

## 複数特徴点の追跡によるカメラ動き補償

上村 浩文<sup>†</sup> 石川 聖二<sup>†</sup> Krystian Mikolajczyk<sup>‡</sup>

<sup>†</sup>九州工業大学 〒804-8550 福岡県北九州市戸畑区 1-1

<sup>‡</sup> University of Surrey Guildford, Surrey, GU2 7XH United Kingdom

E-mail: <sup>†</sup>(uemura,ishikawa)@ss10.cntl.kyutech.ac.jp, <sup>‡</sup>k.mikolajczyk@surrey.ac.uk

あらまし 本論文では、複数特徴点の追跡によるカメラ動き補償法を提案する。提案法では、特徴点の検出に FAST 特徴点検出器を用い、特徴点の追跡に KLT 法及び SIFT デスクリプタを用いたロバストな特徴点追跡手法を提案し、これを用いる。各特徴点の追跡情報を画像の荒いセグメンテーション情報と結び付けることにより、フレーム間における背景の動きを複数の平面射影変換モデルとして推定する。動き補償を行うことにより、動画像中からカメラによる動きを補償した移動物体のみの動き情報の推定を行う。複数の異なる環境において移動カメラによって撮影した動画像に対して本手法を適用し、効果的な動き補償が行えるという結果を得た。

キーワード FAST, KLT, SIFT, 特徴点追跡, 平面射影変換, オプティカルフロー, カメラ動き補償

## Camera Motion Compensation Based on Multi Feature Points Tracking

Hirofumi UEMURA<sup>†</sup> Seiji ISHIKAWA<sup>†</sup> and Krystian MIKOLAJCZYK<sup>‡</sup>

<sup>†</sup> Kyushu Institute of Technology, 1-1 Sensuityo, Tobata, Kitakyushu, Fukuoka, 804-8550, Japan

<sup>‡</sup> University of Surrey, Guildford, Surrey, GU2 7XH, United Kingdom

E-mail: <sup>†</sup>(uemura,ishikawa)@ss10.cntl.kyutech.ac.jp, <sup>‡</sup>k.mikolajczyk@surrey.ac.uk

**Abstract** This paper describes an approach to local feature tracking and robust estimation of background motion under camera motion. The main contribution is a robust feature tracking algorithm based on KLT tracker and SIFT as well as a method for estimating dominant planes in the scene. We employ FAST feature detector to obtain large number of features for every frame. The motion vectors for the features are estimated using optical flow and SIFT based matching. The features are combined with the segmented parts on the image to estimate dominant homographies, and then separated into static and moving ones regardless of the camera motion. The camera motion compensation shows very good accuracy on a number of test sequences.

**Keywords** FAST, KLT, SIFT, Feature point tracking, Homography, Optical flow, Camera motion compensation

### 1. まえがき

動画像中の人物動作認識は、コンピュータビジョンの世界でより大きな興味を集める研究題材となってきた。これまでに提案された動作認識の手法[2,3,8,9,12,14]では、高い動作認識率を実現しているが、使用した映像は固定カメラにより撮影されたものが多く、背景も動きの無い単純なものが多い。しかし、現実には撮影される映像では固定カメラから得られる映像よりも、移動カメラから得られるものが多く、背景も動いていることが一般的である。Laptev&Perez は[4]において、実際の映画から切り出した映像に対して、時空間方向を考慮した Boosting を提案したが、カメラが固定されているシーンのみを用いていた。Yilmaz&Shah は[16]において、初めて移動カメラの映像を用いた動作認識を行ったが、複数のカメラや同一人物を複数の視点から撮影した映像を必要としており、

