

グローバル動きを用いた高速動画像モザイク手法  
A Fast Video Stitching Method Using Global Motion Estimation

清水 智行<sup>†</sup>  
Tomoyuki Shimizu

米山 暁夫<sup>†</sup>  
Akio Yoneyama

滝嶋 康弘<sup>†</sup>  
Yasuhiro Takishima

## 1. まえがき

本研究では、H.264 で符号化された複数の動画像を入力し、動き補償フレーム(P フレーム)に関して、空間方向に接合して 1 つの動画像として高速合成する動画像モザイク手法を提案する。

複数の静止画からモザイク画像を生成する手法として、KLT[1]等を用いて画素レベルでの特徴点抽出を利用したものがある。このような手法をフレーム毎に実行することによってパノラマ動画像を生成することは可能である。

しかし、画素単位での特徴点抽出は多くの計算量を必要とする。従って、複数の動画像の位置合わせを各フレームに対して行うためには、より高速な位置合わせ方式が必要となる。仮に、カメラを雲台に固定した場合でも微細な動きによる位置ずれが発生する可能性があるため、フレーム毎の位置合わせを行うことが必要である。

そこで本研究では、各カメラの動きを考慮して、圧縮符号化された動画像データの高速モザイクを実現する手法を提案する。

## 2. 高速位置合わせ手法

本稿では、I フレームでの位置合わせは手動で与えられているものと仮定し、後続の P フレームに関して議論する。

本手法では、フレーム  $n$  における位置合わせを表す射影行列  $\mathbf{R}_n$  を補正することによって後続のフレーム  $n+1$  における射影行列  $\mathbf{R}_{n+1}$  を求める。 $\mathbf{R}_{n+1}$  の計算は 2 つの段階に分かれる。第 1 段階では位置合わせを表す射影行列を求めて大まかな位置合わせを行い、第 2 段階で微調整を行う。

### 2.1 第 1 段階: グローバル動きを用いた位置合わせ補正

動画像  $B$  の座標系における任意の点  $P_B^n$  が動画像  $A$  の座標系における点  $P_A^n$  に写像されるものとする。このとき、射影行列を  $\mathbf{R}_n$  とすると、式(1)の関係が成立する。

$$\vec{P}_A^n = \mathbf{R}_n \vec{P}_B^n \quad (1)$$

同様に、フレーム  $n+1$  について、

$$\vec{P}_A^{n+1} = \mathbf{R}_{n+1} \vec{P}_B^{n+1} \quad (2)$$

点  $P_A^n$  がフレーム  $n+1$  において  $P_A^{n+1}$  に移動するとき、動画像  $A$  のフレーム  $n+1$  とフレーム  $n$  の間のグローバル動きを  $\mathbf{V}_A^n$  とすると、式(3)が成立する。グローバル動きは符号化データの動きベクトル(以下、MV)を利用することによって、高速に求めることが可能である[2, 3, 4]。

$$\vec{P}_A^{n+1} = \mathbf{V}_A^n \vec{P}_A^n \quad (3)$$

本稿では、これらの行列はアフィン変換とし、最小二乗法によって計算することとする。動画像  $B$  について同様に、

$$\vec{P}_B^{n+1} = \mathbf{V}_B^n \vec{P}_B^n \quad (4)$$

式(3), (4)を式(2)に代入して、

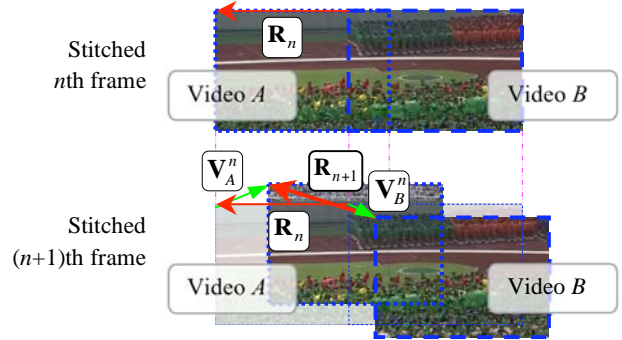


図 1: 位置合わせとグローバル動き

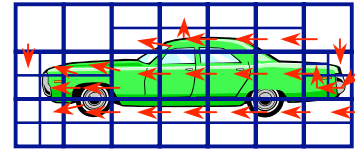


図 2: 可変ブロックサイズ動き補償と移動物体

$$\mathbf{V}_A^n \vec{P}_A^n = \mathbf{R}_{n+1} \mathbf{V}_B^n \vec{P}_B^n \quad (5)$$

