時間制約に基づく時系列パターンマイニング手法を用いた映像からのイベント抽出

Extracting Semantic Patterns in Video Using Time-constrained Sequential Pattern Mining

白 浜 公 章[†], 出 野 構 一[†], 上 原 邦 昭[†]

Kimiaki Shirahama[†], Koichi Ideno[†] and Kuniaki Uehara[†]

Abstract We developed a technique for video data mining to extract ‘semantic patterns’ associated with semantically relevant events in videos. First, several types of raw-level metadata are derived from the raw video data in each shot. The metadata is then sequentially aggregated into a multistream. Then, sequential patterns, each of which is a temporally ordered set of raw-level metadata, are extracted from the multistream. The sequential patterns are reduced to likely semantic patterns using two types of time constraints that we introduce: ‘semantic event boundaries’ and ‘temporal localities’. We developed a prototype system and demonstrated the effectiveness of extracted semantic patterns.

キーワード：ビデオデータマイニング，データマイニング，時系列パターン，時間制約，メタデータ，イベント

1. ま え が き

近年のマルチメディア技術の進歩により、大量のデジタル化された映像が、個人のパソコンやデータベースに蓄積されている。こうした環境の中で、ユーザが効率的に映像を視聴するためには、意味内容に基づいて映像を処理することが重要になってきている。具体的には、動画像、音声、テキストなど複数のメディアから得られた低次の情報を、高次の意味内容へと統合するための知識が必要である¹⁰⁾。そこで、データマイニング技術を用いて映像から有用な知識を発見する‘ビデオデータマイニング’が注目されている。

データマイニングとは、膨大なデータから過去には知られていなかった、興味深いパターンを発見する技術である。例えば、マーケティングのための顧客の購買パターン、ネットワークへの不法侵入防止のためのアクセスパターンのように、発見されたパターンはユーザに利益をもたらすと言われている。これにならうと、本論文のビデオデータマイニングは、映像からイベント（例えば、会話や戦闘などの意味的にまとまりのある区間）を表現するための‘意味的なパターン’を発見し、意味内容に基づいた映像検索のための有用な知識を発見することになる。

映像の意味内容は、‘空間的特徴’、‘時間的特徴’と呼ばれる二種類の特徴をもっている。空間的特徴とは、単一のビデオフレームによって表される意味内容で、ビデオフレームに映し出されている登場人物、オブジェクト、場所などが挙げられる。時間的特徴とは、時間的に連続するビデオフレームから構成されるシーケンスによって表される意味内容で、登場人物の行動、オブジェクトの動き、場所の変化などが挙げられる。ただし、デジタル化された映像から直接的に空間的特徴や時間的特徴を得ることはできないため、前処理せずに意味的なパターンを抽出することは困難である。そこで、まず映像から空間的特徴、時間的特徴を特徴づける‘メタデータ’を抽出しなければならない。そして、抽出されたメタデータにデータマイニング技術を適用して意味的なパターンを発見することになる。

本論文では、以下のようなアプローチをとる。まずショットごとに様々な種類のメタデータを抽出し、時間順に並べてマルチストリームを作成する。そして、マルチストリームから、意味的なパターンを時系列パターンとして抽出する。この際、不適切な時系列パターンが抽出されることを防ぐために、映像に特有の時間制約を用いた時系列パターンマイニング手法を導入する。最終的に、抽出された意味的なパターンを用いたイベント検索の性能を評価する。

2. 関 連 研 究

本章では、ビデオデータマイニングアプローチの特徴を、映像処理とデータマイニングの二つの観点から述べる。映

2005 年 12 月 14 日, MIPR 2005 で発表

2006 年 1 月 20 日受付, 2006 年 6 月 12 日再受付, 2006 年 7 月 7 日採録

[†] 神戸大学 大学院 自然科学研究科

(〒 657-8501 神戸市灘区六甲台町 1-1, TEL 078-803-6220)

[†] Graduate School of Science and Technology, Kobe University

(1-1, Rokkodai-cho, Nada-ku, Kobe, 657-8501, Japan)

像処理の観点からは、映像の意味内容に基づく検索、ブラウジング、要約などのアプリケーションが開発されている。しかしながら、既存のアプリケーションでは、ニュース映像¹¹⁾、スポーツ映像⁶⁾、固定された監視カメラからの映像⁹⁾のような、‘ルールをもつ (rule-dependent)’ 映像を対象としている。例えば、ニュース映像では、インタビューする人とインタビューされる人のショットが交互に繰り返されるという、インタビューイベントに特有のルールがある。また、スポーツ映像では、観客の拍手や歓声が起こった後にテロップのスコアが変化するという、ゴールイベントに特有のルールがある。さらに、固定監視カメラからの映像では、オブジェクトが激しく動いているときは隣接フレーム間の差分が大きくなるという、異常な状況を発見するためのルールがある。このように、ルールをもつ映像を対象とした場合、ルールに従ってイベントを検出しさえすれば、意味内容に基づいた処理を容易に実現できる。

これに対して、本論文では、‘ルールをもたない (rule-independent)’ 映像を意味内容に基づいて処理することを目的としている。例えば、映画の戦闘イベントについて考えると、出現する登場人物、使用する武器、場所、時刻、天気など様々な要因によって、幾通りもの戦闘イベントが想定される。このため、意味内容に基づいて処理するためには、どのような特徴量 (メタデータ) がイベントを特徴づけるために有用かを知ることが不可欠である。そこで、画像、映像、音声処理の分野で有用性が認められたメタデータを用いて、ルールをもたない映像からの意味的なパターン抽出を試みる。

これまでに、映画を対象としたいいくつかのアプローチが提案されているものの、フレームワークが提案されているのみであったり¹⁰⁾、メタデータを抽出するのに多大な人手を要する¹⁾ など、実用的なアプローチは報告されていない。また、映画には映像文法と呼ばれる編集ルールが存在するが、編集ルールは編集者の意図や考えを正確に表現するためのシンタックスのみを扱ったものであり、イベントを表現するためのルールではない¹⁾。

一方、データマイニングの観点からは、有限種類のシンボルで定義されたストリームからの時系列パターン抽出が研究されている。この研究では、いかにして巨大な探索空間から効率よくパターンを抽出するかが問題となる。以下では、ストリームの次元数による探索空間の大きさと探索空間を削減するための手法について概説する。

