coding Algorithms

The reason for developing these algorithms is that we need to focus on the importance of features which our human visual system seems to emphasize. Also, finding features is extremely important for pattern recognition such as in the case of target recognition and identification. In addition to finding features, we also need highly efficient algorithms to represent these features with a minimum number of bytes in order to yield a high compression ratio.

The finding of features was done with the help of filters applied to all color planes. This was done so as to make the maximum possible use of the available information in finding features. The result of this step is a set of feature values. From these feature values, we then attempt to group them together to efficiently represent the features of an image.

For vector quantization, our implemented technique is based on an unbalanced binary tree of values; the purpose of doing so is to reduce the time necessary to find the closest member of that table from the table's size to the logarithm of that size. In addition, the construction of this binary tree could be cut off if the fit errors were sufficiently low, thus improving compression efficiency as well as speed. This resulting quantization may then be compressed by Huffman coding.

Huffman coding is a well-known and highly-efficient technique for context-less symbols; this was used for further compression of the features. Though arithmetic coding is probably more efficient, it is protected by certain IBM patents. The information was encoded differentially; what was encoded was the difference from the previous value, since fewer bits are needed to specify a small number of symbols than a large one, as is the case here. In practice, however, only the signs and the exponents were Huffman-coded; the fractional parts were inserted unchanged directly into the coded data stream. This keeps down the size of the "Huffman table" of information necessary for setting up for doing Huffman decoding.

So far, we have tested the effectiveness of these algorithms in comparison with our old methods using simple representations of features such as lines and circles since all features essentially can be distilled into these basic and fundamental features. Figure 1 shows simple circles in grayscale; Figure 2 shows simple lines in grayscale, and finally Figure 3 represents a double color spiral. Table 1 shows that the improvement in feature compression efficiency can be substantial, ranging from 2 to about 10 times.

