

習慣的な動画閲覧行動の再利用による 動画閲覧経験の拡張

高嶋 章雄^{†1} 田中 譲^{†2}

動画閲覧は、監視カメラの記録分析、ユーザ観察実験の記録分析、スポーツの試合記録分析など、映像分析の方法によって獲得できる情報が変化するという双方向メディアとしての利用が注目されている。自分や他者の行う習慣的な動画閲覧行動 (= 動画閲覧スタイル) を再利用することで、効率的、あるいは発見的な情報獲得が期待される。本研究では、動画閲覧のスキルやノウハウといった、経験的かつ暗黙的な知識に基づき確立される個人々の動画閲覧スタイルを表出化し、それらを再利用することで閲覧経験の追体験を可能にし、さらに複数の閲覧経験を組み合わせることで、ユーザに新たな動画閲覧経験を与える環境を提案する。動画閲覧経験を拡張するための環境構築にあたり、動画コンテンツのドメイン知識を利用せずにユーザの閲覧スタイルを再利用する手法、およびその結果得られる動画閲覧経験を組み合わせる新たな経験を活用する手法についてケーススタディの結果を交えて述べる。

Enhancing Video Viewing Experience by Reusing Habitual Viewing Behaviors

AKIO TAKASHIMA^{†1} and YUZURU TANAKA^{†2}

Video viewing styles, which are considered to be habitual behaviors during video viewing, are used to externalize one's viewing skills or know-how about video viewing; they allow users to experience videos through these skills. In order to allow users to experience videos in various viewing styles, we have developed a system called the video viewing experience reproducer (VVER), which determines the user's viewing styles and reuses them. This system also allows users to compose Video Viewing Experiences which are brought by Video Viewing Style in order to create new experiences. This paper describes the notion of this reusing and composing, and then reports the results of case studies.

1. はじめに

人間がある状況下で似通った行動を選択する場合、その行動は個人々の経験的知識に基づき選択されたものと考えられる。このような習慣的な行動は、協調フィルタリングに利用されるような一般的な行動もあるが、大半が個人々で特有のものである。21世紀の情報環境においては、計算機が、ユーザの習慣的行動から嗜好を推測したり、他者の習慣的行動を追体験させたり、いくつかの行動を組み合わせさせたりすることで、情報に対する新たな接し方をユーザに提供することが可能となると考えられる。

本研究課題では特に、動画閲覧における個人々の習慣的な閲覧行動 (以下、動画閲覧スタイル (Video Viewing Style)) を題材とし、その閲覧スタイルを活用する情報環境を提案する。たとえば、サッカーの試合を閲覧する場合、サッカーチームの監督が試合を分析するために閲覧する方法と、スタープレイヤーのファンが映像を閲覧する方法とが異なる場合が考えられる。早送りや、スロー再生など、行動に一貫性がある場合、その動画閲覧行動を選択する知識を抽出することで、他者の動画閲覧経験を追体験したり、個人々に特化した動画要約をしたり、さらには複数の動画閲覧の経験方法を組み合わせる新たな閲覧経験を創出したりすることが可能となる。

動画閲覧はこれまで、たとえば映画のように「作品」としての動画を、閲覧者が「消費」という1方向的な関係が主であった。しかし近年では、監視カメラの記録分析、ユーザ観察実験の記録分析、スポーツの試合記録分析など、映像分析の方法によって獲得できる情報が変化するという双方向メディアとしての動画像の利用が注目され始めている¹⁾。このような目的においては、自分や他者の動画閲覧スタイルを再利用することで、効率的、あるいは発見的な情報獲得が期待される。

このように、本研究では動画閲覧スタイルを再利用すること、またその結果得られる動画閲覧経験を組み合わせることにより、ユーザが動画像を様々な経験することができる環境の構築を目指す。個人々の動画閲覧における習慣的な行動は、経験的かつ暗黙的な知識に基づき獲得されるものと考えられるが、そのような知識を正確に記述することは困難である。そのため、本研究では、知識そのものを取り扱うのではなく、経験的知識に基づいて行われた

^{†1} 東京工科大学
Tokyo University of Technology

^{†2} 北海道大学
Hokkaido University

であろう行動と、行動が選択された状況とを関連付けて取り扱うというアプローチをとる。

以下、2章で関連研究について述べた後、3章で動画閲覧スタイルの再利用、4章で動画閲覧経験の組合せの方法について説明する。それらのアプローチに基づいて作成したシステム、Video Viewing Experience Reproducer を用いたケーススタディについて5章で述べ、続く6、7章で考察しまとめる。

2. 関連研究

本章では、関連研究として、嗜好抽出技術、追体験支援技術、知識メディア技術について述べる。

嗜好抽出に基づく情報推薦は、Amazon.com に代表されるオンラインショップのアイテム選択などにおいて広く用いられている。嗜好抽出技術には、ユーザが自身の興味に関する情報を直接入力する直接的手法と、計算機がユーザの行動履歴からユーザの興味情報を推定する間接的手法が利用される²⁾。本研究で着目する間接的手法を利用した例としては、テキストブラウザ上でのマウス操作から、テキストに対する興味を抽出するものや³⁾、セグメント化されたビデオシーンの選択的な閲覧履歴に基づき動画をスキミングするもの⁴⁾などが報告されている。このような手法は、たとえばwebページのリンクをクリックする動作とクリック後に表示されるwebページとの関係のように、推薦される情報が単純で、かつユーザ操作との対応が明確である場合には効果的であるが、スキルやノウハウといった暗黙的な情報を推薦するには向かない。動画閲覧スタイルなどの暗黙的な情報を共有するために、本研究では、科学的な発見は対象知 (knowing what) ではなく方法知 (knowing how) によって起こるという Polanyi の主張⁵⁾ をベースに、暗黙知を暗黙知のまま、つまりはユーザによる動画閲覧スタイルを、動画の自動的な再生として、ユーザに伝えることを試みる。

