

MPEG-2 bitstream からの動領域抽出法の検討

田邊 集[†] 秦泉寺久美^{††,†††} 渡辺 裕^{††} 富永 英義^{†,††}

[†] 早稲田大学 理工学部 電子情報通信学科

〒1169-8555 東京都新宿区大久保3-4-1

^{††} 早稲田大学 大学院 国際情報通信研究科

〒169-0051 東京都新宿区西早稲田1-3-10

^{†††} 日本電信電話株式会社 NTT サイバースペース研究所

〒239-0847 神奈川県横須賀市光の丘1-1

E-mail: tshu@tom.comm.waseda.ac.jp

あらまし 様々な帯域の混在するネットワーク上で MPEG 動画像配信を行なう際に、動画像のビットレートを伝送路の帯域に合わせて変換するトランスコーディングに関する研究が現在盛んに行われている。しかし、これまでのトランスコーダはビットレート削減に伴う画質の劣化は全体に一律であった。そこで本研究では、MPEG-2 のビットストリームから実時間で動領域を抽出し、動領域と背景領域で画質を変化させることで、より効率的なトランスコーディングの実現を目的としている。

キーワード 動領域、前景領域、カメラパラメータ、トランスコーダ、MPEG

A Study on Moving Area Detection from MPEG-2 Bitstream

Shu TANABE[†], Kumi JINZENJI^{††,†††}, Hiroshi WATANABE^{††}, and Hideyoshi TOMINAGA^{†,††}

[†] Department of Electronics, Information and Communication Engineering, Waseda University
Okubo 3-4-1, Shinjuku-ku, Tokyo, 169-8555 Japan

^{††} Graduate School of Global Information and Telecommunication Studies, Waseda University
Nishiwaseda 1-3-10, Shinjuku-ku, Tokyo, 169-0051 Japan

^{†††} NTT Cyber Space Laboratories, NTT Corporation
Hikarinooka 1-1, Yokosuka-shi, Kanagawa, 239-0847, Japan

E-mail: tshu@tom.comm.waseda.ac.jp

Abstract Streaming of MPEG video on the network where various bandwidth are intermingled has been getting common. Thus the research on transcoding, which changes the bit rate of video according to the bandwidth of a transmission line, has been conducted. In the conventional transcoding, degradation of image quality caused by bit rate reduction diffuse across the whole image. We focus on realization of more efficient transcoding from MPEG-2 bitstream by changing image quality between foreground and background, which were extracted in realtime from the bitstream.

Key words moving area, foreground, camera motion parameter, transcode, MPEG

1. はじめに

近年の MPEG フォーマットの動画像の増加により、様々な帯域の混在するネットワーク上で MPEG 動画像のリアルタイム伝送が行われている。この際、いったん符号化されてしまっているストリームをネットワークに適したフォーマットへ変換するトランスコーディングに関する研究が盛んに行われている。

これまでのビットレート削減トランスコーダは動画像を伝送帯域に適したビットレートに変換するための再圧縮処理器であり、ビットレート削減に伴う画質の劣化は全体に一律であった。

動画像配信を効率的に行うという観点において、画質は最も重要な要素の一つである。ビットレートの削減に伴い画質も全体に劣化してしまうのでは、画像符号化方式として効率的ではない。ビットレートの削減を行う過程で、ユーザの興味領域 ROI (Region Of Interesting) の画質を部分的に保つことができれば、ユーザの好印象を得ることができる。

そこで、符号化されている MPEG-2 ビットストリームに含まれる情報から実時間で動領域を抽出することによって、動領域以外の領域の画質を落とし、動領域の画質を保った低ビットレートの MPEG-2 ビットストリームを生成する手法を検討の対象とする。想定システムを図 1 に示す。

2. 従来手法

動画像から人や物などの物体を抽出する研究は盛んに行われており、一般にはオブジェクト抽出、動領域抽出などと呼ばれる。従来手法としては大きく分類すると、

(1) 背景差分法などに代表される画素領域での処理による手法 [1]

(2) 動きベクトルなどを用いた圧縮領域での処理による手法 [2]

の 2 種類がある。ここではそれらの問題点を述べる。

前者の画素領域での処理を用いる手法は、精度は比較的よいものの、符号化されたデータに対して処理を行う場合、フルデコード、つまり画素値までデコードしなければならず、さらに領域抽出処理時間

