MERL – A MITSUBISHI ELECTRIC RESEARCH LABORATORY
http://www.merl.com

Rate-Distortion Modeling for Multiscale Binary Shape Coding Based on Markov Random Fields

Anthony Vetro Yao Wang Huifang Sun

TR-2003-31 March 2003

Abstract

The purpose of this paper it to explore the relationship between the rate-distortion characteristics of multiscale binary shape and Markov Random Field (MRF) parameters. For coding, it is important that the input parameters that will be used to define this relationship be able to distinguish between the same shape at different scales, as well as different shapes at the same scale. In this work, we consider an MRF model, referred to as the Chien model, which accounts for high-order spatial interactions among pixels. We propose to use the statistical moments of the Chien model as input to a neural network to accurately predict the rate and distortion of the binary shape when coded at various scales.

Published in IEEE Transactions on Image Processing, March 2003.

Rate-Distortion Modeling for Multiscale Binary Shape Coding Based on Markov Random Fields

Anthony Vetro*, *Member, IEEE*, Yao Wang, *Senior Member, IEEE*, and Huifang Sun, *Fellow, IEEE*

EDICS: 2-MODL

Contact Author Information: Dr. Anthony Vetro, Mitsubishi Electric Research Labs, 558 Central Ave, Murray Hill, NJ 07974. Tel: +1-908-363-0504 Fax: +1-908-363-0550, Email: avetro@merl.com.

Manuscript received April 2001; revised June 2002.
A. Vetro and H. Sun are with are with Mitsubishi Electric Research Laboratories, Murray Hill, NJ 07974.
Y. Wang is with the Department of Electrical Engineering, Polytechnic University, Brooklyn, NY 11201.

Abstract

The purpose of this paper it to explore the relationship between the rate-distortion characteristics of multiscale binary shape and Markov Random Field (MRF) parameters. For coding, it is important that the input parameters that will be used to define this relationship be able to distinguish between the same shape at different scales, as well as different shapes at the same scale. In this work, we consider an MRF model, referred to as the Chien model, which accounts for high-order spatial interactions among pixels. We propose to use the statistical moments of the Chien model as input to a neural network to accurately predict the rate and distortion of the binary shape when coded at various scales.

Keywords

Markov Random Fields, Rate-Distortion, Multiscale, MPEG-4, Shape Coding

I. INTRODUCTION

In recent years, the interest in object-based coding has grown. This is partly due to the creation of the MPEG-4 standard [1], which allows for arbitrarily-shaped objects to be coded. The key technical development that allows object-based functionality to be achieved is an efficient and flexible means of coding the shape of an object. The MPEG committee has adopted a context-based arithmetic encoding (CAE) algorithm for this purpose [2]. With this coding scheme, the binary shape data may be coded at multiple scales, which allows for the shape coder to achieve a trade-off between rate and distortion.

For texture coding, a variety of models have been developed that provide a relation between the rate and distortion, e.g., [3], [4]. These models are most useful for rate control of texture information. Given some bit budget for an object or frame, one can find a quantizer value that meets the specified constraint on the rate. Additionally, such models can be used to analyze the source or sources to be encoded in an effort to optimize the coding efficiency in a computationally efficient way. In the case of shape coding, however, no such models exist. The relationship between the rate and distortion for shape coding is very different from that for texture coding. It is clear though that such a model would be very useful in the analysis stage of an object-based encoder. For example, the model can be used for optimal bit allocation among objects, including allocation between shape and texture data. Also, it has been reported in [5] that shape coding can consume a significant percentage of the total bit-rate. In this case, a shape model can help to improve the stability of a buffer by accurately predicting the number of shape bits that will be used in the next frame.

In this paper, we consider a probabilistic approach for modeling the rate-distortion characteristics of shape. This study is restricted to intra-coded shape, i.e., shape that is coded independent of data in other

frames. Working in this domain allows us to focus on the spatial interactions among pixels, which is central to the class of context-based coders that we are dealing with. We propose to model the shape based on the statistics (or moments) that one can extract from the data. To model the shape coding process, we would like to have a distribution whose samples resemble the type of data that we are trying to code. Furthermore, due to the nature of the shape coding process under consideration, we establish two criteria for the statistics: (i) they should be able to distinguish between different shapes at the same scale, and (ii) they should be able to distinguish between the same shape at different scales.

Traditionally, Markov Random Fields (MRF's) have been extensively used in image processing applications, such as image restoration and segmentation [6], [7], [8], [9], [10]. Because of their ability to model global properties using local constraints, MRF's have become very popular and seem to provide a natural fit to our multiscale shape modeling problem as well. In [11], the suitability of two specific MRF models were compared. The first MRF model accounted for a pair-wise interaction between pels, and for the binary case, is typically referred to as the auto-logistic model. The second MRF model accounted for higher order spatial interactions and is referred to as the Chien model. This study concluded that only the Chien model was able to satisfy the two criteria stated above. Further details can also be found in [12].

To put this problem into a broader context, the problem is formally stated as follows. Let $(R, D)_k$ denote the rate-distortion values for a binary shape that is coded at resolution k. Using input parameters, b_i, and modeling function, $f(\cdot)$, we consider the approximation,

$$(R, D)_k \approx f(b_i) \tag{1}$$

Here, we apply this formulation to the modeling of the shape coding process. For the input parameters, the sufficient statistics of the Chien model are used, and for the modeling function, we investigate the use of a multi-layer neural network to map these input parameters to estimates of the rate and distortion at different scales.