応用の範囲はそれほど広くないと考えられる。Veit ら[13]や Yang ら[15]は、動き情報と画像のセグメンテーション情報を組合せ、シーン中の平面領域の推定やカメラ動きの推定を行ったが、事前に精度良く分割された移動領域や3次元復元された情報が必要であった。ロボットナビゲーションのために平面領域を推定する手法として、オプティカルフロー情報を用いた繰り返し処理を用いた手法[10]がある。しかし、これまでに人の動作認識のためのカメラ動き補償を、移動カメラ一台のみで実現する手法は無かった。

本論文では、人の動作認識への応用を目的としたカメラ動き補償法を提案する。提案法は、複数特徴点を追跡することにより得られるオプティカルフロー情報を利用して、背景の動きを推定する。特徴点の検出には、FAST 特徴点検出器[11]を用いる。特徴点の追跡には、KLT 法[6]及び SIFT[5]を用いたロバストな特徴点

追跡手法を提案し、これを用いる。画像は、色情報を元に領域分割を行い、追跡した特徴点情報を組み合わせることにより、背景領域を複数の平面領域に分割し、各平面領域に対して背景領域の動きを射影変換モデルとして推定を行うことにより、より正確な背景領域の動き推定を実現する。

2 節では、特徴点の検出および追跡手法についてその詳細を述べる。3 節では、画像の平面領域への分割及び背景の動き推定について述べる。4 節では、複数環境で撮影した映像に対して提案法を適用し、その効果を考察する。

## 2. 特徴点の追跡

提案法では、1. FAST による特徴点の検出、2. KLT 法による各特徴点におけるオプティカルフローの推定、3. SIFT による追跡失敗点の検出と削除を行うことにより、ロバストな特徴点の追跡を実現する。

### 2.1. 特徴点の検出

一般的に用いられる特徴点検出器として、ハリス・ラプラシアン [7]がある。特徴点検出器を用いることにより、画像中からコーナーなど特徴の強い点・領域を検出することが可能である。Rosten&Drummond により提案 [11]された FAST は、非常に多数の特徴点を比較的短時間のうちに検出することを実現した特徴点検出器である。

提案法では、FAST を用い追跡対象となる特徴点を検出する。FAST は、閾値  $Th$  を変化させることにより、検出する特徴点数を制御することが可能である。ただし、後に述べる KLT 法は、追跡する点数が増えるに従い計算コストも増えるため、あまりに多くの特徴点を追跡するのは処理時間の観点から好ましくない。また、各特徴点間の距離が近すぎること、処理の上では無駄が大きい。

そこで、特徴点の検出において特徴点の個数  $n < N$ 、特徴点間の距離  $d > D$  となるように  $Th$  の調整を行う。まず、初期閾値を  $Th_0$  とし特徴点の検出を行う。次に、検出された特徴点をラスタスキャンし、選択された点の半径  $D$  内に存在する点を削除する。最後に、残った点の数  $n$  を計測し  $n < N$  となるまで処理を繰り返すことにより、希望する特徴点間の距離  $D$  及び特徴点数  $N$  を自動的に得ることが可能となる。

### 2.2. 特徴点の追跡

特徴点の追跡には、KLT 法を用いる。大きく動いた特徴点の追跡を効果的に行うため、提案法ではピラミッド画像上で KLT 法による特徴点の追跡を行う。

まず、もっとも解像度の低い画像  $I_1$  を上でオプティカルフローを計算し、 $I_1$  における対応点  $p_1$  を求める。一つ解像度の上がる次の画像  $I_{i-1}$  では、 $p_i$  を初期位置

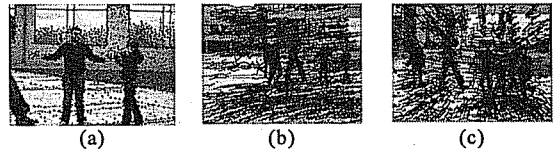


図1 特徴点の検出と追跡。(a)検出結果、(b)追跡結果(パン動作)、(c)追跡結果(ズーム動作)。

としてオプティカルフローの計算を行う。以上の処理を最も解像度の高い画像  $I_0$  まで再帰的に行うことにより、大きく動く特徴点にも対応することが可能となる。