がかかるので、二段階の処理になり、実時間での動領域抽出は困難であるという問題点があり、後者にはカメラにパン、チルト、ズームなどの動きがあると抽出が困難であるという問題点がある。

提案手法が想定しているシステムはトランスコーダであるので、実時間処理が不可欠である。また、カメラモーションに限られてしまうのでは動画像配信として実用価値が低い。そこでこの 2 つの問題点を解決できる手法を提案する。

3. 提案手法

本研究で提案する手法の特徴として、以下のようなものを挙げる。

● 実時間処理

入力画像として、MPEG-2 符号化方式で符号化された動画像を対象としているので、その動画像をフルデコードすることなく、ビットストリーム中に含まれる情報から実時間で動領域を抽出する。これは従来の画素領域処理に比べて高速処理につながる。

● カメラモーションへの対応

カメラモーションに対応するために、まずカメラモーションを判定し、カメラモーションごとに動領域判定方法を切り替えることによって対応させる。これによって従来の圧縮領域処理では対応していなかったカメラモーションへの適応を可能にする。

3.1 カメラモーション

高速処理とカメラモーションへの対応を考えて、カメラモーションによって処理を切り替え、それぞれフルデコードなしでの動領域判定法を提案する。ここでカメラモーションの判定に文献 [3] を用いる。ただし、この手法ではカメラモーションの判定を GOP 単位で行っている。本手法ではこれをフレーム単位で判定する。これによって多少精度が劣化する。しかし GOP 単位の判定を行うと、カメラモーションの切り替えの場所が GOP の途中であった場合、その GOP 全体がどちらかのカメラモーションと判定されてしまうという問題が生ずる。その結果、カメラモーションごとに処理を切り替えると、誤った処理を行ってしまう可能性がある。そこでフレーム単位の判定を採用した。ただし、B ピクチャではカメラモーションの誤判定 (特にパン) が多いため、前の I、P ピクチャと同じカメラモーションをとるものとした。

3.2 カメラ固定時の動領域抽出手法

3.2.1 条件

動領域について、4 マクロブロック以上の大きさを持っているものとした。これは最後の整形処理で縮小処理を行ったときの孤立点とならないための最低の大きさである。また、あまり小さい動領域はユーザの興味領域とはならない場合が多いことによる。

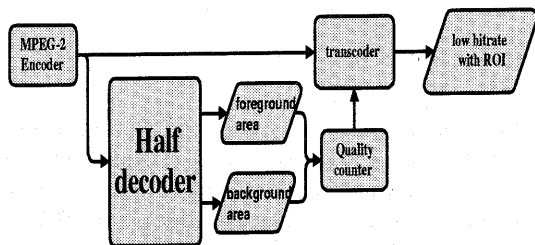


図 1 想定システム図

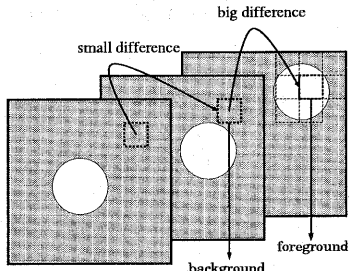


図2 動き補償予測誤差利用の概念図

3.2.2 動き補償予測誤差

動領域判定には概念的には動き補償予測誤差の値を用いる。動き補償予測の処理の中で、動きベクトルを求める際に f_code に基づいた範囲だけマクロブロックをマッチングさせ、その中で最も差分の小さいマクロブロックを選ぶ。そのマクロブロック同士の係数の差を動き補償予測誤差と定義する。今まで背景であった領域に動領域が来た場合、そのマクロブロックとの差分は非常に大きくなると考えられるので、そのマクロブロックは動領域の可能性が高いと考える。概念を図2に示す。

ところが、動き補償予測誤差が求められるのはデコード処理でも終段に相当するため、始めに述べたような高速処理には向かない。そこで実際の演算には輝度成分のDCT係数最高次係数位置を用いる。これは「動き補償予測誤差が大きいブロックは0でないDCT係数(有効係数)が高次まで存在する確率が高い」、「量子化パラメータに依存せず、動き補償予測誤差に比例して輝度成分の有効係数の最高次数位置は高くなる」[4]という先行研究による。動き補償予測誤差の二乗誤差と輝度成分のDCT係数最高次係数位置の関係を図3に示す。これは動画像中のすべてのマクロブロックに対して関係をプロットしたものである。図3より、動き補償予測誤差が大きいほど、最高次数位置も高くなる確率が高いことがわかる。