追体験支援研究において、Nilsson らは、ウェアラブルコンピュータを利用した追体験支援技術を用いることで、ユーザの経験や知識を共有できることを主張しており⁶⁾、本研究の目的もこれに近いものである。小酒井らは他者の経験を実世界に反映させることで自身の行動選択に有効な情報を与えるという主張に加え、追体験コンテンツ (ユーザが作成するアノテーションなど、経験がどのようなものであったかを定義する二次的な情報) を編集・統合する環境を構築している⁷⁾。本研究は、動画閲覧の経験方法自体を編集する点で独創的である。

ユーザ個人や他者の知識を利用するアプローチは、情報推薦や追体験支援研究だけでなく広く行われるものであるが、それらを組み合わせ活用する研究については発展途上であ

り、今後の情報環境において重要な役割を担うものと考えられる。知識メディア技術と呼ばれるこの技術は、Tanaka により、知識を知的リソースとして外在化するもの、またそれらの編集を可能にするものとして定義されている⁸⁾。この定義のもとで、様々なwebサービスを自在に連携させたり⁹⁾、可視化プロセスにおける操作を部品として組み合わせたりする環境が構築されている¹⁰⁾。本研究は、人間が日常生活において獲得するノウハウを、知識メディアによる外在化と編集の対象ととらえるものであり他に類を見ない。詳細について次章に述べる。

3. 動画閲覧スタイルの再利用

3.1 知識メディアとしての動画閲覧スタイル

本研究では、動画閲覧スタイルを知識メディアとしてとらえている。動画閲覧スタイルはすでに述べたように各ユーザの習慣的な閲覧操作であるが、具体的には、早送りや巻き戻し、スロー再生などの速度変化操作を対象としている。本研究では、動画閲覧スタイルに基づき、自動的な速度変化をとともなう動画をユーザが経験すること (動画閲覧経験 (Video Viewing Experience)) により、動画閲覧スタイルを構築するものとなったノウハウを享受できるものとする (図1)。

たとえばサッカーチームの監督が、あるプレイヤーの特徴的なフェイントを繰り返し閲覧することは、その監督が持つ閲覧ノウハウに基づいた動画閲覧スタイルであり、そのスタイルに基づき他の試合の録画ビデオの一部が自動的に同様に繰り返し再生されることは、監督自身 (あるいは他者) が、監督の閲覧ノウハウを新たな動画表現として経験するものといえる。

3.2 動画閲覧スタイルの抽出

本研究では、動画閲覧スタイルを、動画の特徴量とユーザによる速度変化操作とを対応付けることで構築する。

動画の特徴量には、画像情報や音情報などのローレベルな特徴のみを考慮し、意味的内容を含めないこととした。コンテンツに基づく動画要約などの既存の研究では、動画の意味的内容を積極的に取り入れる研究も多いが^{11),12)}、これらはある種のドメイン知識を考慮したものである。たとえばサッカーのビデオにおいてシュートシーンが重要であるとか、喚声の大きいシーンが重要である、という知識に基づくことで、ある程度一般的に許容されうる動画要約を行うことは可能である。しかし本研究で対象とする動画閲覧スタイルは、個人々によって様々なものであり、たとえばシュートシーンよりも中盤の攻防が重要であるとする閲覧スタイルも考えられる。対象とする動画を特定のジャンルに制限したとしても、

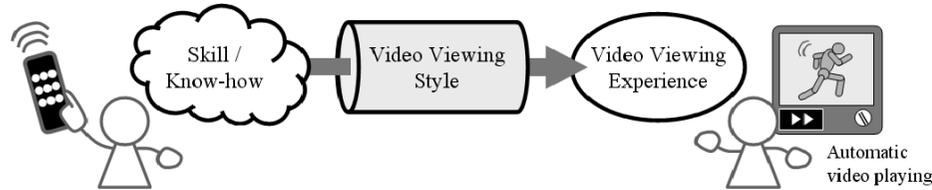


図 1 知識メディアとしての動画閲覧スタイル
Fig. 1 Video viewing style as knowledge media.

個々人でとらえ方のことなる動画の意味的内容を、すべてドメイン知識としてあらかじめ準備しておくことはほぼ不可能である。どのシーンにも存在するローレベルな情報を利用することで公平性を保つと考え、意味的な内容を排除した。

ユーザの動画像に対する操作は、これまでの VCR (Video Cassette Recorder) を踏襲したものにとどまらず、映像のサイズや速度を細かく変化させたり、映像の一部分だけ速度を変化させたりする操作が、研究としては可能になってきている¹³⁾。本研究でも、新たな動画閲覧経験を提供するという点で、様々なユーザ操作や視線追跡データなどを、閲覧スタイル抽出に利用することを視野に入れているが、現状では早送りやスロー再生などの VCR で利用されてきたシンプルな速度変化のみを対象としている。詳細について次節に述べる。