It should be pointed out that context modeling is not a new problem. For example, outside the coding domain, this kind of problem has been addressed to predict the reliability of an algorithms output [13]. The authors of this work use three contextual variables as input: Gabor components, entropy and signal/noise ratio. Based on these variables, the prediction function is determined from training using a logistic regression model. Although the problem is quite different from what is considered here, the philosophy is indeed the same.

The rest of this paper is organized as follows. The next section provides a brief review of the CAE shape coding algorithm adopted by MPEG-4, and also provides an overview of MRF's and the Chien model,

in particular. Section III describes the proposed method to estimate the rate-distortion characteristics at various scales. Finally, results of our modeling are provided in Section IV and a summary of this work is provided in Section V.

II. BACKGROUND

A. Shape Coding

This section reviews the basics of the CAE shape coding algorithm as an example of a context-based coder. It should be mentioned that context-based coders have long been used for image and data compression, such as in the JBIG standard for coding bi-level images [14], in the coding of gray-level images [15], [16], [17] and in the compression of multifont Chinese characters [18]. However, since the emphasis of this paper is on shape modeling, we only review relevant parts of the CAE algorithm. This algorithm has been adopted by MPEG-4 to code the shape of arbitrarily shaped objects and has demonstrated superior performance in comparison to other proposed algorithms over a wide range of testing conditions. Further details of the algorithm can be found in [1], [2].

The CAE algorithm works at a macroblock level, and specifies whether a pixel in a macroblock belongs to an object or not using an arithmetic coder. As with any context-based coder, the arithmetic coder makes use of the conditional probability that this pixel belongs to the object given the status of its adjacent pixels in a causal neighborhood. The algorithm may operate in a variety of modes. However, for the purpose of this paper, it is sufficient to describe the intra mode only. In intra mode, three different types of macroblocks are considered: transparent, opaque, and border blocks. Transparent blocks do not contain information about the object of interest, opaque blocks are located entirely inside the object, and border blocks cover only part of the object. Transparent and opaque blocks are signaled as a macroblock type. For the border blocks, a template of 10 pels is used to define the casual context for predicting the shape value of the current pel. This context is shown in Fig. 1, and is computed according to

$$C = \sum_k c_k \cdot 2^k. \tag{2}$$

The context number is used to access a probability table, and finally, the sequence of probabilities are used to drive an arithmetic encoder.

When the shape blocks are coded at full-resolution (16×16), this algorithm is able to achieve a lossless representation. To reduce the bit-rate, distortion can be introduced by down-sampling the original block by a factor of two or four. The distortion is computed as the difference between the original shape and the reconstructed shape.

It should be noted that the motivation for exploring R-D models for shape coding is to avoid performing the down-sampling, up-sampling and actual coding to get the R-D characteristics of the binary shape over various scales. Accurate estimates of these characteristics can play a key role in optimizing the bit allocation for the binary shape among blocks and between shape and texture coding. Also, knowing the expected rate for shape can stabilize the buffer occupancy, especially for low bit-rate video coding.

B. Markov Random Field Models

In the following, we provide a brief overview of MRF's and the specific model under consideration, the Chien model. A discrete Gibbs random field (GRF) provides a global model for an image, X, by specifying a probability mass function of the form

$$P(X) = \frac{1}{Z} e^{-U(X)} \tag{3}$$

The function $U(X)$ is called the *energy function* and the normalizing constant Z is called the *partition function*. The energy

$$U(X) = \sum_{c \epsilon \mathcal{C}} V_c(X) \tag{4}$$

is a sum of clique potentials $V_c(X)$ over all possible cliques $\mathcal{C}$. The value of $V_c(X)$ depends on the local configuration of the clique c. One point to keep in mind is that $P(X)$ measures the probability of occurrence of a particular pattern. The more probable configurations are those with lower energies.

In [19], [20], a model that preserves the fine structures of an image is presented. The model relies on three parameters, edge, line and noise parameters, denoted by e, l and n, respectively. The set of cliques are 3×3 squares, so the neighborhood of a given site is 5×5. Different local configurations are classified by the dimension and orientation of the pattern. By considering all possible patterns over the 3×3 block and taking into account symmetric and rotational considerations, it turns out that 51 distinct configurations exist. Each of the 51 configurations is associated with a potential, $C(i)$, which is a linear combination of the three parameters:

$$C(i) = \epsilon(i)e + \lambda(i)l + \eta(i)n \tag{5}$$

Edge and line constraints help to define a system of equations that ultimately assign energy that is contributed for each parameter. As shown in the above equation, the energy associated with the edge, line and noise parameters are $\epsilon(i)$, $\lambda(i)$ and $\eta(i)$, respectively. Configurations that are not subject to edge

and line constraints are assigned noise. Further details on the possible configurations and their assigned energies can be found in [19], [20].

The distribution of the Chien model is a Gibbs field whose energy is linear with respect to its parameters and is given by,

$$U(X) = eN_0(X) + lN_1(X) + nN_2(X) \tag{6}$$

where

$$\begin{aligned} N_0(X) &= \sum_{i=1..51} \epsilon(i)\zeta_i(X) \\ N_1(X) &= \sum_{i=1..51} \lambda(i)\zeta_i(X) \\ N_2(X) &= \sum_{i=1..51} \eta(i)\zeta_i(X) \end{aligned} \tag{7}$$

and $\zeta_i(X)$ is the number of configurations of type i in X.

The Chien moments given by eqn. (7) are considered sufficient statistics of this particular MRF model and will be used in the rest of the paper. In terms of computation, they are quite simple to extract from the data. For example, given image X, we compute $\zeta_i(X)$, which denotes the number of i-type configurations contained in X, $i = 1..51$. Each of these configurations are assigned a cost in terms of edge, line and noise energy, which are given by $\epsilon(i)$, $\lambda(i)$ and $\eta(i)$, respectively [19]. Given these energy values and $\zeta_i(X)$ for each configuration, the moments are easily determined using eqn. (7). To estimate the model parameters, namely e, l and n, one may employ the *histogram method* [10] or a Markov Chain Monte Carlo Maximum Likelihood (MCMCML) estimation [20].

