ビデオオントロジーの構築による映像イベントの体系化

杉原 ちえり† 上原 邦昭‡
†神戸大学工学部 ‡神戸大学大学院工学研究科

1 はじめに

近年，大容量の情報蓄積技術やブロードバンドの普及に伴って，映像コンテンツは急激に増加している．大量の映像の中から手動で映像を探すのは難しいため，ユーザが見たい映像を自動で探しだせるシステムが必要である．本研究では，映画の映像から特定のイベントを抽出することを考える．イベントの自動抽出システムにより，ユーザは映像中の自分の見たい部分のみをピンポイントで見ることが可能になる．

イベント抽出の問題点は２つ存在する．１つはコンピュータと人間の認識が異なるという点である．コンピュータは映像を色や音のデータ（RGB, 音量など）としてしか認識できないが，人間はオブジェクト (車，人など）やシーン（ドライブ，会話など）といった多様な認識が可能である．

２つめはコンピュータの知識が不足している点である．コンピュータは映像の意味内容に関する知識を持っておらず，ある程度，人手によりデータを与える必要がある．しかし，映像中に存在する物全ての知識を与えることは不可能なため，どのようにして網羅的，一般的なデータの記述を行うかが重要となってくる．

この２点を解決するために，ビデオオントロジーを導入する．ビデオオントロジーは映像分野に対するオントロジーである．代表例として LSCOM[1] が挙げられる．LESCOM は，約 1000 個の概念を持ったニュース映像に対するオントロジーであり，大量のニュース映像から災害や戦争などのイベントを検索するために利用される．本論文ではドラマや映画におけるビデオオントロジーを考える．オントロジーを構築することにより，意味内容に関する網羅的な知識を与え，オブジェクト，イベント認識の精度を向上させることが可能となる．

2 ビデオオントロジーと Raising

オントロジーとは対象とする分野を，概念とその関係を用いて体系化したものであり，コンピュータと人間が理解を共有できるように記述される．オントロジーの構築方法論はいくつか存在するが，本論文では Uschold と King の方法論 [2] を元に，オントロジーの構築を行った．Uschold と King の方法論によるビデオオントロジーの構築方法は次のとおりである．

1. 存在している概念を対象とする分野（ドラマや映画）から全て抜き出す．
2. 抜き出した概念をクラスタリングにより分類し，同義語の削除と重要語の選出を行うことによって，語彙の整理を行い，概念を決定する．
3. 決定した概念を元にオントロジーの階層構造を構築する．概念を検討し，上位にあたる概念（一般的な概念）と下位にあたる概念（特殊な概念）を，数の比率を考慮しながら決定する．
4. 概念の名前の決定と，概念定義を行う．

ビデオオントロジーの１つ目の問題は，オントロジーを構築する際のイベントの粒度をどう設定するかという点である．イベントの粒度というのは，どれを１つのイベントとして設定するかという大きさである．映像をイベントごとに区切る場合，区切り方は観点により大きく異なってしまう．例えば，パーティという大きい粒度のイベントは，会話，食事等の小さい粒度のイベントに分割可能である．

イベントの粒度を統一するために，本稿では SNAP-SPAN オントロジー [3] を利用した．SNAP オントロジーとは，同時世界のオントロジーであり，映像のキーフレーム内に存在する物質を概念として構築する．概念を抜き出す基準として，登場人物（動物）と人数，場所，注目している物体，時間，フレームサイズを用いている．SPAN オントロジーは継続世界のオントロジーであり，映像の１ショット内に存在するオブジェクトの変化を概念として扱う．概念を抜き出す基準として，行動，カメラワークを用いている．

また，SNAP オントロジーと SPAN オントロジー間で概念の粒度が同じになるように，語彙の整理と概念の決定を行う．例えば，SNAP の “人” という概念に対応する SPAN の “人の動作” の概念，人の一部である “腕” という概念に対応する “腕の動き” の概念など，SNAP で決定した各概念ごとに動きや変化をまとめていく．

もう１つの問題は，映像からの概念抽出には各概念の確率モデルが必要となるため，オントロジー中の全ての概念を映像から抽出することは困難であるという点である．そこで，Raising[4] という手法を用いて下位概念から一般的なイベントを抽出するという方法を考える．Raising とは，概念の階層を上げるという意味である．ビデオオントロジーにおいて，概念の階層を１段上げる（概念の親に当たる概念で置き換える）ことを考えると，子に当たる概念が抽出できた場合，親概念のイベントも記述することが可能になる．Raising を用いることによって，イベントの網羅的な記述と，映

Organization of Video Events by Building Video Ontology
Chieri Sugihara† Kuniaki Uehara‡
†Department of Computer Science and Systems Engineering, Kobe University ‡Graduate School of Engineering, Kobe University