3.3 Video Viewing Experience Reproducer

上述のアプローチに基づき、ユーザの動画閲覧スタイルを抽出し動画閲覧経験を提供するシステム、Video Viewing Experience Reproducer (以下 VVER) を構築した (図 2)。VVER は Association Extractor と Behavior Applier からなる。Association Extractor では、まず動画の特徴量を抽出し、どのような動画像特徴量を持つ映像フレームがどのように操作されるかを分類する分類器を作成する。Behavior Applier は、Target Video の各映像フレームの動画特徴量を算出し、その値を分類器にかけ、それぞれのフレームをどのように再生するかを決定する。以下に詳細を述べるとともに、システム利用プロセスについて説明する。

Association Extractor

Association Extractor では、ユーザによる複数の Training Video の閲覧ログと各ビデオの動画特徴量とを用いた教師付き学習により分類器が作成される。

動画特徴量としては、動画の各フレームにおける以下の 5 種類の特徴量抽出を行い、計 84 次元の特徴量を利用している。

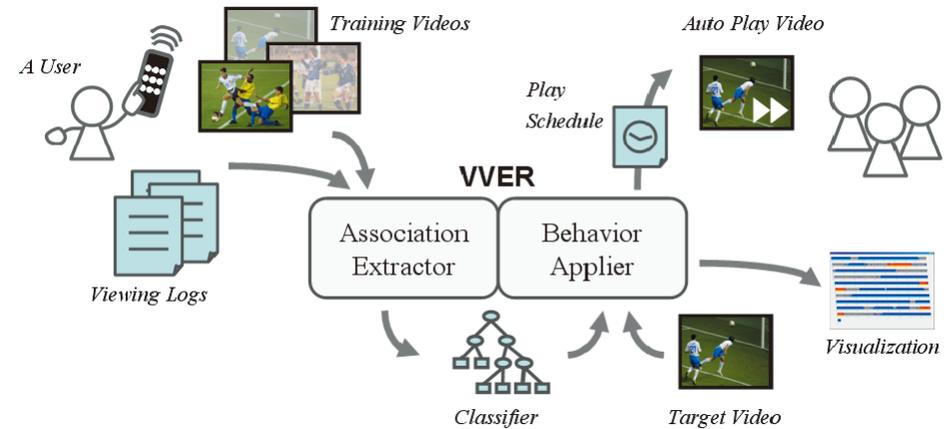


図 2 Video Viewing Experience Reproducer および動画閲覧スタイル再利用プロセス概観
Fig. 2 The Video Viewing Experience Reproducer and the overview of reusing viewing style.

- フレーム内に現れる色の統計値
- 最頻色の空間的な中心や分散
- オプティカルフローデータ
- 動作中のオブジェクトの数
- 音域の分布情報

フレーム内に現れる色の平均値や分散は、最もシンプルな動画フレームの情報と考えられ、HSV 表色系での値を利用している。最頻色の空間的な中心や分散は、たとえばサッカーのフィールドがどの程度映像に含まれるかを知る一手段となる (図 3 (a))。現状では、最頻色の指定は、色空間における特定の範囲 (e.g. $30 < H \leq 50, 80 < S \leq 90, 75 < V \leq 95$) を手動で設定しているが、あらかじめ学習に使う全動画における最頻色を自動的に抽出して利用することを検討している。オプティカルフローデータは、まず映像フレームを 12 分割し、各領域に含まれる特徴点を抽出する。その点が次フレームにおいてどう変化するかを算出し、各領域における変化量と方向の平均や分散値を特徴量として利用している (図 3 (b))。動作オブジェクトの数は、CAMSHIFT 法¹⁴⁾を用いて抽出した動作オブジェクトを、その大きさに応じて 3 種類に分類してカウントした (図 3 (c))。音域の分布は、対象フレームにて再生される音がどの音域 (e.g. 0-32 Hz, 32-64 Hz, ..., 8,000-10,000 Hz) に分布しているかを測定している (図 3 (d))。

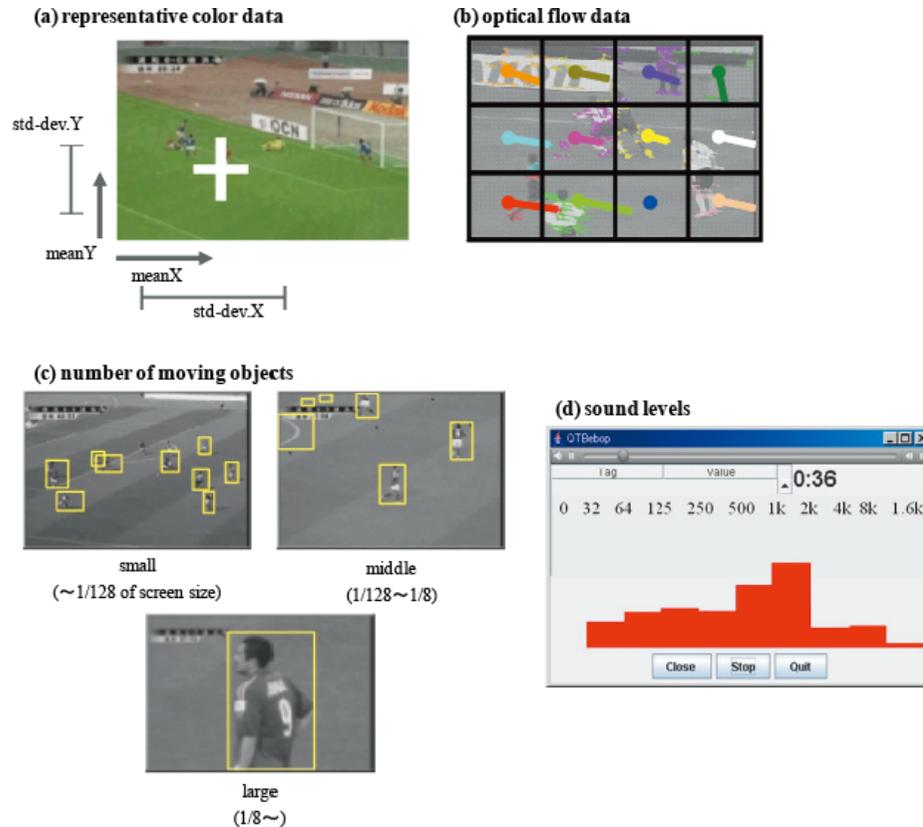


図 3 動画特徴量の種類 (著作権の問題から、一部モザイクを施している)
Fig. 3 Example of the types of video feature.