III. Rate-Distortion Modeling for Shape Coding

To estimate the rate and distortion at various scales, we choose the statistical moments of the Chien model, i.e., N_0, N_1 and N_2, as our set of input parameters. As shown in [11] these parameters exhibit the favorable properties discussed earlier; that is, they are able to distinguish between the same shape at different scales and different shapes at the same scale. Such properties were not observed with the auto-logistic model, which is the reason we focus on the Chien model only in this paper.

Two sets of sample binary images are shown in Figs. 2 and 3. The first set of images in Fig. 2 are used to examine the Chien moments at the encoding resolution, where we aim to understand the relation between these moments and the rate. The reduced resolution images in this set are referred to as images with a conversion ratio (CR) of either 2 or 4. On the other hand, the second set of images in Fig. 3 are used to examine the Chien moments at the same resolution, i.e., after the successive down-sampling and up-sampling at a specified CR. With these images, we study the relation between the Chien moments

and the distortion. The reconstructed images in this set are denoted as $\hat{2}$ or $\hat{4}$, and the difference image between the original and reconstructed images is denoted by either $(1-\hat{2})$ or $(1-\hat{4})$.

The Chien parameters and moments corresponding to the sample images in Figs. 2 and 3 are given in Tables I and II, respectively. Table I associates the corresponding rate to the first set of sample images, while Table II associates the distortion to the second set of sample images. The Chien moments are calculated according to eqn. (7), the rate is obtained by coding the shape image with the CAE algorithm, and the distortion between the original and reconstructed images is calculated as the sum of absolute pixel differences, i.e., the total number of different pixels in the binary image. The Chien parameters are only provided as supplemental information and are estimated using the MCMCML approach described in [20].

From these tables, it is clear that the Chien parameter values and moments depend heavily on the particular segmentation map that is being considered. In Table I, we observe that the moments are generally well-behaved and decrease in value with lower resolution. The one anomaly is with the Dancer sequence at CR=2, where the N_0 and N_1 values do not follow the decreasing pattern from CR=1 to CR=4.

In Table II, the extracted Chien moments agree with our intuition for the most part in that noisy boundaries are smoothed out in the reconstructed images. This is evident with the Coastguard and Dancer images, where thin lines have either broadened or completely disappear. For these images, we observe that the values of N_1 corresponding to the line energies completely disappear in the smoothed out image. Also, the values of N_2 corresponding to the noise are significantly dampened. The Foreman image however, illustrates a different and interesting case in contrast to the examples above. In this image, the original is already very smooth and due to the down/up-sampling operations the reconstructed image becomes more noisy. This can be verified through close inspection of Fig. 3. As expected, the value of N_2 is significantly increased - it jumps from a relatively small value of 15 to 244.

From Tables I and II and the above observations, we see that the rate and distortion associated with a shape strongly varies with the Chien moments (and parameters) for that shape. The relationship is however not obvious and definitely non-linear. This is expected though, primarily because the data is binary and its transformation between scales is a non-linear operation. For instance, in MPEG-4, down-sampled pixels are computed using a majority-type operation among a 4-pixel neighborhood, while up-sampled pixels are computed based on the adaptive context of a 12-pixel neighborhood. We propose a neural network approach to model the relationship. It is well-known that neural networks provide a powerful framework that is capable of performing non-linear regression [21].

There are many types of neural networks that may be used for this purpose; some may be better suited than others. However, the aim of our work is not to provide an exhaustive study on the different types of neural networks that may be used to model the rate-distortion characteristics of shape. We only hope to demonstrate their applicability to the problem. Therefore, we adopt a simple multilayer feed-forward network as shown in Figure 4, where the input to this network are the statistical moments of the Chien model that are calculated from the original shape image, and the output is the estimated rate and distortion of the shape data at different scales. In this network, we have one hidden layer. The commonly used sigmoid function is used as the activation function for nodes in the hidden layer, while a linear function is used for nodes in the output layer. With a fixed structure and fixed activation functions, the whole network represents a complex non-linear function, where the weights between each node can be seen as the parameters or coefficients of this function. For this function to yield the desired output, back-propagation is applied to determine suitable weights and biases of the network using supervised training [22].

Thus far, we have only described the basic tool and framework that we propose to use for our modeling problem. However, within this framework, there is still a significant amount of flexibility regarding its implementation and use. For instance, the number of nodes in the hidden layer has not yet been mentioned. If too few nodes are chosen, then the network will not be capable of representing the desired function. On the other hand, if too many nodes are chosen, the network will be able to memorize all the examples in the training set by forming a large look-up table, but will not generalize well to inputs that have not been seen before. In other words, like all statistical models, the neural network is subject to over-fitting when there are too many parameters in the model. Since there are no methods to determine the suitable number of hidden nodes, the number is decided experimentally.

Another important aspect that we need to consider is the input to our network. In the experiments that we used to determine the suitability of the Chien parameters for modeling, we extracted parameters corresponding to each down-sampled scale for the rate, and parameters corresponding to the reconstructed shape at each scale for the distortion. However, one of the primary objectives in modeling the shape is to keep computational complexity low. Therefore, if it is possible to predict the rate-distortion characteristics for each scale using parameters extracted from the full-scale only, we can even avoid the down-sampling process.