### 2.3. 追跡失敗点の検出と削除

KLT 法による追跡では、明るさの変化や動きによるぶれ、テクスチャの変化などにより間違った対応点を示すオプティカルフローを返す場合がある。追跡失敗点を検出するために、SIFT による特徴点間の対応確認を行う。ただし、連続するフレーム間ではスケール変化は無視できるほど小さいことが期待できるため、SIFT で行うキーポイントのサンプリングは行わず、FAST により検出した点をキーポイントとして扱う。

前フレームの特徴点  $p_{i-1}$  及びオプティカルフローより求めた現フレームの特徴点  $p_i$  をそれぞれキーポイントとして、画像の勾配ヒストグラム  $q_{i-1}$ ,  $q_i$  を計算する。勾配ヒストグラム間の距離  $Q$  を計算することにより 2 点間の対応を求めることが可能である。勾配ヒストグラム間の距離として Bhattacharyya 距離を用いる。Bhattacharyya 距離を用いることにより、 $Q$  のレンジが 0.0~1.0 に正規化されるため、対応の正否を決める閾値  $T_Q$  を容易に決定できる。ここで、 $Q$  は 0.0 に近いほど一致しているものとする。

SIFT によるマッチングの結果、 $Q > T_Q$  となった特徴点は追跡失敗点として削除される。また、追跡中に画面外に出てしまった点も削除される。その後、削除された特徴点と同じ数だけの点を新たに検出し、新規特徴点として追跡を行うことにより、常に希望する追跡点数  $N$  を保持することが可能となる。図 1(a)に特徴点を検出した画像を、(b)、(c)にカメラをパン及びズームした場合の特徴点追跡例を示す。

## 3. 背景の動き推定

フレーム間の対応点情報を用い、背景の動きによる画像間の見え方の変化を射影変換によりモデル化する。見え方の変化を表現した射影変換モデルは、同時に背景の動きを表現していると考えられることができる。しかし、一つの射影変換モデルでは一つの平面の幾何変換のみしか表現できないため、複数の平面が組み合わされたシーンでは、背景の動きを完全に表現することが難しい。例えば、カメラを平行移動させながら撮影した場合、カメラに近い物体はより早く移動するように

見え、遠い部分はゆっくり移動するように見える。また、地面から垂直に立った壁などは、地面とは違った向き・速さで移動するように見える。

この問題を解決するため、シーンを平面要素ごとに領域分割し、各平面要素に対して射影変換モデルを推定する。平面要素への領域分割を単眼カメラからの映像で実現するため、色情報を用いた領域分割を行う。領域分割後、各領域を平面と仮定しそれぞれの領域の動きを推定し、動き情報を基にした領域統合を行うことにより、シーンを最大3つの平面領域に分割する。これはシーン中に存在する平面として、地面と二つの壁を想定したものである。

### 3.1. 画像の領域分割

画面の領域分割には、RGB色空間上における Mean Shift を用いた領域分割法[1]を用いる。画像領域上で大きさ $(R_w, R_h)$ の矩形窓を設け、同時に RGB 色空間上に半径 $R_c$ の円形窓を設け、Mean Shift により画像を局所領域に分割する。分割された領域は、画像空間上で位置が近いかつ RGB 色空間上で近い色を持つピクセルが統合されたものとなる。 $R_w=11, R_h=11, R_c=40$ として図1に示した画像を領域分割した結果を図2に示す。

### 3.2. 平面の動き推定

領域分割により、画像は $M$ 個の領域 $S_1, S_2, \dots, S_M$  ( $m=1, 2, \dots, M$ )に分割される。ここで、特徴点 $\mathbf{p}_n$ がどの領域に含まれるかを表す行列 $\mathbf{S}_p$ を以下に定義する。

$$\mathbf{S}_p = \begin{matrix} & \mathbf{p}_1 & \dots & \mathbf{p}_N \\ \mathbf{S}_1 & \begin{bmatrix} 0 & 1 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 1 \end{bmatrix} \\ \vdots & \\ \mathbf{S}_M & \end{matrix} \quad (1)$$

