

市街地映像マップの構築のための車載カメラ映像間対応付け

佐藤准嗣[†] 高橋友和[†] 井手一郎[†] 村瀬 洋[†]

[†] 名古屋大学大学院情報科学研究科

〒 464-8603 愛知県名古屋市千種区不老町

E-mail: †{jsato,ttakahashi,ide,murase}@murase.m.is.nagoya-u.ac.jp

あらまし ITS 技術の一環として、市街地映像と位置情報を同時に蓄積したデータベース（市街地映像マップ）の構築が考えられる。我々は、全方位カメラと普及型 GPS を搭載した多数の一般車両を長期にわたり自由に走行させ、各フレームに位置情報を付与した映像を収集、統合して映像マップを手軽に構築することを目指している。そのためにはまず、撮影時期の異なる大量の映像群から同一地点を撮影したフレーム同士を対応付けることと、一般に測定誤差が大きい普及型 GPS による位置情報の高精度化を行う必要がある。そこで本稿では、映像間を固有空間上で DP マッチングすることによるフレームの対応付けと、対応付いた各フレームに付随する GPS 座標を平均化することによる位置情報の高精度化を目的とする。実験では約 9ヶ月間に渡り収集した 40 組のデータを用い、映像間の高精度な対応付けと、映像に付随する位置情報の精度向上に関して、本手法の有効性を示した。

キーワード 全方位画像, 固有空間法, DP マッチング

Alignment of Car-mounted Video Images for the Construction of a Street Image Map

Junji SATO[†], Tomokazu TAKAHASHI[†], Ichiro IDE[†], and Hiroshi MURASE[†]

[†] Graduate School of Information Science, Nagoya University

Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8603 Japan

E-mail: †{jsato,ttakahashi,ide,murase}@murase.m.is.nagoya-u.ac.jp

Abstract As part of an ITS technology, we are creating a database which is composed of street images and their location (Street Image Map). We aim at the automatic construction of a Street Image Map, by aligning a large amount of data, which are pairs of image and coordinates, collected by numerous cars running freely during a long period with a omni-directional camera and a general GPS. For the construction, it is necessary to align the images taken at a same location from a large amount of images taken at a different time, and to improve the precision of the coordinates measured by a general GPS which has a large margin of error. In this paper, our purpose is to align images taken at a same location by comparing image frames by DP matching in an eigen space, and to improve the precision of the coordinates by calculateing the average of the coordinates of the aligned images. Experiments with 40 data pairs showed the effectiveness of our method.

Key words omni-directional image, eigen method, DP matching

1. はじめに

近年、最先端の情報通信技術を用いて人と道路と車両とを情報ネットワークで繋ぎ、交通事故や渋滞等の道路交通問題の解決を図る、ITS（高度道路交通システム）の技術開発が進められている。この中でナビゲーションシステムの高度化は、ドライバにより良い情報を提供するという観点でとても重要である。また一方で、現代社会は都市開発や新たな建築物の建設が盛ん

に行われており、都市の様子は刻一刻と変化している。そのためナビゲーションシステムに利用する地図の更新も頻繁に行う必要があるが、迅速に対応することは難しいのが現状である。

ナビゲーションシステムの高度化の一環として、車載カメラから市街地の景観を撮影した市街地映像と、撮影した地点の位置情報を同時に蓄積したデータベース（市街地映像マップ）の構築を考えている。これは、地図上の任意の地点と、そこから任意の方向を向いたときの風景画像を対応付けたデータベースで

あり、同一地点に対して異なる時期の様々な風景画像を蓄積するものである。これが実現できれば、市街地の風景画像をキーにその位置を取得したり、位置をキーにその地点の画像を表示したりすることが可能となる。また、ある地点の異なる時期の画像を簡単に取得することができるので、刻一刻と変化する都市の状況を把握することが容易になると考えられる。さらに地図上の任意地点の映像を集めることができれば、地図上である経路を設定した時に、その経路を通った時の任意視点での風景映像を生成することも可能となる。

このような映像マップの手軽な構築に向けて、多数の一般車両にカメラと普及型 GPS を搭載して長期にわたり自由に走行することで、映像の各フレームに位置情報を付与したデータを収集し、それらの情報を統合して映像マップを構築する手法を検討している [1]。この手法により、低コストなシステムで多数の車両が自由に走行するだけで、任意位置の異なる時期での映像をデータベース化した市街地映像マップが精度良く構築されることが期待される。映像の取得には、撮影地点から任意の方向を向いたときの画像を効率よく取得するため、全方位カメラを用いる。位置情報の取得には、ディファレンシャル GPS やキネマティック GPS 等、数 mm ~ 数 cm の測位精度を持つ高精度 GPS を用いることも考えられるが、大変高価であるため現時点では本目的のためには利用できない。そこで本研究では、一般のカーナビゲーションシステムに搭載される程度の普及型 GPS を想定する。その際、普及型 GPS では測定誤差が 10m 近くあるため、単純に同じ GPS 座標の画像を集めても正しい映像マップは構築できないという問題がある。また、異なる時期に自由に走行して撮影した大量の映像群から同一地点の映像を対応付けることは簡単ではないということも問題である。そこで本研究は以下の 2 つを目的とする。

