

共同注視状況における複数人物頭部カメラの位置姿勢推定

高瀬恵三郎[†] 近藤 一晃^{††} 小泉 敬寛[†] 中村 裕一^{††}

[†] 京都大学大学院工学研究科 〒 606-8318 京都府京都市左京区吉田本町

^{††} 京都大学 学術情報メディアセンター 〒 606-8318 京都府京都市左京区吉田本町

E-mail: [†]{takase,koizumi}@ccm.media.kyoto-u.ac.jp, ^{††}{kondo,yuichi}@media.kyoto-u.ac.jp

あらまし 我々は、体験型の学習形態における共同注視の閲覧・分析を支援するために、個人視点映像を入力とした SLAM により参加者の視線を推定し可視化することを目指している。本稿では、入力映像が SLAM 初期化アルゴリズムの前提条件を満たしていることを評価する方法、および、複数視線の推定結果を統合する方法について報告する。環境マーカーの導入とともに、初期化処理の途中結果などを用いた種々の評価値の設計や統合のための共通座標系の導入について検討を行った。検証実験の結果、良い評価値が得られた場合には前提条件を満たしている傾向が強いことが定性的に確認できた。

キーワード 共同注視, 個人視点映像, SLAM

SLAM of multiple head mounted cameras in conjugate gazing situations

Keizaburo TAKASE[†], Kazuaki KONDO^{††}, Takahiro KOIZUMI[†], and Yuichi NAKAMURA^{††}

[†] Graduate School of Engineering, Kyoto University Yoshida Honmachi, Sakyo-ku, Kyoto, 606-8318 Japan

^{††} Academic Center for Computing and Media Studies, Kyoto University Yoshida Honmachi, Sakyo-ku, Kyoto, 606-8318 Japan

E-mail: [†]{takase,koizumi}@ccm.media.kyoto-u.ac.jp, ^{††}{kondo,yuichi}@media.kyoto-u.ac.jp

Abstract The novel way of browsing and analyzing the enormous logs of the experimental study, which aims to learn the knowledge and experience in the fields, is expected for evaluation. The conjugate gaze is thought as one of the important reference. Therefore, we intended to visualize it by estimating the participants FPVs by using the SLAM. In this paper, we described the method that to evaluate whether the input video is appropriate for the SLAM initialization, and the method that to consolidate these results to visualize conjugate gaze. In the experiments, we confirmed that the evaluation results indicated appropriate input from the FPVs for the SLAM initialization, and the conjugate gaze was detected from the estimated gaze.

Key words Conjugate Gaze, First Person View, SLAM

1. はじめに

近年、体験学習やワークショップといった体験を通じた学習形態が広く用いられるようになってきている。このような学習形態では、参加者が主体的にものごとに触れることができるため、座学などの教授的な知識獲得とは異なる「体験や経験に基づいた学習効果」が期待される。しかし、その学習効果は筆記テストを行うと言った方法では簡単に評価することが難しく、体験学習の最中に起きた出来事をつぶさに分析・評価することが必要である。そうした分析の対象となる重要な出来事の一つとして、共同注視 [1] [2] が挙げられる。共同注視とは、複数人物の注視の関係を示しており、認知科学、発達心理学などの分野でよく用いられている。狭義では、互に見つめあった二人のうち片方が視線を他の対象に移したとき、他方もそれに倣って同

じ対象に視線を移す現象を示す。体験学習ではより多様なインタラクションが発生するため、同一物体に複数の視線が集中したり、その状態から視線が発散したりといった現象も共同注視とみなされる。例として、ガイド付きの博物館鑑賞ツアーを考えてみよう。ガイドに対する共同注視からは、説明時に参加者の注目を集められていたか、話を聞いていなさそうな参加者は誰か、などが分かる。また、共同注視から展示物への興味の傾向や、展示物のどの部分に興味を持って注視していたのかを分析することによって、学習効果の評価や、展示や説明方法の再検討にも役立つと考えられる。

共同注視の従来分析・評価手法は、対象の場面を撮影記録した動画を分析者が何度も見返すというものであった。このような方法は長大な動画を逐一閲覧しなければならず、分析者への負担が非常に大きく、非効率である。そこで本稿では、共

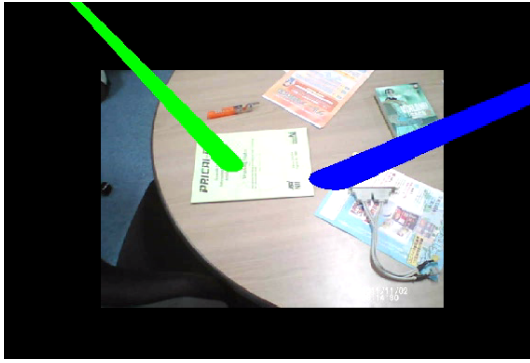


図 1 複数視線の可視化．ある参加者の個人視点映像が背景に書き込まれ，2つの円錐が他の参加者の視線の位置関係を表している．視線を一つのビューに可視化することができれば，共同注視の有無やその特性などを簡単に把握することができる．

Fig. 1 Visualizing relationships of multiple view lines for browsing conjugate gazing situations.