There are a number of papers in the literature that investigate the relation between MRF's at different scales [23], [24], [25], [26], [27]. Although many different approaches and philosophies have been pre-

sented in these papers, none have focused on the binary shape problem that we have considered. Rather, the focus has been image restoration and segmentation, or the development of a more general multiscale framework that can be used in a wide range of detection, estimation and synthesis applications. In [24], for example, Bouman and Shapiro propose a novel multiscale random field (MSRF) model that is composed of a series of random field models progressing from coarse to fine scale. The series of fields form a Markov chain in scale, where it is assumed that points in each field are conditionally independent given their coarser scale neighbors.

Although the main focus of the paper by Bouman and Shapiro is to provide an efficient sequential MAP estimator and apply it to a multiscale segmentation problem, some key points related to our framework are evident. First, since the MSRF is a Markov chain in scale, the distribution of the field at a fine scale given all the coarser scale fields is only dependent on the next coarser scale. This implies that the distribution of the finer scale contains information about all coarser scales and supports the possibility of estimating the rate and distortion of coarser scales from the original shape only. This is quite intuitive and expected. As a second point, the relationship between the fields at different scales are specified by transition densities. The transition densities are dependent only on a local neighborhood of pixels at the next coarser scale. Along these lines, we may think of our neural network as a framework that learns and specifies the relation between the distributions at various scales. For the moment, there is no formal evidence to prove this notion, but this analogy is quite interesting and is certainly a topic for further study. The outcome of this research could potentially lead to a better understanding of the multiscale relationship of MRF's and more efficient estimators.

In summary, there is reason to believe that the neural network should be able to model the rate and distortion at coarser scales based on the sufficient statistics at the finest scale. The accuracy of this claim will be explored in the next section, where we consider the Chien parameters extracted from the full-resolution only as input to our neural network.

IV. Experimental Results

The objectives of the experiments in this section are twofold. First, our primary aim is to demonstrate that MRF parameters, and specifically the Chien model parameters, can lead to accurate estimates of the rate and distortion at various scales. Second, we would like to provide support for the claims in the previous subsection that indicate the possibility of predicting these characteristics from data observed at the full-resolution only.

For the experiment, we consider a number of binary objects that represent a wide array of object

characteristics. In all, seven clips have been selected, where each clip is composed of 25 consecutive frames. Four of the seven clips are used for training, and the other three are used for testing. The training set consists of 1 object from each of the following test sequences: Akiyo (frames 1-25), Coastguard (frames 50-75), Container (frames 1-25) and Singer (frames 1-25). From Akiyo, we have a stationary head-and-shoulder object, from Coastguard, the object is a large moving boat, from Container, the object represents a small moving boat, and from Singer, the object represents a slim silhouette of a human body with waving arms. The test set consist of 1 object from the following sequences: Foreman (frames 160-185), Coastguard (frames 4-29) and Dancer (frames 1-25). The object from Foreman is also a head-and-shoulder object, but with very different characteristics compared to Akiyo and the object is leaving the scene during these frames. In the testing clip used for Coastguard, the boat is just beginning to enter the scene and is a much different shape than that used in training. The object from the Dancer sequence consists of two human bodies touching hands.

The Chien model parameters that are used as input to the neural network are shown in Figure 5. To provide maximal benefit to the neural network, each of the input parameters have been scaled. The edge, line and noise parameters have been scaled by a factor of 1000, 10 and 100, respectively, so that the parameters are on the same order of magnitude. This is done since the activation function in the first layer is a sigmoid function and operates best with inputs on the order of 1. Larger values would saturate the function, hence the whole network would have difficulty in approximating a non-linear function. From this plot, we can see that a wide range of data characteristics are represented.

As stated earlier, the weights and biases within the network are trained using back-propagation. This is an iterative procedure and for our training data required approximately 200 iterations to converge. In our experiments, we used 20 nodes in the hidden layer, but only noticed small difference when simulated with 10 or 15 nodes. The output layer consists of 5 nodes, three nodes corresponding to the rate at full, half and quarter scale, and 2 nodes corresponding to the distortion between the reconstructed binary maps from half and quarter scales. The results of our experiments are illustrated in Figures 6 and 7.

The plots in Figure 6 provide a comparison of the actual rates and output rates of the neural network for each of the three scales. In these plots, the first 100 frames correspond to the training set and the next 75 frames correspond to the testing set. We can see that that the neural network does indeed provide close estimates to the actual rate and follows the trends quite well for both the training and testing data.

Similarly, the plots in Figure 7 provide a comparison of the actual distortions and output distortions of the neural network for the reconstructed binary maps from half and quarter scales. Although the estimates

for the distortion are not as close as those for the rate, they are still good and provide consistent results for all of the three test sequences at the different scales. However, it should be noted that the distortion is much more difficult to approximate compared to the rate. This is mainly due to the loss of information in the down-sampling combined with a complex up-sampling process that accounts for neighboring dependency. We should note that it may be possible to achieve better predictions if observations were also made from the lower resolutions scales. In this way, the current neural network could be modified to accept a larger input vector consisting of input parameters from both the original and reduced resolution. Alternatively, a different neural network could be used at each scale.

Overall, the results shown here demonstrate that MRF parameters can be used to model multiscale rate-distortion characteristics of binary shape, and do so with information provided at the full-resolution only.

