

画像の標本化過程のモデルに基づく 高精度・頑健な平面追跡

伊藤 栄介 岡谷 貴之 出口 光一郎^{†1}

カメラに対して相対的に運動する平面を、画像上で精度良く安定して追跡することは、コンピュータビジョンの中心的課題である。これまで、照明変化、オクルージョンやモーションブラーなど、追跡の精度や安定性を損なう要因が色々指摘され、それらへの対策が提案されてきている。本稿では、追跡対象の平面が視線に対して大きく傾いたり、あるいは遠く離れた場合に生じる画像上での実効的な解像度の低下もまた、追跡に悪影響を与えることを指摘する。そしてこの解像度の低下を、平面の姿勢に依存して決まる異方性ガウス関数の、追跡対象とするテンプレートへの畳み込みによってモデル化する。この解像度低下のモデルを Malis らの提案した ESM 法に組み込むと、従来方法よりもより高精度かつロバストな平面追跡が実行できることを、実験によって示す。

Accurate and Robust 2D Visual Tracking based on a Model of Image Sampling Process

EISUKE ITO, TAKAYUKI OKATANI
and KOICHIRO DEGUCHI^{†1}

It is one of the central issues in computer vision to track a planar object moving in space relatively to the camera in an accurate and stable fashion. Many studies have been conducted that point out several factors deteriorating tracking accuracy and stability and propose solutions to them. In this paper, we first point out that various forms of degradation in nominal image resolution, which typically occur in a situation where the planar object has a significantly oblique pose with respect to the viewing direction or in a situation where it moves to a distant location from the camera, can impair tracking performance. We then propose to model such degradation in image resolution by a convolution of the template with an anisotropic Gaussian function whose shape is determined by the current pose of the target plane. Incorporating this imaging model to ESM developed by Malis et al., we demonstrate through several experiments that it can perform more accurate and robust tracking of a planar object than the conventional method.

1. はじめに

本稿では、単一カメラに対して相対的に運動する平面あるいはその一部を、画像上で追跡する問題を考える。

このような平面追跡の問題は、コンピュータビジョンの基礎的アルゴリズムの一つである。例えばそのための一つの方法である LK 法は、動画像を使った 3 次元復元の目的で、シーンにある複数の特徴点まわりの微小平面を、画像上で追跡するのに良く利用される。また、拡張現実感 (AR) では、シーンに静止した平面を手持ちのカメラ等で撮影し、仮想的な物体がそのシーンに存在するように見える画像を合成するために、平面の追跡を行ってカメラの運動を計算することが行われる。

平面追跡の方法は 3 つに大別される。一つは、画像の濃淡を直接使う方法で、テンプレートと画像間の各画素での密な濃淡の差分を最小化するように、反復法によって幾何学的な運動パラメータを推定する方法である。代表的なものに LK 法⁶⁾ や Malis らの ESM 法¹⁾ がある。もう一つは、特徴点に代表される画像の局所特徴を抽出し、テンプレートと画像間におけるこれら特徴の疎な対応を求めて、そこから幾何学的な運動パラメータを直接計算する方法である。この方法は最もロバストであり、オクルージョンやフレーム間の移動量の大小を基本的に気にする必要がないのが利点だが、一般に推定精度を向上させるのに多数の対応が必要となり、そのせいもあって計算量が大きくなる傾向がある。この方法は一つ目の方法の初期値を求めるのに使われることもある。最後の方法は、学習に基づく方法で、平面が姿勢を変化させたときのテンプレートのあらゆる見えをあらかじめ学習しておき、これに基づいて運動パラメータを推定する。運動パラメータの数が少ない場合はよいが、そうでない場合には次元の呪いが様々な問題を生む。近年、再近傍探索を高速化する pull-back と呼ぶ操作によってこの問題を緩和する方法が Tian ら⁹⁾ によって提案されているが、まだまだ問題が多い。

AR 等のリアルタイム性が要求されるアプリケーションでは、特に精度が高いことと計算量の小ささが求められる。この点で、上の 3 つの方法のうち 1 番目の方法が、ロバスト性において 2 番目の方法に劣るものの、最も優れている。中でも Malis らの提案した ESM (Efficient

^{†1} 東北大学
Tohoku University

Second-order Minimization) 法¹⁾ は、収束範囲の広さによる追跡性能のロバスト性と計算量の小ささをうまく両立した方法として、その性能が認められている。

ESM 法はその後、照明変動にともなう見えの変化に対処するための拡張²⁾ や、モーションブラーを扱えるようにする拡張⁸⁾ が施され、追跡性能の一層の向上が図られてきた。

本稿では、これと同じ動機に基づき、ESM 法 (あるいは一般の濃淡ベースの平面追跡法) をよりロバストにする方法を述べる。われわれは、追跡対象とする平面が視線方向に対して大きく傾いたり、あるいは遠ざかるとき、その画像上の実効的な解像度が低下する問題に注目する。濃淡ベースの方法では、追跡対象のテンプレートと、最新の画像の濃淡を推定すべき平面の運動パラメータで幾何学的に変形させたものと比較し、その差を最小にするように運動パラメータを求める。実効的な解像度の低下が起こると、仮に正しい運動パラメータを与えても、最新の画像の濃淡はテンプレートのそれと決して一致しなくなる。実際、これが原因で追跡性能が低下することは実験的にも確認できる。

