解 説

映像情報検索とその評価技術の最前線

KDDI 研究所

帆足啓一郎 *hoashi@kddilabs.jp*
菅野　勝 *sugano@kddilabs.jp*
松本一則 *matsu@kddilabs.jp*

映像情報検索の時代

地上波ディジタル放送の開始や，HDD や DVD などの大容量メディアを搭載したレコーダの普及に伴い，一般ユーザでも大量の高画質映像データを蓄積できる環境が急速に広まっている．また，ブロードバンドネットワーク環境の拡大に伴い，オンデマンド型の映像配信サービスが盛んになるなど，ユーザが閲覧できる映像データはここ数年急速に増加している．

しかし，ユーザが気軽に大量の映像データを蓄積することができるようになる一方，蓄積された大量のデータの中から閲覧したい映像を探しだすことが難しくなるのは言うまでもない．近年では EPG（電子番組表）に含まれるキーワードを基に自動録画する機能や番組単位での検索を可能にする機能を持つ機器も利用できるが，通常 EPG は番組単位でしか付与されていないため，たとえば録画したニュース番組の中から，興味のある話題だけを見たいなどといったニーズには対応が困難である．

また，映像データの細かい中身に対するメタデータの形式として，MPEG-7 や TV-Anytime などが規定されているが，テレビ番組などで発生する細かい事象のすべてに対してメタデータを付与する作業は高コストであり，同様にしてすでに流通している映像データにメタデータを付与することも膨大な作業コストが必要となる．そこで，大量の映像データの中身を解析して検索を行う content-based な映像情報検索技術のニーズが急速に高まっている．

本稿では，近年研究が盛んに行われている content-based 映像情報検索技術，および同技術の評価方法の最新動向について解説する．具体的には，映像情報検索技術の現状を解説し，映像情報検索技術評価の最新動向として，映像情報検索のための大規模テストコレクションの代表的な取り組みである TRECVID を紹介する．そして，最後に今後の映像情報検索技術の発展に向けた課題などを述べる．

映像情報検索技術の現状

映像情報検索技術という観点からは，前述した content-based 映像検索やメタ情報の効率的な付与のため，映像情報のインデクシング技術の研究が盛んに行われている．これまで，映像情報のインデクシング技術としてカット点検出や無音検出などの映像データの物理構造を定義するインデクシング（構造的インデクシング）が数多く検討されており，カット点表示などは商用の映像アーカイブシステムでも一般的な機能となっているが，構造的インデクシングではより高度な検索要求，たとえば特定イベントや話題などの検索を実現することができないため，これらを可能にするために，映像データの意味構造を定義するインデクシング（意味的インデクシング）が求められている．これまで人手が介在する必要があった意味的インデクシング技術においては，映像データの増加に伴う映像制作コストの増大を抑制するため，自動化や半自動化により実現できることが望まれている．代表的な意味的インデクシングとしては，スポーツ映像からのハイライト生成[1]，イベント検出[2]，ニュース映像からのアンカーショット検出[3]，話題分割[4]，および映画のジャンル分類[5] などが挙げられる．

これらの既存文献では，実証実験に使用した映像データにおいてはある程度の結果が得られることが報告されているが，少数の映像データにおける限定された評価手法の範囲内での結果しか示されていないものがほとんどである．これは，たとえばあらゆるニュース映像に対して十分な精度を保証するための映像データおよび正解（ground truth）の入手が困難であること，さらに意味的インデクシング技術の評価は一般的に主観的側面に立