3.2.3 有意ブロック

有意ブロックとは符号化されているブロックのことを指す。MPEG-2では予測誤差の小さいブロック

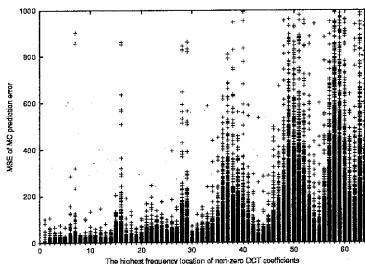


図3 動き補償予測誤差とDCT係数最高次位置の関係

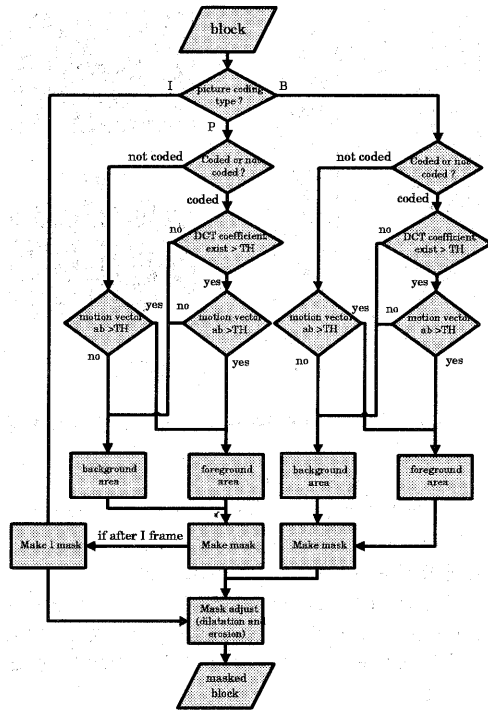


図4 カメラ固定時の動領域抽出アルゴリズム

は有意ブロックとは判定されず、符号化されずにスキップされてしまう場合がある。このスキップされた理由についても2つの解釈ができる。したがって、スキップの状態と動きベクトルの絶対値を関連させて判定材料に加える。

- (1) 背景領域で予測誤差が少ないためにスキップされた。(動きベクトル小)
- (2) 動領域だが動き補償したらうまくマクロブロックが重なり、予測誤差が小さくなったためスキップされた。(動きベクトル大)

3.2.4 処理手順

処理手順は以下の通りである。これらの処理は各ブロック(8*8画素)について行う。カメラ固定時の動領域判定法のアルゴリズムを図4に示す。

- (1) まずはじめに $picture_coding_type$ より、ピクチャタイプによって閾値を変化させる。
- (2) ブロックごとにそのブロックが有意ブロックかどうかを $coded_block_pattern$ から判断する。
- (3) DCT係数最高次係数位置で閾値処理。
- (4) 動きベクトルの絶対値で閾値処理。
- (5) ここで有意ブロックであり、DCT係数、動きベクトルともに閾値以上の領域を動領域とする。
- (6) また、有意ブロックでなくても動きベクトルが閾値以上の領域も動領域とする。

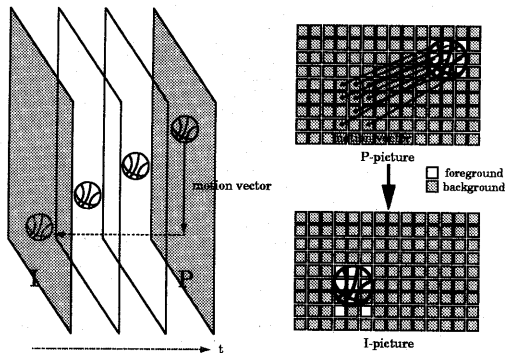


図5 I-picture の処理

(7) 後述する整形処理によって最終的な動領域を得る。

3.2.5 Iピクチャでの処理

Iピクチャでは動きベクトルやDCT係数の差は存在しないため、ビットストリームに含まれる情報から動領域を判定することは困難である。

そこでそのIピクチャを参照しているPピクチャを用いる。このPピクチャで抽出された動領域が、予測の際に参照しているIピクチャの領域を動領域とした。

つまり、図5に示すように、I-pictureの後の初めてのP-pictureの動領域判定の終了後、動領域と判定されたマクロブロックを、そのマクロブロックが持っている動きベクトル分だけずらすことで、そのずらした領域を前のI-Pictureの動領域とする。

