Directional residue prediction with motion alignment for video coding

K. Zhang, S. Ma, D. Zhao and W. Gao

A directional residue prediction method is proposed, in which motion-aligned neighbouring residues are used in the directional prediction for residues in an inter-block. Corresponding motion estimation strategies are also designed. Experiments show that the bit rate saving can be up to 20% with a negligible decoding complexity increase.

Introduction: Block-based motion compensation (MC) plays a crucial role in predominant video coding standards such as H.264 [1]. Recently, some researchers [2, 3] attempted to combine motion compensation with the intra-prediction adopted by H.264. Although these ideas inspire us greatly, they have not shown an apparent gain yet. In this Letter, we propose a new technique named directional residue prediction with motion alignment (DRP-MA), which also introduces the concept of directional intra-prediction into the residual domain. Different from [3], in which reconstructed residues of the neighbouring blocks are used in the residue prediction, DRP-MA calculates motion-aligned neighbouring residues to predict residues in the current block. Accordingly, an optimised motion estimation (ME) strategy is also designed to take full advantage of the potential capability within DRP-MA.

Correlations among residues: A certain number of correlations still exist among residues after motion compensation [4, 5]. Roughly, the residual signal can be viewed as a stationary first-order Markov process (AR(1) process), just like the original image signal, except that the correlation coefficient is smaller [4]. Moreover, the residual signal even retains a geometric structure to some extent [5]. These properties prompt us to utilise transform coding on residues. Nevertheless, transform coding can only reduce redundancies among residues within an individual block. To further exploit residual correlations beyond the border of a block, we propose the DRP-MA method.

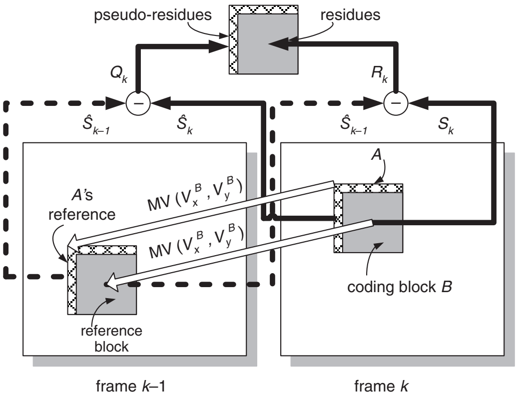

Fig. 1 *Generation of pseudo-residues. Frame k − 1 is reference frame for frame k*

DRP-MA: DRP-MA aims to further attenuate the energy of residues after MC. Intuitively, since the residual signal and the original signal share similar statistical properties [4, 5], we can introduce the directional intra-prediction concept [1] into the residual domain. In general, the procedure of DRP-MA involves three major steps described as follows.

First, the pseudo-residues are generated as depicted in Fig. 1. Suppose the transform block size is $N \times N$, and $S_k(i, j)$ is a pixel in one $N \times N$ block B of frame k, then the residue $R_k(i, j)$ can be formulated as

$$\begin{aligned} R_k(i,j) &= S_k(i,j) - P_k(i,j) \\ &= S_k(i,j) - \hat{S}_{k-1}(i+V_x^B, j+V_y^B), \quad (i,j) \in B \end{aligned} \tag{1}$$

where $P_k(i, j)$ is the motion-compensated prediction for $S_k(i, j)$. $\hat{S}_{k-1}$ represents the reconstructed image of the reference frame $k-1$. (V_x^B, V_y^B) is the motion vector (MV) of block B. Furthermore, A is defined as the one-pixel-width region, adjacent to the left and/or top borders of block B. Since blocks containing A are all coded previously, the reconstructed pixels in A denoted as $\hat{S}_k$ are available when coding block B. By motion-aligning with block B, A's reference is localised and the pseudo-residues denoted by Q_k are generated as

$$\begin{aligned} Q_k(m,n) &= \hat{S}_k(m,n) - \hat{S}_{k-1}(m+V_x^B, n+V_y^B), \\ &(m,n) \in A \end{aligned} \tag{2}$$

Secondly, the directional residue prediction is performed. As shown in Fig. 2, four prediction modes are adopted with the same meaning as H.264/AVC [1], except that they are in the residual domain. In addition, not using residue prediction is also treated as a special mode. Abstractly, we can view the prediction for $R_k(i, j)$ as a functional on the signal Q_k, denoted as $\Psi(i, j, Q_k)$. Then the attenuated residue $r_k(i, j)$ should be

$$r_k(i,j) = R_k(i,j) - \Psi(i,j,Q_k), \quad (i,j) \in B \tag{3}$$

Thirdly, the new motion-compensated prediction $P_k^*(i, j)$ is obtained. To avoid the clipping problem, $r_k(i, j)$ is not yielded by (3) directly. Instead, we let

$$r_k(i,j) = S_k(i,j) - P_k^*(i,j), \quad (i,j) \in B \tag{4}$$

and

$$P_k^*(i,j) = Clip(\hat{S}_{k-1}(i+V_x^B, j+V_y^B) + \Psi(i,j,Q_k)), \quad (i,j) \in B \tag{5}$$

where *Clip* is the clipping operator.

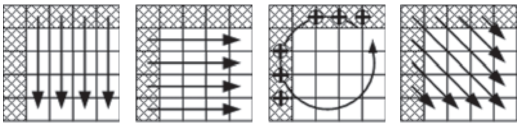
mode 0: vertical　mode 1: horizontal　mode 2: DC　mode 3: diagonal

Fig. 2 *Directional residual prediction modes for 4 × 4 blocks. Squares shaded by crossed lines represent pseudo-residues. Others are ordinary residues*

In practice, blocks in one motion partition, which may sometimes be larger than $N \times N$, are assigned the same residue prediction mode. Therefore, the mode information can be entropy coded and transmitted with the motion vector information of this partition together.

