DCT領域 Distributed Video Codingにおける尤度推定手法

橋本亮司† 简井弘† 尾上孝雄† 猪間知宏††
† 産業技術総合研究所情報科学部門 〒565-8520 大阪市吹田区
†† Sharp株式会社 〒260-8520 千葉市美浜区

E-mail: †[hashimoto.ryoji,tsutsui,onoye]@ist.osaka-u.ac.jp, ††ikai.tomohiro@sharp.co.jp

あらまし 近年、新しい画像符号化パラダイムとして、Distributed Video Coding (DVC) が注目されている。DVC は既存の MPEG-2 などと比べて、低演算量で符号化が可能である一方、符号化効率の面で実用化への課題がある。本稿では、DVC の符号化効率に影響を与える誤り訂正率において、誤り訂正率の入力ビットの尤度を推定する手法を提案する。提案手法では、周波数成分ごとに原画像と復号器で生成した予測画像の誤差を、コーディングによりモデル化する。具体的には、復号画像である前の前後のキーフレームを用いてモデル化した後、各ビットの尤度を算出する。ソフトウェアシミュレーションで評価を行った結果、提案手法は平均誤差 1.6%で誤り率を推定可能であった。

キーワード Distributed video coding, 尤度推定, 誤り訂正, DCT

Likelihood Estimation for Transform Domain Distributed Video Coding

Ryoyo HASHIMOTO†, Hiroshi TSUTSUI†, Takao ONOYE†, and Tomohiro IKAI††
† Graduate School of Information Science and Technology, Osaka University
1—5 Yamadaoka, Suita, Osaka, 565—0871 Japan
†† Sharp Corporation, 1—9—2 Nakase, Mihama—Ku, Chiba, 261—8520 Japan
E-mail: †[hashimoto.ryoji,tsutsui,onoye]@ist.osaka-u.ac.jp, ††ikai.tomohiro@sharp.co.jp

Abstract Distributed Video Coding (DVC), which is a new image compression paradigm, attracts a lot of attention from video researchers. While computational cost of DVC encoding is lower than that of MPEG-2, low coding efficiency has been an issue for practical DVC applications. In this paper, a likelihood estimation method for transform domain DVC is proposed, which uses Cauchy distribution as a virtual channel model. In the decoder, virtual channel is estimated by utilizing error between forward and backward predicted images for each frequency component. Likelihood is obtained from estimated virtual channel. Simulation results show that the proposed method can estimate the error rate with 1.6% error on average.

Key words Distributed video coding, Likelihood estimation, Virtual channel, DCT

1. はじめに

DVC (Distributed Video Coding) は、近年注目されている新しい動画像符号化方式である。DVC は 1970 年代に提案されたスレパン＝ワルフ [1] とウェナー＝ジフ [2] の理論を基盤としている。現在、広範に利用されている MPEG-2 [3], H.264 [4] では、符号化器は、符号化パラメータと原画像と予測画像の誤差である予測誤差信号を、符号化する。復号器は、符号化パラメータを用いて予測画像を取り、予測誤差を加算することで、復号画像を得る。これに対して、DVC では、原画像を通信路符号化を用いて符号化する。復号器側では、符号化器から送信された誤り訂正符号を用いて、自身で生成した予測画像の訂正を行うことで、復号画像を得る。DVC は、このように、MPEG-2, H.264 の符号化において演算量の大部分を占めていた予測画像の生成（動き検出など）を符号化時に行わないため、従来の方式に比べて、符号化に要する演算量が少ない。一方で、復号時には予測画像の生成、誤り訂正という複雑な処理が必要になるため、復号に要する演算量は増加する。このように複雑な処理が符号化器側から復号器側へ移動したため、演算資源に制約があり、低消費電力が要求されるモバイル端末等での画像符号化への応用が期待されている。DVC の実用化への課題として、従来の符号化方式に比べて符号化効率が低いことが挙げられる。このようなため、DVC の符号化効率の改善のためにさまざまな研究が行われている。