今、1次元ストリームと m 次元マルチストリームがあると仮定する。さらに、1次元ストリーム、マルチストリームの各ストリームは n 種類のシンボルで構成されているとする。ここで、時間長が k のパターンを考えると、1次元ストリームでは n^k 種類、マルチストリームでは n^{mk} 種類ものパターンが存在する可能性がある。このように、ストリームの次元数が増えるほど、探索空間は急激に巨大化することがわかる。この点を考慮して、文献5)では主成分分

析 (PCA) を用いて、文献6)ではマルチストリームの時間的順序を保存したままシンボルを並び替えて、1次元ストリームに変換してからパターンを抽出する手法が提案されている。

また、探索空間を削減する手法として、主に‘モデルアプローチ’と‘アプリアプローチ’がある。モデルアプローチは、抽出するパターンを事前に指定して探索空間を削減する手法である。代表的なモデルアプローチとして、固定時間幅のウィンドウを用いてパターンの時間長を指定する手法³⁾、最小記述長原理などの評価基準を用いてパターンの最適時間長を指定する手法⁵⁾、(パターンに含まれるシンボル数)+1個の状態をもつ隠れマルコフモデル (HMM) を用いてパターンの構造を指定する手法⁴⁾ などがある。一方、アプリアプローチは、パターンに含まれるシンボル数を一つずつ増やしながらか、再帰的にパターンを抽出する手法である¹⁾²⁾⁶⁾。具体的には、シンボル数が k 個のパターンを用いて $k+1$ 個のパターンを生成し、そのうち有用性が認められたパターンのみを選択する手法である。

本論文では、アプリアプローチを用いて、マルチストリームからパターンを抽出する手法を提案する。本手法の特徴は、映像データに特有の二種類の事前知識 (イベント境界と時間的偏り) を組んでいる点である。これによって、不適切な時系列パターンが抽出されなくなるだけでなく、パターン抽出の効率も格段に向上させることができる。

また、時系列パターン抽出には、パターンとマルチストリームをマッチングさせて、パターンのインスタンス (事例) を発見することが必要不可欠である。このとき、各ストリームに繰り返して現れる同一シンボルから、どのシンボルを組合せてインスタンスを構成すればよいかが重要な問題となってくる。本論文では、シンボル間の時間的関係を考慮しながら、マルチストリーム中の各シンボルが重複して選択されないようにして、インスタンスを発見する手法を提案する。

3. メタデータ

本論文では、図1に示されているように、ショットごとに空間的特徴 (spatial aspect) と時間的特徴 (temporal aspect) を表す13種類のメタデータを抽出している。各メタデータは、メタデータの種類を表す大文字のアルファベットと、その属性値を表す数字が組合されたシンボルで表現されている。そして、各ショットごとに抽出されたメタデータを時間順に並べることによって、13次元のマルチストリームを得ている。

図1の *stream 1* から *stream 8* は、ショットの空間的特徴を表すメタデータで構成されている。*stream 1*, *stream 2*, *stream 3* は、キーフレーム*の H (色相), S (彩度), V (輝度) 成分における色分布を表している。H, S 成分

* 本論文では、ショットの空間的特徴は真ん中のビデオフレーム (キーフレーム) によって表され、と仮定している。

		Shot No.	1	2	3	4	5	6	..
Spatial aspect	stream 1	CH:	CH4	CH7	CH3	CH2	CH7	CH6	..
	stream 2	CS:	CS4	CS4	CS4	CS4	CS4	CS4	..
	stream 3	CV:	CV2	CV2	CV3	CV0	CV1	CV0	..
	stream 4	LN:	LN3	LN3	LN2	LN1	LN1	LN2	..
	stream 5	LL:	LL0	LL3	LL1	LL0	LL1	LL1	..
	stream 6	LB:	LB2	LB0	LB3	LB2	LB2	LB0	..
	stream 7	SA:	SA2	SA2	SA1	SA1	SA2	SA2	..
	stream 8	LA:	LA0	LA2	LA1	LA0	LA0	LA0	..
Temporal aspect	stream 9	SL:	SL4	SL4	SL3	SL4	SL4	SL1	..
	stream 10	MS:	MS3	MS3	MS2	MS4	MS5	MS4	..
	stream 11	MV:	MV1	MV4	MV0	MV1	MV0	MV0	..
	stream 12	SM:	SM1	SM1	SM1	SM1	SM1	SM1	..
	stream 13	AM:	AM1	AM3	AM2	AM3	AM3	AM2	..

図 1 ショットごとに抽出されたメタデータからなるマルチストリームの例
An example of a multistream which consists of raw level metadata derived from each shot.

は、キーフレームに映し出された主要オブジェクトや背景を識別するために、V 成分は、キーフレームの明るさを識別するために有用である¹²⁾。stream 4, stream 5, stream 6 は、エッジから抽出された直線の数、長さの分布、角度の分布を表している。直線数はキーフレームに含まれるオブジェクトの数に比例して多くなる傾向があり、長さや角度の分布は人工物（ビルや窓）や自然の風景を識別するために有用である¹²⁾。stream 7 は、キーフレーム中での最大の肌色領域の面積を表し、映し出された人物の大きさを特徴づけている。これに対して、stream 8 は最も大きな明るい領域の面積を表し、大きな領域に集中する人工的な光（レーザーや閃光）の存在を特徴づけている。

図 1 の stream 9 から stream 13 は、ショットの時間的特徴を表すメタデータから構成されている。stream 9 は、ショットの持続時間を表している。一般に、緊迫したイベントではショットの持続時間が短く、ロマンチックなイベントでは長くなると言われている。stream 10 と stream 11 は、MPEG の動きベクトルから算出された、ショット内のビデオフレームにおける画素値（輝度、色差）の変化量と変化の方向を表している。これらのメタデータによって、オブジェクトの動きや場所の変化が特徴づけられる。stream 12 は、ショット内の音声ストリームに最も多く出現する音の種類を表している。音の種類は、GMM (Gaussian Mixture Model)¹⁴⁾ を用いて “人の声”, “音楽”, “無音” のいずれかを判定している*。stream 13 はショット内の音声ストリームにおける最大の振幅値を表し、銃声、爆発などの効果音の特徴づけている。

4. 時間制約に基づく時系列パターンマイニング

本章で述べる時系列パターンマイニング手法は、マルチ

* ただし、良好な結果が得られなかったため、最終的には人手によって音の種類を判定している。

ストリームから時系列パターンを抽出するために、パターンの生成と検証を繰り返す。すなわち、マルチストリームからパターンのインスタンスを発見して、有用性が認められたパターンのみを選択する。しかしながら、考えるすべてのパターンに対して有用性の検証を行うことは、多大な計算コストを要するために現実的ではない。そこで、検証するパターンの数（すなわち、探索空間）を削減するために、アプリアオリアプローチに従って、再帰的に検証すべきパターンを生成している。