一方ユーザ操作としては、簡単のため、現状では、1倍速よりも早い速度で再生する Skip、あるシーンを巻き戻して1倍速以下のスピードで見直す Re-examine の2種類に限定し、通常の速度で再生する Normal を合わせて計3種類とした。

分類器作成には WEKA システム¹⁵⁾ を使用し、上述の84次元の動画特徴量を説明変数、3種類のユーザ操作を目的変数とした。分類器作成アルゴリズムとしては、情報利得比が最大となる変数を順次選択して決定木を作成する C4.5 アルゴリズム¹⁶⁾ を採用しているが、これに限定するものではない。様々なアルゴリズムを選択できるようにシステムを拡張する予定

である。

Behavior Applier

Behavior Applier は、Target Video の動画特徴量を Association Extractor と同様に計算し、それを分類器にあてはめることで各フレームをどのような速度で再生するかを記した Play Schedule を作成する。Target Video は Play Schedule に基づいて自動的に速度を変えながら再生される。現在の実装においては、Skip と判断されたフレームを5倍速で、Re-examine と判断されたフレームを0.5倍速で、それ以外のフレームを1倍速で再生するようマッピングしているが、これらの速度の決定はユーザが自由に変更できるよう改良する予定である。また、Behavior Applier は、Play Schedule におけるノイズ (Skip と判断された連続するフレームの中に1フレームだけ Re-examine と判断されたフレームが存在する場合など) を除去する機能と、Play Schedule を確認する可視化機能を有している。

このように VVER では、Training Video を用いてあらかじめ分類器を作成し、それに基づき Target Video を自動的に再生することで、ユーザに動画閲覧経験を与えている。

4. 動画閲覧経験の組合せ

本研究では、特定の動画閲覧スタイルに基づいた動画閲覧経験をユーザに与えるだけでなく、他の動画閲覧スタイルに基づく動画閲覧経験を複数組み合わせることで、新たな動画閲覧経験を創出する環境を提供する。実際には、動画の自動再生時における速度変化を記した Play Schedule を組み合わせることで、新たな動画閲覧経験をユーザに与えるものである。組合せ方法として $A \cap B := \{x | x \in A \text{ and } x \in B\}$, $A \setminus B := \{x | x \in A \text{ and } x \notin B\}$, $A \cup B := \{x | x \in A \text{ or } x \in B\}$ という3種類を定義する。ここで A, B は、あるビデオに含まれる全動画フレームがそれぞれどのようなユーザ操作 (Skip, Re-examine, Normal) に関連付けられるかを表す集合であり、A と B 2人のユーザの集合を示している。x はそれらの集合に含まれる特定のフレームにおける関連を示す。ユーザ操作の関連を組み合わせる例としては以下のようなものが考えられる。

$$S_{USER1} \cap S_{USER2} \cap S_{YOU} \tag{ex.1}$$

$$(S_{USER1} \cap S_{USER2}) \setminus S_{YOU} \tag{ex.2}$$

$$(S_{USER1} \cap S_{USER2}) \cup S_{YOU} \tag{ex.3}$$

図4上部に Play Schedule の例を、下部に ex.1 ~ ex.3 で示された組合せの例を示す。図の帯状のグラフが Play Schedule を表し、グラフの左端が Target Video の最初のフレーム、右端が最後のフレームに相当する。動画閲覧スタイルに基づき、Target Video がどのよう

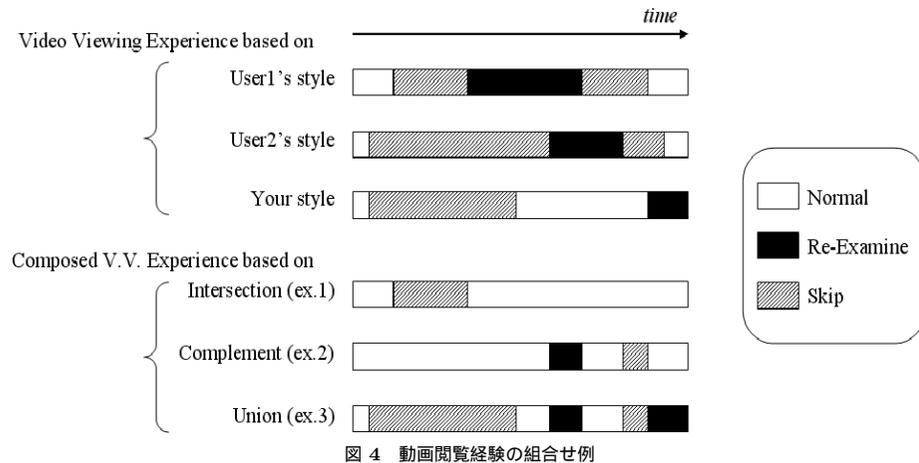


図 4 動画閲覧経験の組合せ例

Fig. 4 Examples of video viewing experience and their composition.

