PIXEL PREDICTION BY CONTEXT BASED REGRESSION

Lingyan Sheng, Antonio Ortega

Signal and Image Processing Institute, Department of Electrical Engineering
Viterbi School of Engineering, University of Southern California

ABSTRACT

We propose a pixel prediction algorithm, which learns a regression function corresponding to each context. A context refers to a group of pixels, that have similar correlations with its neighboring pixels. We propose to form a pixel's feature vector by its neighboring pixels' ratios, so that they better capture the pixel properties described by the regression weights. Then we use K-means clustering to classify the feature vectors of all pixels into several contexts. Clustering reduces pixel randomness within each context, thus reducing prediction error. We apply three regression algorithms, the least square, quantile and lasso regression, which assume different loss function and regularization. Experimental results demonstrate that all context based regression methods have outperformed conventional pixel predictors. Among them, quantile regression, which assumes $l1$-norm loss function has the best result. It has 3.1% less bits per pixel (bpp) than least square prediction with 12 neighboring pixels.

Index Terms— lossless image coding, least square, quantile regression, lasso regression

1. INTRODUCTION

Pixel prediction in lossless image compression has two main categories. One category is linear pixel predictors, such as least square (LS) and least entropy (LE), which assume a fixed linear prediction function for each image and infer its prediction coefficients and context grouping thresholds adaptively. Some researchers recently proposed to adaptively generate the prediction function for each image by genetic algorithm [1], but the algorithm is too complicated to be realistic. However an image may contain different contexts. A context refers to a group of pixels, that have similar correlations with its neighboring pixels. Pixels in different contexts behave differently with a fixed prediction function. The other category is nonlinear predictors , such as GAP in CALIC [2], MED in LOCO-I [3] [4]. They have a specific prediction function with respect to the pixel's context, which is represented by the properties of its neighboring pixels. However the prediction scheme is general with all images.

We propose a new pixel prediction method using context based regression. For each pixel in the image, the method first builds a vector of ratios of its neighboring pixels. Then it clusters the vectors into different contexts. We learn a regression function within each context.

One contribution of this paper is building a feature vector of each pixel's neighboring pixels, and using K-means clustering to classify the pixels into different contexts. Our algorithm combines the advantages of linear and nonlinear predictors by learning a different prediction function for each context, and adaptively selects prediction function coefficients. For each pixel in an image, we first build a feature vector of ratios of its neighbors. After deriving the vectors of all pixels, we cluster them into several contexts by K-means clustering [5]. The ratios of neighboring pixels detect the horizontal, vertical and diagonal edges of different scales. They capture the properties of the regression coefficients, thus reducing pixel randomness in each context.

The other contribution of this paper is using regression with different loss function and regularization.We learn three types of regression function in each context, LS, quantile regression [6] and lasso regression [7]. The latter two regressions have never been used in image coding to the best of our knowledge. Quantile regression assumes l_1 loss function, and Lasso regression solves the least square regression with a lasso penalty. The corresponding predictors are named C-LS, C-Quantile and C-Lasso.

Twelve adjacent pixels serve as features in regression. At the decoder, we first infer the context by the neighboring pixels' distance to the centers of K contexts. Then the pixel value is estimated by the regression function of the particular context. The encoder needs to send the centers of K contexts and the regression coefficients of the contexts. We test C-LS, C-Quantile and C-Lasso on four images. All three algorithm outperform standard pixel predictors. C-Quantile has the best compression performance with $q = 0.5$, where q stands for quantile. l_1 loss function is less sensitive to large error than squared loss function. The gain of quantile regression may come from context clustering, which reduces large prediction error in each context.

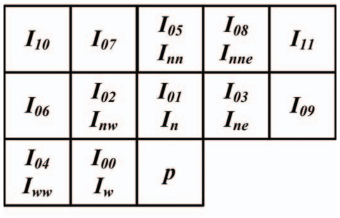
The remainder of this paper is organized as follows. Section 2 presents the design of our algorithm. Section 3 gives the experimental results and analysis. Section 4 concludes the paper.

978-1-4673-0046-9/12/$26.00 ©2012 IEEE

2. ALGORITHM DESCRIPTION

2.1. Context Definition

For each pixel $I[i,j]$ at location (i,j), we define its seven neighboring pixels in Fig. 1.



I_{10}	I_{07}	I_{05} I_{nn}	I_{08} I_{nne}	I_{11}
I_{06}	I_{02} I_{nw}	I_{01} I_{n}	I_{03} I_{ne}	I_{09}
I_{04} I_{ww}	I_{00} I_{w}	p		

Fig. 1. seven and twelve neighboring pixels

I_n, I_w, I_{ne}, I_{nw}, I_{nn}, I_{ww}, I_{nne} mean north, west, north east, north west, north north, west west and north north east respectively. In [2], the following quantities estimate the gradient of the intensity function at current pixel $I[i,j]$.