- 異なる時期に自由に走行して収集した映像群から、同一地点を映すフレーム同士の高精度な対応付け

- 映像中の各フレームに付随する位置情報の高精度化

映像間の対応付けのために、概ね同じ GPS 座標位置を走行している異なる時期に撮影した映像系列間で、映像照合により、同一地点を映すフレーム同士を対応付ける。映像の照合には固有空間法と DP マッチングを用いる。固有空間法によって映像の高次元の特徴を低次元に変換した後、撮影車両の速度の違いによる映像の時間的な伸縮を吸収するために、固有空間上で DP マッチングをすることで対応付けを行う。これにより同一地点の異なる時期でのフレームを対応付けることができる。そして対応付けられたフレームに付随する位置情報を平均化し、各フレームの位置情報の高精度化をねらう。

2. 関連研究

2.1 映像と位置情報の対応付け

現在、映像と位置情報を対応付けてデータベースとし、位置に基づいて対応する映像を表示するシステム（リアルマップシステム）が提案されている [2]。同様に、車両の位置情報とそれに関連付けられた画像データを収集し、デジタルマッピングを行う技術をモービルマッピングという [3]。しかし、これらのシ

ステムには専用の撮影装置が必要であり、例えば日本全国の市街地映像データを定期的に収集してデータの更新を行う場合、かなりのコストを要すると考えられる。これに対して本研究は、多数の一般車両に全方位カメラと普及型の GPS を搭載して自由に走行するだけで、収集したデータの自動的な統合によって高精度に映像と位置情報を対応付けた映像マップの構築を行う。これによって、低コストで自動的に映像マップの構築・更新を行うことが可能になると考えられる。

また、GPS データと画像特徴を用いて同一地点画像の検出を行う手法も提案されている [4]。しかしこの手法は、GPS データによる位置の絞込みはするが、1 枚の画像特徴のみを用いているため、景観が変化した場合は誤検出が多くなると考えられる。本手法は映像全体で整合を取りながら同一地点を対応付けるため、精度がよくなるといえる。

2.2 市街地映像と地図情報・実世界との対応付け

市街地映像と地図情報とをマッチングする研究も報告されている [5], [6]。これらは映像から建物の両端を表すエッジのパターンを作成し、これを地図から得た境界パターンと対応付けている。これにより、実世界映像を構造化した映像データベースの構築、都市の 3 次元モデルの構築を行う。これらの研究は、本研究で構築・更新された映像マップデータベースを、より有効に利用する手法であると考えられる。本研究により、異なる時期に撮影された同一地点の映像を統合してデータベース化することが可能となるため、これを利用することで、精度良く任意時期の実世界映像の構造化が可能となる。

全方位画像から市街地空間の 3 次元構造を復元する手法も提案されている [7]。本研究は全方位カメラを利用して映像マップを構築するため、このデータに上記手法を適用することで、異なる時期の市街地空間モデルを再現でき、また、最新のモデルへの更新も簡単であると考えられる。

3. 市街地映像マップの構築

3.1 手法概要

図 1 の手順に従って、市街地映像マップの構築を行う。データ収集では、全方位カメラによる市街地の映像と、同時に GPS からの位置情報を取得する。そして大量にある映像から同一区間を走行して得た映像を切り出す。つまり概ね類似する GPS 座標を持つ映像区間を検出し、これを対応付けに用いる。次元圧縮処理では、後の映像間対応付けでの計算時間短縮、使用メモリ量削減のために、映像の高次元の特徴ベクトルを低次元に変換する。映像間対応付けでは、切り出した同一区間の複数映像間でフレーム単位で対応付けを行い、同一地点を示すフレームを得る。最後に、対応付けした同一地点を示す複数のフレームに付随する GPS 座標を平均化し、各フレームの GPS 座標の精度向上をねらう。以降の節で各処理について詳しく説明する。

3.2 映像と位置情報データの収集

全方位カメラを車載し、走行中に映像を撮影する。このとき全方位カメラは、撮影車両の周囲 360°を撮影できるように、車両の上部に鉛直上向きに取り付ける。また映像と同期させて GPS からの位置情報を取得する。

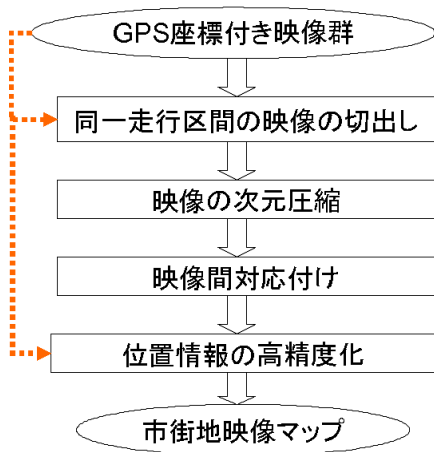


図 1 市街地映像マップの構築の処理手順

映像と GPS との同期は専用のハードウェアシステムを作ることできるが、本研究では予備実験のため、ある基準点を利用して時刻合わせを行った。まずその基準点の正確な座標を取得し、映像でその地点を通過した時刻と、GPS 座標が基準点の座標と一番近づいた時刻とを合わせる。

一般に GPS による測位更新時間と撮影カメラのフレームレートとは同期していない。そこで、GPS 座標が付与されていないフレームに対しても線形補間によって仮想的に座標を付与する。

3.3 同一区間の映像の切り出し

収集したデータは多数の車両が自由に走行して得たものである。そこで、映像間対応付けによって同一地点同士を対応付ける前処理として、大量のデータの中から同一区間を走行して撮影した映像を切り出す必要がある。GPS の誤差は 10m 以下なので、データの GPS 座標を用いれば走行した経路の判別は可能である。そこで、GPS 座標によって概ね同じ経路を走行している部分の映像の切り出しを行う。これにより、収集したデータから、異なる時期に同一の区間で撮影した映像を複数取得することができる。

