BEST VIEW SELECTION AND COMPRESSION OF MOVING OBJECTS IN IR SEQEUNCES

Namrata Vaswani and Rama Chellappa

Center for Automation Research
Department of Electrical and Computer Engineering
University of Maryland, College Park
MD 20742, USA
{namrata,rama}@cfar.umd.edu

ABSTRACT

A system for selecting a single best view image chip from an IR video sequence and compression of the chip for transmission is presented. Moving object detection was done using the algorithm described in [1]. Eigenspace classification has been implemented for best view selection. Fast algorithms for image chip compression have been developed in the wavelet domain by combining a non-iterative zerotree coding method with 2D-DPCM for both low and high frequency subbands and compared against existing schemes.

1. INTRODUCTION

Remote surveillance of battlefields is an important component of Future Combat Systems (FCS). Typically, multiple infra-red(IR) sensors are placed in the field for image acquisition from different orientations. The moving target is detected in each frame from each sensor but since the channel bandwidth available for transmission is very low (300bps), all the frames cannot be transmitted. So we develop algorithms for selecting and compressing a single best view image of the target. The bestview selection and compression have to be performed in real-time and on low power computing hardware and hence have to be computationally very simple .

The problem of moving target detection is formulated as one of segmenting an image function using a measure of its local singularity as proposed by Shekarforoush in [1]. Eigenspace classification has successfully been used for pose detection [3] and face recognition applications [2]. We formulate the best view selection problem as a pose matching problem in eigenspace. A wavelet zerotree coding scheme for compression is presented in [4] but since it uses VQ, it cannot be used for our application. [5] presents an embedded predictive wavelet image coder(EPWIC). But again it uses arithmetic coding which is not suitable for hardware implementation. So we have developed algorithms using the Haar wavelet transform followed by non-iterative zerotree coding[4] and 2D-DPCM for all subbands. The computational complexity of these algorithms is only marginally higher than simple scalar quantization (SQ) of the entire image. The performance of the algorithms has been analysed in terms of the bpp, PSNR and the theoretically achievable entropy rate if we were to do entropy coding.

Prepared through collaborative participation in the Advanced Sensors Consortium (ASC) sponsored by the U.S. Army Research Laboratory under the Federated Laboratory Program, Cooperative Agreement DAAL01-96-2-0001. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

2. EIGENSPACE CLASSIFICATION

An eigenspace is constructed offline using different views (back, side and front) of the object in our case army tanks. We work on the assumption that the side view is the 'best view' since it has most of the identifying features (See figure 1(a)). Each image chip which is above a certain size threshold is classified in eigenspace and the one that is closest to the side view is chosen as the best view.

Distance Metric: The distance metric can either be the simple Euclidean distance(ED) or the Mahalanobis distance (MD). The latter gives better results since it gives more weight to those directions where the noise variance is lower.

We have developed a modified Mahalonobis distance metric , the class normalized distance(CND). In this case the distance (along an eigenspace direction) from a particular class is normalized by the variance of the training set members of *that class* only rather than the global variance (eigenvalue) in that direction. This effectively suppresses intra-class variance.

3. WAVELET IMAGE COMPRESSION

The Haar wavelet has been used because of its simplicity in hardware implementation. Also it is suitable for small sized images as in our application. Four different schemes for encoding the wavelet coefficients were implemented and their results compared. In all cases the LL subband was encoded using the 2D Predictive DPCM scheme discussed below.