ここで、 $\mathbf{S}_p(S_m, \mathbf{p}_n)=1$ であれば、特徴点 $\mathbf{p}_n$ は領域 $S_m$ に含まれることを示す。複数領域の境界付近に存在する特徴点は、複数領域に含まれることがある。

次に、領域 $S_m$ の動きを表す射影変換モデル $\mathbf{H}_m$ を、ロバスト推定の一つである段階的マッチング[17]を用いて推定する。もし、 $\mathbf{H}_m$ を推定するのに十分な数の特徴点が領域に存在しない場合、 $\mathbf{H}_m$ は単位行列とする。

次に、 $\mathbf{H}_m$ を用いて領域 $S_m$ を $M \leq 3$ となるまで領域の統合を行う。そのために、特徴点 $\mathbf{p}_n$ の動きがどの射影変換 $\mathbf{H}_m$ に含まれるかを表す行列 $\mathbf{H}_p$ を以下に定義する。

$$\mathbf{H}_p = \begin{matrix} & \mathbf{p}_1 & \dots & \mathbf{p}_N \\ \mathbf{H}_1 & \begin{bmatrix} 0 & 1 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \\ \vdots & \\ \mathbf{H}_M & \end{matrix} \quad (2)$$

ここで、前フレームの特徴点 $\mathbf{p}_n(t-1)$ を $\mathbf{H}_m$ により投影した点 $\mathbf{p}_n(\mathbf{H}_m|t-1)$ と、 $\mathbf{p}_n(t-1)$ を追跡した結果得られた点 $\mathbf{p}_n(t)$ との距離を $d(\mathbf{p}_n, \mathbf{H}_m)$ とし、 $d(\mathbf{p}_n, \mathbf{H}_m) < T_d$ である場

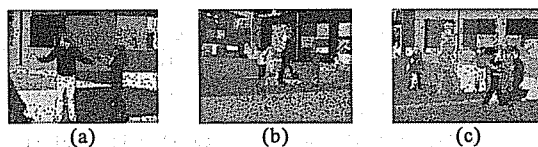


図2 Mean Shift による領域分割

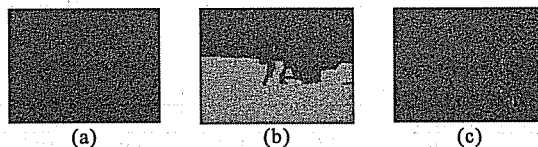


図3 推定された平面領域

合、点 $\mathbf{p}_n$ の動きは $\mathbf{H}_m$ で表現可能であると考えることができ、点 $\mathbf{p}_n$ を $\mathbf{H}_m$ のインライア $\mathbf{H}_p(\mathbf{H}_m, \mathbf{p}_n)=1$ とし、それ以外を0(アウトライア)とする。

次に、以下の計算により行列 $\mathbf{H}_s$ を求める。

$$\mathbf{H}_s = \mathbf{S}_p \cdot \mathbf{H}_p^T = \begin{matrix} & \mathbf{H}_1 & \dots & \mathbf{H}_M & \sum_S \\ \mathbf{S}_1 & \begin{bmatrix} 4 & 34 & \dots & \dots \\ 2 & 21 & \dots & \dots \\ 3 & 56 & \dots & \dots \\ 24 & 1 & \dots & \dots \end{bmatrix} & \dots & \dots & 42 \\ \vdots & \\ \mathbf{S}_M & \dots & \dots & \dots & \vdots \\ \sum_H & 33 & 112 & \dots & \dots \end{matrix} \quad (3)$$

