訂　正　確　認　報　告　書

訂正承認日	2016 年　4 月　25 日	訂正申請日	2016 年　3 月　15 日
題名	Low-Complexity Video Compression Algorithm and Video Encoder LSI Design		
著者氏名	劉　欽　　Qin Liu		
報告者氏名	池永　剛	確認者氏名	吉江　修

訂正確認報告書
題名：Low-Complexity Video Compression Algorithm and Video Encoder LSI Design
著者指名：　劉　欽
報告者指名：　池永　剛

１. 訂正箇所と訂正内容
（１）訂正箇所　Abstract
訂正内容：　記述の訂正
具体的内容：
~~As of 2009, H.264, the latest block-oriented commotion compensation based codec standard, has been developed by ITU-T VCEG together with ISO/IEC MPEG. The H.264/AVC is to create a standard capable of providing good video quality at substantially lower bit rates than previous standards.~~

H.264 is the latest block-oriented motion compensation based video codec standard, which has been developed by ITU-T VCEG and ISO/IEC MPEG. Comparing with previous video coding standard, H.264/AVC achieves substantially lower bit rates with- out significant video quality loss.

（２）訂正箇所　Chapter 1, Section 1.1.1, page 1;
訂正内容：　記述の訂正
具体的内容：
~~Starting in the late 1970s to the early 1980s, to manipulate and add effect to the video, several types of video production equipment were introduced, such as time~~
中略
~~consumers'world, whose architecture was a time-based and streaming data formats.~~

From the late 1970s, some video production equipment, such as Time base correctors (TBC) and digital video effects (DVE) were designed to convert analog video input to digital video format, which was easy to add effect to video. After digital processing, it had to be converted back to analog form for the output device. It was the early attempt for the digital video.
Until 1986, the birth of Sony D-1 format represented the digital video world's arrival. Instead of high-band analog forms, the video signal is stored, processed in digital form, and used by the huge television networks.
In around 1990, Apple Computer first introduced digital video into consumers' world, whose architecture was a time-based and streaming data formats.

（３）訂正箇所　Chapter 1, Section 1.1.2, page 2;
訂正内容：　記述の訂正
具体的内容：
~~Video is basically a three-dimensional array of color pixels. Two dimensions serve as spatial (horizontal and vertical) directions of the moving pictures, and one dimension represents the time domain. A data frame is a set of all pixels that correspond to a single time moment. Basically, a frame is the same as a still picture.~~
~~Video data contains spatial and temporal redundancy. Similarities can thus be encoded by merely registering differences within a frame (spatial), and/or~~

~~between frames (temporal). Spatial encoding is performed by taking advantage of the fact that the human eye is unable to distinguish small differences in color as easily as it can change in brightness. With temporal compression only the changes from one frame to the next are encoded as often a large number of the pixels will be the same on a series of frames.~~

Image is a two-dimensional array of color pixels. Each pixel is most commonly made of by RGB color space, which stands for red, green, and blue. Video has one more dimension for time domain. The raw video data consists of a set of still two-dimensional still picture in temporal order.

The raw video data needs too much space for recording. Both similarities between inner frames and inter frames can be utilized to compressed the raw video data. Both spatial redundancy and temporal redundancy can save the bits for recording the similar pixels.

（４）訂正箇所　Chapter 1, Section 1.1.2.1, page 2-3;
訂正内容：　記述の訂正
具体的内容：

~~There are four mainstream methods for spatial compression, discrete cosine trans- form (DCT), vector quantization (VQ), fractal compression, and discrete~~

中略

~~some re- search, but are typically not used in practical products (except for the use of wavelet coding as still-image coders without motion compensation).~~

In current image and video compression standard, discrete cosine transform and discrete wavelet transform are those two mainstream spatial compression methods.

For image and video application, human eye can't perceive the small high-frequency components, which can be discarded in the lossy compression method. Thus, discrete cosine transform transfers the image data into frequency domain, analyzes it and discards those neglectable high-frequency components. Discrete cosine transform is applied in many image and video compression standards, such as JPEG, MPEG, H.261, and H.263.

