

マルチカメラを用いた3次元シーンフロー推定に関する研究

画像認識講座

水野 宏基

指導教授

藤吉 弘亘

1. はじめに

近年、ユビキタス社会におけるユザビリティの向上として、人の自然動作に近いインターフェイスであるジェスチャ認識の実現が期待されている。本研究では、ジェスチャ認識に応用が可能である技術として、3次元シーンフローを用いたモーション推定法を提案する。

2. 提案手法の流れ

対象とする3次元シーンフローは、複数の画像中のオプティカルフローから求められる[1]。一般に2次元オプティカルフローは多くのノイズやアウトライアが含まれる。そのため、復元される3次元シーンフローにも多くのノイズが含まれ、アウトライアが発生する。そこで、本研究では物体の運動が短時間では剛体であると仮定し、Subspace拘束を用いた3次元シーンフローの修正を行う。次に、復元した3次元シーンフローから物体のモーションパラメータを推定する。図1に本手法の流れを示す。

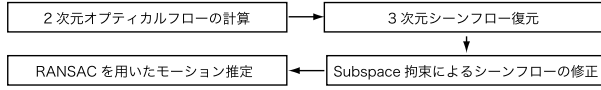


図1：提案手法の流れ

3. 3次元シーンフローの復元

運動モデルと観測行列 シーン内の運動が剛体であると仮定した場合、ある時刻 j における対象物体中の点 $\mathbf{x}_i$ のシーンフロー $\mathbf{v}_{ij}$ は、次式のように、平行移動ベクトル $\mathbf{t}_j$ と、回転速度ベクトル $\boldsymbol{\omega}_j$ を用いて表現することができる。

$$\begin{aligned} \mathbf{v}_{ij} &= \mathbf{t}_j + \boldsymbol{\omega}_j \times \mathbf{x}_i \\ \mathbf{v}_{ij} &= [v_{x_{ij}}, v_{y_{ij}}, v_{z_{ij}}]^T \quad \mathbf{x}_i = [X_{x_i}, X_{y_i}, X_{z_i}]^T \\ \mathbf{t}_j &= [t_{x_j}, t_{y_j}, t_{z_j}]^T \quad : \text{平行移動ベクトル} \\ \boldsymbol{\omega}_j &= [\omega_{x_j}, \omega_{y_j}, \omega_{z_j}]^T \quad : \text{回転速度ベクトル} \\ i &: \text{座標番号} \quad j: \text{フレーム番号} \end{aligned} \quad (1)$$

式(1)を展開すると次式となる。

$$\begin{bmatrix} v_{x_{ij}} \\ v_{y_{ij}} \\ v_{z_{ij}} \end{bmatrix} = \begin{bmatrix} s_{x_j} \\ s_{y_j} \\ s_{z_j} \end{bmatrix} \mathbf{q}_{i(6 \times 1)} \quad (2)$$

$$\begin{aligned} s_{x_j} &= [t_{x_j}, 0, 0, 0, -\omega_{z_j}, \omega_{y_j}] \\ s_{y_j} &= [0, t_{y_j}, 0, \omega_{z_j}, 0, -\omega_{x_j}] \\ s_{z_j} &= [0, 0, t_{z_j}, -\omega_{y_j}, \omega_{x_j}, 0] \\ \mathbf{q}_i &= [1, 1, 1, X_{x_i}, X_{y_i}, X_{z_i}]^T \end{aligned}$$

ここで、式(2)は単位時間の空間中の一点のみの運動を表現した式である。 N 点 M フレームの情報を扱うために式(2)を以下のように拡張する。

$$\begin{bmatrix} \mathbf{V}_x \\ \mathbf{V}_y \\ \mathbf{V}_z \end{bmatrix}_{(3M \times N)} = \begin{bmatrix} \mathbf{S}_x \\ \mathbf{S}_y \\ \mathbf{S}_z \end{bmatrix}_{(3M \times 6)} \mathbf{Q}_{(6 \times N)} \quad (3)$$

$$\mathbf{Q}_{(6 \times N)} = [\mathbf{q}_1, \dots, \mathbf{q}_N] = \begin{bmatrix} 1 & \dots & 1 \\ 1 & \dots & 1 \\ 1 & \dots & 1 \\ X_{x_1} & \dots & X_{x_N} \\ X_{y_1} & \dots & X_{y_N} \\ X_{z_1} & \dots & X_{z_N} \end{bmatrix}$$

$$\mathbf{S}_x = \begin{bmatrix} s_{x_1} \\ \vdots \\ s_{x_M} \end{bmatrix}, \quad \mathbf{S}_y = \begin{bmatrix} s_{y_1} \\ \vdots \\ s_{y_M} \end{bmatrix}, \quad \mathbf{S}_z = \begin{bmatrix} s_{z_1} \\ \vdots \\ s_{z_M} \end{bmatrix}$$