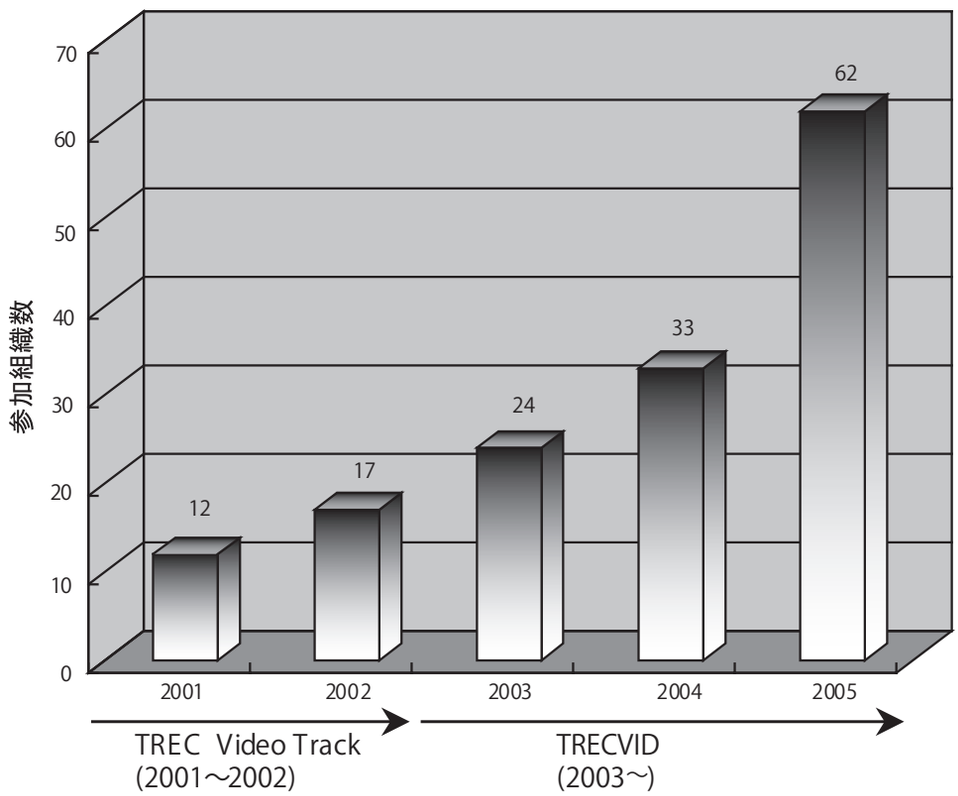

図 -1　TREC Video Track および TRECVID への参加組織数

つことが多いため，評価尺度も確立されていないことに起因する．また，スポーツ映像や TV 映像を実験に用いている場合は，著作権の問題からそれらを共通的に利用することができないため，方式の優位性を示すための比較実験を行うことも困難である．したがって，大規模かつ共通的な映像データを実験に利用できるための環境を構築することが望まれている．

映像情報検索技術の評価に関する最新動向

前述の背景により，テキスト情報検索分野における TREC や NTCIR のような，共通の評価プラットホームが映像情報検索分野にも必要となっている．そのような要望から TREC Video Retrieval Workshop（TRECVID）が始まった[☆1]．現在では，TRECVID で提供されるテストコレクションが，映像情報検索技術の性能を評価するためのデータとして，デファクトスタンダードになっている．本章では，映像情報検索技術評価に関する最新動向の代表例として，TRECVID での取り組みについて紹介する．

TREC Video Retrieval Workshop（TRECVID）

TRECVID は，当初 TREC の中の 1 つのサブタスクとして，TREC 2001 と TREC 2002 に含まれていたが，2003 年からは，独立したワークショップとして開催されており，今年の TRECVID 2005 で 3 回目の開催となる．映像情報検索研究への関心の増加に伴い，TRECVID の規模も年々拡大しており，TRECVID 2005 には 62 組織（主催者情報による）が参加を表明している．**図 -1** に，2001 年から 2005 年までの TREC Video Track および TRECVID への参加組織数の推移を示す．

TREC と同様，TRECVID も大量の実験用データとともに，共通のタスクおよび各タスクに対する評価基準を設定している．参加者は，TRECVID が配布する実験用データに基づいてタスクに取り組み，与えられた期限までに各タスクの実験結果を主催者である NIST（National Institute of Standards and Technology）に送付する．NIST では，タスクごとに設定された評価基準に従ってすべての実験結果を評価し，その結果を参加者に返送するとともに，TRECVID で報告を行う．

TRECVID は，映像情報検索技術評価のためのテスト

☆1　http://www-nlpir.nist.gov/projects/trecvid/

	学習データ	評価データ
TRECVID 2003	ABC World News Tonight (1998.1~4, 58 files) CNN Headline News (1998.2~4, 59 files)	ABC World News Tonight (1998.4~6, 58 files) CNN Headline News (1998.4~6, 59 files)
TRECVID 2004	ABC World News Tonight (1998.1~6, 116 files) CNN Headline News (1998.2~6, 118 files)	ABC World News Tonight (1998.10~12, 60 files) CNN Headline News (1998.10~12, 68 files)