な速度で再生されるかを推測した結果がグラフの色で表現されている。

ex.1 は、すべてのユーザに共通する動画閲覧スタイルを抜き出す操作である。図の例では、すべてのユーザが前半部分にあるシーンをスキップすると推測されている。より多くのユーザに共通する閲覧スタイルを抜き出すことで、一般的に許容される動画閲覧経験を提供することができる。

ex.2 は S_{USER1} と S_{USER2} の動画閲覧スタイルから共通する部分を抜き出し、 S_{YOU} の閲覧スタイルとの差分をとる操作である。これは、 S_{YOU} 以外のユーザは共通して行いがちなユーザ操作、かつ S_{YOU} は行わないユーザ操作を経験させるものである。これにより、自分は知らないが、一般的に行われがちなユーザ操作に気づくことができると考えられる。

ex.3 では、 S_{USER1} と S_{USER2} の動画閲覧スタイルから共通する部分を抜き出し、 S_{YOU} の閲覧スタイルとの和をとることで、 S_{YOU} の閲覧スタイルをベースにしながらも、他者に共通するユーザ操作を取り入れることが可能となる（ただし Union では、Skip, Re-examine などの操作に優先順位を定義しておく必要がある）。

5. ケーススタディ

VVER を利用して行った 2 種類のケーススタディについて以下に述べる。なお、ケーススタディに用いた動画コンテンツは 3 種類で、サッカーおよび相撲のテレビ中継と、建物の

エントランスに設置された監視カメラ映像である。監視カメラの映像は、パンやズームなど編集がなされていない固定カメラで記録したものであり、映像分析の用途を想定することができる。テレビ中継は、編集者の意図が含まれたカメラ割りやシーン構成となっているが、ハードディスクレコーダの普及が進んでいることを背景に、閲覧スタイルを一般人が日常的に活用することを想定し、対象とした。しかし、ドラマや映画など、作品性の強いものは、本研究の対象としては十分ではないと考え、ユーザ操作が現れそうなスポーツ映像を選択している。

以下に述べるケーススタディでは、サッカーと相撲に関して、複数のゲームや取組からランダムに 5 分間のビデオをそれぞれ 8 本切り出した。ただし放送局は 1 局に限定している。監視カメラについては同日に撮られた映像から、5 分間の動画像をランダムに 8 本切り出した。各 8 本の動画のうち、5 本を Training Video として分類器の作成に用い、3 本を Target Video として利用した。被験者は 4 人で、全員、普段ビデオデッキや DVD プレイヤなどを介した映像に触れる機会を持っていた。

なお、現状のシステムでは、特徴量抽出と分類器作成は統合されていない。実験においては、 320×240 サイズの Training Video と Target Video に対し、特徴量抽出にそれぞれ 7 分程度、5 本のビデオを用いた分類器作成に 10 分程度の時間を要した（CPU: Intel Core2Duo 2.33 GHz, メモリ: 2 GB）。

5.1 動画閲覧スタイルの再利用

ユーザが、自身の動画閲覧スタイルに似通った動画閲覧経験を得ることを目的として、動画閲覧スタイルを再利用する実験を行った。

前節で述べた 3 種類の動画コンテンツそれぞれについて、Training Video を被験者に各 3 分間で閲覧してもらい、コンテンツごとに分類器を作成した。表 1 に、各ユーザのコンテンツに対する興味（5: 非常にある, 3: どちらでもない, 1: まったくない）と、観察された閲覧スタイルを示す。なお、表中の閲覧スタイルの記述における Oop (Out of Play シーン) は、サッカーではボールが外に出てプレイが止まっている状態、相撲では立ち会い前の仕切りなどが映されているシーンを指す。Ip (In play シーン) はゲームや取組がまさに行われている状態のシーンを指し、Rp (Replay シーン) はスローモーションや通常速度でのリプレイシーンを指す。

サッカーと相撲の動画においては、全被験者に共通する閲覧スタイルとして、Oop を早送りして Skip し、Rp を通常速度で再生しがちであることが観察された。監視カメラ映像では全被験者が、人が映っていないシーンを Skip した。コンテンツの興味と動画閲覧スタ

表 1 ユーザごとのコンテンツに対する興味と動画閲覧スタイル
Table 1 The profile and video viewing style of each user.

	コンテンツに対する興味 (5:非常にある 1:まったく無い)			動画閲覧スタイル		
	サッカー	相撲	監視カメラ	サッカー	相撲	監視カメラ
SubA	4	1	1	Oop を Skip, Rp を Normal, Ip のうち、中盤の攻防を Skip, シュートシーンを Re-examine	Oop を Skip, Rp を Normal	人がいない場面を Skip, 人がいるシーンを Normal
SubB	2	1	2	Oop を Skip, 稀に Rp を Skip	Oop を Skip, Rp を Normal	人がいない場面を Skip, 人がいるシーンを Normal
SubC	5	3	1	Oop を Skip, Rp を Normal, Ip のうち、中盤の攻防を Skip, ゴールシーンを Re-examine ファウルシーンを Re-examine	Oop を Skip, Rp を Normal	人がいない場面を Skip, 人がいるシーンを Normal または Re-examine
SubD	4	4	1	Oop を Skip, Rp を Normal, 稀にゴールシーンを Re-examine	Oop を Skip, Rp を Normal 解説者が話すシーンを Normal,	人がいない場面を Skip, 人がいるシーンを Normal または Re-examine