Discrete wavelet transform also transforms the image data into frequency components, same as DCT. However, its main merit is its temporal resolution, which captures both frequency and temporal information. That is why discrete wavelet transform can be applied in JPEG 2000 image compression standard.

（５）訂正箇所　Chapter 1, Section 1.1.2.2, page 3-4;
訂正内容：　記述の訂正
具体的内容：

~~One method used by various video formats to reduce temporal redundancy is motion compensation. For many frames of a movie, the only difference between one frame~~

中略

~~entropy coder will take advantage of the resulting statistical distribution of the motion vectors around the zero vector to reduce the output size.~~

Motion compensation is one commonly used method to reduce temporal redundancy in video compression. Video captures frames in a high speed. Thus, each frame and next frame are very similar. Motion Compensation utilizes this feature, to recover the next frame from previous reference frame and the differences between two frames.

Block based motion compensation method will split each frames into different sized macroblocks. For example, in MPEG standard, the size of macroblock is 16 × 16 pixels. Search algorithm will find a similar 16 × 16 block in the reference frame. The shift of the positions between original block and target block is the motion vector. The difference between the original block and target block is the compensation.

Both the compensation and the motion vector are encoded in the final data stream.

（6）訂正箇所　Chapter 1, Section 1.2.1, page 4-5;
訂正内容：　記述の訂正
具体的内容：

~~As of 2009, H.264/AVC/MPEG-4 Part 10 is the latest block-oriented commotion compensation based codec standard developed by the ITU-T VCEG together with~~

中略

~~specified decoding. Adaptive encoder selection between the 4 × 4 and 8 × 8 transform block sizes for the integer transform operation.~~

H.264 is the latest video codec standard, which is widely used in all kinds of video application. It achieves higher compression rates without significant video quality loss.

However, a lot of new features with intensive computation complexity are added into H.264 standard.

For example, for the multiple reference frames feature, up to 16 previously encoded frames are able to be taken as reference frames. If "B pictures" are included, the multiple reference frames will be extended to 32 frames for both back and forth direction.

To predict moving regions more precisely, variable block-size motion compensation method is introduced in H.264, which allows the predicted block segmentation to vary from 16 × 16, 16 × 8, 8× 16, 8 × 8, 8 × 4, 4 × 8, to 4× 4.

Each macroblock in every partition will have its own motion vector, which can point to different position in different reference pictures.

A 4 × 4 block size is added to entropy coding part. Encoder will select between the 4 × 4 and 8 × 8 transform block sizes, according to final coding efficiency of the residual data.

（7）訂正箇所　Chapter 1, Section 1.2.2, page 5-6;
訂正内容：　記述の訂正
具体的内容：

~~Same as previous standard, H.264 does not explicitly define a CODEC but rather defines the syntax of an encoded video bitstream together with the method of de-~~

中略

~~block. A filter is applied to reduce the effects of blocking distortion and the reconstructed reference picture is created from a series of block Fn′.~~

In Figure 1.1(26), a data flow of H.264 encoder is shown, which consists of two main paths: "forward" path and "reconstruction" path. Each macroblock in each input frame F_n will be encoded in intra mode or inter mode. In intra mode, prediction of macroblock is searched in already-encoded slice in current frame, while in inter mode, the prediction is searched from one or two reference frames before or after the current frame. ME will have the prediction macroblock's motion vector.

After subtracting the prediction macroblock in motion compensation phase, the residual data D_n is sent to transform and quantization part. The processed coefficients are reordered and do the entropy coding. The final data with some side information are packaged into Network Abstraction Layer data area.

To simulate the decoder part's reconstruction work for better motion compensation, the reverse of transform and quantization will generate the residual data D_n'. After deblocking filter, the reconstructed reference picture F_n' is utilized in the next frame's motion estimation and motion compensation phase.