われわれは、画像取得時の標準化過程の考察により解像度の低下をモデル化し、このモデルに基づいて平面追跡をよりロバストにする方法を提案する。例えば AR への応用において、追跡対象の平面が視線に対し大きく傾く場合でも、従来方法よりもロバストに追跡を継続でき、合成物体の画像合成を安定して行えることを示す。

2. 平面追跡法に関する既存研究

ESM は、Malis ら¹⁾ により提案された平面追跡の方法である。ここでは、ESM の基本的なアルゴリズムを要約し、照明変化²⁾、およびモーションブラー⁸⁾ にそれぞれ対応できるようにした拡張についてまとめる。

2.1 ESM の原理

追跡対象となるテンプレート画像を I^* 、現在の画像を I と書く。 I^* は、図 1 のように、追跡対象とする平面 (の矩形領域) を、真正面から適当な距離でカメラで撮影して得た画像に等しい。カメラと平面の相対位置姿勢が変化し、それにもなつて I が変化するとき、 I 上で I^* を追跡したい。

平面を異なる視点から撮影した 2 枚の画像間の座標変換は平面射影変換で与えられる。 I^* の点 p^* を I の点 p に変換する射影変換を、 3×3 行列 H を使って

$$p = w(p^*; H) \propto Hp^* \quad (1)$$

と書く。以降、 p および p^* は、文脈に応じて一つの記号で同次座標と非同次座標の 2 つを使い分ける。

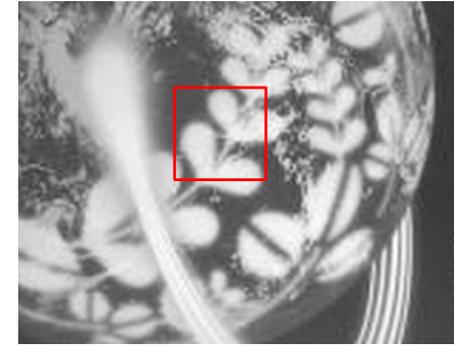


図 1 追跡対象となる平面とテンプレート画像 (赤枠) の例。

平面の同じ点の画像上の明るさが変化しないとすれば、画像変化は平面射影変換のみで記述され、テンプレート I^* と最新の画像 I の関係は次のように表現できる。

$$I^*(p_i^*) = I(w(p_i^*; H)), \quad i = 1, 2, \dots, q. \quad (2)$$

ただし i はテンプレート画像内の画素のインデックスで、総数を q としている。

推定したいのは毎フレームごとの H であり、 H の自由度は 8 ある。ESM では不定な H のスケール倍の自由度を $\det H = 1$ とすることで拘束し、8 成分のベクトル x で $H = H(x)$ を

$$H(x) = \exp \left(\sum_{j=1}^8 x_j G_j \right) \quad (3)$$

と表す。 G_j はリー代数の基底である。

追跡は、初期値 \hat{H} が与えられている条件の下で、増分 $H(x)$ を推定することで行う。具体的には、次のようにテンプレート画像の画素数 q と同じ成分数を持つベクトル $y(x) = [y_1(x), y_2(x), \dots, y_q(x)]^T$ を、その成分を

$$y_i(x) = I(w(p_i^*; \hat{H}H(x))) - I^*(p_i^*) \quad (4)$$

のように定義し、この y の大きさを最小化する

$$x_0 = \underset{x}{\operatorname{argmin}} \|y(x)\|^2 \quad (5)$$

を求める。その後 $\hat{H} \leftarrow \hat{H}H(x_0)$ とし、これを毎フレームごと収束するまで何度か反復し、

目的の射影変換を推定する． Baker らの分類⁶⁾によれば，以上の方法は Forward Compositional 法に相当する．

上の最小化は，ニュートン法に基づく反復計算にて行う．ESM の特徴は，ヘッセ行列を陽に計算することなく，2次微分の項を表現することにある．ガウスニュートン法では，ヘッセ行列をヤコビ行列の積で近似するが，これよりも近似精度が高い．具体的には式(5)の解 \mathbf{x}_0 は，次の式を解くことに帰着される．

$$\mathbf{J}_{esm} \mathbf{x}_0 = -\mathbf{y}(0), \quad (6)$$

ただし， \mathbf{J}_{esm} の i 列目の成分は以下の式で表される．

$$\mathbf{J}_{esm} = \frac{1}{2}(\mathbf{J}_{I^*} + \mathbf{J}_I)\mathbf{J}_w\mathbf{J}_H \quad (7)$$

ここで， \mathbf{J}_{I^*} と \mathbf{J}_I はそれぞれ， $I^*(\mathbf{p}^*)$ と $I(w(\mathbf{p}^*; \hat{\mathbf{H}}\mathbf{H}(\mathbf{x})))$ の \mathbf{p}^* に関する（空間）微分， \mathbf{J}_w はテンプレート画像の座標情報， \mathbf{J}_H はリー代数の基底を並び変えた定数行列である．これらのうち \mathbf{J}_{I^*} と \mathbf{J}_w ， \mathbf{J}_H は値が一定であり，事前の計算により得られる．追跡時には \mathbf{J}_I と $\mathbf{y}(0)$ のみ毎回計算することになる．