※Oop: Out of Play シーン, Ip: In play シーン, Rp: リプレイシーン

イルの間に明確な関連があるとはいえないが、すべてのコンテンツにあまり興味を持たない SubB は Re-exam をほとんど行わなかった。また、サッカーの閲覧に非常に興味があると答えた SubC は、他の被験者と異なりファウルシーンを Re-exam する様子が観察されており、個々人が他者と異なる閲覧スタイルを持つことが確認された。

各被験者は Training Video の閲覧をした後日、自分自身の閲覧スタイルに基づき自動的に再生される Target Video を、各コンテンツにつき 3 本ずつ、計 9 本閲覧した。閲覧後に、「自分の閲覧スタイルに基づき自動再生される動画像を見てどう感じたか」「その他の参考意見」について質問をしたところ、全員が自分の行った閲覧スタイルに近いと感じたと述べた。

被験者へのインタビューに加え、閲覧スタイルがどの程度再現されているかを確認する定量的実験を行った。これは、人間の、強要された動画閲覧スタイル (Intentional Viewing Style) を介した動画閲覧経験と、正解セットを利用して学習させた分類器 (Intentional Classifier) を介した動画閲覧経験とが、どの程度似ているかを探るものである。

実験手順としては、まず正解セットとなる動画シーンと閲覧操作の関連を定義する。たとえば監視カメラの場合、人が 1 人で映っている場面は Normal、2 人以上が映る場面は Re-exam、それ以外の人々が映らない場面は Skip とそれぞれ関連付ける。Training Video のシーンを手作業で分析し、各シーンを上述のとおり関連付け、Viewing Log (図 2) の正解セットを作成する。この正解セットをもとに作成された分類器を Intentional Classifier と

呼び、Target Video の自動再生に利用する。一方人間の閲覧スタイルについては、その閲覧の方略に明確な意図があると仮定し、上記の正解セットを作るのと同様の関連付けに基づいて、Target Video の特定のシーンを特定の操作で閲覧する Intentional Viewing Style で動画を閲覧することを被験者に強要する。なお、比較の基準として、Target Video についてもシーン分析を手作業で行い、正解セットを準備する。

理想的には、Target Video を Intentional Classifier を介して見た結果 (以下 ClassifierSet) も、Intentional Viewing Style を介して見た結果 (以下 HumanSet) も、Target Video の正解セット (以下 CorrectSet) と同様の結果になるべきである。ClassifierSet および HumanSet が CorrectSet からどの程度のずれを生じるかを算出するために、各ユーザ操作に値を設定した。具体的には Re-exam は 1、Normal は 0、Skip は -1 としている。これは、ユーザの動画フレームに対する注目度 (Re-exam は注目度が高く、Skip は低い) を考慮したものである。たとえばあるフレームが、CorrectSet で Re-exam と関連付けられている際、誤って Skip として推測する場合と、Normal として推測する場合に差をつけることが可能となる。なお、ClassifierSet と HumanSet それぞれの CorrectSet に対するずれは、以下の式で定義した。

表 2 コンテンツごとのエラー値
Table 2 Error value of each content.

	ClassifierSet	HumanSet_SubA	HumanSet_SubB	HumanSet_SubC	HumanSet_SubD
サッカー	0.142	0.201	0.098	0.120	0.103
相撲	0.090	0.028	0.089	0.135	0.083
監視カメラ	0.049	0.060	0.039	0.072	0.034

$$error_{classifier} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m (A_{ij} - C_{ij})^2$$

$$error_{human} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m (A_{ij} - H_{ij})^2$$

なお n は評価に用いた Target Video の総数, m は各 Target Video のフレームの総数, A は CorrectSet, C は ClassifierSet, H は HumanSet を表す. コンテンツごとに 3 本の Target Video を用いた評価結果を表 2 に示す. 値は前述したユーザ注目度に応じたユーザ操作間の距離に相当する. 結果から, ClassifierSet におけるエラー値が HumanSet におけるエラー値と同程度の値を示していることが分かる. これは, ユーザが一貫した閲覧スタイルを行う場合, VVER を利用することで, 人間が行うのと同程度の閲覧経験を再現できることを意味している. また同時に, ユーザは特定のシーン (人が存在する, ゴールが決まる) に対して操作を行っているが, システムは, シーンのドメイン知識を利用することなく, 動画特徴量だけを用いてユーザ操作を再現したともいえる.

5.2 動画閲覧経験の組合せ

動画閲覧経験を組み合わせて作る新たな動画閲覧経験が, ユーザにどのような印象を与えるかを調査するため, 前節で作成した各被験者の動画閲覧スタイル, Target Video を用い, 同じ 4 被験者にインタビューを行った. 組合せ操作としては下記の 3 種類とした. なお, cmp.2 の記号の添え字は, SubD が閲覧する場合のものになっている.

$$S_{SubA} \cap S_{SubB} \cap S_{SubC} \cap S_{SubD} \quad (\text{cmp.1})$$

$$(S_{SubA} \cap S_{SubB} \cap S_{SubC}) \setminus S_{SubD} \quad (\text{cmp.2})$$

$$S_{SubA} \cup S_{SubB} \cup S_{SubC} \cup S_{SubD} \quad (\text{cmp.3})$$

実験では前節で述べた 9 本の Target Video それぞれにおいて, 上記の 3 種類の組合せを

行い, 動画が自動で再生される様子をユーザに見てもらい, 「気がついた点, 感想」を自由に述べてもらった.