追加参考文献

[26] Iain E. G. Richardson. H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation on Multimedia. Wiley, 2003. 4, 57

（文献番号は、順次繰り下げる）

（８）訂正箇所　Chapter 1, Section 1.2.3, page 6;

訂正内容：　記述の訂正

具体的内容：

~~Block-matching algorithm (BMA) for motion estimation is the mainstream algorithm for video compression, has been adopted by many standards such as MPEG- 1/2/4, H.263/264 and AVS. The key problem for BMA is to find the best matched motion vector in their reference frames for every macroblock.~~

~~Full search method searches each search candidate point sequentially in the square search window. Although this method can get the best motion vector, the calculation burden is huge. Especially for the H.264, it adopts some effective features such as the variable block sized motion estimation (VBSME) and multiple reference frame (MRF) to improve the video compression quality.~~

Same as previous MPEG and H.263 standard, block matching algorithm is also adopted by H.264 in the motion estimation part. Motion estimation is intent to find the best motion vector in the reference frames, while block matching algorithms focus on how to search it. By some researchers' statistical analysis for the whole computation time, motion estimation part will occupy around 93%.

The brutal-force full search method searches each candidate point in the search window one by one. Although it gets the best result, the huge computation complexity is not acceptable for real H.264 application, which increases the complexity by some new features previously mentioned above for higher video quality, such as variable block sized motion estimation and multiple reference frames.

（９）訂正箇所　Chapter 1, Section 1.3, page 8;

訂正内容：　記述の訂正

具体的内容：

~~Although many fast motion estimation algorithms have been proposed to reduce the huge calculation complexity instead of full search, the motion estimation time still can not satisfy the critical real-time applications' needs.~~

It is still a critical and urgent research topic to further reduce the computation complexity for motion estimation algorithm.

（１０）訂正箇所　Chapter 1, Section 1.3, page 9;

訂正内容：　記述の訂正
具体的内容：
~~After trans- form and quantization, if all the coefficients of the block's residue data are zero, this block is called all-zero block (AZB). Provided that an AZB can be detected early, the process of transform and quantization on an AZB can be skipped, which reduces significant redundant computations.~~

All-Zero block (AZB) means the macro block's transformed and quantized residual data be- come to be an all-zero array. If this kind all zero block can be detected early, transform and quantization on all zero block is not necessary and can be saved.

（１１）訂正箇所　Chapter 2, Section 2.1, page 13-14;
訂正内容：　記述の訂正
具体的内容：
~~Block-matching algorithm (BMA) for motion estimation (ME) is the mainstream algorithm for video compression, has been adopted by many standards such as~~
中略
~~these algorithms are easy to trap into the local minimum point, which will degrade the video quality.~~

Block matching algorithm is widely adopted by motion estimation part in MPEG standard and AVS standard. The block-matching algorithm is targeted to search the best motion vector in the search area for each macroblock. In H.264, full search algorithm should search all the search candidate point in each search window for 7 type block sizes and all the reference frames, which means a very heavy burden for computation complexity.
Due to the heavy computation burden for full search motion estimation, three categories of fast search methods are proposed. The first category is aimed to reduce the number of search candidate points. Different fixed search patterns are utilized to find the best search point in the search window, such as Three step search (3SS)(12), 2D-logarithm search (LOGS)(12), new three-step search (N3SS)(13), four step search (4SS)(19), block-based gradient descent search (BBGDS)(18), diamond search (DS)(40), hexagon-based search (HEXBS)(39).
These search algorithms use some fixed search patterns to search, and utilize previous search result to optimize next search steps, which can reduce the search candidate points a lot without significant video quality loss. However, these fixed search patterns cannot be suitable for every scenario, while most of these algorithms are easy to trap into the local minimum point, which will degrade the video quality.