DVC における符号化効率の改善として、DVC を画素領域ではなく、DCT (Discrete Cosine Transform) 領域で行う手法 [5] が提案されており、画素領域に比べて高い符号化効率が得られている。しかしながら、DCT 領域で符号化を行うことによって、
いくつもの問題が生じている。その１つとして、誤り訂正における尤度推定が挙げられる。画像領域のDVCでは、原画像と予測画像の差誤（予測誤差信号）をラプラス分布を用いてモデル化することで、尤度の推定を行っている[6],[7]。一方、DCT領域では、低周波成分と高周波成分では予測誤差信号の分布の傾向が異なることが予想される。このため、画像領域と同様のモデル化では、尤度の推定精度が悪化し、誤り訂正が完了しない可能性があり、結果として符号化効率が低下してしまう。

そこで本稿では、DCT領域DVCにおける尤度推定手法を提案する。提案手法では、各周波度成分ごとに予測誤差信号の分布をコーシー分布を用いてモデル化する。モデル化時には、復号器で得られる情報のみを用いて、予測誤差信号の分布を推定する。また、これまで尤度推定に関しては、予測誤差分布とモデルの近似精度、RD（Rate-Distortion）特性などが用いられてきた[8]。本稿では、推定した誤り率と実際の誤り率との誤差を用いて評価を行う。評価の結果、提案手法では平均誤差1.2%で誤り率の推定が可能である。

本稿の構成は以下の通りである。2章でDVCにおける符号化復号処理と尤度推定方法について説明する。3章で提案尤度推定手法について詳述した後に、4章でその評価を行う。最後に、5章で結論を述べる。

2. Distributed Video Codingにおける符号化復号と尤度推定

2.1 Distributed Video Codingにおける符号化復号

図1に[5]で提案されているDCT領域におけるDVC符号化器、復号器の構成を示す。符号化器では、原画像をキーフレームとWZ（Wyner-Ziv）フレームの2種類に分けて符号化を行う。本稿では、簡素化のため奇数番目のフレームX_{t-1}をキーフレーム、偶数番目のフレームX_tをWZフレームとして符号化するものとする。キーフレームは、フレーム内予測のみを用いて符号化されるフレームであり、MPEG-2、H.264における1ピクチャに相当する。一方、WZフレームは原画像を4×4ブロック単位に分割したうえでDCT、量化を行う。この結果を各周波数成分、各ピクセルブロック単位に分割し、それぞれをSW（Slepian-Wolf）符号化して得られる信号列を復号器に送信する。

復号器では、キーフレームの復号は従来と同様の方法で行われ、復号画像X_{2t-1}が得られる。WZフレームの復号では、まず最初に復号したキーフレームの画像から予測画像の生成を行う。予測画像は、前後のキーフレームに対して動きを推定を行い、得られた動きベクトルを用いて、以下の式に基づきフレーム補間を行うことで生成される。

\[ Y_{2t}(x,y) = \frac{1}{2}(X_{2t-1}(x-mv_x,y-mv_y)+X_{2t}(x+mv_x,y+mv_y)) \]  \hspace{1cm} (1)

得られた予測画像に対して、符号化器と同様、DCTを行い、SW復号に用いる補助情報を生成する。次に生成した補助情報を対し、尤度を推定する。尤度推定の詳細に関しては、次節で述べる。次に推定した尤度と符号化器から送信された信号列を用い

図1 DCT領域DVCのシステム構成

2.2 Distributed Video Codingにおける尤度推定

尤度推定は、誤り訂正の前処理として行われ、誤り訂正へ力が入されるピットの確からしさを設定する。一般に誤り訂正符号の復号法である尤度復号、最大事後確率復号では、ピットの確からさを利用して誤り訂正を行う。尤度が正しく推定されなかった場合には、誤り訂正に要する符号長が増加する。あるいは完全に誤り訂正が行えないとした問題が生じ、DVCの符号化効率の低下を招く。

