Motion Adaptive Deinterlacing of Video Data with Texture Detection

Yanfei Shen, Dongming Zhang, Yongdong Zhang and Jintao Li, Member IEEE

Abstract — *In this paper, we propose a motion adaptive deinterlacing method with texture detection for scan-rate conversion of video data. The basic idea of this method is to classify the missing pixel to four different regions including motion smooth region, motion texture region, static smooth region, and static texture region based on the results of motion detection and texture detection, then four different interpolation methods are used to produce the missing pixel. Extensive simulations conducted for video sequences indicate that the performance of the proposed method is superior to the previous methods based on the direct merging method, bilinear interpolation, ELA, and vertical temporal median method. Moreover, our proposed method requires lower computational complexity and provides better object and subject performance*[1].

Index Terms — Deinterlacing, Video processing, Texture Detection, Scan-rate conversion

I. INTRODUCTION

As the video industry transitions from analogy to digital, more and more video processing equipment will also need to transition from analogy to digital. The current analogy television standards, NTSC, PAL, etc., are based on interlaced formats. As these standards transition to digital, the demand for progressive material will increase, causing a directly proportional increase in the demand for video processing products with high quality deinterlacing, which converts interlace sequences to progressive sequences.

There are many deinterlacing methods, commonly grouped in two main categories: motion compensated and non-motion compensated methods. Motion compensated (MC) method provide the highest reconstruction quality, but they are computationally more expensive because they require the estimation of two-dimensional motion vector and pixel shifting calculations. On the other hand, non-motion compensated (non-MC) methods are cheaper and can achieve a good compromise between performance and complexity. An extensive review of deinterlacing technology is made by Gerard de Haan in [1].

The Simplest non-MC methods are based on time or space replication, the missing lines of the actual field replicate the lines form the previous field (temporal replication) or replicate the lines of the actual field (spatial replication). Better results are achieved by averaging known data using interpolation instead of replication. Although non-MC is simpler, the line crawling will occur in the motion area.

The motion compensated methods utilize motion estimation to find the most similar blocks in the neighboring fields and calculates its motion vectors. Then a new field is reconstructed from neighboring field. Block matching requires a large buffer size to locate the current macroblock and reference macroblock. Additionally, it also needs a lot of computational power to calculate the sum of absolute difference value. Although MC methods have greater potential to produce better results, they are also more complex since they require a motion estimation engine.

In this paper a motion adaptive deinterlacing method with texture detection for video format conversion is proposed. It consists of advanced motion detection method based on pixel, spatial filter for smooth region, two different vertical temporal filters for static texture region and motion region. In section II, we first review some previous deinterlacing methods briefly. In section III, the proposed method is described in detail. We show the extensive experimental results in section IV. The conclusion is made in section V.

II. SUVERY OF DEINTERLACING TECHNIQUES

In this section, several previous deinterlacing methods are briefly described that include conventional edge-based line-averaging (ELA) method, Vertical temporal (VT) median filtering method, motion adaptive deinterlacing method, and motion compensated method.

A. *The ELA method*

The ELA method utilizes directional correlation among neighbor pixels to linearly interpolate the missing pixel. A 3x3 window is used as shown in Fig.1.

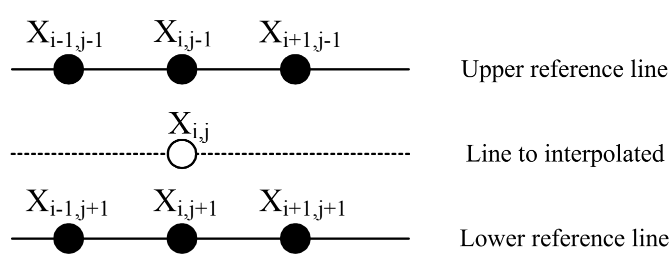


Fig. 1. 3x3 window for ELA method

$$\begin{cases} a = \left| X_{i-1,j-1} - X_{i+1,j+1} \right| \\ b = \left| X_{i,j-1} - X_{i,j+1} \right| \\ c = \left| X_{i+1,j-1} - X_{i-1,j+1} \right| \end{cases} \quad (1)$$

[1] This work was supported in part by the National Natural Science Foundation of China Under Grant No.60302028

