Signal Processing: *Image Communication* 28 (2013) 581–596

Contents lists available at SciVerse ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Mode dependent down-sampling and interpolation scheme for high efficiency video coding

Qingbo Wu, Hongliang Li *

School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

ARTICLE INFO

Article history:
Received 1 June 2012
Received in revised form
20 November 2012
Accepted 17 March 2013
Available online 27 March 2013

Keywords:
Video coding
High definition
H.264/AVC
Down-sampling

ABSTRACT

In this paper, a mode dependent down-sampling and interpolation scheme is proposed to improve the coding efficiency of the intra prediction module. In the proposed method, we elaborately design the down-sampling structures and interpolation schemes for each directional intra prediction mode by minimizing the spatial prediction distance. The sampled pixels are predicted with a traditional directional intra prediction scheme, and the non-sampled pixels are predicted from the interpolation of their neighboring reconstructed sampling pixels. Both the residuals of the sampled and non-sampled pixels are encoded at last. Experimental results show that the proposed method achieves an average 7.52% bitrate reduction relative to KTA reference software. Since the down-sampling structure and interpolation method is only related to the intra mode, there is no additional overhead at the encoder.

1. Introduction

In the past decades, video coding technologies have greatly promoted the development of digital multimedia contents related industry. Lots of these services, like TV broadcasting, network video, DVD, Free-viewpoint TV (FTV) [1], etc., have deeply changed our lives. International video coding standards play an important role on these advanced technology promotion. Recently, the state-of-the-art video coding standard H.264/AVC [2–4] has been widely used and achieved remarkable success. To satisfy the rapid increasing demand for high-definition (HD) and ultra-HD (UHD) video contents, a higher requirement for more efficient video coding has been brought forward.

Recently, the next generation video coding standard named High Efficiency Video Coding (HEVC) is stepping up in development. In order to better adapt to HD contents, many novel coding tools are adopted in HEVC. Meanwhile, the main hybrid coding framework and some classical schemes are reserved with respect to H.264/AVC. For the intra-frame prediction module, there are up to 36 intra modes available in the latest HEVC draft [5]. The novel angular intra prediction (AIP) still inherits the main idea of directional intra prediction (DIP) in H.264/AVC. Similarly, the spatial correlation and structural regularity are also used to design the just-noticeable difference model [6].

Some inherent flaws in DIP have been reported which limits the intra prediction performance. As investigated in [7,8], two issues negatively affect the performance of DIP. Firstly, for HD contents, smaller partition block is preferred due to complex texture, which introduces more bits to signal intra mode information for each block. Secondly, along with the distance increase between the reference pixel and the pixel to be predicted, the prediction quality could degrade significantly for high texture, high detail contents in HD video. Two kinds of strategies are widely discussed in the literatures to solve these problems. The first strategy focuses on reducing spatial redundancy by refining the conventional intra prediction structure. Choi et al. [7] modified Intra16 $\times$ 16 vertical and horizontal modes into a line-by-line structure. In [8], a bi-MB level horizontal spatial prediction was proposed. The

* Corresponding author. Tel.: +86 28 61830586; fax: +86 28 61830064.
E-mail addresses: wqb1985@yahoo.cn (Q. Wu), hlli@uestc.edu.cn (H. Li).

0923-5965/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.image.2013.03.003

bi-directional intra prediction (BIP) is proposed in [9], where 9 extra bi-directional intra modes were introduced. These methods improve the prediction accuracy by introducing more correlated reference pixels or elaborate modes. This inspires us to design a flexible interpolation scheme for each directional intra mode.

The second strategy tries to reduce the spatial redundancy with the down-sampling based coding (DBC) scheme [10–17]. In [11], an efficient super resolution technique is proposed to reconstruct the down-sampling image by exploiting the inter-resolution correlation and inter-frame correlation. In [12], the wavelet transform based down-sampling is adaptively executed according to local region's smoothness. In [14–17], different sampling rates are explored to adapt to various local characteristics. A resample-based intra prediction method is further proposed in [17]. Particularly, in [13], Wu et al. proposed an interesting adaptive down-sampling scheme with directional prefiltering. A good performance can be achieved by the DBC schemes in low bit-rates. However, it is hard to extended the DBC methods to the medium or high bit-rates applications, which is widespread for HD content. In addition, the uniform down-sampling (UDS) scheme is also unsuitable for the high detail regions. This makes us consider introducing appropriate down-sampling structures for different local characteristics.