3.1. 2D Predictive DPCM for LL Subband Encoding

The LL subband contains the maximum information and thus more bits are allocated for its encoding. But it is also the most highly correlated subband and this fact needs to be exploited to maximize compression. A 2D-DPCM scheme is used for encoding the LL subband. The current pixel is predicted based on a linear combination of the causal nearest neighbors. The predicted value of the pixel, $\hat{X}_{n,n}$ is obtained as

$$\hat{X}_{n,n} = l(\bar{Q}) = \bar{w}.\bar{Q} = \sum w_k Q_k \quad (1)$$

The predictor coefficients $\bar{w}$ are calculated to minimize the mean squared prediction error over the set of linear estimators (LMSE) as follows[5]

$$\bar{w} = E[X_{n,n}.\bar{Q}]E[\bar{Q}\bar{Q}^T]^{-1} \quad (2)$$

where $X_{n,n}$ is the pixel to be predicted, Q_i are the quantities based on which the pixel would be predicted(in this case the nearest neighbors, $X_{n-1,n}$ and $X_{n,n-1}$), and $\bar{w}$ are the predictor coefficients. Instead of quantizing the pixel value, the error between the actual and the predicted value $(X_{n,n} - \hat{X}_{n,n})$ is quantized, which requires lesser bits since the error would be much smaller than the original pixel value if the prediction is good. Calculation of LMSE predictor coefficients can be done offline using a set of similar images.

3.2. Scalar Quantization

Simple scalar quantization(SQ) of the wavelet coefficients and 2D-DPCM encoding of the LL coefficient is done. Variable bits are allocated to the subbands based on their variances [4].

3.3. Zerotree Coding

A modification of the zerotree coding(ZT) scheme described in [4] is used. For computational simplicity, the VQ is replaced by a uniform scalar quantizer with variable bits allocated to different subbands. When a coefficient is decided as insignificant, all its descendants are also assumed to be insignificant. Thus only the escape codes [4] for the zerotree root can be transmitted instead of transmitting the entire zerotree. In our implementation we do a run length coding of the zerotree root map before transmission.

3.4. 2D Predictive DPCM on Wavelet Subbands

The residual correlation in the wavelet coefficients can be exploited to design a 2D-DPCM scheme for the LH and HL bands. The prediction for the current pixel is obtained based on its horizontal (for LH) and vertical (for HL) neighbors and the parent coefficient at the same location. But the prediction coefficient for the parent subband was small indicating the low correlation between the subbands and hence in the final version, we used only adjacent pixels for prediction. The performance of this scheme is the worst because a lot of coefficients below the zeroing threshold are actually 'noise' and cannot be predicted by the previous pixel.

3.5. Combined Zerotree and DPCM coding

We propose to combine zerotree coding and the DPCM encoding (ZT/DPCM) of wavelet coefficients to achieve maximal compression. First a simple zerotree coding is applied to the subbands. This is followed by DPCM coding of the 'non-zeroed' coefficients.The value of a 'zeroed' neighbor is predicted as follows. If we predict $C_{x,y}$ based on $C_{x-1,y}$ which is 'zeroed' and the zeroing threshold is T, we estimate $C_{x-1,y}$ as follows

$$S = C_{x-2,y} + C_{x-1,y-1}$$

$$\hat{C}_{x-1,y} = \begin{cases} 0 & \text{if } S = 0 \\ -T & \text{if } S < 0 \\ +T & \text{if } S > 0 \end{cases}$$

This is based on the assumption that since the next coefficient is non-zero, the previous one would be close to the threshold.

DPCM combined with zerotree coding works much better because the noisy coefficients have been set to 'zero' and we do not try to predict their value.

3.6. Magnitude Prediction & Zerotree Coding

As discussed in [5], the correlation between the parent and child coefficients is not too high but the variance of the child coefficient is strongly dependent on the magnitude of its parent and neighboring coefficients. Hence, the magnitude of the child coefficient($C_{x,y}$) is predicted based on the parent coefficient($P_{x,y}$) and nearest neighbors again using DPCM on the magnitudes i.e.

$$|\hat{C}_{x,y}| = w_1|P_{x,y}| + w_2|C_{x-1,y}| + w_3|C_{x,y-1}| \quad (3)$$

where the predictors w_i are obtained using equation 2.

