車載カメラ映像の時空間マッチングによる自車位置推定

小野晋太郎 *1　松久亮太 *2,1　川崎洋 *3　池内克史 *2,1
東京大学 生産技術研究所 先進モビリティ研究センター（ITS センター）*1
東京大学 大学院情報学環・学際情報学府 *2
鹿児島大学 大学院理工学研究科 *3

車載カメラを用いた相対自車位置決定の一手法を提案する．ある走行区間の車載カメラ映像を「検索キー」として与えると，予め蓄積された車載カメラ映像から同じ区間に相当する映像を「検索結果」として返すものである．従来法では良い初期値を必要としたり，車両側方への変位や日照条件の変化に応じきれない等の問題があった．本手法では周辺の建物が GPS 電波の障害となることを逆手に取り，建物列の見え方を時系列的に記述する「時空間特徴量」と連続 DP マッチングを導入することで問題解決を試みた．実験では，日照条件が近い場合で 2 m，異なる場合で 4 m 程の精度で GPS に頼ることなく位置を決定できた．

Vehicle Localization using Spacetime Matching of On-vehicle Video

Shintaro ONO*1　Ryota MATSUHISA*2,1　Hiroshi KAWASAKI*3　Katsushi IKEUCHI*2,1
Advanced Mobility Research Center (ITS Center), Institute of Industrial Science, The University of Tokyo*1
Interfaculty Initiative in Information Studies, The University of Tokyo*2
Graduate School of Science and Engineering, Kagoshima University*3

Abstract We propose a method to estimate relative position of the self vehicle using on-vehicle video camera. When a short-length on-vehicle video is given as a "searching query," this system returns a part of video stream from video database as a "searching result." Existing methods had some problems, i.e. requirement for good initial value, difficulty in dealing with lateral displacement of a vehicle and lighting condition. Our method take advantage that buildings around obstruct the GPS wave, and solve the problems by introducing "spacetime feature" where the appearance of the buildings are described in time-series. The experimental result showed that the system could determine the position with less than 2 m error when the lighting condition was similar.

Keyword: *On-vehicle camera, Positioning, Localization, Spacetime feature, Continuous DP*

1. はじめに

1.1 背景

自動車等のの自己位置を求めることは，ITS の様々なアプリケーションを支えるもっとも基本的な問題の一つと言える．良く知られた GPS は絶対位置を直接求めることができるが，電波の弱い地域での精度向上は，補正データの重畳，準天頂衛星などインフラ側の整備に頼らざるを得ない．これに対し，近年になって一般車両にも搭載されつつあるカメラや距離センサなどは，直接に位置を得るものではないが，計測対象との相対位置・姿勢などを得ることができ，精度を向上できる余地も大きい．特にカメラは駐車・後退時の運転支援をはじめ，事故記録や風景記録など実用から趣味の範囲まで活用されており，今後も様々な目的で設置が進むことが予想され，様々な地域・時間における映像が一般車両によって日常的に記録される時代も近づいていると言える．

本稿では，車載カメラを用いた相対的な自車位置決定法の一つを提案する．予め蓄積されたある範囲の車載カメラ映像データベース (DB) に対し，ある走行区間の車載カメラ映像が「検索キー」として与えられる

と，同じ区間に相当する映像を「検索結果」として返すものである．このように車載カメラ映像どうしを対応づけることができれば，DB 上の映像に対する相対自己位置が求められる（DB 上で絶対位置が既知である場合は絶対位置も求まる）．このほかにも，都市活動のモニタリング，街並み変化（新規建設など）の検出，歩行者や車両などを分離した街並み景観画像の生成，さらには三次元地図構築の省力化や観光用コンテンツの整備などにも応用の可能性が開ける．

1.2　本手法の位置づけと関連研究

カメラや距離センサを利用して自己位置を得る手法は自立走行ロボットの分野で研究が進んでおり，SLAM (Simultaneous Localization And Mapping)，VSLAM (Visual SLAM) といった分野を形成している．これらの手法は，画像であればピクセル精度の整合が取れるように位置を推定することができる（実際の精度は手法により異なる）が，予め初期位置を必要とすることが一般的であり，屋外では手動あるいは GPS で与えることが多い．しかし実際には，図 1 のように，電波状況の良くない場所における GPS の信頼性は十分でなく，得られる位置情報は「地区」程度の水準である．