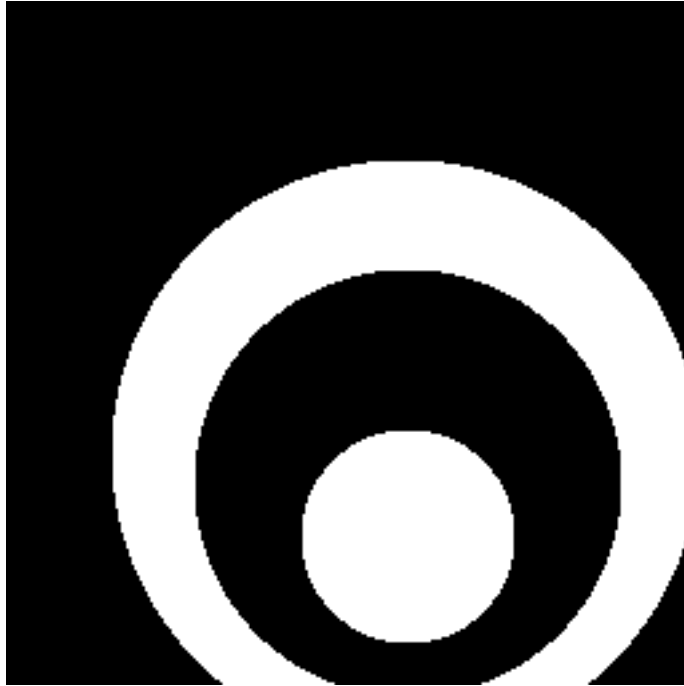


Figure 1 - "Circles" from the University of Waterloo Bragzone web-site

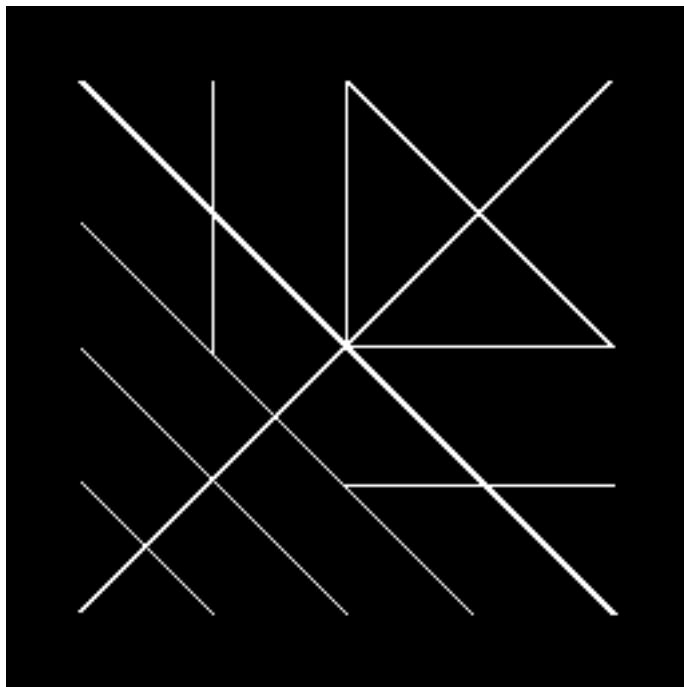


Figure 2 - "Crosses" from the University of Waterloo Bragzone web-site



Figure 3 - Pair of logarithmic color spirals

**Table 1: Improvement based on the New Algorithms**

	Circles	Crosses	Log Spirals
# Pixels	65536	65536	262144
# Compress Bytes (Old method)	1000	2727	19296
# Bits/Pixel	0.122	0.333	0.589
# Compress Bytes (New Algorithms)	454	279	3215
# Bits/Pixel	0.055	0.034	0.098
<b>Improvement</b>	<b>2.20</b>	<b>9.77</b>	<b>6.00</b>

As these new algorithms are targeted for compressing high-resolution images, they are not appropriate for low-resolution ones where the number of pixels within each image is small and every single pixel is important; thus we attempt to compress this class of images with totally different techniques. This class of image is common for a very important application — IR images — which tend to have much lower resolutions than the more typical sort of images. Of the different approaches we have explored, a neural-network-based technique seems to be promising at this stage.

### Objective Picture Quality Metrics

Since for many applications, a human being is either the ultimate receiver of the compressed/decompressed images and videos or is involved in the image evaluation and distribution processes, subjective evaluation is

therefore a very important measure of the results of compression/decompression. This is important because traditional objective measures such as peak signal-to-noise ratio (PSNR) are inadequate to fully characterize the results, especially since measures like this one do not take into account human-perception features.

However, subjective measurement suffers from experimental difficulties and inaccuracies. It is prone to personal biases and ambiguities, which may vary with the subjects and the test conditions. In addition, it is time-consuming, inconvenient, and potentially very costly. Using subjective measurements to evaluate the results of different settings of the same compression process or algorithm to optimize its settings can be difficult for the above reasons. Similarly, it is hard to compare the results of one compression algorithm against another to determine which one is superior. Traditional objective measurements such as the PSNR can only provide rough guidance, because one can easily show that an image compressed with a higher PSNR can appear to be inferior in subjective judgment when compared with the same image compressed with a lower PSNR with some different technique or parameter set. Furthermore, a single number simply cannot be used to fully characterize the entire image, since an image is subjected to many different aspects of human evaluation.

Though the establishment of objective measures in the compression community is still in its infancy and thus at this point these measures cannot replace subjective evaluation, they do, however, provide another perspective to quantify the compressed/decompressed results. We have decided on eight different objective picture-quality metrics ranging from descriptions based on random errors to structural errors; from statistical to biological considerations, and from global to algorithm-specific errors. Some of these metrics take into account the human visual perception. These eight metrics include the CCITT formula for image error, the Weber-Fechner weighted measure, errors at the block boundaries, locally-correlated errors, errors near edges, variances of the different images, errors pooled from channels produced by filters with different spatial frequency, orientation-finding analogous to that in biological vision, and finally the use of image-parsing trajectories to control spatial-error distribution. These measurements of distortion are performed at the pixel level. From these pixel-level distortions, one can then construct a single number for the entire image for each particular type of distortion. Since there are eight metrics, therefore there will be eight of these numbers. It is evident that such an approach is more comprehensive than simply using the PSNR. We are exploring how we can intelligently combine these numbers to construct a single overall number that can replace the PSNR, while being more meaningful than the PSNR or the weighted mean-square error (WMSE). We are currently working on this task, and we believe that these metrics will be indispensable tools for us to gauge our progress, as well as that of our colleagues in the compression community.