ESM は，平面射影変換をリー群に基づいて表現し，かつ最小化すべきコストを高精度に近似することで，計算の高速性と収束範囲を両立した優れた方法である．文献^{1),6)}にあるように，Baker らの Inverse Compositional 法よりも計算量の面では若干劣るが，収束範囲において勝るとされる．

2.2 照明変動の影響への配慮

上の基本的なアルゴリズムは，輝度の差の二乗和を最小にしていた．そのため，平面が運動し照明との位置関係が変化すること等で生じる輝度変化の影響を受けやすかった．Malis らは文献²⁾で，この影響に対処する方法を述べている．

追跡対象の平面上での光の反射は，拡散反射，鏡面反射，環境光の変化，そしてカメラのバイアスによって決定され，次式ようになる．

$$I'(\mathbf{p}_i; \alpha_i, \eta_i, \beta) = \alpha_i I_d(\mathbf{p}_i) + \eta_i I_s(\mathbf{p}_i) + \beta \quad (8)$$

追跡対象平面表面の拡散反射や鏡面反射のパラメータは未知である為，拡散反射と鏡面反射の項を次のように一つにまとめて推定を行う．

$$I' = \gamma I + \beta \quad (9)$$

さらに，平面パターンは一様な光の影響を受けているとは限らない為，追跡領域を格子状に

数分割し，それぞれの領域に異なる γ を与えて推定を行う．

画像を n 分割した場合，ESM で推定すべきパラメータ数は，射影変換の自由度分の 8，画像分割分の γ の n ，そして環境光とカメラバイアスの影響を考慮した β の 1，合計して $n+9$ である．このパラメータを $\theta = [\tilde{\mathbf{x}}^T, \tilde{\gamma}^T, \tilde{\beta}]$ とする．分割した j 番目のブロックにおける i 番目のピクセルの輝度を

$$I'(w(\mathbf{p}_{ij}^*; \mathbf{H}); \gamma_j, \beta) = \gamma_j I(w(\mathbf{p}_{ij}^*; \mathbf{H})) + \beta \quad (10)$$

と定義すると，最小化すべき関数は以下のように書き換えることができる．

$$f(\theta) = \frac{1}{2} \|\mathcal{I}'_{ij} - \mathcal{I}_{ij}^*\|^2 = \frac{1}{2} \|\mathbf{d}(\theta)\|^2 \quad (11)$$

この式の最小化は非線形問題であり，オリジナルの ESM 同様に線形化し，反復によって θ を計算する．

2.3 モーションブラーへの対処

カメラのシャッタースピードが長い，あるいは平面が高速に運動する場合，画像上でモーションブラーが発生する．ESM に限らず，テンプレートと最新画像の輝度差を最小にすることで運動パラメータを推定する方法では，このようなモーションブラーは追跡性能を一般に低下させる⁸⁾．

Mei ら⁷⁾は，モーションブラーをテンプレートに反映させることで，追跡性能の低下を防ぐ方法を示した．具体的には，式(4)の輝度差 y_i を次のように修正する．

$$y_i(\mathbf{x}, \mathbf{u}) = I(w(\mathbf{p}_i^*; \hat{\mathbf{H}}\mathbf{H}(\mathbf{x}))) - \mathcal{I}_b^*(\mathbf{p}_i^*; \mathbf{u}). \quad (12)$$

ここで $\mathcal{I}_b^*(\mathbf{p}_i^*; \mathbf{u})$ はテンプレート \mathcal{I}_b^* にモーションブラーを施したもので

$$\mathcal{I}_b^*(\mathbf{p}_i^*; \mathbf{u}) \equiv \int_0^1 \mathcal{I}^*(e^{-t\hat{\mathcal{H}}} e^{-t\mathcal{H}(\mathbf{u})} \mathbf{p}_i^*) dt \quad (13)$$

である． $\mathcal{H}(\mathbf{x}) = \sum_j^8 x_j \mathbf{G}_j$ であり， $\hat{\mathcal{H}}$ は $\hat{\mathbf{H}}$ に対応する同様のリー代数の基底和である．これらの指数関数は \mathbf{p}_i^* の射影変換を与える．これらを $-t$ 倍し，上の式のように t について線積分を計算することで，射影変換の自由度を持つモーションブラー（ブラーが直線状の最も単純な場合だけでなく，回転や拡大，さらに一般の射影変換の自由度を持つブラー）を表現することができる．

論文⁷⁾では， \mathbf{u} を平面の運動 \mathbf{x} とは無関係に求める場合と，ブラー自体の運動が平面の運動に一致する ($\mathbf{u} = \mathbf{x}$ となる) 場合の両方を考えている．前者は求めるパラメータの自由

度が大きくなる分、不安定さが想像されるが、実験結果もその通りとなっている。基本的にはブラーと運動そのものが一致する後者の考えるモデルが実用的だろう。

なお Park らは⁸⁾、フレーム間の平面の動きがパラメータに関して線形で、かつカメラのシャッター時間が既知であるという仮定の下で、オリジナルの ESM のわずかな拡張でモーションブラーを扱えるようにできることを示した。テンプレート画像の変形をパラメータ θ とし、シャッターが t_0 から t_1 の時間に開いているとし、ブレ画像 $I(\theta)$ を次式で表す。

$$I(\theta) = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} w\left(T, \frac{t}{t_1}\theta\right) dt \quad (14)$$