本手法の位置づけは，このような言わば km オーダの初期値から，画像間の対応点（共通特徴点）探索などが可能な言わば m オーダの位置情報を得ることである．ジャイロセンサはこのような位置推定に適当であるが，車載カメラ映像を直接に用いれば，インターネット上に多数存在する，過去に撮り貯められた映像のみのデータも活用できる．

車載カメラ映像どうしを対応づけるもっとも直接的な方法としては，撮影画像の画素値どうしをそのまま比較して相関を取るといった方法が考えられ，[1] などの例もある．このような方法は入力映像と比較対象の映像がほぼ同じ走行軌跡上から撮影されたものである場合には適用可能だが，車両側方への変位（すなわち走行車線の違い）が大きい場合には対応できないほか，日照条件の変化に対応しにくい，計算量が膨大になる等の問題がある．[2] では SVM を用いた学習により 2 系列の映像を対応づけているが，実験区間は 350m ほどである．

本手法では，2 次元画像情報のみに頼るのではなく，シーンの構造特徴などに注目する．そもそも都市部において GPS 電波の阻害要因となっているのは建物列であり，それを逆に積極的に利用して対応付けを行うことを考える．具体的には，建物列の見え方（色やシルエット）の変化を時系列的に取り出したものを「時空間特徴量」として定義し，これを指標として連続 DP マッチングにより対応付けを行う．

なお本手法では，将来的にはカメラが更に安価になり，様々な方へ向けて設置されることが十分に考えられることを見越して，入力画像として全方位画像を用いている．但し，実験では全視野を参照しない場合についても評価を行っている．

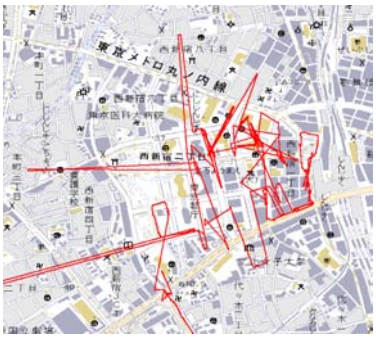
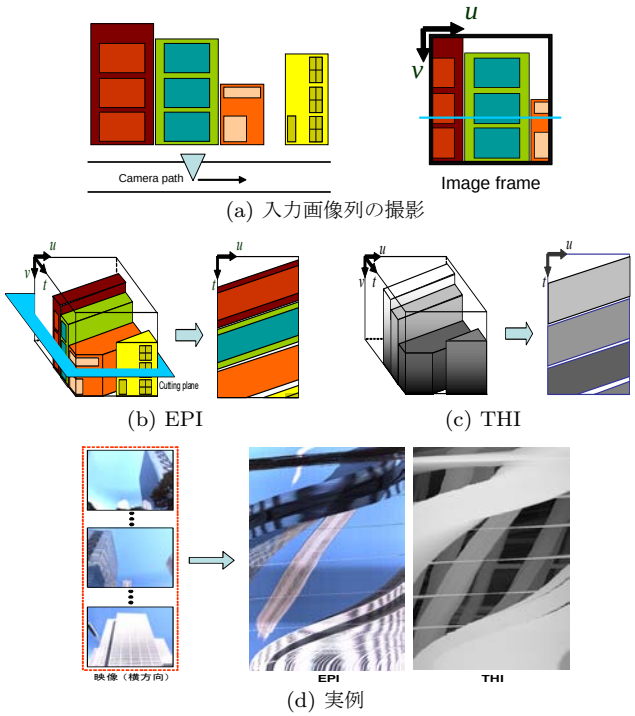