Yanfei Shen, Dongming Zhang, Yongdong Zhang, and Jintao Li are with the institute of computing technology, Chinese Academy Sciences.(e-mail: syf@ict.ac.cn , dmzhang@ict.ac.cn, zhyd@ict.ac.cn, jtli@ict.ac.cn)

Contributed Paper
Manuscript received October 10, 2006 0098 3063/06/$20.00 © 2006 IEEE

$$X_{i,j} = \begin{cases} (X_{i-1,j-1} + X_{i+1,j+1})/2, if \min(a,b,c) = a \\ (X_{i,j-1} + X_{i,j+1})/2, if \min(a,b,c) = b \\ (X_{i+1,j-1} + X_{i-1,j+1})/2, if \min(a,b,c) = c \end{cases} \quad (2)$$

According to Fig. 1., the ELA method can be defined in the equation (1) and (2).

The ELA method provides good performance in most cases, however, due to misleading edge directions, often interpolation errors becomes larger in the areas of high frequency components. These unwanted artifacts significantly deteriorate visual quality. In order to remedy these problems and improve the ELA method, several approaches were proposed. In [2], the ELA method was combined with line doubling method for performance improvement. In [3], the ELA method was extended by taking horizontal edges into consideration. In [4], the ELA method was extended by introducing additional measurements for estimating the edge direction more correctly.

B. Vertical temporal median filtering method

VT median filtering method for deinterlacing can implicitly adapt to motion or edges, the simplest version is the three tap VT median [5], as illustrated in Fig. 2.

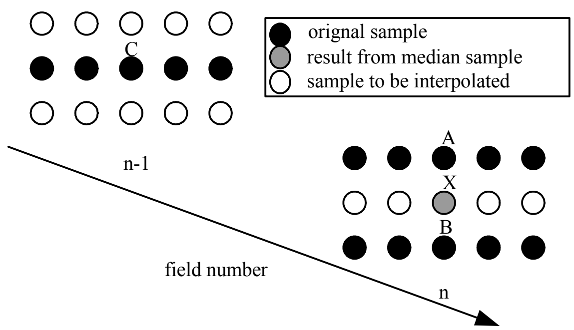


Fig.2. VT median filtering

The interpolated samples are found as the median luminance value of the vertical neighbors (A and B), and the temporal neighbor in the previous field (C):

$$X = med(A,B,C) \quad (3)$$

Where $med(A,B,C)$ is defined as

$$med(A,B,C) = \begin{cases} A, (B < A < C) \vee (C < A < B) \\ B, (A \le B \le C) \vee (C \le B \le A) \\ C, otherwise \end{cases} \quad (4)$$

The underlying assumption of VT median filtering method is that, in case of stationary, X is likely to have a value between that of its vertical neighbors in the current field, which results in temporal interpolation. However, in case of motion, intra field interpolation often results, since then the correlation between the samples in the current field is likely to be the highest. Median filtering automatically realizes this intra/inter switch on pixel basis. The major drawback of VT median filtering for deinterlacing is that it distorts vertical details and introduces alias.

C. Motion adaptive deinterlacing method

The motion adaptive deinterlacing method combines the advantages of both intra-field deinterlacing and inter-field deinterlacing. It detects the motion areas firstly, and then adopts intra-field deinterlacing in motion areas and inter-field deinterlacing in static areas. The block diagram is shown in Fig.3 [6]. High-resolution and flicker-free picture can be realized in both static area and motion area. The image quality is the same as that of intra-field interpolation. Motion adaptive deinterlacing relies on accurate motion detection. Any erroneous detection will cause artifacts as spots in the video. Based on the previous work, a motion adaptive deinterlacing method with texture detection is presented in [7] where the consistent of texture is considered.

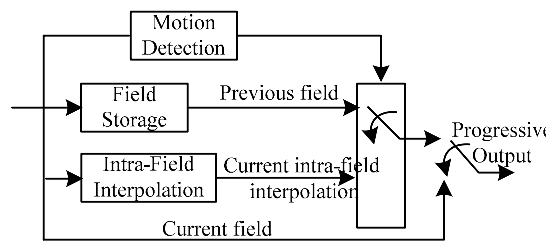


Fig.3. Motion adaptive deinterlacing

D. Motion compensated method

The motion compensated method utilizes motion estimation to find the most similar blocks in the neighboring fields and calculates its motion vectors [8]. Then a new field is reconstructed from neighboring field. Block matching needs a large buffer size to locate the current macroblock and reference macroblock. Additionally, it also needs a lot of computational power to calculate the Sum of Absolute Difference (SAD) value. Although motion compensated methods have greater potential to produce better result, they are also more complex since they require a motion estimation engine.

