雑音に強い動画像符号化のためのフレーム間類似度判定法

6Ｆ－8

前田潤治

日本アイ・ビー・エム(株) 東京基礎研究所

1　はじめに

動画像を扱うアプリケーションでは、その膨大な情報の圧縮技術が必須となる。動画像情報圧縮技術は大きくは通信系と蓄積系に分類されるが、本稿ではリアルタイム性を要求される通信系に焦点を当てる。

現在、動画像圧縮技術はMPEG-1,2、H.261[1]など、内容が何であろうと画面を正方形のブロックに機械的に分割し、その正方形ごとに個別に処理を加えるブロックベースのものが全盛である。(これらの標準では、処理の内容によって対象となる正方形は、厳密には「ブロック」「マクロブロック」などと呼び方が変わるが、本稿ではこれらをまとめて単に「ブロック」と呼ぶことにする。)

これらの手法を特徴づける要素技術として二次元離散的コサイン変換（ＤＣＴ）がある。ＤＣＴは注目しているブロック、または注目しているブロックと前のフレーム中にある参照ブロックとの差分を符号化する技術であるが、計算時間、圧縮率、画質のいずれの観点からも、ＤＣＴを行なわずに済むものならその方が理論的には望ましい。具体的には隣接フレーム間のブロック同士の類似度が高ければ、時間的に前のブロックの情報のみを使うことによってＤＣＴを省くことができる。しかし現実には、避け得ない雑音のために隣接フレーム間の類似度が不当に低く評価されてしまい、不必要なＤＣＴが行なわれてしまう。そこで本稿では、雑音の悪影響を受けにくいフレーム間類似度判定法を提案する。

2　対話型動画像通信システム

本稿で焦点を当てているのはテレビ電話やテレビ会議システムといった、対話型動画像通信システムである。これらはまさに現在広く実用になりつつあり、普及させるための標準化[1][2]の動きも盛んである。[1]や[2]で想定しているのは、典型的には肩上人物像（テレビ電話）や複数の人が横一列に並んだ画像（テレビ会議システム）である。これらにおいては、カメラは通常固定されていて背景は不動であり、前景（多くの場合人間）も一部（目や口）を除いて動きが少ない。背景にあたるブロックは前のフレームの同じ位置のブロックと類似度が極めて高いため、このブロックは「変化がない」として通信せずに済ませることができる。また、同じ位置ではなくても、あるブロックが前のフレームのブロックと極めて似ている場合には、相互の位置関係のみを伝送することでＤＣＴを避けることができる。

実用に供されている、ブロック同士の類似度判定基準としては、

$$D = \sum_i |a_i - b_i| \qquad (1)$$

(iはブロック中の画素の位置、a_i, b_iは比較している二つのブロックの画素iの値）などがあり、このDがあるしきい値より小さくなれば、類似度が十分高いとしてＤＣＴを省略する。

3　実用のシステムの問題点

式(1)で表される指標は、計算が簡単でソフトウェアでも容易に実現できる、という利点を持つが、実用化する段階では大きな障害を抱えている。まず、室内でのテレビ会議システムでは避け得ない蛍光灯の影響である。蛍光灯は非常に短い間隔で明滅を繰り返しており、人間には検知できないが、カメラを通すと、同じシーンでも、フレームによって全体の明るさが違ったものが取り込まれてしまう。このような二つのフレームに式(1)を適用すると、明るさの差が全画素にわたって足し込まれ、非常に大きな値となってしまう。

もう一つの障害は、コンピュータを構成する各部品から発生する雑音である。これは孤立点雑音であり、やはり式(1)のしきい値処理に深刻な影響を与える。

4　ベクトル演算による類似度の判定

本稿で提案する類似度判定法では、比較したいブロックをベクトルと見なし、ベクトル間の角度をもって判定基準とする。

"A measure of resemblance for robust video codec"
Junji Maeda,　maeda@trl.ibm.co.jp
Tokyo Research Laboratory, IBM-Japan Ltd., 1623-14, Shimotsuruma, Yamato, Kanagawa, 242, Japan