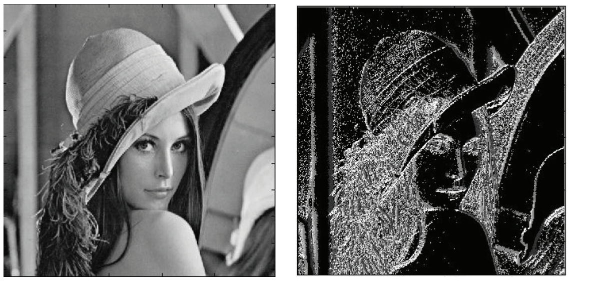
$$\begin{aligned} d_h &= |I_w - I_{ww}| + |I_n - I_{nw}| + |I_n - I_{ne}| \\ d_v &= |I_w - I_{nw}| + |I_n - I_{nn}| + |I_{ne} - I_{nne}| \end{aligned} \quad (1)$$

$d_h - d_v$ is used to detect the magnitude and orientation of edges in the input image. However, we define a more flexible method to represent a pixel's context. For each pixel, we build a vector of ratios of the neighboring pixels, as follows:

$$\begin{array}{ll} \text{horizontal} & \frac{I_w}{I_{ww}}, \frac{I_n}{I_{nw}}, \frac{I_{ne}}{I_n}, \frac{I_{nne}}{I_{nn}}, \\ \text{vertical} & \frac{I_w}{I_{nw}}, \frac{I_n}{I_{nn}}, \frac{I_{ne}}{I_{nne}}, \\ \text{diagonal} & \frac{I_n}{I_w}, \frac{I_{nw}}{I_{ww}} \end{array} \quad (2)$$

The first line in (2) represents how the pixel values vary along the horizontal axis, and the second line represent how they vary vertically. The last two values represent diagonal changes. We cluster the pixels into K groups by K-means. Each group tends to summarize a context in the image. This context classification method is more flexible than the fixed edge types defined in [2], because it allows a number of clusters defined by the user, and could give different weights to pixel ratios. We use ratios of neighboring pixels instead of absolute differences, because the linear regression function is a weighted sum of the neighboring pixels. The ratio gives a more accurate estimate of weights of different contexts than absolute differences.

To evaluate the effectiveness of K-means clustering on defining image context, we apply it to Lena. We set the number of clusters as twenty, and hope to retrieve the smooth areas and the various edges. The clustering result is shown in Fig. 2(b).



(a) Lena

(b) clustering result

Fig. 2. Lena (a) original image, (b) clusters by k-means

Fig. 2(b) shows the smooth area are clustered as the same context and most of the edges are found by K-means. Especially edges with different directions are clustered into different contexts.

2.2. Boundary Pixels

We set the neighboring pixels outside the image to have 0. The pixels at the boundary have some or all of its neighboring pixels as 0. This phenomenon makes boundary pixels close to each other in K-means clustering and likely to be clustered together. The partially zero vectors in a context will lead to singularity problem when building regression functions. To avoid this trouble, we predict the boundary pixels using the approach proposed in CALIC [2].

2.3. Optimal Number of Clusters

When clustering is applied to image coding, overhead bits are required to transmit the cluster information for the decoder to recover the context. As the number of clusters approaching optimum, the residual information decreases, while the information to represent clusters increases. We are looking for a cluster number that achieves the lowest sum of two information amounts.

Assume K is the number of clusters. To represent which cluster each pixel belongs to, we need to transmit the center of each cluster. The decoder first obtains the vector of the ratios in (2), then calculates its distance to each center to determine the closest as its context. For each cluster, we need to transmit nine parameters to represent the center. Thus, in order to represent the contexts, $9K$ parameters have to be sent.

2.4. Building Context Based Pixel Predictor

For a pixel p, the twelve neighboring pixels as shown in Fig. 1 form a feature vector involved in the regression function.

In each of the K contexts, a prediction function is derived by LS, quantile regression and lasso regression. Suppose presented with a dataset $(x_i, y_i), i = 1, 2, \ldots, N$, x_i is a twelve dimensional vector of independent variables and y_i

is dependent observed data. The regression function has the form $f(x, \beta)$, where β holds twelve adjustable parameters.