3.3 カメラモーションがある場合

3.3.1 条件

カメラモーションにはズーム、シェイク、パン、チルトなどがあるが、本研究ではパニングに限って行っている。また、撮影者は動物体を追って撮影しているものとした。

3.3.2 動きベクトル

パニング区間では全体に動きベクトルが発生し、動き補償予測誤差の値は全体的に大きくなると考えられるので、前景領域判定にはこの値を用いない。パニングの場合には動きベクトルの値を使う。まず、各マクロブロックの動きベクトルのx方向成分について閾値処理を行い、絶対値の少ない領域を前景領域とする。これは撮影者が動物体を追って撮影していることから、背景に比べて前景領域のカメラ内での動きが少ないと考えられるからである。さらに、Bピクチャでマクロブロックに関して2つの動きベクトルがあった場合、2つのベクトルの角度差から動きの連続性を得て、連続性のあるのはカメラモーションによるものだとその領域を背景領域にする。フレーム内処理としてピクチャタイプとマクロブロックタイプに応じて以下のように処理を場合分けし、

マクロブロック単位で背景領域か人物領域かを判定する。

- P-Picture

前方向予測が行われている場合は、動きベクトルのx方向成分の大きさを閾値処理する。予測が行われていない場合は、カメラが動物体を追って撮影したために動物体が止まって見えたものとして動領域とする。Intra符号化されている場合は、探索範囲に参照マクロブロックがなかったと判定されたとして背景領域とする。

- B-picture

前方向・後方向予測が行われている場合はP-Pictureと同様に動きベクトルのx方向成分の大きさを閾値処理する。双方向予測が行われている場合は、信頼性がある場合には動きベクトルを閾値処理し、信頼性のない場合には判定には用いない。intra符号化されている場合は判定には用いない。

信頼度については、前方向予測と後方向予測ベクトルのなす角度が理想的には180度だが、45度の誤差をとり、135度から225度のときに信頼性があるとした。動きベクトルの大きさにおける閾値処理では、画面全体の動きベクトルのx方向成分について平均を取り、その値をフレームでの最適閾値とした。最適閾値を境に、小さいものを動領域マクロブロック、大きいものを背景マクロブロックとした。

- I-picture

I-pictureに関してはビットストリームに有用な情報が存在しないため、カメラモーションがないときと同様の処理から動領域を求める。

またこの処理で得られた領域に後述する整形処理を行い、領域を整形する。この手法の前提としてカメラは動物体を追って撮影しているとので、動領域はフレームごとに大きな移動は見られないと考えられる。そこで得られた動領域に対して膨張処理を行った後、前フレームでの動領域とAND処理を行うことで誤検出領域を削除する。

3.3.3 処理手順

処理手順は以下のとおりである。これらの処理は動きベクトルがマクロブロック単位でしか値を持たないため、動領域判定もマクロブロック単位(16*16画素)で行う。パニング時の動領域判定アルゴリズムを図7に示す。