III. THE PROPOSED METHOD

In order to improve both object quality and subject quality of deinterlaced video, we propose a motion adaptive deinterlacing method with texture detection for video format conversion, the block diagram of the proposed method is shown in Fig. 4., The interpolated pixel is firstly classified based on the results of the motion detection and texture detection, then four different deinterlacing methods are respectively used to produce the interpolated pixels belonging to motion smooth region, motion texture region, static smooth region or static texture region.

A. Motion detection

Motion information is calculated by using the difference of the luminance of pixels in three neighbor reference fields [9]. Fig.5 shows how motion information is extracted from neighbor reference fields, where x and y is the relative field row and column coordinates, t is the field period. The shadow rectangle represents the interpolated pixel.

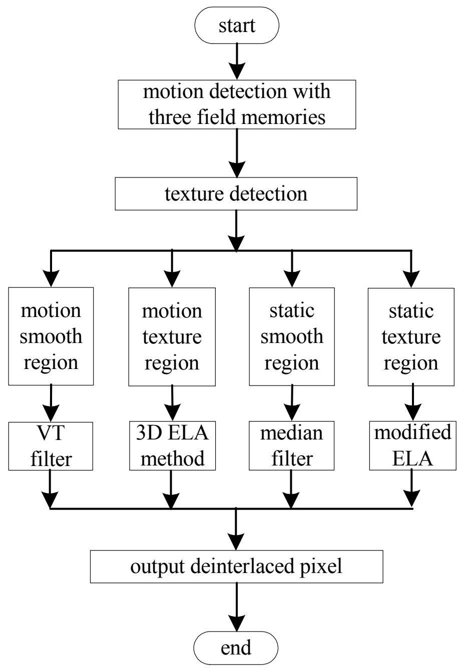


Fig.4. Block diagram of proposed deinterlacing method

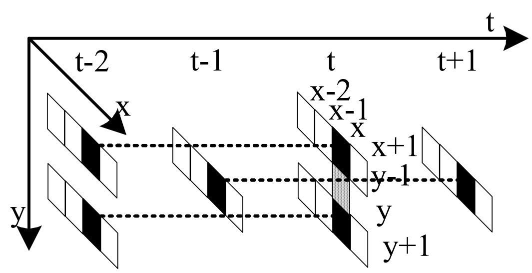


Fig.5. Motion detector with three field memories

In order to estimate the motion information of the interpolated pixel, the pixel difference is defined as follows:

$$d(x, y-1) = abs(p(x, y-1, t) - p(x, y-1, t-2)) \quad (4)$$

$$d(x, y+1) = abs(p(x, y+1, t) - p(x, y+1, t-2)) \quad (5)$$

$$d(x, y) = abs(p(x, y, t+1) - p(x, y, t-1)) \quad (6)$$

and

$$D = \max(d(x, y-1), d(x, y+1), d(x, y)) \quad (7)$$

Where D is the largest of absolute pixel difference, motion decision md is made based on the following equation:

$$\begin{aligned} &md(x, y, t) = 0 \ \ if \ D < th \\ &md(x, y, t) = 1 \ \ \text{otherwise} \end{aligned} \quad (8)$$

Where th is threshold, which is decided by experiment, $md(x, y, t) = 0$ represents that the interpolated pixel belongs to stationary region, otherwise the interpolated pixel belongs to motion region.

B. Texture detection

The region that the interpolated pixel belongs is classified into smooth region and texture region. The variance of pixel around the interpolated pixel is calculated, as shown in Fig.6.

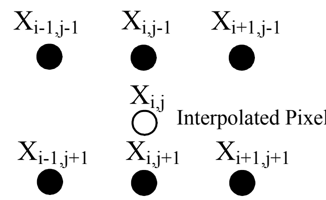


Fig.6. Interpolated pixel and its neighbor pixel

$$Var = \sum_{k=-1}^{1} ((x_{i-k,j-1} - avg)^2 + (x_{i-k,j+1} - avg)^2) \quad (9)$$

$$avg = (\sum_{k=-1}^{1} (x_{i-k,j-1} + x_{i-k,j+1})) / 6 \quad (10)$$

If Var is larger than threshold, which is decided by experiment, the region is classified to texture region, otherwise, the region is classified to smooth region.

