行動状況により検索可能な体験映像提示手法の検討

志村 将吾 † 平野 靖 ‡ 梶田 将司 ‡ 間瀬 健二 ‡
† 名古屋大学大学院情報科学研究科社会システム情報学専攻
‡ 名古屋大学情報連携基盤センター

1. はじめに

計算機，ビデオカメラやマイクロフォンなどの小型化や，ハードディスクの大容量化から，ウェアラブルな装置を用いて体験を常時記録する研究が行われている [1][2]. 常時記録を行うことにより，出来事を撮り逃してしまうことがなくなる．また，記録された映像を閲覧することで，その体験を想起することが可能である．

しかし，常時記録によって長時間の映像を取得してしまうため，ユーザが閲覧したいと思う特定のシーンを発見することが困難になる．生理センサを身に付けて自動的にインデキシングを試みる研究があるが [2]，ユーザの負担が大きく常時記録には適さないシステムである．

ある一日において，ユーザが閲覧したいと思う体験は，普段とは少し違った事である可能性が高い．したがって，普段の行動をモデル化することによって，普段の生活とは異なった出来事を検出することが可能になる．そこで，著者らは,「行動」という活動単位に着目して行動のモデル化を行った．

本稿で提案する行動のモデル化によって，行動を分類した実験結果を示す．

2. 行動について

本稿で述べる行動とは，デスクワークやミーティングなどのことを表しており，これらはいくつかのサブ行動によって構成されるものである．例えば，デスクワークという行動は，キーボードを叩くサブ行動や書類などを扱うサブ行動などによって構成され，ミーティングという行動は会話というサブ行動などによって構成されている．デスクワークは自分の机で行い，講義や演習などはある決まった教室で行われるように，行動は場所が強く関連しているものである．さらに，行動を表す指標として音情報が考えられる．例えば，ミーティングでは話し声が多く，デスクワークでは話し声は少なくなり，キーボードの打鍵音や紙の資料などを扱う音が増える．

ここで，通学や通勤という行動を例に挙げる．これらは，毎日ほぼ同じ時刻に自宅を出発し，ほぼ同じ時間をかけて学校や会社に到着するという行動である．つまり，行動の開始時刻と行動に費やした時間も，行動をモデル化する上で重要な情報であると考えられる．

本研究では，場所，音情報，行動の開始時刻，そして行動に費やした時間を用いて行動をモデル化する．

3. 記録システムと行動のモデル化

3.1 行動のモデル化

本記録システムでは，映像は 5fps で記録し，音声は 44.1kHz，16bit，モノラルで収録する．さらに，場所の推定を行うために，無線 LAN のアクセスポイントの MAC アドレスと電波強度をタイムスタンプと共に記録する．場所の認識に関しては，RFID などのタグを用いる方法も挙げられるが，システム利用のためにタグなどを設置する必要がある．一方無線 LAN は，広く普及しており，システムを使用するために特別な準備の必要がなく，手軽に利用可能である．また，GPS の利用も考えられるが，屋外，屋内を問わず場所の認識を行いたいことから，本システムには採用しなかった．

取得された無線 LAN のアクセスポイントの状態から場所の推定を行う．タイムスタンプも記録されているため，ある場所に入った時刻 T_i とその場所から出た時刻 T_o も取得できる．したがって，行動に費やした時間は $T_o - T_i$ で計算できる．

行動は，いくつかのサブ行動によって成り立っており，サブ行動の出現時間の割合によって行動の分類が可能であると考える．そこで音情報を，無音，キーボードの打鍵音，紙をめくる音，人間の音声，そしてその他の音の 5 カテゴリに分類し，それぞれの出現時間の分布を行動のモデル化に利用する．