さらに、時系列パターンと映像の意味内容を対応づけるために、‘イベント境界’、‘時間的偏り’と呼ばれる時間的特徴を考慮する。イベント境界とは、二つの隣り合うイベント間の境界であり、具体的には、登場人物の行動の変化や場所の変化など、意味内容の重要な変化が起こる時点である。例えば、図 2 (a) の一番左のイベントでは、登場人物 A は車を運転しているが、次のイベントでは車から降りて会話している。すなわち、shot 5 と shot 6 の間にはイベント境界 (SEB1) が存在していることになる。

時間的偏りとは、時間的に近い位置にあるショットほど意味的な結びつきが強いという特徴である⁶⁾¹³⁾。例えば、図 2 (a) の shot 10, shot 11, shot 15 は同一の会話イベントに属しているが、shot 10 と shot 11 は時間的に非常に近い位置にあり、意味内容に関しても強い結びつきをもっていると考えられる。しかしながら、shot 15 は shot 10 と shot 11 から遠く離れており、意味的な結びつきをもっているとは考えにくい。このようなイベント境界と時間的偏りを時間制約とすることによって、不適切なパターンが抽出されることを防ぐことができる。

4.1 アルゴリズムの概要

マルチストリーム S を構成するストリーム s において、時間 t に出現するシンボル v を (v, t) と表わすことにする。例えば、図 2 (b) の stream A で時間 1 に出現するシンボル A2 は、 $(A2, 1)$ と表わすことができる。さらに、 S 内の任意の二つのシンボルに対して、‘シリアル (serial)’、もしくは‘パラレル (parallel)’という時間的関係を定義する**。図 2 (b) を用いて説明すると、 $(A2, 1)$ と $(C3, 1)$ はパラレルであり、二つのシンボルは同じ時間に出現している。一方、 $(C3, 1)$ と $(C4, 2)$ はシリアルであり、二つのシンボルは異なる時間に出現している。ただし、両者は必ずしも同じストリームに出現する必要はなく、 $(C4, 2)$ と $(E1, 4)$ もシリアルである。すなわち、任意の二つのシンボル (v_1, t_1) と (v_2, t_2) に対して、シリアル、パラレル関係は以下のよう

$$\begin{aligned} t_1 \neq t_2 &\rightarrow \text{serial}, \\ t_1 = t_2 &\rightarrow \text{parallel}. \end{aligned} \quad (1)$$

S における時系列パターン p_l を、時間的に順序づけられた l 個のシンボルからなるシーケンスとして定義する。以

** これらの用語名は文献 3) と同じであるが、定義は全く異なる。

図 2 イベント境界と時間的偏りの例 (a) と、それらに基づいた時間制約 SEB , $TDT = 2$ を用いた 4-パターン $p_4 = (A2, nil), (C3, parallel), (C4, serial), (E1, serial)$ のインスタンスの発見例 (b)
An example of semantic event boundaries and temporal localities (a), and an example of finding instances of 4-pattern $p_4 = (A2, nil), (C3, parallel), (C4, serial), (E1, serial)$, using SEB and $TDT = 2$ (b).

後, p_l を l -パターンと呼ぶ。例えば, 図 2 (b) において, 丸で囲まれた四つのシンボルからなるシーケンスは 4-パターン p_4 である。一般的に, p_l は以下のように定式化できる。

$$p_l = (v_1, nil), (v_2, tr_2), (v_3, tr_3), \dots, (v_l, tr_l). \quad (2)$$

ここで, $i = 2, \dots, l$ のとき, (v_i, tr_i) は一つ前のシンボル (v_{i-1}, tr_{i-1}) と tr_i という時間関係をもっている (すなわち, $tr_i = serial$ もしくは $tr_i = parallel$)。例えば, 図 2 (b) の p_4 は, $p_4 = (A2, nil), (C3, parallel), (C4, serial), (E1, serial)$ と表すことができる。

p_l のインスタンスとは, マルチストリーム S において, p_l に含まれるシンボルが時間順序どおりに出現している事例である。すなわち, p_l のインスタンスは, p_l に含まれる各シンボルの S における出現箇所を表すシーケンス, $(v_1, t_1), (v_2, t_2), (v_3, t_3), \dots, (v_l, t_l)$ と定式化できる。例えば, 図 2 (b) には, p_4 の三つのインスタンスが示されているが, 最も左にあるインスタンスは $(A2, 1), (C3, 1), (C4, 2), (E1, 4)$ と表すことができる。

本論文では, 意味的な結びつきのない p_l のインスタンスが発見されないように, 映像の時間的特徴に基づいた二種類の時間制約を用いている。具体的には, p_l にシリアル関係をもつシンボル $(v_i, tr_i), (v_{i+1}, serial)$ が含まれているとき, これらのシンボルの S における出現箇所 $(v_i, t_i), (v_{i+1}, t_{i+1})$ (ただし, $t_i < t_{i+1}$) には, 以下のような時間制約が課せられる。

イベント境界 (v_i, t_i) と (v_{i+1}, t_{i+1}) の間に, ' SEB (Semantic Event Boundary)' が存在することは許されない。例えば, 図 2 (b) の真ん中のインスタンスは, $(C3, 5)$ と $(C4, 6)$ の間に $SEB1$ が存在するので発見されない。このように, インスタンスの発見範囲をイベント内に制限すれば, 一貫性のある意味内容を p_l に対

応づけることができる。

時間的偏り (v_i, t_i) と (v_{i+1}, t_{i+1}) の時間的距離が, ' TDT (Temporal Distance Threshold)' を越えることは許されない (すなわち, $t_{i+1} - t_i \leq TDT$)⁶⁾。例えば, 図 2 (b) で $TDT = 2$ と設定されている場合, 最も右にあるインスタンスは, $(C4, 11)$ と $(E1, 15)$ の時間的距離が $4 > TDT$ と離れているので発見されない。すなわち, TDT を越えて出現しているシンボルには, 意味的な結びつきがないとみなされる。

これまで述べてきたことを踏まえると, 本論文で提案するマイニングアルゴリズムは, 以下のように要約できる。

- (1) パターンのシンボル数 l を $l = 1$ として初期化する。そして, マルチストリーム S に出現する各種シンボルを 1-パターン p_1 とみなして, 評価尺度 f を満たす p_1 を選択する。
- (2) l を 1 増やして, $(l - 1)$ -パターン p_{l-1} から l -パターン p_l を生成する。
- (3) 各 p_l に対して, S 内で SEB と TDT を満たすインスタンスを発見する。そして, p_l のインスタンス数 (すなわち, p_l の頻度) と f を用いて, p_l の有用性を検証し, 有用性が認められた p_l のみを選択する。
- (4) もし, p_l が一つも選択されなければ, アルゴリズムを終了する。さもなければ, (2) に戻って, 反復的にパターンの生成と検証を行う。