同注視の閲覧・分析を支援し，時間的，労力的な負担を軽減するために，複数人物の視線情報を単一の動画上に可視化する方法を提案する(図1)．視線を可視化するためには，眼球位置・視線方向を推定する必要がある．本稿では，これらを頭部位置・方向で近似できるとみなし，以下の方法で推定を行う．

(1) 頭部に小型カメラを装着して個人視点映像として記録する．

(2) 個人視点映像から VisualSLAM (以降，単に SLAM と呼ぶ) を用いて頭部位置・方向を推定する．

ただし，実際の共同注視シーンにこの方法を単純に適用しただけでは大きく2つの問題が生じる．第一に，SLAM が正しく動作するためには入力映像が適切な前提を満たしている必要があり，そうでない場合には推定結果の信頼性は全く保証されない点である．前提を満たすように映像を記録することは参加者の行動を大きく阻害することになってしまい，体験学習として本末転倒である．そのため，入力個人視点映像を SLAM が正しく動作する前提を満たしている部分とそうでない部分に分離し，前者のみを SLAM に適用することが求められる．第二に，個々の個人視点映像を用いた SLAM の結果を統合することが難しい点である．SLAM では初期化によって特徴点マップが作成される(以降単にマップと呼ぶ)．座標系の取り方はカメラ毎，初期化フレーム位置毎に異なるため，複数のマップを一つの座標系に統合する必要がある．しかし複数のマップ中の特徴点の対応を確認するのは困難であるため，推定されたマップを用いて各 SLAM 座標系の位置合わせを行うことは容易ではない．

本稿では，これらの問題に対して，入力映像の特徴や SLAM 処理の途中結果を用いた評価関数により主に SLAM 初期化ステップを対象とした前提チェックを行う手法，及び環境マーカーを用いた複数 SLAM 座標系に対する共通座標系の構成方法を提案する．

2. 関連研究

頭部位置・方向を推定する従来手法としては，環境カメラ

で撮影された人物映像を用いるもの [3] [4] [5] [6] や，磁気センサ・加速度センサなどの計測装置を頭部に装着するもの [7] がある．環境カメラを利用した方法には，人物の全身像・上半身・顔(頭部)などの検出とステレオ視を組み合わせていることが多い．これらの手法はいずれも限定的な空間では効果的であるが，屋外などの広範囲なフィールドで行われる体験学習やワークショップへ導入するには多数の環境カメラが必要となり，設置の手間やカメラキャリブレーションなどのコストが大きくなる．一方，磁気センサは環境中の磁性体の有無などによって特性が変化するため，事前にキャリブレーションを行うことが必要な上，場合によっては利用不可能なこともある．また測定可能な範囲が狭いという欠点もある．加速度センサによって得られる情報は積算されることで初めて位置・方向として利用可能となるため，累積誤差の影響が大きく，適宜それを補正するような別の機構が必要である．

本研究では，体験学習やワークショップの主催者が簡単に構成でき，参加者への負担や行動の制限ができる限り少ない環境で，視線を自動的に可視化することを目指している．そのため，上記のような手法は不適である．個人視点映像を用いた手法では，たかだか参加者分の小型カメラを用意だけで収録が可能であり，さらに参加者の体験の内容が個人視点映像中に映り込むため，視聴・閲覧に基づいた分析に有用である．本稿では環境中にマーカーを貼り付けて使用するが，費用・手間・技術面でのコストは環境カメラを設置するコストと比べて非常に小さなものである．さらに，共同注視を分析・評価したい対象が事前に限定されているならば，その対象にのみマーカーを貼り付けるだけでよい．

本提案において重要な点は特定のアルゴリズムを用いて自動処理を行う場合に，入力データがアルゴリズムの前提条件を満たしているかどうかを自動判定することにある．高度に制御された実験環境ではなく，一般のシーンを対象とした場合には種々の悪状況が発生する．悪状況であることを自動判定することができれば，例えば信頼性の高い結果のみをユーザーに提示する，結果の信頼性とアルゴリズム適用範囲のトレードオフをユーザーが制御する，といったように推定結果の用途を大きく広げることができる．このような観点から考えると，本提案は西谷らが提案した協調的認識 [8] に近い概念を持っている．彼らの提案では，認識途中結果や悪状況を検出する別アルゴリズムを用いて悪状況を検出し，ユーザーにそれを訂正するように促すことで物体認識の精度を向上させている．