C. Four deinterlacing methods for different regions

Based on the results of motion detection and texture detection, four different filters for deinterlacing are used. In this section, these filters are introduced respectively.

1) Vertical temporal filter for motion smooth region

For motion smooth region, vertical temporal filter is the best linear approach. The filter is designed such that the contribution from the neighbor fields is limited to higher vertical frequencies. As a consequence, no motion artifact is present for smooth region. The interpolated pixel is calculated by using equation (11):

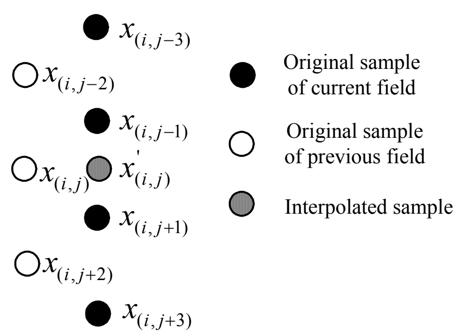


Fig.7. Vertical temporal filter

$$X'_{(i,j)} = (\sum_{k=-3}^{3} x_{(i,j+k)} h(k)) / 18 \quad (11)$$

$$h(k) = \{1, -5, 8, 10, 8, -5, 1\} (k = -3, -2, -1, 0, 1, 2, 3)$$

$h(k)$ is the coefficients of vertical temporal filter, original samples and interpolated sample are shown in Fig. 7.

2) 3D ELA method for motion texture region

ELA method is one well known spatial domain interpolation method, which utilizes directional correlation to interpolate a missing line of the current field. The ELA method gives a good result in regions where the edge can be estimated correctly. In order to take motion information into deinterlacing method, the ELA method is extended three dimensions, as shown in Fig. 8.

Six directional changes of temporal domain are calculated as following:

$$D(k,l,t) = \left| x_{(i+k,j-l,n-1)} - x_{(i-k,j+l,n+1)} \right| \tag{12}$$

Where $k = (-1,0,1)$, $l = (-1,1)$, and t represent temporal direction, (i, j) represent the position of interpolated pixel, n represent field number.

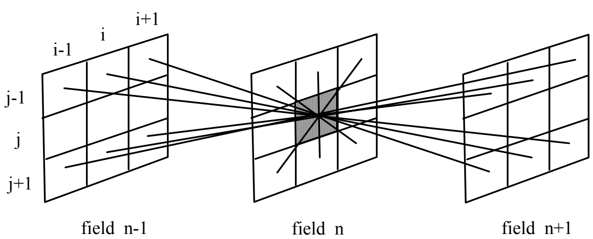


Fig.8. 3D ELA method

Three directional changes of spatial domain are calculated as follows:

$$D(k,l,s) = \left| x_{(i-k,j-l,n)} - x_{(i+k,j+l,n)} \right| \tag{13}$$

Where $k = (-1,0,1)$ and $l = 1$, and t represent spatial direction. The final interpolated direction is defined as following:

$$\theta = \min(D(k,l,t), D(k,l,s)) \tag{14}$$

The interpolated pixel is calculated by averaging two neighbor pixels along the direction θ.

3) Median filter for static smooth region

For static region, the simplest deinterlacing method is directly merging two neighbor fields into one frame. However, because they are sampled at different time, the variety of noise of luminance will degrade the performance of this simple deinterlacing method. Experimentally, vertical temporal median filtering method has been proven to be a very attractive method for deinterlacing in this case. The interpolated samples are found by median filtering the spatial neighbors in vertical direction and the corresponding sample in the previous field, as description in section II.

4) Modified ELA for static texture region

The ELA method uses a directional correlation between the pixels in the adjacent original scan lines. In order to take static characteristics into this deinterlacing method, conventional ELA method are extended.

First, to check the edge direction at pixel $X_{(i,j)}$, as shown in Fig. 9, the difference D is calculated as follows:

$$D(k) = \left| X_{(i-k,j-1)} - X_{(i+k,j+1)} \right| \tag{15}$$

Where $k = \{-2,-1,0,1,2\}$ respectively represents five edge directions.