（１２）訂正箇所　Chapter 2, Section 2.2, page 15-17;
訂正内容：　記述の訂正
具体的内容：
~~Some impact of aliasing on the motion-compensated prediction in hybrid video coding (such as H.264) have been analyzed. Thomas Wedi (33) did some~~
中略
~~each other for Y (jω). It causes aliasing problem, which incurs the distortion of signal. Especially for high frequency signal it will bring more error for prediction.~~

Thomas Wedi(35) did some mathematical analysis on the aliasing problem on motion compensation prediction in hybrid video coding. Aliasing problem is introduced in the space-discrete signals. According to his prediction error signal's equation, two conclusions can be drawn by Wedi: First, aliasing is caused by the high frequency signal. Secondly, the impact of aliasing on the pre- diction error vanishes at full-pel displacements and reaches the summit at half-pel displacements.

Figure 2.1 in Zhenyu Liu's paper(22) shows the frequency aliasing problem more explicitly.

1-D signal $x(t)$ and its frequency spectrum is denoted as $X(j\omega)$. In Fig.2.1, original signal's frequency spectrum $X(j\omega)$ and subsampling signal's frequency spectrum $Y(j\omega)$ are shown respectively. The neighbor waves overlap each other for $Y(j\omega)$. It causes aliasing problem, which incurs the distortion of signal. Especially for high frequency signal it will bring more error for prediction.

（１３）訂正箇所　Chapter 2, Section 2.3, page 18;
訂正内容：　記述の訂正
具体的内容：
~~The masks can be applied separately to the input image's 2 × 2 block, to produce separate measurements of the gradient component in each orientation (Gx and Gy). These can then be combined together to find the absolute magnitude of the gradient at each point. The gradient magnitude is given by G. G' is an approximate magnitude which is much faster to compute. G' is adopted in this algorithm.~~

To measure the gradient component in X and Y orientation, the masks is to be applied separately to the 2 × 2 block. It is easy to combine two orientation's measurement to calculate the absolute magnitude of the gradient G at each point. G' is adopted to simplify the calculation of G in this algorithm.

（１４）訂正箇所　Chapter 3, Section 3.1, page 33-34;
訂正内容：　記述の訂正
具体的内容：
~~Block-matching algorithm (BMA) for motion estimation (ME) is the mainstream algorithm for video compression, has been adopted by many standards such as~~

中略

~~(UMHS) (8) (7) algorithm. They combine many methods together and achieve both fast speed and image quality.~~

Full search is brutal force method to search all the candidates in the search window for all the 7 types block sizes and commonly 5 reference frames. That means even for the QCIF sequence, there are (2*16+1)*(2*16+1)*99*41*5=22101255 candidate points for each frame.

To reduce the search points, many fast search algorithms are proposed. Diamond search algorithm(DS)(40), Adaptive rood pattern search (ARPS) (24) algorithm, Predictive motion vector field adaptive search technique (PMVFAST) (32), Hexagon-based search algorithm(HEXBS)(39) and Unsymmetrical cross Multi Hexagon grid Search (UMHS) (6) (7) algorithm are representative works. Especially UMHS combines many methods together and achieves both fast speed and image quality, and has been adapted by h.264 reference software. But these algorithms are so complicated, not fast enough, not easy to be transplanted to hardware implementation.

（１５）訂正箇所　Chapter 3, Section 3.2.2, page 35;
訂正内容：　記述の訂正
具体的内容：
~~Motion Prediction is an important part in the whole motion estimation. If a good motion vector predictor can be made, it means that the point to start to search is much nearer the best point. Less search points will be examined and it has higher possibility to get the best point.~~

The start point for searching best motion vector is very critical for the whole motion estimation. Better start point means less search points to be checked. Thus, prediction of motion vector is adopted by many motion estimation algorithms.

