

NAIST-IS-MT9551094

修士論文

固有空間内での軌跡の類似性に基づく
動画像検索に関する研究

藤本 泰史

1997年2月14日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科において
修士(工学) 授与の要件として提出された修士論文である。

提出者： 藤本 泰史

指導教官： 横矢 直和 教授
鹿野 清宏 教授
竹村 治雄 助教授

固有空間内での軌跡の類似性に基づく 動画像検索に関する研究*

藤本 泰史

内容梗概

本論文では、動画像データベースの中から入力画像系列に類似した系列を検索する内容に基づく動画像検索の実現を目指して、画像系列間の類似性を固有空間内の軌跡の類似性によって評価する手法を提案する。本手法は画像系列毎に固有空間と固有空間内の軌跡をあらかじめ学習しておき、入力画像系列と学習した固有空間との距離、ならびに固有空間内の軌跡の類似度を評価することで類似する画像系列を検索する手法である。また、一つの画像系列が一つのジェスチャを表現する実動画像を対象とした実験を行ない、画像系列の検索が可能であることを示す。

キーワード

内容に基づく画像検索, 動画像検索, 固有空間, 軌跡の類似度, ジェスチャ認識

*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 修士論文, NAIST-IS-MT9551094, 1997年2月14日.

Retrieval of Image Sequences Based on Similarity of Trajectories in Eigenspaces*

Yasufumi Fujimoto

Abstract

In this thesis, a method for retrieving an image sequence in a database based on its contents is proposed. The method evaluates the similarity between image sequences based on the similarity of trajectories in eigenspaces. Two criteria to evaluate the similarity are employed. One concerns the distance between an input image sequence and the eigenspaces, each of which is previously organized from an image sequence. The other concerns the similarity of trajectories in the eigenspaces. Experimental results show the feasibility of the proposed retrieval method with real image sequences which represent hand gestures.

Keywords:

content-based image retrieval, image sequence retrieval, eigenspace, similarity of trajectories, gesture recognition

*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT9551094, February 14, 1997.

目次

1. はじめに	1
2. 関連研究と本研究の位置づけ	4
2.1 画像の検索方法	4
2.2 内容に基づく静止画像検索の従来研究	5
2.3 内容に基づく動画画像の検索と本研究の方針	7
3. 画像系列の特徴づけ	9
3.1 KL 展開に基づく特徴抽出	9
3.2 固有空間の構成法	11
3.3 固有空間内での学習軌跡	15
4. 動画画像の検索手法	17
4.1 画像系列の内容の類似度	17
4.2 固有空間内の軌跡の類似度	19
5. 実験および結果	22
5.1 入力画像の準備	22
5.1.1 画像入力装置	22
5.1.2 画像の前処理	22
5.2 分節化された入力画像系列への適用	25
5.2.1 既知の動作を検索対象とした場合	25
5.2.2 未知の動作を検索対象とした場合	32
5.3 分節化されていない入力画像系列への応用	35
5.3.1 提案手法の応用方法	35
5.3.2 シミュレーション実験	38
6. むすび	45
謝辞	48

目 次

3.1	特徴軸の抽出	10
3.2	画像ベクトル	11
3.3	平均ベクトルの線形分解	14
3.4	(K+1)次元固有空間の構成	14
3.5	固有空間と学習軌跡	16
4.1	固有空間との距離	18
4.2	軌跡を構成する点の対応づけ	21
4.3	点の対応づけの手順	21
5.1	入力画像の前処理方法	23
5.2	8種類の学習画像系列	26
5.3	入力画像系列 s_1	27
5.4	入力画像系列 s_1 と固有空間との距離	27
5.5	固有空間 $\mathbf{E}^{(1)}$ 内での 入力画像系列 s_1 の軌跡 $\mathbf{t}_1^{(1)}$	30
5.6	固有空間 $\mathbf{E}^{(8)}$ 内での 入力画像系列 s_1 の軌跡 $\mathbf{t}_1^{(8)}$	30
5.7	固有空間 $\mathbf{E}^{(1)}$ 内での軌跡間の距離	31
5.8	固有空間 $\mathbf{E}^{(8)}$ 内での軌跡間の距離	33
5.9	固有空間 $\mathbf{E}^{(4)}$ 内での軌跡間の距離	34
5.10	入力画像系列の分割	37
5.11	時系列の検索結果	37
5.12	入力画像系列	39
5.13	入力軌跡 $\mathbf{t}_r^{(1)}$ と学習軌跡 $\mathbf{T}^{(1)}$ との対応点	44

表 目 次

5.1	入力画像系列と固有空間との距離の順位	28
5.2	固有空間内の軌跡の類似度評価に用いる閾値 (31 次元)	39
5.3	部分入力画像系列 L_r と固有空間との距離の順位	40
5.4	部分入力画像系列 L_r に対する検索結果	40
5.5	部分入力画像系列 L_5 の軌跡 $\mathbf{t}_5^{(1)}$ と学習軌跡 $\mathbf{T}^{(1)}$ の対応点 . . .	43
5.6	部分入力画像系列 L_6 の軌跡 $\mathbf{t}_6^{(1)}$ と学習軌跡 $\mathbf{T}^{(1)}$ の対応点 . . .	43
5.7	部分入力画像系列 L_7 の軌跡 $\mathbf{t}_7^{(1)}$ と学習軌跡 $\mathbf{T}^{(1)}$ の対応点 . . .	43
5.8	部分入力画像系列 L_8 の軌跡 $\mathbf{t}_8^{(1)}$ と学習軌跡 $\mathbf{T}^{(1)}$ の対応点 . . .	43

1. はじめに

近年，計算機の処理能力の向上，光ディスク等の安価かつ高速な記憶媒体の大容量化，データ通信の高速化といった背景に伴い，計算機を用いて静止画像や動画画像のデータを扱う機会が増大し，その蓄積量は膨大なものとなりつつある．このため，画像を対象とした入力処理技術，認識技術，蓄積技術，検索技術，出力処理技術は今後ますます重要になると思われる．中でも，膨大な量を有するデータベースの中から所望の静止画像や動画画像を効率的に検索する手法の必要性が高まるものと考えられる．

画像を検索する一般的な手法としては，各画像に対して画像番号や画像名，画像ファイル名などの識別子を与えておき，これを検索の手がかり（検索キー）とする方法がある [48]．しかし，この方法では画像を検索する際に，これらの画像識別子を覚えておく必要があり，大規模なデータベースからの検索を考えた場合には限界がある [47]．これに対して，画像自身を検索キーとした検索方法が考えられる．この方法は，検索キーである画像と類似する画像をデータベースの中から検索する方法であり，内容に基づく画像検索（Content-based Image Retrieval）と呼ばれる．静止画像の内容に基づいて画像を検索する手法としては，画像の色情報やテキスト情報をを用いるものなどが提案されている [7]．一方，動画画像は静止画像とは異なり，一連の動きがある意味を表現している．そのため，この動き情報が動画画像検索におけるキーとして重要であると考えられる．しかし，動画画像に含まれる動き情報を用いた検索手法はこれまでに確立されておらず，動き情報を含んだ動画画像をキーとして，類似した動画画像の検索が実現できれば，ジェスチャ認識などへの応用が考えられるため，重要な要素技術の研究として意義深い．

本論文では，内容に基づく動画画像検索手法の実現の基礎として，動画画像データベースの中から入力の画像系列に類似した系列を検索する手法を提案する．まず，動画画像の類似性について考察する．動画画像は静止画像の系列と考えることができる．そのため，ある二つの画像系列の類似性は，

- 1). 各画像系列間での個々の静止画像の類似性
- 2). 類似している静止画像の出現順序の類似性

に基づいて評価が可能であると考えられる。すなわち、類似した画像が同じような順序で出現した場合に、画像系列が類似するとみなす。この考えに基づき動画画像検索を実現するためには、まず、個々の画像系列がどのような特徴を有するか、すなわち、どのような画像がどういう順序で出現するかを表現することが必要となる。それらの特徴を利用することで、画像の内容とその出現の順序に基づく類似性評価が可能となる。

本研究では、個々の画像系列の特徴づけのために、Karhunen-Loève 展開 (KL 展開) に基づき固有空間を構成する [44][49]。このとき、時間的に隣接する画像の内容が類似し相関が高ければ、固有空間内で対応する位置も隣接する [8] [26] [27] [59]。そのため、一つの画像系列は固有空間内では、連続した一つの軌跡として表現される [33] [34]。これにより、固有空間内での軌跡の類似度を評価することで時系列の画像に含まれる運動の類似性を評価可能となる [5] [6]。

提案手法では、動画画像の検索は、あらかじめ固有空間が求められた画像系列を対象に、入力画像系列との類似性を画像の内容とその出現順序によって二段階で評価することによって行う。まず、画像内容の類似性を評価するために、CLAFIC (CLAss-Featuring Information Compression) 法 [58] に基づき、各固有空間と入力画像の距離を求める。その後、内容が類似していると判断された系列に対して、固有空間内での軌跡の類似性を評価する。

上述のアプローチを採用する提案手法の特徴としては、以下の点が挙げられる。

- 画像系列を構成する個々の画像を低次元の固有空間で表現でき、構成画像の内容と完全に一致する画像が学習したデータベース中に存在しなくても、画像の内容の類似性を評価できる。
- 画像系列毎に固有空間を個別に学習し、固有空間との距離を類似性の評価尺度として用いているため、すべての学習データを対象とした計算が不要であることから、学習する画像系列の追加が容易である。

本論文では、以下の各章において画像系列毎の固有空間を構成し、動画画像の軌跡を得る方法、固有空間との距離を用いた画像系列の内容の類似度を評価する方法、ならびに軌跡の類似度を評価する方法について述べる。最後に、ジェスチャ

を表現する実動画像を対象とした実験を通して、提案手法を用いた動画像の検索が可能であることを示す。

各章の概略は以下の通りである。第2章において関連研究について述べるとともに、本研究の位置づけ、ならびに方針について述べる。

さらに第3章において、個々の画像系列を特徴づける方法について述べる。具体的にKL展開を用いて、一つの画像系列から一つの固有空間を構成する方法について述べ、動画像の軌跡を得る方法について述べる。

第4章では、動画像の検索の枠組について述べる。まず、本研究における動画像検索について述べた後、画像系列の内容の類似度を評価する方法、ならびに固有空間内での軌跡の類似度を評価する方法について述べる。

第5章では、8種類のジェスチャを示す動画像を用いた実験および結果を示す。まず、入力画像系列が分節化されている場合の実験により、提案手法の有効性を確認する。その後、入力画像系列が分節化されていない場合への応用の可能性について実験により考察する。

最後に第6章で、結論および今後の展望について述べる。

2. 関連研究と本研究の位置づけ

本章では、画像検索の代表的な方法について述べ、本研究の位置づけを明らかにする。文字や数値などの記号情報を対象とした従来のデータベースは、記号情報を手がかりとした検索法により、比較的容易に検索を行なうことができる。これに対し、画像を対象としたデータベースの場合には、あらかじめ与えておいたキーワードによる検索のみでなく、画像処理やパターン認識技術により抽出された特徴を検索キーとするなどの柔軟な検索方法が求められている [56] [57].

以下では、まず画像データベースからの検索方法について整理し、内容に基づく画像検索の従来手法を概観する。次に、動画像検索の手法について考察し、本論文で提案する手法の概略を示す。

2.1 画像の検索方法

本節では、画像データベースから所望の画像を検索する方法について整理する。画像検索の初歩的な方法としては、各画像に対して与えられている文字や数値の付帯情報を索引として検索する方法が挙げられる。この方法は、画像データベース中に蓄積されている各画像に対して画像番号や画像名、画像ファイル名などの識別子を登録時に与えておき、その識別子を検索キーとして検索する方法である。そのため、従来のデータベース技術を適用することができ、一意に目的とする画像に到達することができる。この形態のシステムを利用したものとしては、例えば、医用画像、リモートセンシング画像などが挙げられる。医用画像の場合では、患者名、疾患の部位、病名、撮影年月日などが、また、リモートセンシング画像の場合では、日付、経度や緯度、気象条件、チャンネル番号などが索引情報とされる [48]. このように、検索の目的が明確なときは有効であるが、画像を検索する際に、これらの画像の索引やデータベースの構成方法を覚えておく必要がある。そのため、利用者が画像データベースの構成方法を知らない場合には、膨大な数の画像の中から所望の画像を効率的に検索するのは困難となる。

この方法に対して、画像の内容に基づいて検索する方法 (Content-based Image Retrieval) が提案されてきている [7]. この方法は、画像データ中に含まれている

意味的な情報（対象物の色や形，配置など）を取り出し，その情報が類似するものをデータベースの中から探す方法である．この場合，画像の識別子を検索キーとする検索に比べ，応用範囲は広く，多様な検索が実現できる．また，データベースの構成方法などをあらかじめ知っておく必要がないため，検索時における利用者の負担を軽減することができる．

2.2 内容に基づく静止画像検索の従来研究

本研究では，動画像は静止画像の時系列の集合と捉えるため，静止画像を対象とした検索手法は，動画像を検索するための要素技術として重要であると考えられる．そこで本節では，静止画像を対象とした内容に基づく画像検索の従来手法について概観する．

内容に基づく画像検索に対する従来の方法は，内容の抽出という観点から，画像の内容を表すキーワードに基づく検索方法，ならびに画像の特徴抽出を自動化する方法の二つのアプローチに大別できる．まず，前者の方法は，データベースの登録者によってあらかじめ各画像に付与されたキーワードを検索キーとして，通常のデータベース管理システムで取り扱う方法である [41]．実用段階にあるデータベースの多くは，このようなキーワードを検索キーとして利用している [16]．例えば，画像内の対象物体（人間，家，自動車など）間の空間的な位置関係などを用いて，「人間の左には家」，「家の上には曇」などのように画像をあらかじめモデル化しておき，検索時は，これらの対象物体と位置関係を記述した文字列の照合を行なう方法などがある [3]．キーワードに基づく検索方法は，利用者のもつイメージに含まれる情報のうち，全体的な情報（場面，テーマ，印象など）や画像内に登場する事物の種別（人間，犬など）を指定するのに向いている．しかしその反面，検索時に指定できるキーワードが画像のデータベースへの登録時に登録者が付与されたものに限定されるため，登録者の主観が検索結果に大きく依存する．また，データベースに新たな画像を登録する度に，各画像にキーワードを付与しなければならない．さらに，形状や位置関係，色彩などの視覚的な情報は自然言語で完全に表現できない多様性をもっていることから，キーワードのみでは多義性・曖昧性の高い画像情報の表現が十分にできないことが問題点として指摘

されている [16][17].