ここで、 $[\mathbf{V}_x/\mathbf{V}_y/\mathbf{V}_z]$ を観測行列と呼ぶ。瞬間的な物体の運動は等速であると仮定した場合、 $\mathbf{S}_x$, $\mathbf{S}_y$, $\mathbf{S}_z$ 各行列

のランクは全て1以下となるため、行列 $[\mathbf{S}_x/\mathbf{S}_y/\mathbf{S}_z]$ のランクは3以下となる。観測行列 $[\mathbf{V}_x/\mathbf{V}_y/\mathbf{V}_z]$ は行列 $\mathbf{Q}$ と行列 $[\mathbf{S}_x/\mathbf{S}_y/\mathbf{S}_z]$ の積で表現されており、小さい方のランクに制限される。よって、観測行列のランクは3以下となる。

Subspace拘束を用いたシーンフローの修正 2次元オプティカルフローにノイズやアウトライアが含まれる場合、シーンフローから得られる観測行列のランクは3以上となる。そこで復元されたシーンフローに対し、Subspace拘束によるランク制限を施すことで精度を向上させる。

Subspace拘束を用いてランク制限を施すために、複数フレームのシーンフローの成分を v_x , v_y , v_z それぞれ観測行列に当てはめる。次に、観測行列に対し、特異値分解を行い次式を得る。

$$\begin{bmatrix} \mathbf{V}_x \\ \mathbf{V}_y \\ \mathbf{V}_z \end{bmatrix}_{(3M \times N)} = \mathbf{U}_1 \mathbf{D} \mathbf{U}_2^T \quad (4)$$

$\mathbf{U}_1$ は $3M \times 3M$ の直交行列、 $\mathbf{U}_2$ は $N \times N$ の直交行列、 $\mathbf{D}$ は特別な対角行列であり、 $\mathbf{D}$ の対角要素は特異値と呼ばれ、ベクトル $\mathbf{d} = [d_1, d_2, \dots, d_{3M}]^T$ で表す。ここで、観測行列のランクを3に制限するには対角成分が $\mathbf{d}' = [d_1, d_2, d_3, 0, \dots, 0]^T$ となる行列 $\mathbf{D}'$ を次式のように求める。

$$\mathbf{D}' = \begin{bmatrix} d_1 & 0 & \dots & \dots & \dots & 0 \\ 0 & d_2 & 0 & \dots & \dots & 0 \\ 0 & 0 & d_3 & 0 & \dots & 0 \\ 0 & \dots & \dots & 0 & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 \end{bmatrix}_{(3M \times N)}$$

この $\mathbf{D}'$ から、 $\mathbf{U}_1$ と $\mathbf{U}_2$ を用いて新しい観測行列を以下のように計算する。

$$\begin{bmatrix} \mathbf{V}_x' \\ \mathbf{V}_y' \\ \mathbf{V}_z' \end{bmatrix}_{(3M \times N)}' = \mathbf{U}_1 \mathbf{D}' \mathbf{U}_2^T \quad (5)$$

得られた新しい観測行列 $[\mathbf{V}_x'/\mathbf{V}_y'/\mathbf{V}_z']'$ は、ランクが3の行列となり、式(1)に示す運動モデルに従うシーンフローに修正することができる。

4. シーンフローを用いたモーション推定

シーン内のある点 $\mathbf{x}_i$ における運動 $\mathbf{v}_{ij}$ が得られた場合、時刻 j における式(1)は次式のように変形できる。

$$\mathbf{v}_{ij} = \begin{bmatrix} -[\mathbf{x}_i]_{\times} & \mathbf{I}_{(3 \times 3)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\omega}_j \\ \mathbf{t}_j \end{bmatrix}$$