以降では, 次の二点について詳しく述べる。一点目はどのようにして $(l - 1)$ -パターン p_{l-1} から l -パターン p_l を生成するかであり, 二点目はどのようにして p_l のインスタンスを発見するかである。パターンの評価尺度 f に関しては, 文献 7) に様々な評価尺度が紹介されているが, 本論文では, 支持度 (support) と確信度 (confidence) を用いている。

4.2 パターンの生成

$(l - 1)$ -パターンから l -パターンを生成するための効率的

図3 4-パターン $p_4 = (A1, nil), (B2, serial), (D3, parallel), (E4, serial)$ の生成

The generation of candidate 4-pattern $cp_4 = (A1, nil), (B2, serial), (D3, parallel), (E4, serial)$.

な手法は、文献2)で提案されている。しかしながら、文献2)では、トランザクションデータについては考慮されているものの、図2(b)のようなマルチストリームは考慮されていない。このため、式(2)のマルチストリームにおける時系列パターンの定義に従って、手法を拡張する必要がある。

図3に、二つの3-パターン $p_3 = (A1, nil), (B2, serial), (D3, parallel)$ と $p'_3 = (B2, nil), (D3, parallel), (E4, serial)$ を用いて、4-パターン $p_4 = (A1, nil), (B2, serial), (D3, parallel), (E4, serial)$ が生成される過程を示す。まず初めに、 p_4 を生成するためには、先頭シンボルを除いて作成されたサブシーケンスと最後尾シンボルを除いて作成されたサブシーケンスからなる3-パターンのペア p_3, p'_3 を、既に構築された3-パターンの集合から発見しなければならない。もし p_3 と p'_3 が存在すれば、共通するサブシーケンスを重ね合わせて p_4 が生成できる。 l -パターンの生成に関しても同様に、サブシーケンスのペアが $(l-1)$ -パターンの集合に含まれていなければならない。このような条件によって、検証しなければならないパターンを大幅に削減することができる。

上記のパターン生成条件の原理は、「抽出されるパターンのどのサブシーケンスもパターンでなければならない」ということである。この原理に基づく、パラレル関係を含んだパターンの生成には、さらなる条件が課せられる。例えば、図3の p_4 に関しては、 $subseq_1 = (A1, nil), (B2, serial), (E4, serial)$ と $subseq_2 = (A1, nil), (D3, serial), (E4, serial)$ が、3-パターンの集合に含まれていなければならない。ここで、 $subseq_1$ と $subseq_2$ は、 p_4 におけるシンボル間の時間的順序を保持したままで、同時に起こりうるシンボルを切り分けて考慮したものである。このように、パラレル関係を含む l -パターンを生成する場合、パラレル関係にあるシンボルを分解して作成されたすべてのサブシーケンスも、 $(l-1)$ -パターンの集合に含まれていなければならない。

4.3 パターンのインスタンスの発見

l -パターン p_l に関して、マルチストリーム S における p_l のインスタンス数、すなわち頻度をカウントする。そして、

図4 3-パターン $p_3 = (A3, nil), (B1, serial), (C2, serial)$ のインスタンスの発見方法

An illustration of locating 3-pattern $p_3 = (A3, nil), (B1, serial), (C2, serial)$.

p_l の頻度から算出される支持度と確信度に基づいて、 p_l の有用性を検証し、有用であると判定された p_l のみを選択する。以下では、 SEB , TDT の時間制約を用いて、 S における p_l のインスタンスを発見するための手法について述べる。

$TDT = 2$ と設定したときの、 $p_3 = (A3, nil), (B1, serial), (C2, serial)$ のインスタンス発見方法を図4に示す。実線矢印で示されているように、 $t = 1$ から $t = 3$, $t = 4$ から $t = 8$, $t = 5$ から $t = 9$ において、 SEB , TDT を満たす三つのインスタンスが発見されている。ただし、 $(A3, 1)$ から $(B1, 3)$ にかけての点線矢印は、「 TDT で制限された区間に複数のシリアル関係が存在する場合、シンボル間の距離が近いほど意味的にも強い結びつきをもっている」という時間制約を表している。すなわち、 $(A3, 1)$ に対して、 $(B1, 2)$ とのシリアル関係のみが選択される。同様の制約は、 $(B1, 2)$ から $(C2, 4)$ にかけての点線矢印にも当てはまる。

一方、同一シンボルを重複して複数のインスタンスに許してしまうと、パターン長の短い $(l-k)$ -パターン p_{l-k} ($k < l$) よりも、パターン長の長い p_l の方が頻繁に出現しているという不自然な結果が生じてしまう。例えば、図4において、 $(A3, 5)$ から $(B1, 6)$ と、 $(B1, 7)$ から $(C2, 8)$ にかけての点線矢印を含めてしまうと、 $t = 4$ から $t = 9$ の区間において、 p_3 のインスタンスは四つ発見されるにも関わらず、 $p_2 = (A3, nil), (B1, serial)$ のインスタンスは三つしか発見されない。そこで本論文では、「同一シンボルは、一つのインスタンスを構成するためにのみ用いられる」という制約を用いる。すなわち、図4の $(B1, 6)$ と $(C2, 8)$ は、 $t = 4$ から始まる p_3 のインスタンスで用いられるので、 $t = 5$ から始まるインスタンスでは、 $(B1, 7)$ と $(C2, 9)$ を用いなければならない。以降、 p_l のどのインスタンスにも用いられていないシンボルを、「未使用シンボル」と呼ぶ。

以上のことを踏まえて、マルチストリーム S における p_l のインスタンスを発見するためのアルゴリズムについて述べる。このアルゴリズムは、 S の先頭から欲張り法の要領で、 $Forward$, $Backward$ フェーズを用いて、 p_l のインスタンスを逐次発見していく。 $Forward$, $Backward$ フェーズは文献2)で提案されたものであるが、本論文では、 SEB , TDT の時間制約だけでなく、時間的距離が最も近いシリアル関係を選択し、かつ未使用シンボルのみから構成され

