Context-based Interframe Coding of MR Images

R. Srikanth
Department of Electrical Engineering
Indian Institute of Science
Bangalore-560012, India.
srikanth@ee.iisc.ernet.in

A. G. Ramakrishnan
Department of Electrical Engineering
Indian Institute of Science
Bangalore-560012, India.
ramkiag@ee.iisc.ernet.in

Abstract

This paper presents a new lossless interframe coding scheme for Magnetic Resonance (MR) images. The existing schemes, like block matching method and uniform mesh-based scheme, are inadeqaute to model the motion field of MR sequence. The above schemes use uniform mesh elements which may comprise multiple motions. We propose a scheme consisting of (a) content-based mesh generation (b) forward motion tracking (c) motion compensation using affine transformation and (d) context-based modeling. By using context-based modeling, intraframe correlation is also exploited in addition to interframe correlation. The obtained average compression ratio of 4.3:1 is better than the values of 4:1, achieved by CALIC, the state of the art lossless intraframe coding scheme and 3:1, by the existing uniform mesh-based interframe coding scheme for MR images. The performance of the existing uniform mesh-based scheme can also be improved by employing context-based modeling.

1. Introduction

Image compression is necessary for efficient archiving and transmission of images. Image compression schemes can be broadly classified as lossy and lossless schemes. A lossy scheme is irreversible in the sense that it cannot faithfully retrieve the original image, whereas, a lossless scheme is reversible. Lossy compression schemes can achieve high compression ratios of the order 10:1 to 30:1, whereas, lossless schemes achieve modest compression ratios of about 2:1 to 4:1. The high compression ratios of lossy schemes are generally at the expense of image quality. Medical imaging generally requires lossless schemes. This is due to the following reasons :

1. Possibility of incorrect diagnosis due to probable loss of useful clinical information caused by lossy compression.

2. Post processing operations like image enhancement may accentuate the degradations caused by lossy compression.

Hence, efficient lossless compression schemes are required for medical images.

Several lossless schemes based on linear prediction and interpolation [5] have been proposed. Recently, context based schemes [10], [3], [2] have gained popularity since they can enhance the performance of linear prediction and interpolation schemes. These schemes exploit the correlation within the image.

Medical image sequence is a data set representing 3-D sampling of some organ. Such sequences contain interframe correlation. If the slice thickness is low (1.5 mm and less), one may get an advantage in trying to exploit interframe correlation. In principle, schemes which exploit both interframe and intraframe correlation are expected to perform better than those based on only intraframe correlation. Roos et al. [5] extended their 2-D hierarchical interpolation scheme to 3-D and found that there is not much gain in extending to the third dimension. They also reported that motion-compensation schemes, based on block-matching algorithms [6], are not efficient for medical image sequences. This could be due to inadequate motion models. Recently, 2-D mesh models based on spatial transformations have gained popularity. Spatial transformations model rotation, translation and scaling of image objects within consecutive frames as opposed to modeling of only the translational motion by block-matching schemes. Nakaya et al. [11] proposed a scheme based on uniform mesh elements (triangles and rectangles). Aria et al. [1] used this model for compressing MR sequences and the residue obtained after motion compensation is compressed by wavelet based zero-tree coder. Hence, this scheme is not suitable for lossless compression. Also, uniform mesh model is inadequate since pixels within each element may have different motions as in block-matching schemes.

In this paper, we propose a nonuniform mesh-based interframe coding scheme for MR sequences. Content-based mesh ensures that multiple motions are avoided within each element. We exploit both interframe and intraframe correlation effectively using spatial transformations and context-based modeling.

2. Pre-processing

A typical MR image consists of two parts :

1. Air part (background)

2. Flesh part (foreground)

The flesh part contains the useful clinical information which needs to be compressed without any loss. On the other hand, the air part does not contain any clinical information. It is only noise and consumes unnecessary bit budget and impairs the performance of a compression scheme. In [2], a scheme is proposed which uses two source models, one for background and the other for foreground, and an improvement in performance is reported. But no justification is given to code the air part as there is no useful information present in it. In this work, we ignore the air part. We generate image masks in such a way that the flesh part is totally included and the pixel values in the air part are made zero. The rest of this section explains an image independent algorithm for mask generation.