DVCにおいては、Y_tに生じる誤り（予測誤差信号）は、通信路の中で生じるものではないが、原画像が仮想的なチャネル（仮想チャネル）を通った際に誤りが生じた（画像の値が変化した）と考えることができる。そのため、この仮想チャネルの適切なモデル化が誤り訂正において重要な課題となる。画像領域DVCにおける原画像X_tと予測画像Y_t間の仮想チャネルは、ラプラス分布でモデル化される[6]。この場合、WZフレームX_tにおける画像X_t(x,y)がある値iとなる確率は以下の式を用いて算出される。

\[ \Pr(X_t(x,y) = i|Y_t(x,y)) = \frac{1}{2\sigma} \exp\left(-\frac{|i - Y_t(x,y)|}{\sigma}\right) \]  \hspace{1cm} (2)

ここで、Y_t(x,y)、2\sigmaがそれぞれラプラス分布の期待値、分散となる。さらに、画像X_t(x,y)の下位からb番目のピクチャブロックのピットをX_{2t}(x,y)とすると、X_{2t}(x,y)が1となる確率は、

\[ \Pr(X_{2t}(x,y) = 1|Y_t(x,y)) = \sum_{i=1}^{2^b} \Pr(X_t(x,y) = i|Y_t(x,y)) \]  \hspace{1cm} (3)

ここで、UはX_t(x,y)の値が1となるX_t(x,y)の集合を示し、

\[ U = \{m \cdot 2^b, m \cdot 2^b-1, \ldots, (m+1) \cdot 2^b-1\} \quad (m = 1, 3, 5, 7, \ldots) \]

となる。例えば、X_t(x,y)∈{0,1,2,3,63}、b=3の場合、

\[ U = \{8, 9, 15, 24, 25, \ldots, 31, 40, 41, \ldots, 47, 56, 57, \ldots, 63\} \]

となる。

尤度復号、最大事後確率復号に用いられる尤度は対数尤度比 (LLR: log-likelihood ratio) で与えられ、X_{2t}(x,y)に対するLLRは

\[ LLR_{2t}(x,y) = \log \frac{\Pr(X_{2t}(x,y) = 1|Y_t(x,y))}{\Pr(X_{2t}(x,y) = 0|Y_t(x,y))} \]  \hspace{1cm} (4)

\[ = \log \frac{\Pr(X_{2t}(x,y) = 1|Y_t(x,y))}{\Pr(X_{2t}(x,y) = 0|Y_t(x,y))} = \log \Pr(X_{2t}(x,y) = 1|Y_t(x,y)) \]  \hspace{1cm} (5)

として算出される。
3. 提案尤度推定手法

本章では、はじめに提案尤度推定手法で用いる仮想チャネルモデルについて述べる。次に、その仮想チャネルモデルのパラメータの推定手法について述べる。

3.1 仮想チャネルモデルの選定

仮想チャネルのモデル化を考えるにあたり、まずDCT後の各周波数成分における予測誤差信号の分布の測定を行った。その結果を図2に示す。測定には、foreman, mobile & calender, news, salesmanの4種類の動画像を用いた、いずれもCIF(352×288)解像度の動画像である。基フレームに関しては、H.264の1フレームとして符号化し、その際の量子化値は26とした。周波数成分のインデックスは図3で示されるH.264におけるジグザグスキャン順に基づくものとする。例えば、ジグザグスキャン順で4番目の周波数成分はAC4と表記される。

図2より予測誤差信号の分布に関して、いずれの周波数成分も予測誤差は0付近を中心とする分布となっているが、裾野の広がり具合は周波数成分によって異なる。低周波成分ほどピークの値が低く、裾野が広い。逆に、高周波成分は裾野が狭く、急峻になっている。一般の自然画像においては、高周波成分が少なく、その値は低周波成分に比べて小さな値を取る傾向にあると言われている。このため、誤差の値も小さくなっていると考えられる。