る p_l のインスタンスを発見できるように拡張している。

今、 p_l の i 番目のシンボル (v_i, tr_i) のインスタンスを $P_{l,i} = (v_i, t_i)$ と表わす。そして、 p_l の先頭から i 個のシンボルからなるサブシーケンス $(v_1, nil), \dots, (v_i, tr_i)$ に対して、インスタンス $P_{l,1}, \dots, P_{l,i}$ が既に見つかっているとする。

Forward phase: p_l の $i+1$ 番目のシンボル (v_{i+1}, tr_{i+1}) のインスタンス $P_{l,i+1}$ を見つける。

$tr_{i+1} = serial$ の場合、 $t_{i+1} > t_i$ である未使用シンボル (v_{i+1}, t_{i+1}) を選ぶ。もし、 (v_i, t_i) と (v_{i+1}, t_{i+1}) のシリアル関係が *SEB*, *TDT* を満たせば、 $P_{l,i+1} = (v_{i+1}, t_{i+1})$ が見つかったことになる。さらに、 $P_{l,i+2}$ を見つけるための *Forward* フェーズを実行する。そうでなければ、*Backward* フェーズを実行する。

$tr_{i+1} = parallel$ の場合、 $t_{i+1} \geq t_i$ である未使用シンボル (v_{i+1}, t_{i+1}) を選ぶ。もし、 $t_{i+1} = t_i$ であれば、 $P_{l,i+1}$ が見つかったことになり、*Forward* フェーズを実行する。 $t_{i+1} > t_i$ であれば、*Backward* フェーズを実行する。

どちらの場合においても、 $i+1 = l$ かつ $P_{l,l}$ が見つかったとき、 p_l のインスタンスが発見されたことになる。

Backward phase: 直前に見つかった (v_{i+1}, t_{i+1}) は、 $P_{l,i+1}$ として認められなかった。そこで、 (v_{i+1}, t_{i+1}) に基づいて、別の $P_{l,i}$ を見つけるためのバックトラックを行う。

$tr_{i+1} = serial$ の場合、 $t'_i > (t_{i+1})$ の直前の *SEB* かつ $t'_i \geq t_{i+1} - TDT$ である未使用シンボル (v_i, t'_i) を選ぶ。

$tr_{i+1} = parallel$ の場合、 $t'_i \geq t_{i+1}$ である未使用シンボル (v_i, t'_i) を選ぶ。

次に、既に見つかっている $P_{l,i-1}$ を用いて、 (v_i, t'_i) を $P_{l,i}$ として認めることができるかどうかチェックする。もし認められれば、 $P_{l,i+1}$ を見つけるための *Forward* フェーズを実行する。そうでなければ、別の $P_{l,i-1}$ を見つけるための *Backward* フェーズを実行する*。

上記の *Forward*, *Backward* フェーズにおいて、指定された未使用シンボル (v_{i+1}, t_{i+1}) , (v_i, t'_i) が存在しなければ、 $P_{l,i+1}$ もしくは $P_{l,i}$ の候補すら存在しないことになる。したがって、 S には p_l のインスタンスが存在しないことが保証され、その時点でアルゴリズムを終了する。最終的に、それまでに発見された p_l のインスタンス数を、 p_l の頻度 $freq(p_l)$ として返す。

5. 実験結果

本論文で提案したアプローチを以下の六つの映像に適用し、意味的なパターンの抽出を行った。

• *video 1* と *video 2* は、監督 *A* によって製作されたサイエンスフィクション映画 (*movie_A1*) からの約 30 分の部分映像で、それぞれ 444 ショット、725 ショットから構成されている。

• *video 3* と *video 4* は、上記の監督 *A* によって製作された別のサイエンスフィクション映画 (*movie_A2*) からの約 30 分の部分映像で、それぞれ 578 ショット、439 ショットから構成されている。

• *video 5* と *video 6* は、別の監督 *B* によって製作されたサイエンスフィクション映画 (*movie_B*) からの約 30 分の部分映像で、それぞれ 444 ショット、453 ショットから構成されている。

上記の映像はすべて MPEG 形式の動画で、フレームレートは 29.97(*fps*) である。ショット境界の検出に関しては、MP-Factory (MPEG Software Development Kit)¹⁶⁾ を用いている。また、メタデータの抽出には、MP-Factory と OpenCV (Open Source Computer Vision Library)¹⁵⁾ を用いている。

まず初めに、ショットの持続時間に基づいて、映像をイベントよりも大きな構成単位である‘セグメント’に分割する。例えば、会話や恋愛などの穏やかなイベントが続くセグメントではショットの持続時間は長く、戦闘などの緊迫したイベントが続くセグメントでは短くなっている。ただし、ショットの持続時間にはばらつきがあるため、状態ごとに確率密度関数が割り当てられた無限状態オートマトンを用いて、ショットの持続時間が特定の確率分布に従う区間をセグメントとして発見している⁸⁾。そして、各セグメントにおいて、ショットの空間的特徴と時間的特徴、および時間的偏りに基づいて、イベント境界を検出している¹³⁾。しかしながら、上記の手法を用いても、六つの実験映像に対して、平均の精度が 54.2%、再現率が 70.8% と良好な結果を得ることはできなかった。このため、検出されたイベント境界を手手で修正して実験に用いている。

5.1 抽出された意味的なパターンの評価

表 1 に、実験映像から抽出された意味的なパターンの例を示す。一番左の列に示されているように、抽出された意味的なパターンは、三つのグループに分類できる。一つ目は、‘アクション’グループであり、登場人物が特定の行動をしているイベント（会話、移動、戦闘）を特徴づける意味的なパターンを含んでいる。二つ目は、‘シチュエーション’グループで、特定の状況下でのイベント（暗い、クローズアップ、緊迫している）を特徴づけるパターンを含んでいる。三つ目は、‘コンピネーション’グループであり、登場人物が特定の状況下で特定の行動をしているイベント（暗い状況下での会話、緊迫した状況下での会話など）を特徴づけるパターンを含んでいる。なお、隣接する二つのシンボル間のシリアル関係を $X - Y$ 、パラレル関係を XY と示している。また、適合率 (P) と再現率 (R) は、以下の式によって計算している。

* ただし、 $i = 1$ のときは、必ず *Forward* フェーズを実行する。

表 1 抽出された意味的なパターンの例
Examples of extracted semantic patterns.