Morphological operations can be effectively used to generate image masks, which contain a value of '1' in the foreground and a value of '0' in the background. The original image is then multiplied with these masks to obtain "background noise free" images while keeping the information in the foreground intact. The algorithm for generating the mask is given below:

1. Binarize the image with a threshold value of '15 '. This value has been decided after studying the background characteristics of various images.

2. Holes may be formed within the foreground. Close these holes using morphological 'closing' operation.

3. Background may contain spurious lines. Use morphological 'erode' operation to remove these lines.

4. The above erosion operation also erodes the boundary of the foreground region. To make sure that the mask spans the entire foreground region, use morphological 'thickening' operation to thicken the boundary of the foreground region.

5. Multiply the original image with the resulting binary mask.

6. Fit the smallest bounding box to the resulting image. Send the position and size of the box to the decoder as a side information.

Figure 1 shows an MR image, its mask and the image obtained after multiplication with the mask. Note that this algorithm ensures that the flesh part, which is to be compressed without any loss, remains intact while the background is suppressed.

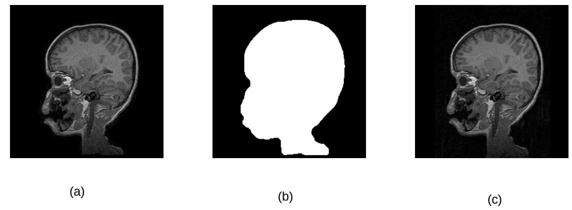

Figure 1: Suppression of background in an MR image using morphological operations. (a) original image (b) the generated mask (c) background suppressed image.

3. Interframe Coding

This section explains in detail the procedures for content-based mesh generation, forward motion tracking and motion compensation of the next frame based on the previous frame.

3.1 Node selection and mesh generation

An image can be motion compensated based on the previous image by estimating the optical flow of each pixel in the image. The corresponding pixel positions need to be sent to the decoder as a side information. This pixel by pixel side information takes a lot of bit budget and impairs the overall performance of the scheme. To reduce the amount of side information, the image is divided into different blocks. In MPEG-1 and MPEG-2 standards, the image is divided into rectangular boxes of equal size. It is assumed that each pixel in a block has the same motion. Only translation motion is assumed and the motion vectors of each block is estimated by block matching algorithm [8]. This translation motion model is inadequate in our case, since it cannot model rotation and scaling of image objects. A promising alternative is motion compensation by 2-D mesh models, which allow for spatial transformations to model translation, rotation and scaling of the image. 2-D meshes can be classified as uniform with equal size elements and nonuniform meshes that adapt to particular scene content. Nakaya et al. [11] proposed motion compensation based on uniform mesh. Aria et al. [1] used this scheme for MR images. This scheme is inadequate as each element may have multiple motions. A better approach is to design a content-based mesh. Wang et al. [9] proposed a scheme based on an optimization frame work. In this work, we propose a simple scheme to generate a content-based mesh. This method consists of selection of node points at salient positions followed by delaunay triangulation. The procedure is explained below :

1. Label all pixels as "unmarked".
2. Find the spatial edge map of the image using "Canny" edge detector.
3. Select a pixel as a "node" if it is "unmarked" and falls on a spatial edge and is sufficiently away from all the previously marked nodes.
4. Go to step-3 until required number of nodes are generated.

The basic idea is to (i) place node points in such a way that mesh boundaries align with object boundaries, (ii) maintain the distance between two node points in such a way that the nodes span the entire image; (iii) form triangular elements by delaunay triangulation.

3.2 Motion Estimation

In uniform mesh coding, an uniform mesh is laid on the current frame and the corresponding nodes in the previous frame are estimated. This procedure is called "backward tracking". In the nonuniform mesh coding, backward tracking cannot be employed since the positions of the node points need to be sent to the decoder. To reduce the side information, we select nodes on the previous frame (which is available both at encoder and decoder) and estimate the corresponding nodes on the current frame. This procedure is called "forward tracking".