Fig.1 西新宿地区（高層ビル街）の GPS 軌跡



(a) 入力画像列の撮影

(b) EPI　(c) THI

(d) 実例

Fig.2 EPI と THI の概念図と実例

2.　時空間特徴に基づく車載カメラ映像の対応付け

2.1　時空間特徴量

対応付けのための指標には，動画像の時間変化を空間的に捉えることのできる「時空間特徴」を 2 種類用いる．一つは動画像処理において古くから知られたエピポーラ平面画像 (Epipolar Plane Image, EPI) [3]，もう一つは以前に我々が提案した時系列高さ画像 (Temporal Height Image, THI) [4] である．図 2 に EPI, THI の概念図を示す．

EPI とは，カメラの移動が単純な水平移動の場合，各撮影フレームのある高さ 1 ライン分だけを切り出して連続的に並べたものであり[1]，シーンの 3 次元復元 [3] や時系列データの対応付け [5] などに利用されている．EPI からはシーンの奥行き情報も（粗く）得ることはできるが，基本的には色情報のみの集合であるため，同

[1] 簡略化した説明．正確には，撮影フレームを並べた「時空間ボリューム」を撮影点間のエピポーラ面で切断し，切り口に現れる画像である．

じ場所であっても日照条件や道路上の走行位置によって変化する．従って EPI 単独では本用途には十分とは言えない．

一方，THI とは，主に建物列の「シルエットの見え方」（高さパターン）の変化を時系列的に記述した時空間画像である．撮影画像から空領域とその他の領域の境界線を求め，対象物の高さ（ピクセル数または仰角）をグレイスケール値で表したものである．THI はシーンの構造情報の推移を追うことができるため，日照条件が異なる映像どうしを対応づけることが可能となる．しかし，カメラから建物群までの距離などにより値が変化するため，値の相対関係や対象までの奥行き値を考慮する必要がある．

ここで我々は，上記の特徴が補完関係にあることに着目した．すなわち，

- 色情報：EPI に含まれる
- 構造情報（高さ）：THI に含まれる
- 構造情報（奥行き）：EPI に含まれる
- 構造情報（幅）：THI およびマッチングの伸縮（後述）に含まれる

このように両者を用いれば，シーンの特徴をほぼ網羅した対応付けが可能になる．これが本手法の最大の特徴である．

図 3 に特徴量の定義を示す．EPI および THI の横軸を u，縦軸（時間軸）を t とすると，基本的には画素値 $E(u,t), T(u,t)$ そのもの[2]が特徴量を表す．実際に対応付けを行うときは t 軸方向を一定間隔 h に区切り，これを対応付けの最小単位（ノード）とする．すなわち，時間方向の分解能は h フレームである．

ノード 1 つあたりの特徴量は，ノード内部の画素値の重み付き和とする．このとき，車両進行方向に対して真横方向では，走行車線の違いなどによる見えの変化が最も少ないと考えられるため，そこに相当する EPI および THI の中央部 $(u = w/2)$ 付近においては重みを多くする．

2.2 連続 DP による映像の対応付け

映像どうしの対応付けには，連続 DP マッチング [6] を用いる．2 系列のデータ（この場合は映像）について，切り出す始点と長さを変化させながら DP マッチングを行うことで，各映像の撮影時の走行速度の違いや始点・終点の違いによる影響を吸収しつつ対応区間を探索することが可能である．

処理の手順は以下の通りである．検索キーとして与える映像を固定長パターン A，探索対象であるデータベース映像のうちの一つを自由長パターン B として，

1. A の先頭ノードに対し，対応付けコスト（ノード間の差異を表す評価値；後述）が最も小さいノードを B から探し，対応付け開始点とする．A のノード m と B のノード n が対応付けられることは，図 4(a) 右側のようなグラフで表される．
2. B の開始点以降の各ノードを終点の候補として，始点〜終点間で対応付けコストの合計が最小となるグラフ上のルートを求める（図 4(b)）．この部分は，始点と終点が既知である通常の DP マッチングと同じである．
3. 各ルートを比較し，その中でも最小コストとなるルートを対応付け結果とする（図 4(c)）．

また，進行方向が逆の映像（対向車線を走行したとき）と対応づける可能性も考慮するため，t 軸を反転させた EPI, THI に対しても同様のマッチングを行い，コストが低い方を採用する．

