

## 隣接性に基づく個人行動記録の構造化類似検索 Associative Retrieval of Personal Experience Records based on Contiguity

小泉 敬寛<sup>†</sup>  
Takahiro Koizumi

中村 裕一<sup>†‡</sup>  
Yuichi Nakamura

### 1. はじめに

コンピュータや情報機器を人間が装着し、情報支援をはじめとする様々な用途に用いる試みが行なわれている。

その1つとして、頭部装着カメラをはじめとする各種のセンサを体に装着し、そこから得られたデータを記憶の補助や経験の共有のために用いる試みがある [1][2][3]。しかしこのような方法で撮り流した映像やセンサデータは長時間のデータとなるため、そのままの形で利用するのは困難である。そこで、何らかの方法で要約を行ったり、効率的な検索を可能にする必要があり、これまで種々の手法が提案されてきた [3][4]。

例えば、堀らは、映像記録だけでなく発汗や脈拍センサによる生体情報やメール、GPSによる位置情報など多量のインデックスを組み合わせたライフログの検索手法を提案した [3]。川嶋らは活動場所に基づいて日常の行動を区分することで、体験映像の整理要約する手法を提案した [4]。大出らはカメラ装着者が何かに注目しようとした場面（注目シーン）を用いる手法を提案した [5]。この手法は、注目シーンが日常行動を想起、要約するのに適していることを利用したものである。久保田らは、環境に備え付けられた広視野のカメラを併用することで、詳細な位置情報や装着者周辺の状況を加えた行動記録を作成する方法を提案した [6]。

しかし、このような要約・構造化を行なったとしても、記録が長時間になるにつれ、必要な情報を探ることが難しくなる。さらに、このような記録を検索する目的には様々なものがあることも考えなければならない。例えば、単に自分の行動を思い出すという目的もあれば、他人の記録から何か未知の情報を得るといったものもあり、各々によって検索すべきデータやそれに適した提示方法が異なる。そのため、予め定義された特徴量による類似検索では必要な情報を検索することが難しい場合がある。

そこで、本研究では個人行動記録中の各断片間の関係を扱うために隣接性を基にすることを提案する。次に、隣接性に基づく検索と類似検索とを同時に行う手法を提案し、簡単な実験によって提案手法が効果的に働くことを示す。

## 2. 個人行動記録の検索

### 2.1 問題点

個人行動記録の用途を考えると、○○をどのようにやったかを思い出したい/ 知りたい、こんな○○を捜したい、というような曖昧な問い合わせが多くなる。そのため、単純な特徴量に基づく類似検索を行おうとすると、特徴量

の選択や検索キーの指定が難しい。このような問題に対する検索手法として、適合性フィードバックによる類似検索がある [7]。適合性フィードバックはユーザが選択したサンプル集合を基にして、類似度を求める関数のパラメータや特徴量の重みを調整しながら類似検索を行なう手法である。これにより、ユーザは必要とするデータに似ていると考えたサンプルを選ぶだけで良いという利点がある。

しかし、ユーザがこのようなサンプルを選択することすらできない場合がある。例えば、「これから行なう作業に必要な道具は何か」を検索をする場合を考えてみる。ユーザは必要となる道具の色や形状を知らないため、検索したいデータに類似する検索キー（例えば画像）を選択することが難しい。また、行動記録の短い断片を見ただけでは、その行動がどのようなものであるかを判断することが難しい場合もある。これらの問題は、単純に特徴量が似ているという関係だけで行動記録を検索するのは不十分であることを表している。そこで、我々は隣接性による関連記録の検索手法を提案する。

### 2.2 隣接性に基づく関連記録の検索

記憶や記録をたどる最も有効な方法の一つは、強く関連する情報を芋づる式に引き出すことである。我々は、この基となる関連性として、メトニミー（換喩）という言葉現象に着目した。メトニミーは、ある対象に言及することにより他方の対象を表す比喩表現であり、二つの物事が強く関連していると人間が認識している際によく現れる。メトニミーの基となる関係を隣接性<sup>§</sup>と呼ぶが、例えば、場所で機関を表す場合（「霞が関」で「官庁」を表す等）、道具で使用者を表す場合（「ピアノ」で「ピアニスト」等）、部分で全体を表す場合（「花」で「草木」等）、その他様々な関係がある。このように隣接性で結び付けられる対象同士は、人間が日常的に関連付けていたり、同時に思い浮かべるものが多い。このような隣接性に基づいた検索を行うことができれば、自分の記憶や他人による記録を効率的に利用することが可能となる。