We use block-matching algorithm to estimate the motion of each node. Take a 16×16 block with the node as the center. We assume that the maximum displacement in any direction of each node is not more than 5 pixels. Move the 16×16 block in the next frame within a region of 5×5 and take the position with minimum mean square difference as the corresponding node point in the next frame. Send the difference between the two positions to the decoder as a side information. Repeat the above procedure for all the nodes and the triangular elements are deformed accordingly.

3.3 Affine Model

The motion field within each triangular element can be represented by an affine model with six parameters $a_{i1}, a_{i2}, a_{i3}, a_{i4}, a_{i5}$ and a_{i6}, where i is the element index.

Let x, y denote the coordinates of a pixel in i-th element and u, v denote the corresponding pixel in the previous frame. Then,

$$u = a_{i1}x + a_{i2}y + a_{i3}$$
$$v = a_{i4}x + a_{i5}y + a_{i6}.$$

We need to find the above six unknown parameters for each triangular element. For each element, we have correspondence between the vertices of a triangular element in the current frame with the vertices of the corresponding triangular element in the previous frame. Hence, we have six equations in six unknowns which can be easily solved for the affine parameters.

3.4 Motion Compensation

Let $im(x, y, 1)$ and $im(x, y, 2)$ denote the previous and current frames, respectively. Let $ip(x, y)$ be the predicted image of $im(x, y, 2)$ based on $im(x, y, 1)$. $im(x, y, 2)$ is compensated based on $im(x, y, 1)$ as follows:

Scan the triangular elements in $im(x, y, 2)$ one after another. Transform the pixel coordinates (x, y) in each element by using the above affine transformation. Let (u, v) be the corresponding coordinates in the previous image. Generally (u, v) will be real numbers. Find the pixel value at (u, v) by bilinear interpolation as given below:

$$\begin{aligned} ip(x,y) &= \lfloor (1-\alpha)[(1-\beta)\ im(X,Y,1) + \\ &\quad \beta\ im(X+1,Y,1)] + \alpha[(1-\beta) \\ &\quad im(X,Y+1,1) + \beta\ im(X+1,Y+1,1)]\rfloor \end{aligned}$$

$$resd(x,y) = im(x,y,2) - ip(x,y)$$

where (X, Y), (α, β) are the integral and fractional part of the coordinate (u, v), respectively and $\lfloor . \rfloor$ represents the rounding operator. The residue, $resd(x, y)$, is entropy coded and sent to the decoder.

The prediction step does not completely remove the redundancy in the image. This is due to the errors in motion model and estimation of motion of node point. To improve the prediction, we employ the procedure adopted in CALIC [10], which is a state of art lossless compression scheme. It employs gradient adjusted prediction (GAP) and context-based coding. By employing these in interframe coding, we can exploit both intraframe and interframe correlation and can reduce the above residue.

Estimate the gradients in image by

$$\begin{aligned} d_h &= |im(x-1,y,2) - im(x-2,y,2)| + \\ &\quad |im(x,y-1,2) - im(x-1,y-1,2)| + \\ &\quad |im(x,y-1,2) - im(x+1,y-1,2)| \\ d_v &= |im(x-1,y,2) - im(x-1,y-1,2)| + \\ &\quad |im(x,y-1,2) - |im(x,y-2,2)| + \\ &\quad |im(x+1,y-1,2) - im(x+1,y-2,2)|. \end{aligned}$$

where, d_h, d_v are the estimates of horizontal and vertical gradients, respectively.
The prediction $ip(x,y)$ is improved by (refer to CALIC [10]) using the above estimated gradients as follows:

IF $((d_v - d_h) > 80)$ {sharp horizontal edge}
$ip(x,y) = im(x-1,y,2)$
ELSE IF $((d_v - d_h) < -80)$ {sharp vertical edge}
$ip(x,y) = im(x,y-1,2)$
ELSE{

$$ip(x,y) = \lfloor (1-\alpha)[(1-\beta)\ im(X,Y,1) + \beta\ im(X+1,Y,1)] + \alpha[(1-\beta)\ im(X,Y+1,1) + \beta\ im(X+1,Y+1,1)] \rfloor$$