もう一つのアプローチは、登録者が画像の構成を解釈してその関係を入力するという前述のアプローチの限界を克服するために、画像処理やパターン認識技術を援用し、画像に含まれる特徴の抽出といった部分を自動化する方法である。まず、データベースの登録時に各画像から特徴量（形状、大きさ、配置、色、濃淡など）を自動的に抽出する。検索時には、利用者から入力された検索キーの特徴量を登録時と同じ方法で抽出し、その特徴量が類似する特徴量を有する画像をデータベースの中から取り出す方法である。しかしながら、画像中の特徴を自動的に抽出する方法は、高い計算コストが必要とされ、画像に含まれる意味的な情報を取り出したり、あるいは、その類似度を定義するのは容易ではない [7]。そのため、取り扱う領域を特化するなどして研究が進められている [9] [35]。

検索に利用される特徴量として、画像の中に含まれる色情報やテクスチャ情報などを用いて検索する手法が提案されている [7] [45]。この中でも、視覚的表現を検索キー（検索イメージ）に用いる、「画像により画像を検索する」手法が盛んに研究されている。検索イメージの表現によって、利用者のイメージのどの部分がどの程度所望の検索に反映されるかが決まるため、概略図やスケッチなどを検索キーに用いる [18] [24] [37]。これらの手法はいずれも、画像内の対象物体とそれらに対応する領域の色、形状や配置などを図式的に指定できるようにし、指定された色、配置などをデータベースにある画像と比較し、類似度を計算することで検索する手法である。この手法は、利用者に検索対象画像のイメージを直接的に表現させるもので、利用者からシステムへのイメージの的確な伝達に有効な手段である。しかし、画像の内容を特定しない場合、何をキーとして採用するかが難しいという問題がある。これに対して、統計的に画像中の特徴量を抽出し、特徴を表す空間（特徴空間）内での距離や類似度を利用方法がある。この方法は、画像を特定せずに機械的に適用することができるが、利用者が感じる主観的な画像間の類似度が反映されているとは限らない。この方法には、商標画像を対象に、画像が有する空間周波数特徴や濃淡の特徴を局所的なパターンの統計的な数値パラメータ列（部分画像の黒画素数を要素にもつベクトル）として表現するもの [12] や、顔画像を対象に KL 展開を用いたもの [38] などがある。さらに、KL 展開を

応用し，登録時に付与されたキーワードを特徴空間内で分離することで，特徴量の類似度評価を改良し，大規模なデータベースでの効率的な検索手法も提案されている [46].

2.3 内容に基づく動画像の検索と本研究の方針

前節では，静止画像を対象とした内容に基づく画像検索のさまざまな研究について述べた．近年の計算機の処理能力の向上により，動画像を扱う機会が増大しつつあるため，静止画像の場合と同様に動画像の内容に基づく検索手法についても必要性が高まると考えられる．しかし，動画像は静止画像に比べてデータ量が大きくなるため，動画像を大規模に蓄積してそれを検索しようとした試みは現在のところほとんどなく，実験的に検索の可能性を調べているのが現状である [19].

まず，内容に基づく動画像検索を実現するための要素技術として，動画像を構造化する技術が挙げられる．動画像をデータベースとして蓄積する際には，意味のある最小単位（ショット）で抽出し，構造化しなければならない．動画像を構造化する従来研究として，ショットの切れ目である「カット」に注目し，カットの変化点を自動的に検出，構造化する方法などが提案されている [31] [40] [52]．これらは，画素値の変化量などに注目しており，動画像の意味を十分に取り出せているとは限らないという問題点が残されている．

動画像がショットの単位で構造化されたとすると，一つのショットは一つの静止画像と対応し，前節の静止画像を対象とした内容に基づく画像検索手法を適用することができる．まず，キーワードに基づく検索方法としては，放送用の動画像を対象に記者がつけた注釈を手入力し，特徴量とすることを提案している [43]．一方，ショットに含まれる特徴を自動で抽出する方法としては，動画像中の文字や記号，色情報などを対象としている研究が挙げられる [15] [23] [32]．また，動画像の解析を行ない，撮影時のカメラの動きを求めることで，動画像を特徴づける方法も提案されている [50][51]．これらがうまく抽出ができれば，有効な検索キーとなり得る．しかし，いずれも十分な精度が得られているとはいえない．

そこで，本研究では，静止画像を対象とした内容に基づく画像検索で有効であった，画像を類似画像により検索する手法を動画像検索に対して適応することを提

案する。所望の「動き」の特徴（動画像）と類似する動画像をあらかじめ蓄積されたデータベースの中から探す。この実現のためには、動画像に含まれる「動き」特徴を抽出し、その特徴量の類似度の評価尺度を定義する必要がある。

本研究では、画像を特定せずに機械的に動画像に含まれる動き特徴を抽出するため、静止画像を対象とした場合と同様に統計量を用いる。具体的にはKL展開に基づく方法を用いる。KL展開は、Karhunen と Loève によって提案された手法であり、情報圧縮手法の一つである。KL展開に基づく方法はパターン認識の分野においても応用範囲が広く [21] [29] [42]，また、近年は顔画像認識などに適用され、高い認識率が報告されている [14] [28] [49]。

本研究では、あらかじめ動画像データベースに蓄積されている動画像（ここでは、学習画像系列）は分節化されているものとする。（ここで分節化された動画像とは、一つの意味を有する動きを表現する単位で切り出された画像の系列を指す。）個々の画像系列の「映っている事物」を特徴づけるために、画像系列毎にKL展開に基づく手法を適用することで、固有空間を構成する。このとき、時間的に隣接する画像の内容が類似し相関が高ければ、固有空間内で対応する位置も隣接する。そのため、一つの画像系列の「動き」は固有空間内では連続した一つの軌跡として表現される。これにより、固有空間内の軌跡の類似度を評価することで時系列の画像に含まれる運動の類似性を評価することができる。そのため、利用者によって動画像（ここでは、入力画像系列）が入力された際には、入力画像系列を固有空間に射影し、固有空間内での軌跡度を評価することで動画像検索を行なう。以下、3章から提案手法について述べる。

3. 画像系列の特徴づけ

本章では、個々の画像系列を特徴づける方法について述べる。画像系列は膨大な情報量を有するため、計算コストなどの観点から、少ない情報量で最も特徴を表現できる特徴パラメータを抽出することが望ましい。また、画像系列を構成する画像を特定せず、機械的に特徴パラメータを抽出できることが望ましい。この実現のために、本研究では個々の画像系列毎に KL 展開に基づいた特徴抽出方法を適用し、画像系列に含まれる「映っている事物」と「動き情報」を画像系列の固有空間内での軌跡として表現する。この結果得られた固有空間と軌跡が検索のための特徴として用いられる。

以下では、KL 展開に基づく特徴抽出の基本原理を説明するとともに、一つの画像系列から一つの固有空間を構成し、画像系列の固有空間内での軌跡を求める方法について述べる。

3.1 KL 展開に基づく特徴抽出

KL 展開は、Karhunen と Loève によって提案された手法であり、パターン認識における情報圧縮手法の一つである。この展開は、多種類のサンプルデータ間の相関関係を解析することにより、少ないパラメータで元のサンプルデータ全体を表現するものであり、この基本原理は主成分分析 [55] と等価である。ベクトルで表せば、各サンプルデータがベクトル空間内において位置ベクトルをもつものと考えられ、多数の座標軸で示されるサンプルデータの座標点を低次元の線形部分空間に射影することにより、情報を圧縮する。

具体的には、サンプルデータの平均座標点を基点として、最も分散の大きい方向を示すベクトル \mathbf{e} を固有値問題を解くことによって求める。2次元平面上にサンプルデータが散在している場合の例を図 3.1 に示す。求められたベクトル \mathbf{e} は、サンプルデータ集合に共通する主成分を表す特徴軸となる。このベクトルはサンプルデータの位置ベクトルの共分散行列の固有ベクトルであるため、これらのベクトルで張られる部分空間は固有空間と呼ばれる。一般に、この固有空間は元のベクトル空間に比べ低次元となるため、固有空間に射影されたデータとして元の

サンプルデータを低次元の情報で圧縮することができる。図 3.1 の場合は、 e によって形成される座標軸を用いて 2 次元情報を 1 次元に圧縮する。構成した固有空間の次元において、元のサンプルデータと再構成されたデータとの差の二乗和が最小であることが知られている [11].

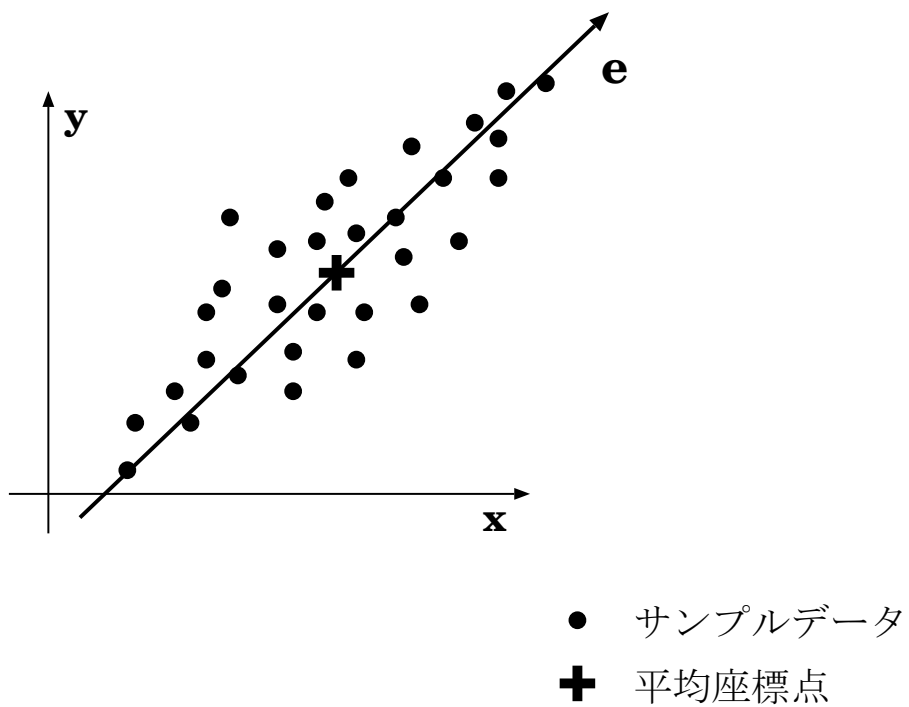


図 3.1 特徴軸の抽出

3.2 固有空間の構成法

本研究では、前節で述べた KL 展開に基づく特徴抽出法を画像系列毎に適用する。まず、一つの画像系列から一つの固有空間を求める方法について述べる。 P 枚の画像からなる動画データ中の画像系列 (学習画像系列) S を考える。画像系列を構成する各画像を、画素の輝度値を要素にもつ画像ベクトル Φ_j ($j = 1, 2, \dots, P$) とする。画像内の画素数を n とすれば、画像ベクトル Φ_j は以下のように、 n 次元空間上の位置ベクトルとして表現される (図 3.2)。以下、学習する画像系列および画像を大文字で表記し、入力する画像系列および画像を小文字で表記する.)

$$\Phi_j = (X_1, X_2, \dots, X_n)^t \quad (j = 1, 2, \dots, P) \quad (1)$$

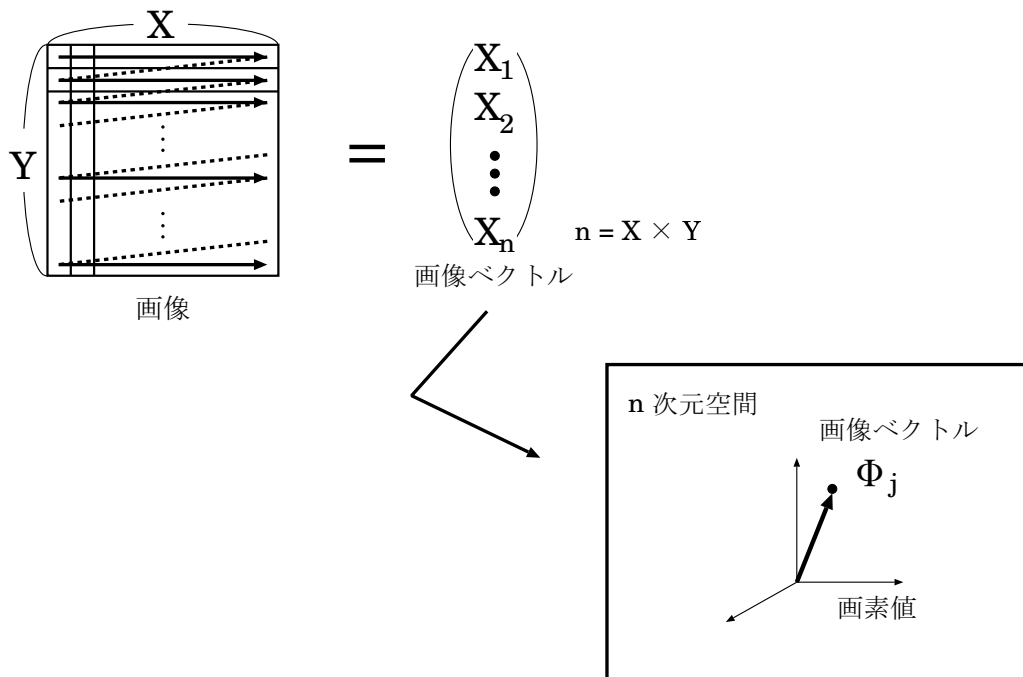


図 3.2 画像ベクトル

このとき、画像系列 $S = \{\Phi_j\}$ に対する固有空間 E は次のように求まる。ま

ず，画像ベクトルの集合 $\{\Phi_j\}$ の平均ベクトル $\bar{\Phi}$ を求める．

$$\bar{\Phi} = \frac{1}{P} \sum_{j=1}^P \Phi_j \quad (2)$$

次に，画像ベクトル集合 S の共分散行列 C を求め，これを固有値分解して固有値 λ_k ，ならびに固有ベクトル \mathbf{u}_k を求める．

$$C = \frac{1}{P} \sum_{j=1}^P (\Phi_j - \bar{\Phi})(\Phi_j - \bar{\Phi})^t \quad (3)$$

$$C\mathbf{u}_k = \lambda_k \mathbf{u}_k \quad (k = 1, 2, \dots, K, \dots, P) \quad (4)$$

