Optik 124 (2013) 6545–6552

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

A low complexity block-based adaptive lossless image compression

Long Yang[a], Xiaohai He[a,*], Gang Zhang[b], Linbo Qing[a], Tiben Che[a]

[a] Image Information Institute, College of Electronics and Information Engineering, Sichuan University, Chengdu 610064, China
[b] Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, China

ARTICLE INFO

Article history:
Received 28 December 2012
Accepted 18 May 2013

Keywords:
Embedded compression
IWT
LHT
FPGA

ABSTRACT

For low power and lossless image compression, in this paper, a low complexity, block-based decomposition of subbands technology is proposed for embedded compression (EC) algorithm, which is ready for being implemented on a single-chip of FPGA. The proposed algorithm is based on high-speed pipeline architecture of 2-D lossless integer wavelet transformation (IWT) with 2-D Lossless Hadamard Transformation (LHT). In the proposed algorithm, the coefficients of a 2-D IWT are decomposed by 4 × 4 blocks to further remove redundancy, compared with direct encoder by EBCOT of JPEG2000. Considering the feature of the 2-D IWT, a different strategy is designed for LL-subband and non-LL subbands, which denotes DC prediction (DCP) and adaptive transformation method (ATM), respectively. DCP is used to remove the correlation between two adjacent blocks of LL-subband, and ATM is used to transform non-LL subbands by 2-D LHT selectivity. After further transformation, the coefficients are decomposed as truncated integer part (TIP) and truncated residue parts (TRP), considering the complexity of hardware implementation, TIP is encoded by Zero Running Length (ZRL) and Exp-Golomb (EG). TRP is encoded by a fixed length (FL) encoder after removed redundancy by the feature of 2-D LHT, when seen as bit patterns [1]. Experimental results show that the proposed EC algorithm can achieve a good compression performance as JPEG2000, and the coding latency can be decreased at an average of 43.9%. Another innovation of this paper is EC's hardware-friendly feature and easy hardware implementation, which are presented by a simple addition or subtraction of the LIWT and LHT, and need a small on-chip memory.

1. Introduction

With the availability of low-cost hardware, Wireless Multimedia Sensor Networks (WMSN) has been potentially applicable to several emerging applications. The WMSN is composed of a large number of dense wireless multimedia node compositions. Each node is independent on the low storage capacity of the low power consumption with sensing processing functions and wireless communication module of intelligent device. When the data is transmitted or stored on WMSN, large amount of information will result in heavy pressure on the transmission bandwidth, storage capacity, and also the power consumption. The block-based video encoding system, like H.264/AVC, MPEG-4, needs more operation on DRAM accessing. DRAM I/O will take more than 60% of system power in a typical multi-chip system [3,19]. So, block-based video encoding algorithm is unsuited for WMSN. Whereas, high compression ratio (CR) of image compression algorithm becomes a key technology to the transmission bandwidth, storage capacity and power saving.

At present, JPEG2000 is a high performance of image compression algorithm, but its implementation was so complex to a FPGA in real-time, especially the EBCOT encoder [2]. Although SPIHT algorithm is a great EC engine, but its buffer size, no matter between DWT and SPIHT, or inside SPIHT engine, both requires image-level access [3]. This cannot reach the requirement of low power consumption. Therefore, Chen et al. [3] proposed a four-tree pipelining scheme to improve SPIHT, but the correlation inside the subbands still cannot be eliminated. Other EC algorithms have been proposed in [4–11]. Generally, those algorithms can be divided into two groups: lossy and lossless. Lossy compression algorithms aim at high compression percentage (CP) compared with lossless compression algorithms. Some lossy algorithms [4–6] used a fixed compression ratio to reduce memory size, but cannot achieve better image quality, or suitable for a closed-loop or for an open-loop video coding with a large group of pictures [7]. Son et al. [5] proposed a lossy EC algorithm by using 1-D Modified Hadamard Transformation (MHT) based on JPEG2000, but its algorithm is only suitable for the LL-subband, other high frequency subbands are encoded in a fixed ratio by truncating the least significant bits directly, so the quality decreased significantly. Hence, many researchers paid attention to lossless algorithms [7–11]. The method in [7] introduced a lossless algorithm based on hierarchical prediction and Significant Bit Truncation (SBT) method, which need lots of

* Corresponding author. Tel.: +86 28 85462766; fax: +86 28 85463466.
E-mail address: nic5602@scu.edu.cn (X. He).

0030-4026/$ – see front matter. Crown Copyright © 2013 Published by Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.ijleo.2013.05.125

memory to buffer the subbands coefficients. The method in [8–10] proposed a lossless algorithm by subband decomposition with 1-D Modified Hadamard Transform (MHT). 1-D MHT could perform a vector decorrelation, but for vectors between two adjacent rows, it can be helpless. Hence, the CP was unable to further improve. Kim and Kyung [11] utilized spatial correlation to process a hierarchical prediction, also required large buffer.