$\mathbf{H}_s$ は、領域 $S_m$ 中の点がどれだけ $\mathbf{H}_m$ に含まれるかを表した行列である。ここで、 $\mathbf{H}_s$ の各行の合計 $\sum_H$ は画像全体における $\mathbf{H}_m$ のインライア数を表す。また、 $\sum_S$ は領域 $S_m$ に含まれる特徴点の合計を表す。インライア数が最大となる $\mathbf{H}_m$ が最も大きな背景の動きを表していると考えられるため、そこに含まれる領域のうち、 $\mathbf{H}_s(S_i, \mathbf{H}_j) / \sum_S \geq 0.8$  ( $i, j=1, 2, \dots, M$ )となるすべての領域を統合し平面領域とする。

以上の処理を、領域数 $M < 3$ となるまで繰り返し行う。推定された平面領域より、再び射影変換モデルのロバスト推定を行った結果を、各平面領域の最終的な動きモデル $\mathbf{H}_m$  ( $m=1, \dots, M < 3$ )とする。図2に示した画像の平面領域を推定した結果を図3に示す。

### 3.3. 動き補償

推定された平面領域の動きモデル $\mathbf{H}_m$ を推定し、各特徴点の動き補償を行う。以下の計算により、動き補償を行い、各特徴点のローカルモーション $\mathbf{v}_n$ を得る。

$$\mathbf{v}_n = \mathbf{p}_n(t) - \mathbf{p}_n(\mathbf{H}_m | t-1) \quad (4)$$

ただし、 $\mathbf{p}_n$ は平面領域 $S_m$ に含まれており、 $\mathbf{p}_n(\mathbf{H}_m | t-1)$ は $\mathbf{p}_n(t-1)$ を $\mathbf{H}_m$ により投影した点である。

## 4. 実験

カメラ動き補償の効果を確認するため、屋内で撮影した2つの映像及び屋外で撮影した3つの映像に対して提案法を適用した。実験で用いた各パラメータの値を表1に示す。

表1 パラメータ設定

徴点数:	N = 1000	[個]
特徴点間距離:	D = 10	[pixel]
勾配ヒストグラム閾値:	TQ = 0.3	
領域分割用矩形窓サイズ:	Rw, Rh = 11	[pixel]
領域分割用色空間窓サイズ:	Rc = 40	
インライア判別閾値:	Td = 1.0	[pixel]

表2 実験結果

(a) カメラ動き補償なし

映像	屋外1	屋外2	屋外3	屋内1	屋内2
前景	6.3	7.5	9.2	0.0	10.7
背景	3.2	9.6	12.9	3.7	14.1

(b) カメラ動き補償あり

映像	屋外1	屋外2	屋外3	屋内1	屋内2
前景	5.6	4.9	5.9	0.0	5.8
背景	0.7	1.1	2.2	0.7	4.2

カメラ動き補償の効果を評価するために、あらかじめ背景領域と前景領域(移動人物)の領域を指定したマップ画像を作成しておき、各領域内の特徴点のオプティカルフローの強度の平均を求め、カメラ動き補償前と補償後における変化を計測した。理想的な結果として、カメラ動き補償後は背景領域のオプティカルフロー強度は減少し、前景領域のオプティカルフロー強度は背景領域ほど減少しないことが望ましい。

#### 4.1. 実験結果および考察

表2(a)にカメラ動き補償なしの場合、(b)にカメラ動き補償ありの場合での実験結果を示す。なお、屋内1の映像は背景のみであったため、前景の値が0となっている。

結果より、すべての環境においてカメラ補償ありの場合に背景領域のオプティカルフロー強度が減少した。その一方で、前景領域のオプティカルフロー強度は背景ほど減少しておらず、移動人物のローカルモーション情報が得られていることがわかる。カメラ補償後の前景・背景それぞれのオプティカルフロー強度の減衰比の平均は、前景において0.3、背景において0.8となった。屋内2の環境において、背景領域のオプティカルフロー強度は、カメラ動き補償後も4.2と高い値を示している。これは、撮影した環境がテクスチャのない一様な壁のあるものであったため、他の環境に比べて特徴点の追跡が上手くいかずに、大きな強度を持ったオプティカルフローがノイズとして多く発生したためであると考えられる。