group	event	pattern	movie_A1		movie_A2		movie_B	
			video 1	video 2	video 3	video 4	video 5	video 6
action	talk	SM1 – SM1	P:56.0 R:100	P:66.7 R:88.9	P:84.2 R:100	P:85.0 R:73.9	P:100 R:91.7	P:75.0 R:93.8
		SM1MV0	P:65.0 R:92.9	P:66.7 R:100	P:72.3 R:100	P:90.5 R:89.5	P:100 R:100	P:65.2 R:93.8
		MV0 – MV0	P:78.6 R:78.6	*	P:70.4 R:100	P:77.3 R:85.0	P:100 R:100	P:65.2 R:93.8
	move	MV4SM2	P:73.3 R:91.7	*	*	*	*	*
		MS5SM2	P:87.5 R:93.8	P:88.9 R:80.0	*	*	*	*
		SM2 – SM2	P:84.2 R:100	P:76.7 R:62.2	P:53.3 R:88.9	P:43.8 R:100	*	*
situation	dark	SL0 – SL0	*	P:81.8 R:90.0	P:60.0 R:100	P:46.7 R:100	*	P:40.0 R:100
		CV2	P:87.5 R:100	*	P:35.5 R:100	P:47.1 R:100	*	P:35.3 R:100
		LN1	P:52.2 R:85.7	P:42.2 R:86.4	*	*	P:46.2 R:100	P:89.5 R:85.0
	thrilling	SA3	P:48.1 R:92.9	P:41.2 R:95.5	P:100 R:61.1	P:66.7 R:82.4	*	P:94.7 R:90.0
		CS4	P:56.2 R:100	P:60.0 R:84.0	P:71.4 R:100	P:71.4 R:83.3	*	*
		CS4SM1	P:30.0 R:100	*	*	*	*	*
combination	talk + thrilling	CV2SM1	P:61.1 R:84.6	*	*	*	*	P:14.3 R:66.7
	talk + dark	CV2SM1	P:61.1 R:84.6	*	*	*	*	P:14.3 R:66.7
	talk + close-up	SA3SM1MV0	P:53.3 R:100	*	P:100 R:47.4	P:90.9 R:62.5	*	P:50.0 R:80.0
	talk + indoor	LN1SM1MV0	*	*	P:92.6 R:65.0	P:75.0 R:100	P:72.7 R:80.0	P:40.0 R:50.0
	dark + close-up	LN2SM1MV0	P:78.6 R:84.6	*	*	*	*	P:73.3 R:100
	dark + close-up + battle	CV2LN1SA3	P:71.4 R:83.3	*	*	*	*	*
		CV2LN1SA3SL0	P:55.6 R:55.6	*	*	*	*	*

$$P = \frac{\# \text{ of output and correct events}}{\# \text{ of output events}} \times 100(\%),$$

$$R = \frac{\# \text{ of output and correct events}}{\# \text{ of correct events}} \times 100(\%). \quad (3)$$

上式において、 $\# \text{ of output events}$ は、意味的なパターンを用いて検索されたイベント数である。 $\# \text{ of correct events}$ は、正しく検索されたイベント数であり、人が実際に映像を見て判定している。ただし、*は意味的なパターンが抽出されなかったことを表している。

まず、movie_A1と movie_A2から抽出された意味的なパターンについて説明する。アクショングループに属するSM1 – SM1, SM1MV0, MV0 – MV0は、会話を特徴づける意味的なパターンである。SM1 – SM1は「人間の声を含むショットの連続」、SM1MV0は「人間の声を含み、かつ動きに方向性のないショット」、MV0 – MV0は「動きに方向性のないショットの連続」と解釈できる。とりわけ、これらのパターンに対する再現率は高く、ほとんどすべての会話イベントで、登場人物は動かずに会話していることがわかる。MV4SM2は移動を特徴づける意味的なパターンであり、「一定の動きの方向をもち、かつ音楽を含むショット」と解釈できる。すなわち、登場人物が歩いたり、もしくは乗り物に乗って移動するイベントでは、頻繁にBGMが使用されている。戦闘を特徴づける意味的なパターンMS5SM2は「動きが大きく、かつ音楽を含むショット」、SM2 – SM2は「音楽を含むショットの連続」、SL0 – SL0は「持続時間の短いショットの連続」と解釈できる。とりわけ、MS5SM2とSM2 – SM2は、戦闘イベントでは人間の声よりも音楽の音量の方が大きいという興味深い特徴を抽出している。

シチュエーショングループに属するCV2は、暗い状況の特徴づける意味的なパターンで、「大部分が暗い色のショット」と解釈できる。CV2の再現率はすべての映像に対して100%であるが、video 3とvideo 4における適合率は低くなっている。これは、検索されたイベントには、黒い服を着た登場人物が出現するイベントが含まれていること

が原因となっている。LN1とSA3は、登場人物のクローズアップを特徴づける意味的なパターンであり、それぞれ「直線の数が多いショット」、「大きな肌色領域をもつショット」と解釈できる。LN1が抽出された理由は、人間の顔は丸いので輪郭からあまり直線が検出されないことによる。SA3に関しては、SAが登場人物の大きさを特徴づけるメタデータのため、抽出されて当然のパターンといえる。ただし、video 1やvideo 2における低い適合率が示すように、人物のクローズアップ以外に、肌色のオブジェクトや背景（岩や砂漠）を含むイベントも検索されてしまうという問題が生じた。

コンビネーショングループに含まれる意味的なパターンは、アクショングループとシチュエーショングループを組合せた複合的なパターンである。例えば、登場人物のクローズアップでの会話を特徴づけるLN1SM1MV0は、会話を特徴づけるSM1MV0とクローズアップを特徴づけるLN1を組合せたパターンである。ただし、室内での会話を特徴づけるLN2SM1MV0は例外で、LN2は直線の数が多いイベントのみを特徴づけるにすぎず、室内での状況の特徴づける意味的なパターンは抽出されていない。しかしながら、LN2がSM1MV0と結びつくと、会話（SM1MV0）している登場人物の背景にある窓や柱から、多くの直線（LN2）が検出されるため、会話の状況が室内と特定できるようになる。

ここで、上記の意味的なパターンを映像検索の観点から考察する。アクショングループとシチュエーショングループの高い再現率からわかるように、これらのグループに属する意味的なパターンを用いれば、検索要求に適合するイベントのほとんどを検索できることがわかる。しかしながら、これらのグループの適合率はあまり高くなく、適合しないイベントまでも検索されてしまう。この点に関しては、新たなメタデータを追加して、より詳細な意味的なパターンを発見する必要があると考えられる。