3. 複数個人視点映像を用いた SLAM における問題点とアプローチ

3.1 SLAM 初期化ポイントの評価

SLAM のアルゴリズムは数多く提案されているがその基本原理はほぼ共通しており，大きく以下の2つのステップで構成される．

初期化ステップ 初期マップを作成する．

逐次処理ステップ 入力画像(フレーム)毎にカメラ自己位置・姿勢を推定する．条件によってはマップを更新する．

SLAM が正しく動作する，すなわち信頼性の高い結果を出力するかどうかは，初期化ステップに大きく依存する．初期マップの作成は，シーンを異なる視点から撮影した 2 枚（以上）の画像（初期化ポイント）によりステレオ法の原理に基づいて行われ，そのときのカメラ運動量も同時に推定される．そのため，よい初期マップが作成されるための前提条件には以下のようなものが考えられる．

前提条件 1 対応点が正しく抽出されること
 前提条件 2 シーンが静止していること
 前提条件 3 視点間に十分な並進移動があること
 前提条件 4 対応点がシーン中に広く均一に分布していること
 前提条件 5 対応点がシーン中に多く分布していること

前提条件 1 はシーンの同一点が画像上で正しくペアリングされていることを意味しており，これが必須であることは自明である．同様に前提条件 2 はステレオ視における前提条件を示している．前提条件 3 は十分なベースライン距離がない場合にはマップの推定精度が低くなってしまいうステレオ視の特性を示している．前提条件 4 はシーンの特定の領域に対応点が集中していると，その領域付近でしか精度が保証されないことから，対応点はできるだけ広く分布している方が良いことを意味する．以上の 4 つは初期マップの推定精度に関係している．一方で，前提条件 5 は逐次処理における対応点追跡に関係する．逐次処理ステップにおいて良い推定をするためには多くの対応点が引き続き追跡でき，推定に用いられることが望ましい．そのためには初期化における対応点が多いことが求められる．ロボットナビゲーションや AR における SLAM [9] では，これらの条件を満たすようにカメラ運動を制御したり環境を選択することが可能である．しかし，体験学習では参加者は自由に行動するため，上記前提条件を満たしていない状況が多々存在する．それを人手で判断して適切な初期化ポイントを与えることは非常に困難かつ大きな時間と労力を要するので，初期化ポイントを評価する指標を設定し，その評価値に応じて自動的に選択することが求められる．

問題の簡単化のために前提条件 2, 4 については明示的には扱わず，まず本稿では前提条件 1, 3, 5 を評価対象とする．評価は対象の初期化ポイントに初期化アルゴリズムを適用する前と後の二段階に分けて行う．以降，それらを事前評価，事後評価と呼ぶこととする．事前評価では，初期化アルゴリズムとは別の方法により前提条件 1 と 3 を対象に評価を行い，低評価であった初期化ポイントは棄却する．評価には環境マーカーの検出を通してカメラの位置・姿勢を推定するアルゴリズムを併用する．まず，前提条件 1 の評価として，環境マーカーが検出されたフレームペアのみを初期化ポイントの候補とする．これは種々の画像処理を行う上で十分な画質であることを保証するものである．次に，前提条件 3 の評価のために，環境マーカーを用いたカメラ位置・姿勢推定の結果を基に，フレームペア間でのカメラ並進移動量 D_c を推定する． D_c がある一定のしきい値 D_{thresh} 以上であるときにそのフレームペアを初期化処理の評価対象とする．

事後評価では，初期化アルゴリズムの途中結果や最終結果を

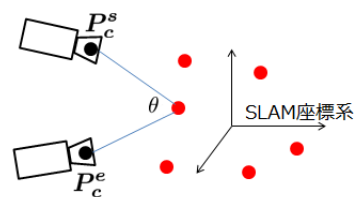


図 2 マップ上の三次元点の見かけ上の移動量
 Fig. 2 Apparent moving amounts of 3D points.

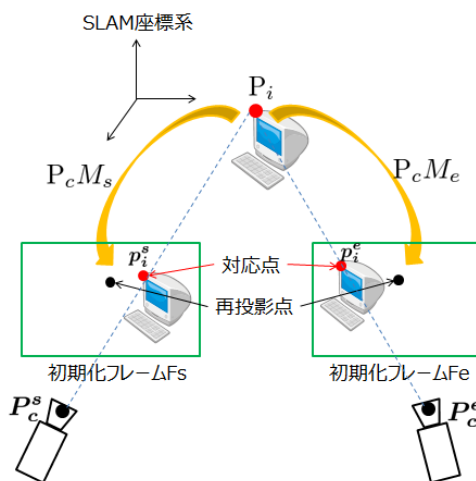


図 3 初期化マップ上の点の画像への再投影
 Fig. 3 Reprojection error of an estimated map.