V. Concluding Remarks

This paper introduced and investigated the problem of rate-distortion modeling of multiscale binary shape data. A probabilistic approach based on MRF's has been proposed. The sufficient statistics of the Chien model were used as input parameters to our modeling function. Since we are dealing with binary data in a multiscale environment, we are forced to deal with non-linear operations on the data, such as down-sampling and up-sampling. In this paper, we have adopted a multilayer feed-forward network to provide us with a function that maps the Chien model parameters to rate-distortion characteristics of the binary shape at various scales. A single network has been trained to predict the rate at three different scales and the distortion between the full-resolution binary map and two coarser scales. The input to the network consisted of the Chien model parameters that were extracted from the full-resolution only. As shown in the experiments, the estimated rates were very close to the actual values over a wide range of objects with varying properties. The estimated distortion values are less accurate, but still quite good.

As noted earlier, the above results on shape modeling apply to intra-coded shapes only. To be used in a video coding environment, a model for inter-shape coding is necessary as well. The extension to inter-coded shapes seems non-trivial at this time, but we hope that the work presented in this paper provides a good treatment of the problem and a solid foundation of ideas for further study. We feel that the main problem is that the methods currently used to model the shape lack support for a time-series of data and make no mention of the temporal interactions among pixels. As a result, the shape models that are developed in this paper are not demonstrated within the context of a rate control algorithm. However, it is our hope that the implications of this work are much broader than just object-based coding. There is

a strong similarity between the problem that we explore here and multiscale modeling and estimation in general, which has application to segmentation, pattern analysis, detection and synthesis problems.

ACKNOWLEDGMENT

The authors would like to thank Dr. Xavier Descombes of INRIA, France for providing software to estimate the Chien model parameters, Prof. Min Wu from the University of Maryland, College Park, for her assistance with the neural network simulations, and Dr. Onur Gueryuz, now with Epson Labs, for his time spent discussing this problem.

REFERENCES

[1] ISO/IEC 14496-2:2001, "Coding of audio-visual objects - Part 2: Visual," 2nd Edition, 2001.

[2] N. Brady, "MPEG-4 standardized methods for the compression of arbitrarily shaped objects," *IEEE Trans. Circuits and Syst. Video Technol.*, vol. 9, no. 8, pp. 1170-1189, Dec. 1999.

[3] T. Chiang and Y-Q. Zhang, "A new rate control scheme using quadratic rate-distortion modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 246-250, Feb 1997.

[4] H.M. Hang and J.J Chen, "Source model for transform video coder and its application - Part I: Fundamental theory," *IEEE Trans. Circuits Syst. Video Technol.*, vol.7, no.2, pp. 287-298, April 1997.

[5] A. Vetro, H. Sun, and Y. Wang, "MPEG-4 rate control for multiple video objects," *IEEE Trans. Circuits and Syst. Video Technol.*, vol. 9, no. 1, pp. 186-199, Feb. 1999.

[6] R. C. Dubes and A. K. Jain, "Random field models in image analysis," *J. Applied Statistics*, vol. 16, no. 2, pp.131-163, 1989.

[7] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Royal Statistic. Soc., Series B*, vol. 36, pp. 192-236, 1974.

[8] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 6, pp. 721-741, 1984.

[9] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 5, no. 1, pp. 25-39, 1983.

[10] H. Derin and H. Elliot, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, no. 1, pp. 39-55, 1987.

[11] A. Vetro, Y. Wang and H. Sun, "A probabilistic approach for rate-distortion modeling of binary shape," *Proc. IEEE Int'l Conf. Acoustics, Speech, Signal Proc.*, Orlando, FL, May 2002.

[12] A. Vetro, "Object-based coding and transcoding," Ph.D. Dissertation, Polytechnic University, Brooklyn, NY, 2001.

[13] B. Chalmond, C. Graffigne, M. Prenat, and M. Roux, "Contextual performance prediction for low-level image analysis," *IEEE Trans. Image Processing*, vol. 10, no. 7, pp. 1039-1046, July 2001.

[14] ISO/IEC 11544:1992, "Coded representation of picture and audio information – progressive bi-level image compression," 1992.

[15] W.B. Pennebaker and J.L Mitchell, *JPEG – Still image data compression standard*, Van Nostrand Reinhold, New York, 1993.

[16] M.J. Weinberger, J.J. Rissanen, and R.B. Arps, "Application of universal context modeling to lossless compression of gray-scale images," *IEEE Trans. Image Processing*, vol. 5, no. 4, pp. 575-586, Apr. 1996.

[17] X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," *IEEE Trans. Communications*, vol. 45, no. 4, pp. 437-444, Apr. 1997.

[18] M.-K. Tsay, C.-H. Kuo, R.-H. Ju, and M.-K. Chiu, "Data compression of multifont Chinese character patterns," *IEEE Trans. Image Processing*, vol. 3, no. 2, pp. 139-146, Mar. 1994.

[19] X. Descombes, J.F. Mangin, E. Pechersky and M. Sigelle, "Fine structure preserving model for image processing," *Proc. 9th SCIA*, Uppsala, Sweden, pp. 349-356, 1995.

[20] X. Descombes, R.D Morris, J. Zerubia and M. Berthod, "Estimation of Markov Random Field parameters using Markov Chain Monte Carlo Maximum Likelihood ," *IEEE Trans. Image Proc.*, vol. 8, no. 7, pp. 954-962, July 1999.

[21] S.Y Kung, *Digital Neural Networks*, Prentice Hall, 1993.

[22] A.E. Bryson and Y.C. Ho, *Applied Optimal Control*, Blaisdell, New York, 1969.

[23] M.R. Luettgen, W.C. Karl, A.S. Willsky and R.R. Tenney, "Multiscale representation of Markov random fields," *IEEE Trans. Signal Proc.*, vol. 41, no. 12, pp. 3377-3395, Dec. 1993.

[24] C. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. Image Proc.*, vol. 3, no. 2, pp. 162-177, Mar. 1994.