よって、式(1), (5)より、

$$\mathbf{R}_{n+1} = \mathbf{V}_A^n \mathbf{R}_n (\mathbf{V}_B^n)^{-1} \quad (6)$$

式(6)で表される関係は図 1 のようになる。すなわち、フレーム  $n$  における動画像  $B$  から動画像  $A$  への射影行列  $\mathbf{R}_n$  は、フレーム  $n+1$  における射影行列  $\mathbf{R}_{n+1}$  とグローバル動き  $\mathbf{V}_A^n$ ,  $\mathbf{V}_B^n$  から簡単な計算で得られる。

### 2.2 H.264 におけるグローバル動きと移動物体

前節で述べた位置合わせにおいて、動画像のグローバル動きを動きベクトルから導出するには、移動物体の動きの影響を排除する必要がある。

従来方式[2][7]では、MPEG-2 を対象として、動きベクトルの分布や予測誤差に基づいて移動物体領域の判定を行う。しかし、H.264 では多彩な符号化モードが採用されており、モード選択の傾向が大幅に異なるため、従来手法をそのまま適用することは困難である。そこで本研究では、H.264 の符号化モードに着目した移動物体抽出を行う。

まず、イントラ符号化モードのマクロブロック(以下、MB)に関しては、動き情報を含んでいないため、除外する。

フレーム間予測モードの MB に関して、H.264 では、可変ブロックサイズの導入により動き補償予測における予測誤差を大幅に減少させ、符号化効率の向上を実現している。特に、移動物体のエッジ付近において、エッジ形状やテクスチャの複雑さに応じたブロックサイズが選択される傾向がある(図 2)。従って、小さなブロックサイズのフレーム間予測 MB を除去することによって、精度よく背景領域を抽出してグローバル動きを求めることが可能となる[5, 6]。

そこで、本研究では、グローバル動きの計算に用いる MB は、 $16 \times 16$ ,  $16 \times 8$ ,  $8 \times 16$  の 3 種類のブロックサイズが

<sup>†</sup> (株)KDDI 研究所, KDDI R&D Laboratories Inc.

選択されているものに限定することとし、 $8 \times 8$ 以下のブロックサイズのMBについては除外することとする。

さらに、[7]と同等の手法により移動物体と背景領域の分別を行う。

### 2.3 第2段階: 位置合わせの微調整

式(6)で求められる射影行列の精度は、グローバル動きの抽出精度に依存する。しかし、圧縮符号化された動画の動き情報から求められるグローバル動きは誤差を含む場合が多い。そこで、第2段階では位置合わせの微調整を行う。

重なり領域においては、射影行列の誤差の影響で動画A, Bの対応点 $P_A^{n+1}$ ,  $P_B^{n+1} = R_{n+1} P_B^{n+1}$ が必ずしも一致しないが、非常に近い位置を示す。そこで、以下の手順で位置合わせの微調整を行う。

1. 重なり領域の各MBにおいて、整数画素精度でブロックマッチングを行い、両MBの画素の差分絶対値和(SAD)が最小となる位置を求める。これは、動きベクトルを求める過程と同様の手順である。なお、誤差補正が目的であるため、探索範囲は $\pm 4$ 画素とする。
2. 1.で得られた対応点の周辺において1/2画素精度のブロックマッチングを行う。画素の補間はH.264のハーフペル平面と同じフィルタを用いる。
3. 2.の結果に基づいて、動きベクトルを用いてグローバル動きを計算する過程と同様の手順で射影行列を求める。

## 3. 実験

提案方式の性能を検証するため以下に示す実験を行った。まず、高解像度の動画の各フレームにおいて、重なり領域が存在するように同一サイズの矩形領域を2箇所切り抜いて、2つの小さな動画を作成し、 $320 \times 240$ 画素にリサイズして、実験用の動画を作成した。使用した高解像度動画はITE標準動画FlamingoesおよびGroupgameである。なお、画像間の位置合わせはフレーム毎に1/2画素精度で少しずつずらしている。

これらの実験用動画をH.264参照ソフトウェアJM9.3で符号化したものを入力として、提案手法(第1段階のみ、第1・第2段階併用)によって位置合わせを行い、誤差の絶対値和及び分散を測定した。符号化条件を表1、実験結果を表2(a), (b)および図2(a), (b)に示す。動作環境はPentium 4 3.2GHzを搭載したWindows PCである。また、第2段階の実装ではMMX命令による最適化を行っている。