このような検索を考えたときにまず問題となるのが、隣接性を検出する方法である。本研究では、そのために、個人行動記録映像とそれが記録された時間、場所の情報を利用する。関連性を人間の行動から検出するために映像、時間、場所の情報を処理するのが本研究のポイントである。これについては3.で述べる。

さらに本研究では、隣接性に基づいて行動記録断片検索する手法として関連性フィルタを用いる。適合性フィードバックを用いて類似検索を行った結果に関連性フィル

<sup>†</sup> 京大大学院工学研究科電気工学専攻

<sup>‡</sup> 京大大学術情報メディアセンター

<sup>§</sup> contiguity. 近接性と呼ぶ場合もある。

タを適用することで、関連の強い行動記録断片を同時に検索結果として提示する。これにより、ユーザが検索キーとする行動記録断片を選ぶのに苦勞する必要がなくなり、関連するデータが萃づる式に検索され、効率的な個人行動記録の検索が可能となる。これについては4.で述べる。

### 2.3 個人行動記録断片 (注目シーンの利用)

本研究で用いる個人行動記録は、頭部装着カメラから撮られた個人視点映像と、天井に設置された広視野のカメラによって行動者周辺の様子が撮られた環境カメラ映像により構成される。したがって、個人行動記録は映像とそれに対応する位置座標や時刻が付加されたものであるが、それを静止画や短時間のビデオセグメント (行動記録断片) の集まりとして考える。この行動記録断片は、できるだけ単一の行動を表し、なおかつ何が映っているのかを人間が理解し易い断片であることが望ましい。これについて、我々は大出らの手法 [5] により得られた注目シーンを行動記録断片の単位とすることにした。これにより、単に移動している間の見づらいシーンや冗長なシーンを取り除き、カメラ装着者が何かに注目していたり、動きを見守っている重要なシーンだけを検索、閲覧することが可能になる。

### 3. 隣接性の検出

行動記録断片が表す情報間の隣接性には、例えば以下のようなものがある。

- 箱 ↔ 鍵 ある物 (箱など) とそれを利用するために必要な物 (鍵など)
- 物 ↔ 位置 ある物と存在する (と思われる) 場所 (定位置)
- 行動 ↔ 結果 ある行動とその結果
- 道具 ↔ 行動 ある道具とそれを使ってする行動

本研究では、隣接性を検出するために、行動記録から種々のイベントを検出し、そのイベントやイベント間の関係を用いる。以下に、既に実装したいくつかのイベント検出方法をあげる。

まず、意味的なイベントの一つとして、カメラ装着者が何かを手に取る、置く、あるいは操作する、といった行為を行う際に手を使うことから、個人視点映像から手領域の検出を行う。また、カメラ装着者が立ち止まっている時は何らかの作業がそこで行なわれていると考えられるので、その位置を停留点として検出する。

#### 手領域

ユーザが作業の中で使う物が一つだけという事は少なく、複数の関連する物を使う場合が多い。そこで、作業の中で同時に使う物同士を関連付ける。使っている物を行動記録断片から直接検出することは困難であるため、手領域を用いて間接的に物を扱うシーンを検出する。

個人視点映像からカメラ装着者の手領域を検出するために、あらかじめカメラ装着者の手の画像を学習データとして与える。学習データから手の色の統計的特徴量を計算し、その値と画像中の各ピクセルのマハラノビス距

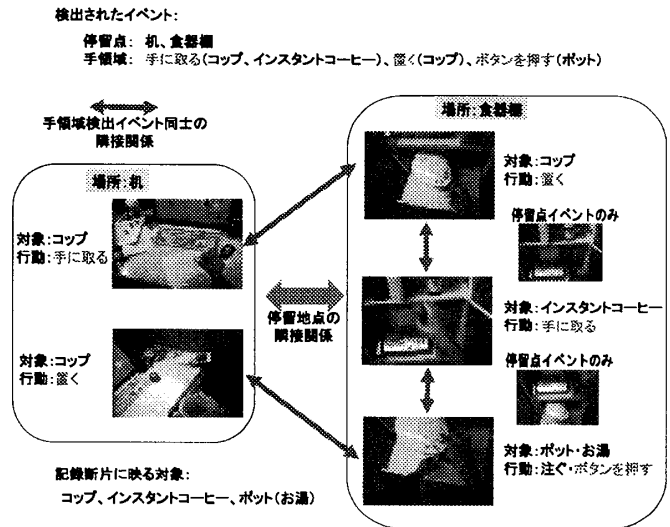


図1 イベントによる隣接性ネットワーク構築例

離を求める、しきい値以下のピクセルに対してラベリング処理を行ない、一定以上の面積の領域を手領域とする。

検出された手領域の動きが大きい時は何かを取る、置く、動かしたと考えられるので、これを動作イベントとする。逆に静止シーンのなかで、手領域も大きな動きがないときは、そこで細かい作業を行なっていると考えられるので、これを作業イベントとする。