$[\mathbf{x}_i]_{\times}$: $\mathbf{x}_i$ の歪対称行列、 $\mathbf{I}_{(3 \times 3)}$: 3×3 の単位行列
これを N 点に拡張すると次式を得る。

$$\begin{bmatrix} \mathbf{v}_{1j} \\ \vdots \\ \mathbf{v}_{Nj} \end{bmatrix} = \begin{bmatrix} -[\mathbf{x}_1]_{\times} & \mathbf{I}_{1(3 \times 3)} \\ \vdots & \vdots \\ -[\mathbf{x}_N]_{\times} & \mathbf{I}_{N(3 \times 3)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\omega}_j \\ \mathbf{t}_j \end{bmatrix} \quad (6)$$

ここで、未知数は $\boldsymbol{\omega}_j = [\omega_{x_j}, \omega_{y_j}, \omega_{z_j}]^T$ と $\mathbf{t}_j = [t_{x_j}, t_{y_j}, t_{z_j}]^T$ の6つとなる。式(6)からは、 $3 \times N$ の式を得ることができるが、歪対称行列 $[\mathbf{x}_i]_{\times}$ のランクは2であるため、 $N=2$ の場合、式(6)は一意に解を決定することができない。よって、3点以上の $\mathbf{v}_{ij}$ と $\mathbf{x}_i$ の組から最小二乗法を用いることで、剛体運動の各パラメータを推定する。

2節で述べた Subspace 拘束により修正した3次元シーンフローは、運動モデルのランクに従った運動となる。このとき、Subspace 拘束ではシーンフローの大きさと方向を修正することは可能であるが、本来存在しないところに得られたシーンフローを除去することはできない。そこで、このようなアウトライアを除去してパラメータを推定するために、本研究では RANSAC を用いる。

5. 実験

提案手法の評価実験として、シミュレーションと実画像を用い、3次元シーンフロー復元とモーション推定を行う。また、本実験では、基準フレームから前後1フレームのシーンフローから観測行列を作成し、シーンフローの復元と修正を行う ($M=2$)。

5.1. シミュレーションによるシーンフロー復元実験

シミュレーションでは、立方体、球、任意の形状の3種類の物体に対し、平行移動、回転運動、平行移動と回転が同時に発生した場合の計9シーケンスのシーンフローを対象とする。シーンフローの大きさと方向に対し、正規乱数を用いてノイズを与え、シーンフローの存在位置に対し、10点に1点の確率でアウトライアを生成する。シミュレーションにより生成したシーンフローと、Subspace拘束を用いた修正シーンフローの例を図2に示す。図2(a)では

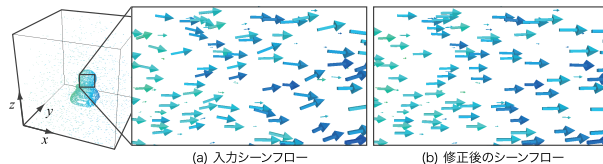


図2: シミュレーション実験結果の例

分散しているシーンフローが、Subspace拘束により、同一の方向へ修正されていることが分かる。シーンフローの精度の評価として、シーンフローの真値 $\mathbf{v}_t$ と復元されたシーンフロー $\mathbf{v}$ との方向の類似度を式(7)より、また大きさに対する評価を式(8)より計算する。

$$\text{方向に対する類似度} : \cos \theta = \frac{\mathbf{v} \cdot \mathbf{v}_t}{\|\mathbf{v}\| \|\mathbf{v}_t\|} \quad (7)$$

$$\text{大きさの差のノルム} : \frac{\|\mathbf{v}\| - \|\mathbf{v}_t\|}{\|\mathbf{v}_t\|} \quad (8)$$

図3に、9シーケンスに対し、角度、大きさに対する真値との類似度を平均した結果を示す。図3(a)はシーンフロー

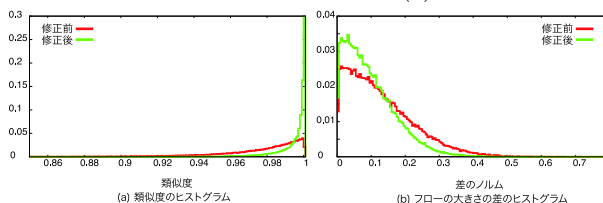


図3: 類似度のヒストグラム

の方向の類似度のヒストグラムであり、修正前は、誤差10度以内のシーンフローが全体の56.7%であるのに対し、修正後は94.2%まで向上した。図3(b)は大きさに対するヒストグラムであり、修正前は、誤差10%以内のシーンフローが全体の46.3%であるのに対し、修正後は、58.3%まで向上した。