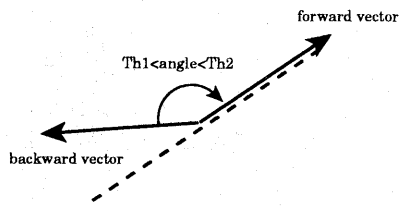


図6 動きベクトル信頼度

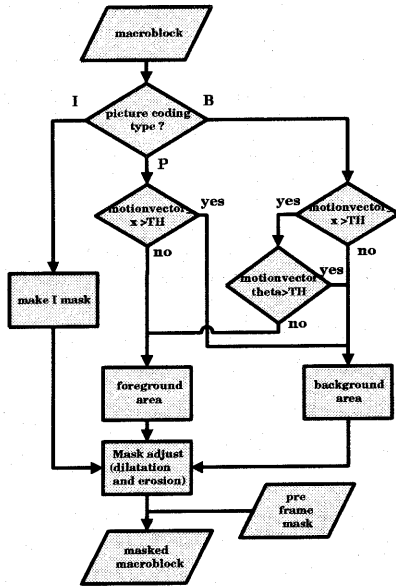


図7 パニング時の動領域抽出アルゴリズム

- (1) まずはじめに picture_coding_type より、ピクチャタイプによって閾値を変化させる。
- (2) 画面全体の動きベクトルを格納し、x成分について閾値処理をする。
- (3) Bピクチャだった場合、動きベクトルの信頼度を求める。
- (4) 動きベクトルのx成分が小さく、信頼度の高い動きベクトルを持つマクロブロックを動領域とする。
- (5) 後述の整形処理を行い、動領域を整形する。
- (6) 前フレームの動領域を膨張処理で膨らませ、求められた動領域とAND処理を行い、誤抽出領域を除去する。

3.4 整形処理

上述の手法で動領域抽出法では誤検出、未検出領域の発生は避けられない。そこで、それらを少しでも減らすためにモルフォロジー理論を元にした整形処理を行う。具体的には得られた動領域に対して膨張処理→収縮処理で穴埋めをし、収縮処理→膨張処理で孤立点の除去を行う。この処理を繰り返すことによって未検出、過剰検出領域をある程度除去することができる。動領域抽出結果と、それに対して膨張収縮処理を4回ずつ行った結果を8に示す。

3.4.1 膨張処理

背景領域ブロックに注目したとき、そのブロックを中心とする3×3ブロックの領域中に1つでも動領域ブロックがあればその注目ブロックを動領域ブロックに変える処理を、全体に独立に行う。

これにより孤立した未検出領域を除去することが

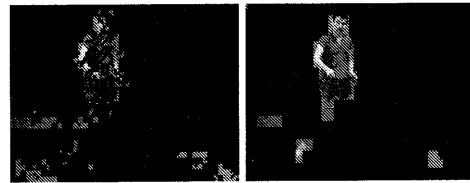


図8 動領域抽出結果(左)と、整形処理結果(右)

できる。

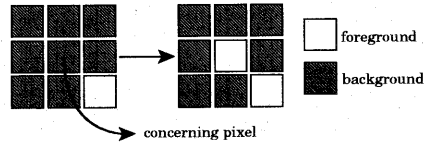


図9 膨張処理

3.4.2 縮小処理

動領域ブロックに注目したとき、そのブロックを中心とする3×3ブロックの領域中に1つでも背景ブロックがあればその注目ブロックを背景ブロックに変える処理を、全体に独立に行う。

これにより誤検出された動領域ブロックを除去することができる。

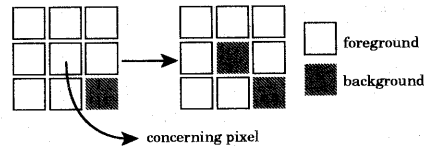


図10 縮小処理

4. 実験結果

4.1 実験条件

今回の実験では動画像の中の動物体は1つで、カメラモーションについてはパニングに限っている。またパニング時は動領域を追って撮影しているものとした。実験画像について表1, 2に示す。

4.2 結果

提案手法に用いた閾値を表3に示す。カメラ固定時の動領域抽出結果を図11にそれぞれ示す。パニング時の動領域抽出結果を図12にそれぞれ示す。

表1 実験画像 skate12M.m2v

MPEG-2 Sequence factor	Parameters
Frame Rate	30[frame/s]
Number of frames in GOP	15[frame]
I/P Frame Distance	3[frame]
Frame size	720*486[pixel]
Bit Rate	12[Mbps]
The number of frames	600[frame]

表2 実験画像 skate12M.m2v の内容

フレーム	0~48	48~199	200~529	530~589	590~600
動物体	なし	あり	あり	あり	なし
カメラ	固定	固定	パン	固定	固定

また、動領域が抽出できたとと思われるフレームの割合を表4に示す。

これを見ると、従来手法では動領域の抽出が不可能であった200frame~529frameのパニングの区間で動領域の抽出が可能になっている事がわかる。

また、カメラ固定の区間において多少提案手法のほうが従来手法より抽出割合の低下が見られる。これはカメラモーションの誤検出により、カメラ固定の区間にも関わらずパニング時のアルゴリズムで動領域抽出処理を行ってしまったため、動領域が抽出できなかったものと考えられる。