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K \geq \dots \geq \lambda_P \quad (5)$$

固有空間 E は，式 4) より求まる固有ベクトル $\{\mathbf{u}_k\}$ を直交基底ベクトルとして構成される．

また， $\nu_{jk} = (\Phi_j - \bar{\Phi})^t \mathbf{u}_k$ とするとき，画像ベクトル Φ_j の KL 展開は，

$$\Phi_j = \bar{\Phi} + \sum_{k=1}^P \nu_{jk} \mathbf{u}_k \quad (6)$$

と表される．このとき，固有値が大きい固有ベクトルほど与えられた画像ベクトル集合を分離するのに有効なベクトルであり，画像ベクトルを表現するのに有用なベクトルである．そのため，固有値が大きい上位 K 個の固有ベクトルを用いて，

$$\tilde{\Phi}_j \approx \bar{\Phi} + \sum_{k=1}^K \nu_{jk} \mathbf{u}_k \quad (7)$$

と近似する．一般に， $\tilde{\Phi}_j$ と Φ_j との残差が K 次元の最小二乗近似の意味で最適であることが知られている．

式 7) は，平均ベクトル $\bar{\Phi}$ を原点とした K 次元の固有空間へ画像ベクトル Φ_j を射影することを意味する．しかし，本研究では，画像系列毎に固有空間を学習するため，各固有空間の原点が異なり，その原点を揃える必要がある．そのために，各固有空間の原点が元の空間の原点を通るようにするために，以下の処理を施す．

平均ベクトル $\bar{\Phi}$ を K 個の固有ベクトル $\{\mathbf{u}_k\}$ とそれに直交するベクトルの方向との線形和で表す (図 3.3)．さらに，その直交ベクトルの方向に画像ベクトル

を K 次元で近似した残差を含め、その直交ベクトルを $(K + 1)$ 次元目の固有ベクトルとする [53].

$$\begin{aligned}
\Phi_j &= \sum_{k=1}^K \nu_{jk} \mathbf{u}_k + \left[\bar{\Phi} + \sum_{k=K+1}^P \nu_{jk} \mathbf{u}_k \right] \\
&= \sum_{k=1}^K \nu_{jk} \mathbf{u}_k + \left[\sum_{k=1}^K \alpha_k \mathbf{u}_k + \sum_{k=K+1}^P \nu_{jk} \mathbf{u}_k \right] \\
&= \sum_{k=1}^K (\nu_{jk} + \alpha_k) \mathbf{u}_k + (\Phi_j^t \mathbf{e}_{K+1}) \mathbf{e}_{K+1}
\end{aligned} \tag{8}$$

ただし,

$$\mathbf{e}_{K+1} = \frac{\bar{\Phi} - \sum_{l=1}^K \alpha_l \mathbf{u}_l}{\left\| \bar{\Phi} - \sum_{l=1}^K \alpha_l \mathbf{u}_l \right\|} \tag{9}$$

$$\alpha_l = \bar{\Phi}^t \mathbf{u}_l \tag{10}$$

ここで,

$$\nu_{jk} + \alpha_k = \Phi_j^t \mathbf{u}_k \tag{11}$$

であるため、 $\mathbf{e}_k = \mathbf{u}_k$ ($k = 1, 2, \dots, K$) とおくと、式 8) は,

$$\Phi_j = \sum_{k=1}^{K+1} \rho_{jk} \mathbf{e}_k \tag{12}$$

$$\rho_{jk} = \Phi_j^t \mathbf{e}_k \tag{13}$$

と表せる。また、式 9) より、 \mathbf{e}_{K+1} は、 K 次元の固有空間に直交するベクトルであるため、画像ベクトル集合を表現する有効性が低いベクトルである。式 12) により、学習する固有空間は全て、原点を通ることになる (図 3.4)。

画像のサイズが 100×100 であるとしても、一枚の画像が有する画素数 n は、 $n = 10000$ となり、画像系列を構成する枚数 P に比べて一般に大きいといえる。さらに、 K は P より小さいため、3.1節で述べたように、学習画像系列を構成する個々の画像 Φ_j は n 次元の空間に比べ、かなり低い $(K + 1)$ 次元の固有空間内に射影された点として表現可能となる。

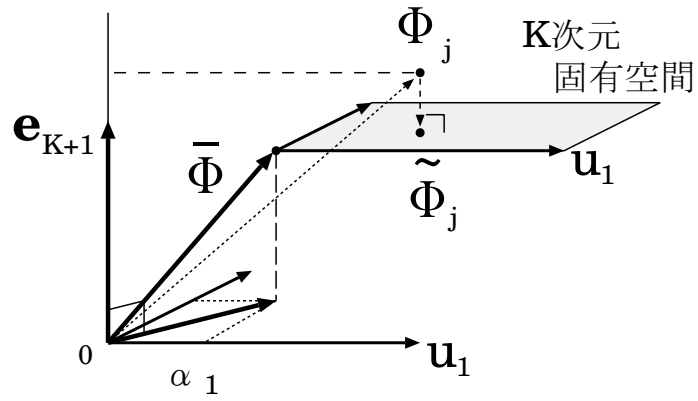


図 3.3 平均ベクトルの線形分解

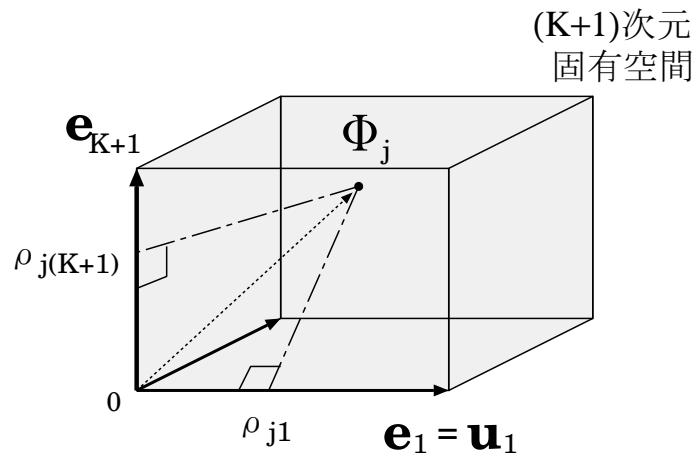


図 3.4 (K+1)次元固有空間の構成

3.3 固有空間内での学習軌跡

元の空間内で隣接するデータは、固有空間内での対応する位置も隣接するという関係がある [29]. そのため、時間的に隣接する画像の内容が類似し相関が高ければ、固有空間内で対応する位置も隣接する [30]. したがって、画像系列を画像の集合とみなせば、固有空間内で対応する点列 (軌跡) となり、その軌跡を画像の動きを表す情報として抽出することができる.

式 (2) は、画像ベクトル Φ_j を $(K+1)$ 個のベクトル $\{e_k\}$ で張られる固有空間へ射影することを意味する. $\{e_k\}$ をノルム 1 の正規直交基底にすれば、 Φ_j は、固有空間 E での $(K+1)$ 次元の位置ベクトル $\Psi_j = (\rho_{j1}, \rho_{j2}, \dots, \rho_{jK}, \rho_{j(K+1)})^t$ へ射影される. ここで、あらかじめ画像ベクトル集合 $\{\Phi_j\}$ のノルムを正規化しておけば、画像間の相関が高いほど固有空間内で対応する点間の距離が小さくなるという関係が成り立つ. つまり、画像系列を構成する画像が連続的に変化すると仮定すれば、画像系列を固有空間内に射影した点は連続した軌跡を描くことになる.

画像系列を構成する画像 Φ_j をすべて固有空間 E に射影することにより、 $(K+1)$ 次元のベクトル Ψ_j の系列、すなわち、固有空間内の点列 $T = \{\Psi_1, \Psi_2, \dots, \Psi_P\}$ として軌跡 (以下、学習軌跡) が得られる.

以上の手順を各画像系列に施し、 P 枚の学習画像系列から固有空間と学習軌跡を求める. 学習する画像系列 $S^{(i)} = \{\Phi_j^{(i)}\}$ ($j = 1, 2, \dots, K, \dots, P; i = 1, 2, \dots, Q$) の数を Q とすると、 Q 個の $(K+1)$ 次元の固有空間 $E^{(i)}$ が構成され、固有空間 $E^{(i)}$ 内の学習軌跡 $T^{(i)}$ が得られることになる (図 3.5).

これにより、以下の二つの特徴量を検索のために用いる.

- | | |
|---|--|
| { | <ul style="list-style-type: none"> ● 固有空間 |
| | n 次元の固有ベクトル $\times (K+1) \times Q$ |
| | <ul style="list-style-type: none"> ● 学習軌跡 |
| | $(K+1)$ 次元の位置ベクトル $\times P \times Q$ |

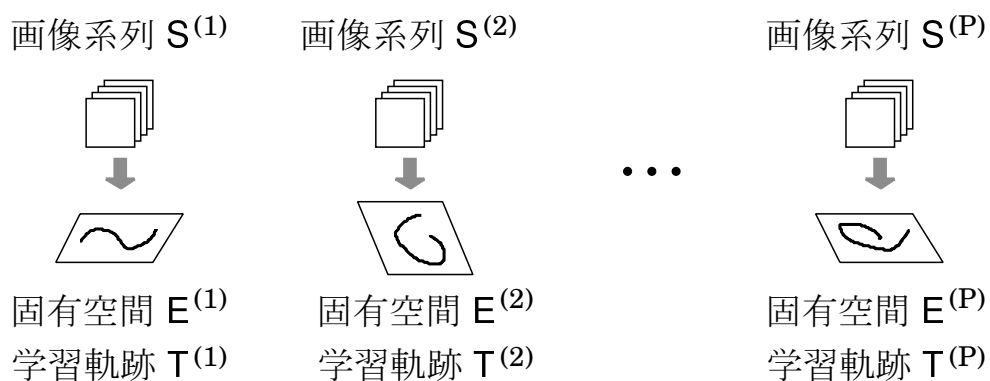


図 3.5 固有空間と学習軌跡

以上のように、動画像データベースの中にある学習画像系列からあらかじめ固有空間と学習軌跡を求めておく。その後、検索対象の画像系列が入力された場合、固有空間に射影し軌跡（入力軌跡）を得る。この入力軌跡と学習軌跡は比較することができるため、類似度を評価することで検索することができる。以下、第4章で動画像データベースの中から動画像を検索するための特徴量の類似度を評価する手法について述べる。

4. 動画像の検索手法

本章では、動画像データベースの中から所望の動画像を検索する手法について述べる。前章で述べた手法により抽出された特徴の類似性を評価する尺度を定義し、類似性を判断する基準について述べる。

本研究では、動画像データベース中からの動画像の検索とは、以下の条件を満たす画像系列を探すことをいう。

- 1). 画像系列を構成する個々の画像の内容が類似
- 2). 時系列における画像の出現順序が類似

本論文では、画像系列を構成する画像の類似性を評価する尺度として、入力画像系列と学習した固有空間との距離を用いる。固有空間との距離が小さいということは入力画像系列を構成する画像が学習した画像と内容が類似するということを意味している。しかしながら、個々の画像が類似しているだけでは画像系列に含まれる運動が類似するとは必ずしもいえない。例えば、全く同じ画像からなる系列であっても、画像に含まれる動作が逆の方向の場合、同一の画像系列であるとはいえない。そこで、固有空間との距離が最も小さいと判断された固有空間内において、さらに固有空間内での軌跡の類似度を計算し、画像の出現順序の類似性を判定することとする。以下では、画像系列の内容、ならびに固有空間内の軌跡の類似性を評価する尺度の定義とその判断基準について述べる。

4.1 画像系列の内容の類似度

本節では、画像系列を構成する個々の画像の内容の類似性を評価する尺度と判断基準について述べる。本論文では、この評価尺度として、検索のために入力された画像系列と学習した固有空間との距離を用いる。

入力された画像系列を構成する個々の画像と、学習された固有空間の距離を求めるためには、部分空間法の一つである CLAFIC 法 [58] を用いる。CLAFIC 法では、まず、各カテゴリの学習パターン群から、その分布を近似するために最適な正規直交基底ベクトルを求める。その後、入力パターンをそのベクトルで張ら

れる各部分空間へ射影し，射影成分を求め，この入力パターンの射影成分が最大となる部分空間を検出する．検出された部分空間を構成したカテゴリに含まれる成分と入力パターンに含まれる成分が類似するとして，入力パターンはこのカテゴリに属すると決定する．この部分空間が本研究における固有空間に対応する．固有空間への射影成分が最大となるのは，固有空間との距離が最小となるのと等価であり，ここでは，入力画像系列と固有空間との距離が最小となる学習画像系列を検出する [10]．

本手法では， q 枚の画像 ϕ_m からなる一つの入力画像系列 $s = \{\phi_m\}$ ($m = 1, 2, \dots, p$) ($p \leq P$) と i 番目の固有空間 $E^{(i)}$ との距離 Δ_i を， ϕ_m とそれを固有空間に射影した $\psi_m^{(i)}$ との距離の総和とする (図 4.1) ．

$$\Delta_i = \sum_{m=1}^p \|\phi_m - \psi_m^{(i)}\| = \sum_{m=1}^p \delta_m^{(i)} \quad (14)$$

この距離 Δ_i が最小となる固有空間 $E^{(i)}$ を検出し， $E^{(i)}$ を構成した学習画像系列 $S^{(i)}$ と入力画像系列 s の内容が類似すると判断する．