前述の時空間特徴量を使うと，A のノード m と B の

[2]EPI の場合は RGB 色情報（0〜255），THI の場合は高さ情報を表すグレースケール値（0〜255）．

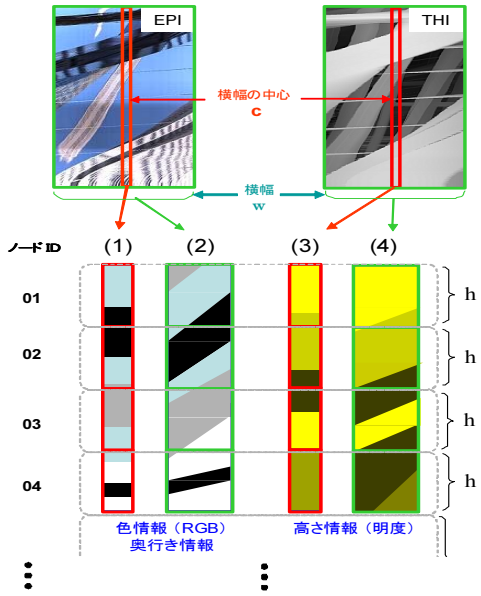
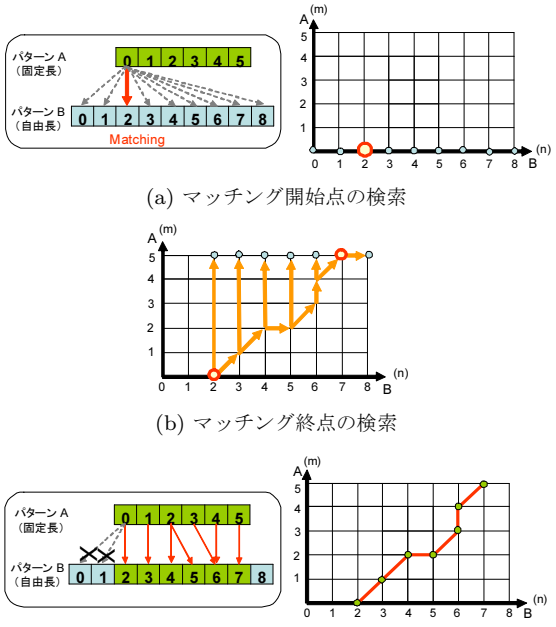


Fig.3 時空間特徴量



(a) マッチング開始点の検索

(b) マッチング終点の検索

(c) マッチング結果

Fig.4 連続 DP による映像間の対応付け

ノード n の対応付けコストは，次のように表される．

$$Cost(m,n) = \sum_{t=0}^{h-1}\sum_{u=0}^{w-1} W_E(u)|E_{A(m)}(u,t) - E_{B(n)}(u,t)| \quad (1)$$

$$+ \sum_{t=0}^{h-1}\sum_{u=0}^{w-1} W_T(u)|T_{A(m)}(u,t) - T_{B(n)}(u,t)| \quad (2)$$

$E_\square, T_\square$ は EPI, THI の当該ノードにおける画素値である．重み係数 W_E, W_T を変化させることにより，例えば日照条件が大きく異なる場合には THI 側の特徴のみを用いるなどの調整が可能となる．W の最適な与え方については今後の課題であり，今回は W_E, W_T は同じ重みとする．

3. 実験

3.1 実験内容

実環境から撮影したデータにより，対応付けの性能を確認する実験を行った．GPS 電波が十分に受信できない環境であっても，走行しているエリア程度（数 km オーダ）の粗い位置は特定されていることを想定し，図 5 のような東京都新宿区内の経路を実験対象とした．撮影経路の道路は片側 2〜3 車線，道幅はおよそ 16 m 程度である．撮影車は制限速度内で流れに沿って走行した．

具体的には，車載カメラ映像 A から各 100 フレーム（約 40 m 相当）の区間 3 カ所を「検索クエリ映像」A_1, A_2, A_3 として選び，別の機会に撮影した車載カメラ映像 B の全区間 12930 フレーム（1.9 km 相当）と対応付けた結果，A_1, A_2, A_3 と同じ区間が得られるかどうかを検証した．A と B は別の機会に撮影したものであり，走行速度（信号待ちの有無も含む），走行車線，走行方向は異なっている．また連続 DP のノードは $h = 5$ フレーム間隔で設定した．