$w(T, \cdot)$ はテンプレート画像を変形させる関数である。

オリジナルの ESM の式 (7) に対応する、このモデルに対する計算式は次のようになる。

$$\mathbf{J}_{esm} = \frac{1}{2} (a\mathbf{J}_{I^*} + \mathbf{J}_I) \mathbf{J}_w \mathbf{J}_H \quad (15)$$

a は露光時間によって決定される定数である。露光時間が非常に短く $t_1 = t_0$ と近似できる場合は $a = 1$ となる。逆に露光時間がフレーム間隔と等しい最大の場合は $t_0 = 0$ で、 $a = 0.5$ となる。

3. 解像度低下への対処

3.1 解像度低下による追跡性能の劣化

上述したような既存方法では、画像の標本化の過程が考慮されることはなかった。対象となる平面がカメラに対して極端に傾く場合や、あるいはカメラから遠方に離れる場合、実効的な画像の解像度が低下し、これによって追跡性能が低下することを、われわれは確認している。

具体的には次のようなことである。テンプレート画像と最新の画像間の差分を測るコストは、 $\sum_i |I^*(\mathbf{p}_i^*) - I(w(\mathbf{p}_i^*; \hat{\mathbf{H}}\mathbf{H}(\mathbf{x})))|^2$ のように、テンプレート画像の各画素について計算される。固定されたテンプレート画像と比較されるのは、推定する射影変換によって最新の画像を変形したものである。図 2(b) に、平面がカメラに対して大きく傾いた場合の $I(w(\mathbf{p}_i; \mathbf{H}(\mathbf{x})))$ を、同図 (c) に、平面がカメラから大きく遠ざかった場合の典型例をそれぞれ示す。図のように、平面がカメラに対して傾くと、追跡領域はカメラの画像上で細長いつぶれた形となる。カメラの解像度は不変であるから、このつぶれた形をテンプレートの矩形に戻すと、輝度分布は図のようにぼやけたようなものとなる。平面がカメラから遠ざかる場合も同じである。



図 2 カメラとパターンの相対的な位置変化によって得られる追跡領域の画像。(a) は平面をほぼ正面から撮った画像、(b) は平面が大きく傾いたときに撮った画像、(c) は平面が遠ざかったときに撮られた画像である。

このことが追跡性能の劣化につながるの直感的にも理解できる。典型的には、追跡が途中で不可能になる場合が最も多い。また、誤った射影変換を推定してしまうなど精度の低下も良く起こる。具体例は、後の実験結果で示す。

3.2 解像度低下のモデル

追跡性能の向上のため、このような解像度の低下をモデル化し、これを追跡アルゴリズムに組み込むことを考える。

式 (12) のモーションブラーの場合と同様にテンプレート画像を修正することとし、具体的には最小化すべきベクトル \mathbf{y} の各成分を次のように変更する。

$$y_i(\mathbf{x}) = I(w(\mathbf{p}_i^*; \hat{\mathbf{H}}\mathbf{H}(\mathbf{x}))) - I_r^*(\mathbf{p}_i^*; \hat{\mathbf{H}}\mathbf{H}(\mathbf{x})). \quad (16)$$

$I_r^*(\mathbf{p}_i^*; \hat{\mathbf{H}}\mathbf{H}(\mathbf{x}))$ は、標本化過程をシミュレートするフィルタをテンプレートに畳込んだもので次のように定式する。

$$I_r^*(\mathbf{p}_i^*; \hat{\mathbf{H}}\mathbf{H}(\mathbf{x})) \equiv I^*(\mathbf{p}_i^*) * f(\mathbf{p}_i^*; \hat{\mathbf{H}}\mathbf{H}(\mathbf{x})). \quad (17)$$

このフィルタは次のように決める。フィルタの画像座標を $[u, v]$ ($\mathbf{p}^* \propto [u, v, 1]^T$) としたとき、 $f(u, v) = f(\mathbf{p}^*; \mathbf{H})$ を、次のように定義する。まず $\mathbf{H} = \hat{\mathbf{H}}\mathbf{H}(\mathbf{x})$ を近似するアフィン変換を

$$\mathbf{H} \sim \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \quad (18)$$

とする．近似は， \mathbf{H} および上のアフィン変換にて画像座標を変換した後の座標のずれの 2 乗距離が最小となるように決める．この \mathbf{A} を用いて，フィルタ $f(u, v)$ を次のようなガウス関数とする．

$$f(u, v; \mathbf{A}, \alpha) = \frac{1}{\sqrt{2\pi|\mathbf{A}^{-1}\mathbf{A}^{\top-1}|}} \exp\left(-\frac{\alpha}{2} \begin{bmatrix} u & v \end{bmatrix} (\mathbf{A}^{\top}\mathbf{A}) \begin{bmatrix} u \\ v \end{bmatrix}\right) \quad (19)$$

