Keywords you should know. 第43回

知っておきたいキーワード

ビデオアノテーション

(正会員) 新田直子†

† 大阪大学 大学院工学研究科
"Video Annotation" by Naoko Nitta (Graduate School of Engineering, Osaka University, Osaka)
キーワード：意味に基づいた検索，メタ情報，アノテーション，シーン，MPEG-7

ビデオアノテーションとは

現在，ウェブ上の大量のテキスト，画像，映像などに対し，GoogleやYahooなどの検索エンジンを用いたキーワードによる検索が身近なものとなっています．ここでの検索は主に，ウェブページ内のテキスト情報と検索エンジンに入力されたキーワードの照合により実現されています．ウェブページ内の画像や映像などに対しては，ページ内において画像や映像の内容が説明されていることが多いため，このようなキーワードによる検索が可能となります．

しかし，テレビで放送される映像や，ディジタルカメラで撮影した画像，映像などには，内容を説明したテキスト情報が存在しないため，色や形などの画像特徴の類似性に基づいた検索が一般的となります．この場合，画像や映像はテキストに比べて情報量が非常に多いため，処理に時間がかかるとともに，映っている人物が誰かなど，意味的な情報は見た目から判断しにくい，という問題が生じます．したがって，現在の検索エンジンのように，キーワードなどを用いて簡単に，意味に基づいた画像・映像検索を実現するためには，各画像や映像に対し，その内容を表すメタ情報をテキスト形式で付与しておくことが重要となります．このようなメタ情報，もしくはメタ情報の付与を一般にアノテーションと呼びます．本稿では特に，映像を対象としたビデオアノテーションについて紹介します．

多くの映像ではその内容が時間的に変化することを考えると，ビデオアノテーションでは，映像全体のみでなく，映像内のある時区間（映像セグメント）に対して意味内容の記述が必要となります．記述される意味内容の基本としては，いつ（WHEN），どこで（WHERE），だれが（WHO），どのように（HOW），なぜ（WHY），何をした（WHAT）といった5W1Hに関する情報が想定されます．**図1**にビデオアノテーションのイメージ図を示します．

現在では，YouTubeなどの動画共有ウェブサイトにおいて，ユーザが人手でアノテーションを行える仕組みも取り入れられていますが，人間の手間を減らすため，画像処理などにより自動的にアノテーションを付与するための研究も多く進められています．以下では，映像としてテレビで放送される放送型映像（以下，単に映像と呼ぶ）を対象に，自動アノテーションを実現する方法について説明します．

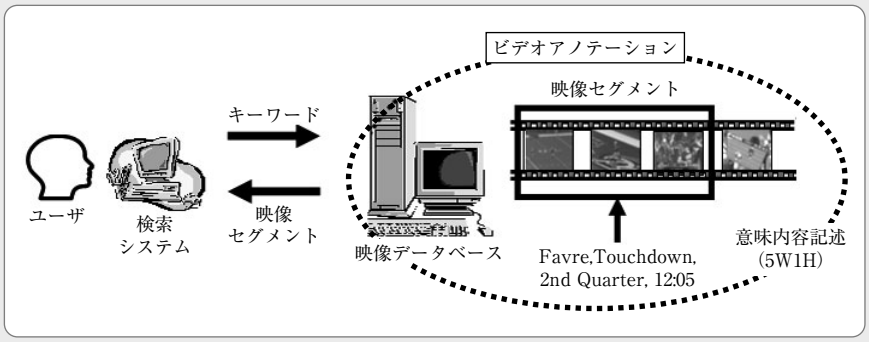

図1 ビデオアノテーションのイメージ図

映像セグメンテーション

まず，映像中のどの時区間にアノテーションを付与するかを考えます．映像は**図2**のように，最下層の一枚一枚の画像であるフレーム，同じカメラで撮影された連続したフレーム列であるショット，意味的なまとまりを持つ連続したショット列であるシーンというように階層化することができ，意味内容に関するアノテーションは多くの場合，意味的なまとまりを持つシーンに対して付与されます．

シーン列は，特定ジャンルや番組の映像において，ある定まった構造を持つ場合があります．例として，スポーツ映像は複数のプレイシーンにより構成され，各プレイシーンは一般に，野球では投球ショット，テニスではサーブショットなど，視覚的に非常に類似したショットから始まる，といった特徴を持ちます．そこで，**図3**のように，特定ジャンルや番組に対して定まる見かけの特徴を予め設定し，映像をシーンに分割する手法が多く提案されています．