表 -1　TRECVID 2003, 2004 における学習データと評価データ概要

コレクションとしては，世界で初めてのものである．データ量も豊富な上，データに含まれる映像データの（研究目的での）利用権も許諾されていることなどから，映像情報検索研究を進めるためには理想的な実験データであるといえる．また，TRECVID のタスクは，構造的インデクシングであるカット点検出にはじまり，意味的インデクシングとしての高次特徴抽出に至るまで，非常に幅が広く，かつタスクの内容や評価基準も TRECVID 参加者の間での議論を経て決定されている．以上の点から，映像情報検索技術の分野においては，TRECVID のテストコレクションがデファクトスタンダードとしての地位を確立しているといえる．

以下，TRECVID のテストコレクションに含まれる映像データ，および共通タスクとその評価基準について，それぞれ紹介する．

TRECVID 実験データ

TRECVID では，各タスクの実験を行うための映像データとして，ニュース番組を中心とした大量の MPEG-1 ファイルが用意されている．TRECVID 2003 では，1998 年 1 ～ 6 月に米国で放送された「ABC World News Tonight」および「CNN Headline News」の 2 つのニュース番組（241 本，約 120 時間分）と，米国議会での議論の模様などが収録されている CSPAN（25 本，約 13 時間分）の映像から構成されている．また，TRECVID 2004 では，TRECVID 2003 の実験データに加え，新たに 1998 年 10 ～ 12 月に放送された「ABC World News Tonight」と「CNN Headline News」（128 本，約 64 時間分）が追加されている．

TRECVID では，上記の映像データを放送日に従って 2 つのグループに大別している．この 2 グループの内，時系列的に古い方を「学習データ（development data）」，新しい方を「評価データ（test data）」と定義している．後述する TRECVID のタスクの多くでは，学習データを基に実験システムの開発ならびにチューニングを行い，評価データに対する実験結果を提出する手順を採っている．**表 -1** に，TRECVID 2003 および 2004 における学習データと評価データの概要を示す．

また，TRECVID の映像データには，以下の情報が提供されている．

- 共通カット点情報（Common shot boundary）
  すべての映像ファイルに対し，カット点検出処理を行った結果が提供されている．このカット点検出結果によって分割された個々の「ショット」が，後述する Feature extraction および Search の各タスクの評価の単位として利用されている．また，各ショットには，キーフレーム（各ショットを表す代表的なフレームの静止画像）も併せて付与されている．
- 音声認識結果
  個々の映像ファイルから抽出されたオーディオ情報に対し，音声認識処理を実行した結果が付与されている．

上記の情報のほか，各映像ファイルのクローズドキャプション情報も提供されてはいるが，後述する TRECVID 2004 のタスクでは，映像解析で使用するテキスト情報は上記の音声認識結果から得られる情報に限定されており，クローズドキャプション情報の利用は禁止されている．

さらに，TRECVID 2003 では，参加者の有志によって，実験データに含まれる映像ファイルに対しアノテーション情報を付与する作業（common feature annotation）が行われている．この作業で付与されるアノテーション情報は，各ショットでの発生事象（例：「人物が歩いている」「飛行機の離陸」），撮影環境（例:「スタジオ」「屋内」「屋外」），および出現する被写体（例：「動物」「人間」「人工物」）など，合計 133 種類に及んでいる．実際の作業では，参加者有志で TRECVID 2003 学習データ（約 60 時間）をファイル単位で分担し，IBM Research が開発したアノテーションツールを利用して行われた．その結果，アノテーションが付与された映像データとしては，世界最大規模のデータが構築された．

TRECVID では，映像ファイルに対する上記のさまざまな情報を参加者に提供することにより，カット点検出や音声認識といった映像解析の要素技術を有しない組織であっても，参加が容易になっていることも大きな特徴

例 1：瞬時カット

例 2：ディゾルブカット

例 3：ワイプ効果付きカット

図 -2　TRECVID 映像データにおけるカット点の例

の 1 つといえる．

TRECVID タスク

前述の通り，TRECVID では毎年複数のタスクを設定しており，タスクごとに評価基準を設けている．以下，筆者らが参加した TRECVID 2004 において設定されたタスクの概要と，各タスクに対する主な参加者のアプローチについて，それぞれ簡単に紹介する．

◆ Shot boundary detection（カット点検出）

［タスク概要］

Shot boundary detection タスク（以下，SBD タスク）の目的は，実験対象映像ファイルに含まれるカット点，すなわち，各ショット間の切替点を自動的に検出することである．TRECVID の SBD タスクでは，カット点を「瞬時カット」と「特殊カット」の 2 種類に大別している．瞬時カットとは，ショットが瞬間的に切り替わるカット点のことであり，特殊カットは，複数フレームにまたがってショットが切り替わるカット点である．特殊カットの例としては，ショット切り替え前の映像がフェードアウトしながら次のショットの映像がフェードインするディゾルブ型のカット点や，ショット切り替え前の映像が次のショットの映像に水平方向や垂直方向に押し出されたり上書きされるワイプ型のカット点などが該当する．**図 -2** に，上記の各カット点の例を示す．