α は使用したカメラのイメージセンサやレンズの性能によって決まる定数である．以下の実験では $\alpha = 2.0$ とした．

なお上では， $f(u, v) = f(\mathbf{p}^*; \mathbf{H})$ の $\mathbf{H} = \hat{\mathbf{H}}\mathbf{H}(\mathbf{x})$ とした．これはモーションブラーのモデル式 (12) におけるテンプレートの修正のやり方を踏襲している．ただしモーションブラーの場合と違い，今の標準化のモデルの場合， $\mathbf{H}(\mathbf{x})$ を落として $\mathbf{H} \sim \hat{\mathbf{H}}$ のように近似しても結果にはほとんど影響がない．モーションブラーの場合は，フレーム間変動 \mathbf{x} がブラーそのものと直接に関係するため， $\mathbf{H}(\mathbf{x})$ の項の存在が本質的に重要である．標準化のモデルの場合，フレーム間変動よりも現在までに蓄積された射影変換，すなわち現在の平面の絶対的な姿勢が支配的である．この近似によって，最適化計算は非常に簡単になるものとなる．

4. 実験結果

提案手法の性能を評価する実験を行った．平面の運動を表す射影変換の精度と追跡の安定性（ロバスト性）について評価した．精度の評価には，追跡対象の平面と正確に 90 度の角度をなす別の平面上にチェスボードパターンを貼り付け，これを用いた．具体的には，チェスボードパターンから平面の運動パラメータ（射影変換）を求め，これを真値の目安として用いて，提案手法によって推定した運動パラメータと比較した．従来手法とは，オリジナルの ESM に Malis らの照明変化モデルおよび Park らのモーションブラーのモデルを加えたもので提案手法とは，これにさらに提案した解像度低下のモデルを加えたものである．

実験を行ったシステムは次の通りである．カメラは Point Grey Research 社の Grasshopper を用いた．これは 640×480 ピクセルの画像を毎秒 60 フレーム撮影する．PC は，3.2GHz の Core i7 を CPU に持ち，nVidia 社の GTX480 を GPU に持つ．ESM 法の主要部分は GPU 上で実行され，リアルタイムでの追跡が十分可能である．GPU 上での ESM 法の実

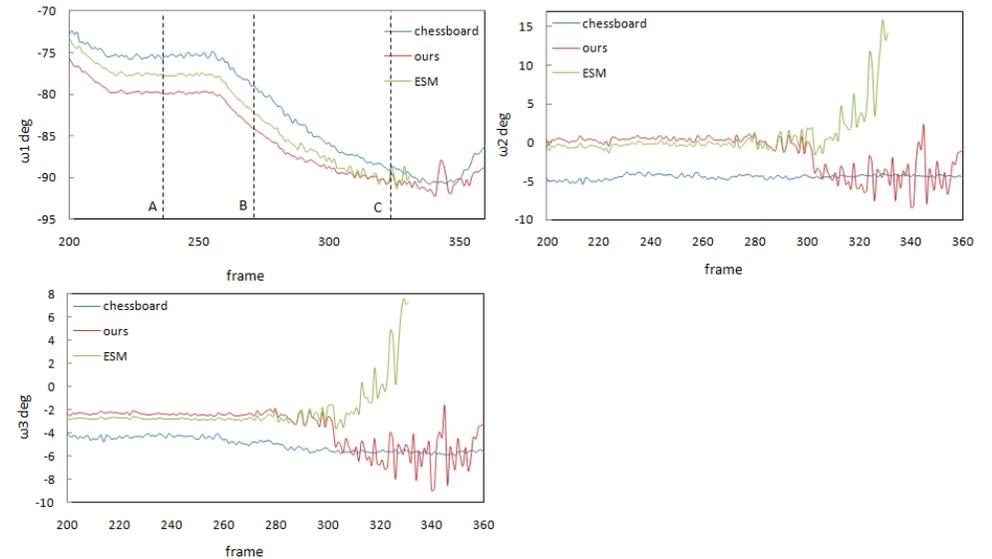


図 3 回転行列をベクトル表記したとき，各成分をフレーム毎に記述したもの．

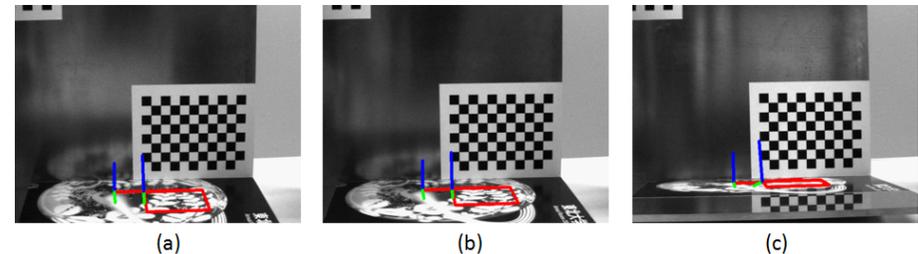


図 4 解像度低下をモデル化していない ESM . (a) ~ (c) は A ~ C と同じフレームでの結果に対応．

装については文献¹⁰⁾を参照されたい．

図 3 に，約 160 フレームの動画シーケンスを対象とした実験結果の一例を示す．ここでは，図 5 に示すテンプレートを追跡している．図 3 は，平面の空間における運動パラメータのうち回転成分を，推定された射影変換から取り出し，その 3 成分をプロットしたもので