（１６）訂正箇所　Chapter 3, Section 3.2.2, page 36-37;
訂正内容：　記述の訂正
具体的内容：
~~the partition with lowest cost will be selected. Big partition will be tested first, then it will turn to smaller ones gradually. The MV search of big parttion is a guide for the small partition. It shows the trend of the movement of the marcoblock. So the 16*16 macroblock's motion vector can be referenced by 16*8 or 8*16 macroblock.~~
~~Reference Frame MVP (Figure 3.2) is to utilize the temporal relation ship of the same macroblocks in neighbor frames. The temporal neighbor of reference frame has the similar MV. So this similarity will be utilized to do MV prediction.~~

In VBMSE, all the 7 types of macroblock partition will be checked, from large partition size to small partition size. The macroblock's motion vector in larger partition can be the motion vector prediction for small partition size, which is called Upper Layer MVP (Figure 3.1(b)). For example, the 16*16 macroblock's motion vector can be the prediction motion vector of 16*8 or 8*16 macroblock. Motion vector of macroblocks with same position in neighbor frames have great similarity, which can be utilized to do MV prediction and to be called Reference Frame MVP (Figure 3.2).

（１７）訂正箇所　Chapter 4, Section 4.1, page 47;
訂正内容：　記述の訂正
具体的内容：
~~Rate-distortion optimization techniques have been widely applied to video encoders (15), and provide the potential for substantial improvements in compression efficiency.~~

Rate-distortion optimization method is targeted to remove the maximum redundancy from a source, subject to a criterion, which is a mainstream method in video encoders.

（１８）訂正箇所　Chapter 4, Section 4.4, page 50-51;
訂正内容：　記述の訂正
具体的内容：
~~Unlike other video coding standards, H.264 has many different Intra and Inter mode choices to code a macroblock. To choose the best mode, it is recommended~~

中略

~~reference frame, motion vector and direction of prediction in B slice by using different lambda values and distortion measurement.~~

Each macroblock has to choose whether it will be encoded in Intra mode or Inter mode. To choose the best mode, H.264 standard recommends to use the Lagrangian multiplier to compute the cost for each mode, and to select the smallest cost's mode. The following equation（15） shows the way to calculate Lagrangian cost, J

$$J = D + \lambda_{md} \times R \qquad (4.5)$$

The Lagrangian cost consists of two parts: Distortion part and Rate part. Distortion part measures the distortion between original frame and reconstructed frame. Rate part means the bits to code the macroblock. Thus, the best mode with smallest cost means smaller distortion and less bitrate, while two factors are simultaneously quantified by the Lagrangian multiplier, λ_{md}

The Lagrangian cost equation with different lambda values and different distortion measurement is also critical and useful to decide the best reference frame, motion vector and prediction direction for B slice.

追加参考文献

[15] K.P. Lim, G. Sullivan, and T. Wiegand. Text description of joint model reference encoding methods and decoding concealment methods. In JVT-N046.doc, Joint Video Team (JVT) of ISO/IEC MPEG ITU-T VCEG, Jan. 2005. 46

（文献番号は、順次繰り下げる）

（１９）訂正箇所　Chapter 5, Section 5.2.1, page 61-62;

訂正内容：　記述の訂正

具体的内容：

~~The 4 × 4 DCT transform in H.264 is given by Eq. 5.2, which is an approximation to the original 4×4 DCT. CXC^T is a 'core' 2D transform, while E is a matrix of~~

中略

~~multiplication by a factor $MF_{(\mathrm{QP}\ \%6, Pos)}$ and a right shift, avoiding any division operations.~~

The DCT transform in H.264 for 4 × 4 macroblock is shown in the following equation（26）. CXC^Tmeans 2D integer transform, which is an approximation to the original 4 × 4 DCT, while ⊗E indicates the scalar multiplication.

where Y_{ij} is a coefficient of previous transform result as the input data; Qstep is a quantiser step size; The rounding operation is round to the nearest smaller integer; and Z_{ij} is the coefficient after quantization as the output data.

In order to simplify the arithmetic operation between transform and quantization, the post-scaling factor E in Eq. 5.2 is combined in forward quantiser. JM reference software utilizes a multiplication by a factor $MF_{(\mathrm{QP}\ \%6, Pos)}$ and a right shift to avoid any division operations.