正規分布、ラプラス分布、コーニー分布を用いて予測誤差信号の分布の近似を行った。近似結果を実分布と合わせて図2に示す。これらの分布はそれぞれ

\[ h(x) = \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{|x-\mu|^2}{2\sigma^2} \right), \]  
\[ f(x) = \frac{1}{2\sigma} \exp\left( -\frac{|x-\mu|^2}{2\sigma^2} \right), \]  
\[ g(x) = \frac{1}{\pi(x-x_0)^2+y^2), \]

で表される。ここで、\( h(x) \)は、平均値、標準偏差、分散、位置母数、尺度母数を示す。図2より、いずれの画像、周波数成分でも正規分布が他の2つの分布に比べて近似精度が低くなっていることが分かる。すなわち、正規分布はDCT領域において仮想チャネルモデルとして適していないといえる。残りの2つの分布に関しては、mobile AC4ではピークの高さが実分布と異なっていることを除いて、比較的良い近似であるといえる。より詳細な評価を行うために、以下の式で与えられる実分布と近似分布の誤差（error）を求める。

\[ \text{error} = \sum \left( p(i) - F(i) \right)^2. \]

ここで、\( p \)は分布の実測値、\( F \)は近似分布であり、\( f,g,h \)のいずれかである。表1に誤差を算出した結果を示す。表1示すようにほとんどの動画像、周波数成分において、コーニー分布の方がラプラス分布よりも誤差が小さい。よって、本手法では、仮想チャネルモデルとしてコーニー分布を用いるものとする。

3.2 パラメータの推定

前節で述べたように、本手法では、仮想チャネルモデルとして、コーニー分布を用いる。しかしながら、復号器では、原画像と予測画像の誤差計算を行うことは実際には不可能である。このため、パラメータ \( x_0, \gamma \)を推定する必要がある。

パラメータを推定する方法として、本研究では、前方から予測した画像と後方から予測した画像を用いる手法を提案する。ここで、前方、後方から予測した画像 \( Y_5(x,y) \)、\( Y_6(x,y) \)は

\[ Y_5(x,y) = Y_{5,0} + (x - mx_1)/y - mv_1/2), \]  
\[ Y_6(x,y) = Y_{6,0} + (x + mx_1)/y + mv_1/2), \]

のように、予測画像生成の際に得られた動きベクトルを用いて算出される。この2つの差分画像を算出し、周波数変換を行う。
ことにより各周波数成分の分布 $p'$を得る。提案手法では、得られた分布 $p'$を用いて実分布 $p$のパラメータの推定を行う。$x_0$は、得られた分布の最頻値とし、$y$は、以下の式のように最頻値とそこからの値の確率を用いて推定する。

$$ p'(\omega, x_0) + p'(\omega, x_0 + k) + p'(\omega, x_0 - k) = \frac{\gamma_0}{\pi(\gamma_0^2 + k^2)} \gamma$$

ここで、$\omega$は周波数のインデックスである。$k$の値に関しては、小さな値を用いるのが好ましいと考えられる。これは、最頻値付近の近似精度が全体の近似精度に大きく影響を与えるためである。

以上のように、パラメータの推定を行った。その結果得られた分布を図4に示す。$k$の値は1としている。

mobileにおいて、DC成分、AC成分のピクの高さに誤差が存在しているが、誤差改善は今後の課題である。

4. 尤度推定手法の評価

これまで、仮想チャネルモードの評価に関しては、実分布とモデルの誤差の大きさやRD特性を用いて評価されてきた。誤り訂正に用いられる尤度は、誤っている信号に対しては、低い値が設定され、正しい信号に対しては、高い値が設定されることが好ましい。これは通信において、尤度が受信信号を実際の信号に近づけたときに推定信頼を有する信号を示すためにある。尤度の大きさに関しては、「より大きなことが望ましい」という定性的な評価のみが可能である。そこで、本稿では誤り率を用いて評価を行う。