IF $((d_v - d_h) > 32)$ { horizontal edge}
$ip(x,y) = \lfloor \frac{1}{2}[ip(x,y) + im(x-1,y,2)] \rfloor$
ELSE IF $((d_v - d_h) > 8)$ {weak horizontal edge}
$ip(x,y) = \lfloor \frac{1}{4}[3ip(x,y) + im(x-1,y,2)] \rfloor$
ELSE IF $((d_v - d_h) < -32)$ {vertical edge}
$ip(x,y) = \lfloor \frac{1}{2}[ip(x,y) + im(x,y-1,2)] \rfloor$
ELSE IF $((d_v - d_h) < -8)$ {weak vertical edge}
$ip(x,y) = \lfloor \frac{1}{4}[3ip(x,y) + im(x,y-1,2)] \rfloor$
}.

The prediction error strongly correlates to the smoothness of the image around the predicted pixel $im(x,y,2)$. To model this correlation, formulate an error energy estimator

$$\Delta = d_h + d_v + |e_w|,$$

where $e_w = im(x-1,y,2) - ip(x-1,y)$ (previous prediction error) and quantized into eight levels. The residue is classified into one of these bins. The above bin levels and threshold settings for GAP are taken from [10]. In addition to this, texture contexts can be formed to capture the texture patterns and behaviour of error. Form the compound context C by combining texture contexts and energy contexts. The texture contexts are formed with 3 causal neighbors in the current frame and 5 neighbours in the previous frame as follows:

$$\begin{aligned} C \;=\; & [im(x,y-1,2)\ im(x-1,y,2) \\ & im(x-1,y-1,2)\ im(x-1,y+1,1) \\ & im(x,y,1)\ im(x,y+1,1) \\ & im(x+1,y,1)\ im(x+1,y+1,1)]. \end{aligned}$$

Quantize C into an 8-ary binary number by using $ip(x,y)$ as a threshold:

$$C(k) = \begin{cases} 0, & \text{if } C(k) \geq ip(x,y) \\ 1, & \text{otherwise.} \end{cases}$$

Classify the error into one of the compound contexts (α,β), where α is the energy context and β is the texture context. These compound contexts are formed by combining the energy contexts and texture contexts. This can be viewed as product quantization of two independent image features. Accumulate the error in each context and maintain the number of occurrences of each context. The mean of errors in each context is the most likely error. Add this mean error as a bias to the earlier prediction. Let $ipc(x,y)$ denote the corrected prediction and is given by

$$\begin{aligned} ipc(x,y) &= ip(x,y) + \lfloor bias \rfloor \\ resdc(x,y) &= in(x,y,2) - ipc(x,y) \end{aligned}$$

where $bias = error(\alpha,\beta)/N(\alpha,\beta)$, where $N(\alpha,\beta)$ is the number of occurrences and $error(\alpha,\beta)$ is the accumulated error of the compound context (α,β) and $resdc(x,y)$ is the error after the improved prediction.

To be able to repeat the process at the decoder, take the accumulated errors up to previous error. This is similar to feedback with a time delay of one unit. Update the errors and counts of each context.

In addition to improving the estimated value of $im(x,y,2)$, we can predict the sign of the residue by using the estimated mean of the present context. The sign is predicted as follows:

IF $\{error(\alpha,\beta) < 0\}$ send $-resdc(x,y)$
ELSE send $resdc(x,y)$.

At the decoder, the reverse operation can be done by maintaining the same context errors and counts. The sign prediction helps in reducing the entropy of the residue since the uncertainity in the sign bit reduces. We classify the residue $resdc(x,y)$ into eight energy contexts as described above and use arithmetic coding in each context to further compress the residue.