[25] M. Crouse, R.D. Nowak and R.G. Baraniuk, "Wavelet-based statistical signal processing using Hidden Markov Models," *IEEE Trans. Signal Proc.*, vol. 46, no. 4, pp. 886-902, Apr. 1998.

[26] K. Timmermann and R.D. Nowak, "Multiscale modeling and estimation of Poisson Processes with application to photon-limited imaging," *IEEE Trans. Info. Theory*, vol. 45, no. 3, pp. 846-862, Apr. 1999.

[27] K. Daoudi, A. B. Frankt and A. S. Willsky, "Multiscale autoregressive models and wavelets," *IEEE Trans. Info. Theory*, vol. 45, no. 3, pp. 828-845, Apr. 1999.

LIST OF FIGURES

LIST OF TABLES

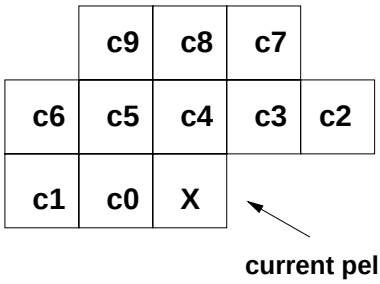


Fig. 1. Illustration of template used to compute intra context. As specified by the standard, padded values are used for pixels that lie outside the macroblock boundary.

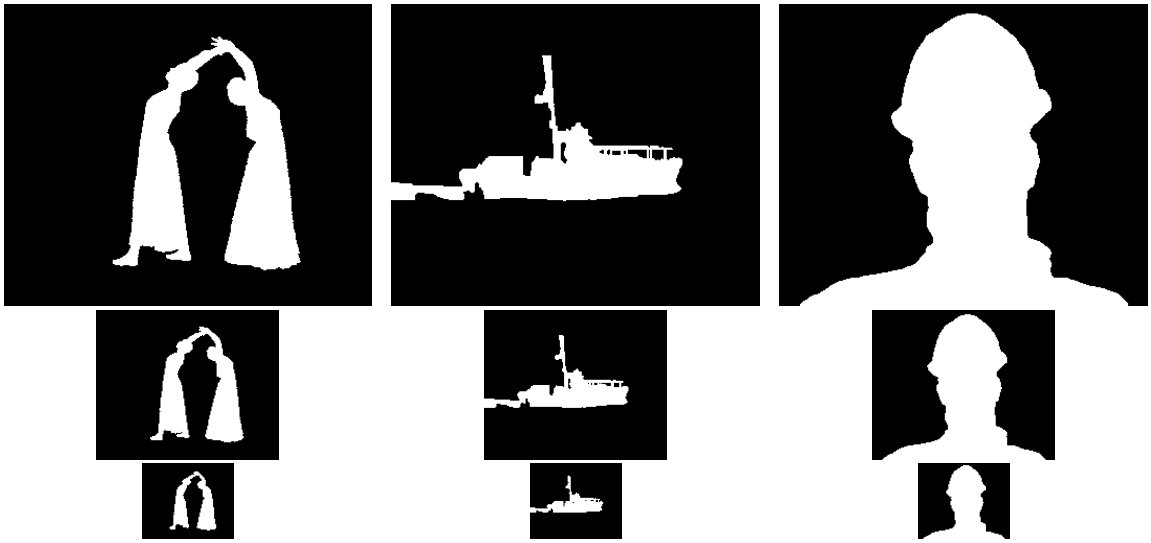


Fig. 2. Test images from the Dancer, Coastguard and Foreman sequences. The first row shows the original shapes at full-resolution. These shapes are then down-sampled by a factor of 2 and 4 according to the MPEG-4 standard; the results are shown in the second and third rows, respectively.