受信機が0を受信した場合において、その尤度を$p$とするとき、推定機側が0を送って0を受信した確率が$p$、1を送って0を受信した確率が$1-p$となる。この場合、誤りが発生した確率は300であり、これを推定した誤り率とみなす。また、1を受信した際にも、同様のことが言える。推定値に関しては、尤度を用いることで0＝1%の値を算出することが可能である。一方、実測値に関しては、ピクト単位で誤り率の算出を行うと常に0、100%のどちらかになる。このため、ピクト単位で評価を行うことは困難となる。本稿では、この問題を解決するため各ピクトの尤度を平均することで評価を行う。この評価では、推定した値と実測値の誤差が小さいほど良い結果であるといえる。

各ピクトブレーン、各周波数成分において、提案手法で推定した誤り率、実際の原画像と予測画像の実分布を用いて推定した誤り率、実測した誤り率の比較を行った。その結果を図5に示す。図より、下位のピクトブレーンほど誤り率が増加し、推定誤差も増加していることが分かる。同一周波数成分においては上位のピクトブレーンの方が下位のピクトブレーンに比べて、画質の影響が大きいことを踏まえると、下位ピクトブレーンにおいて、推定誤差が少ない方が望ましいといえる。提案手法では、上位ピクトブレーンにおいて推定誤差は少ない。また、異なる周波数成分において、周波数成分の下位ピクトと周波数成分の上位ピクトでは、どちらが画質に大きな影響を与えるかは明らかではなく、これらの詳細な評価は今後の課題である。
5. 結 論

本稿では、DCT 領域における尤度推定手法の提案とその評価を行った。提案手法は実分布を利用した場合とほぼ同様の誤り訂正が可能であるといえる。

なお、ピットプレーン、周波数成分ごとの平均推定誤差を表 2 に示す。提案手法では、周波数成分ごとの平均に関しては、AC4における推定誤差が最も大きい。これは、予測誤差分布の近似結果（図 4）において近似結果が良好ではない mobile AC4 の下位ピットプレーンの推定誤差が原因であると考えられる。全体の平均推定誤差に関しては、提案手法の平均誤差は 1.6% となり、実分布と実測値の平均誤差が 1.2% であることを踏まえると、提案手法は、実分布と同様高い精度で誤り率が推定可能となっている。

また、提案した尤度から算出した LLR を誤り訂正部への入力として、誤り訂正処理を行った後の、誤り率に関しても評価を行った。今回の評価条件を表 3 に、評価結果を図 6 に示す。図より、提案手法は実分布を利用した場合とほぼ同様の誤り訂正が可能であるといえる。

参考文献

表2 平均推定誤差 [%]

<table>
<thead>
<tr>
<th></th>
<th>DC</th>
<th>AC4</th>
<th>AC15</th>
<th>平均</th>
</tr>
</thead>
<tbody>
<tr>
<td>ビットプレーン</td>
<td>提案手法</td>
<td>実分布利用</td>
<td>提案手法</td>
<td>実分布利用</td>
</tr>
<tr>
<td>5</td>
<td>0.44</td>
<td>0.04</td>
<td>0.08</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>0.44</td>
<td>0.33</td>
<td>0.43</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>0.52</td>
<td>0.65</td>
<td>1.51</td>
<td>0.90</td>
</tr>
<tr>
<td>2</td>
<td>0.72</td>
<td>0.35</td>
<td>5.55</td>
<td>1.43</td>
</tr>
<tr>
<td>1</td>
<td>1.26</td>
<td>0.62</td>
<td>10.61</td>
<td>4.14</td>
</tr>
<tr>
<td>平均</td>
<td>0.68</td>
<td>0.43</td>
<td>3.60</td>
<td>1.40</td>
</tr>
</tbody>
</table>

図6 誤り訂正結果の評価

**図の説明**

- mobile AC4 ビットプレーン
- salesman AC4 ビットプレーン