次に、*movie_A1*と*movie_A2*から抽出された意味的なパターンの汎用性を検証するために、*movie_B*から抽出されたパターンと比較する。まず、会話を特徴づける意味的なパターンに関しては、*movie_A1*と*movie_A2*から抽出された三種類のパターンと同じパターンを抽出することができた。すなわち、すべての実験映像に対して、会話イベントは共通のパターンを用いて検索できるといえる。一方、戦闘を特徴づける意味的なパターンに関しては、BGMを強調した戦闘イベント *MS5SM2*と *SM2-SM2*が *movie_A1*と *movie_A2*から発見されたが、*movie_B*からは抽出することができず、*SL0-SL0*のみが抽出された。また、緊迫したイベントを特徴づける意味的なパターン *CS2*（「大部分が淡い色のショット」と解釈できる）は、*movie_A1*と *movie_A2*からは抽出できたが、*movie_B*からは抽出できなかった。理由としては、淡い色を表すメタデータ *CS2*が映像に依存したメタデータであるため、そもそも *movie_B*に *CS2*は含まれていなかった。

5.2 マイニング手法の性能評価

以下では、4章で提案したマイニング手法の性能について考察する。図5に、本マイニング手法において生成された *video 1*の *l*-パターンの数と、検証の結果、抽出された *l*-パターン数を示す。必要なパラメータとして、 $TDT = 2$ 、(最小支持度) = 20、(最小確信度) = 0.6と設定した。

本マイニング手法では、最悪の場合、 $2^{l-1} \times (N_{p1})^l$ 個の *l*-パターンを生成しなければならない。ここで、 N_{p1} は1-パターン数であり、*l*-パターンに含まれる各シンボルに対して、パラレルかシリアルかの時間関係を含めた $2 \times N_{p1}$ 通りの選び方が存在する*。図5においては $N_{p1} = 59$ であり、例えば、3-パターンを抽出するためには、最悪の場合、 $2^2 \times 59^3 = 821516$ 個の3-パターンを生成しなければならない。これに対して、本マイニング手法では7732個と格段に少ない。さらに、 $l > 3$ のとき、*l*-パターン数は指数関数的に急増するが、実際に生成された *l*-パターン数は $l = 3$ をピークに減少している。このように、本マイニング手法は、生成すべきパターンを大幅に絞り込んだ効率的な手法となっている。

ただし、検証を行っているにも関わらず、多数の無意味なパターンが抽出されてしまうことがある。とりわけ、ほとんどの時間に頻出しているシンボルは特徴のないシンボルであり、パターンに含めても意味がない。しかしながら、このような頻出するシンボルを含んだ無意味なパターンのインスタンスは発見されやすく、頻度が高いパターンを重要視する支持度と確信度を用いて検証しているため、多数の無意味なパターンが抽出されてしまうという問題がある。例えば、図5の実験では、*video 1*についての13次元のマルチストリームから最終的に15262個のパターンが抽出された。このうち10983個のパターンは *LA0*というシンボ

図5 *video 1*から生成された *l*-パターン数と抽出された *l*-パターン数
Numbers of generated *l*-patterns and extracted *l*-patterns from *video 1*.

ルを含んでいたが、*LA0*は *video 1*の約八割のショットに現れる、何の意味内容もないメタデータであった。

このような無意味なパターンが多数抽出されることによる弊害として、多数の無意味なパターンにメモリー領域が占拠されてしまい、アルゴリズムを終了できなくなってしまうという問題が生じる。そこで本論文では、ほとんど一種類のシンボルしか出現しないストリームはマルチストリームから事前に除いて、パターン抽出に悪影響を及ぼす要因を排除している。すなわち、*video 1*についての13次元マルチストリームから、*LA*に関するストリームを除いた12次元のマルチストリームを作成した。そして、この12次元のマルチストリームに対して、パラメータを低く設定して($TDT = 2$ 、(最小支持度) = 15、(最小確信度) = 0.5)、パターンを抽出した。結果として、パラレル関係にある二つのシンボルが周期的に出現する、興味深い意味的なパターンを発見することができた。

- (1) *SL0AM0-SL0AM0-SL0AM0*: 持続時間が短く、かつ音量の小さいショットの繰り返しであり、アクションが主体の緊迫したイベントを特徴づけている。
- (2) *MS0MV0-MS0MV0-MS0MV0*: 動きが小さく、かつ動きに方向性のないショットの繰り返しであり、会話イベントを特徴づけている。

次に、本論文で提案した時間制約 (*SEB*, *TDT*) の有用性を検証する。表2に、*SEB*, *TDT*を様々に変化させたときに抽出されたパターン数を示す。*TDT*は大きく設定するにつれて、抽出されるパターン数が急増していることがわかる。このため、大きく設定しすぎると、時間的に大きく離れたシリアル関係を含んだパターンのインスタンスが発見されてしまう。つまり、一つのイベント内でも、話題が変わる前後のショットを関連づける不適切なパターンが抽出されてしまう。

* 先頭シンボルの選び方は、時間関係が *nil* で固定されているため N_{p1} 通りである。

表 2 様々な *SEB*, *TDT* 設定下において抽出されたパターン数の比較

Comparison of numbers of extracted patterns, in various settings of *SEB* and *TDT*.

with <i>SEB</i>							
<i>TDT</i>	0	1	2	4	8	16	∞
# of patterns	1429	3225	6716	22736	122694	250374	252281
without <i>SEB</i>							
<i>TDT</i>	0	1	2	4	8	16	∞
# of patterns	1429	3829	698608	The algorithm cannot finish.			

また, *SEB* を用いない方が, 抽出されるパターン数が圧倒的に多くなることがわかる. すなわち, *SEB* を用いずに抽出されるパターンは, 異なるイベント間にも関わらず, たまたまシンボル間に共起性があるために, 意味内容の異なる複数のイベントを関連づけてしまっている. 特に, $TDT \geq 4$ で, かつ *SEB* を用いない場合, 上記の不適切なパターンが多数抽出されてメモリー領域を占拠してしまい, アルゴリズムが終了できなかった. 以上の点から, *SEB*, *TDT* の時間制約は, パターンに適切な意味内容を関連づける点においても, パターン抽出の効率においても合理的であるといえる.

5.3 意味的なパターンを用いた映像検索システム

最後に, 意味的なパターンの応用例として, 映像の意味内容に基づく検索システムを示す. 図 6 で, ユーザは, 索引付けされていない大量の映像が蓄積されたアーカイブから, カーチェイスイベントを検索しようとしている. ここで, カーチェイスを特徴づける意味的なパターンを, 「直線の数が少なく, かつ動きが大きいショット」と仮定している. そして, このパターンを含むイベントが, 画像, 音声, 映像処理などの技術を用いて検索され, ユーザに返される. その結果, ユーザは, 検索されたイベントの中から, 効率的にカーチェイスイベントを見つけることができる. このシステムの最大の特徴は, 複数の映像に有効である汎用的な意味的なパターンを抽出していれば, 索引付けされていない別の映像からでも, 同じ意味内容のイベントを検索できる点である.