SBD タスクの評価基準としては，全正解ショット切替点のうち，検出することができたショット切替点の割合（再現率，Recall）と，検出されたショット切替点中の正解ショット切替点の割合（適合率，Precision）が採用されている．また，特殊カットに対しては，上記の指標に加え，Frame-recall と Frame-precision という指標も採用されている．Frame-recall および Frame-precision は，カット点の有無だけでなく，区間としてどの程度正確に検出できているかを示す指標である．

また本タスクでは，カット点検出の精度を示す上記の各指標に加え，カット点検出に必要な処理時間（complexity）についての参加者からの報告が必要となっている．具体的には，カット点検出に要した総処理時間，MPEG デコード処理時間，および実験で利用した計算機の CPU の情報が報告事項として求められている．

［主な手法］

TRECVID で提供されているような MPEG 符号化された映像データからのカット点検出技術としては，デコードを必要とせずに符号化パラメータを使用するものが一般的に主流となりつつあるが，TRECVID の SBD タスクにおいては，より時間をかけて高精度な検出を実現するために，デコードを経た画像データ上の特徴を使う組織が多い．比較的よく利用される特徴は HSV や YUV などの色ヒストグラムであり，画面全体というよりはたとえばフレームを 16 分割したブロック内のヒストグラムを用いる．また，特にディゾルブやフェードなどの特殊カット点を検出するため，移動窓（RMIT 大学など）や複数ペアワイズ（IBM など）などを用いて数フレームから十数フレームに渡る特徴変化を観測することが多く，これらの組織は精度の面では上位についている．これに対し筆者らの方式は，DCT 係数の DC 成分から構

成される画像（DC 画像）を簡易復号し，DC 画像上で得られる輝度や色差，エッジなどの特徴を利用して特殊カットを含むカット点の判定を行う．精度の面では参加した 18 組織のうち 6 位程度であるが，処理の面では参加組織の中で最も高速であった．

◆ Story segmentation（話題分割）

［タスク概要］

Story segmentation タスク（以下, SS タスク）の目的は，映像データを意味的な単位，すなわち「話題」に分割することである．TRECVID 2004 では，実験データとしてニュース番組を利用している．ニュース番組は，一般的には複数の話題によって構成されているが，SS タスクでは，評価データに含まれる個々のニュース番組を構成する話題の境界（話題分割点）を自動的に検出することが目的である．従来，テキスト情報検索の分野でも話題分割の研究は進められてはいるが，本タスクでは，テキスト情報に基づく従来の話題分割技術に対し，映像から得られる特徴の利用方法について評価することが目的の 1 つであるとされている．

本タスクでは，以下の 3 つの実験条件を設定しており，参加者は，上記 3 条件のそれぞれに該当する実験結果を送付することが義務付けられている．

- Audio + Video
  映像から抽出される映像特徴量ならびに音響特徴量のみに基づく話題分割．
- Audio + Video + ASR
  上記の特徴量に加え，音声認識結果から得られるテキスト情報を利用した話題分割．
- ASR Only
  音声認識結果から得られるテキスト情報のみを利用した話題分割．

本タスクでは，参照話題分割点の前後 5 秒以内，合計 10 秒の区間内に検出された話題分割点を正解とみなし，再現率と適合率をそれぞれ算出し，評価基準として採用している．

［主な手法］

本タスクの参加者の多くは，ニュース番組における話題分割点の基準として，ニュース番組のメインキャスターが映っている「アンカーショット」を抽出し，新たにアンカーショットが発生した個所を話題分割点として検出する手法によって，話題分割を行っている．

IBM Research と Columbia 大の合同チームは，アンカーショット抽出結果に加え，さまざまな中レベル特徴量（人物検出，撮影環境（室内 or 室外），話者変化検出など）を抽出し，話題分割点と相関の高い中レベル特徴量を，Information Bottleneck 理論に基づいて推定し，話題分割に適用する方式を採用し，約 61％の F-measure を達成した．