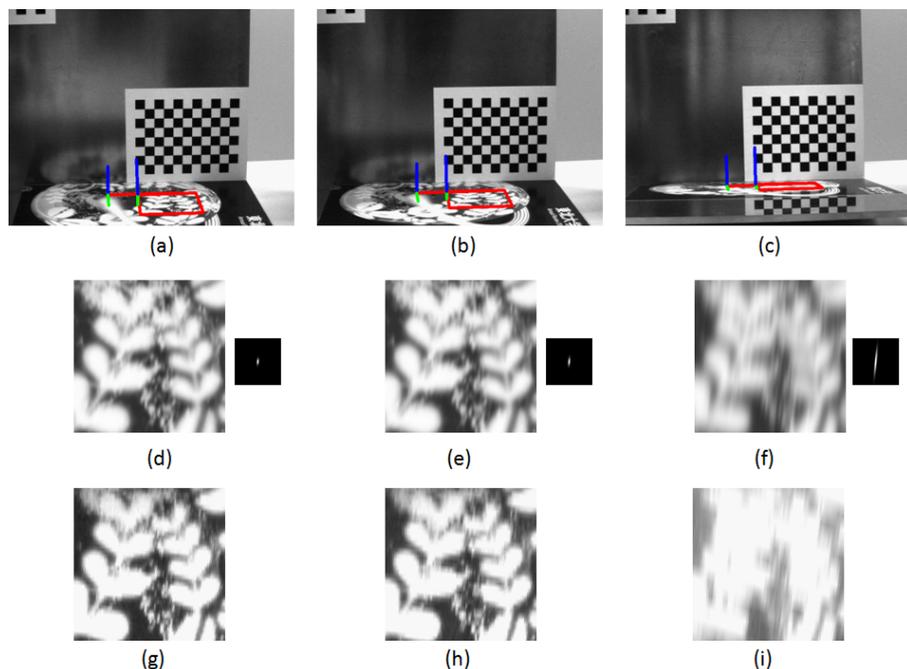


図 5 推定領域とテンプレート画像の変化．(a)～(c) は A～C の画像に対応．左にある軸はチェスボードパターンから推定された回転軸，右にある軸は傾いた画像から推定された回転軸を表している．(d)～(f) はフィルタによって計算されたテンプレート画像，(g)～(i) は推定された領域である．

ある．射影変換からこの回転成分を取り出す方法は次のようにした．まず，カメラの内部パラメータを事前のキャリブレーションで求めておき，これを K とする．推定した射影変換を H とするとき， $H \propto K[r_1, r_2, t]$ のように表せる．ただし r_1 および r_2 は回転成分に対応する回転行列の 1, 2 列ベクトルで， t は並進成分のベクトルである．これに基づき，まず $K^{-1}H$ を計算し，結果の 3×3 行列の 1, 2 列ベクトル \tilde{r}_1 および \tilde{r}_2 を抽出する．これらを元に新たに行列 $[\tilde{r}_1, \tilde{r}_2, \tilde{r}_1 \times \tilde{r}_2]$ を作成し，これを特異値分解し，その特異値をすべて強制的に 1 に置き換えた行列を求めて，これを目的の回転行列の推定値とした．こうして得られた回転行列を回転軸・角度表現の 3 成分ベクトルに変換し，その成分を各フレームごとにプロットしたのが図 3 である．

動画シーケンスは，平面を視線方向に対するその角度が徐々に 90 度に近づくように動

かしており，後のフレームになるほど追跡が難しくなっている．図 3 から，従来手法（解像度低下を考慮しないもの）は，約 310 フレーム付近で追跡精度が低下し始め，330 フレーム付近で完全に追跡に失敗しているのがわかる．追跡失敗時には，ESM 法の反復計算が収束しなくなった．一方提案手法は，同じ画像に対しても安定して追跡を継続できている．シーケンスの最後の部分で推定値は大きく振動し必ずしもその精度は高くはないが，視線になす平面の角度が 90 度にかなり近い困難な領域での追跡であり，無理はない．反対に，視線と平面の傾きがそれほどでもない領域では，提案手法よりも従来手法の方がチェスボードパターンから求めた推定量に近く，従来手法の方がよさそうである．ただしチェスボードパターンによる推定量がより正確であるという保証はないので，どちらが優れているかははっきりとは言えない．

図 4 に，特定のフレームでの従来手法の推定結果を可視化したものを示す．同図 (a)-(c) は図 3 中の A, B, C のフレーム位置での結果である．図 3 を見ると，A, B のフレームでは良好に追跡されているが，C のフレームで大きな誤差が発生している．このことは図 4 から確かめられる（平面上の座標フレームを緑，赤，青色の線分で表現しているが，(c) では緑の軸が明らかに正しくなくなっている）．一方，同じフレームに対する提案手法の結果を図 5 に示す．(a)-(c) は同様に A～C 各フレームでの推定結果を可視化したもので，特に (c) を図 4(c) と見比べると，提案手法がより正確であることが分かる．また図 5 には，解像度低下のモデルに基づいて修正したテンプレートを中段に，画像を推定パラメータで変換したものを下段にそれぞれ示している．両者が各フレームで近くなっていることから，提案した解像度低下のモデルが適切であったことが確かめられる．なお，(c) では撮影画像が照明変動の影響を受け，テンプレートよりも全体的に明るくなっているが，照明変動のモデル化を行っていることからうまく対処できている．