用いて評価値を導出する．マーカーとカメラが正対に近い幾何関係にある場合，環境マーカーを用いたカメラ位置・姿勢の推定精度はあまりよくないことが知られている．そこで事後評価では，初期化処理で推定された初期化マップとそのときのカメラ運動量から，見かけ上の移動量 θ (図 2 を参照) を導出し前提条件 3 を再評価する．ただし，この評価値はステレオ視としての良さを示したものであり，前提条件 1 と 2 が満たされている時にのみ信頼できる．そこで初期化マップの各点を画像上に再投影したときの誤差 σ を追加の評価値として用いる (図 3) ．つまり θ と σ の両評価値が良い時のみを前提条件 3 が満たされている状況とする．前提条件 5 については，単純に対応点の数 N_c を評価値とする．

3.2 複数 SLAM 座標系の統合

異なる初期化ポイントにより推定されたマップは，一般に原点，座標軸，スケールなどが異なるため，複数の SLAM 座標系を統合する仕組みが必要となる．ただし，マップ上の特徴点分布を用いた位置合わせが困難であることは先に述べた通りである．そこで本提案では，マーカー中心を原点として実際のシーンと同じスケールを持つ環境マーカー座標系を基準座標として用いることで各 SLAM 座標系間の関係を決定する．環境マーカーを用いたカメラ位置・姿勢推定の結果を直接利用することもできるが，前節で述べたように精度があまりよくないという問題がある．そのため，環境マーカーからはその検出結果のみを用いることとし，関係を記述する変換行列の

導出には初期化マップを用いる．具体的には，環境マーカーの4端点の座標値 P_i は SLAM 座標系上ではステレオ法により算出でき，またマーカー座標系上では既知であるので，それらを一致させるような変換行列 M_m^{sj} を推定する．複数の SLAM 座標系の変換行列 $M_{s_i}^{sj}$ はそれら変換行列の積により表すことができる．

4. 初期化ポイントの評価と SLAM 座標系統合のアルゴリズム

4.1 初期化ポイントの事前評価

まず入力映像を連番画像に分解し，それら全てに対して環境マーカーの検出を行う．検出には ARToolKit [11] を用いた．次に環境マーカーが検出されたフレーム群から任意の2枚を取り出してフレームペアを作成する．ただし，組み合わせ爆発を防ぐため，以下に示すような手順でペア数の削減を行う．

(1) 環境マーカーが N フレーム以上連続して検出されているクラスタを取り出す．

(2) 各クラスタ C_i について，クラスタ先頭から N_s フレームおきに代表フレーム $F_{R_{ij}}$ を設定する． j は i 番目のクラスタにおける代表フレームのインデックスを示す．

(3) N_d フレーム以下の距離である2枚の代表フレームによりフレームペア

$$F_{pair} = (F_s, F_e) \text{ s.t. } F_s, F_e \in F_{R_{ij}}, |F_s - F_e| < N_d \quad (1)$$

を作成する．

続いて各フレームペアにおけるマーカー座標系における三次元カメラ位置 x_s, x_e を用いてカメラ並進移動量 $D_c = \|x_s - x_e\|_2$ を推定する．カメラ位置の推定にはマーカーの検出と同様に ARToolKit を用いた．この値が閾値 D_{thresh} よりも大きいフレームペアのみを初期化処理および事後評価の対象とした．

4.2 初期化ポイントの事後評価

対応点の数 N_c

本稿で用いている SLAM アルゴリズムは Klein らによって提案された PTAM [9] とした．PTAM の初期化処理では，まず Shi-Tomashi らの手法 [10] によって初期化開始フレーム F_s で特徴点を検出する．検出された特徴点はテンプレートマッチングにより初期化終了フレーム F_e に向かってトラッキングされ，最後まで追跡された特徴点が対応点として採用される．評価値 N_c はそれら対応点の数であり，SLAM 初期化処理の途中結果を取り出すことで容易に算出できる．

再投影誤差 σ

マップ上の三次元点 P_i ，初期化フレームペアを撮影した2つのカメラ位置・姿勢は初期化処理により推定されており，カメラ射影行列 P_c も既知であるので再投影誤差 σ は以下のように算出できる．

$$\sigma = \frac{1}{2N_c} \sum_{i=1}^{N_c} (\|p_i^s - P_c M_s P_i\|_2 + \|p_i^e - P_c M_e P_i\|_2) \quad (2)$$