6. む す び

本論文では, 映像からイベントを特徴づける意味的なパターンを抽出するためのビデオデータマイニングアプローチについて述べた. 4 章では, 意味的なパターンをマルチストリームにおける時系列パターンとして抽出するために, 時間制約に基づいた時系列パターンマイニング手法を提案した. しかしながら, 本マイニング手法を用いても, 多数の無意味なパターンが抽出されるという問題が残されている. そこで, さらに検討すべき点として, アルゴリズムの観点からとマルチストリームの取り扱い方の観点から, 以下のような点が挙げられる.

アルゴリズムの観点からは, パターンの有用性を検証するための評価尺度として, 本マイニング手法では支持度と

図 6 意味的なパターンを用いた映像検索システム
A video retrieval system by using semantic patterns.

確信度を用いたが, これら以外にも, 様々な評価尺度 (例えば, ϕ -coefficient, interest factor, J-measure など) が提案されている. 今後は, 様々な評価尺度を比較して, 意味的なパターンを抽出するために最適な評価尺度を選択する必要がある⁷⁾.

また, マルチストリームの取り扱い方の観点からは, 無意味なパターンが抽出されないようにマルチストリームを再構成させることである. 本論文では, 無意味なパターンの原因となっているシンボルが現れるストリームを単純に除いて検討したが, 文献 6) では, 時間軸上で連続して出現しているシンボルを一つのシンボルに置き換える操作を行い, ほとんどの時間に頻出しているシンボルを含むパターンが抽出されないようにしている. このように, 映像から作成されたマルチストリーム内のすべてのシンボル (メタデータ) を用いるのではなく, それらの中でも重要なシンボルだけを用いて意味的なパターンを抽出するアプローチについても, 今後検討する予定である.

〔文 献〕

- 1) Y. Matsuo, K. Shirahama and K. Uehara: "Video Data Mining: Extracting Cinematic Rules from Movie", Proc. of 4th International Workshop on Multimedia Data Mining (MDM/KDD 2003) pp. 18-27 (2003)
- 2) R. Srikant and R. Agrawal: "Mining Sequential Patterns: Generalizations and Performance Improvements", Proc. of 5th International Conference on Extending Database Technology (EDBT 1996), pp. 3-17 (1996)
- 3) H. Mannila, H. Toivonen and A. Verkamo: "Discovering Frequent Episodes in Sequences", Proc. of the International Conference on Knowledge Discovery and Data Mining (KDD 1995), pp. 210-215 (1995)
- 4) D. Chudova and P. Smyth: "Pattern Discovery in Sequences under a Markov Assumption", Proc. of 8th International Conference on Knowledge Discovery and Data Mining (KDD 2002), pp. 153-162 (2002)
- 5) Y. Tanaka, K. Iwamoto and K. Uehara: "Discovery of Time-Series Motif from Multi-Dimensional Data Based on MDL Principle", Machine Learning, Volume 58, Issue 2-3, pp. 269-300 (2005)
- 6) X. Zhu, X. Wu, A. Elmagarmid, Z. Feng and L. Wu: "Video Data Mining: Semantic Indexing and Event Detection from the Association Perspective", IEEE Transactions on Knowledge and Data Engineering, Volume 17, Issue 5, pp. 665-677 (2005)
- 7) P. Tan, V. Kumar and J. Srivastava: "Selecting the Right Interestingness measure for Association Patterns", Proc. of 8th International Conference on Knowledge Discovery and Data Mining (KDD 2002), pp. 32-41 (2002)
- 8) J. Kleinberg: "Bursty and Hierarchical Structure in Streams",

Proc. of 8th International Conference on Knowledge Discovery and Data Mining (KDD 2002), pp. 91–101 (2002)

- 9) J. Oh and B. Bandi : “Multimedia Data Mining Framework for Raw Video Sequences”, Proc. of 3rd International Workshop on Multimedia Data Mining (MDM/KDD 2002), pp. 1–10 (2002)
- 10) D. Wijesekera and D. Barbara : “Mining Cinematic Knowledge: Work in Progress”, Proc. of the International Workshop on Multimedia Data Mining (MDM/KDD 2000), pp. 98–103 (2000)
- 11) J. Pan and C. Faloutsos : “VideoCube: A Novel Tool for Video Mining and Classification”, Proc. of 5th International Conference on Asian Digital Libraries (ICADL 2002), pp. 194–205 (2002)
- 12) A. Mojsilovic and B. Rogowitz : “Capturing Image Semantics with Low-level Descriptors”, Proc. of the 2001 IEEE International Conference in Image Processing (ICIP 2001), pp. 18–21 (2001)
- 13) Y. Rui, T. Huang and S. Mehrotra : “Constructing Table-of-Content for Videos”, ACM Multimedia Systems Journal, Volume 7, Issue 5, pp. 359–368 (1999)
- 14) L. Lamel and J. Gauvain : “Speaker Recognition with the Switchboard corpus”, Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP 1997), pp. 1067–1070 (1997)
- 15) “OpenCV: Open Source Computer Vision Library”, Intel (<http://www.intel.com/research/mrl/research/opencv/>)
- 16) “MPFactory: MPEG Software Development Kit”, KDDI (<http://w3-mcgav.kddilabs.jp/mpeg/mpfs40/indexe.html>)

しらはま きみ あき
白浜 公章 2003 年, 神戸大学工学部情報知能工学科卒業。2005 年, 同大学院自然科学研究科博士前期課程修了。現在, 同大学院自然科学研究科博士後期課程在学中。時系列データを対象としたデータマイニングと映像処理に関する研究に従事。

いで の こう いち
出野 構一 2005 年, 神戸大学工学部情報知能工学科卒業。現在, 同大学院自然科学研究科博士前期課程在学中。時系列データを対象としたデータマイニングと映像処理に関する研究に従事。

うえ はら くに あき
上原 邦昭 1978 年, 大阪大学基礎工学部情報工学科卒業。1983 年, 同大学院博士後期課程単位取得退学。大阪大学産業科学研究所助手, 講師, 神戸大学工学部情報知能工学科助教授, 同都市安全研究センター教授を経て, 現在, 同大学院自然科学研究科教授。人工知能, 特に機械学習, マルチメディアデータベース, 自然言語によるヒューマンインタフェースの研究に従事。