本研究は予備実験のため、切り出す区間の始点と終点を指定して経路を特定した。しかし大量の映像から切り出しを行うためには、GPS 座標系列の連続 DP [11] 等を用いる必要がある。

3.4 映像の特徴次元圧縮

固有空間法 [12] を用い、図 2 の手順に従って映像の特徴ベクトルを次元圧縮する。これは、高次元の特徴ベクトルを低次元の特徴に変換することで、計算量や使用メモリ量の削減をねらうものである。以下に詳しく説明する。

3.4.1 特徴ベクトルの抽出

映像の各フレームの各画素の R 値、G 値、B 値を要素として持つベクトルをそのフレームの特徴ベクトルとし、映像をこの特徴ベクトルの系列と考える。このときフレーム全体の画素を使用するのではなく、図 3 に示すようなマスクで使用領域を限定する。図 3 (右) の白画素が使用領域である。使用領域を限定するのは、メモリ領域と計算時間の短縮のためだけでなく、以下の理由がある。

- 全方位画像は円形である

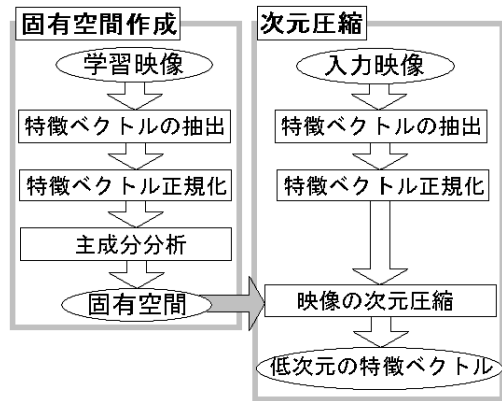


図 2 映像の次元圧縮の処理手順

- 円の外側：人や駐車車両が多く、マッチングに適さない
 - 円の中心付近：空の領域であり、マッチングに適さない
 - 円の上側と下側：前後の車両が映るため適さない
 - 全方位カメラのコードが映る部分を除去する必要がある
- マスクは人手で領域を指定したが、空領域を自動で除去するなど、上記の必要性を考慮したより良い領域選択法も考えられる。マスクの画素数を N_0 としたとき、特徴ベクトルは各画素の R 値、G 値、B 値それぞれを用いるため、各特徴ベクトルの次元数は $N = 3N_0$ となる。

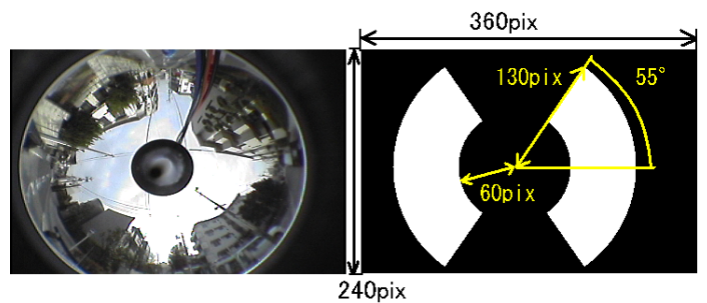


図 3 (左): 全方位画像

(右): 特徴ベクトル抽出に用いるマスク ($N_0 = 25,538$)

映像中のあるフレームから特徴ベクトルを抽出するには、フレームとマスクをラスタスキャンし、マスクの白画素位置に対応するフレームの画素の RGB 値を並べ、以下のようなベクトル \hat{x} を作成する。

$$\hat{x} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_N]^T \quad (1)$$

3.4.2 特徴ベクトルの正規化

撮影した映像は日時によって天候が異なるため、照明条件が異なることになる。この違いを吸収するために画像の正規化を行う。すなわち、以下の処理によって元の特徴ベクトル \hat{x} の平均値を 0、ノルムを 1 にして、ベクトル x とする。

$$x = [x_1, x_2, \dots, x_N]^T \quad (2)$$

$$x_n = \frac{\hat{x}_n - \bar{x}}{\sqrt{\sum_{i=1}^N (\hat{x}_i - \bar{x})^2}} \quad (n = 1, \dots, N), \quad \bar{x} = \frac{1}{N} \sum_{i=1}^N \hat{x}_i \quad (3)$$

3.4.3 主成分分析

固有空間法は，主成分分析によって得られた K 個の主成分より，固有値の大きい順に第 1 主成分から第 $k (< K)$ 主成分のみを用いて新たに固有空間を作成する．そして各データをこの固有空間に線形射影することで，情報損失量を最小にして特徴ベクトルを低次元に圧縮する．これまでの手法で，映像中のあるフレームの特徴ベクトル x を得ることができた． x が一般的な市街地映像の性質を表現できるだけ十分多くあるとすると，この複数の特徴ベクトルから以下の手順で固有空間を求めることができる． x が N 次元ベクトルであり，それが M 個用意されたとき，行列 X を以下のように定義する．

$$X = [x_1, x_2, \dots, x_M] \quad (4)$$

求めるべき固有空間の次元を k とすると， X の共分散行列 $Q = XX^T$ の，対応する固有値の大きい順に k 個の固有ベクトル

$$\{e_1, e_2, \dots, e_k\} \quad (5)$$

によって張られる部分空間が，求めるべき固有空間である．

3.4.4 映像の次元圧縮

特徴ベクトル x は映像中のフレーム毎に抽出される．つまり映像のフレーム数を p とすると，

$$\{x_1, x_2, \dots, x_p\} \quad (6)$$

が映像を表す．これらの特徴ベクトルを，式 (5) によって張られる固有空間に投影することで k 次元ベクトルに変換し，以下のような固有空間上の座標系列を得る．

$$\{y_1, y_2, \dots, y_p\} \quad (7)$$

3.5 固有空間上での DP マッチングによる映像間対応付け

異なる時期に同一の区間で撮影した映像間で，フレーム単位の対応付けを行う．この処理によって，同じ地点を映しているフレーム同士を対応付けることができる (図 4)．撮影時の車両速度の違いによる映像の時間的な伸縮を吸収し，映像全体で整合をとるため，対応付けには DP マッチングを用いる．