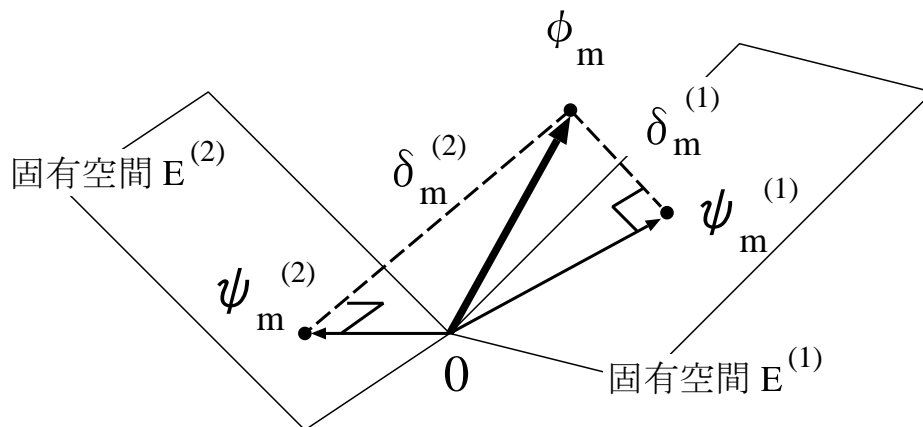


図 4.1 固有空間との距離

4.2 固有空間内の軌跡の類似度

前節の処理によって求まる固有空間において、固有空間内の軌跡の類似度を評価する。固有空間内の軌跡は、画像を表す順序づけられた離散的な点の集合と捉えることができる。軌跡の類似性を厳密に評価するためには補間などの処理が必要となると考えられるが、ここでは簡単のため、軌跡の類似度を軌跡を構成する点の対応づけで近似することとする。軌跡を構成する点の対応づけは、順序を考慮した多対多の対応づけ問題に帰着でき、順序を守りつつ対応を求める必要がある。また、入力画像系列が学習画像系列より短い場合にも対応できるように考慮する必要がある。

本論文では、点の対応付けの尺度として2点間の距離を用い、以下の方法で対応点を決定する。固有空間 $\mathbf{E}^{(i)}$ 内の2点 $\psi_a^{(i)}$ と $\psi_b^{(i)}$ 間の距離にはユークリッド距離 $d_{a,b}^{(i)}$ を用いる。

$$d_{a,b}^{(i)} = \left\| \psi_a^{(i)} - \psi_b^{(i)} \right\| = \sqrt{\sum_{k=1}^{K+1} (\rho_{ak}^{(i)} - \rho_{bk}^{(i)})^2} \quad (15)$$

4.1節の処理により、入力画像系列 s との距離が最小となる固有空間 $\mathbf{E}^{(i)}$ における学習軌跡を $\mathbf{T}^{(i)} = \{\Psi_j^{(i)}\}$ ($j = 1, 2, \dots, P$) とする。また、 $\mathbf{E}^{(i)}$ に射影された入力画像系列 $s = \{\phi_m\}$ の軌跡を $\mathbf{t}^{(i)} = \{\psi_m^{(i)}\}$ ($m = 1, 2, \dots, p$) ($p \leq P$) とする。このとき、 $\mathbf{t}^{(i)}$ と $\mathbf{T}^{(i)}$ との間における対応づけは、順序関係を保存させるため、 $\mathbf{t}^{(i)}$ を階層的に細分化することでなされる (図 4.2)。詳細な手順を図 4.3 に示す。手順の概略を以下に示す。まず、入力軌跡を構成する点のうち、両端の点とその真中の点を学習軌跡に対応づける。次に、対応づけられた2点の間に対し、同様に対応づける。これを再帰的に繰り返し、入力軌跡のすべての点を学習軌跡に対応づける。

図 4.3の手順により、固有空間 $\mathbf{E}^{(i)}$ 内の学習軌跡 $\mathbf{T}^{(i)}$ と検索対象の入力画像系列に対応する入力軌跡 $\mathbf{t}^{(i)}$ の距離 D を求め、この距離 D が閾値以下であるとき、画像が出現する順序も類似しているとみなし、学習画像系列 $S^{(i)}$ を検索結果とする。なお、 D が閾値を越える場合には、入力画像系列と類似する動画はデータベース中に存在しないものと判断される。

本章では、2つの類似度を評価する尺度を定義し、類似性を判断する基準について述べた。これにより、本研究における動画像の検索は以下ようになる。画像系列が入力されると、まず、あらかじめ学習した各固有空間との距離を計算する。次に、この距離が最小となった固有空間 $\mathbf{E}^{(i)}$ 内において、入力画像系列に対応する軌跡と学習軌跡の距離 D を計算する。距離 D が閾値以下とされた場合は、動画像データベースの中から学習画像系列 $S^{(i)}$ を出力する。それ以外の場合は、データベースの中には入力の画像系列に類似する学習画像系列が存在しないとして、未知の画像系列を示す $S^{(0)}$ を出力する。次章では、本手法の有効性を示すために、実画像を用いた実験を行ない、その結果について述べる。

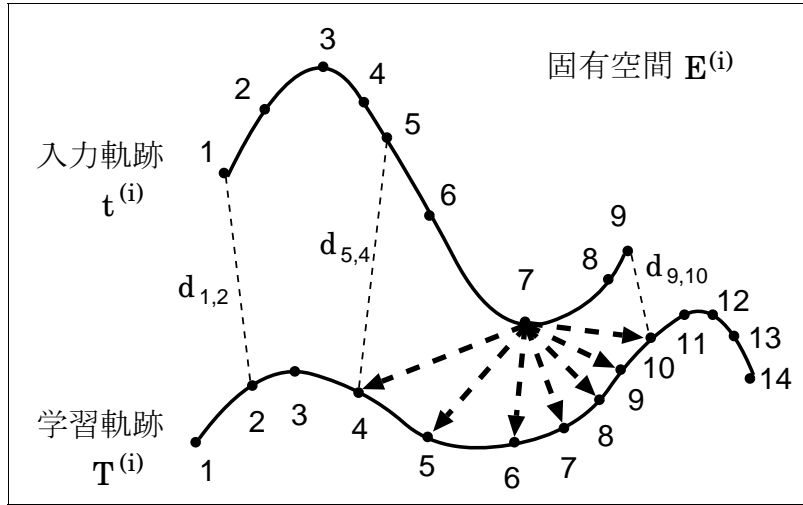


図 4.2 軌跡を構成する点の対応づけ

[step. 1] 初期設定)

$m = \frac{p}{2}, 1, p$ の順番で, $\psi_m^{(i)}$ からの距離が最小となる $\Psi_j^{(i)}$ ($1 \leq j \leq M$) を求め, $\psi_m^{(i)}$ と $\Psi_j^{(i)}$ を対応づける.

[step. 2] $\psi_{m_1}^{(i)}$ が $\Psi_{j_1}^{(i)}$ に, $\psi_{m_2}^{(i)}$ が $\Psi_{j_2}^{(i)}$ に対応づけられていたとする.

$\psi_m^{(i)}$ ($m = \frac{m_1 + m_2}{2}$) からの距離が最小となる $\Psi_j^{(i)}$ ($j_1 \leq j \leq j_2$) を求め, $\psi_m^{(i)}$ と $\Psi_j^{(i)}$ を対応づける.

[step. 3] すべての $\psi_m^{(i)}$ が $\Psi_j^{(i)}$ に対応するまで [step. 2] を繰り返す.

[step. 4] 対応づけられた 2 点 $\psi_m^{(i)}$ と $\Psi_j^{(i)}$ 間の距離の総和 $D(\mathbf{T}^{(i)}, \mathbf{t}^{(i)})$ を求める.

$$D(\mathbf{T}^{(i)}, \mathbf{t}^{(i)}) = \sum_{m=1}^p d_{corr[m],m} \quad (16)$$

ただし, $j = corr[m]$ は, $\psi_m^{(i)}$ に対応する点が $\Psi_j^{(i)}$ であることを表す.

図 4.3 点の対応づけの手順

5. 実験および結果

本章では、前述の個々の画像系列を特徴づける方法および検索手法を用いて動画像データベースから所望の画像系列を検索する実験を行ない、その結果について述べ、提案手法の有効性を検証する。

実験のために、学習画像系列として8種類のジェスチャの動画像を用意した。まず、一つの入力画像系列に含まれる動作が一つの意味をもつように分節化されている場合に対して本手法を適用した実験とその結果について述べる。ここでは、入力されたジェスチャが、既知の動作としてデータベースに存在する場合と存在しない場合について実験し、固有空間内の軌跡の類似度評価における閾値の設定が可能であるかについて考察する。次に、より柔軟な検索キーの入力のために、入力画像系列が分節化されていない場合に対する提案手法の拡張の可能性について述べる。

5.1 入力画像の準備

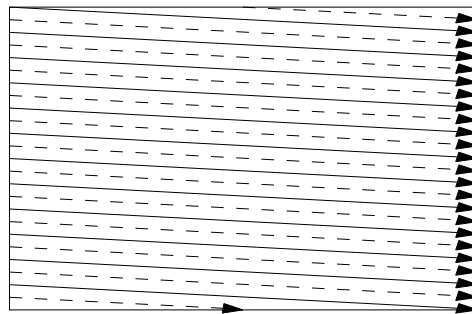
5.1.1 画像入力装置

実験に用いた画像の撮影には、ビデオカメラ (SONY Handycam PRO 41 万画素の CCD × 3, Hi8 方式) を使用した。また、撮像した画像を電子ファイルとして取り込むのには、SGI 製の Sirius Video 装置を利用した。本装置は 30 フレーム / 秒または、60 フィールド / 秒の速度で動画像をメモリに格納することができ、格納した画像をハードディスクに連続して記録することができる。

5.1.2 画像の前処理

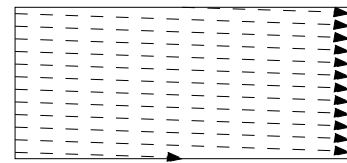
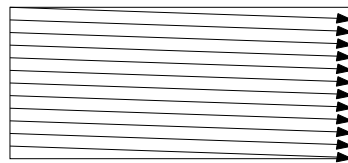
画像入力に使用した Sirius Video 装置のアナログ・ビデオ・インタフェースは、1 フレーム内に時系列の異なる 2 枚の画像が含まれるインタレース信号である。そこで、実験で使用する画像が同一時刻に撮像された画素のみで構成されるようにするために、図 5.1 に示すようなフィールド単位の画像分割と、走査線方向に 1/2 に縮小する処理を行なった。

カメラで撮影された画像
(インタレース)



分割

Sirius Video で取り
込んだ分割画像



横幅 $\times 1/2$

実験で使
用した画像

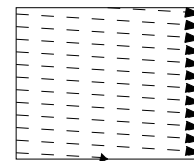
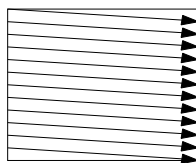


図 5.1 入力画像の前処理方法

このようにして得られた RGB 表色系のカラー画像を，式 (17) によりグレイスケール 256 階調 (画素値:Z) の濃淡画像 (PGM 形式: Portable Graymap File Format) に変換する.

$$Z = 0.299 \times R + 0.586 \times G + 0.114 \times B \quad (17)$$

以上の処理により，フィールド毎に分割後作成される 2 種類の画像列が得られるが，そのうち統一してどちらか一方を選んで実験に使用した.

5.2 分節化された入力画像系列への適用

まず、入力画像系列が分節化されている場合に対する提案手法の有効性を確認する。以下に、入力画像系列に含まれる動作と同じ動作があらかじめ構成された動画データベースに存在する場合と存在しない場合についての実験について述べる。これにより、固有空間内の軌跡の類似性評価における閾値が設定できるかどうかについて考察する。

5.2.1 既知の動作を検索対象とした場合

入力画像系列と同じ動作を表す画像系列がデータベースに存在する場合を考える。学習画像系列 $S^{(i)} = \{\Phi_j^{(i)}\}$ ($i = 1, 2, \dots, 8; j = 1, 2, \dots, 60$) の一部を図 5.2 に示す。 $S^{(i)}$ はそれぞれ大きさ 360×238 、256 階調の濃淡画像 60 枚で構成されている。各画像系列は、カメラ固定、照明条件一定のもとで背景は一定として撮像されたものである。また、一つの画像系列は同一人物が椅子に座りながら一つのジェスチャをしている動画画像である。

まず、前述の方法により、学習画像系列それぞれに対して、固有空間 $\mathbf{E}^{(i)}$ と学習軌跡 $\mathbf{T}^{(i)}$ を得る ($i = 1, 2, \dots, 8$)。ここでは、式 (12) における K を 30 とし、固有空間は 31 個の固有ベクトルで張られる空間とした。

さらに、入力画像系列として、それぞれ

$s_1 \cdots$ 戻る」	$s_2 \cdots$ いいえ」
$s_3 \cdots$ 「バイバイ」	$s_4 \cdots$ ぼんざい」
$s_5 \cdots$ 右へ」	$s_6 \cdots$ 左へ」
$s_7 \cdots$ じっちへこい」	$s_8 \cdots$ パチパチ」

のジェスチャを意味する、学習画像系列とは別の画像系列 s_l ($l = 1, 2, \dots, 8$) を用意して実験を行った。また、 s_l は、学習画像系列 $S^{(i)}$ と同一の条件 (人物も同一) 下で $S^{(i)}$ と別に撮像した動画画像であり、 $S^{(i)}$ と同様に 60 枚からなる画像系列である。

まず、戻る」のジェスチャをした画像系列 s_1 (一部を図 5.3 に示す) を入力とした場合の実験結果を以下に示す。



(a) $S^{(1)}$ [まる]

(b) $S^{(2)}$ [いいえ]



(c) $S^{(3)}$ [バイバイ]



(d) $S^{(4)}$ [ばんざい]



(e) $S^{(5)}$ [右へ]



(f) $S^{(6)}$ [左へ]



(g) $S^{(7)}$ [こっちへこい]



(h) $S^{(8)}$ [パチパチ]