別のテクスチャを持つ平面を同様に追跡した実験結果を図 6 に示す．ほぼ同様に，提案手法の方が従来手法よりも安定した追跡を実現できた．図 7，図 8 は，従来方法および提案手法による特定のフレームに対する追跡結果を可視化したものである．

さらに我々は，平面を手でより自由に動かして約 1000 フレームからなる動画を撮影し，従来手法と提案手法を適用した．提案手法は全フレームで安定して追跡を行えたが，従来手法は図 9 に示す 6 フレームで追跡に失敗した．それぞれの画像には，右上から順にテンプレート画像，追跡領域，フィルタも表示している．

平面を使った AR を行う場合，追跡対象の平面を視線に対し大きく傾けたり遠くに配置する状況も想定され，提案手法はそのような場合に特に役立つ．一例として図 10 を示す．

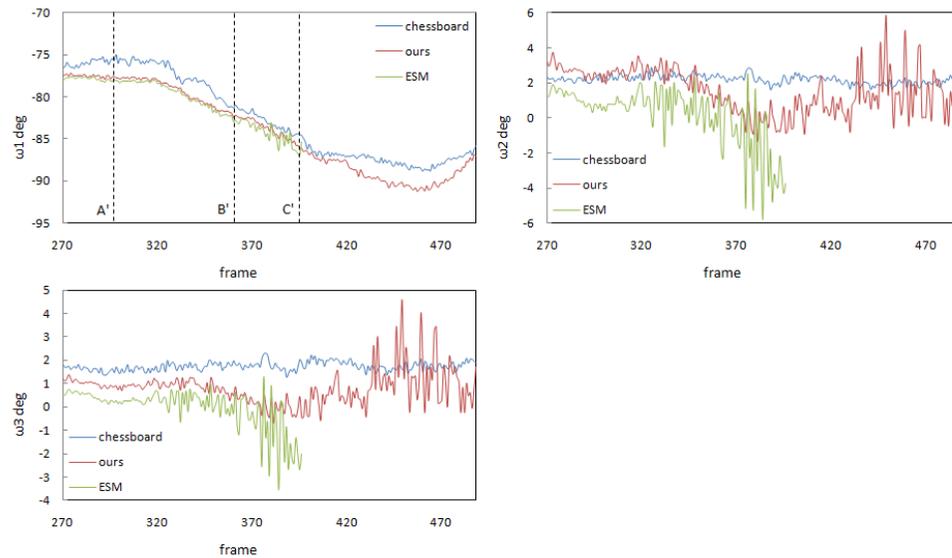


図 6 回転行列をベクトル表記したとき、各成分をフレーム毎に記述したもの。

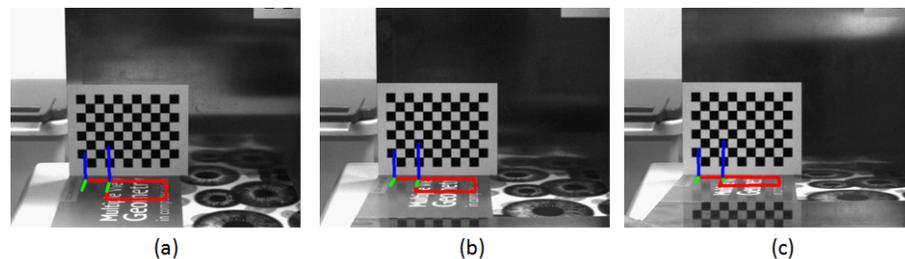


図 7 解像度低下をモデル化していない ESM. (a)~(c) は A'~C' と同じフレームでの結果に対応。

従来の手法で精度良く追跡できないような画像においても、提案手法は比較的精度良く追跡できる。したがって、提案する解像度低下のモデルを組み込んだ ESM を使うことで AR の活用の場面を一層拡大できる。