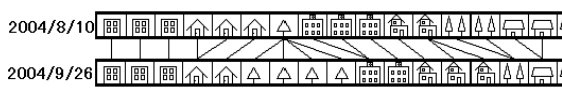


図 4 DP マッチングによる映像間対応付け

前節によって次元圧縮された 2 つの画像系列 $\{y_1, y_2, \dots, y_p\}$ と $\{y'_1, y'_2, \dots, y'_q\}$ に対して，以下の通りに DP マッチングを行うとする．ここでは画像間の相違度として固有空間上のユークリッド距離を用いる．つまり

$$d(i, j) = \|y_i - y'_j\|^2 \quad (8)$$

である．まず $D(1, 1) = d(1, 1)$ とし，漸化式

$$D(i, j) = \min \begin{cases} D(i-1, j) + \omega_1 \cdot d(i, j) \\ D(i-1, j-1) + \omega_2 \cdot d(i, j) \\ D(i, j-1) + \omega_3 \cdot d(i, j) \end{cases} \quad (9)$$

を計算する．ここで $\omega_1, \omega_2, \omega_3$ は相違度に対する重みであり，実験時にはそれぞれある定数を指定する．漸化式を更新する際，選んだフレーム番号対 (i, j) を記録しておく．最後に $D(p, q)$ を取り出し，これが 2 映像間の相違度になる．このとき $D(p, q)$ を計算するまでに選んだフレーム番号対 (i, j) の系列が，2 映像のフレーム間の対応を表す．

3.6 平均化による位置情報の高精度化

GPS による測定値の精度は，ばらつきの尺度である標準偏差により表されるのが一般的である．標準偏差は平均値からのずれの大きさであるため，系統誤差 (平均値と真の値とのずれ) を含めることはできない．しかし，長時間の観測を行えば系統誤差は小さく抑えられるため，長期間にわたって同一地点で測位された測定値の平均は，真値に収束すると考えられる．そこで本研究では，標準偏差により GPS 座標の誤差を評価し，その平均値を高精度な座標と考える．

DP マッチングは 1 対 1 の対応付けであるため，前節の映像間対応付けで c 個の映像を使用した場合，ある 1 つの映像と残りの $c-1$ 個の映像との対応付けを行う．これによって，映像中の任意地点に対して，相互に類似した c 枚のフレームとそれに付随する GPS 座標を c 個得ることができる．

映像 A_1 と映像 A_2 を対応付けしたとき， A_1 のある 1 つのフレームと A_2 の複数のフレームが対応付けられる場合がある．この場合は複数フレームの中央のフレームを対応するフレームとして抽出する．

このようにして得た c 枚のフレームの組は同一地点を映していると考えられるため， c 個の座標は同一地点で観測された座標であるといえる．よってこの c 個の座標の平均は，より精度の高い座標であると考えられる．従ってこの平均化した座標を元の c 枚のフレームに付与することによって，映像の各フレームの位置情報を高精度化することができると考えられる．また c が十分に大きい場合は， c 個の座標の標準偏差が，測位された座標の誤差の程度を表す．

4. 実験と考察

4.1 映像間対応付けの実験

4.1.1 実験概要

本手法を用いて，まず映像間の対応付けの実験を行い，フレーム単位でどの程度正確に対応付けられているかを確認した．実験には以下のデータを使用した．

- 撮影映像：2004 年 8 月から 2005 年 4 月までの約 9 ヶ月にわたり同一経路 (図 8 の白点の系列、約 1.2km) を走行して撮影した 40 データ (平均約 7,000 フレーム、全て異なる日時)
- 全方位カメラ：VS-C14N (ガイストン社製、図 5, 6)
- カメラ取り付け位置：撮影車両の中央、高さ約 165 cm，鉛直上向き (図 7)
- 映像のフレームレート：30 fps
- GPS の測位更新時間：1 sec.
- 撮影車両の速度：40 km/h 以下
- 各フレーム画像のサイズ：360 × 240 pixels
- 特徴ベクトルの次元：圧縮前 64,458 次元，圧縮後 20 次元

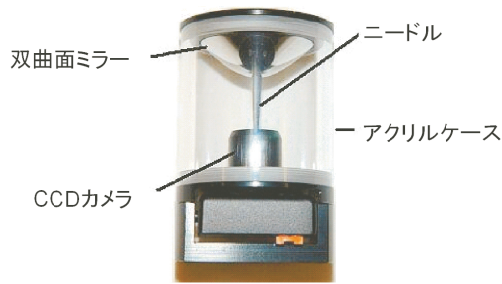


図 5 全方位カメラ



図 6 車載した全方位カメラ

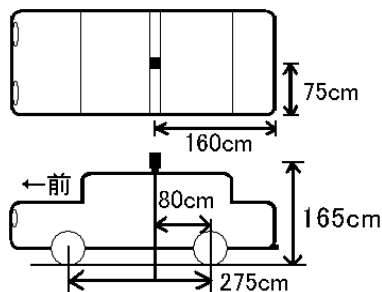


図 7 全方位カメラ取り付け位置

次元圧縮処理の固有空間の作成には、上記のデータとは異なる日に同一経路で撮影した映像から 1,000 フレームを取り出し、学習サンプルとして用いた。

対応付けを行った区間は図 8 中の 2 つの区間 A, B であり、それぞれ GPS 座標によって切り出した。

- 区間 A : 約 147 m, 約 550 フレーム
- 区間 B : 約 170 m, 約 1,300 フレーム

区間 A は片側 1 車線の道路である。経路上に建設中の建物があり、映像中でこの建物が映っているフレームは約 100 フレーム (区間 A 全体の約 18.2%) を占めている。撮影開始時点では更地であったが、半年後には建物の形ができてきている。