cmp.1 では, 全員が共通する操作を行うシーンにその操作を適用し, 共通しない部分は Normal として再生する. 表 1 にも示したように, 相撲と監視カメラの映像における Skip 行動が, 全被験者に共通していた. 結果として, どの被験者も自分の閲覧スタイルとの違いが分からないという意見を述べた. サッカーの映像に対して被験者 A は, 「中盤のシーン (被験者 A が Skip しがちな, 両ゴールエリアが現れない In Play シーン) が普通のスピードで再生されるものの, 特に問題はない」と述べた.

cmp.2 は, 4 章で述べた ex.2 のように, 一般的に行われることが多いものの, 自分が行わない操作を経験できる組合せである. 被験者 B は全コンテンツを通して Re-exam をほとんど行っていないが, この組合せの結果, 他の 3 人の閲覧スタイルに沿ってサッカーのシュートシーンがスローで再生されることに興味を持った. なお, cmp.2 では他者と同じ操作をする場合は, 1 倍速で再生される設定であったため, 全員がスポーツ映像の Out of play シーン, および監視カメラに何も映らないシーンが早送りされないことに苛立っていた.

cmp.3 は, 被験者すべての操作を再現しようとするものであるため, だれか 1 人の閲覧スタイルが, 対象フレームを Re-exam や Skip すべきと判断した場合, それらが選択され, 速度変化の激しい自動再生が提示された (なお, コンフリクトが生じる場合には Re-exam, Skip, の順で優先順位を定めていた). 結果として, 被験者 B と C は, 特にサッカー映像について, ダイジェスト版の映像を見ているようだと言った. 監視カメラの映像において, 被験者 A は, 一部の人間の出入りがスローで再生された際, 特徴的な歩き方だと言って見入っていた.

6. 議論・考察

前章のケーススタディの結果に基づき, 動画閲覧スタイルの再利用, その結果得られる閲

覧経験の組合せについて考察する。

動画閲覧スタイルの再利用

前章で述べたケーススタディでは、被験者の閲覧スタイルを再利用して Target Video を同様のスタイルでユーザに経験させることができることを確認した。ドメイン知識をいっさい使わず、ユーザにとってある程度満足のいく結果が得られた点は、動画の利用方法が多様化していくであろう今後の情報環境において、重要な意味を持つと考えられる。

ここで、良い結果を得るには、ある程度ジャンルを特定した動画像を用意し、かつユーザが一貫した行動をとる必要がある。今回のケーススタディでは、サッカーと相撲、監視カメラの3種類の映像それぞれに分類器を作成し、動画閲覧スタイルを再利用することを可能にした。補足実験として今回抽出した動画閲覧スタイルの一部を別のコンテンツに適用することを試みた。サッカーおよび相撲の閲覧をもとに作成した閲覧スタイルは、互いに影響を与えたが、ユーザにとって意味のある自動再生をコンスタントに得るには至らなかった。監視カメラで得られた閲覧スタイルは、サッカーと相撲の映像に対しては、すべてを Skip してしまう場合などがあり、こちらも十分に機能していない結果となった。動画のジャンルに応じてユーザの操作が変化することも報告されており¹⁷⁾、閲覧スタイルを学習する際には、動画に付与されるメタデータなどの情報を利用し、ある程度ジャンルを絞る必要があると考えられる。

動画閲覧経験の組合せ

Play Schedule を組み合わせ、新たな動画閲覧経験を提供することについて、4章で組合せの例を述べ、5.2節でケーススタディの結果を示した。ケーススタディでは、一部の被験者において思いがけず興味を持てる経験をした様子が見られた。一方で、各被験者に共通して、特に emp.2 の組合せ結果において、自分自身の閲覧スタイルとは異なる閲覧経験が提示された際、ネガティブな印象をいただくことがあった。これは組み合わせる場合だけでなく、他者の閲覧スタイルに基づく経験をそのまま享受するときにも起こりうるものである。動画閲覧経験がユーザにとって思いがけず面白いものと感じられるか、苛立ちを覚えさせるものになるかを事前に判断することは容易ではないが、ユーザがある程度それを把握できる仕組みが必要と考える。VVER における Play Schedule を可視化する機能などを利用し、動画閲覧経験をユーザが把握できるような仕組みを取り入れることで、ユーザが閲覧経験をイメージしながら組合せを試行錯誤できるかもしれない。

組合せ後の閲覧経験の再利用

本稿では閲覧経験の組合せに関して、Play Schedule を組み合わせ、新たな動画閲覧経

験を提供することについて述べたが、この組合せは Play Schedule 作成に使われた Target Video にしか適用できない。つまり Target Video が変わるたびに組合せを計算しなおす必要が生じる。これを解決するためには、閲覧経験を組み合わせた結果を、スキル、ノウハウとして活用する必要がある。組合せ後の閲覧経験 (= Play Schedule) を機械学習のためのユーザ操作ログととらえ、再び分類器の学習に利用することでこれを実現できると考え、現在システムの改良を進めている。

7. 結論・展望

本研究では、動画閲覧のスキルやノウハウといった暗黙的な知識を、「動画閲覧スタイルの再利用」および「動画閲覧経験の組合せ」によってユーザが活用するための環境を構築した。

本研究課題の特色は、動画像に対する習慣的な体験方法である閲覧スタイルを、共有・再利用の対象としてとらえ、知識活用の支援を試みる点にある。人間の行動とその行動に至った状況とを細かく対応付けることができれば、逆に、ある状況に直面した人間に対し、どのような行動をとるべきか、どのような行動の選択肢が可能であるかを提示することができる。行動主義的立場でユーザの要求を完璧に予測し満たすことは容易ではないが、ユーザの意思決定を補助したり、新たな経験を与えることでユーザの興味を広げたりすることが可能となるであろう。

