映像ストリームにおけるバースト検出に基づくトピック発見

白浜 公章 † 上原 邦昭 ‡
† 神戸大学大学院経済学研究科 ‡ 神戸大学大学院工学研究科

1 はじめに

近年，ウェブ上や個人の HDD 内に蓄積された大量の映像の中から，ユーザが視聴したいイベントを効率的に検索するための映像検索技術の開発が活発に行われている．映像検索においては，映像を意味的に一貫性のあるイベントに分割するための“映像セグメンテーション”が重要になってくる．従来の映像セグメンテーションでは，色や動きなどの“信号レベルの特徴量”を用いていた．この方法は，ニュースやスポーツなどに対しては，イベントごとに撮影場所，カメラ配置，編集方法などに規則性が存在するため有効である．

しかしながら，映画やドラマなどにおいては，上記の規則性が存在せず，信号レベルの特徴量では，適切にイベントに分割できないことが多々ある．例えば，散歩イベントを考えてみる．このイベントでは，登場人物が，任意に動いたり立ち止まったりしながら，場所を変えて行くため，色や動きに規則性はない．そのため，散歩イベントを，意味的に一貫性のあるイベントと判定することは不可能である．

これに対して，本論文では，映画やドラマなどのイベントにおいて，登場人物という”意味レベルの特徴量”に規則性があることに着目する．例えば，上記の散歩イベントでは，「ほとんどのショットに，散歩している人物が出現している」という点に規則性がある．そこで，登場人物の出現パターンに基づいて，映像をイベントに分割するアプローチを提案する．そして，登場人物の例外的な出現パターン（“バースト”）を含んだイベントを，その人物が興味深い行動を行っている“トピック”として抽出する．

2 基本コンセプト

本章では，本論文の映像セグメンテーション，及びトピック抽出に関するコンセプトについて概説する．図 1 から分かるように，映像は，様々なカメラから撮影されたショットをつなぎ合わせて制作されている．そして，例えば人物 *A* に注目すると，*A* が出現しているショット（*shot 1*, *shot 3*）と出現していないショット（*shot 2*）が存在することが分かる．このように，映像中の 1 人の登場人物に注目すると（“注目人物”），映像を，図 1 の下部に示されているような注目人物の出現区間と非出現区間を表わす一次元時系列（“出現・非出現区間シークエンス”）に変換することができる．

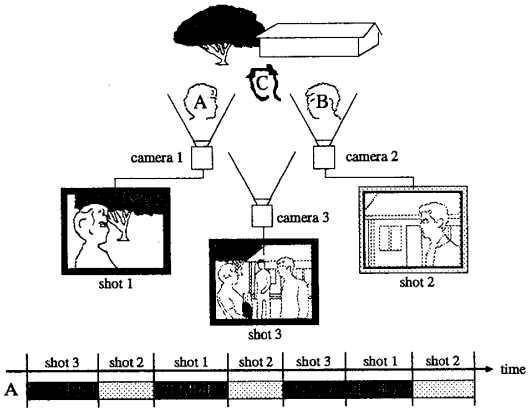

図 1: 注目人物の出現・非出現区間シークエンス．

出現・非出現区間シークエンスを用いると，映像中の様々なイベントを特徴づけることができる．例えば，図 2 の *Event 3* において，出現区間長，非出現区間長が極端に短くなっている．これは，*Event 3* のようなスリリングなイベントでは，緊迫感を出すために，ショットの時間長が非常に短く設定されているからである．このような「意味内容に応じたショットの時間長のリズム」[1] が，出現区間長，非出現区間長に反映される．

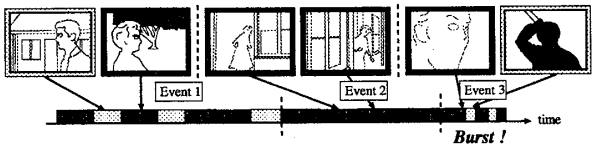

図 2: 映像セグメンテーションとバースト検出．

さらに，出現と非出現の“発生率”を考慮すると，以下のようなイベントが特徴づけられる．図 2 の *Event 1* のような会話イベントでは，注目人物が出現しているショットと話し相手が出現しているショットが交互につなぎ合わされているため，出現と非出現の発生率がほぼ等しくなる．一方，*Event 2* のような注目人物が単独で行動しているイベントは，注目人物の行動を様々なカメラから撮影したショットで構成されるため，出現の発生率が非常に高くなる．

本論文では，上記の議論から，イベント内では，注目人物の出現区間長，非出現区間長，出現と非出現の発生率が類似していると仮定する．さらに，映像をイベントに分割した後で，注目人物の例外的な出現パターン（“バースト”）を含むイベントを“トピック”として抽出する．ここで，「意味内容に応じたショットの時間

Topic Extraction based on Burst Detection in Video Streams
†Kimiaki Shirahama ‡Kuniaki Uehara
†Graduate School of Economics, Kobe University
‡Graduate School of Engineering, Kobe University

長のリズム」[1]に基づいて，以下の2種類のバーストを定義する．1種類目は，*Event 3* のようなスリリングなイベントに代表される高速なショットの切り替えに対応して，「時間長の短いイベントに，多数の出現区間が存在する」というバーストである．2種類目は，恋愛のイベントに代表されるゆったりとしたショットの切り替えに対応して，「時間長の長いイベントに，少数の出現区間しか存在しない」というバーストである．