図 5.2 8種類の学習画像系列



図 5.3 入力画像系列 s_1

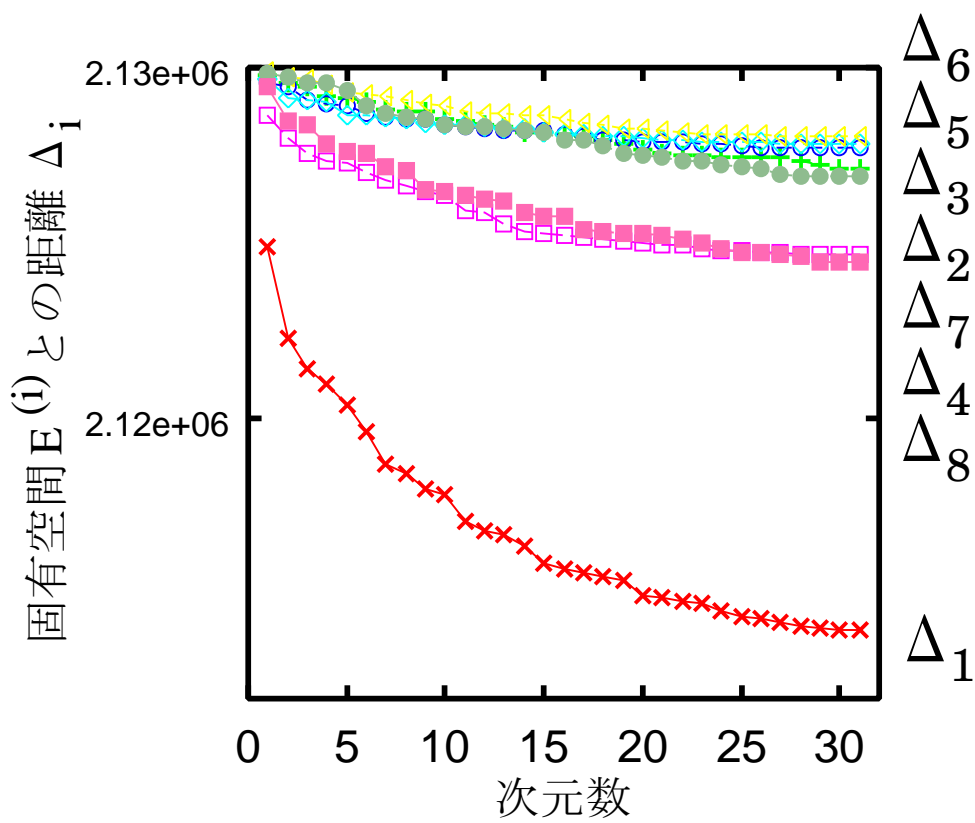


図 5.4 入力画像系列 s_1 と固有空間との距離

表 5.1 入力画像系列と固有空間との距離の順位

		固有空間							
		$\mathbf{E}^{(1)}$	$\mathbf{E}^{(2)}$	$\mathbf{E}^{(3)}$	$\mathbf{E}^{(4)}$	$\mathbf{E}^{(5)}$	$\mathbf{E}^{(6)}$	$\mathbf{E}^{(7)}$	$\mathbf{E}^{(8)}$
入 力 画 像 系 列	s_1	1	5	6	3	7	8	4	2
	s_2	3	1	5	2	7	8	4	6
	s_3	2	5	1	3	7	8	6	4
	s_4	2	6	5	1	8	7	4	3
	s_5	3	5	7	2	1	8	6	4
	s_6	2	4	7	3	8	1	6	5
	s_7	3	4	7	5	6	8	1	2
	s_8	2	5	6	3	8	7	4	1

入力画像系列 s_1 と固有空間 $\mathbf{E}^{(i)}$ との距離の計算結果を図 5.4 に示す。図 5.4 の横軸は式 (12) の k であり、各画像を近似するのに用いる固有ベクトルの数である。縦軸は、 s_1 と固有空間 $\mathbf{E}^{(i)}$ との距離 Δ_i を示している。距離が小さい線ほど、入力画像系列と学習画像系列の内容が類似することを意味する。図 5.4 より、固有空間 $\mathbf{E}^{(1)}$ との距離が最小であることが分かる。そのため、入力画像系列 s_1 が固有空間 $\mathbf{E}^{(1)}$ を構成した学習画像系列 $S^{(1)}$ (「まる」) の内容と類似すると判断される。

同様に、すべての入力画像系列 s_l と固有空間 $\mathbf{E}^{(i)}$ との距離の計算結果を表 5.1 に示す。表 5.1 は、31 次元における距離の値を小さいものから昇順に順位づけをした結果である。表 5.1 より、 s_1 の場合と同様に、入力画像系列と同じジェスチャを表す画像系列から構成された固有空間との距離が最小となることが分かる。このことより、8 種類の入力画像系列は、それぞれ同じジェスチャをした学習画像系列の内容と最も類似すると判定される。

次に、固有空間内での軌跡の類似度を評価する。入力画像系列 s_1 との距離が最小となる固有空間 $\mathbf{E}^{(1)}$ 内の軌跡 $t_1^{(1)}$ の様子を図 5.5 に示す。図 5.5 のそれぞ

れの軸は固有空間 $\mathbf{E}^{(1)}$ を構成する固有ベクトルである。固有空間は、実際には $(K + 1)$ 次元空間 ($K = 30$) であるが、表示の都合上固有値が大きい 3 個の固有ベクトル ($\mathbf{e}_1 \sim \mathbf{e}_3$) で張られる空間を表示してある。

ここで、他の固有空間での軌跡の様子を比較することにする。表 5.1 より、入力画像系列 s_1 と固有空間との距離が 2 番目に近いと判断された $\mathbf{E}^{(8)}$ 内での軌跡 $\mathbf{t}_1^{(8)}$ を例として、図 5.6 に示す。固有空間 $\mathbf{E}^{(8)}$ は学習画像系列 $S^{(8)}$ (「パチパチ」) から構成されたものである。図 5.6 と比較しても、固有空間 $\mathbf{E}^{(1)}$ 内での学習軌跡 $\mathbf{T}^{(1)}$ と入力の軌跡 $\mathbf{t}_1^{(1)}$ が類似していることが分かる。

各固有空間内の定量的な軌跡の類似度評価の有効性を示す。画像系列 s_1 から s_8 までをそれぞれ、固有空間 $\mathbf{E}^{(1)}$ 内に射影し軌跡 $\mathbf{t}_1^{(1)}$ から $\mathbf{t}_8^{(1)}$ とした。学習軌跡との類似度を評価した結果を図 5.7 に示す。 s_l の軌跡と学習軌跡 $\mathbf{T}^{(1)}$ との距離を $D_l^{(1)}$ とする。図 5.7 の横軸は、点の対応づけの評価基準であるユークリッド距離を計算するのに用いた次元数である。用いる次元が異なると、対応づけが異なる場合がある。縦軸は、固有空間 $\mathbf{E}^{(1)}$ 内の学習軌跡 $\mathbf{T}^{(1)}$ と入力画像系列の軌跡 $\mathbf{t}_l^{(1)}$ との距離である。図 5.7 より、学習軌跡 $\mathbf{T}^{(1)}$ と入力画像系列 s_1 の軌跡 $\mathbf{t}_1^{(1)}$ の類似度が一番高くなっていることが分かる。これは、固有空間 $\mathbf{E}^{(1)}$ を構成した学習画像系列と同じ動作を表現する画像系列の類似度が一番高いことを意味している。これより、本手法の軌跡の類似度評価が有効であることが分かる。また、軌跡の類似度を表す距離が、ある閾値以下であるとできれば、軌跡の類似性を評価できることが分かる。

以上の結果より、入力画像系列 s_1 は固有空間 $\mathbf{E}^{(1)}$ との距離が最小であり、かつ、軌跡の類似度が閾値以下であれば、学習画像系列 $S^{(1)}$ すなわち、「まる」のジェスチャをしている画像系列であると検索することができる。

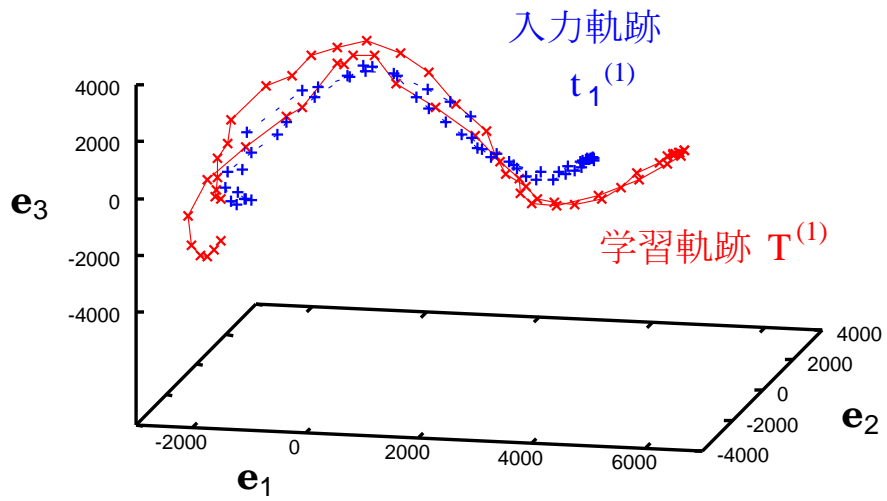


図 5.5 固有空間 $\mathbf{E}^{(1)}$ 内での 入力画像系列 s_1 の軌跡 $t_1^{(1)}$

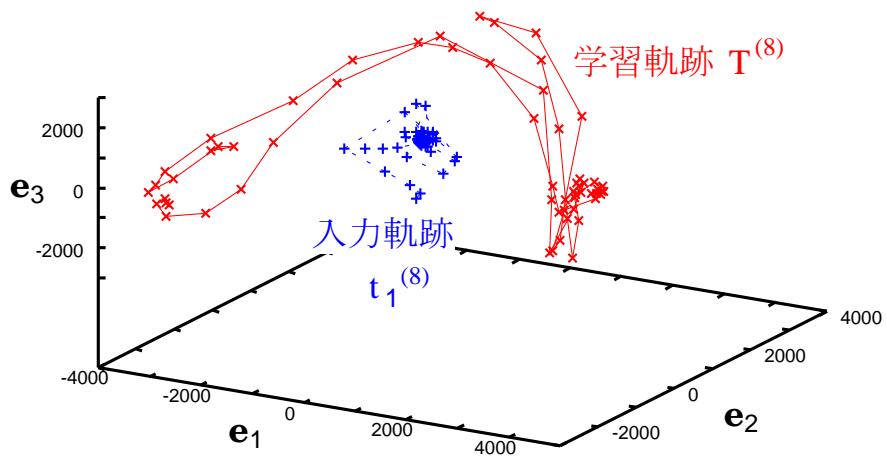


図 5.6 固有空間 $\mathbf{E}^{(8)}$ 内での 入力画像系列 s_1 の軌跡 $t_1^{(8)}$

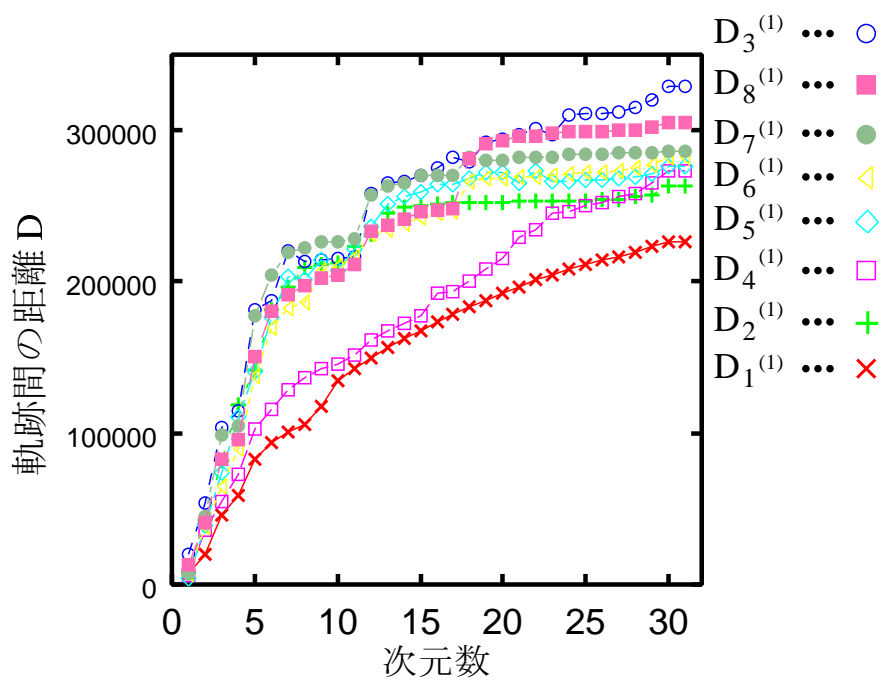


図 5.7 固有空間 $\mathbf{E}^{(1)}$ 内での軌跡間の距離

5.2.2 未知の動作を検索対象とした場合

次に、入力画像系列が表す動作の種類が学習の画像系列が表す動作の中に存在しない場合を考える。ここでは、上述の8種類の学習画像系列のうち1種類だけデータベースに存在しないものとし、存在しない種類のジェスチャを入力とする。例として、データベース中に学習画像系列 $S^{(1)}$ が無いものとし、画像系列 s_1 を入力とする場合を考える。

まず、 s_1 と固有空間との距離を考える。固有空間 $\mathbf{E}^{(1)}$ が存在しないことになるため、表 5.1より、固有空間との距離が最小となるのは、固有空間 $\mathbf{E}^{(8)}$ となる。このとき、固有空間 $\mathbf{E}^{(8)}$ 内の軌跡の類似度を評価した結果を図 5.8に示す。 $S^{(8)}$ に類似する s_8 の固有空間内の軌跡の類似度を比較する。固有空間 $\mathbf{E}^{(8)}$ 内の s_1 , s_8 の軌跡と学習軌跡 $\mathbf{T}^{(8)}$ との距離をそれぞれ、 $D_1^{(8)}$, $D_8^{(8)}$ とする。図 5.8より、 $D_1^{(8)}$ は $D_8^{(8)}$ に比べ全体的に大きい値となっている。固有空間 $\mathbf{E}^{(8)}$ 内では s_8 の軌跡が学習軌跡と類似すると判断できなければならず、 $D_8^{(8)}$ は閾値以下である。これに対し、 $D_1^{(8)}$ は閾値以上であるときに、学習軌跡と類似しないとすることができる。そのため、軌跡の類似度を表す距離 $D_1^{(8)}$ と $D_8^{(8)}$ を分割できるような閾値を設定することができれば、入力画像系列が学習データベースには存在しない未知の動作を示す画像系列であると判断することができる。