Second, the minimum difference $D(l)$ is chosen as the highest correctional direction, the candidate pixel value P_{c1} is calculated using the average of two neighbor pixel along chosen direction, as shown in equation (16). Anther candidate pixel value P_{c2} is calculated along the vertical direction using bilinear interpolation, as shown in equation (17).

$$P_{c1} = (X_{(i-l,j-1)} + X_{(i+l,j+1)})/2 \tag{16}$$

$$P_{c2} = (X_{(i,j-1)} + X_{(i,j+1)})/2 \tag{17}$$

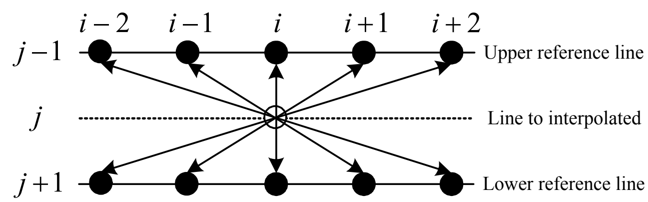


Fig.9. 5x5 window for ELA method

Because the interpolated pixel belongs to static region, the corresponding pixel in the previous field is selected as the third candidate pixel value P_{c3}.

Finally, the median of above three candidate pixels is the last interpolated pixel.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Performance comparison

In order to evaluate the performance of the proposed deinterlacing method, the proposed method is simulated using some well-known video sequences, which are converted from progressive sequences, as shown in Fig. 10. The original 60 frames/second progressive frame is decimated into 60 fields/second interlaced fields [6]. After deinterlacing, the 60 fields/second interlaced fields are reconstructed into 60 frames/seconds. Then we compare the original progressive sequences and the output of different deinterlacing methods.

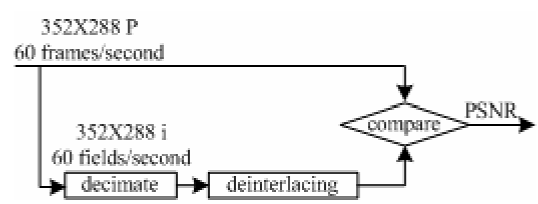


Fig.10. Performance measure method

From PSNR comparison as shown in table I and the subjective quality comparison as shown in Fig.11, the proposed deinterlacing method has better performance than the conventional bilinear, ELA, vertical temporal median filter methods and motion adaptive interpolation with horizontal motion detection for deinterlacing [6]. For large motion test sequences as foreman, mobile and template, the proposed method can accurately detect motion and texture region and adopt vertical temporal filter and 3D ELA deinterlacing method for most region, the proposed method not only use intra field interpolation as 3D ELA for motion texture region, but also use inter field filter as VT filter for motion smooth region, while motion adaptive method only use enhanced directional intra field interpolation with median ELA for motion region, so the proposed method has better performance than simple bilinear method, ELA method and motion adaptive method.

For low motion test sequences as news, akiyo etc, motion adaptive directly use forward pixel value to fill the current field, but there may be noise signal due to different sampling time of forward and current field. Our proposed method can use median filter and modified ELA method to reduce the influence of noise signal based on the characteristic of texture distribution. From the results, our proposed method can gain obvious better performance for these test video sequences than motion adaptive method.

Table I PSNR (dB) RESULTS OF DIFFERENT DEINTERLACING METHOD FOR VARIOUS SEQUENCES

Sequence	Bilinear	ELA	VT Median	Motion Adaptive	Proposed
Foreman	29.90	30.50	32.69	33.07	33.22
Paris	26.62	25.27	30.31	30.22	30.84
Mobile	25.47	23.38	25.46	25.30	26.49
Akiyo	37.91	36.54	42.62	41.63	44.16
Tempete	29.33	27.02	31.21	32.21	32.68
News	33.28	31.27	38.29	36.88	41.01
Container	28.41	27.62	34.25	36.11	38.10
Coastguard	28.54	29.07	29.71	29.45	30.59