3 アルゴリズム

本章では，前章の映像セグメンテーション，及びトピック抽出に関するコンセプトを定式化する．出現・非出現区間シークエンスは，$X = x_1, x_2, \cdots, x_N$ $(x_i = (a_i, d_i), a_i \in \{A, D\}, d_i \in \Re)$ と表現できる．ここで，a_i は x_i が出現区間 (A) か非出現区間 (D) を表わし，d_i は x_i の区間長を表わしている．ゆえに，映像セグメンテーションとは，X を $K(\ll N)$ 個のイベント $E = e_1, \cdots, e_K$ $(e_i = x_a, \cdots, x_b)$ に分割することに他ならない [2]．

イベント e_i 内の x_j $(a \le j \le b)$ に関して，出現区間長，非出現区間長，出現と非出現の発生率の類似性を，以下の確率モデルを用いて評価する．

$$p(e_i) = \Pi_{j=a}^{b} \; p_A(d_j|a_j = A)p(a_j = A) \; p_D(d_j|a_j = D)p(a_j = D). \qquad (1)$$

上式において，$p(a_j = A)$ と $p(a_j = D)$ は，出現，非出現の発生確率である．また，$p_A(d_j|a_j = A)$，$p_D(d_j|a_j = D)$ は，それぞれ出現区間長 d_j （$a_j = A$ のとき），非出現区間長 d_j （$a_j = D$ のとき）が観測される確率を表す指数分布である．すなわち，e_i 内で出現区間長，非出現区間長，出現と非出現の発生率の類似していれば，e_i 全体を観測できる確率 $p(e_i)$ が高くなる．最終的に，動的計画法を用いて，X 全体が最も高確率で観測できる最適な K 個のイベントの配置を求めることができる．

上記の映像セグメンテーションを行った後で，イベント e_i がバーストを含んでいるかどうか，以下の評価関数を用いて検証する．

$$BI(e_i) = \frac{T_A^{e_i}}{T^{e_i}} \times \int_0^{\infty} |\lambda_A^{e_i} e^{-\lambda_A^{e_i} x} - \bar{\lambda}_A e^{-\bar{\lambda}_A x}| \, dx, \qquad (2)$$

上式の第一項は，注目人物が長時間出現していればいるほど，e_i において重要な役割を担っているという重みを表わしている．第二項は，映像全体における平均出現区間長 $(1/\bar{\lambda}_A)$ と e_i における平均出現区間長 $(1/\lambda_A^{e_i})$ との逸脱度を表わしている．最終的に，しきい値以上の $BI(e_i)$ をもつイベントを，バーストを含んでいると判定し，トピックとして抽出する．

4 実験結果

本論文では，サスペンス，SF，コメディ，ドラマというジャンルからの4本の映画に対して実験を行った．ここで，それぞれの映画における主人公を注目人物とした．映像セグメンテーションに関しては，約80%のイベントが，意味的な一貫性をもつという良好な結果が得られた．主な要因としては，以下の2点が挙げられる．1点目は，実際の映画では，図2のようにイベントごとに出現区間、非出現区間長が均一ではないが，確率モデルを用いることにより，全体的に出現時間長，非出現時間長が長い（もしくは，短い）イベントにロバストに分割できた点である．

2点目は，注目人物と他の人物やオブジェクトとの“かかわり”を特徴づけるために，出現と非出現の発生率が非常に有用であった点である．具体的には，注目人物が話題の中心であるときは，ほぼ全てのショットに出現する．そして，他の人物と出会ったり，オブジェクトを発見すると，それらを映したショット（注目人物が出現していない）に切り替わるようになる．ゆえに，映像を意味的に解析するには，注目人物が出現していないこと（見えない情報）も重要である．

トピック抽出に関しては，注目人物が殺害されるトピック，逃走するトピック，おかしな行動をするトピック，恋愛のトピックなどを抽出することに成功した．しかしながら，注目人物の出現と非出現だけからでは，おかしな内容の会話やダンスのイベントなどをトピックとして抽出することはできなかった．このようなイベントに関しては，注目人物の動きや発話内容，他の人物の出現・非出現などを考慮する必要がある．

本論文の最大の問題点は，いかにして映像から注目人物の出現・非出現区間シークエンスを導出するかという点にある．図2から，ショットごとに，画面上における注目人物の向きや大きさが大きく異なっていることが分かる．そのため，従来の人物認識手法では，注目人物の出現・非出現を正確に注釈付けることは困難である．そこで，映像の時間的特徴を考慮した注目人物の認識手法を現在開発中である．この手法の詳細に関しては，別で議論する [3]．

5 まとめ

本論文では，注目人物の出現パターンに基づいて，映像をイベントに分割する手法，及びバーストを含んだイベントをトピックとして抽出する手法を提案した．今後，複数の登場人物やオブジェクトの出現パターンを組み合わせてイベントを詳細に解析するためには，それらを意味的に矛盾なく組み合わせるための“オントロジー [4]”が必要になってくると考えられる．

参考文献

[1] J. Monaco: “How to Read a Film”, Oxford University Press, 1981.

[2] J. Himberg et al.: “Time Series Segmentation for Context Recognition in Mobile Devices”, In Proc. of ICDM 2001, pp. 203 – 210, 2001.

[3] 清水，他：“制約充足問題に基づく，顔の向きによらない登場人物の認識”，第70回情報処理学会全国大会，2008.

[4] 杉原，他：“ビデオオントロジーの導入による映像イベントの体系化”，第70回情報処理学会全国大会，2008.