ストレージデバイスやセンサ類の発展、ユビキタス環境実現の取り組みを背景に、日常生活における人間の様々な行動履歴を蓄えておくことが可能になりつつあり、さらに今後は行動履歴だけでなく、その行動にいたった状況をもセットで蓄積することが可能となると考えられる。つまり状況と行動との関連付けを利用することで、たとえば料理のコツや、マシンの修理など、日常生活の様々な行動のもととなる知識を活用することが可能となり、知識管理の新たなアプローチとなりうると期待できる。本研究はそのアプローチを動画閲覧というドメインで行う意義深いものである。

参考文献

- 1) Yamamoto, Y., Nakakoji, K. and Takashima, A.: The Landscape of Time-based Visual Presentation Primitives for Richer Video Experience, Human-Computer Interaction, *INTERACT 2005*, Costabile, M.F. and Paterno, F. (Eds.), Rome, Italy, pp.795-808, Springer (Sep. 2005).
- 2) 土方嘉徳: 情報推薦・情報フィルタリングのためのユーザプロファイリング技術, 人

- 工知能学会誌, Vol.19, No.3, pp.365-372 (2004).
- 3) Sakagami, H. and Kamba, T.: Learning Personal Preferences on Online Newspaper Articles from User Behaviors, *Proc. 6th International World Wide Web Conference, In Computer Networks and ISDN Systems*, Vol.29, pp.1447-1456 (1997).
 - 4) Yu, B., Ma, W.-Y., Nahrstedt, K. and Zhang, H.-J.: Video Summarization Based on User Log Enhanced Link Analysis, *ACM MM '03*, pp.382-391 (2003).
 - 5) Polanyi, M.: *Tacit Dimension*, Peter Smith Pub Inc. (1983).
 - 6) Nilsson, M., Drugge, M. and Parnes, P.: Sharing experience and knowledge with wearable computers, *Pervasive 2004: Proc. Workshop on Memory and Sharing of Experiences* (2004).
 - 7) 小酒井一稔, 長尾 確: 個人用知的移動体における追体験支援システム, 第19回人工知能学会全国大会論文集, 3A3-06 (2005).
 - 8) Tanaka, Y.: *Meme Media and Meme Market Architectures: Knowledge Media for Editing, Distributing, and Managing Intellectual Resources*, IEEE Press (2003).
 - 9) 藤間 淳, 田中 譲: Web文書の直接編集による複数アプリケーション間の連携と情報探索シナリオの多重化, 電子情報通信学会論文誌 D, Vol.J89-D, No.12, pp.2612-2624 (2006).
 - 10) 杉淵剛史, 田中 譲: 関係データベースモデルに基づくデータベース可視化フレームワークの提案と実装, 電子情報通信学会論文誌 D, Vol.J90-D, No.3, pp.918-932 (2007).
 - 11) Nakamura, Y. and Kanade, T.: Semantic analysis for video contents extraction—spotting by association in news video, *Proc. 5th ACM International Conference on Multimedia*, Seattle, pp.393-401 (1997).
 - 12) Ekin, A., Tekalp, A.M. and Mehrotra, R.: Automatic soccer video analysis and summarization, *IEEE Trans. Image Processing*, Vol.12, No.7, pp.796-807 (2003).
 - 13) Takashima, A., Yamamoto, Y. and Nakakoji, K.: A Model and a Tool for Active Watching: Knowledge Construction through Interacting with Video, *Proc. INTERACTION: Systems, Practice and Theory*, Sydney, Australia, pp.331-358 (2004).
 - 14) Bradski, G.R.: Computer vision face tracking as a component of a perceptual user interface, *Workshop on Applications of Computer Vision*, pp.214-219 (1998).
 - 15) WEKA. <http://www.cs.waikato.ac.nz/ml/weka/>
 - 16) Quinlan, J.R.: *C4.5: Programs for machine learning*, Morgan Kaufmann Publish-

ers, CA (1993).

- 17) Li, F.C., Gupta, A., Sanocki, E., He, L. and Rui, Y.: Browsing digital video, *Proc. CHI 2000*, ACM, pp.169-176 (2000).

(平成19年10月11日受付)

(平成20年4月8日採録)



高嶋 章雄 (正会員)

東京工科大学コンピュータサイエンス学部助教。1999年九州工業大学情報工学部知能情報工学科卒業。2003年奈良先端科学技術大学院大学情報科学研究科博士課程修了。博士(工学)。2004年8月から2005年3月まで東京大学先端科学技術研究センター学術研究支援員。2005年から北海道大学ベンチャービジネスラボラトリ学術研究員, 2008年より現職。HCI,

情報可視化等の研究に従事。



田中 譲 (正会員)

北海道大学大学院情報科学研究科教授。1950年京都に生まれる。1974年京都大学大学院工学研究科電子工学専攻修士課程修了。工学博士(東京大学)。1970年代にはデータベース理論の研究に従事し, 1970年代後半から1980年代前半にデータベース・マシンの研究に従事。1985年10月より1年間, IBMワトソン研究所客員研究員。1987年より知識メディアの研究を開始し現在に至る。1990年より北海道大学大学院工学研究科教授, 2004年に機構改革により現職。1996年より北海道大学知識メディアラボラトリ長。1998年から2000年京都大学社会情報学専攻教授を併任。2004年より国立情報学研究所客員教授。『Meme Media and Meme Market Architectures』(IEEE Press & John Wiley, 2003)等の著書あり。