ここで， M_s, M_e は SLAM の初期化により推定された，SLAM 座標系からそれぞれ F_s, F_e 取得時のカメラ座標系への変換行列， p_i^s, p_i^e は P_i の画像 F_s, F_e 上での対応点である (図3)．

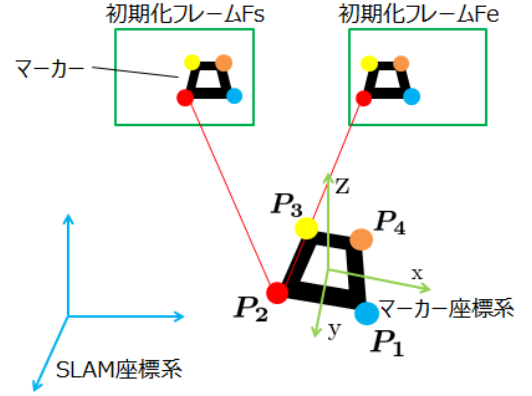


図4 SLAM 座標系とマーカー座標系間の変換の推定

Fig. 4 Estimation of a transformation between the SLAM and the marker coordinate systems.

見かけ上の移動量 θ

マップ上の三次元点 P_i の見かけ上の移動量は各カメラから P_i を見たときの2ベクトルの為す角 (図2) として定義できるので，

$$\theta = \frac{1}{N_c} \sum_{i=1}^{N_c} (\cos^{-1} \frac{(P_i - P_c^s) \cdot (P_i - P_c^e)}{\|P_i - P_c^s\|_2 \|P_i - P_c^e\|_2}) \quad (3)$$

と定式化される． P_c^s, P_c^e は各カメラの中心座標を示す．

4.3 マーカー座標系と SLAM 座標系との変換行列

環境マーカーを共通の座標系として設けたため，複数 SLAM 結果の統合は，SLAM により推定された環境マーカー座標系への変換行列 M_m^{sj} ($j = 1, 2, \dots$) を導出することに帰着する (図4)．

$$\begin{bmatrix} P_1 & P_2 & P_3 & P_4 \end{bmatrix} = M_m^{sj} \begin{bmatrix} \frac{h}{2} & \frac{h}{2} & -\frac{h}{2} & -\frac{h}{2} \\ \frac{h}{2} & -\frac{h}{2} & -\frac{h}{2} & \frac{h}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (4)$$

P_i は SLAM 座標系上での環境マーカー4端点の三次元位置を示しており，フレームペアでの環境マーカー検出結果からステレオ視により推定する． h は正方形である環境マーカーの1辺の長さである．ただし，上式で求めた変換行列の信頼性は， P_i の精度，つまり初期化の良さに依存することに注意する．

5. 実験

5.1 目的及び方法

提案手法が初期化ポイントの良さを正しく評価していること，及び複数 SLAM 結果を正しく統合できていることを確かめるために検証実験を行った．入力映像となる個人視点映像は以下のような条件下で収録した (図5)．

- 1名がホワイトボードを用いて他の2名に説明を行うシーンを再現 (実際に起こり得る共同注視の場面を想定)
- 大きさ $h = 150mm$ の環境マーカーをホワイトボード上に設置

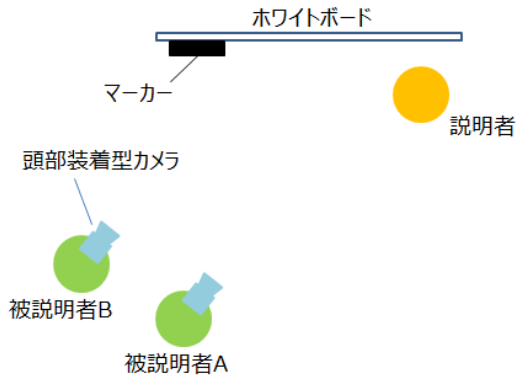


図 5 実験環境の設定

Fig. 5 Geometric configuration of the recording situation.