2.4.1. *Least Square*

LS looks for a linear function $f(x, \beta)$ that minimizes the sum of squared error.

$$\min_{\beta \in \mathcal{R}} \sum_i (y_i - f(x_i, \beta))^2 \tag{3}$$

Solving (3), we obtain an estimate of the conditional expectation function $E(y|x)$.

2.4.2. *Quantile Regression*

In quantile regression, instead of minimizing $\sum_i (y_i - f(x_i, \beta))^2$, we minimize

$$\min_{\beta \in \mathcal{R}} \sum_i \mathcal{C}(y_i - f(x_i, \beta)) \tag{4}$$

where $\mathcal{C}(y_i - f(x_i, \beta))$ is the loss function defined as:

$$\begin{aligned} &\mathcal{C}(y_i - f(x_i, \beta)) \\ &= \begin{cases} q(y_i - f(x_i, \beta)) & \text{if } y_i \geq f(x_i, \beta) \\ (1-q)(f(x_i, \beta) - y_i) & \text{otherwise} \end{cases} \\ &\text{where } 0 < q < 1 \end{aligned} \tag{5}$$

The solution to (4) gives qth quantile of y. If we let $q = 0.5$, solving (4) gives an estimate of the conditional median of y. Quantile regression can be reformulated as a linear programming problem, which can be solved by simplex method or interior point method [8].

2.4.3. *Lasso Regression*

The lasso regression solves the following optimization problem.

$$\begin{aligned} &\min_{\beta \in \mathcal{R}} \frac{1}{2N} \sum_{i=1}^{N} (y_i - f(x_i, \beta))^2 \\ &\text{subject to } \lambda P_\alpha(\beta) \leq t \\ &\text{where } t \geq 0 \end{aligned} \tag{6}$$

where

$$P_\alpha(\beta) = \alpha \|\beta\|_{l_1} \tag{7}$$

The optimization problem can be solved using quadratic programming or more general convex optimization methods. [9] proposed a fast algorithm by cyclical coordinate descent,computed along a regularization path. The method reduces the computation time for image prediction to a few seconds. The l_1-regularized formulation tends to prefer solutions with fewer nonzero parameters, effectively reducing the number of variables upon which the given solution is dependent.

2.5. Evaluation of the Predictor

We sum two information amounts, $H_T + H_R$ bits as the metric to evaluate pixel predictors. H_R is the amount of residual (prediction error) signal's information. For both proposed and compared methods, we apply CALIC's [2] edge-based context separation method to reduce the residual entropy. The equation to calculate H_R is given as:

$$H_R = -\sum_C \sum_{n_{dc} \neq 0} n_{dc} log_2(n_{dc}/N_C) \tag{8}$$

n_{dc} is the number of prediction error $d_C(= -255, \ldots, 255)$ occurrences under context C (K contexts) and N_C is the number of pixels that belong to context C.

H_T includes the amount of information that represents the clusters and the regression parameters in each cluster. $9K$ parameters are needed to transmit the centers of the clusters. For each cluster, we need 12 parameters to transmit the information of regression parameters. In total, there are $21K$ parameters transmitted besides the residual information. Conventionally, LS(12) predictor requires 130 bits as H_T to send 12 parameters. There are $\frac{21K \times 130}{12}$ bits for context based regression algorithms.

As the number of clusters approaches optimum, the residual information H_R decreases, while the overhead H_T increases. The best number of clusters is the one that minimizes $H_T + H_R$. To reduce complexity, we test cluster number that are multiples of 10.

2.6. Complexity

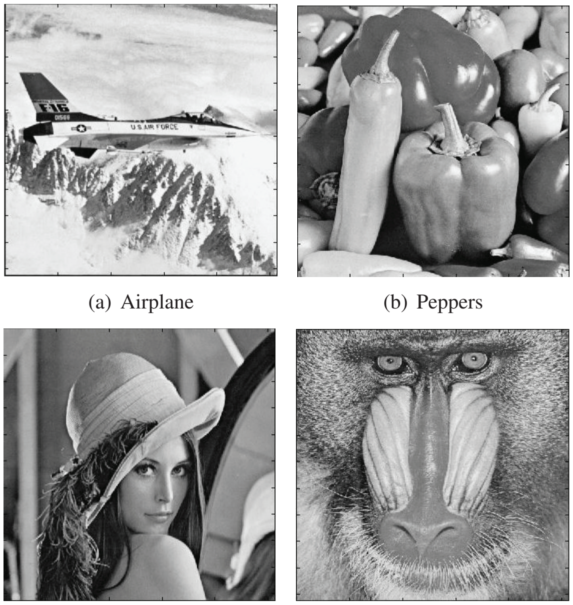
Since quantile and lasso regression have fast solutions, the main complexity of the encoder lies in K-means. The time complexity of K-means is $O(INKd)$ [10], where N is the number of pixels in the image, d is the dimension which equals 12, K is the number of clusters, and I is the number of iterations. At the decoder, the main complexity lies in finding the corresponding context for each pixel. The time complexity for context searching is NKd.