追加参考文献

[26] Iain E. G. Richardson. H.264 and MPEG-4 Video Compression:Video Coding for Next-Generation on Multimedia. Wiley, 2003. 4, 57

（文献番号は、順次繰り下げる）

（２０）訂正箇所　Chapter 5, Section 5.2.3, page 68;

訂正内容：　記述の訂正
具体的内容：
~~The concept of adapting the transform size to the block size of motion-compensated prediction has proven to be a promising coding tool within the H.264/AVC video coding layer design. Gordon(26) proposed a seamless integration of a new 8x8 transform into H.264.~~

In (27), a new 8 × 8 transform is introduced into H.264 by Gordon. Thus, the transform size can adapt to the block size of motion-compensation prediction, which has been proven to be a more promising feature for video coding.

（２１）訂正箇所　Chapter 5, Section 5.2.4, page 73;
訂正内容：　記述の訂正
具体的内容：
~~It is not necessary to perform transformation and quantization to the detected AZBs. Furthermore, the reconstructed values of these AZBs are acquired before inverse quantization and inverse transformation, thus, numerous unnecessary calculations of these blocks can be saved.~~
~~To examine how much calculation is reduced, the total number of operations required for the proposed method will be calculated and compared with that of the other methods respectively.~~

If the AZBs are detected beforehand, transform and quantization are not necessary. Thus, transform and quantization calculation of these AZBs can be saved. To evaluate the effect of this proposed AZBs algorithm, the total number of operation is calculated and compared with conventional algorithms.

（２２）訂正箇所　Chapter 5, Section 5.3.1, page 79;
訂正内容：　記述の訂正
具体的内容：
~~Thus, the input values at the input of transform can be approximated by a separable covariance γ(m,n) = σf2ρ|m|ρ|n| , where m and n are the horizontal and vertical distances, respectively, between two pixels, and |ρ| is the correlation coefficient. ρ = 0.6 is commonly set as 0.6 in the simulations. Huang(12) use Laplacian distribution for simple estimation. As huang said, the generalized Gaussian distribution is the most accurate representation of transformed coefficients.~~

Thus, Huang(11) approximates the input values by a separable covariance $\gamma(m,n) = \sigma_f^2 \rho^m \rho^n$. m means horizontal distances between two pixels, while n is vertical distances. |ρ| stands for the correlation coefficient. ρ = 0.6 is commonly set as 0.6 in the Huang's simulations. Huang(11) use Laplacian distribution for simplifying the computation. As Huang said, the generalized Gaussian distribution is the most accurate representation of transformed coefficients. Thus, in our derivation, Gaussian distribution is utilized.

（２３）訂正箇所　Chapter 6, Section 6.3.5, page 92;
訂正内容：　記述の訂正
具体的内容：
~~To reduce the calculation burden of the full-search block motion estimation, many effective methods have been proposed: three step search algorithm~~

~~(3SS)(13), the 2D- logarithm search algorithm (LOGS), the new three-step search algorithm (N3SS)(14), the four step search algorithm (4SS)(19), the block-based gradient descent search algorithm (BBGDS)(18), the diamond search algorithm (DS)(38), and the hexagon- base search algorithm (HEXBS)(37). However, all of these algorithms easily become trapped at the local minimum point, which degrades the video quality.~~

このパラグラフは削除。

２．訂正理由

序論や各章の導入部分において、動画像圧縮技術、H.264標準仕様、動き予測処理などの一般的な解説（歴史的背景、用語説明など）に関する記述において、インターネット(Wikipediaなど)、教科書、標準化仕様書などから、不用意な引用が見られたので、それらを訂正するように指導した。(訂正箇所1-11, 13-19, 23)

また、5章などにおいて、提案技術と種々の従来技術と比較するため、従来技術のポイントや共通する利点などが記載されている。これらに関連する論文は参考文献に記されていたが、段落レベルの不用意な引用が見られたので、訂正するように指導した。(訂正箇所12,20,21,22)

３. 訂正を認めた理由

（１）適切に訂正されている。
（２）訂正部分は、序論と各章の導入部分の一般的な解説、及び従来技術との比較を明確にするため記述された部分であり、訂正内容も技術的意味を変更しているわけではない。よって、博士論文としての価値は損なわれておらず訂正は妥当と判断する。