In this paper, we propose a novel mode dependent down-sampling and interpolation (MDDI) scheme which can replace the conventional intra prediction structure in H.264/AVC. In our proposed MDDI scheme, we divide the pixels in current block into two subsets, the down-sampled and non-sampled ones. For the down-sampled pixels, we employ the DIP scheme to compute the prediction value. Then, the directional interpolation will be used to predict the non-sampled pixels. An adaptive down-sampling structure is designed according to the intra mode directions. To guarantee the reconstruction quality, both the residuals of the sampled and non-sampled pixels are transmitted to the decoder. Since the down-sampling and interpolation scheme is only related to the intra mode, there is no additional syntax modification. Experimental results show that significant RD performance improvement can be achieved compared to both the DIP and UDS schemes.

The reminder of this paper is organized as follows. We investigate the prediction distance characteristics for down-sampled image in Section 2, followed by a detailed description for the proposed MDDI scheme in Section 3. Experiment results are shown in Section 4. At last, we draw the conclusion of this paper in Section 5.

2. Analysis of prediction distance characteristics for down-sampling scheme

2.1. Analysis of intra prediction accuracy

The main task of intra prediction module is to remove the spatial redundancy between the correlated pixels in current block. Firstly, we will investigate the quantitative relation of intra prediction error and the pixels' correlation. The prediction error between the ith pixel and its jth reference sample is measured with square error $SE_{i,j}$, and their Pearson correlation coefficient is labeled by $\rho_{i,j}$. It should be noted that if there are multiple reference samples for the ith pixel, we count their weighted average value as the jth reference sample. The weights are determined by the intra prediction mode. To make the investigation cover all directional intra prediction modes and more spatial distances, we count each intra mode's $SE_{i,j}$ and $\rho_{i,j}$ values for all intra 8×8 blocks. The statistical results are shown in Fig. 1.

There are four CIF sequences with different local features involved in this investigation. The sequence *foreman* has simple foreground and its background contains rich structural information. The foreground and background in *mother–daughter* are both simple. The sequence *FOOTBALL* contains complex foreground and *stefan* has complex background. In each sequence, the similar linear relationship always can be found between the $SE_{i,j}$ and $\ln(\rho_{i,j}^{-1})$ for all directional intra modes. It is consistent with our intuitive understanding that the higher correlation in term of $\rho_{i,j}$ brings lower prediction error in term of $SE_{i,j}$. Without losing of generality, we can formulate the kth intra block's sum of prediction error $D(k)$ as

$$\begin{aligned} D(k) &= \sum_{i=1}^{N} SE_{i,j}(k) \\ &= \sum_{i=1}^{N} [a \cdot \ln(\rho_{i,j}^{-1}(k)) + b] \end{aligned} \tag{1}$$

where a and b are linear fitting parameters for each sequence.

Based on the discussion in [18], we know that the autocorrelation coefficient of 1-D stationary Markov process can be represented as a 1-D distance power of the successive random variables' correlation coefficient. Then, we can further rewrite (1) as

$$\begin{aligned} D(k) &= \sum_{i=1}^{N} SE_{i,j}(k) \\ &= \sum_{i=1}^{N} [a \cdot \ln(\rho^{-S(i)}) + b] \\ &= -a \cdot (\ln \rho) \cdot \sum_{i=1}^{N} S(i) + N \cdot b \end{aligned} \tag{2}$$

where ρ, with $|\rho| < 1$, is a correlation coefficient parameter, $S(i)$ is the ith pixel's 2-D prediction distance model which would be discussed in the following subsection.

From the observation in Fig. 1, we know that $a > 0$ and $\ln \rho < 0$. Then, the problem of minimizing the sum of intra prediction error $D(k)$ is equivalent to minimize the sum of prediction distances in current block.

2.2. Prediction distance model

In H.264/AVC, 8 directional intra prediction modes are designed to remove the spatial redundancy as shown in Fig. 2. During this process, it is a common assumption that the intra prediction residual grows larger as the spatial distance increase. Based on this hypothesis, the intra mode determination in DIP is equivalent to minimize the sum of prediction distance (SPD) between the reference pixels and the pixels under prediction in current block, which