区間 B は 1 車線の一方通行道路である。この区間では建物の大きな変化がないが、街路樹が季節によって葉をつけていたり落葉していたりといった変化はある。

DP マッチングに関して、式 (9) において相違度に対する重みは $(\omega_1, \omega_2, \omega_3) = (2, 1, 2)$ とした。これは予備実験において最も精度が高かった値であるが、さらに精度が上がる重みの付け方もあると考えられる。

映像間対応付けの後、対応付いた同一地点を示す複数のフレームに付随する GPS 座標を平均化する。そして平均座標を各フレームに付与することで、映像の位置情報を高精度化する。

4.1.2 実験結果

表 1 に区間 A での結果の一部を、表 2 に区間 B での結果の



図 8 撮影を行った経路 [14]

一部を示す。表には映像名と撮影日時を示し、2 つの映像間で対応付けを行った。相違度は DP マッチングで計算された映像間の相違度を示す。

成功率の評価は目視で行い、対応付いたフレームが最も類似したものと対応付いていた場合は正解、対応付いたフレームより類似するものが存在していた場合は不正解とした。そして全フレーム中の正解フレームの割合を成功率とした。映像の区間の切り出しは GPS 座標により行うが、誤差があるため開始と終了フレームはそれぞれ映像間でずれ (平均 20 フレーム程度) がある。そこで、両方の映像に同一地点のフレームが存在する部分を評価の対象とした。

表 1 区間 A の対応付け結果の一部

映像名 : 撮影日時	映像名 : 撮影日時	相違度	成功率 (%)
A ₀ : 2004/08/10	A ₁ : 2004/08/17	121.1	98.8
"	A ₂ : 2004/09/14	167.1	92.2
"	A ₃ : 2004/10/25	151.7	94.3
"	A ₄ : 2004/11/09	140.4	86.7
"	A ₅ : 2004/11/26	193.3	73.5
"	A ₆ : 2004/12/20	121.8	93.9
"	A ₇ : 2005/01/11	152.9	92.3
"	A ₈ : 2005/02/07	202.2	80.2
"	A ₉ : 2005/03/01	242.0	72.5
"	A ₁₀ : 2005/04/14	161.7	92.4
平均		165.4	87.6

対応付けの結果は概ね良好な成功率を得ることができた。GPS 座標による走行区間の切り出しでは開始フレームにも終了フレームにも平均 20 フレーム程度のずれがあったが、DP マッ

表 2 区間 B の対応付け結果の一部

映像名：撮影日時	映像名：撮影日時	相違度	成功率 (%)
B ₀ ：2004/08/10	B ₁ ：2004/08/17	493.4	96.7
"	B ₂ ：2004/09/14	1026.5	86.2
"	B ₃ ：2004/10/25	406.1	97.3
"	B ₄ ：2004/11/09	452.9	95.2
"	B ₅ ：2004/11/26	391.5	97.2
"	B ₆ ：2004/12/20	615.3	93.6
"	B ₇ ：2005/01/11	586.8	92.3
"	B ₈ ：2005/02/07	597.2	94.3
"	B ₉ ：2005/03/01	717.3	93.3
"	B ₁₀ ：2005/04/14	508.4	94.8
平均		467.2	94.1

チングの結果，両方の映像に同一地点のフレームがある部分ではほぼ正確に（表 1，2 に示すように）対応付いていた．一方の映像で車が停止した場面においては，動いている映像の 1 フレームと停止している映像の複数フレームが完全に対応付いており，撮影車両が停止する場合でも成功することを確認した．

区間 A で 11 月下旬以降に評価が低くなっているのは，それまで更地であった場所に建物が建った地点でずれが大きいためである．また映像 A₄，A₅ は，この地点に大きな路上駐車車両があり，それを避けるため，大きく車線変更を行ったことも原因と考えられる．それ以外の時期も建設現場は常に少しずつ変化があったため，その地点でずれが生じていることがあった．

区間 B では全体として，車が左右にずれて走行している地点で少しの対応付け誤りが発生していた．なお映像 B₂ はある地点で大きく誤りが生じていたが，これも駐車車両を避けて車が横に動いた地点で生じていた．

図 9 に，区間 A 上のある地点に対応付いたフレームの一部を示す．各フレームの下に映像名とそのフレーム番号を示す．どれもほぼ同一地点を映しているといえ，撮影日時によって天候が異なっていることがわかる．

図 9 で示した区間 A 上のある地点に対して，それぞれのフレームに付随する GPS 座標とその平均座標を図 10 に示す．これは全てのフレームで対応付けが正解であった地点である．中心の四角点が平均座標であり，点在する丸点が各データを示す．矢印は撮影車両の進行方向を示す．