#### 5. まとめ

複数特徴点の追跡によるカメラ動き補償法を提案した。提案法では、FASTを用いて検出した特徴点をKLT法による追跡を行い、SIFTの勾配ヒストグラムによるマッチングにより追跡失敗点を削除することによ

り、多数の特徴点をロバストに追跡した。また、色情報を用いた画像領域の領域分割を行い、各領域と追跡点の情報を合わせて用いることにより、動き情報に基づいた背景平面領域の推定を実現した。推定した背景平面領域の動き情報を複数の射影変換モデルで推定し、動き補償を行うことにより移動人物のローカルモーションを得ることに成功した。5つの異なる環境において提案法を適用し、提案法の効果を確認した。

#### 文献

- [1] D. Comaniciu and P. Meer, Mean Shift: A Robust Approach toward Feature Space Analysis, PAMI Vol.24 No.5, pp.603-619, May 2002.
- [2] P. Dollar, V. Rabaud, G. Cottrell and S. Belongie, Behavior Recognition via Sparse Spatio-Temporal Features, VS-PETS, pp.65-72, Oct 2005.
- [3] C. Fanti, L. Zelnik-Manor and P. Perona, Hybrid Models for Human Motion Recognition, CVPR'05, pp.1166-1173, June 2005.
- [4] I. Laptev and P. Perez, Retrieving Actions in Movies, ICCV'07, Oct 2007.
- [5] D. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, IJCV Vol.60 No.2, pp.91-110, Nov 2004.
- [6] B. D. Lucas and T. Kanade, An Iterative Image Registration Technique with An Application to Stereo Vision, IJCAI, pp.674-679, 1981.
- [7] K. Mikolajczyk and C. Schmid, A Performance Evaluation of Local Descriptors, PAMI Vol.27 No.10, pp.1615-1630, Oct 2005.
- [8] J. C. Niebles, H. Wang and L. Fei-Fei, Unsupervised Learning of Human Action Categories using Spatial-Temporal Words, BMVC'06, pp.1249-1258, 2006.
- [9] S. Nowozin, G. Bakir and K. Tsuda, Discriminative Subsequence Mining for Action Classification, ICCV'07, pp.1919-1923, Oct 2007.
- [10] N. Ohnishi and A. Imiya, Model-Based Plane-Segmentation Using Optical Flow and Dominant Plane, LNCS Vol.4418, 2007.
- [11] E. Rosten and T. Drummond, Machine Learning for High-Speed Corner Detection, ECCV'06, pp.430-443, May 2006.
- [12] C. Schuldt, I. Laptev and B. Caputo, Recognizing Human Action: A Local SVM Approach, ICPR'04, pp.32-36, Aug 2004.
- [13] T. Veit, F. Cao and P. Bouthemy, An a contrario Decision Framework for Region-Based Motion Detection, IJCV Vol.68 No.2, pp.163-178, June 2006.
- [14] S. Wong and R. Cipolla, Extracting Spatiotemporal Interest Points using Global Information, ICCV'07, Oct 2007.
- [15] A. Yang, S. Rao, A. Wagner and Y. Ma, Segmentation of a Piece-Wise Planar Scene from Perspective Images, CVPR'05, pp.154-161, June 2005.
- [16] A. Yilmaz and M. Shah, Recognizing Human Actions in Videos Acquired by Uncalibrated moving Cameras, ICCV'05, pp.150-157, Oct 2005.
- [17] 金澤 靖, 金谷 健一, 段階的マッチングによる画像モザイク生成, 信学論 D-II Vol.J86-D-II No.6, pp.816-824, 2003.