また，対応付け手法の特性を確かめるため，映像 B においては

- B_O：手を加えないデータ
- B_L：低画質カメラ想定データ（解像度を 1/8 にしたもの）
- B_F：高速走行想定データ（フレームを 1 ずつ間引いたもの）
- B_{LF}：B_L, B_F の両方

のようなデータを作成して影響を検証した．図 6 は，それぞれに相当する EPI の例である．

撮影には全方位カメラ Ladybug2 [7] を用いた．2048 × 1024 ピクセル，30 fps の全方位画像を取得後，進行方向に対して左側，右側それぞれを 512 × 512 ピクセルの透視投影画像に変換して EPI, THI を作成した（従って，全方位カメラに限定された手法ではないことに注意されたい）．

実験では，車載カメラが多数設置されている場合とそうでない場合を想定し，左右両側の EPI, THI を使用した場合と，片側のみを使用した場合を試した．

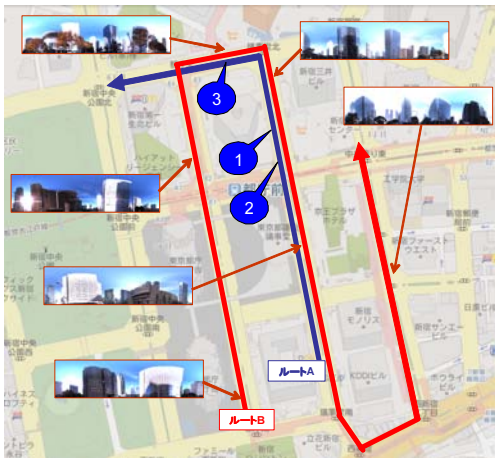


Fig.5 撮影ルート A, B および検索区間 A_1, A_2, A_3

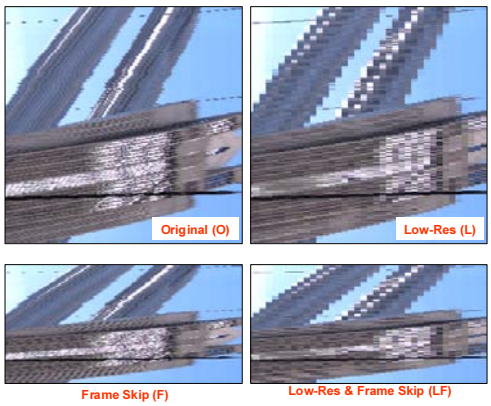


Fig.6 低解像度，高速走行などの検証条件に相当する EPI

3.2 実験 1：日照条件が近い場合

まず A_1 区間の映像と B 全体の映像で対応付けを行い，同じ区間が得られるかどうかを検証した．A_1 は見通しが良く，晴天で日照条件も B に似ている．図 7 にこの区間の EPI, THI を示す．車両の進行方向は逆向きであり，対向車線を走行しているため，横方向には 3〜10 m 程度ずれた地点から撮影した映像どうしを対応づけることになる．

図 8 は A_1 に対応づけられた区間 $\hat{B}_1$ を示したものである．また表 1 は $\hat{B}_1$ と，真の対応区間 B_1 との差分フレーム数を示したものである．もともと走行車線が異なるため真の対応区間を厳密に定義することは難しいが，ここでは入力の全方位画像 A_1 と B を目視で比較することにより真の対応区間を求めた．

最も良い条件である B_O の場合は，誤差が 1 ノード分（5 フレーム）に納まった．これは，A_1 撮影時の速度から換算すると，車両進行方向の距離にしておよそ 2 m である．横方向に 3 m 以上離れた点から撮影した映像どうしを外部機器に頼らず対応づけた結果であることを考えると，十分に良いと言える．