3.2 サブ行動の検出方法

記録された音から，s 点のサンプリングデータを t[ms] ずつ移動させながら切り出し，以下に示す方法を用いて，各時刻におけるサブ行動を決定する．

まず，s 点のサンプリングデータの平均パワー A_p を計算し，無音か有音であるか閾値によって判定する．閾値の決定には，周囲を無音の状態にして録音されたデータを用いる．このデータから s 点のサンプリングデータを t[ms] ずつ移動させて，各区間ごとの平均パワー A_s を計算する．各区間の平均パワーの平均 $\overline{A_s}$ と標準偏差 σ_{A_s} を計算し，式 (1) を満たせば無音，それ以外は有音と判定する．

$$A_p < \overline{A_s} + 3 \times \sigma_{A_s} \tag{1}$$

有音であると判定された区間については，ハニング窓を用いて FFT を行う．周波数変換されたデータに対して，キーボードの打鍵音か紙をめくる音であるかの判定を行う．判定方法は次の通りである．事前にこれらの音を静かな状況で n 音ずつ録音し，サンプル点数 s 個，ハニング窓を用いて FFT を行い，周波数領域に変換した各データ $F_n(\omega)$ から各周波数値における平均スペクトルを計算する．各データごとに各周波数における平均スペクトルとの 2 乗誤差 E_n を計算し，誤差の平均 $\overline{E_n}$ と標準偏差 σ_{E_n} を計算する．判定対象のデータ $F_t(\omega)$ と，このように事前に用意した平均スペクトルとの 2 乗誤差

Experience Movie Presentation Method Using Action Situation Query

Shogo Shimura, Yasushi Hirano, Shoji Kajita and Kenji Mase
†Graduate School of Information Science, Nagoya University
‡Information Technology Center, Nagoya University

表 1: 検出精度

	keyboard	paper	voice
再現率	97.0%	99.0%	78.5%
適合率	86.6%	100.0%	95.1%

正解数 (秒) keyboard:100 回, paper:100 回, voice:25 秒

E_t を計算し，E_t が式 (2) を満たす場合に，キーボードの打鍵音や紙をめくる音として判定する．式 (2) を 2 音とも満たす場合は，音の分布が正規分布であるとして，平均 $\overline{E_n}$ と分散 $\sigma^2_{E_n}$ を用いて，E_t における値が大きいものに決定する．

$$\overline{E_n} - 3 \times \sigma_{E_n} \leq E_t \leq \overline{E_n} + 3 \times \sigma_{E_n} \quad (2)$$

どちらの音でもないと判断された場合は，人間の音声であるかの判定を行う．判定の方法は，人間の音声の特徴である倍音構造を利用する．まず，スペクトルがピークになっている周波数 F_i を求め (式 (3) が真)，その中からスペクトルの値が大きいものから順に上位 3 つの周波数を選ぶ．周波数の小さい方から F_1，F_2，F_3 とし，F_1 が 100[Hz] 以上，かつ比 F_2/F_1，F_3/F_1 が整数倍であれば人間の音声と判定する．しかし，実際は完全に整数倍になることは少ないので，m のマージンを含んだ上で判定を行う (式 (4))．式 (4) の $D(x)$ は引数の小数部分を返す関数で，本式の値が真であれば人間の音声として判定し，偽であればその他の音として判定する．

$$(F_{i-1} - F_i < 0) \wedge (0 < F_i - F_{i+1}) \quad (3)$$

$$(1 - m \leq D(\tfrac{F_2}{F_1}) \leq m) \wedge (1 - m \leq D(\tfrac{F_3}{F_1}) \leq m) \quad (4)$$

4. 実験と考察

4.1 サブ行動の検出実験

3.2 で述べた検出方法を用いて，サブ行動がどの程度の精度で検出可能であるか検証する実験を行った．著者らの研究室における普段の生活の中で，キーボードを 100 打鍵，紙を 100 回めくり，そして他者が 25 秒発話した状況を約 6 分 30 秒間にわたって録音した．各パラメータは $s = 4096$，$t = 10$，$n = 50$，そして $m = 0.25$ とした．検出結果を表 1 に示す．