表1: 符号化条件

プロファイル	Baseline (IPPP フレーム構造)
レート歪み最適化	On
QP	28 (固定)
参照フレーム数	1
動き探索範囲	$\pm 16$
符号化フレーム数	447

表2(a): 実験結果(第1段階のみ)

入力動画	Flamingoes	Groupgame
誤差絶対値平均(x, y) [画素]	(1.073, 0.425)	(0.776, 0.583)
誤差分散(x, y)	(1.591, 0.146)	(0.641, 0.271)

表2(b): 実験結果(第1・第2段階併用)

入力動画	Flamingoes	Groupgame
誤差絶対値平均(x, y) [画素]	(0.575, 0.420)	(0.766, 0.318)
誤差分散(x, y)	(0.143, 0.099)	(0.144, 0.049)

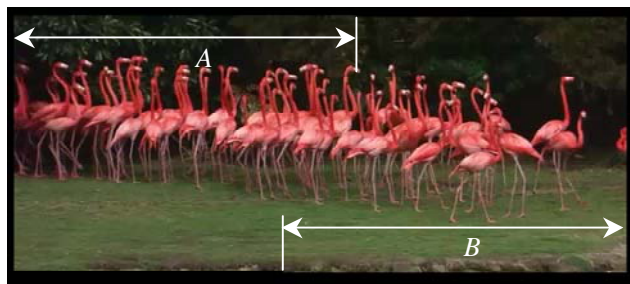


図2(a): 実験結果(Flamingoes)

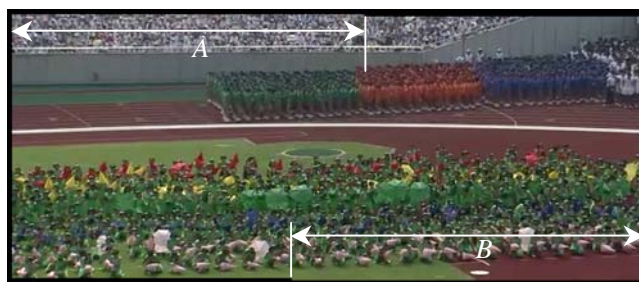


図2(b): 実験結果(Groupgame)

表2(a), (b)より、いずれのサンプル動画像についても位置合わせの誤差が0.3~1.1画素程度となっている。第1段階のみでは誤差の分散が大きく、フレームによっては非常に誤差が大きくなるが、第2段階による微調整を併用することによって分散が小さくなり、どのフレームにおいても小さい誤差で抑えられている。

処理速度については、第1・第2段階併せて1フレームあたりの位置合わせの所要時間が1ミリ秒未満となっており、毎秒30フレームの動画の位置合わせにおいてもリアルタイム処理に十分耐えることが確認できた。

## 4. 結論

本研究では、H.264の符号化情報を利用して、高速に複数の動画のモザイクングを実行する手法を提案した。本手法を用いることによって、平均して1/2画素前後の誤差で高速に位置合わせを行うことが可能となる。非動き補償フレームでの位置合わせを併せた検討が今後の課題である。

## 参考文献

- [1] S. Birchfield, "KLT: An Implementation of the Kanade-Lucas-Tomasi Feature Tracker," <http://www.ces.clemson.edu/~stb/klt/>.
- [2] H. Shen et al, "Mosaic Based View Enlargement for Moving Objects in Moving Pictures," IEEE Proc. ICME 2004.
- [3] Y. Li et al, "Robust Panorama from MPEG Video," IEEE Proc. ICME 2003, vol. I, pp.81-84, 2003.
- [4] M. F. Fu et al, "Fast Global Motion Estimation Based on Local Motion Segmentation," IEEE Proc. ICIP 2003, vol. II, pp. 367-370, 2003.
- [5] A. Chang et al, "Fast Multi-block Selection for H.264 Video Coding," IEEE Proc. ISCAS '04, vol. III, pp.817-820, 2004.
- [6] 清水他, 「H.264/AVCにおけるグローバル動き探索に関する考察」, 信学総大, D-11-51, 2005年3月.
- [7] 加藤他, 「MPEG符号化データに基づく移動物体検出追跡アルゴリズム」, 信学総大, D-11-10, 2000年3月.