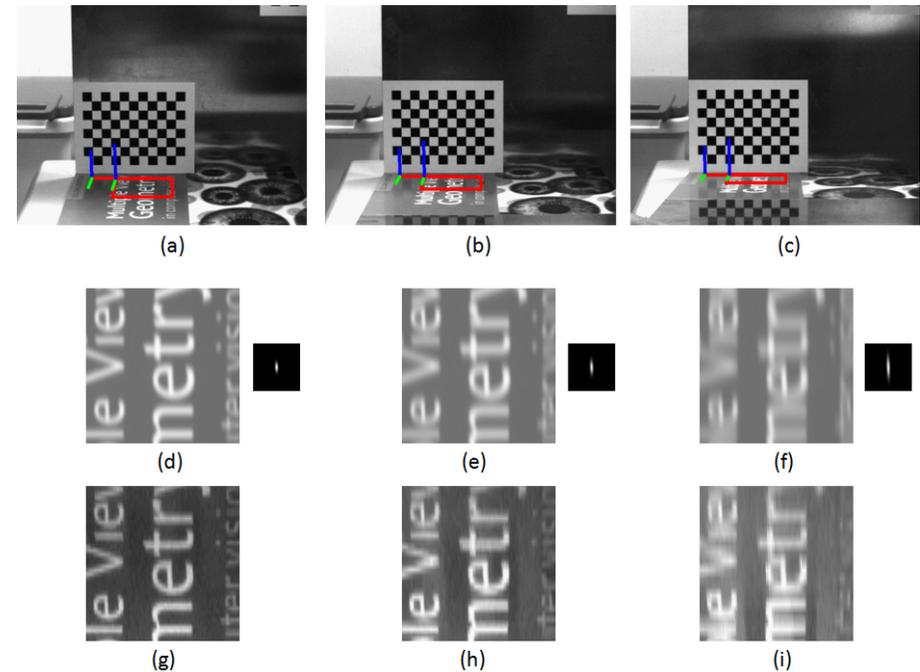


図 8 推定領域とテンプレート画像の変化。(a)~(c) は A'~C' の画像に対応。左にある軸はチェスボードパターンから推定された回転軸、右にある軸は傾いた画像から推定された回転軸を表している。(d)~(f) はフィルタによって計算されたテンプレート画像、(g)~(i) は推定された領域である。

5. おわりに

本研究では、平面追跡において解像度低下が問題となることを指摘し、解像度低下をモデル化することを提案した。平面が大きく傾いていたり遠くに見えていたりする場合、テンプレート画像と追跡領域の画像が大きく異なり ESM の計算が収束しなくなる。従来の手法では画像の標準化の過程をまったく考慮していないことが原因で、追跡が失敗していた。そこで、我々はテンプレート画像にフィルタを畳込むことで解像度低下をモデル化した。これによって、従来の手法が困難な領域でも安定かつリアルタイムで平面を追跡できることを確認できた。今後の課題として、ESM 法と自然に融合する方法を考え、高速化を図ることが必要である。



図 9 1000 フレーム中、従来手法で追跡に失敗した 6 フレーム。提案手法はこの全てで成功した。

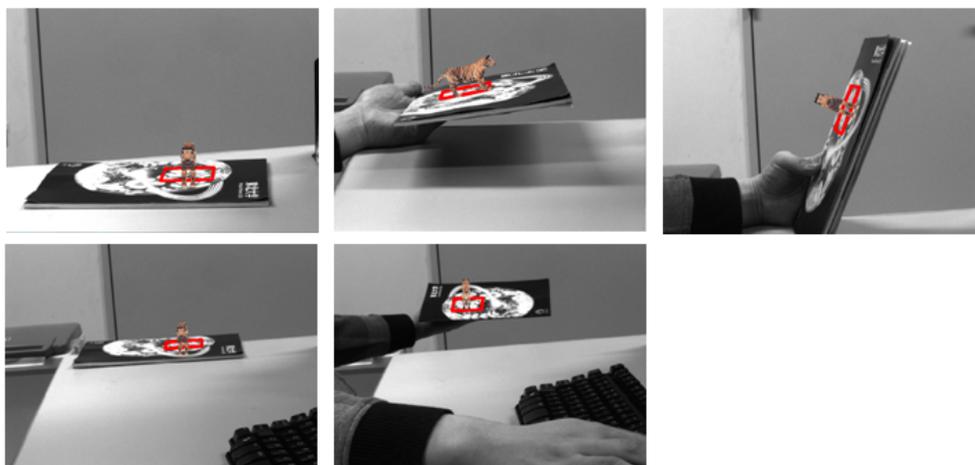


図 10 虎のモデルを重畳した画像。平面が傾いたり遠方に見えている場合でも安定した画像の合成が可能となった。

参考文献

- 1) S.Benhimane,E.Malis, “Homography-based 2D Visual Tracking and Servoing”,The International Journal of Robotics Research Vol.26,No.7,pp.661-676,2007.
- 2) G. Silveira, E. Malis, “Real-time Visual Tracking under Arbitrary Illumination Changes”, In Proc.CVPR,pp.1-6,2007.
- 3) E. Malis,“An efficient unified approach to direct visual tracking of rigid and deformable surfaces”, In Proc. IEEE/RSJ International Conference on Intelligent Robots Systems, pp. 2729 - 2734,2007.
- 4) E. Malis, “Improving vision-based control using efficient second-order minimization techniques”, In Proc. IEEE International Conference on Robotics and Automation, Volume 2, Pages 1843-1848,2004.
- 5) L.Yuan,J.Sun,L.Quan,and H.-Y.Shum, “Image Deblurring with Blurred/Noisy Image Pairs”,ACM Transactions on Graphics,Vol.26, No.3,Article 1,2007.
- 6) S.Baker and I.Matthews, “Lucas-Kanade 20 years on: A unifying framework”, International Journal of Computer Vision,56(3):221-225,2004.
- 7) C.Mei and I.Reid,“Modeling and generating complex motion blur for real-time tracking”,In Proc. CVPR,2008.
- 8) Y.Park,V.Lepetit,W.Woo,“ESM-Blur: Handling and Rendering Blur in 3D Tracking and Augmentation”,In Proc. International Symposium on Mixed and Augmented Reality,2009.
- 9) Y.Tian, S.Narasimhan,“A Globally Optimal Data-Driven Approach for Image Distortion Estimation ”,In Proc. CVPR,2010.
- 10) E.Ito,T.Okatani,K.Deguchi,“GPU を使った高速高精度な平面トラッキング”,CVIM,172-30,2010.