像から取り出さなければいけない概念数の削減という利点が得られる．

実験で構築したSNAP，SPANオントロジーの一部を図1，図2に示す．このオントロジーを用いたRaisingにより，イベントを抽出することを考える．例えば，SNAPオントロジー中の“犬”という概念を，Raisingによって階層を上げると“動物”という概念となる．SPANオントロジーと組み合わせて，概念が検出されたショット内に“動物”の水平移動があった場合は「動物(犬)が歩く」というイベントが検出され，垂直移動があった場合は「動物(犬)が跳ぶ」というイベントが検出可能となる．

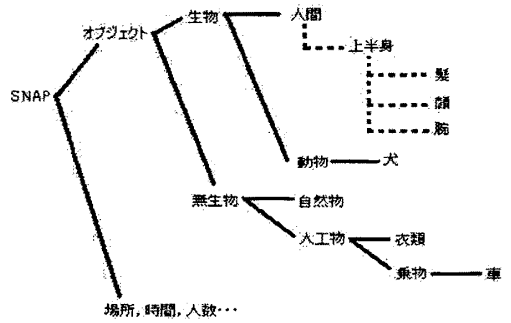

図1: SNAPオントロジー（直線はis-a関係，点線はpart-of関係を表す）

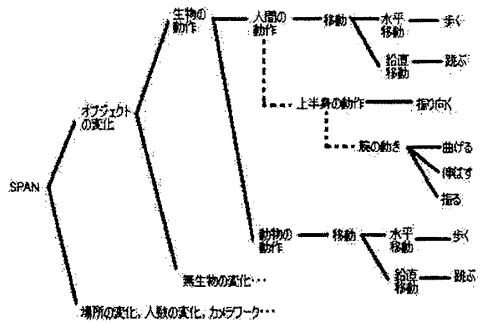

図2: SPANオントロジー

3 実験結果

実験にはホームドラマの一部の映像を用いた．まず映像をショットに分割し，そのうち1/3のショットを用いてオントロジーを構築した．次にオントロジーの構築に利用したショットのキーフレームから，2次元HMMによる学習[5]を行う．学習データには，キーフレームをマクロブロックに分割して，各ブロックに下位概念を割り当てたものを利用する．そこから，マクロブロックの特徴量（テクスチャ，色）とマクロブロック間の位置関係を元に2次元HMMによる確率モデルを構築する．

図3，図4は構築したモデルを元に，テスト用のキーフレーム(左)に下位概念を割り当てた例（右）である．図3では，キーフレームから犬（茶色部分），葉(緑色部分）という下位概念が抽出できたことが分かる．しかし，図4では人の顔が抽出できず，その他にあたる概念（黒色部分）であると認識されてしまった．2次元HMMでは，特徴量に色情報を利用しているが，各概念の色は映像の明るさ(昼の映像か夜，夕方の映像か）によって大きく異なってしまう．そのため，明るさの異なる映像ではきちんと認識が出来なかったと考えられる．また，マクロブロックを用いた特徴量の抽出を行っているため，ブロック内にノイズ（ブロックに割り当てた概念とは異なる概念の一部分）が入りやすいという欠点ももっている．

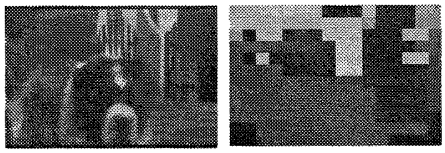

図3: 下位概念の抽出（成功例）

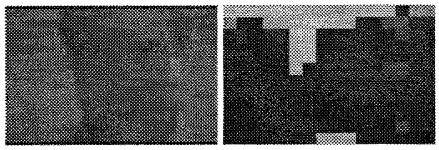

図4: 下位概念の抽出（失敗例）

4 今後の課題

今後の課題として，SNAPオントロジーにおける下位概念抽出の精度向上，ショット映像からのSPANオントロジーにおける下位概念の抽出，SNAP，SPANを組み合わせたイベントの記述できるオントロジーの構築が挙げられる．

参考文献

[1] J. R. Smith, “Large-Scale Concept Ontology for Multimedia”, IEEE Multimedia, Volume 13, pp 86-91, 2006

[2] M. Uschold and M. Gruninger, “Ontologies: Principles, methods and applications”, Knowledge Engineering Review, pp 93-155, 1996

[3] T. Bittner and B. Smith, “Granular Spatio-Temporal Ontologies”, In Proc. of AAAI Spring Symposium on FASTR, pp.12-17, 2003

[4] X. Zhou and J. Geller, “Raising, to Enhance Rule Mining in Web Marketing with the Use of an Ontology”, H. O. Nigro, S. E. G. Cisaro and D. H. Xodo (eds.), Data Mining with Ontologies: Implementations, Findings and Frameworks, pp 18-36, Information Science Reference, 2007

[5] 出野，白浜，上原, “映像検索のためのビデオオントロジーに基づいた自動アノテーション”, 映像情報メディア学会冬期大会, 2006