また悪条件を仮定した場合 (B_L, B_F, B_{LF}) でも，この程度の解像度低下や高速走行では対応付けが破綻することはなく，誤差が数 m 程度の増大にとどまった．精密な自車位置特定には十分ではないが，より精密な

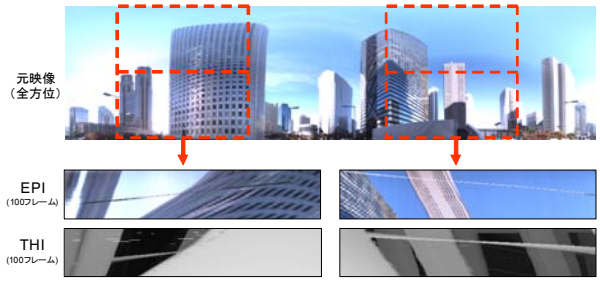


Fig.7 A_1 区間の撮影画像と EPI, THI

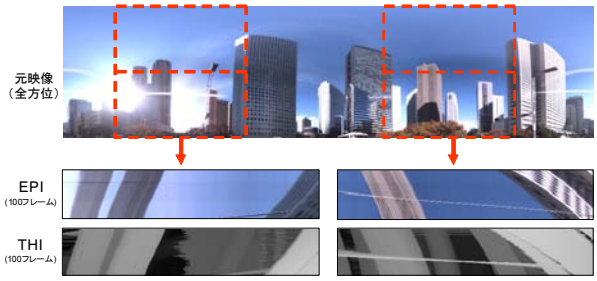


Fig.9 A_2 区間の撮影画像と EPI, THI

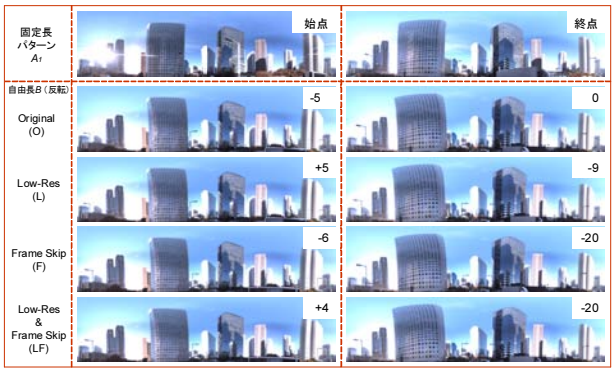


Fig.8 A_1 に対応付けられた区間 $\hat{B}_1$（左右参照時）．$\hat{B}_1$ は A_1 の対向車線から撮影されているため，左右を反転表示している．

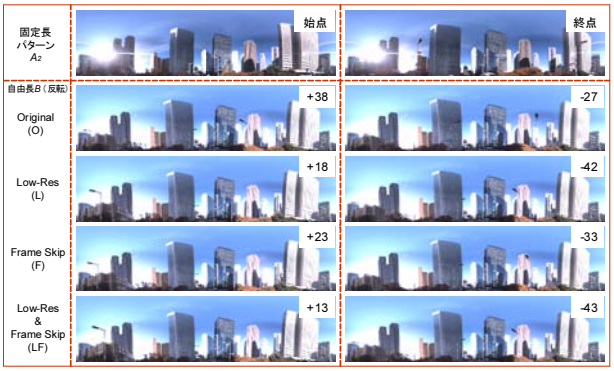


Fig.10 A_2 に対応付けられた区間 $\hat{B}_2$（左右参照時）．$\hat{B}_2$ は A_2 の対向車線から撮影されているため，左右を反転表示している．

Table 1 A_1 に対応づけられた区間 $\hat{B}_1$ と，真の対応区間 B_1 とのずれ（フレーム数）

映像データ使用方向	評価項目	条件バリエーション B_O	B_L	B_F	B_{LF}
左右両側	始点側誤差	−5	+5	−6	+4
	終点側誤差	0	−9	−20	−20
左側のみ	始点側誤差	−5	+5	+9	+9
	終点側誤差	0	−5	−15	−20
右側のみ	始点側誤差	−40	+5	−26	+4
	終点側誤差	−8	−5	−15	−15