## Compression of the MICOM "Tank" Video

We have been working with the U.S. Army Missile Command Center (MICOM) to assist in the image-compression part of the "Takeall" project, whose aim is automated target recognition with as little human assistance as possible. Our goals are twofold. First, we want to demonstrate our technology in real-world applications. Second, understanding that LDRD funding is only seed money, we need to secure funding for the future.

The performance of image-compression technologies is often very image-dependent; there is good reason to believe that there is no single compression algorithm that can be effectively used on all images. The compression technologies we have been working on, such as those described above, have been aimed at processing high-resolution and color images; thus, we need to adapt our current technologies as well as come up with new compression techniques to compress the very low-resolution and grayscale images from MICOM. MICOM also specifies that each image from the video sequence is preferably compressed individually and independently, and thus the sort of high-compression-ratio technology that uses frame-to-frame redundancy such as the Motion Picture Experts Group one (MPEG) cannot be applied. We have been given an IR image sequence of a moving tank containing 200 frames of 12-bit 40x56 pixel images. This extremely low-resolution and noisy sequence represents a serious compression challenge. MICOM has already accumulated a total of 21 different compression algorithms from industry, universities, and other research institutes in order to evaluate these algorithms for their needs. Currently, MICOM is aiming at a 10:1 compression ratio with excellent image quality for these images. This apparently low requirement is in fact hard to achieve, since the images are both small and grayscale, meaning that easy-to-compress flat areas will be relatively small and also meaning that one cannot take the shortcut of compressing some of the color information with greater loss than some of the other color information. This is because a common shortcut in color-image compression is to compress the chroma (hue + saturation) with greater loss than the intensity, because our eyes have more sensitivity to changes in the latter than in the former.

Though our current compression technologies do yield reasonable results when applied to the "tank" sequence of images from MICOM, we believe that they are not appropriate for this class of images. One possible direction is to take a totally different approach to compression — instead of traditional transform-based processing such as JPEG or our proposed human-perception-based processing, as we have successfully demonstrated for compressing high-resolution images, we view compression as a modeling process or problem. Let us take a simple example — a circle. It will not be very efficiently compressed with JPEG and it will be more efficiently compressed with our feature-based compression algorithms. However, if we look at the circle as something to model, then one can easily conclude that the most efficient method is to describe the circle by the following equation.

$$(x-x_0)^2 + (y-y_0)^2 = R^2$$

Then, not only can we describe the image by using three numbers,  $x_0$ ,  $y_0$ , and  $R$ ; but we can also describe many different circles by using different values for these three numbers.

However, exact and simple closed-form mathematical modeling is very unlikely for real-life images, since they are far more complex than could be described by such a compact analytical model. The next best thing is to use empirical modeling, an area we have had experience and success in, one example being modeling such complicated phenomena as plasma processing using artificial neural networks (NN). The advanced nonlinear modeling capabilities of artificial neural networks are essential for modeling complex behavior, such as that of an image. This modeling approach represents a very different paradigm for looking at the process of compression.

### **JPEG vs. Neural Network based Compression for the "Tank" Sequence**

One characteristic of images is that the information in adjacent pixels is usually similar in value. Block-based techniques, such as JPEG, attempt to exploit this property. In a block-based technique, the image is described by blocks which contain some small number of pixels, i.e., 8\*8 pixel blocks. Though grouping into blocks is not a compression technique by itself, it is a useful starting point for the real work of compression. In JPEG, each of these 8\*8 blocks is further processed with a Discrete Cosine Transform (DCT) in both dimensions, and is quantized by dividing each transformed value by some quantization coefficient that is a function of position in the transform and rounding to the nearest integer. The lossiness of JPEG is a result of quantization. Further compression is achieved through lossless techniques such as differential coding, run-length encoding (RLE) and Huffman encoding.