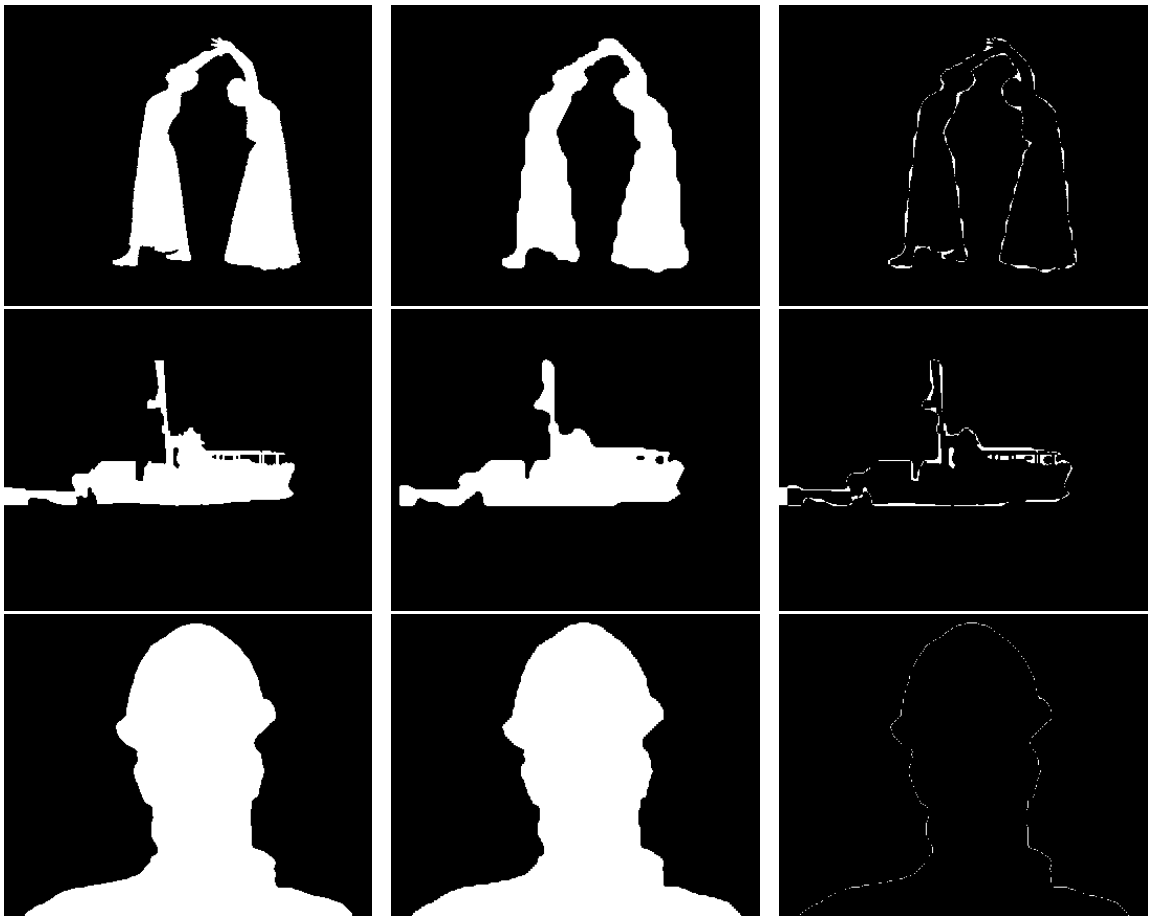


Fig. 3. Test images from the Dancer, Coastguard and Foreman sequences. The first column shows the original shapes at full-resolution. These shapes are then down-sampled and up-sampled according to the MPEG-4 standard; the results are shown in the second column. For Dancer, Coastguard and Foreman the down-sampling factor is 4, 4 and 2, respectively. Finally, the difference between the first and second column shapes are shown in the last column.

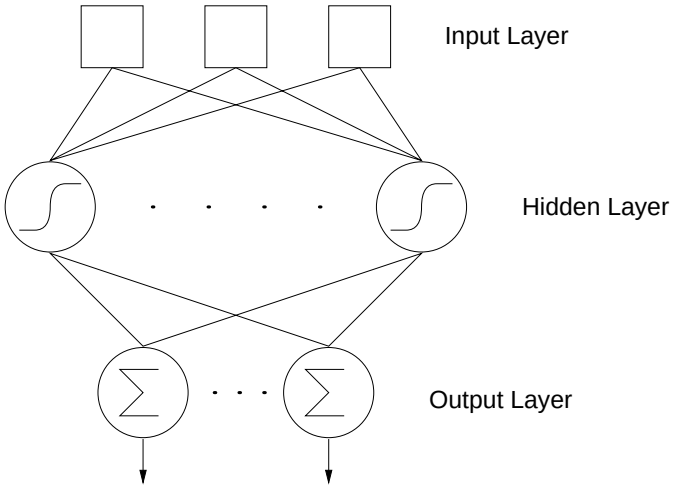


Fig. 4. Multilayer feed-forward network with one hidden layer.

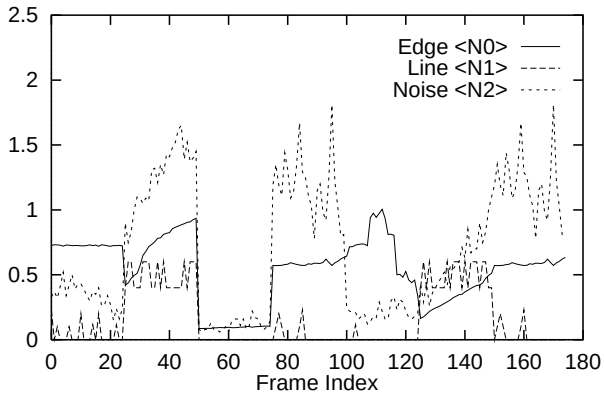
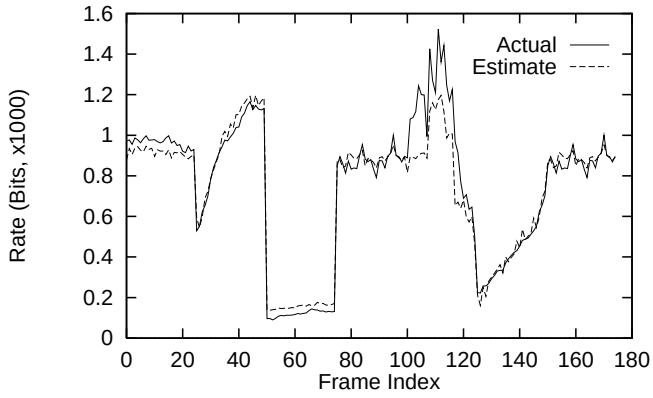
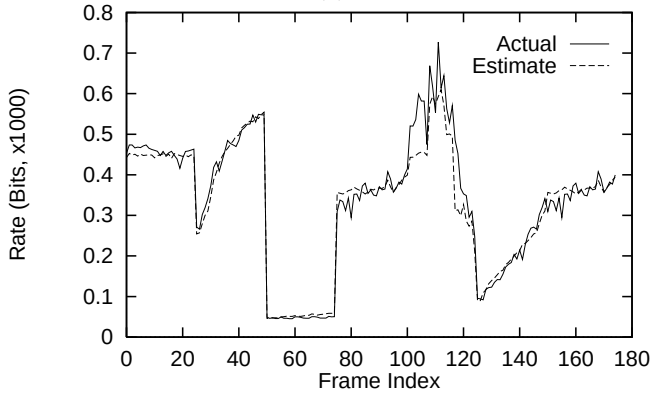