4. Results and Discussion

We compare the proposed ”Context-based interframe coding using nonuniform mesh” with CALIC and uniform mesh-based interframe coding. We have found that the performance of uniform mesh-based coding improves by incorporating context-based coding as explained in the previous section. We compare the performance of our scheme with the original uniform mesh-based coding proposed by Nakaya et al [11] and Aria et al [1]. We have used 256×256, 8-bit MR sequences with slice thickness of 1 mm provided by National Institute of Mental Health and NeuroSciences (NIMHANS), Bangalore.

frames	calic	uniform	context based uniform	non-uniform	context based nonuniform
1, 2	5.71	3.97	5.28	5.51	5.86
2, 3	5.34	3.82	5.06	5.25	5.58
15, 16	3.62	2.79	3.49	3.6	3.78
16, 17	3.62	2.72	3.45	3.55	3.76
28, 29	3.28	2.54	3.14	3.3	3.48
29, 30	3.25	2.47	3.07	3.23	3.39

Table 1: Compression ratios. (Note: "frames 1, 2" means that frame 2 is compensated based on frame 1.)

Table-1 compares the performances (compression ratios) of the above mentioned schemes. The results include the side information for motion vectors. The compression ratios are calculated as follows:

$$compressionratio(CR) = \frac{256 * 256 * 8}{si + nob}$$

where si is side information and nob is the number of bits for residue after arithmetic coding.

Figure 2 shows the original two consecutive MR images and the direct pixel to pixel differences of the original images. Figure 3 shows nonuniform mesh on slices 1 and 2. Figure 4 shows the residues for the uniform mesh-based tecniques without and with context-based modeling and nonuniform mesh-based motion compensation schemes with context-based modeling. Clearly, the modified scheme exploits the intra and inter frame correlation more effectively and reduces the entropy of the residues.

The following reasons may account for the superior performance of the proposed scheme:

1. The uniform mesh model is inadequate since each element may have multiple motions.
2. CALIC effectively exploits intraframe correlation but not interframe correlation.
3. By incorporating context based models in interframe coding (both uniform and nonuniform), both inter and intraframe correlation are exploited.
4. We generate nonuniform mesh in such a way that only the object in the image is meshed and the air region is left out. This straightaway improves the performance of the scheme. This kind of mesh coding can be considered as "object based coding" employed in MPEG-4 and this is achieved without any additional shape information.

5. Conclusion

A lossless context-based nonuniform mesh-based coding is proposed. Both intra and interframe correlation can be effectively exploited by employing context-based models in mesh-based coding schemes. The proposed scheme can also be used for lossy compression schemes. Since residue contains very little information, it can be quantized coarsely without degrading the quality of the image. Hence high compression ratios can be achieved.

References

[1] Aria Nosra, Nader Mohsenian, Michael T.Orchard and Bede Liu, "Interframe Coding of Magnetic Resonance Images", IEEE transactions on Medical Imaging, vol.15, no.5, 639-647, October, 1996.

[2] Marius Midtvik and Ingvil Hovig, "Universal Coding of MR images", IEEE transactions on Medical Imaging, vol.18, no.9, pp.795-800, September, 1999.

[3] T.V.Rambadran and Keshi Chen, "The Use of Contextual Information in the Reversible Compression of Medical Images", IEEE transactions on Medical Imaging, vol.11, no.2, pp. 185-197, June, 1992.

[4] J.J.Rissanen, "Universal Coding, information, prediction and estimation", IEEE transactions on Information Theory, vol. IT-30, pp.629-636, 1984

[5] P.Roos and M.A.Viegever, "Reversible 3-D decorrelation of medical images", IEEE transactions on Medical Imaging, vol.12, no.3, 413-420, 1993.

[6] P.Roos and M.A.Viegever, "Reversible Interframe Compression of Medical Images: A Comparison of Decorrelation Methods", IEEE transactions on Medical Imaging, vol.10, no.4, 538-547, December, 1991.

[7] Stephen Wong, Loren Zaremba, David Gooden, "Radiologic Image Compression - A Review", Proceedings of the IEEE, vol.83, no.2, pp. 194-219, Febrauary 1995.