同様に、 $S^{(2)} \sim S^{(8)}$ の一つを学習画像系列に存在しないとして、それぞれ $s_2 \sim s_8$ を入力とする場合を考える。このとき、それぞれ固有空間 $\mathbf{E}^{(2)} \sim \mathbf{E}^{(8)}$ が存在しないこととなり、距離が2番目である固有空間が最も近い固有空間となる。 $s_3 \sim s_8$ を入力とした場合、 s_1 の場合と同様に、固有空間内の軌跡の距離の値が、正解と判断すべき場合と比べて大きいことが確認できた。しかし、 s_2 を入力とした場合、図 5.9に示すように、評価に用いる次元数が小さいときには、正しい結果が得られたが、次元数が大きいとき、固有空間内の軌跡の距離の値が、正解と判断すべき s_4 の場合より小さくなった。これは、軌跡の距離を求める点の対応づけの際に、固有値が小さい固有ベクトルの方向の距離が大きく影響しているためであると考えられる。そのため、軌跡の距離評価を固有値の大きい固有ベクトル、すなわち、軌跡の特徴をよく表現する軸の方向に対して大きな重みをかけるなど、軌跡の類似度の評価法を改善する必要がある。

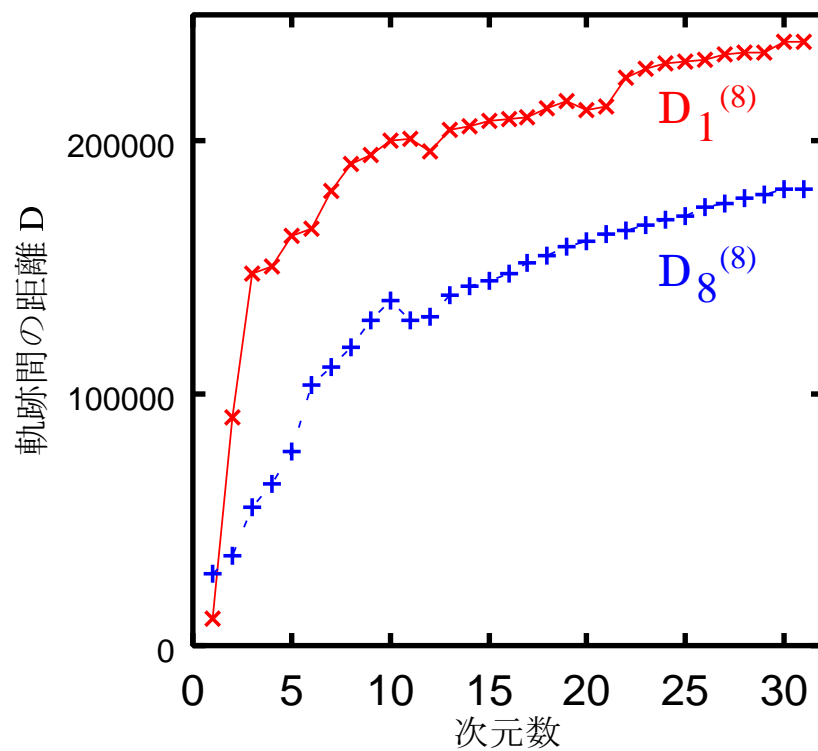


図 5.8 固有空間 $\mathbf{E}^{(8)}$ 内での軌跡間の距離

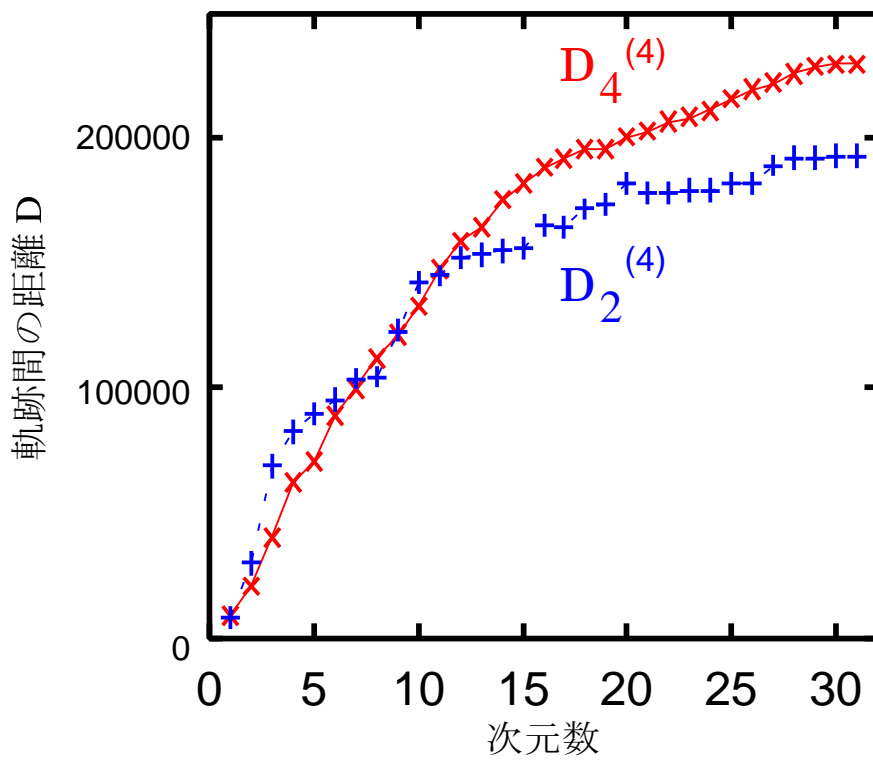


図 5.9 固有空間 $\mathbf{E}^{(4)}$ 内での軌跡間の距離

5.3 分節化されていない入力画像系列への応用

以上では、一つの入力画像系列が一つの意味のある動きを表現するように、分節化されている場合について考察したが、検索キーとなる入力画像系列は、常に分節化されているとは限らない。そのため、一般的な検索手法の実現のためには、入力画像系列が分節化されていなくても、含まれる動き情報を取り出し、柔軟に検索のための入力とできることが望まれる。

そこで本章では、入力画像系列が明確に分節化されていない場合に対する本手法の応用について考察する。例えば、人間が「歩き」という動作を認識しようと思えば、ある程度の時間、蓄積された情報量が必要であると考えられる。そのためここでは、まず、ある程度の時間に連続した画像系列を一つの入力として順次検索を実行する。これにより得られる時系列の検索結果から入力の動き情報を推定し、同時に動画データベースの中から類似する画像系列を検索結果とする方法を考える。以下に、分節化されていない画像系列を入力とした場合の本手法の応用方法ならびに、そのシミュレーション実験について述べる。

5.3.1 提案手法の応用方法

本節では、分節化されていない画像系列を入力とした場合への提案手法の応用方法について述べる。学習画像系列 $S^{(i)}$ ($i = 1, 2, \dots, Q$) から動画データベースを第3章で述べた方法を用いて構成する。ここでは、一般性を失わずに、入力画像系列 s は時間軸方向に長く、その s の中のある場所に入力である一連の意味を表現する動きが含まれるとする。

まず、入力画像系列 s を時間軸に沿って B 枚毎に分割し、それぞれ B 枚から構成される部分入力画像系列 L_τ ($\tau = 1, 2, \dots, T$) とする (図5.10)。ただし、 B は学習画像系列を構成する枚数 P より少ないとする。

ここで、分割された部分入力画像系列 L_τ それぞれに対して提案手法を適用する。一つの L_τ に対する結果として、学習画像系列 $S^{(i)}$ のどれであるか、あるいは未知の画像系列 $S^{(0)}$ であるかが検索結果 $R_\tau = \{S^{(r)} | r = 0, 1, 2, \dots, Q\}$ として出力される。そのため、すべての部分入力画像系列 L_τ を入力として処理を実行すると、時系列の検索結果 R_τ ($\tau = 1, 2, \dots, T$) を得ることができる (図5.11)。

また, L_τ に対する検索結果が $R_\tau = \{S^{(r)} | r = 1, 2, \dots, Q\}$ であるとき, 固有空間 $E^{(r)}$ 内の軌跡の類似度評価における点の対応づけ関係も記録しておく.

この結果を用いて, 以下の条件をみたすとき, 入力画像系列 s に対する総合的な検索結果とする.

- (1). 部分入力画像の検索結果 R_τ が連続する. ($R_\tau \neq S^{(0)}$)
- (2). 固有空間内の軌跡の対応点の一方向である.
- (3). 入力軌跡が対応する学習軌跡の長さが, 学習軌跡全体の長さに十分近い.

まず, ①) において, R_τ が $\{S^{(r)} | r = 1, 2, \dots, Q\}$ であり, かつ, 連続するかを評価する. 次に, ②) では, 入力画像系列に含まれる動きが, 部分的に反復されている場合を取り除くため, 固有空間 $E^{(r)}$ 内において, 連続した部分入力画像系列間で, 対応づけられた学習軌跡の点が連続かどうか評価する. 最後に, ③) において, 入力画像系列に含まれる動きの意味が学習画像系列の動きの意味の類似性を評価するために, 対応づけられた学習軌跡の長さが, 学習軌跡の全体に近いかどうかを評価する. ここでは, 対応づけられた学習軌跡の点の番号を元に評価する.

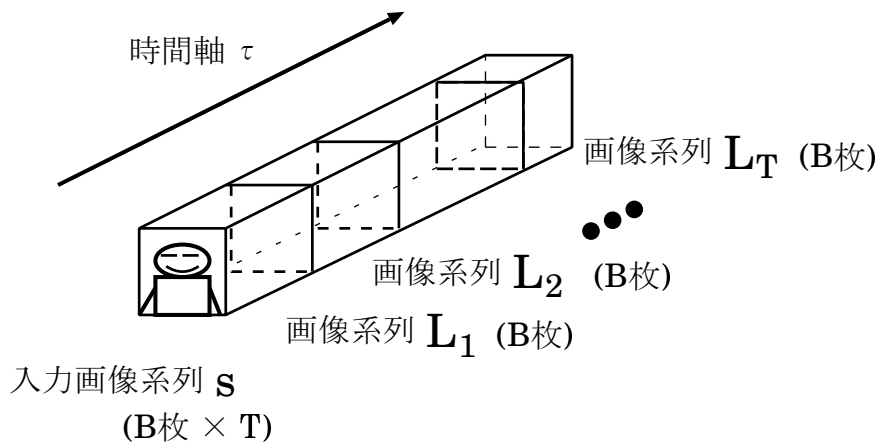


図 5.10 入力画像系列の分割

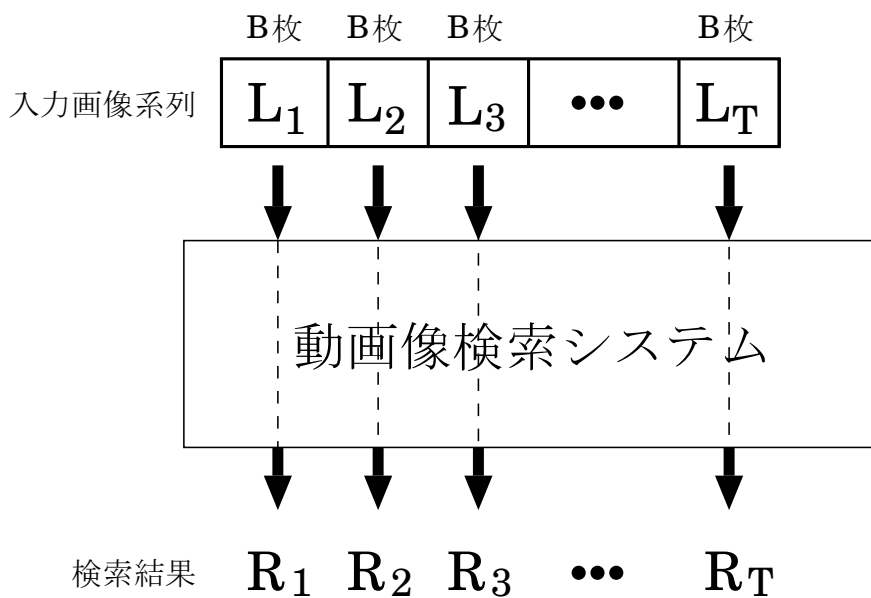


図 5.11 時系列の検索結果

5.3.2 シミュレーション実験

分節化されていない入力への提案手法の拡張の可能性を確認するため、シミュレーション実験を行なった。学習画像系列として、前述の実験で用いたものと同じ8種類のジェスチャを表現する画像系列を用意した。また、前述の実験と同じ固有空間 $\mathbf{E}^{(i)}$ と学習軌跡 $\mathbf{T}^{(i)}$ を用意した ($i = 1, 2, \dots, 8$)。さらに、固有空間 $\mathbf{E}^{(i)}$ 内の軌跡の類似度を評価する際に用いる閾値 (31 次元) を以下のように前述の実験結果を利用して、表 5.2 のように設定した。ここでは、固有空間 $\mathbf{E}^{(1)}$ 内の閾値の設定方法を例として述べる。まず、固有空間 $\mathbf{E}^{(1)}$ 内の軌跡の類似度評価の閾値を求める。図 5.7 において、固有空間 $\mathbf{E}^{(1)}$ 内の軌跡間の距離 $D_1^{(1)}$ は閾値以下で、 $D_2^{(1)}$ は閾値以上と設定されなければならない。そのため、次元数が 31 のときの $D_1^{(1)}$ と $D_2^{(1)}$ のほぼ中間の値 (240000) として閾値が得られる。次に、本実験で用いる閾値に正規化する。前述の実験結果より得られた閾値 60 枚からなる入力画像系列の場合の閾値であり、入力画像系列を構成する画像の数 (10) で正規化 ($240000 \times \frac{1}{6}$) して、最終的な閾値 (40000) とする。

