CH-002

時空間マクロブロックタイプパターンを用いた圧縮動画検索

Compressed Video Search with Spatio-temporal Macro Block Type Patterns

祖泉 大河 * 岡村 和磨 * 森田 啓義 * 眞田 亜紀子 †
Taiga Soizumi Kazuma Okamura Hiroyoshi Morita Akiko Manada

1 はじめに

デジタルビデオカメラ関連技術の著しい進化によって高解像度での映像の撮影が身近となってきている．また，テレビ放送においては，視聴者は HD 映像の地上デジタル放送に対応したテレビやレコーダーによって容易に高解像度の映像を録画・保存することができる．加えて，近年では 4K や 8K といわれる解像度がさらに高い映像が放送され始められており，今後これらの高画質映像も同様に容易に入手できるようになると考えられる．しかし，このように高画質な映像ではデータ量が膨大になる上に，保存する映像数が増すにつれ，ユーザが見たいシーンを高速に検索する技術の開発が急務である．

従来の動画検索ではタイトルなどのキーワードを検索キーとしていたが, Derphins ら [1] は歩く，手を振るといった動作を含むビデオクリップ (短い動画) を検索キーとする手法を提案している．

本研究では，圧縮動画における符号化パターンが動作のパターンとして抽出できることを期待する．これにより，人の動作を含むビデオクリップを検索キーとしながら，圧縮動画の復号をせず高速で検索対象の動画から同じ動作を検索するシステムの提案を目指す．

2 従来手法

Derpanis らの研究 [1] では検索キー動画と検索対象の動画それぞれから，空間情報だけでなく時間情報も利用した動きの特徴量（Spatiotemporal Orientation）を抽出する．そして，検索対象の動画内で検索キー動画を全ての時空間位置でスライドさせながら特徴量の類似度の計算を行い，同じ動作が現れる時空間的な位置を特定する．圧縮動画を復号化した時空間画像をボクセル単位で相関をとる解析を行っており，動作している人の外観の変化や背景・前景のノイズに頑健で，GPU に基づくリアルタイム処理が可能である．

しかし，実験に使用している動画は，検索キー動画では解像度が 50×25，検索対象動画では解像度が 144×180 で 10fps という低品質の動画であり，より高品質な映像をリアルタイムに処理することは困難である．

3 時空間 MBT パターンを用いた動体領域の推定

3.1 マクロブロック

MPEG-2 では 16×16 などの画素の集まりをマクロブロックと呼び，圧縮の際にはマクロブロックごとに符号化方式を決定する．

3.2 マクロブロックサイズ

通常マクロブロックの大きさは 16×16 画素だが，マクロブロックの大きさは可変であり，分割されて 16×8 画素となる場合がある．分割マクロブロックは動体の境界付近に発生しやすいことが報告されている [2]．しかし，複雑な領域においても分割マクロブロックが発生してしまう傾向があり，分割マクロブロックの検出のみでは静止している物体の領域まで動体として検出してしまう．

3.3 MBT

MPEG-2 ではフレーム間の差分を利用したフレーム間予測が用いられており，マクロブロックごとにどのフレームを利用するかを決定し符号化する．符号化方式は以下の 4 種類がある．

- 順方向予測符号化
  過去の画像から予測する符号化方式
- 逆方向予測符号化
  未来の画像から予測する符号化方式
- 双方向予測符号化
  順方向予測と逆方向予測の予測を平均する符号化方式

*電気通信大学大学院情報理工学研究科
†湘南工科大学工学部情報工学科

図 1: MBT の解析例

表 1: ピクチャタイプと使用可能な符号化方式

	順方向予測	逆方向予測	双方向予測	フレーム内符号化
I ピクチャ	×	×	×	○
P ピクチャ	○	×	×	○
B ピクチャ	○	○	○	○

- フレーム内符号化
  他のフレームを参照せずフレーム内の情報のみから符号化する．

符号化の際に選択された予測方式が MBT（マクロブロックタイプ）として記録される．後述するように，時空間における MBT の変遷を追うことで動体の検出が可能となる．

3.4 ピクチャタイプ

使用可能な符号化方式はフレームごとに異なる．I ピクチャ，P ピクチャ，B ピクチャの 3 種類のピクチャタイプがあり，それぞれのピクチャタイプごとに使用可能な符号化方式が定められている．

I ピクチャ，P ピクチャ，B ピクチャの割り当てはエンコーダによって異なるが，本研究では図 2 に示すようにピクチャタイプが割り当てられた GOP（Group of Pictures）を周期とする動画を用いている．

3.5 時空間 MBT パターン

動体が映っている領域の MBT に注目すると P ピクチャではフレーム内符号化，連続する B ピクチャにおける前半では順方向予測，後半では逆方向予測となる傾向がある．