5.2. シミュレーションによるモーション推定実験

シミュレーションで復元された3次元シーンフローからモーション推定の実験を行う。修正前のシーンフローの全点から推定した結果と比較する。それぞれの運動に対し、推定されたモーションパラメータの誤差の平均を表1に示す。Subspace拘束により修正したシーンフローに対し、

表1: モーション推定誤差

	修正前 全点	修正前 RANSAC	修正後 RANSAC
平行移動	1.21E-04	1.16E-03	4.79E-04
回転運動	1.86E-02	8.34E-04	3.70E-04
平行移動 + 回転運動	1.05E-02	2.20E-03	1.21E-03

RANSACを用いた場合に最も良い精度でパラメータを推定できていることが分かる。これは修正を行うことで、シーンフローに含まれるノイズの影響を低減することができ、RANSACによりアウトライアを削除することができたと考えられる。

5.3. 実画像を用いた実験

実画像を用いた実験を行う。実験環境は、図4(a)に示すように、3次元空間中に配置した5台のカメラを用いる。対象とする物体は、図4(b)に示すように、ルービックキューブを持った人間の腕部とする。オプティカルフローの計算にはブロックマッチング法を用い、Vedulaらの手法[1]を用いてシーンフローを復元する。本実験では、 y 軸方向に1秒間に約15cm平行移動した場合と、 x 軸を中心に約180度回転した場合の2つのシーケンスを対象とする。図4(c)のオプティカルフローを入力とした際のシーンフロー復元例を図5に示す。図5(a)では分散しているシーンフロー

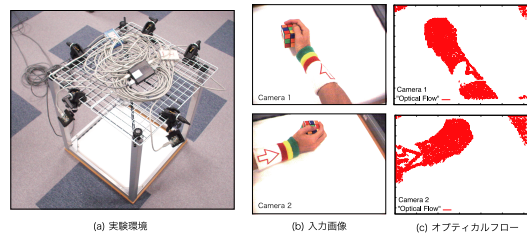


図4: 実験環境と入力データ

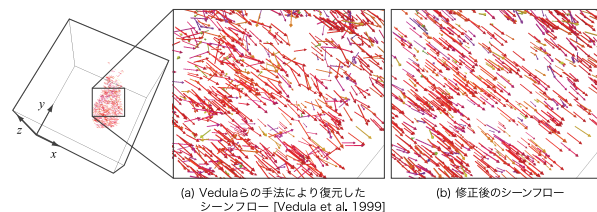


図5: 平行移動における3次元シーンフローの例

が、Subspace拘束により、同一の方向へ修正されていることが分かる。表2に、平行移動と回転運動のモーション推定結果を示す。平行移動では、最も数値の高いパラメー

表2: モーション推定結果

	ω_x [deg/f]	ω_y [deg/f]	ω_z [deg/f]	t_x [mm/f]	t_y [mm/f]	t_z [mm/f]
平行移動	-0.494	0.380	-0.244	0.818	-4.523	-0.078
回転運動	-5.490	-0.114	0.978	-0.246	5.901	2.197

f: frame

タは y 軸方向の移動であり、1frame(1/30sec)で約4.5mm移動と推定された。これは、1秒間で13.5cmの平行移動となり、高い精度で移動量を推定ができていることが分かる。次に回転運動では、 x 軸周りの回転が1frameで約5.5度と推定された。実験に用いたシーケンスの真値は6度であり、高い精度でモーションが推定できていることが分かる。しかし、平行移動していないにもかかわらず、 t_y 成分が高い値を示している。これは、上方からこの運動を観測するため、その表面の運動が t_y 成分の運動として観測されることが原因である。この問題はカメラの台数を増やすことで解消できると考えられる。

6. おわりに

本研究では、Subspace拘束により修正した3次元シーンフローから物体のモーションを推定する手法を提案し、その有効性を確認した。今後は、モーションセグメンテーションを用いた複数物体への対応を検討する予定である。

参考文献

- [1] S. Vedula, S. Baker, P. Rander, R. Collins, T. Kanade, "Three-Dimensional Scene Flow", Proceedings of the 7th International Conference on Computer Vision, Vol. 2, September, pp. 722 - 729, 1999.

研究業績

- [1] 水野宏基, 藤吉弘巨, 岩堀祐之. "マルチカメラを用いた3次元シーンフローによるモーション推定", 画像の認識・理解シンポジウム (MIRU2006), p. 529 - 534, 2006.
(他 研究会技術報告1件, 学会口頭発表2件)