図 10 に示された，対応付けられた複数の座標の標準偏差は 6.86m であった．これが，この地点で長期間観測して得られた GPS 座標の誤差であると考えられる．このような標準偏差を全フレームで平均したものを表 3 に示す．

表 3 各地点に対応付けられた複数の GPS 座標の標準偏差

区間	標準偏差 [m]
A	5.06
B	6.48

4.1.3 考 察

表 2 に示した結果は，車線変更や建築物に変化のない経路での映像間対応付けの精度である．つまり映像の局所的な変化が少ない区間であり，平均で 94.0%と，高精度に対応付けが行わ

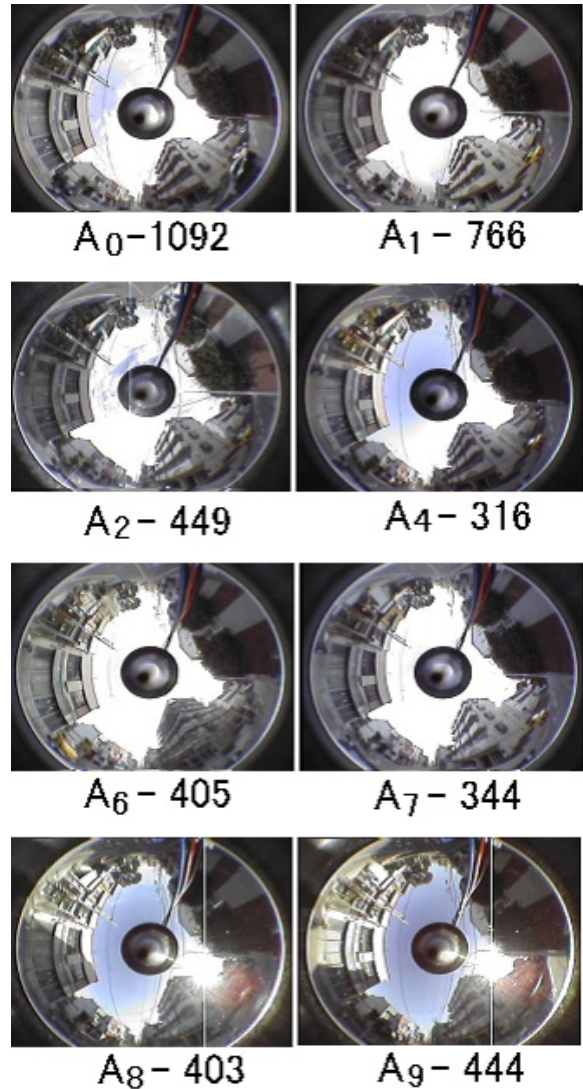


図 9 対応付けられた区間 A 上のある地点

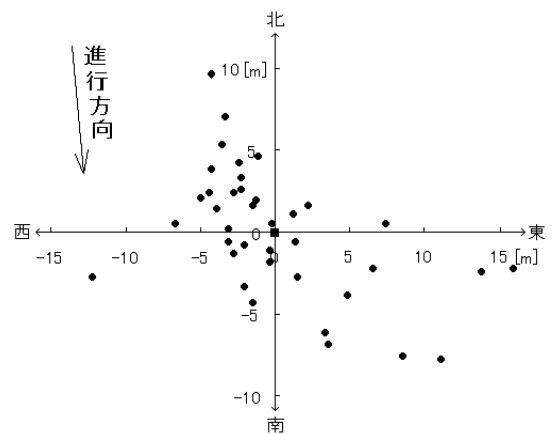


図 10 対応付けられたあるフレームに付随する GPS 座標

れていることがわかる．また表 1 に示した結果は，車線変更や建築物に変化のあった経路での対応付けの精度である．映像の変化が多かった区間であるが，平均 87.6%と，かなり高い精度で対応付けできている．このことから，異なる時期の同一地点における映像を対応付けることが可能となったといえる．

同一地点の GPS 座標は標準偏差で 5~7m のばらつきがあったが、これらの平均化により各地点の平均座標を得た。一般に、GPS 座標を長期にわたって計測した値は真値に近づくと考えられる [10]。もしこの仮定が正しければ、この平均座標はより高精度になっていると考えられる。よってばらつきのある測定値を、平均化によって高精度な値に収束させることが可能となると考えられる。

a) 映像間対応付け誤りの要因

区間 A は経路上に建築物を建設中の地点を含んでおり、これは映像中の全フレームの約 18.2% を占めている。撮影の初期ではあまり変化がなかったため対応付けも精度良く行われていたが、11 月下旬にはそれまで更地であった場所に大きなマンションのシルエットが現れたため、この地点で大きく対応付け誤りが発生していた。しかしこの建築物の前後ではうまく対応付けができており、映像中の局所的な変化のために全体的に対応付けが破綻するという事はなかった。

また、11 月下旬にはこの建物の前に建設用の車両が駐車しており、これを避けるために車線変更を行ったことで映像が左右に動いた。これも対応付け誤りを発生させた要因と考えられる。特に映像 A₅ はかなり大きく左右に動いたため、それに伴って誤差も大きくなっていった。これは区間 B でも同様であり、車両が車線変更や追い越しのために左右に動くことで映像に左右のずれが生じることが、対応付け誤りの一因となっている。

図 11 に、映像 B₀ と映像 B₁ で対応付け誤りのあった地点を示す。上の 2 つの画像が誤って対応付けられたフレームであり、左下の画像が正解のフレームである。車が左右にずれたことがこの誤差を発生させた一因であると考えられる。