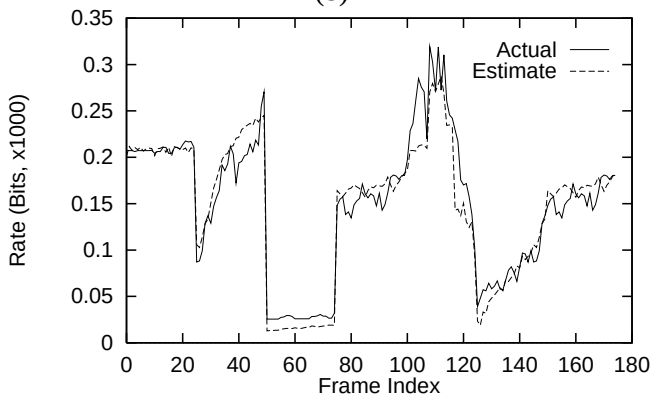
Fig. 5. Traces of Chien input parameters for each of the seven clips used in the experiment. The parameters corresponding to edge, line and noise have been scaled by a factor of 1000, 10 and 100, respectively. The first four clips (corresponding to the first 100 frames) are used for training, while the following three clips (corresponding to the remaining 75 frames) are used for testing.



(a)

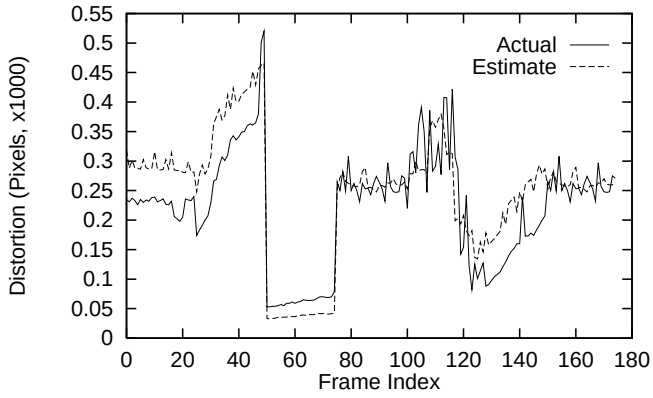


(b)

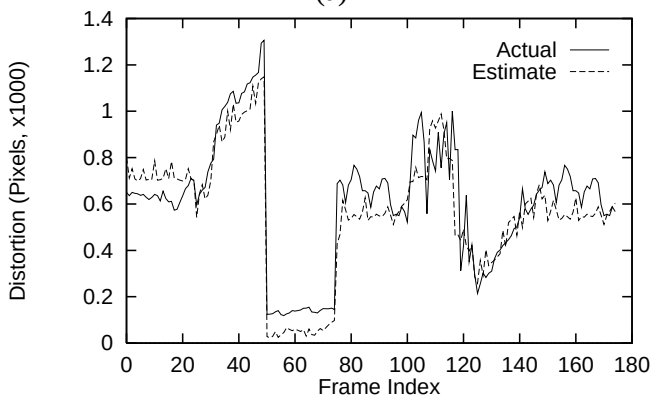


(c)

Fig. 6. Comparison of actual rate and output rate of neural network: (a) full-scale, (b) half-scale, and (c) quarter-scale. The first four clips (corresponding to the first 100 frames) are part of the training set, while the following three clips (corresponding to the remaining 75 frames) are part of the test set.



(a)



(b)

Fig. 7. Comparison of actual distortion and output distortion of neural network. Distortion is measured between original and reconstructed binary map from (a) half-scale, and (b) quarter-scale. The first four clips (corresponding to the first 100 frames) are part of the training set, while the following three clips (corresponding to the remaining 75 frames) are part of the test set.

TABLE I

COMPARISON BETWEEN CHIEN PARAMETERS, CHIEN MOMENTS, AND RATE.

Image		Parameters			Moments			Actual
Map	CR	e	l	n	$< N0 >$	$< N1 >$	$< N2 >$	Rate
Dancer	1	2.24	6.60	1.48	1302	1.15	198	2000
	2	0.97	2.46	1.62	62.2	3.62	85	871
	4	0.84	113.0	1.51	302	0.0	50	410
Coast-	1	2.51	5.95	1.58	1134	6.0	150	1316
guard	2	1.42	4.09	1.36	516	2.0	112	665
	4	0.79	2.92	1.38	204	2.0	88	317
Foreman	1	7.21	31.3	3.20	925	0.0	15	1290
	2	2.11	71.7	2.00	452	0.0	16	612
	4	3.38	20.9	2.20	215	0.0	15	280

TABLE II

COMPARISON BETWEEN CHIEN PARAMETERS, CHIEN MOMENTS, AND DISTORTION.

Image		Parameters			Moments			Actual
Map	CR	e	l	n	$< N0 >$	$< N1 >$	$< N2 >$	Dist.
Dancer	1	2.24	6.60	1.48	1302	1.15	198	
	$\hat{4}$	2.40	34.1	1.68	1251	0.00	110	
	$(1-\hat{4})$	1.15	2.33	1.28	952	185	921	1218
Coast-	1	2.51	5.95	1.58	1134	6.0	150	
guard	$\hat{4}$	4.28	45.5	2.10	961	0.0	68	
	$(1-\hat{4})$	1.73	3.37	1.31	1147	208	496	1710
Foreman	1	7.21	31.3	3.20	925	0.0	15	
	$\hat{2}$	1.27	41.5	1.44	854	0.0	244	
	$(1-\hat{2})$	2.14	3.35	0.88	262	108	811	319