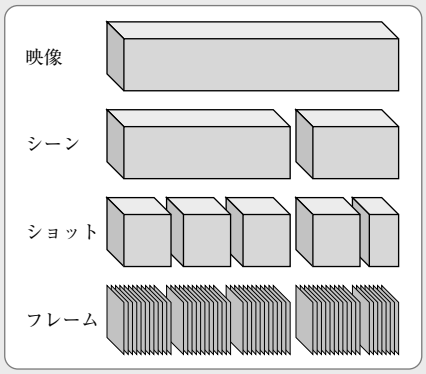

図2　映像の階層化

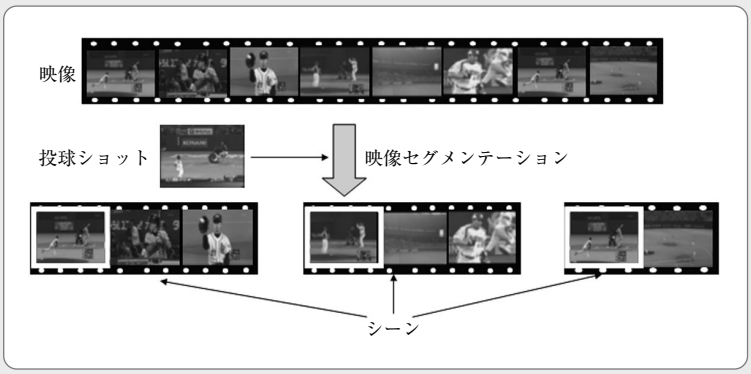

図3　映像をシーンに分割する手法

ルールに基づくアノテーション

映像には画像の他に，音声，音楽，効果音，雑音などのさまざまな音響，画像に存在する字幕やテロップ，音響情報の写しであるクローズドキャプションなどのテキスト情報が存在します．

例えば，野球のホームランのシーンでは，打者が打ったボールが観客席に入る様子が画像に映される他，観客の歓声が上がり，実況中継のアナウンサが「ホームラン」という単語や打者名など，そのシーンに関するキーワードを発話する，というように，シーンの内容によって決まったパターンが見られます．したがって，重要なシーンがどのような特徴を持つかを予めルールとして設定しておくことにより，自動的なアノテーションが可能となります．

また他に，シナリオや電子番組表，ウェブテキストといった映像と独立して作成されるテキスト情報も重要な情報源となります．例えば，スポーツの試合に対しては，プレイ名や選手名がその発生時刻などとともに記述された試合結果情報がウェブ上に存在します．この発生時刻と，画面上で試合の進行状況を伝えるテロップから，文字認識により抽出した時刻情報の対応付けなどにより，該当シーンにプレイ名や選手名をアノテーションとして付与することも可能です．

このようにさまざまな情報の利用は，画像処理に必要な計算量を減らした上で，効率的に信頼性の高い意味内容の自動獲得を実現します．

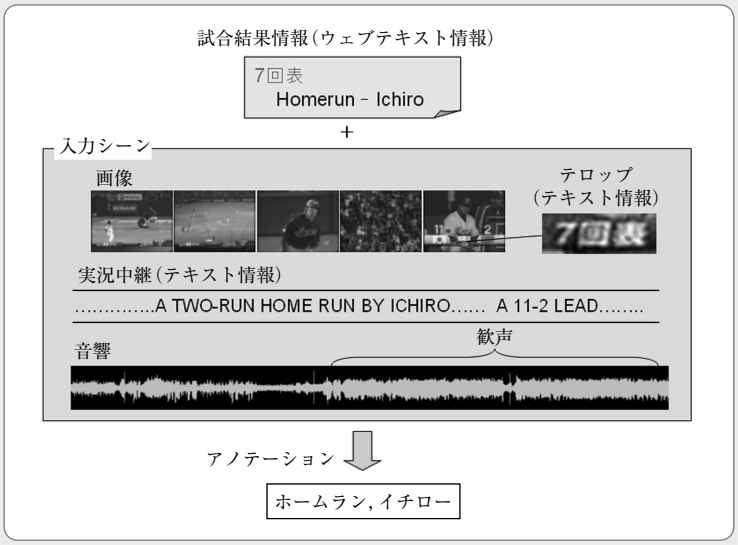

図4　ルールに基づくアノテーション例

学習に基づくアノテーション

ルールに基づいたアノテーションでは，各シーンが持つ特徴をルールとして予め人手で設定する必要があります．しかし，各シーンに対し，さまざまな例に共通する特徴を発見するのは簡単ではありません．