入力画像系列として、図 5.12 に示すような合計 100 枚からなる画像系列 ($B = 10$) を構成した。入力画像系列の構成方法を以下に示す。

- (i). 前述の実験に用いた入力画像系列 s_l ($l = 1, 2, \dots, 8$) の合計 480 枚の画像から無作為に 100 枚取り出し、順に並べた。
- (ii). 図 5.12 の斜線の位置の画像と、 s_1 の画像 60 枚とを入れ換え、100 枚の画像系列を 10 枚ずつの部分入力画像系列 L_τ ($\tau = 1, 2, \dots, 10$) に分割した。このとき、分割された 10 枚の部分入力画像系列 L_τ の先頭が s_1 の先頭とにならないように注意した。

これにより、画像系列 L_3 の 7 枚目から画像系列 L_9 の 7 枚目までが「まる」のジェスチャを表す s_1 となる。

以上の設定のもとでの検索実験の結果を以下に示す。部分入力画像系列 L_τ ($\tau = 1, 2, \dots, 10$) と固有空間 $\mathbf{E}^{(i)}$ との距離の計算結果を表 5.3 に示す。表 5.3 は、31 次元における距離の値を小さいものから昇順に順位づけをした結果である。

表 5.2 固有空間内の軌跡の類似度評価に用いる閾値 (31 次元)

	$E^{(1)}$	$E^{(2)}$	$E^{(3)}$	$E^{(4)}$	$E^{(5)}$	$E^{(6)}$	$E^{(7)}$	$E^{(8)}$
閾値	40000	20000	35000	40000	25000	35000	30000	33000

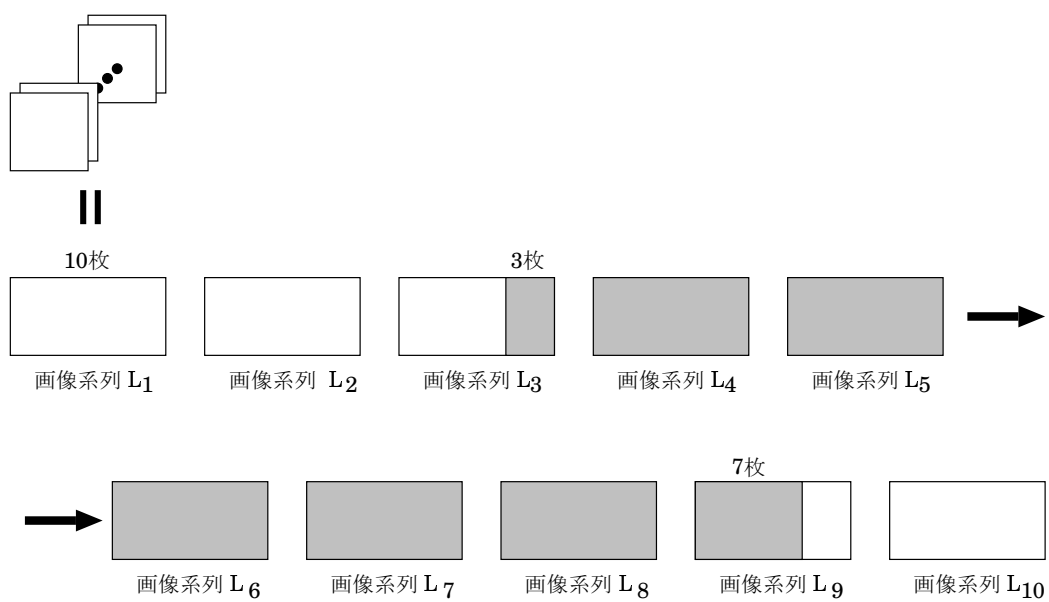


図 5.12 入力画像系列

表 5.3 部分入力画像系列 L_τ と固有空間との距離の順位

		固有空間							
		$E^{(1)}$	$E^{(2)}$	$E^{(3)}$	$E^{(4)}$	$E^{(5)}$	$E^{(6)}$	$E^{(7)}$	$E^{(8)}$
部分 入力 画像 系列	L_1	1	4	5	3	8	7	6	2
	L_2	1	4	7	3	6	8	2	5
	L_3	1	5	7	2	6	8	3	4
	L_4	1	5	7	2	8	6	4	3
	L_5	1	5	8	4	6	7	3	2
	L_6	1	4	7	3	6	8	5	2
	L_7	1	4	7	3	6	8	5	2
	L_8	1	5	7	3	8	6	4	2
	L_9	1	5	6	2	8	7	4	3
	L_{10}	2	6	4	1	8	7	5	3

表 5.4 部分入力画像系列 L_τ に対する検索結果

		検索結果
部 分 入 力 画 像 系 列	L_1	$S^{(0)}$ 未知
	L_2	$S^{(0)}$ 未知
	L_3	$S^{(0)}$ 未知
	L_4	$S^{(0)}$ 未知
	L_5	$S^{(1)}$
	L_6	$S^{(1)}$
	L_7	$S^{(1)}$
	L_8	$S^{(1)}$
	L_9	$S^{(0)}$ 未知
	L_{10}	$S^{(0)}$ 未知

次に、固有空間内での軌跡の類似度を評価する。それぞれの部分入力画像系列に対し距離が最小となった固有空間内での軌跡の類似度が閾値以下となったのは、表 5.3 中の順位を四角で囲った L_5, L_6, L_7, L_8 の固有空間 $\mathbf{E}^{(1)}$ 内の軌跡間の距離であった。これにより、部分入力画像系列 L_r に対する検索結果は表 5.4 のようになる。

また、固有空間 $\mathbf{E}^{(1)}$ 内の入力部分画像系列 L_5 から L_8 に対応する軌跡と学習軌跡間の距離を求める際の点の対応づけの結果をそれぞれ表 5.5 から表 5.8, および、図 5.13 に示す。ここで、固有空間 $\mathbf{E}^{(1)}$ 内の学習軌跡 $\mathbf{T}^{(1)} = \{\#1, \#2, \dots, \#60\}$ とし、入力画像系列 L_5, L_6, L_7, L_8, L_9 の固有空間 $\mathbf{E}^{(1)}$ 内の軌跡をそれぞれ、

$$\mathbf{t}_5^{(1)} = \{\#41, \#42, \dots, \#50\},$$

$$\mathbf{t}_6^{(1)} = \{\#51, \#52, \dots, \#60\},$$

$$\mathbf{t}_7^{(1)} = \{\#61, \#62, \dots, \#70\},$$

$$\mathbf{t}_8^{(1)} = \{\#71, \#72, \dots, \#80\}$$

とする。

ここで、前節で定めた 3 つの条件を評価する。まず、表 5.4 の検索結果より、結果が連続しているのは、 L_4 から L_8 の部分だけである。

次に、出現順序の一方向性を調べる。図 5.13 より、部分入力画像系列間で学習軌跡の対応づけられた点の番号が単調に増加していることが分かる。これより、入力画像系列 L_5 から L_8 の軌跡が学習軌跡の一部だけで繰り返し対応づけられたのではなく、また、 L_5 から L_8 の軌跡を一つの軌跡とみなした時の順序と学習軌跡 $\mathbf{T}^{(1)}$ の順序が同じであることが分かる。

さらに、図 5.13 より、 L_5 から L_8 にかけての入力軌跡は学習軌跡の $\#15$ から $\#50$ までに対応づけられている。そのため、学習軌跡の全体の長さに十分近いと判断すれば、 L_5 から L_8 は学習画像系列 $S^{(1)}$ と類似する内容の画像系列であるとみなすことができる。この場合、入力として「まる」の動作をしたものと判断し、「まる」の画像系列 $S^{(1)}$ を検索結果とすることができる。

また、学習画像系列に $S^{(1)}$ が存在しない場合は、固有空間内での軌跡間の距離がいずれも閾値以上となり、入力画像系列 s は、未知の画像系列であると判断

される。

以上の結果より，本節の実験結果を考察する． L_4 の検索結果は未知の画像系列であると判断されている．これは，データベースの状況や固有空間内の軌跡の類似度評価の閾値の設定によっては，「まる」，「ぼんざい」，「パチパチ」などと判断される．実際， L_4 は，「手を下ろしている」状態から「手を上げ始めている」状態への遷移を表しており， L_4 の部分だけでは，いずれの動作を表しているのか判断できない．しかし，本手法を用いて $L_5 \sim L_8$ の検索結果を総合することにより，入力画像系列 s に対して，同じ動作を表す画像系列 $S^{(1)}$ を検索できていることが分かる．これより，入力の画像系列が分節化されていない場合においても提案手法を応用することで， $S^{(1)}$ と同じ種類の動作と推定し，同時に動画像データベースの中から類似する画像系列を最終的な検索結果とすることができることが分かる．このことから，検索のための柔軟な入力方法が実現できると考えられる．

一方，本実験では，入力画像系列 s を $B(= 10)$ で分割して実験を行ない，本手法の応用の可能性を示したが， B を変化させた場合の考察が必要であると考えられる．

表 5.5 部分入力画像系列 L_5 の軌跡 $\mathbf{t}_5^{(1)}$ と学習軌跡 $\mathbf{T}^{(1)}$ の対応点

入力軌跡 $\mathbf{t}_5^{(1)}$	#41	#42	#43	#44	#45	#46	#47	#48	#49	#50
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
学習軌跡 $\mathbf{T}^{(1)}$	#15	#16	#17	#17	#18	#19	#19	#21	#21	#21

表 5.6 部分入力画像系列 L_6 の軌跡 $\mathbf{t}_6^{(1)}$ と学習軌跡 $\mathbf{T}^{(1)}$ の対応点

入力軌跡 $\mathbf{t}_6^{(1)}$	#51	#52	#53	#54	#55	#56	#57	#58	#59	#60
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
学習軌跡 $\mathbf{T}^{(1)}$	#22	#23	#25	#33	#33	#33	#33	#33	#33	#33

表 5.7 部分入力画像系列 L_7 の軌跡 $\mathbf{t}_7^{(1)}$ と学習軌跡 $\mathbf{T}^{(1)}$ の対応点

入力軌跡 $\mathbf{t}_7^{(1)}$	#61	#62	#63	#64	#65	#66	#67	#68	#69	#70
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
学習軌跡 $\mathbf{T}^{(1)}$	#33	#33	#33	#33	#33	#33	#36	#39	#39	#40

表 5.8 部分入力画像系列 L_8 の軌跡 $\mathbf{t}_8^{(1)}$ と学習軌跡 $\mathbf{T}^{(1)}$ の対応点

入力軌跡 $\mathbf{t}_8^{(1)}$	#71	#72	#73	#74	#75	#76	#77	#78	#79	#80
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
学習軌跡 $\mathbf{T}^{(1)}$	#44	#44	#44	#45	#46	#46	#47	#48	#50	#50

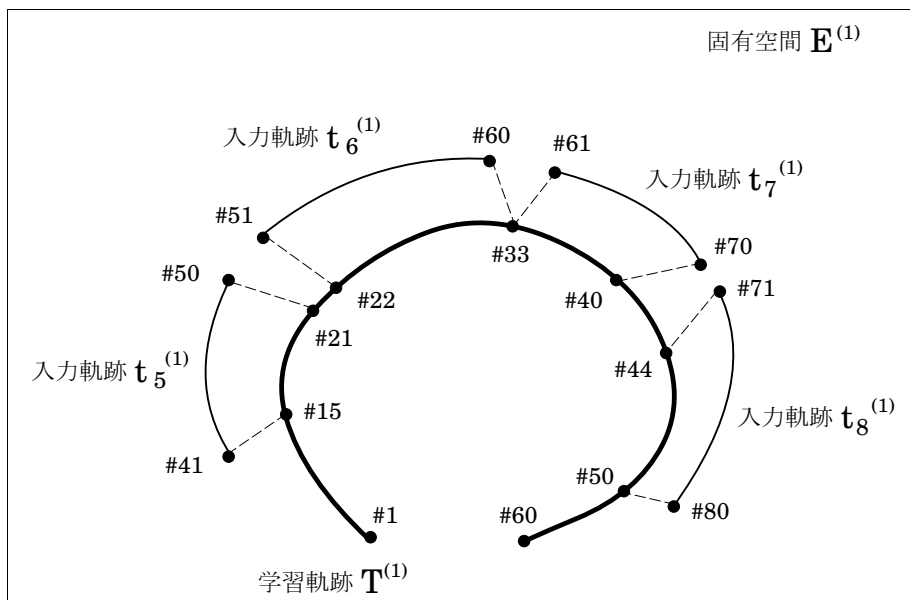


図 5.13 入力軌跡 $\mathbf{t}_r^{(1)}$ と学習軌跡 $\mathbf{T}^{(1)}$ との対応点

6. むすび

本論文では、内容に基づく動画像検索への一つのアプローチとして、動画像の内容が類似し、含まれる運動が類似しているものを動画像データベースの中から検索する手法を提案した。本手法は、画像系列間の類似性を固有空間内の軌跡の類似性を評価することで類似画像系列を検索する手法である。また、実画像を用いた実験により動画像検索手法への有効性を示した。さらに、本手法を応用することで分節化されていない入力に対しても動き情報を推定し、同時に類似画像系列を検索できることを示した。

本手法の特徴として、以下の点が挙げられる。

- 画像系列毎に固有空間を構成するため、新たな動作の学習が容易である。
- 画像系列に含まれる運動を固有空間内での軌跡としてみなすことで、移動物体の追跡などを行なうことなく、運動を記述することができる。
- 検索においては、射影などの比較的小さな計算量だけですむため、高速化が期待できる。

本手法の応用として、ユーザのジェスチャを用いたインタフェースへの応用などが考えられる。具体的には、計算機はユーザの動きを常に監視し、計算機上でのタスクとあらかじめ結びつけられたジェスチャがユーザによって行なわれたと判断すれば、そのタスクを実行するなどが考えられる。例えば、第5.3節で述べた手法を適用することにより、以下のようなシステムが考えられる。まず、計算機上でのタスクと結びつけられたジェスチャを表す画像系列から固有空間と学習軌跡をあらかじめ構成しておく。ユーザの計算機への入力として、計算機に備えつけた CCD カメラでユーザの動作を常時撮像する。ここで、第5.3節と同様に、ある枚数の画像が蓄積される度に提案した検索を行なう。この検索結果と同じ結果が以前からある程度連続し、かつ固有空間内の学習軌跡と類似すると判断できれば、既に学習されたジェスチャが入力されたと判断する。このシステムにより、判断されたジェスチャと対応する計算機上のタスクを実行する。