Although JPEG provides moderate compression, neural-network approaches using either nonlinear or linear processing elements may outperform it, because of the linear nature of the underlying DCT. That is because JPEG tries to exploit intra-block correlations found by re-mapping the image into the DCT domain. Again the underlying problem can be expressed as a sort of a data-modeling problem. The basic assumption is that the dimension of the space in which the image blocks live in is lower than the dimension of the space in which they are originally represented. For example, consider a block-based image representation where each block is 8x8 pixels. The image is now described by a collection of points living in a 64-dimensional space, with each point representing a 8x8 block, and thus all the blocks live in a 64-dimensional space. However, it will probably be the case that the set of these points will not cover the entire space. What is needed is a transformation that maps these points into another space where some of the dimensions do not require so much resolution in order to perform an

adequate reconstruction. JPEG uses DCT for such a transformation or mapping. Neural networks with nonlinear processing elements include the space of linear transforms. (It has been shown that a network with linear hidden units theoretically performs a principal components analysis (PCA), However, a PCA will usually return the solution in more ordered form.) With a nonlinear transform, however, it may be possible to describe the space using fewer dimensions than a linear transform. In addition, standard JPEG does not directly exploit block-to-block or inter-block redundancy, whereas a neural-network approach can exploit both inter- and intra-block redundancy, and thus may achieve higher efficiency, as least in certain classes of images.

### Our Neural-Network Approach

In our recent work, we have used this block-based approach with auto-associative neural networks for image compression. An auto-associative neural network generates both the forward transformation and the inverse transformation from a given space into another space learned by the network. Compression is obtained and controlled by the dimension of the intermediate space (the number of hidden units in the intermediate layer) and the levels of quantization of the activations of this hidden layer. In this approach, the image is coded as the combination of the points in the new lower-dimensional space and the inverse transform (which consists of the weights and network architecture between the intermediate layer and output layer.) However, although the inverse transform can be considered overhead, its impact can be lessened by transmitting once for multiple frames. To accomplish this, the transform must be learned over multiple frames. The consequences of the latency imposed by this approach is dependent on the application.

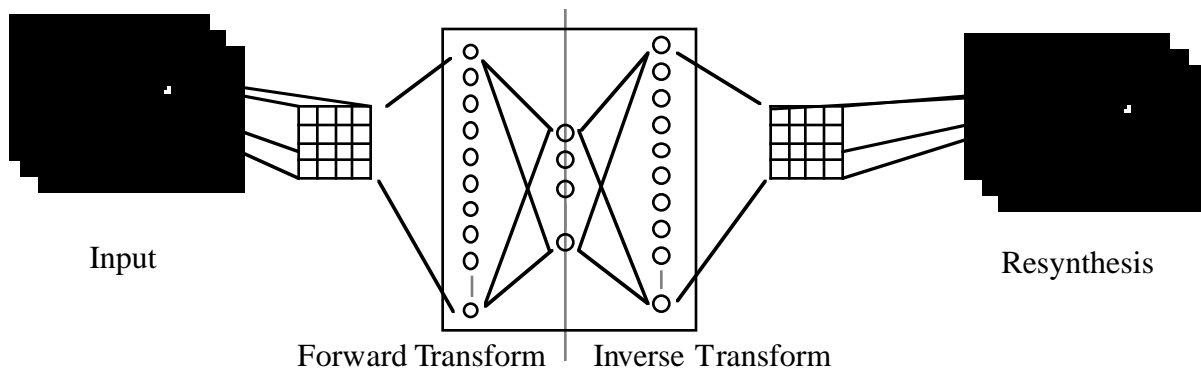


Figure 4 - Block based image representation with an auto-associative neural network for compression. The network architecture and weights between the intermediate (middle) layer and output layers are transmitted along with the activities of the intermediate layer.

Using the NN method described, we have obtained the following results on a series of IR images involving a tank shown in Table 2. The image

sequence contains 200 frames of 12-bit 40x56 pixel images. In each of the experiments, the networks were trained on the first 50 of 200 frames. Sample images obtained for networks at each of the compression rates follow in Figures 5-9. In each of these figures, the original, the neural-network reconstruction, and the difference image are displayed. A difference image is obtained by subtracting the reconstruction from the original and re-scaling the maximum and minimum values to fit between 0 and 255 (which actually *magnifies* the differences). These difference images provide a much more sensitive test for evaluating the compression results, since we only concentrate on the differences without interference from other things, and the magnification of these differences emphasize any minute artifacts. From a subjective point of view, we have observed that for this set of images, the compression technique is more sensitive to the number of hidden units than to the quantization level of the hidden-unit activation-output level. This is illustrated by analyzing the degradation in image quality between Figures 5, 7 and 8. Please note that the compression ratios reported here represent conservative estimates because other compression techniques (such as RLE and Huffman encoding ) have not been used in conjunction with our neural network approach to further enhance the compression.