図 2: GOP（Group of Pictures）

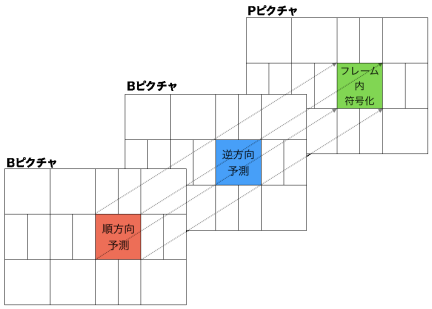

図 3: 時空間 MBT パターン

図 4: 時空間 MBT パターン検出例

図 4 に時空間 MBT パターンの検出例を示す．両腕を交互に前後させるボクシング動作（同図左）における腕の部分や，両手を身体の前で叩く動作（同図右）の際に開かれた腕に時空間 MBT パターンが発生していることがわかる．

したがってピクチャタイプが BBP という並びの 3 枚のフレームにおいて，「順方向予測→逆方向予測→フレーム内符号化」というパターンを示しているマクロブロックを検出することによって，動体の領域の検出が可能となる．

3.6 動体領域の推定

マクロブロックサイズと時空間 MBT パターンを組み合わせて動体領域を推定する方法を提案する．

まず動体領域に発生する時空間 MBT パターンを検出する．これを起点とし，その周囲にある分割マクロブロックを動体領域と判定する．この際，上・下・左・右の 4 マクロブロックだけでなく，右上・左上・左下・右下の 4 マクロブロックも合わせた 8 マクロブロックを周囲のマクロブロックとして扱う．そしてこの時点で動体領域と判定されている領域のさらにその周囲においても分割マクロブロックを動体領域と判定する処理を再帰的に繰り返す．この再帰処理が終了した時点で得られた領域が推定される動体領域となる．

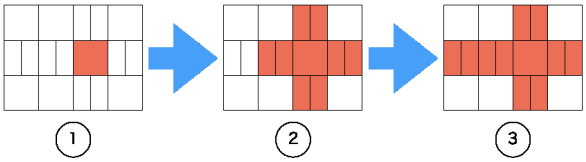

図 5: 動体領域の推定

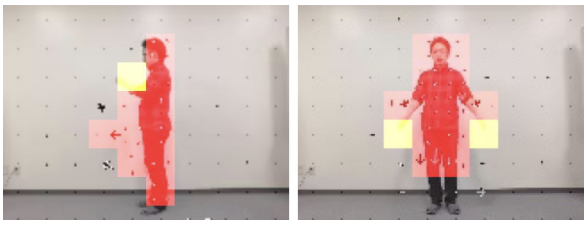

図 6: 動体領域の推定適用例

4　投球シーンの検出

4.1　検出方法

4.1.1　検出方法概要

図 7 のイメージのように野球の試合映像において，投球シーンは決まったカメラで撮影されており，カメラが固定されていることを利用する．したがって，各フレームにおいてカメラが固定されているか判定，投球動作が行われているかの判定を行い，両方とも満たすフレームを投球シーンが行われているフレームであると判定する．投球動作は 1，2 秒かかる動作であるため，常に再生時間にして直近 1 秒分に相当するフレームを用いて両判定を行う．

また，映像を切り替える瞬間であるカット点は直後の MBT に影響を与えるため，カット点から時間的に 3 枚未満のフレームでは投球シーンの検出を行わない．加えて，投球シーンを重複して数えてしまうことを避けるために，一度投球シーンを検出した際には再生時間にして 10 秒分に相当するフレームの間でも検出を行わない．

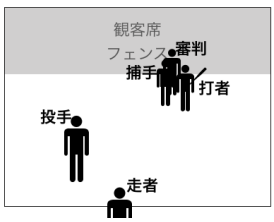

図 7: 投球シーンのイメージ

4.1.2　カメラ固定判定

投球シーンは固定されたカメラで撮影されるため，投球シーンではフレーム内の動きベクトルは零ベクトルが多くを占める．直近 1 秒間分のフレームで動きベクトルを解析し，零ベクトルが半数以上になっているフレームの数が 9 割以上であれば，カメラが固定されていると判定する．ただし，動きベクトルがノイズ的に発生することもあるため，零ベクトル優位になっているフレームがあればその前後も，零ベクトル優位として数える．

4.1.3　投球動作判定

投球シーンでは投手は画面左下の領域に位置し，映る人物の中で最も画面内での占有面積が大きい．P ピクチャごとに直近 10 セットの BBP ピクチャ列において動体領域の推定を行い，画面左下における動体領域のマクロブロック数が画面左上と画面右下よりも多く，かつ画面左下における移動体領域の占める割合が 5%以上であるフレームを数える．カメラ固定判定と同じく，該当するフレームがあればその前後も該当フレームとして数える．最終的に該当しないと判定されたフレームが 1 枚以内であれば投球動作が行われていると判定する．