提案手法の問題点として、以下の点が挙げられる。まず、本手法では、入力画像系列と距離が最小である固有空間内でのみ、軌跡の類似度を評価する。そのため、入力画像系列との距離がほぼ同じである固有空間が複数ある場合に、所望でない固有空間との距離が最小と判断された場合は、固有空間内の軌跡は類似しないと判断され、未知の画像系列や別の画像系列だと判断されてしまうことがある。これは、入力画像系列と学習画像系列を構成する画像が多少の相違がある場合に生じるものと考えられる。この解決策として、固有空間との距離が閾値以下と判断された複数の固有空間で軌跡の類似度を評価するなどが考えられる。この閾値は、固有空間との距離の最小値と評価された値を基準にして設定する。

また、本手法では、画像系列毎に固有空間を構成するため、各固有空間毎の軌跡の類似度評価における閾値の設定が容易でない。しかし、画像系列に含まれる運動が固有空間内の軌跡によって表現されており、実験より固有空間毎に軌跡の類似性が大きく異なることが分かる。そのため、固有空間内の軌跡の類似性の評価を工夫することで、動画像の検索が可能であると考えられる。さらに、逆にいうと、検索の際の類似度の定義がユーザによって異なるので、個人に適応できるように、検索の過程において閾値を変化させることにより検索結果の幅を広げたり狭めたりすることもできるとも考えられる。これに対して、与えられた全ての画像系列を用いて固有空間を一つ構成する方法も考えられるが、この場合、画像中に含まれる運動を分離可能な方向のベクトルを抽出するなどの工夫が必要であると考えられる [4]。また、特に学習のデータベースを拡張する際には、固有空間を再構成し直す計算時間の側面から困難であるという問題もある。

本論文では、固有空間を構成する際の次元 K を経験的に 30 として実験を行ない、本手法の有効性を示したが、 K を変化させた場合の考察が必要であると考えられる。さらに、最適な K を自動的に求める方法の検討も必要である。同様に、入力画像系列を分割する B を変化させた場合の考察が必要である。

今後の課題としては、各固有空間の固有値を用いるなどして軌跡の類似度評価を改良することや、各固有空間毎に設定する閾値を検討することが挙げられる。また、固有空間との距離と軌跡の類似性を総合的に評価する方法を検討することが挙げられる。これと同時に、入力の画像系列が分節化されていない場合に対す

る検索におけるさらなる考察や、背景との差分やモザイクをかけるなどの前処理された画像系列への適用などについても検討する予定である。

謝辞

本研究を進めるにあたり，終始暖かい御指導を頂いた ソフトウェア基礎講座 横矢直和 教授に厚く御礼申し上げます。

副指導教官としてご助言を頂いた 音情報処理講座 鹿野 清宏 教授，並びに ソフトウェア基礎講座 竹村治雄 助教授に深く感謝致します。

本研究への助言や研究環境の整備に協力して頂いた ソフトウェア基礎講座 岩佐英彦 助手， 山澤一誠 助手に深く感謝致します。

日々の研究室の活動を支えて下さった ソフトウェア基礎講座事務補佐員 大隈 (旧姓 村上) 和代 女士に心より感謝申し上げます。

本研究の実験を行なうに際し，快く被験者として協力下さった， 茶園篤 氏と，本研究の実験の遂行を協力下さった， 栗飯原述宏 氏に感謝致します。

最後に，ソフトウェア基礎講座の諸氏には，多大なる御協力を頂きました。ここに感謝の意を表します。

参考文献

- [1] M. J. Black and A. D. Jepson: “EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation,” Proc. ECCV’96, pp. 329–342 (1996).
- [2] A. F. Bobick and A. D. Wilson: “A State-based Technique for the Summarization and Recognition of Gesture,” FiFth Int. Conf. on Computer Vision, pp. 20–23 (1995).
- [3] S. K. Chang and C. W. Yang, “An Intelligent Image Database System,” IEEE Trans. Softw. Eng., Vol.14, No.5, pp. 681–688 (1988).
- [4] Y. Cui, D. L. Swets and J. J. Weng: “Learning-Based Hand Sign Recognition Using SHOSLIF-M,” Proc. of Int. Conf. on Computer Vision, pp. 631–636 (1995).
- [5] 藤本, 岩佐, 横矢, 竹村: “固有空間内の軌跡の類似性に基づく動画像識別法,” 信学情報・システムソサイエティ大会, SD-7-2, pp. 533–534 (1996).
- [6] 藤本, 岩佐, 横矢, 竹村: “固有空間内の軌跡の類似性に基づく動画像検索,” 信学技報, PRMU96-110, pp. 49–56 (1996).
- [7] V. N. Gudivada and V. V. Raghavan: “Content-Based Image Retrieval Systems,” IEEE Computer, Vol.28, No.9, pp.18–22 (1995).
- [8] 畠, 岩井, 谷内田: “KL 展開と隠れマルコフモデルによるジェスチャ認識,” 情処研報, CVIM103-15, pp. 115–122 (1997).
- [9] 林, 河合, 上原, 豊田: “画像データベースの意味検索の試み – 道路地図データベースの略地図による検索 –,” 情処研報, 44-2 (1986).
- [10] 石川, 有木: “正規化複数部分空間法による顔認識と話者認識の統合,” 信学技報, PRMU96-41, pp. 31–38 (1996).

- [11] 岩波書店: “情報科学辞典,” (1990).
- [12] 加藤, 下垣, 藤村: “画像対話型商標・意匠データベース TRADEMARK,” 信学論, Vol.J72-DII, No.4, pp. 535-544 (1989).
- [13] 桐島, 佐藤, 千原: “身振りの動画像認識によるユーザインタフェース,” 信学技報, PRU95-190, pp. 1-6 (1996).
- [14] 小松, 有木: “部分空間法を用いた向きによらない顔の切り出しと認識,” 信学技報, PRU95-191, pp. 7-14 (1996).
- [15] 倉掛, 桑野, 新井, 小高: “認識技術を用いた映像中キーターゲットインデクシングの検討,” 信学技報, 画像工学, IE95-150, pp. 15-20 (1996).
- [16] 栗田, 下垣, 加藤: “主観的類似度に適応した画像検索,” 情処論, Vol.31, No.2, pp. 227-237 (1990).
- [17] 黒川, 洪: “形状情報を用いた画像の類似検索システム,” 情処論, Vol.32, No.6, pp.721-729 (1991).
- [18] 前田, 椋木, 美濃, 池田: “検索者の意図を的確にくみとる画像検索システム,” 1995 信学 秋季大会, A-173, pp. 176 (1995).
- [19] 美濃: “知的映像メディア検索技術の動向,” 人工知能学会誌, Vol.11, No.1, pp. 3-9 (1996).
- [20] 宮森, 粕谷, 富永: “ドキュメントアーキテクチャの拡張に基づく構造化ビデオの検討,” 信学技報, IE-96-37, pp. 17-24 (1996).
- [21] 茂木, 有木: “部分空間法による認識をベースとした手書き文字の切り出し,” 信学技報, PRU94-97, pp. 1-8 (1995).
- [22] 茂木, 寺西, 有木: “文字らしさを表す複数の部分空間を用いた手書き文字の切り出しと認識,” 信学技報, PRU95-105, pp. 13-18 (1995).

- [23] 茂木, 有木: “ニュース映像中の文字認識に基づく記事の索引付け,” 信学技報, IE95-153, pp. 33-40 (1996).
- [24] 椋木, 美濃, 池田: “対象物スケッチを利用した風景画像検索,” 第22回画像電子学会年次大会 (1994).
- [25] H. Murakami and B.V.K. Vijaya Kumar: “Efficient Calculation of Primary Images from a Set of Images,” IEEE Trans. PAMI, Vol.PAMI-4, No.5, pp. 511-515 (1982).
- [26] 村瀬, S. K. Nayar: “2次元照合による3次元物体認識-パラメトリック固有空間法-,” 信学論, Vol.J77-D-II, No.11, pp. 2179-2187 (1994).
- [27] 村瀬, S. K. Nayar: “2次元照合による3次元物体認識とその学習-パラメトリック固有空間法-,” 信学技報, PRU93-120, pp. 31-38 (1994).
- [28] 村瀬: “CVCV-WG 特別報告: コンピュータビジョンにおける技術評論と将来展望 (VI) - 固有空間法による画像認識 -,” 情処研報, CV97-9, pp. 59-66 (1995).
- [29] 村瀬, S. K. ナイヤー: “多重解像度と固有空間表現による3次元物体のイメージスポッティング,” 情処論, Vol.36, No.10, pp.2234-2243 (1995).
- [30] H. Murase and R. Sakai: “Moving object recognition in eigenspace representation: gait analysis and lip reading,” Pattern Recognition Letters, Vol.17, No.2, pp. 155-162 (1996).
- [31] 村田, 中村, 大田: “映像におけるカット変わりの自動検出 - フェードイン、フェードアウト、ディゾルブの場合 -,” 情処全大 50(前), Vol. 2, 6D-7, pp. 107-108 (1995).
- [32] 長坂, 田中: “カラービデオ映像における自動索引付け法と物体探索法,” 情処論, Vol.33, No.4, pp.543-550 (1992).

- [33] 長屋, 関, 岡: “ジェスチャ認識のための動作軌跡特徴の提案,” 信学技報, PRU95-142, pp. 45-50 (1995).
- [34] 長屋, 関, 岡, 向井: “A Proposal of Pattern Space Trajectory for Gesture Spotting Recognition,” 画像の認識・理解シンポジウム (MIRU'96) 講演論文集, Vol.II, pp. 157-162 (1996).
- [35] 中林, 吉田, 福村: “天気図検索システムと等圧線図システムの統合,” 信学技報, PRL85-81, pp. 41-48 (1985).
- [36] 西村, 岡: “ジェスチャ動画の逆時間動作や静止動作をスポットティング認識するための Non-monotonic 連続 DP,” 信学技報, PRU96-32, pp. 49-56 (1996).
- [37] 西山, 松下: “画像の構図を用いた絵画検索システム,” 情処論, Vol.37, No. 1, pp. 101-109 (1996).
- [38] 尾田, 赤松, 深町: “あいまい顔イメージの検索に対する KL 展開の適用性,” 信学論, Vol.J79-A, No.2, pp. 288-297 (1996).
- [39] H. Ogawa: “Karhunen-Loeve Subspace,” Proc. IAPR Int. Conf. on Pattern Recognition, pp. 75-78 (1992).
- [40] 大辻, 外村, 大庭: “映像カット自動検出方式の検討,” テレビ技法, Vol. 16, No. 43, pp. 7-12 (1992).
- [41] 坂内正夫: “画像検索技術,” 信学誌, Vol.71, No.9, pp. 911-914 (1988).
- [42] 櫻井, 岩佐, 竹村, 横矢, 加藤: “全周距離データを用いた固有空間法による顔認識,” 信学技報, PRU95-193, pp. 23-28 (1996).
- [43] 柴田, 金: “シーン記述に基づく映像の要約再生,” 信学技報, 画像工学, IE95-152, pp. 27-32 (1996).
- [44] 島田, 大倉, 塩野, 橋本: “複数特徴部分空間法による手書き類似漢字識別,” 信学論, Vol.J78-D-II, No.10, pp. 1460-1468 (1995).

- [45] R. K. Srihari: "Automatic Indexing and Content-Based Retrieval of Captioned Images," IEEE Computer, Vol.28, No.9, pp.49-56 (1995).
- [46] D. L. Swets and J. J. Weng: "Efficient Content-Based Image Retrieval using Automatic Feature Selection," Proc. Int. Symposium on Computer Vision, pp. 85-90, (1995).
- [47] H. Tamura and N. Yokoya: "Image database systems: a survey," Pattern Recogn., Vol.17, No.1, pp. 29-43 (1984).
- [48] テレビジョン学会編: "テレビジョン・画像情報工学ハンドブック," オーム社, 7-10 (1990).
- [49] M. A. Turk and A. P. Pentland: "Face Recognition Using Eigenfaces," Proc. IEEE Conf. on CVPR, pp. 586-591 (1991).
- [50] 上田, 宮武, 吉澤: "認識技術を応用した対話型映像編集方式の提案," 信学論, Vol.75-D-II, No.2, pp. 216-225 (1992).
- [51] 上田, 宮武, 炭野, 長坂: "動画像解析に基づくビデオ構造の視覚化とその応用," 信学論, Vol.76-D-II, No.8, pp. 1572-1580 (1993).
- [52] 上堀, 村田, 中村, 大田: "画像特徴と音響特徴を利用したCM映像の自動的構造化手法," 信学技報, PRU95-159, pp. 9-12 (1995).
- [53] M. Uenohara and T. Kanade: "Vision-Based Object Registration for Real-Time Image Overlay," Proc. CVRMed'95, pp. 13-22 (1995).
- [54] 大和, 村瀬: "HMM を用いた MPEG データからの動画像検索の検討," 信学技報, PRMU96-34, pp. 65-72 (1996).
- [55] 柳井: "多変量データ解析法 - 理論と応用 - ," 朝倉書店, 3 (1994).
- [56] 横矢, 田村: "画像データベース研究の動向," 電総研彙報, Vol.45, No.9, 10, pp. 57-71 (1981).

- [57] 横矢: “画像データベース,” 画電学誌, Vol.12, No.2, pp. 81-89 (1983).
- [58] S. Watanabe and N. Pakvasa: “Subspace Method in Pattern Recognition,” Proc. Int. Joint Conf. on Pattern Recognition, pp. 25-32 (1973).
- [59] 渡辺, 李, 谷内田: “インタラクティブシステム構築のための実時間ジェスチャ認識の一手法,” 情処研報, CVIM103-14, pp. 107-114 (1997).