#### 停留点

本棚と本のような関係、作業台と作業の関係など、場所により決められた行動は多い。カメラ装着者が何か作業をしていると考えられる停留点を停留イベントとし、手領域によるイベントと関連付けることで、場所と物とを関連付けることができる。

そのために、環境映像からカメラ装着者の移動追跡を行う。検出した移動軌跡から、カメラ装着者が一定時間動かなかつた場所を停留点とし、これを停留イベントとする。

#### 隣接性ネットワーク

検出されたイベントを用いて、仮想的に隣接性ネットワークを構成する。隣接性ネットワークは、イベントが検出された行動記録断片のそれぞれをノードとし、隣接性によって結び付けたネットワークである。

図1は、隣接性として単純に隣り合った同種のイベントを持つ記録断片同士を結び付けた例である。このように、行動記録断片間の関連を定義することで、意味的に関連する行動記録断片同士を間接的に結び付けることができる。例えば、手領域検出によるイベント同士の関係から、「コップ」と「インスタントコーヒー」と「ポット」を関連付けることができる。

### 4. 隣接性を用いた適合性フィードバック

#### 4.1 適合性フィードバック

適合性フィードバックの枠組みによりユーザが行動記録断片を検索する手順は図2のようになる。

まず、システムによりランダムに、あるいはあらかじめ

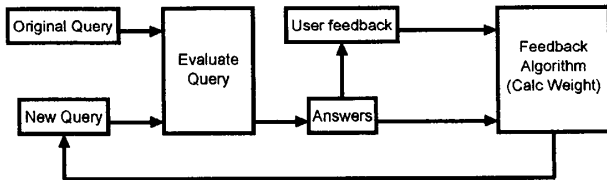


図2 適合性フィードバックによる検索手順

指定された方法によって選択された行動記録断片が初期集合としてユーザに提示される。ユーザは、この初期集合の中から画像的、場所的、時間的に似ていると思った行動記録断片を選択する。選択された行動記録断片の集合を  $Rel = \{x_i, \dots\}$  とする。この  $Rel$  に含まれる行動記録断片の特徴量の統計的情報から、適合性フィードバックによって、類似尺度の重みを調整する。次に、システムは、更新された類似尺度の重みを用いて行動記録断片の類似検索を行なう。新たに重み付けされた類似尺度による類似検索から得られた行動記録断片を見て、ユーザはそこから再び、画像的、場所的、時間的に必要な行動記録断片に似ていると思ったものを選択する。以上の方法で、目的とする行動記録断片が見つかるまで、欲しい行動記録断片に近い行動記録断片を選ぶことを繰り返す。

この方法はユーザが選択したデータに共通点があれば、特徴量  $x_{id}$  の値が似ているものがあるという予測に基づく。紙面の都合上詳細な説明は省くので、詳しくは [7][8] 等を参照されたい。

#### 4.2 隣接性の利用

本研究では 3. で述べたような隣接性を単独または複合的に用いて関連性フィルタを構成する。例えば、作業中に必要な物が普段どこに置いてあるのかを示す行動記録断片を検索したい場合は 3. の、道具  $\leftrightarrow$  行動の関係と、物  $\leftrightarrow$  位置の隣接性を組み合わせることにより、一連の作業中に用いた物が普段置いてある場所を示す行動記録断片が検索されやすくなる。

まず、行動記録断片間の隣接性  $\gamma$  を表すために、次のような関連行列  $A^\gamma$  を用いる。

$$A^\gamma = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

$a_{ij}$  には 0 ~ 1 の値が入り、行動記録断片  $x_j$  に対して  $x_i$  の隣接性の強さを表す<sup>1)</sup>。  $a_{ij}$  の具体的な求め方は、それぞれの隣接性毎に個別に定義される。関連行列は何種類かの隣接性ごとに作成しておき、検索目的に合わせて各々のフィルタの重みを調整する。

ここで、個人行動記録  $X$  の各行動記録断片の適合度 (ユーザの検索要求に対して類似検索で得られた値) を  $r = \{r_1, r_2, \dots, r_N\}$  とする。これを用いて、隣接性による新しい適合度  $r' = \{r'_1, r'_2, \dots, r'_N\}$  を次のように計算