- 映像は $736 \times 480 \text{ pixel}$, 25 fps , 約 6874 kbps に設定
- 収録時間はおよそ 10 分

実験では被説明者 2 名の個人視点映像を対象に検証を行った。初期化ポイントの評価値

初期化が正しく行われていれば、続く逐次処理ステップも含めてマップ・カメラ位置姿勢が正しく推定されるはずである。それを、図 6 のように入力画像上に描画した SLAM 座標系の挙動から目視で定性的に判定する。正しく初期化されていれば、SLAM 座標系はシーン中に固定されて描画されるが、そうでない場合は移動・変形して描画される。以上の結果とそのときの初期化ポイント評価値との相関性を検証することで提案手法の妥当性を確認する。各パラメータは経験的に $N = 5$ フレーム, $N_s = 30$ フレーム, $N_d = 800$ フレーム, $D_{\text{thresh}} = 30 \text{ mm}$ とした。

複数 SLAM の統合

統合の良さを可視化ビューを用いて定性的に評価する。実験では 2 名分の個人視点映像を記録しているので、一方 (A) の個人視点映像に他方 (B) の視線を SLAM の推定結果に基づいて描画した図 7 のような可視化ビューが作成できる。描画された注視先と実際の注視先の画像上での距離が 150 ピクセル以内にある場合統合できている (Success) とし、そうでない場合統合できていない (Failure) と判定する。

5.2 結果と考察

初期化フレームペアの評価

入力画像上に描画した SLAM 座標系の挙動は、概ね以下の 3 パターンであった。

パターン A 逐次処理ステップに入ってからすぐに SLAM 座標系がシーン中で大きく移動・変形するもの (図 8(a))

パターン B SLAM 座標系がシーン中で小さく移動、変形するもの

パターン C SLAM 座標系が安定的にシーン中に固定されているもの (図 8(b))

評価値の大小に対して、これらのパターンがどのように表れたかを調査したところ以下のような結果を得た。

前提条件 3,5 が共に良い評価値であったときの結果例を表 1

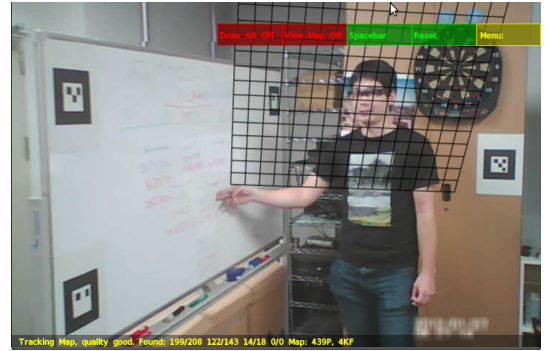


図 6 入力画像上に描画した SLAM 座標系 (メッシュ部)。メッシュの広がり XY 平面を示す。

Fig. 6 The SLAM coordinate system drawn as the mesh in the input image.



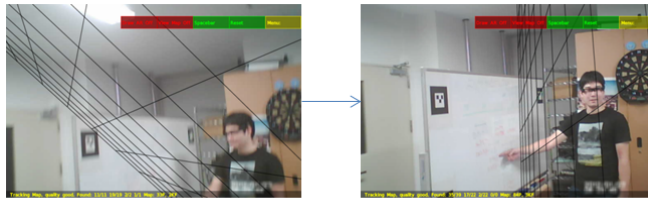
図 7 実験結果の検証用に作成した簡易な可視化ビュー。

Fig. 7 An example of visualization results of estimated gazing points.

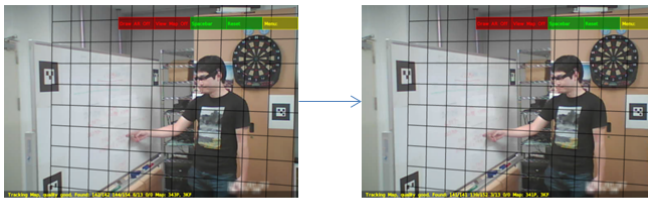
に示す。PTAM がホモグラフィ変換に基づいて初期化を行うのに対応点が最低 4 点が必要であることを考えると、100 点以上の対応点は十分に多いといえる。またカメラからシーン中の点までの実際の距離は平均して約 3m ほどであった。これより、その見かけ上の移動量 θ はたかだが数度と見積もることができるので、その範囲内で大きな値を持つ初期化フレームペアを選択している。対応する SLAM 座標系の挙動はパターン B とパターン C が多く現れており、評価値が正しく機能していることが確認できる。

初期化フレームペアの中には θ が上限値を大きく上回るものがいくつか見られたので追加で調査を行った (表 2)。実際のシーン中の点までの距離から考えて、このような θ は本来現れるはずがない。SLAM 座標系の挙動からも推定に失敗していることが見て取れる。詳しく検証を行って見たところ、対応点の推定に失敗していたり、図 9 に示すように動物体上の対応点を初期化に使っていることが多く確認できた。つまり前提条件 1 と 2 が満たされていない状況である。これが原因で実際とは異なる構造のシーンとして推定してしまったため、結果として

σ は小さいのに θ があり得ない値になってしまったと推察できる。



(a) パターン A の例



(b) パターン C の例

図 8 入力画像上に描画した SLAM 座標系の挙動

Fig. 8 Behavior pattern examples of drawn SLAM coordinate systems.