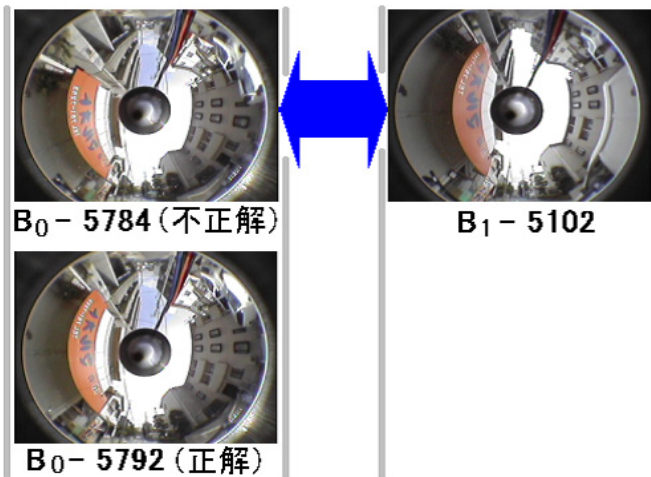


図 11 誤りのあったフレーム対

b) 映像間の相違度と対応付け成功率の関係

実験結果において、映像間の相違度の大きいものの成功率が低い傾向にあることがわかる。もしも対応付けが破綻して精度のとても低い結果が得られた場合、対応付いたフレームは同一地点とはかけ離れたものとなり、位置情報の平均化の際に精度を大きく落とすこととなる。これを防ぐために、映像間の相違度を見て極端に大きかったときはその結果を破棄するような処

理を行えば、誤った対応付けデータを後の位置情報の高精度化処理で使用しないようにすることが可能である。

c) 固有空間上でのユークリッド距離による評価

図 12 に、映像 A₀ と映像 A₁ を対応付けたときの固有空間上の距離による評価を示す。これは、DP マッチングによる映像間対応付けの結果が、固有空間上で最も近いフレーム同士で対応付けられているかどうかの確認である。最も近いフレームと何フレーム離れて対応付けられたかを、DP マッチングに用いた全フレームに対して計算した。図より、最も固有空間上で近いものと対応付けられたフレームが全体の 25.2%、最も近いものと 1 フレーム以内に対応付いたものが全体の 53.9%、最も近いものと 2 フレーム以内に対応付いたものが全体の 71.6% である。これは目視での 98.8% という評価とはかなり異なる結果となった。またこの図では、DP マッチングの結果と固有空間上で最も近いものとの差が 5 フレーム以内のものが全体の 88.6% となっている。しかし本実験で用いた映像では、5 フレームの差は目視してもかなり異なる地点であると認識できるため、固有空間上での距離による評価と目視でのフレーム対応付けの評価にかなり相違があると考えられる。

そこで固有空間上の距離による評価結果を目視で確認してみた。固有空間上で最も近いフレーム同士を対応付けた結果を見たところ、固有空間上で類似するものは必ずしも目視による評価と一致せず、あるフレームに複数のフレームが対応したり、対応するフレームが前後していたりすることがあった。つまり、固有空間上で距離が近くても、元の画像では必ずしも最も類似するとは限らないといえる。これは、車載カメラからの映像は連続しているため前後のフレームがかなり似通っており、次元圧縮の際に細かな情報が失われたためであると考えられる。

以上から、固有空間上での 1 フレームずつの類似度が必ずしも高くなくても、映像全体で DP マッチングによって整合性をとった対応付けを行っているため、結果として全体的に高い評価を得られたのであると考えられる。

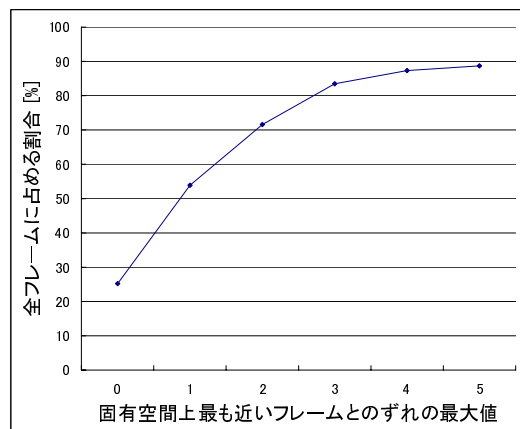


図 12 映像間対応付けの固有空間上での評価

d) 映像間対応付けによる位置合わせの精度

映像間対応付けが精度良く行われれば、同一地点の画像を複数の映像から得ることができる。このとき同一地点と認識された複数のフレームに付与された高精度な位置情報は、フレーム

の標準化誤差を含んでいる。つまり、撮影車両の速度とカメラのフレームレートの関係により、対応付けを行う2つの映像中に全く同一地点から撮影したフレームがあるとは限らない。車両速度40km/h以下でフレームレート30fpsのとき、連続する2つのフレームの撮影地点間の距離は0.4m以下である。つまり2つの映像間で同一地点と認識されたフレームも、0.4m以下の距離の誤差を含むことになる。よってGPS座標の平均化によって求められた高精度な位置情報も、撮影車両の進行方向に対して0.4m以下の誤差を含んでいるといえる。

e) GPS座標の誤差

実験結果より、複数の映像から対応付けられた同一地点のGPS座標のデータには、標準偏差にして5~7m程度の誤差があることが確認された。これは使用したGPSの測位精度が10m以下であることと一致する。ここで、ある地点で長期間計測したGPS座標の平均が正しい座標に収束すると考える。するとこの結果は、本手法によって、映像中の任意フレームに付随するGPS座標の精度を向上させられる可能性があることを示している。つまり長期間にわたる位置座標の平均化によって、同一地点と認識されたフレームに付随するGPS座標は、映像間対応付けの精度により、車両の進行方向に対してフレームの標準化誤差（本実験で用いたデータでは0.4m程度）に収めることができると考えられる。