<sup>1)</sup> 現在は、この  $a_{ij}$  には、隣接性の有無だけを 0 か 1 のみが入力される。

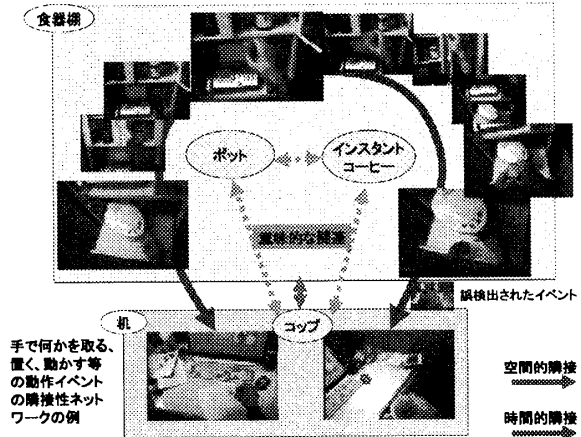


図3 動作イベントと停留イベントによる隣接性ネットワーク

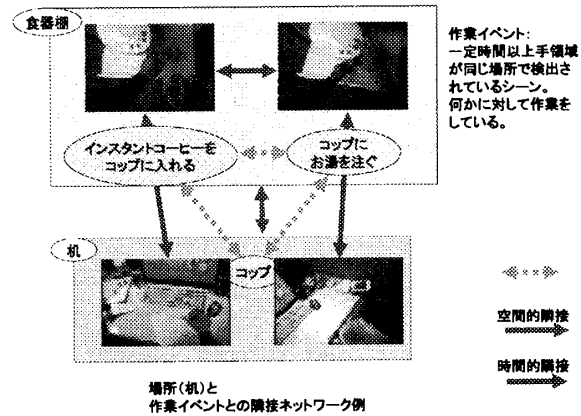


図4 作業イベントとの隣接性ネットワーク

する。

$$r' = \sum_{\gamma} (w_{\gamma} A_{\gamma} r)$$

ただし、 $w_{\gamma}$  は隣接性  $\gamma$  に与える重みである。

この手法によって、類似検索の結果上位に表れた行動記録断片  $x_i$  に強く関連する行動記録断片が検索結果に現れる。

### 5. 実験

個人視点映像は、頭部に装着された小型カメラによって撮影され、背中に背負ったリュックの中にあるノート PC に装着された MPEG2 エンコーダカードによって、MPEG2 形式の映像ファイルと保存される。環境カメラ映像は、SXGA の解像度を持つ IEEE1394 デジタルカメラ (1280 x 960, 7.5fps) により撮影され、MPEG 形式で保存される。

この実験環境を用いて取得された短時間の個人行動記録 (50 秒程度) に対して行った実験例を示す。まず、注目シーンの検出を行なった結果、36 個の行動記録断片を検出した。その内訳は、動作に関わるイベントが 12 個、作業に関わるイベントが 2 個、停留イベントが 2 箇所となった。そのうち、食器棚と机の間を移動している間に誤検出された動作イベントが 1 個あった。

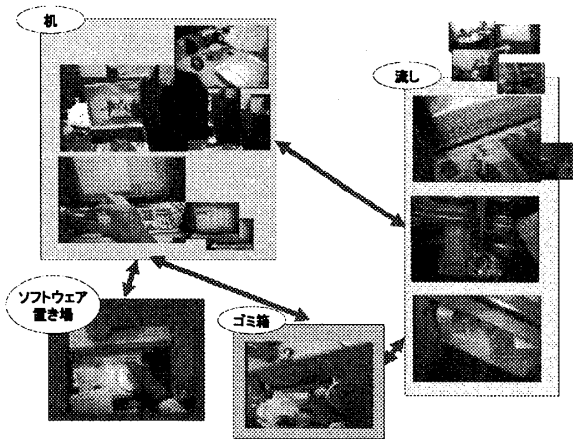


図5 検索結果

動作イベントと停留イベントの時間的・空間的隣接性による隣接性ネットワークは図3のようになった。何かを取る、置く、動かすといった動作をしている行動記録断片同士を時間的な隣接関係により結びつけることで、連続した作業の中で手にした物が映る行動記録断片を関連付けることができていることがわかる。

また、停留イベントによる空間的な隣接関係によっても、本来場所が離れている机と食器棚での行動記録断片を結びつけることができる。それとは対照的に、食器棚と机の間を移動している間に誤検出された動作イベントを持つ行動記録断片は他との隣接性に乏しく、検索結果からはずれることになる。このように、机にあるコップと食器棚のポット、インスタントコーヒーの容器の三つを意味的に関連付ける隣接性ネットワークが構築できている。