These results have been reported in a meeting at MICOM with attendance from personnel from both the compression group and the automatic-target-recognition (ATR) group. They are very pleased with our results for the "tank" sequence and they commented that our results for this particular image sequence is better than any of the algorithms they have evaluated. Our demonstrated compression ratio at ~18:1 for the tank images (Figure 9) is 80% better than their targeted 10:1 compression ratio. We are planning to continue working with MICOM on some more image sequences.

**Table 2: NN Compression of Tank Sequence**

NN Hidden Units	Quantization	Compression
20	8	4.63:1 (Figure 5)
10	12	6.25:1 (Figure 6)
8	8	9.26:1 (Figure 7)
10	8	11.58:1 (Figure 8)
10	4	17.89:1 (Figure 9)

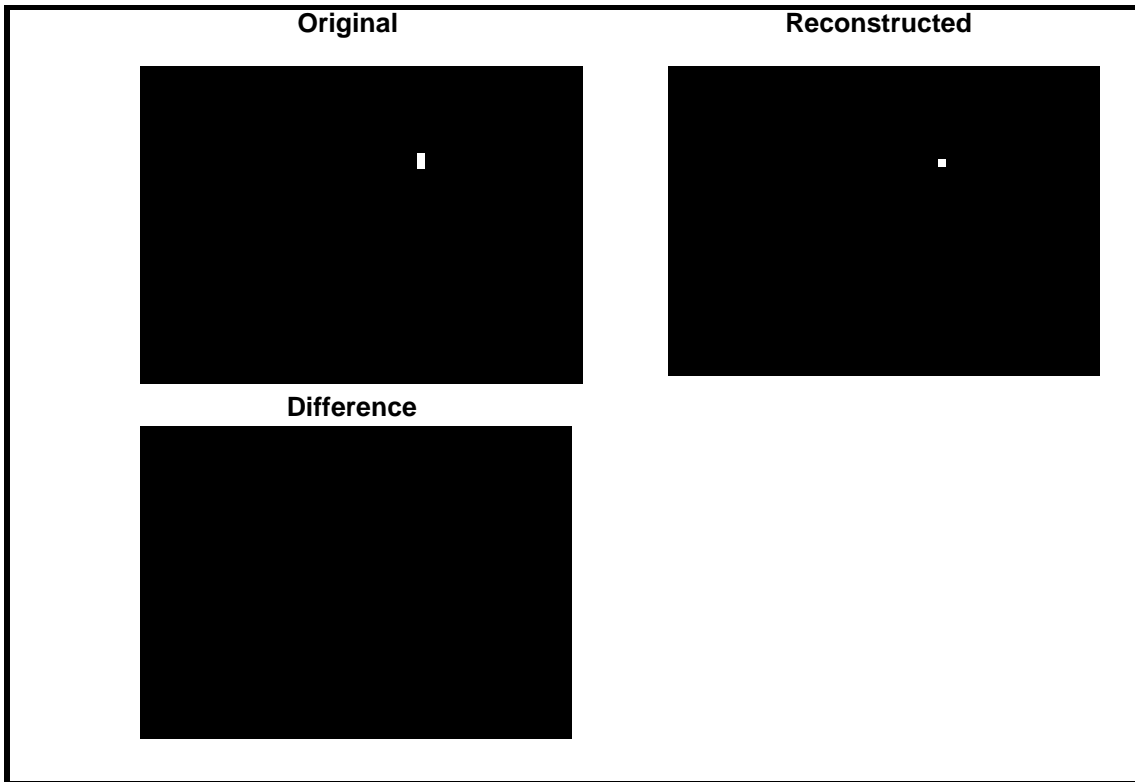


Figure 5 - Neural network compression at 4.63:1

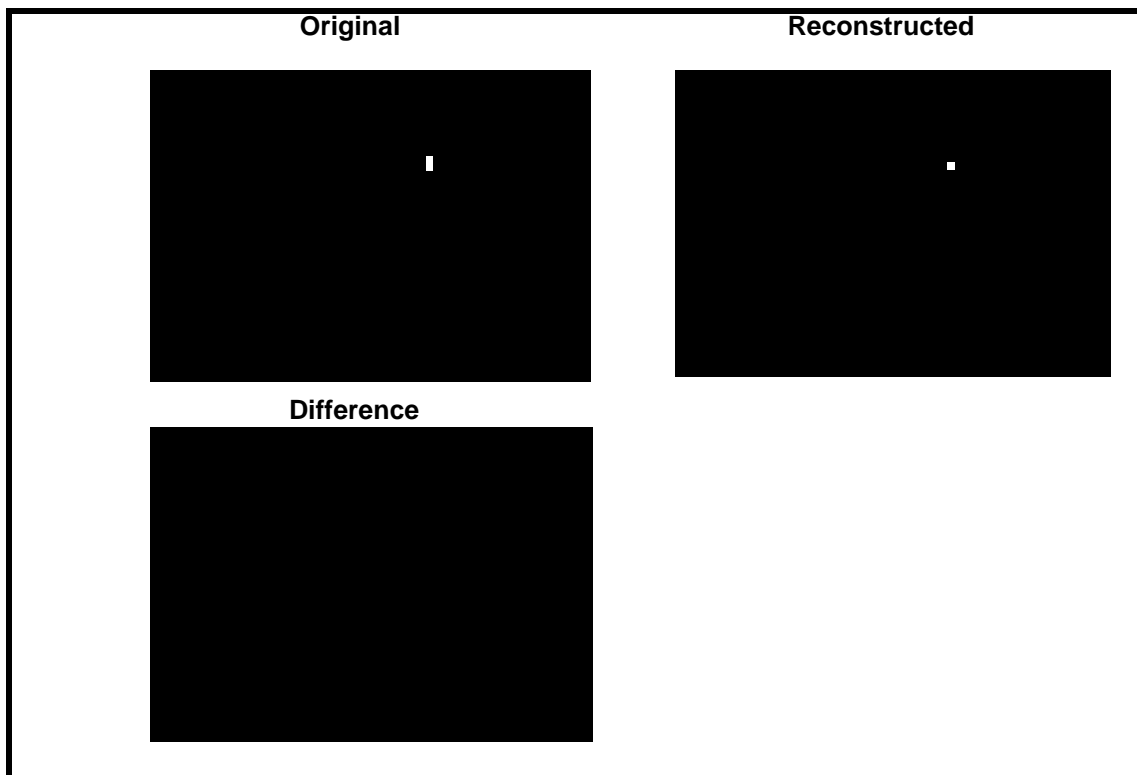


Figure 6 - Neural network compression at 6.25:1

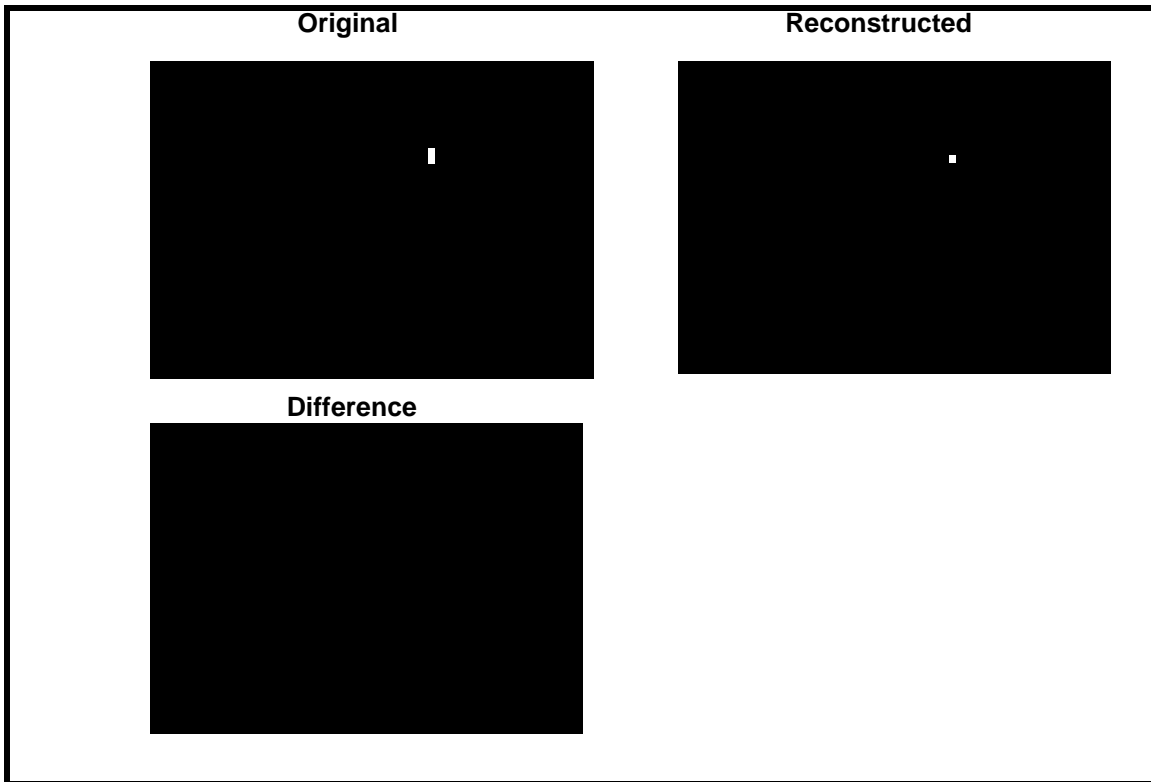


Figure 7 - Neural network compression at 9.26:1

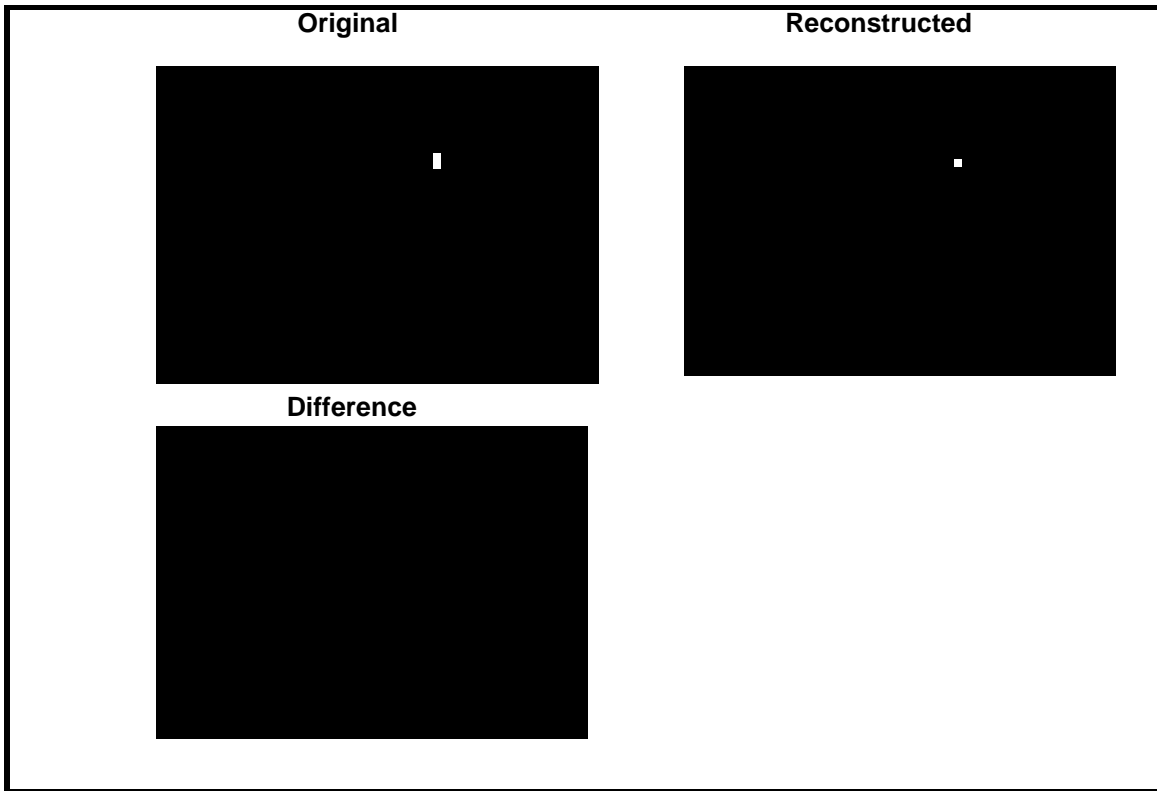


Figure 8 - Neural network compression at 11.58:1

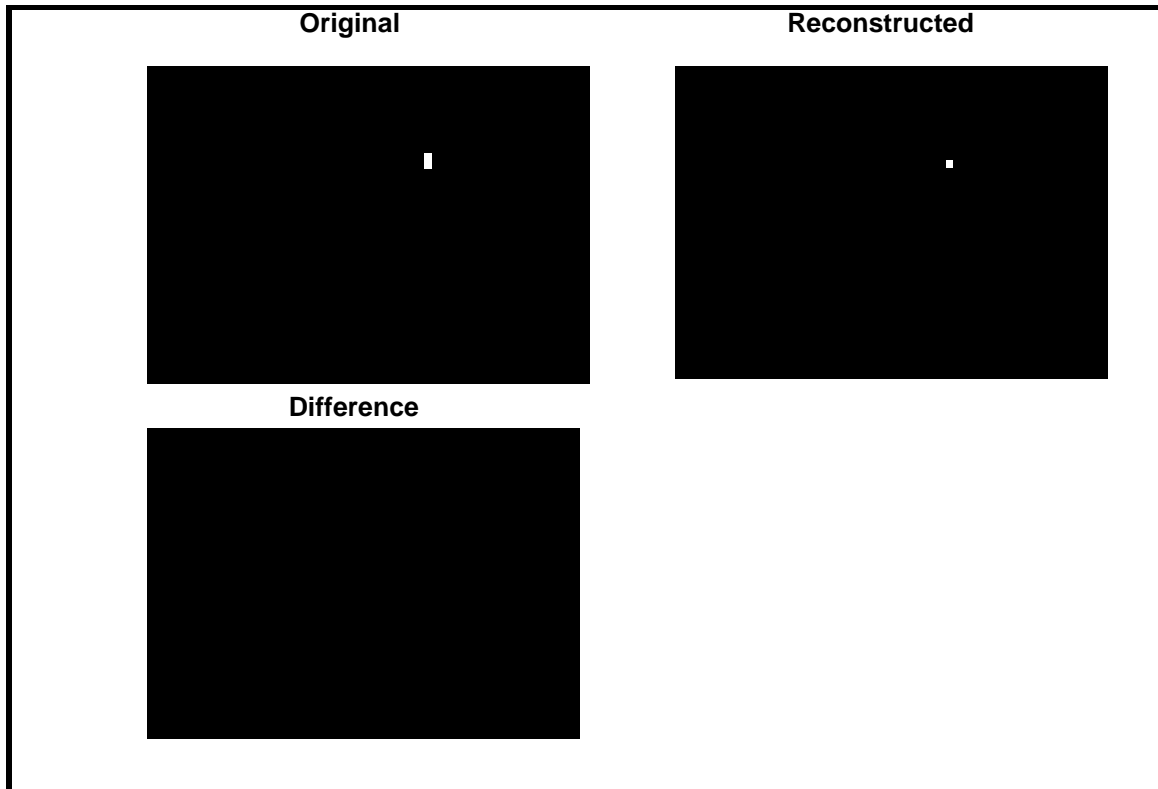


Figure 9 - Neural network compression at 17:89:1

## CONCLUSION

We have finished an enhanced version of some feature-detection and -representation algorithms resulting in a significant improvement (factors of 2 to 10) over our old method. To obtain detailed and objective evaluation of compressed/decompressed images, we have outlined an approach, taking into the account of human visual perception, of developing a set of picture-quality metrics for such an evaluation. We have also begun working closely with MICOM. Our neural-network-based compression technique performed on the tank sequence from MICOM has yielded promising results. These results have been reported in a meeting at MICOM with attendance from personnel from both the compression group and the automatic-target-recognition group. They are very pleased with our progress in the "tank" sequence. The MICOM associative director has sent a letter to our deputy associative director, Dr. Milton Finger, at our Laboratory, commending the positive results of our project.



*Technical Information Department • Lawrence Livermore National Laboratory*  
*University of California • Livermore, California 94551*