Table 1: Experimental results

Sequences	Size	Frame rate (Hz)	Number of frames	Optimised		Ad hoc	
				Δ*PSNR* (dB)	Δ*Bits* (%)	Δ*PSNR* (dB)	Δ*Bits* (%)
Container	QCIF	15	150	0.55	−10.73	0.48	−9.57
Foreman	QCIF	15	150	0.44	−7.67	0.23	−4.18
Bus	CIF	30	100	0.29	−5.69	0.11	−2.33
Crew	CIF	30	100	0.81	−17.86	0.68	−15.37
Foreman	CIF	30	100	0.54	−11.44	0.39	−8.55
Vectra	CIF	30	50	0.56	−11.30	0.27	−5.67
City	4CIF	60	70	0.46	−12.65	0.30	−8.55
Crew	4CIF	60	70	0.78	−20.03	0.67	−17.37
Soccer	4CIF	60	70	0.32	−8.19	0.20	−5.10

ME strategy: To take full advantage of the potential capability of DRP-MA, we apply an optimised ME strategy for DRP-MA. The five modes, including the one not using residue prediction, stand for five MC methods essentially, thus the optimal MVs for different modes may be different. For each MC method, its specific ME is carried out to obtain the optimal MV for this mode. After the five types of ME are performed, the mode with the minimal ME RD-cost and its optimal MV are selected. In the ME procedure, some reconstructed sample values in region A are unavailable if the motion partition size is larger than $N \times N$, and block B is not at the top-left corner of the partition. To address this problem, predicted sample values in region A are used as approximations.

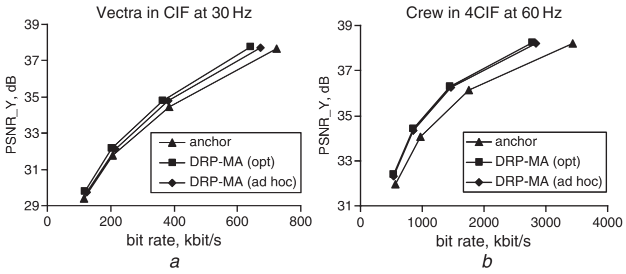

Fig. 3 *RD curves*

Since the computational complexity of the optimised ME strategy is high, we also test an ad hoc ME strategy for a comparison. First, a small set of points are checked to estimate the probable DRP-MA mode. Only the ME for this mode is performed. After the only MV is obtained, the DRP-MA mode with the minimal ME RD-cost on it will be chosen as the final one. Obviously, the ad hoc strategy is not optimal, since the probable mode may be inaccurate and the MV found may not be proper to other modes. Nevertheless, this strategy involves only a little computational overhead on the encoder side. It also provides us with a very simple, fast ME algorithm for DRP-MA. In practical applications, more sophisticated fast ME can be adopted.

Results: We implemented DRP-MA both with the optimised and the ad hoc ME strategies on JM 13.2 [6] and used the original JM codec as an anchor. Experiments were carried out in the H.264 baseline profile for IPPP sequences without CABAC or the 8 × 8 transform. We configured the encoder as follows. One reference frame is used, full search ME with search range 32 is chosen, and only the first frame is the I-frame. The testing QPs are 28, 32, 36 and 40. Table 1 demonstrates the average differences in PSNRs ($\Delta PSNR$) and in bit rates ($\Delta Bits$) [7]. Fig. 3 shows RD curves. It can be seen that DRP-MA improves the RD performance significantly. The average gain is about 0.53 dB and the bit rate saving is up to 20% with the optimised ME. When using the ad hoc ME, an average 0.37 dB gain can also be achieved with only 15% checking times more than the anchor encoder. This result indicates that the proposed DRP-MA is an efficient coding tool and the development of an efficient fast ME algorithm for DRP-MA is feasible. In addition, the computational overheads for DRP-MA are very slight on the decoder side.

Conclusion: A DRP-MA is proposed. Experimental results exhibit a significant gain. Further improvements in coding performance by adopting more prediction modes and sophisticated fast ME algorithms for DRP-MA will be the subject of future work.

9 May 2008
Electronics Letters online no: 20081300
doi: 10.1049/el:20081300

K. Zhang and D. Zhao (*Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, People's Republic of China*)

E-mail: kzhang@jdl.ac.cn

S. Ma and W. Gao (*Institute of Digital Media, Peking University, Beijing 100871, People's Republic of China*)

K. Zhang: Also with Graduate University of the Chinese Academy of Sciences, Beijing 100190, People's Republic of China

References

1 ITU-T: 'Recommendation H.264, http://www.itu.int/rec/T-REC-H.264-200503-I/en accessed March 2005
2 Andersson, K.: 'Combined intra inter prediction coding mode'. VCEG-AD11, 30th meeting, October 2006
3 Chen, S., and Yu, L.: 'Re-prediction in inter-prediction of H.264'. VCEG-AG20, 33rd meeting, October 2007
4 Chen, C., and Pang, K.: 'Hybrid coders with motion compensation', *Multidimens. Syst. Signal Process.*, 1992, **3**, pp. 241–266
5 Tao, B., and Orchard, M.: 'Gradient-based residual variance modeling and its applications to motion-compensated video coding', *IEEE Trans. Image Process.*, 2001, **10**, pp. 24–25
6 Joint Video Team (JVT), reference software JM 13.2
7 Bjontegaard, G.: 'Calculation of average PSNR differences between RD-curves'. VCEG-M33, 13th meeting, April 2001