比較したい二つのブロックを A、Bとし、それぞれに含まれる画素数を n、一つ一つの画素を a_i、b_i（ただし $1 \leq i \leq n$、$a_i \in A$、$b_i \in B$）と表現する。a_i と $b_i(1 \leq i \leq n)$ は対応する位置の画素値である。

A、Bそれぞれの画素値を同じ順に一列に並べたものは n 次元ベクトルと見なすことができる。

$$\vec{a} = (a_1, a_2, ..., a_n) \quad (2)$$

$$\vec{b} = (b_1, b_2, ..., b_n) \quad (3)$$

$\vec{a}$、$\vec{b}$、およびそれらの間の仮想的な角度θの間には下の関係が成り立つ。

$$\cos\theta = \frac{(\vec{a}, \vec{b})}{|\vec{a}||\vec{b}|} = \frac{a_1b_1 + a_2b_2 + ... + a_nb_n}{\sqrt{a_1^2 + a_2^2 + ... + a_n^2}\sqrt{b_1^2 + b_2^2 + ... + b_n^2}}$$

上式の両辺は非負なので、$0 \leq \theta \leq \frac{\pi}{2}$ と制限できる。平方根の計算を避けるために上式の両辺を 2 倍して変形すると

$$\theta = \frac{1}{2}\cos^{-1}\left(\frac{2(a_1b_1 + a_2b_2 + ... + a_nb_n)^2}{(a_1^2 + a_2^2 + ... + a_n^2)(b_1^2 + b_2^2 + ... + b_n^2)} - 1\right)$$

となる。上記θの範囲で単調な関数をはずすと

$$R = \frac{a_1b_1 + a_2b_2 + ... + a_nb_n}{(a_1^2 + a_2^2 + ... + a_n^2)(b_1^2 + b_2^2 + ... + b_n^2)} \quad (4)$$

を得る。この値が大きいほど A と Bの類似度は高いと判定できる。

式（2）（3）のベクトルはそれぞれブロック A、Bの中の画素値の並び方を表現していると見ることができる。したがって、図1からも分かる通り、全体の明るさのみがことなるベクトルはその角度にはほとんど違いがないことになる。

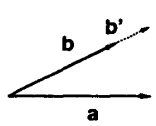

図 1: 蛍光灯の影響

また、ブロック内のごく少数の画素が特異な値をとっても、式（4）によってその影響は吸収されてしまうことも分かる。

5 実験

式（1）の Dと式（4）の Rでは数値の単位、次元が違うので、直接、客観的な比較を行なうことはできない。

そこで、シーン内容として、天井を撮りっぱなしにしたもの（シーン内容が変化しない）と肩上人物像の 2 種類を用い、天井のシーンですべてのブロックが「ＤＣＴを行なう必要なし」と判定されるしきい値の中でもっとも厳しい（ＤＣＴの必要ありと判断されやすい）ものを肩上人物像に適用した。

実験は蛍光灯で照明された室内で、カメラ内蔵ノート型パソコンによって取り込まれた動画像に対して行なわれた。

その結果、式（4）を用いた判定法では式（1）を用いたものに比べてＤＣＴの回数が約半分に減少していた。ただし、この数値はシーンの内容に左右されるものなので、数値自体に意味はない。

6 まとめ

通信系動画像圧縮において、隣接フレーム間の類似度を不当に低く判断させ、不必要なＤＣＴを引き起こす原因として蛍光灯のちらつき、部品から発生する孤立点雑音を挙げ、それらの影響を排除して類似度を判定する手法を提案した。実験では、従来からの代表的な手法より高性能であった。

式（4）から分かる通り、この判定法は処理対象の形を正方形とは限定しない。比較の対象とする二つの領域の形状が同じで、中に含まれる画素同士の対応が取れれば、形状そのものは自由である。これは本手法が次世代の動画像情報符号化技術 [3] の基本的ツールの一つとして利用できる可能性を示唆している。

参考文献

[1] ITU-T Recommendation H.261 "Video codec for audiovisual services at p × 64 kbits"

[2] ITU-T Draft Recommendation H.263 "Video coding for narrow telecommunication channels at < 64kbits/s"

[3] Haibo Li, et al., "Image Sequence Coding at Very Low Bitrates: A Review," IEEE Trans. Image Processing, Vol. 3, No. 5, September 1994.