4.2　投球シーン検出実験

4.2.1　実験環境

提案手法のプログラムを C 言語を用いて作成した．また，実装にあたっては Open CV[3] ならびに FFmpeg[4] のライブラリを用いた．計算機実行環境を以下に示す．

- MacBook Pro(Retina, 15-inch, Mid 2015)
- OS X El Capitan バージョン 10.11.6
- プロセッサ　2.5 GHz Intel Core i7
- メモリ　16GB 1600 MHz DDR3

映像には解像度が 720 × 480 画素，フレームレート 30fps の SD 映像のものを使用した．試合は，映像 1，映像 2 が高校野球の試合で，映像 3 はメジャーリーグで行われた試合である．映像 1 には試合だけでなく野球中継の次の番組が含まれている．映像 2 は野球中継だけの映像である．映像 3 にはニュース映像が含まれている一方で，野球中継は試合の途中から始まっているため投球シーン数が映像 1，映像 2 に比べて少ない．球場は映像 1 と映像 2 が同一であるが，映像 3 は異なる球場である．

表 2: 投球シーン検出結果

	TP	FP	FN	Precision	Recall
映像 1	193	5	17	0.975	0.919
映像 2	262	16	21	0.942	0.926
映像 3	118	6	0	0.952	1.000

表 3: 先行研究との比較

	Precision	Recall
本研究	0.95	0.95
[5]	0.99	0.97
[6]	0.60	0.81

プログラムによって検出された投球シーンと元の映像と照らし合わせ，目視で投球シーンの検出結果を判断した．検出結果の判定基準としては，投手が投球動作に入ってから球が本塁に到達するまでの間に検出できた場合は TP（正検出）とした．一方で牽制なども含めて投球シーンではないところでの検出は FP（誤検出），一連の投球シーンが映っていたにもかかわらず検出できていない場合は FN（未検出）とした．

4.2.2 実験結果

表 2 は実験結果と結果をもとに計算した Precision（適合率）と Recall（再現率）である．

本研究で得られた Precision と Recall の平均値について，他の投球シーンを検出する先行研究での Precision と Recall それぞれの平均値と比較を行った．[5] はマクロブロックの色情報を用いて投球シーンを検出する手法，[6] は音響解析を用いて投球シーンを検出する手法 [6] である．

本研究での結果と手法 1 を比較すると，Precision は約 4%，Recall は約 2%ほど劣る結果となった．しかし，手法 1 では試合映像以外の映像が含まれていない映像で，同一球場の映像のみの研究であることに留意する必要がある．一般的な試合映像に対する研究においては，音響解析を用いた手法 2 に対して Precision は約 35%，Recall は約 14%高い結果が得られた．

4.2.3 実行時間

通常の再生時間と本プログラムの実行時間，また再生時間に対する実行時間の割合は表 4 のようになった．

プログラムの実行結果はいずれも映像の再生時間の 60%未満である．これは本アルゴリズムを用いれば，市販のコンピュータで SD 映像のリアルタイムな処理が十分に可能であることを示している．また，残りの 40%余りの時間を利用することによってさらなる精度向上を期待することができる．

表 4: 再生時間と実行時間の関係

	再生時間	実行時間	実行時間割合
映像 1	6378 秒	3478 秒	54.5%
映像 2	7529 秒	4180 秒	55.5%
映像 3	4688 秒	2600 秒	55.5%

5 まとめ

動体が存在する時空間領域では特有の MBT のパターンを示すことがわかった．また，この時空間マクロブロックパターンと小マクロブロックの特性を組み合わせることにより，リアルタイムでの処理が可能で，球場に依存しない高精度な投球シーン検出法を提案することができた．

今回の実験では投球シーンについての考察を十分に行うことにより，投球シーンの検出プログラムを実装したため，このプログラム自体は投球シーン以外の検出に用いることはできない．しかし，動体領域の特定アルゴリズムは同じ GOP 構成を持つビデオデータについては一般的に用いることができるものである．今後は，投球シーンに限らず様々な動画において本研究で提案した動体領域を利用し，投球に限らない動作の検出に関する研究を行う．

参考文献

[1] K. G. Derpanis *et al.*, “Action spotting and recognition based on a spatiotemporal orientation analysis,” Trans. on PAMI, vol. 35, pp. 527-540, 2013.

[2] 単鴻, “MPEG2 動きベクトルを用いた複数動体の検知・追跡システム,” 修士論文，電気通信大学 (2008)

[3] OpenCV, http://opencv.jp/

[4] FFmpeg, https://www.ffmpeg.org

[5] 志水一也, “ MPEG2 を用いたスポーツ映像の自動解析によるハイライトシーンの抽出, ” 電気通信大学大学院情報システム学研究科修士論文, 2006.

[6] 三上弾, 紺谷精一, 森本正志, “ 突発音検出と教師なし動きクラスタリングを用いた野球映像からの投球イベント検出, ” 電子情報通信学会論文誌 D, Vol.J90-D, No.2, pp.526-534, 2004.