そこで**図5**のように，例えばいろいろな試合の映像からたくさん集めたホームランシーンといった学習用データから，色やエッジなどの画像の情報，音量などの音響の情報，キーワードの有無などのテキストの情報などを取出し，共通のパターンを統計的に自動で学習するといったアプローチについても盛んに研究されています．このようなアプローチは，どのような学習用データを用意するか，また学習方法やどのような情報を取出すかによって性能が大きく左右されるという問題もあるものの，ルールを設定する人間の負担を減らし，かつ汎用性の高いルールを獲得できるという利点があります．

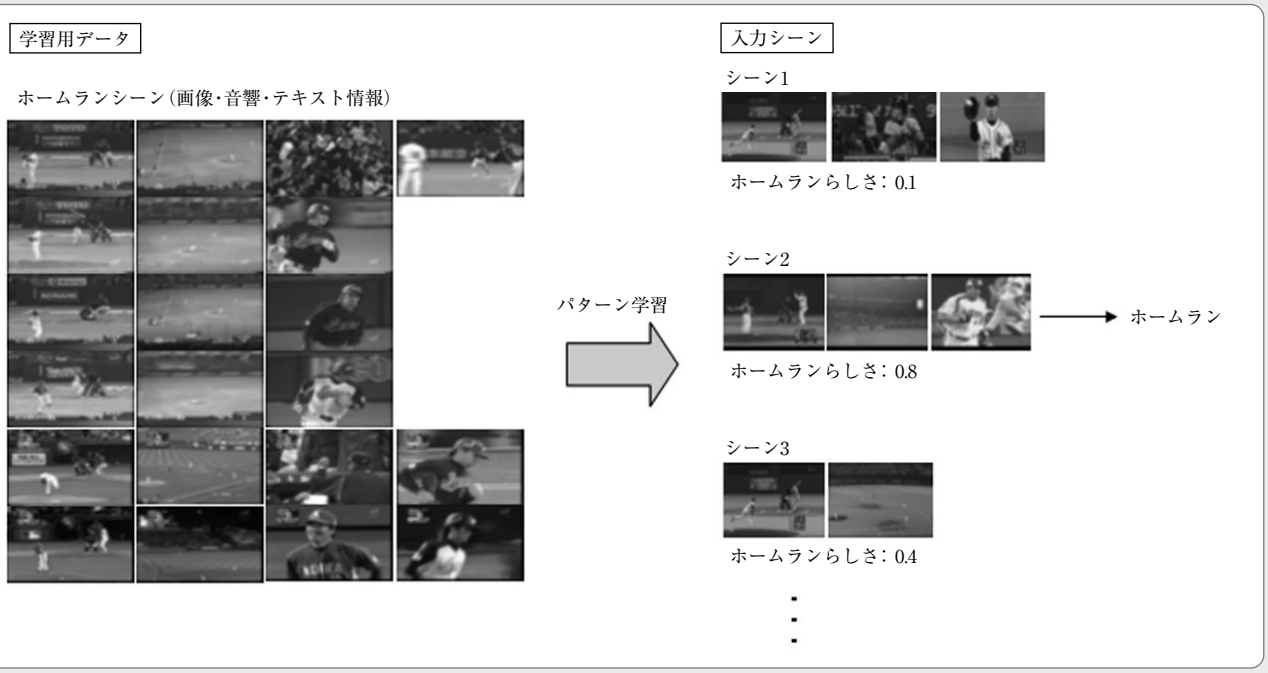

図5　学習に基づくアノテーション例

む　す　び

ビデオアノテーションの記述方式として，MPEG-7（Multimedia Content Description Interface）が国際標準化されており，主に画像や音響から取出される色，形，音量といった低レベルな情報から，5W1Hのような高レベルな意味内容情報まで幅広い情報の記述が可能となっています．今回紹介したビデオアノテーションをMPEG-7などで記述することにより，今後，映像を対象とした意味内容に基づいた検索や編集などのアプリケーションのさらなる進展が期待されます．

（2009年5月25日受付）

新田　直子（にった　なおこ）　1998年，大阪大学基礎工学部情報工学科卒業．2003年，同大学大学院博士課程修了．2002年～2004年，日本学術振興会特別研究員．2003年～2004年，コロンビア大学客員研究員．現在，大阪大学大学院工学研究科講師．メディア理解に関する研究に従事．博士（工学）．正会員．

キーワード募集中

この企画で解説して欲しいキーワードを会員の皆様から募集します．ホームページ（http://www.ite.or.jp）の会員の声より入力可能です．また電子メール（ite@ite.or.jp），FAX（03-3432-4675）等でも受け付けますので，是非，編集部までお寄せください．

（編集委員会）