(a) Original

(b) Bilinear

(c) ELA

(d) VT median

(e) Motion Adaptive

(f) Proposed

Fig.11 Subjective quality comparison of frame#38 of Paris CIF sequence

B. Analysis of implementation complexity

The implementation complexity of the proposed deinterlacing method is similar to that of motion adaptive method. The data bandwidth and memory requirement for deinterlacing is the same for both methods. In order to detection field motion information, four fields need to be stored including forward-forward field, forward field, current field and backward field, while the motion adaptive method need an additional field buffer for morphological operation. In motion detection and texture detection, six pixels are read form memory each time, so the buffer can be reused. In vertical temporal filter for motion smooth region, seven pixels from current and previous field are required to read from memory. In 3D ELA method for motion texture region, eighteen pixels are required to calculate the interpolated pixel. For median filter of static smooth region, only three pixels are needed. For modified ELA for static texture region, ten pixels are required. So totally eighteen pixel buffer sizes are needed to calculate the interpolated pixel. The comparison of memory access and buffer size is listed in table II, where n means the pixel numbers to be processed. Additionally in our proposed method, morphological operation is not used, so motion detection, texture detection and deinterlacing processing can be pipelined to accelerate in hardware implementation.

It is obvious that motion compensated deinterlacing is the most complex and time-consuming method, while the proposed algorithm possesses less complexity and memory access frequency than the motion compensated deinterlacing, and has the same complexity as intra field deinterlacing and motion adaptive method.

Table II PSNR (dB) RESULTS OF DIFFERENT DEINTERLACING METHOD FOR VARIOUS SEQUENCES

Method	External Memory Access		Buffer Size(byte)
	Read	Write	
Bilinear	n	n	1
ELA	2n	n	10
Motion Compensated	>10n	n	528
Motion Adaptive	5n	n	25
Our Proposed	4n	n	18

V. CONCLUSION

In this paper, a motion adaptive deinterlacing method with texture detection for video data is proposed. It consists of advanced motion detection method based on pixel, spatial filter for smooth region, two different vertical temporal filters for static pixel and motion pixel. Simulation results show that our proposed algorithm gets better performance than exiting method.

ACKNOWLEDGEMENT

The authors wish to acknowledge the anonymous reviews for their comments and suggestions that helped in improving the presentation of the paper.

REFERENCES

[1] Gerard De Haan and Erwin B. Bellers, "Deinterlacing - an overview," *Proceedings of the IEEE*, Vol. 86, No. 9, pp. 1839–1857, 1998.

[2] M.-H. Lee, I.-H. Kim, K.-K. Ryu, and D:I. Song, "A new algorithm for interlaced to progressive scan conversion based on directional correlations and its IC design," *IEEE Trans. Consumer Electronics,* vol. 40, no. 2, pp. 119-125, 1994.

[3] C. J. Kuo, C. Liao, and C. C. Lin, "Adaptive interpolation technique for scanning rate conversion," *IEEE Trans. Circuits systems for Video Technology,* vol. 6, no. 3. pp. 317-321, Jun. 1996.

[4] T. Chen, H. R. Wu, and Z. H. Yu, "An efficient edge line average interpolation algorithm for deinterlacing," in *Proc. Visual Communications and Image Processing,* SPIE vol. 4067, pp. 1551-1558,2000.

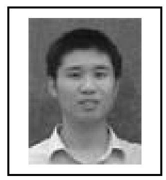
[5] M. Annegarn, T. Doyle, P. Frencken, and D. Hees, "Video signal processing circuit for processing an interlaced video signal," Apr. 1998. U.S. Patent 4740842.

[6] Shyh-Feng, Yu-Ling Chang, and Liang-Gee Chen, "Motion Adaptive Interpolation with Horizontal Motion Detection for Deinterlacing," *IEEE Trans. on Consumer Electronics*, Vol. 49, no. 4, November 2003.

[7] Yang Yuhong, Chen Yingqi, and Zhang Wenjun, "Motion Adaptive Deinterlacing Combining with Texture Detection and Its FPGA Implementation", Proceeding of 2005 IEEE International Workshop on VLSI Design and Video Technology, pp. 316-319, May, 2005

[8] B. Bhatt, D. Birks, and D. Hermreck, "Digital Television: Making it work," *IEEE Spectrum*, vol. 34, pp. 19–28, Oct. 1997.

[9] You-Young Jung, Seungjoon Yang and Pilho Yu, "An Efficient De-interlacing Technique Using Two Types of Motion Information," *IEEE Trans. on Consumer Electronics*, Vol. 49, no. 3, August 2003.



Yanfei Shen received the B.S. degree and the M.S. degree in computer science from the key laboratory of multimedia and network communication, Wuhan University, China, in 1999 and in 2002. He is currently working as associate researcher in Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. His research interests include video coding technology, video processing, digital TV, and associated VLSI architecture.