紙をめくる音については良好な精度が得られた．キーボードの打鍵音の再現率，人間の音声の適合率についても良好な精度が得られた．

本研究で想定している研究室内での行動は，デスクワークとミーティングであり，これらの音情報の特徴は大きく異なるため，本実験で得られた精度で十分に行動を分類できると考えている．しかし，さらに精密な精度を必要とする行動分類の出現に備えて，サブ行動の検出精度を向上させる方法を考察する．

本実験において，キーボードの打鍵音として誤って抽出した音は，紙をめくる音の直前と直後のみに発生した．紙をめくる音については適合率が高いため，紙をめくる音に付随して発生したキーボードの打鍵音は，誤って抽出されたものとして処理することで，さらに精度を向上させることができると考えられる．また，人間の音声については，適合率が高いという特徴がある．したがって，誤抽出が少ないという本検出結果のみを考慮すれば，人間の音声として検出された時間を約 1.3(100/78.5) 倍すれば，ほぼ正確な時間を算出できる．

表 2: 分類結果

		分類結果 [%]		
	行動	Deskwork	Meeting	その他
正解	Deskwork	100%	0%	0%
	Meeting	0%	100%	0%

4.2 行動の分類実験

サブ行動の時間的な出現の割合によって，行動を分類する実験を行った．デスクワークを記録したデータ 16 個 (約 75 時間) とミーティングを記録したデータ 15 個 (約 23 時間) について，3.2 で述べた検出方法を用いてサブ行動を検出し，1 時間あたりの各サブ行動の出現時間を計算した．本実験では，人間の音声の出現時間のみで十分に分類できると考え，他のサブ行動の出現時間については分類の際に用いなかった．評価には Leave-One-Out 法を用い，テストデータ 1 個を除いた学習データから，デスクワークにおける人間の音声の平均出現時間 E_d とその標準偏差 σ_d と，ミーティングにおける人間の音声の平均出現時間 E_m とその標準偏差 σ_m を計算し，テストデータの人間の音声の出現時間 T_t が式 (5)，(6) を満たすか否かによって判定した．式 (5) が真，式 (6) が偽ならば，デスクワークと判定し，式 (5) が偽，式 (6) が真ならば，ミーティングと判定する．式 (5)，(6) が共に偽であれば，その他の行動と判定し，式 (5)，(6) が共に真であれば，人間の音声の出現時間が正規分布であるとして，T_t における値の大きい方に決定する．分類結果を表 2 にまとめる．

$$E_d - 3 \times \sigma_d \leq T_t \leq E_d + 3 \times \sigma_d \quad (5)$$

$$E_m - 3 \times \sigma_m \leq T_t \leq E_m + 3 \times \sigma_m \quad (6)$$

すべてのテストデータにおいて正しく分類され，本モデル化の有効性が示された．

サンプル数の増加によって，結果が変化することも考えられ，引き続き実験を継続する必要がある．

5. おわりに

本研究では行動という単位に注目し，さらにそのモデル化を行った．本論文で提案した行動のモデル化により，行動の分類が可能なことを，実験によって確認した．

今後の課題としては，引き続きデータを収集し，さらに多くのサンプルを用いて行動のモデル化を行うことが挙げられる．また，モデル化を行った上で，体験映像を提示するインタフェースを構築する必要がある．

謝辞

本研究は文部科学省「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」プロジェクトの支援により行われた．

文献

[1] 志村 将吾, 平野 靖, 梶田 将司, 間瀬 健二, “体験記録における日記を用いた感情記録インタフェース”, 情報処理学会研究報告 (ヒューマンインタフェース), HI-115, pp. 61-68 (2005-9).

[2] 堀 鉄郎, 相澤 清晴, “ライフログビデオのためのコンテキスト推定”, 電子情報通信学会技術研究報告, Vol. 103, No. 514, pp. 67-72 (2003-12).