この結果から得られた5~7m程度の誤差は、GPSの誤差によるものの他に以下の原因が考えられる。

- データ収集時に、映像とGPSの同期をとる際に数フレームのずれが生じる可能性がある
- 映像間対応付けの誤りによって異なる地点のフレームが対応付いてしまう

データ収集時に生じる誤差は、さらに多くのデータを集めて平均化することによって小さくすることが可能である。映像間対応付けに誤りが生じるとGPS座標の平均化の際に精度を落としてしまう。これはデータを増やすことでは解決できないため、映像間対応付けの精度を上げることが必要となる。本手法では、複数のデータのうちの1つのデータとその他の映像との対応付けを行った。このとき選ぶ1つのデータによって、対応付け結果が大きく異なる可能性がある。特に映像で局所的な変化がある場合は対応付け誤りが発生しやすい。そこで、収集したデータから平均映像を生成し、これと各データを対応付けることで、映像の局所的な変化に強い対応付けが実現されることが期待される。

5. む す び

本研究では、市街地映像マップの構築に向けて、映像間対応付けと位置情報の高精度化を目指した。長期間にわたって車載映像とGPS座標を同時に取得したデータを用い、同一区間を走行して得た複数の映像同士を対応付けることで、異なる時期の同一地点の画像を対応付け、それらの位置情報を平均化した。映像の取得は全方位カメラを用いた。同一経路の映像の切り出しはGPS座標を用いて行い、映像間の対応付けは固有空間上でDPマッチングすることで行った。

- 映像間対応付けは、経路上で変化の少ない映像では94.0%と高い精度で行うことができ、異なる時期の同一地点の画像を対応付けることが可能となった。また建築工事や車線変更により映像に変化の多い経路でも、平均87.6%と高い精度で対応付けを行うことができた。

- 位置情報の高精度化に関して、対応付けで同一地点と認識されたフレームに付随するGPS座標には5~7mの誤差があったが、対応付けが精度良く行われればこれを、車両進行方向に対して0.4m以下に収められることを確認した。

今後の課題として、より大量のデータに対する適用等が挙げられる。また本手法を適用したデータから市街地映像マップを構築するために、映像群から任意の経路の映像を切り出す手法や、異なる車線の映像間対応付け、データベースとしてのインタフェース作成等が挙げられる。

謝 辞

日頃より熱心に御討論頂く名古屋大学村瀬研究室諸氏に感謝する。データ収集にご協力頂いた野田和広氏に感謝する。本研究の一部は文部科学省科学研究費補助金および21世紀COEプログラム補助金による。

文 献

- [1] 佐藤准嗣, 高橋友和, 井手一郎, 村瀬洋: “車載全方位カメラとGPSを用いた市街地映像マップの構築”, 2005年信学会総合大会, D-12-43, Mar. 2005
- [2] 神田準史郎, 脇本浩司, 田中聡: “街並み映像を利用した携帯電話向け道案内システム”, 2004年信学会総合大会, A-17-21, Mar. 2004
- [3] 山野芳樹: “モバイルマッピングシステムによる道路空間データ収集手法の開発”, APA no.77-1, (財)日本測量調査技術協会, Nov. 2000
- [4] 角谷直樹, 山崎俊彦, 相澤清晴: “GPSと画像特徴を用いた車載ログ映像の類似検索”, 2005年信学会総合大会, D-11-133, Mar. 2005
- [5] 川崎洋, 谷田部智之, 池内克史, 坂内正夫: “実世界映像の自動構造化によるインタラクティブ利用”, 信学論(D-II), vol.J82-D-II, no.10, pp.1561-1571, Oct. 1999
- [6] 上原将文, 全炳東: “数値地図を利用した広域都市3次元モデル構築”, 信学論(D-II), vol.J84-D-II, no.8, pp.1921-1924, Aug. 2001
- [7] 宮川勲, 石川裕治, 若林佳織, 荒川賢一: “車両運動投影モデルに基づく全方位画像系列からの市街地空間の3次元構造復元”, 情報処理学会コンピュータビジョンとイメージメディア研究会論文誌, vol.45, no.SIG13(CVIM10), 2004
- [8] 八木康文, 谷内田正彦: “実時間全方位視覚センサ”, 情報処理学会コンピュータビジョンとイメージメディア研究会論文誌, vol.42, no.SIG13(CVIM3), 2001
- [9] 横矢直和, 山澤一誠, 竹村治雄: “全方位ビデオカメラを用いた視覚情報メディア”, 情報処理学会コンピュータビジョンとイメージメディア研究会論文誌, vol.42, no.SIG13(CVIM3), 2001
- [10] 坂井丈泰: “GPS技術入門”, 東京電機大学出版局, 2003
- [11] 伊藤慶明: “時系列バターンの任意部分区間的高速マッチング手法Shift CDP法”, 信学論(D-II), vol.J86-D-II, no.9, pp.1267-1277, Sept. 2003
- [12] H. Murase, S. Nayar: “Visual Learning and Recognition of 3-D Object from Appearance”, Int. J. Comput. Vision, vol.14, pp.5-24, 1995
- [13] 田村秀行: “コンピュータ画像処理”, オーム社, 2003
- [14] 昭文社 Super Mapple Digital Ver.4 地図使用承認 ©昭文社第47G009号