In [5], a computationally intensive algorithm based on this magnitude prediction is used which is not suitable for our application. We propose a very simple scheme in which the child coefficient is scaled by its magnitude prediction and the scaled coefficient($C_{x,y}/|\hat{C}_{x,y}|$) is uniformly quantized. This is equivalent to adaptive non-uniform quantization of the high frequency subbands with a very low computational overhead. Since zerotree coding has been done, it prevents blowing up of values due to almost zero predicted magnitude.

3.7. Analysis

The aim of any compression scheme is to minimize the mean squared error (or maximize the PSNR) and the entropy per pixel (entropy rate, ER). In SQ, we code each pixel independently and hence are not exploiting the correlation in the image and so the entropy rate is higher.

Randomness and hence entropy rate of a pixel will be minimized if it is coded based on all past pixels on which it depends, i.e. (for a 1D signal)

$$h(X_n) > h(X_n|X_{n-1}) > h(X_n|X_{n-1},...,1) \quad (4)$$

If we assume a one step Markov model,

$$h(X_n|X_{n-1}...1) = h(X_n|X_{n-1}) \quad (5)$$

For 2D data (assuming a Markov Random Field model), this translates to $X_{n,n}$ depending only on $X_{n-1,n}$ and $X_{n,n-1}$. Now the quantization MSE will be minimized for a given bit rate if the mean square value of the quantity to be quantized is minimum. Hence instead of quantizing $X_{n,n}$, in 2D Predictive DPCM, we predict a value ($\hat{X_{n,n}}$) based on past values and quantize ($X_{n,n} - \hat{X_{n,n}}$) . $\hat{X_{n,n}}$ is calculated as discussed in equation (1) to minimize $E[X_{n,n} - \hat{X_{n,n}}]^2$ and hence the quantization MSE over all linear estimators. Also for a given quantization step size (fixed MSE), reduced data variance means reduced entropy.

In zerotree coding, the PSNR is higher than simple SQ because the zeroing error is lower than the quantization error for high frequency subbands which are coarsely quantized. Zeroing also reduces entropy since the number of symbols to be compressed is reduced.

Class	CND	ED	MD
tank2	4	10	24
tank6	15	20	35
tank9	15	19	28
btank12	30	44	56
sftank5	17	21	31

Table 1. Eigenspace Classification Results

The 2D MRF model with second order dependencies (correlations) is very good for the LL subband but is not so exact for the wavelet subbands and the prediction fails completely for very small values (only noise). This is the reason why DPCM on wavelet subbands gives the worst PSNR values. Combined zerotree and DPCM (ZT/DPCM) gives best results both for PSNR and entropy rate. The noisy coefficients are zeroed and hence not predicted and thus quantization error remains low. Because of LMSE prediction, the entropy is minimum and zerotree coding further reduces the entropy rate by reducing the number of symbols to be coded.

4. RESULTS

An eigenspace of Front, Side and Back views of various tanks is constructed and the class means for each class are precalculated. In table 1, results for distances from the perfect side view ('tank2') class are shown. Distance of a query tank2(side view) image(figure 1(a)) is minimum while the distance of tank12(back view) (figure1(c)) is much higher. The distances of tank6 and tank9 which are fronto-side views are somewhere in between these two values. Hence tank2 is the 'best view' in this case. As can be seen from the distance values, there is maximum inter-class variation in the CND and hence it is the best distance metric for our application.

The results of compression using the four schemes are shown in table 2. The compressed images are shown in figure 1. Since the original images have been obtained using low quality IR sensors, they are a little blurred and hence the compressed images are also very blurred.

In the table we have compared the total bpp, PSNR, bpp for RLC coding and entropy rate for 3 different images (two from IR tank data and one a visual Lena image) with a total of 0.5bpp being allocated to various subbands proportional to their subband variance. Since for this low value of bpp, the highest frequency subbands get negative bits allocated to them(which are set to zero), the actual bpp obtained is higher than 0.5.