Table 2 A_2 に対応づけられた区間 $\hat{B}_2$ と，真の対応区間 B_2 とのずれ（フレーム数）

映像データ使用方向	評価項目	条件バリエーション B_O	B_L	B_F	B_{LF}
左右両側	始点側誤差	+38	+18	+23	+13
	終点側誤差	−27	−42	−33	−43
左側のみ	始点側誤差	+143	+138	−157	−162
	終点側誤差	+68	+58	−237	−242
右側のみ	始点側誤差	+18	+28	+13	−43
	終点側誤差	−32	−42	−33	−20

手法に対する入力値としては良い結果と考えられる．

3.3 実験 2：日照条件が異なる場合

次に，日照条件が異なる場合の対応付け実験を行った．図 9 にこの区間の EPI, THI を示す．A_2 区間は直射日光が写り込んでいるため，左側では白飛び，右側では黒つぶれが生じている．

図 10 は A_2 に対応づけられた区間 $\hat{B}_2$ を，表 2 は $\hat{B}_2$ と真の対応区間 B_2 との差分フレーム数を示したものである．A_1 に比べると全体的にずれが大きくなっている．特に左側の映像データを使用した場合は誤差が大きく，悪条件を仮定した場合もより大きく影響を受けている．左側では，白飛びにより色情報が抜けてしまっているほか，本来は建物が存在する箇所の屋上部も空領域に侵食されたように観測されるため，高さ情報も正確に抽出できていないことが原因と考えられる．

逆に，右側や左右両側の映像データを使用した場合は，一部を除いて $\hat{B}_2 \supseteq B_2$ であり，また全ての場合において $\hat{B}_2 \cap B_2 \neq \phi$ の結果が得られている．図 10 からも分かるとおり，これでも対応づけられた区間の映像から共通特徴点を探索するなどの処理は十分に適用可能であると言える．また，悪条件を仮定した場合でも結果が大きな影響を受けないという点も，実験 1 と同様の傾向が見られる．

このような日照条件による誤差は，対応付けコストのパラメータを調整することで改善することも可能である．表 3 は，ノードの長さを $h = 3$ フレームとし，THI，EPI 内部での重み付けを

$$W_E(u) = W_T(u) = \begin{cases} 1 & (u = w/2 \text{ のとき}) \\ 0 & (\text{その他}) \end{cases} \quad (3)$$

Table 3 表 2 において重みパラメータを変更した場合

映像データ使用方向	評価項目	条件バリエーション B_O	B_L	B_F	B_{LF}
左右両側	始点誤差	+29	+20	+29	+17
左右両側	終点誤差	+7	+1	+9	−9

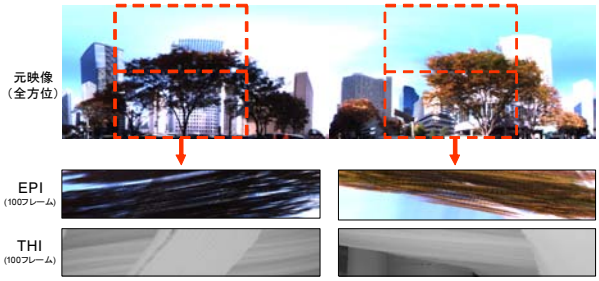


Fig.11 A_3 区間の撮影画像と EPI, THI

として対応付けを行った結果である．

表 2(a) と比べ，特に終点の誤差が減少していることが分かる．特徴量を中央領域に限定することで，誤差の主な原因となっていた太陽光の影響が減少したと考えられる．終点誤差に大きく効果が出た理由は，太陽光が後方にあり，終点に向かうほど横方向（中央）から離れていったからだと推定される．また，ノードの長さを短くすることで，誤差の原因が分散されたと考えられる．逆に始点の方は誤差が増えている．原因の候補としてはノードあたりの情報が減ったことによる影響が挙げられる．これらのパラメータ調整の傾向調査は今後の課題である．