表3 実験に用いた閾値

threshold	value
common	
repetition of erosion	4
repetition of dilatation	4
camera motion fix	
DCT coefficients threshold in P-picture	40
DCT coefficients threshold in B-picture	50
motion vector threshold in P-picture	5
motion vector threshold in B-picture	3
camera motion panning	
motion vector threshold in P-picture	2
motion vector threshold in B-picture	3
motion vector angle	$135 < angle < 225$

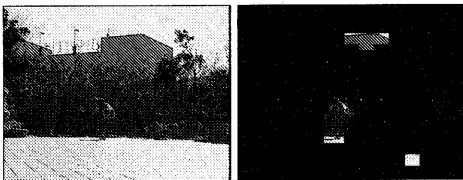


図11 元画像(左), 動領域(右)(カメラ固定時)

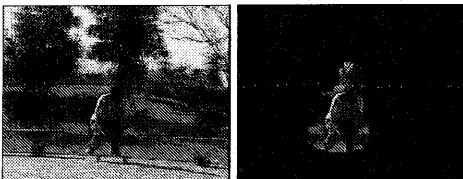


図12 元画像(左), 動領域(右)(パニング時)

表4 動領域抽出割合(%)

フレーム	0~48	48~199	200~529	530~589	590~600
従来手法	0	69	0	95	0
提案手法	0	65	77	89	0

4.3 考察

• 全体的に P-frame での判定精度が B-frame に対して高かった。これはある程度はなれたフレームからのほうがその物体の動きを反映するような予測をする確率が高いからだと考えられる。

• 本手法では閾値処理を行っているため、画像によって動領域の判定精度が閾値によって異なる、同じ動画内でも一定の閾値では領域抽出が困難な場合があった。さらに空などの動きベクトルが散ってしまう領域では、背景領域を動領域と誤検出してしまっていることがあった。

• カメラの動きを限定しているため、パニング以外のカメラモーション、たとえば手ぶれなどが加わると判定が困難であった。

5. まとめ

本論文では、効率的なトランスコーディングを目的とした、動画像 MPEG-2 ビットストリームからの動領域抽出法について述べた。提案手法では、従来実時間処理では不可能であったカメラモーションのある動画像からの動領域抽出処理が可能になった。

今後の研究課題として以下のようなものを考えている。

• パン以外のカメラモーションへの適応
GMV 算出する [5], [6] ことによってさまざまなカメラモーションに対応させていきたい。

• 閾値の決定法の検討
一定の閾値処理では限界があるので、動的に閾値を変える手法を検討したい。

• トランスコーダへの実装
実際にトランスコーダに手法を実装しての符号化実験を行い、画質の評価、符号量の評価、さらに量子化ステップの変更アルゴリズムなどの検討なども必要である。

文 献

- [1] 境田真一, 金次保明, "背景差分と時空間 watershed による領域成長法を併用した動画像オブジェクトの抽出," 信学技報, IE99-151, Mar. 2000.
- [2] 米山暁夫, 中島康之, 柳原広昌, 菅野 勝, "MPEG ビットストリームからの移動物体の検出" 電子除法通信学会論文誌 D-II, vol. J81-D-II, NO. 8, pp1776-1786, 1998
- [3] 土橋健太郎, "手ぶれを考慮したカメラワーク検出に関する検討," 平成 12 年度 早稲田大学卒業論文
- [4] 大迫史典, 八島由幸, 小寺 博, 渡辺 裕, 島村和典, "動的演算量スケーラブルアルゴリズムによるソフトウェア画像符号化," 電子情報通信学会論文誌 D-II, vol. J80-D-II, NO. 2, pp444-458, 1997
- [5] 秦泉寺久美, 石橋 聡, 小林直樹, "カメラモーション抽出によるスプライト自動生成," 電子情報通信学会論文誌, vol. J82-D-II NO.6, pp1018-1030, Jun. 1999.
- [6] 秦泉寺久美, 渡辺 裕, 小林直樹, "スプライト作成のためのグローバルモーション算出法と符号化への適応," 電子情報通信学会論文誌, vol. J83-D-II NO.2, pp535-544, 2000.