Image	Coder (b=0.5)	Total bpp	PSNR	Entropy Rate (Non-zero)	RLC bpp
lena	ZT/DPCM	0.5066	29.40	0.0851	0.2286
	ZT	0.5066	29.30	0.1542	0.2286
	SQ	0.7947	13.07	0.3156	
	DPCM	0.7947	25.34	0.1508	
tank2	ZT/DPCM	0.5628	31.73	0.0920	0.2757
	ZT	0.5628	31.61	0.2112	0.2757
tank12	ZT/DPCM	0.5232	31.75	0.0880	0.2649
	ZT	0.5232	31.65	0.2045	0.2649

Table 2. The bpp, PSNR[$10 \log_{10} 255^2/MSE$] and entropy rate for 3 sample images using Zerotree coding(ZT),Zerotree & DPCM (ZT/DPCM), Scalar Quantization(SQ) & only DPCM (DPCM) coding

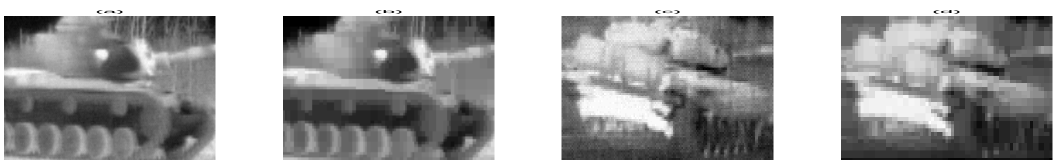

Fig. 1. (a) Tank2(Side view) Image (b) Tank2 compressed by ZT/DPCM (c) Tank12(Back view) (d) Tank12 compressed by ZT/DPCM

As can be seen from the table, the PSNR for ZT coding is higher and the entropy rate (ER) lower than that of simple SQ. DPCM on wavelet coefficients gives the worst PSNR while ZT/DPCM is best both in terms of PSNR and ER. The PSNR value for ZT/DPCM is only marginally higher than the rest because DPCM coding is done only for subbands with more than 2 bpp allocated and at low bit rates this happens only for a few subbands. Also, as can be seen from the value of RLC bpp, almost half the bits are used up in encoding the zerotree information. More efficient binary encoding schemes can be employed to reduce this value and this could considerably improve the bpp. The magnitude prediction scheme has been tested on individual subbands and a significant reduction in MSE has been observed but results for the entire image have not been obtained as yet.

5. CONCLUSIONS

A best view of a detected target chip in an IR sequence is selected for transmission using classification in eigenspace. A new scheme combining non-iterative zerotree coding with 2D-DPCM for LL and also for high frequency subbands has been developed which gives better results than simple scalar quantization both in terms of bpp and PSNR at a marginally increased computational cost. Our compression algorithms are extremely fast and the results can be further improved by using some form of entropy coding (since the entropy rate of our scheme is significantly lower) and replacing run length coding with more efficient binary coding techniques. Also the zeroing thresholds can be calculated for required PSNR values Although the schemes discussed above have been applied to IR sequences, they are equally valid for visual images as well as can be seen from the results on the Lena image.

6. REFERENCES

[1] H.Shekarforoush and R. Chellappa, “A Multi-fractal Formalism for Stabilization and Activity Detection in FLIR Sequences”, *Proceedings of ARL Federated Lab, 4th Annual Symposium*, pp. 305-309, March 2000

[2] M. Turk and A. Pentland, “Eigenfaces for Recognition”, *Journal of Cognitive Neuroscience*, vol. 3, no. 1, 1991

[3] Murase and Nayar, “Visual Learning and Recognition of 3-D Objects from Appearance”, *International Journal of Computer Vision*, vol 14, pp 5-24 1995

[4] Seung-Kwon Paek and Lee-Sup Kim, “A Real-time Wavelet Vector Quantization Algorithm and its VLSI architecture”,*IEEE Trans on CSVT*, vol. 10, pp. 475-489, April 2000.

[5] Robert W. Buccigrossi and Eero P. Simoncelli,“Image Compression via Joint Statistical Characterization in the Wavelet Domain ”, *IEEE Trans Image Processing*, vol. 8, pp. 1688-1700, Dec 1999