さらに、机の上のコップを取るシーンの行動記録断片と作業イベントとの隣接性によって構成される隣接性ネットワークは図4のようになる。このように作業イベントとの隣接性により、場所の離れた机の上でのコップが映る行動記録断片から、そのコップと関連する作業を写す行動記録断片を関連付けていることがわかる。

次に、2時間程度の個人行動記録を用いて検索実験を行なった例を示す。この実験では、動作イベントと停留イベントによる時間的・空間的隣接性による隣接性ネットワークを構成して、関連記録の検索を行なった。検索のための初期間い合わせとして、カメラ装着者が自分の机で作業をしている行動記録断片を適当に数個選択して、それに関連する記録を検索した。スコア順に現れた行動記録断片の上位100程度を見て、その中から有用な行動記録断片を手手で要約しまとめたものが図5である。

検索結果の上位には、選択した行動記録断片と同じような机の上で作業をしている間の行動記録断片が大量に現れた。これは、問い合わせとして、同じ場所で映像的に似通っているものを選択したため、同じ場所での似たようなシーンの行動記録断片のスコアが高くなったためである。しかし、その中でも動作イベントを含む行動記録断片はより上位に現れやすい傾向があった。これは、隣

接関係にある同じ机で作業をしているシーンの行動記録断片同士がお互いを強め合ったためであると考えられる。

次に、隣接関係にある別の場所での動作イベントを含む行動記録断片が検索された。これは、空間的に隣接関係にある動作イベントにより強められたためであると考えられる。これにより、カメラ装着者が自分の机から移動して行なった作業の記録を知ることができる。

今回の実験の動作イベントと停留イベントのように複数の意味的イベントを組み合わせて隣接性ネットワークを構成することで、関連の強い行動記録断片はお互いに強め合い、検索されやすくなると言える。

## 6. おわりに

本研究では、個人行動記録の検索のために、相互に関連があるデータの集合を効率よく検索する手法として、隣接性に基づく構造化検索手法を提案した。本手法を用いることで、類似検索では扱いにくい関連データ探索を効果的に行うことができる。実験では、いくつかの個人行動記録に対して本手法を適用することで、目的のデータを効率よく検索することができることを示した。

今後の課題としては、様々な検索内容に対してどのような隣接性が必要になり、その隣接性をどのように検出するのかを検討し、実装する必要がある。また、検索内容に応じて適切な隣接性を用いた関連性フィルタを選択する枠組みを考える必要がある。

## 参考文献

- [1] Jim Gemmell, Gordon Bell and Roger Lueder. MyLifeBits: a personal database for everything. *Communications of the ACM*, vol. 49, Issue 1 (Jan 2006), pp. 88-95, 2006.
- [2] 河野恭之, 河村竜幸, 上岡隆宏, 村田賢, 浮田宗伯, 木戸出正継. ウェアラブル日記の実現に向けて - 日常記憶の検索・編集・整理・共有機構 -. In *PRMU2003*, 2003.
- [3] 堀鉄郎, 河崎晋也, 石川尊之, 相澤清晴. ライフログ応用に向けたコンテキストに基づく映像・データ検索. In *MVE2003*, 2003.
- [4] T. Kawashima, K. Yoshikawa, K. Hayashi, Y. Aoki. Situation-based Selective Video-Recording System for Memory Aid. In *IEEE Proc. of Int. Conf. on Image Processing, III*, pp. 835-838, 1996.
- [5] Y. Nakamura, J. Ohde, and Y. Ohta. Structuring personal activity records based on attention - analyzing videos from head-mounted camera. In *15th Int'l Conference on Pattern Recognition Track4*, pp. 220-223, 2000.
- [6] S. Kubota, Y. Nakamura, and Y. Ohta. Detecting scenes of attention from personal view records - motion estimation improvements and cooperative use of a surveillance camera. In *IAPR Workshop on Machine Vision and Applications*, pp. 209-213, 2002.
- [7] Michael Ortega-Binderberger, and Sharad Mehrotra. Relevance Feedback in Multimedia Databases.
- [8] Leejay Wu and Christos Faloutsos and Katia P. Sycara and Terry R. Payne. FALCON: Feedback Adaptive Loop for Content-Based Retrieval. *The VLDB Journal*, pp. 297-306, 2000.
- [9] 山梨正明. 認知文法論. ひつじ書房, 1995.