F_s	F_e	N_c [ペア]	θ [degree]	σ [pixel]	評価
2841	2900	152	7.07223	2.09685	パターン B
4635	4665	189	1.05715	0.412803	パターン C
4665	4695	101	6.03455	0.893766	パターン B
4695	4755	159	2.52042	0.526665	パターン B
5535	5589	158	1.57918	0.230519	パターン C
6760	6820	199	3.44435	0.900979	パターン C
7217	7247	133	1.02545	0.887781	パターン B
7277	7337	140	1.12902	0.875489	パターン C
10435	10525	100	4.13619	0.526976	パターン C
10578	10698	233	1.53138	0.269084	パターン C
12653	12743	189	1.14484	0.726423	パターン C
15382	15412	327	1.28978	0.40292	パターン C
15382	15532	199	1.21909	0.259602	パターン C
17083	17113	306	1.75974	0.28291	パターン C
17083	17143	237	1.20378	0.283883	パターン C
17083	17173	211	1.39547	0.266601	パターン C

表 1 前提条件 3 および 5 についての評価値が共に良い場合の SLAM 座標系の挙動

Table 1 Results for good evaluations about the assumption 3 and 5.

複数カメラの位置姿勢推定結果の統合

前提条件 3,5 が共に良い評価値であったときの結果例を表 3 に示す。14 ペア中 8 ペアで図 10 に示すような比較的良好の結果となったが、実用に耐える精度には至っていない。環境マーカーが画像中に現れる大きさは比較的小さいため、4 端点のみ

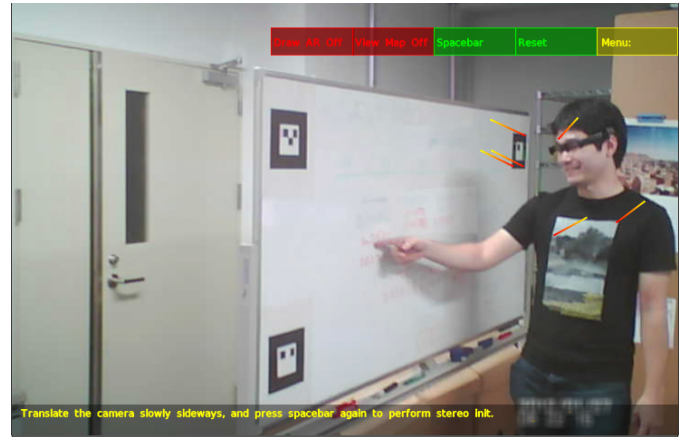


図 9 表 2 の状況における初期化の例。初期化時にトラッキングされている動物体上の点の割合が大きい。

Fig. 9 An example of situations in table 2 with the assumption 2 being bad.

F_s	F_e	N_c [ペア]	θ [degree]	σ [pixel]	評価
1090	1116	12	21.3068	5.072	パターン A
1102	1116	15	36.1543	3.38393	パターン A
1105	1116	18	29.8972	2.45261	パターン A
2767	2788	10	18.7092	6.21924	パターン A
2777	2811	12	12.1092	0.75993	パターン A
2777	2815	10	26.4954	3.22686	パターン A
2777	2822	10	31.2623	1.34828	パターン A
2784	2815	13	24.8266	0.562989	パターン A
2784	2838	13	27.3064	7.19601	パターン B
2784	2841	12	84.7019	1.44854	パターン A
2799	2871	15	19.1901	1.11664	パターン B
2811	2907	10	13.9453	2.97173	パターン A
4605	4635	14	17.6988	4.93134	パターン C
5035	5040	10	27.3438	4.21279	パターン A
7879	7943	10	38.8841	5.2505	パターン A

表 2 θ の値が上限よりも大きい場合の画像中の SLAM 座標系の挙動
Table 2 Results for θ much larger than estimated upper limitation.

を用いる本手法では限界があるとする。ただし、本評価方法は初期化以降の逐次処理も含めたものであるため、どちらが原因で低精度となったのかは検証すべきである。そのためには定性的な評価ではなく、シーンの構造やカメラの移動の真値を計測するとともに、推定値がどの程度正しいかを定量的に分析する必要がある。

6. おわりに

本稿では、個人視点映像を入力とした SLAM アルゴリズムに対して、初期化の前提条件を満たしていることを評価する手法、および、複数の SLAM 推定結果を統合する手法について検討・提案を行った。前提条件についての評価値は、環境マーカーと初期化処理の結果を併用して導出した、対応点の正しさ・マップ推定の正しさ・対応点の数、などである。複数推定結果の統合問題は環境マーカーを共通の座標系として用いることで



図 10 推定された可視化ビューの例

Fig. 10 Examples of estimated and truth gazing points.