3.4 実験 3：遮蔽が大きい場合

最後に，遮蔽が大きい場合について実験を行った．区間 A_3 では左右両側に大きな街路樹があり，視界の多くを占めている．

この場合は図 11 のように，EPI は両側とも大部分が樹木で覆われている．THI でも左側については奥にある建物の屋上部分が観測されているが，B の当該区間ではその建物も完全に樹木の背後に隠れていた．このような場合に正しく対応付けを行うことは不可能であり，本手法の限界と言える．

提案方式の明確な性能限界の検証については今後の課題したい．また，解決策として，時空間画像を用いて街路樹のような手前の障害物を映像から消去し，該当部分の補間まで行う手法を栗林ら [8] が提案しており，そのような方法を組み合わせる事を考えている．

4. まとめ

本論文では，km オーダの位置関係しか分かっていないような複数系列の車載カメラ映像データに対し，m オーダの精度で同じ場所が観測されているシーンを対応付けにより探し出す手法を提案した．

具体的には，

- エピポーラ平面画像 (EPI) と時系列高さ画像 (THI) から，対象シーンの色情報と構造情報（高さ・奥行き・幅）が含まれた「時空間特徴量」を導入した．
- この時空間特徴量の差異を指標として，連続 DP により複数系列の映像間で対応付けを行う方法を提案した．
- 検証のため，走行方向，速度，車線，解像度，フレームレート，日照条件などが異なる 2 系列の車載カメラ映像を，それぞれ検索クエリ映像，データベース映像として与え，対応付け実験を行った．
- 結果，日照条件が近い場合であれば進行方向に対して約 2 m の誤差で対応付けが可能であった．直射日光による白飛びは大きな対応付け誤差をもたらすが，逆に黒つぶれには強く，検索クエリ映像と対応づけられた結果映像がほぼ包含関係となった．より精度の高い位置推定手法に対して初期条件を与える手法として有効であることが期待される．
- 画像全体を覆うような樹木などは特徴量への影響が大きく，何らかの補間方法を導入しない限りは，改善に限界がある．

今後の課題としては，検証に用いる映像を増やし，ノードの長さや重みパラメータ W のような特徴量についての傾向を調査し，適当な値の決定を自動化したい．また，今回の検証において対応付けが出来なかった大きな街路樹があるシーンについて，先に述べた背景補間手法 [8] などを用いるなどの解決法を考え，利用可能な範囲を広げてより汎用的にしたいと考えている．

謝辞

本研究の一部は，新エネルギー・産業技術総合開発機構 エネルギー ITS 推進事業 (P08018)，ならびに国土交通省国土技術政策総合技術研究所の援助を受けて行われた．

参 考 文 献

1) H. Uchiyama, D. Deguchi, T. Takahashi, I. Ide and H. Murase: “Ego-localization using streetscape image sequences from in-vehicle cameras”, 2009 IEEE Intelligent Vehicles Symposium, pp. 185–190 (2009).
2) 三浦, 森田, ヒルド, 白井 : “Svm による物体と位置の視覚学習に基づく屋外移動ロボットの位置推定”, 日本ロボット学会誌, pp. 792–798 (2007).
3) R. Bolles, H. Baker and D. Marimont: “Epipolar plane image analysis: approach to determining structure from motion”, IJCV, **1**, pp. 7–55 (1987).
4) J. Wang, S. Ono and K. Ikeuchi: “Proposal of temporal height image and matching on-vehicle camera image and 3d building model by thi”, IEICE transactions on information and systems, **J92-D**, 8, pp. 1197–1207 (2009).
5) 三上, タンダ, 小野, 川崎, 大沢, 池内 : “Epi 解析を利用した歪みのない複数ビデオカメラ画像の統合”, 電子情報通信学会論文誌, **J89-D**, pp. 1336–1347 (2006).
6) 岡 : “連続 DP を用いた連続音声認識”, 音響学会音声研資料, **78**, pp. 145–152 (1978).
7) Point Grey Research Inc. : Ladybug2 http://www.ptgrey.com/products/ladybug2/.
8) 栗林, 川崎, 小野, 池内 : “移動カメラ映像の障害物除去のための時空間画像フィルタの提案”, 第 12 回画像の認識・理解シンポジウム (MIRU2009) (2009).