In this paper, we focus on low power consumption, low complexity, and high CP of lossless image compression algorithm. Therefore, we propose a pipelined architecture algorithm based on single-chip, and no need off-chip memory. According to the proposed algorithm, only two lines of image data must be buffered to perform 2-D IWT. After first step of decorrelation, a block decomposition technology which based on 2-D LHT is used to perform further decorrelation either inside the block or between two adjacent blocks selectively. To achieve this purpose, an adaptive transformation method is proposed to detect whether the current block needs to further transformation or not. Then the coefficients of 2-D LHT are separate into truncated integer parts and truncated residue parts. Considering the complexity of hardware implementation, the encoder is based on pixel level instead on bit-plane. So, the Zero Running Length and Exp-Golomb encoder are utilized for the integer parts. The residue part, after redundancy is removed, is encoded by a Fixed Length Encoder.

The rest of this paper is organized as follows: in Section 2, the background of our proposed algorithm is introduced. In Section 3, the details of the proposed algorithm are described. In Section 4, the implementation results are presented. Finally, conclusions are drawn in Section 5.

2. Background and related work

2.1. Overview of wavelet transform

It is widely known that the discrete wavelet transform (DWT) is famous for its excellent decorrelation properties, but the limited precision of computer arithmetic prevents the DWT from performing truly lossless compression [20]. Whereas, Integer Wavelet Transform (IWT) not only implements lossless encoding, but also is hardware-friendly [21]. In our proposed algorithm, 5/3 lifting wavelet is adopted; and the classical lifting scheme (LS) was introduced in [17]. There are two benefits to do this: first, it can achieve a reversible lossless integer transformation. Second, its feature of hardware-friendly can implement pipeline architecture on FPGA [18]. The forward process of 5/3 LWT is defined as follows:

$$Y(2n+1)^{predction} = \begin{cases} X(2n+1) - \dfrac{X(2n)+X(2n+2)}{2} & normal \\ X(2n+1) - X(2n) & odd\_end \end{cases} \quad (1)$$

$$Y(2n)^{update} = \begin{cases} X(2n) + \dfrac{Y(2n+1)+1}{2} & even\_begin \\ X(2n) + \dfrac{Y(2n-1)+Y(2n+1)+2}{4} & normal \end{cases} \quad (2)$$

where Y is the 1-D transformation result of X, the "*normal*" in Eqs. (1) and (2) denotes the symmetric periodic extension of the prediction and the update separately; the "*odd_end*" and the "*even_begin*" represent the embedded extensions for the problem of boundary extension [14]. When updating the first data of the image, the "*even_begin*" in (2) is transformed into

$$Y(0)^{update} = X(0) + \frac{Y(1)+1}{2} \quad (3)$$

$Y(1)$ can be calculated by the "*normal*" in (1)

$$Y(1)^{predction} = X(1) - \frac{X(0)+X(2)}{2} \quad (4)$$

Hence, according to Eqs. (3) and (4), only three data ($X(0), X(1), X(2)$) can perform a row transformation. For the same reason, column transformation also needs three data to perform. So, there is no need to wait for the entire image data to finish transformation, only three rows of the image data can implement a 2-D IWT, it also means that two rows of the data must be buffered. For the multi-level 2-D IWT, it is executed by iteratively performing the LL-subband of last stage.

2.2. Overview of Hadamard transform

Hadamard transformation is widely used in frame-recompression EC, like [4,5,8–10] employ 1-D MHT. Because each element of Hadamard transformation matrix is either plus or minus if the normalization factor is ignored [1]. This advantage is beneficial for hardware implementation. Although the above paper proposed subband recompression or intra-frame recompression by using the decorrelation of 1-D MHT, which achieved a certain effect, the correlation in the orthogonal direction of 1-D MHT still cannot be removed. So, 2-D Lossless Hadamard Transformation (LHT) has been on the agenda, it means that a block-based transformation is utilized. 2-D LHT not only eliminates correlation of inside the block, but also make DC component more concentrated, this is convenient for us to further remove the correlation between two blocks. However, 2-D LHT require buffer more data to perform transformation. With the development of VSLI, buffering several lines of image data is easy to be implemented. Another problem by using 2-D LHT is that the Dynamic Range (DR) of the transformed coefficients is larger than 1-Ds. For the 1-D LHT, the DR of transformed coefficients will be four times larger than the input, compare with the 16 times of the 2-D LHT, according to the determinant of Hadamard matrix. For example, let S and L represent the input and transformed 4×4 coefficients matrix, respectively, the 2-D LHT can be expressed as

$$L = HSH^T \quad (5)$$

where

$$H = H^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \quad (6)$$

because of

$$\det(H) = 16 \quad (7)$$

so

$$S' = H^T L H = H^T H S H^T H = 16S \quad (8)$$

Eq. (8) means that the output coefficients of 2-D LHT will be 16 times larger than the input [12]. In the presence of the problem, Philips and Denecker [13] introduced a lossless version of Hadamard transformation, which decomposed H as

$$H = A \cdot B \quad (9)$$

let

$$|\det(B)| = 1 \quad (10)$$

then the 2-D LHT can be defined as

$$L' = BSB^T \quad (11)$$

so

$$L'' = B^T L' B = B^T B S B^T B = S \quad (12)$$