F_s	F_e	N_c [ペア]	θ [degree]	σ [pixel]	注視位置の正しさ
2841	2900	152	7.07223	2.09685	Success
4635	4665	189	1.05715	0.412803	Failure
4665	4695	101	6.03455	0.893766	Success
4695	4755	159	2.52042	0.526665	Success
5535	5589	158	1.57918	0.230519	Success
6760	6820	199	3.44435	0.900979	Success
7217	7247	133	1.02545	0.887781	Success
7277	7337	140	1.12902	0.875489	Failure
10435	10525	100	4.13619	0.526976	Success
10578	10698	233	1.53138	0.269084	Success
12653	12743	189	1.14484	0.726423	Failure
15382	15412	327	1.28978	0.40292	Failure
15382	15532	199	1.21909	0.259602	Failure
17083	17113	306	1.75974	0.28291	Failure
17083	17143	237	1.20378	0.283883	Failure
17083	17173	211	1.39547	0.266601	Failure

表 3 前提条件 3,5 が共に良い評価値であったときの可視化ビュー作成結果

Table 3 Results of gazing point correctness with good evaluations for the assumption 3 and 5.

解決する。

検証実験を通して、評価値が良い場合には SLAM 初期化や注視点可視化が比較的正しく行われることは定性的に確かめられたが、同時に対応点の正しさの評価が不十分であったこと、可視化精度は実用にはまだ達していないことなどが課題として明らかとなった。また、評価値が悪いときの挙動や初期化結果のみ対象を絞った定量的な評価が必要であることも確認できた。今後はそれらの検証を行うとともに、本稿では扱わなかった前提条件や逐次処理における評価値の設計などを通して、より確度の高い可視化ビューの作成について検討を続ける必要がある。

将来的には可視化結果から共同注視のパターンなどを分析し、その結果に基づいた自動推定手法によりさらなる支援を行うことを計画している。

文 献

- [1] 栗山直子, 寺井あすか, 安原正晃, 徳永健伸, 山岸侯彦, 楠見孝, 共同問題解決時の二者の視線一致が共有知識の形成に及ぼす効果, 日本認知科学会大会発表論文集, 27th, pp. 2-31, 2010.
- [2] 鈴木紀子, 馬田一郎, 岩澤昭一郎, 井ノ上直己, 神谷俊郎, 伊藤慎宣, 鳥山朋二, 小暮潔, 共同作業における発話・視線行動に関する分析, 日本認知科学会大会発表論文集, 24th, pp. 506-509, 2007.
- [3] 田口進也, 田中 昭二, 頭部検出器と多視点カメラを用いた人物検出方式の検討, 電子情報通信学会技術研究報告, Vol. 107, No. 538, pp. 7-10, Mar., 2008.
- [4] 中田篤志, 角康之, 西田豊明, 非言語情報の出現パターンによる会話状況の特徴抽出, 情報処理学会研究報告. HCI, ヒューマンコンピュータインタラクション研究会報告, No.13, pp. 1-8, Nov., 2009.
- [5] 前田真希, 小川剛史, 清川清, 竹村治雄, ウェアラブル拡張現実感のための赤外線カメラのステレオ計測と姿勢センサを用いた位置・姿勢推定, 日本バーチャリアリティ学会論文誌, Vol. 10, No. 4, pp. 459-466, Dec., 2005.
- [6] 菅野裕介, 佐藤洋一, 顔変形を伴う 3 次元頭部姿勢の単眼推定, MIRU2007, pp. 311-317, Jul., 2007.
- [7] Masakatsu Kurogi, Takeshi Kurata, Personal Positioning based on Walking Locomotion Analysis with Self-Contained Sensors and a Wearable Camera, International Symposium on Mixed and Augmented Reality (ISMAR2003), pp. 103-112, Oct., 2003.
- [8] 近藤一晃, 西谷英之, 中村裕一, 協調的物体認識のためのマンマシンインタラクション設計, 電子情報通信学会論文誌, Vol. J94-D, No.8, pp. 1206-1215, Aug., 2011.
- [9] Georg Klein and David Murray, Parallel Tracking and Mapping for Small AR Workspaces, International Symposium on Mixed and Augmented Reality (ISMAR2007), Nov., 2007.
- [10] Jianbo Shi and Carlo Tomashi, Good Features to Track, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR94), pp. 593-600, Jun., 1994.
- [11] Kato, H., Billingham, M, Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System, Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR 99), pp. 85, Oct., 1999.