3. EXPERIMENTAL RESULTS

We test our algorithm on four images, Airplane , Peppers, Lena and Balloon (Fig. 3).

We compare the context based regression pixel predictors with two linear predictors (least square (LS)), both 4 and 12 pixel versions, offset inclusive) and two nonlinear predictors (CALIC's gradient adaptive predictor (GAP) and JPEG-LS' median edge detector (MED)).

Tab. 1 shows the experimental results for four test images (all 512x512 pixels, 8 bpp, luminance only) and the total information including H_R and overhead H_T. For LS(4), $H_T = 50$ bits, and for LS(12), $H_T = 130$ bits. $H_T = 0$ for GAP and MED.



(a) Airplane (b) Peppers

(c) Lena (d) Baboon

Fig. 3. Test Images (a) Airplane, (b) Peppers, (c) Lena, (d) Baboon

Table 1. Residual Entropy of the predictors ($H_T + H_R$[bpp])

predictor	Lena	Baboon	Airplane	Peppers
LS(4)	4.290	5.952	3.907	4.634
LS(12)	4.289	5.938	3.891	4.509
GAP(4)	4.167	5.967	3.764	4.569
MED(3)	4.317	6.034	3.837	4.759
C-LS	4.104	5.897	3.705	4.391
C-Quantile	4.091	5.894	3.692	4.375
C-Lasso	4.114	5.898	3.717	4.421

We summarize the context based regression functions in Tab. 2 including its optimal number of clusters K.

Table 2. H_T, H_R[bpp] and optimal number of contexts

predictor	Lena	Baboon	Airplane	Peppers
C-Quantile (N)	20	20	40	30
C-Quantile (H_R)	4.034	5.877	3.658	4.349
C-Quantile (H_T)	0.017	0.017	0.035	0.026

3.1. Discussion

All context based algorithms have better compression performance than conventional methods. C-Quantile is the best among them. Compared with LS, quantile regression infers the conditional median of the predicted value, so is less sensitive to large errors. Since we have classified the image into different contexts, pixels in a particular context tend to give small errors.

4. CONCLUSION

In this paper, we demonstrated that by clustering an image into contexts and train a regression function on each context, we can improve the pixel prediction. The advantage of context based methods lies in the fact that different contexts can better explore the image and reduce pixel randomness in each context, thus reduces prediction error. Our predictor evaluation criterion includes residual information H_R and the predictor's information H_T, which includes parameters to represent the contexts and the regression function for each context. We learn three types of regression function LS, quantile and lasso regression in each context. Context based algorithms all outperform conventional pixel predictors (LS, MED, GAP). They may also be extended to lossy coding and video coding.

Future work includes designing a better and faster clustering algorithm to classify the image pixels, because there are 262144 pixels to train for an 512x512 image, and K-means clustering takes about an hour for $K = 30$. Trying other types of regression analysis or regularization term may further improve the compression rate.

5. REFERENCES

[1] S. Takamura, M. Matsumura, and Y. Yashima, "A study on an evolutionary pixel predictor and its properties," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, 2009, pp. 1921 –1924.

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

[3] M.J. Weinberger, G. Seroussi, and G. Sapiro, "Loco-i: a low complexity, context-based, lossless image compression algorithm," *Data Compression Conference*, pp. 140 – 149, 1996.

[4] JPEG-LS, "Lossless and near-loss compression of continuous tone still images, iso-iec 14495-1," 2000.

[5] G. A. F. Seber, *Multivariate Observations*, John Wiley & Sons, Inc., 2004.

[6] Roger Koenker and Kevin F. Hallock, "Quantile regression," *Journal of Economic Perspectives*, vol. 15(4), pp. 143–156, 2001.

[7] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[8] Roger Koenker, *Quantile Regression*, Cambridge University Press, Cambdige, UK, 2005.

[9] Jerome Friedman, Trevor Hastie, and Robert Tibshirani, "Regularization paths for generalized linear models via coordinate descent," 2009.

[10] Mary Inaba, Naoki Katoh, and Hiroshi Imai, "Applications of weighted voronoi diagrams and randomization to variance-based k-clustering: (extended abstract)," in *Proceedings of the tenth annual symposium on Computational geometry*. 1994, pp. 332–339, ACM.