[8] A.M.Tekalp, Digital Video Processing, Englewood Cliffs, Printice-Hall, 1995.

[9] Yao. Wang and Ouseb Lee, "Active Mesh - A feature Seeking and Tracking Image Sequence Representation Scheme", IEEE transactions on Image Processing, vol.3, 610-624, September, 1994.

[10] Xiaolin Wu and Nasir Memon, " Context-Based, Adaptive, Lossless Image Coding", IEEE transactions on Communications, vol.45, no.4, pp 437-444, April, 1997.

[11] Yuichiro Nakaya and Hiroshi Harashima, "Motion Compensation Based on Spatial Transformations", IEEE transactions on Circuits and Systems for Video Technology, vol.4, no.3, pp 339-354, June, 1994.

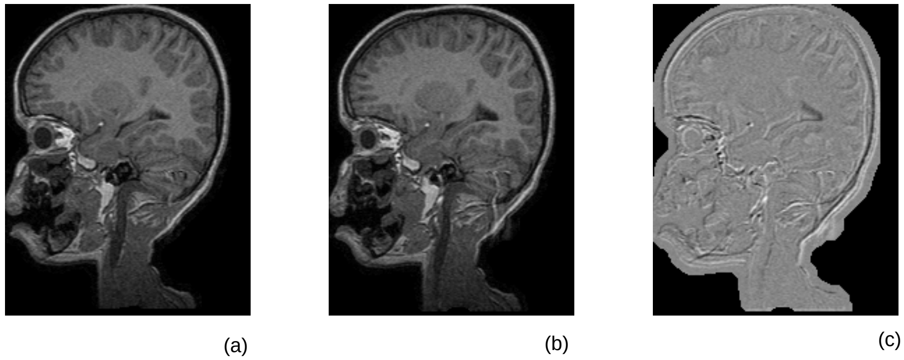

Figure 2: Correlation between successive slices. (a) slice 1 (b) slice 2 (c) direct difference between pixels of slices 1 and 2.

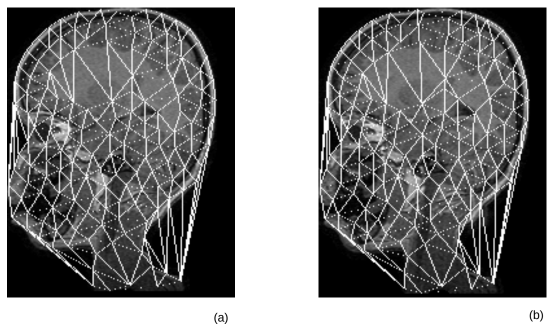

Figure 3: Nonuniform mesh on (a) slice 1 (b) slice 2

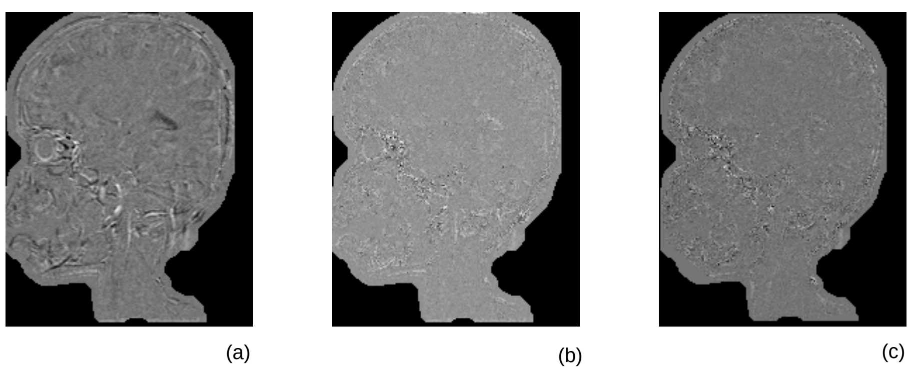

Figure 4: Performance comparison. Residue after (a) uniform mesh-based compensation (b) context and uniform mesh-based compensation (c) context and nonuniform mesh-based compensation. The residues have been level shifted to include both negative and positive differences.