また，筆者らは，アンカーショット検出などといったニュース番組特有の映像解析結果は使用せず，各ショットから動き情報，色配置情報などの汎用的な低レベル特徴量を抽出し，サポートベクターマシン（SVM）を利用して話題分割点が含まれるショットを識別する手法に基づき，話題分割を行った．その結果，TRECVID 2004 の Story segmentation タスク全参加者の中で最高の F-measure（約 69%）を達成した．

ID	Feature（和訳）
28	Boat/ship（船舶）
29	Madeleine Albright（オルブライト元・米国国務長官）
30	Bill Clinton（クリントン元・米国大統領）
31	Train（列車）
32	Beach（砂浜）
33	Basket scored（バスケットボールの得点シーン）
34	Airplane takeoff（飛行機の離陸）
35	People walking/running （2 人以上の人間が歩いている／走っている）
36	Physical violence（暴力シーン）
37	Road（道路）

表 -2　TRECVID 2004 Feature extraction タスク feature 一覧

◆ Feature extraction（高次特徴抽出）

［タスク概要］

Feature extraction タスク（以下, FE タスク）の目的は，指定された事象（feature）が出現する個所を，分析対象映像データから検出することである．具体的には，前述の Common shot boundary によって設定されたショットの中から，課題として与えられた feature が出現するショットを検出するタスクである．TRECVID 2004 では，「Boat/ship（船）」「Bill Clinton（クリントン元大統領）」「People walking/running（2 人以上の人間が歩いている／走っている）」など，10 件の feature が課題として設定されている．**表 -2** に，TRECVID 2004 FE タスクにおける feature の一覧を示す．

本タスクの評価基準は，送付された検索結果（スコアによって順位付けられた最大 1,000 件のショット）の再現率，適合率，および平均適合率（Average precision）である．これらの指標の算出に必要な正解データは, TREC でも採用されている「プーリング方式」によって構築されている．具体的には，参加者が送付した実験結果に含まれるショットのみを評価対象として，

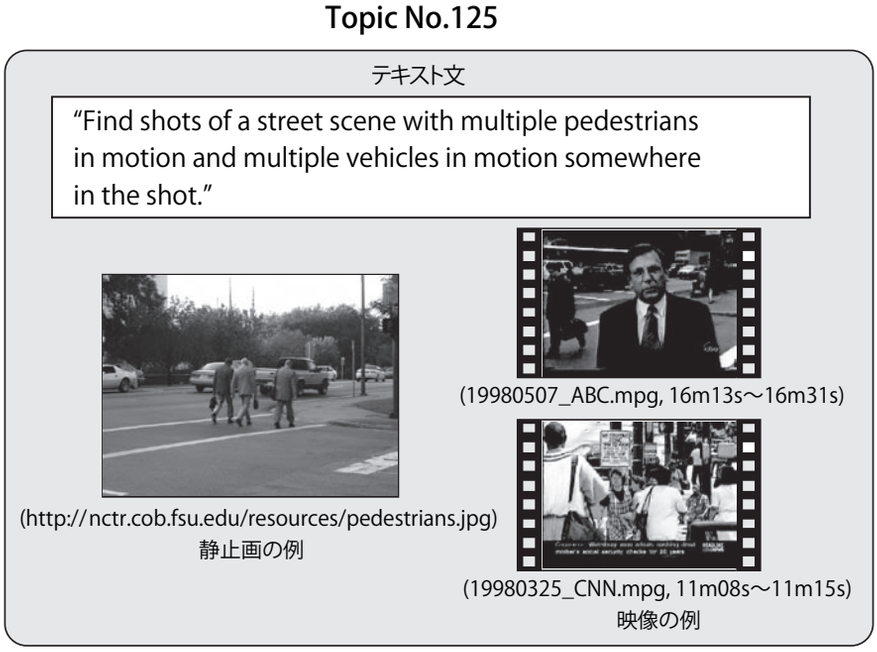

図 - 3　TRECVID 2004 Search タスクの Topic 例

TRECVID の評価者が目で検出対象 feature の出現有無を確認した結果が，正解データとして採用されている．

［主な手法］

本タスクでは，参加者のほとんどが映像データから抽出する視覚的な特徴と，音声認識結果から得られるテキスト情報を組み合わせた手法を採用している．

Carnegie Mellon 大は，各ショットから検出された低レベル特徴量（カラーヒストグラムなど），および音声認識結果や画面上の字幕スーパーに対する OCR 結果から得られるテキスト情報を基に，SVM によって common feature annotation で付与されたすべての事象の有無を検出し，その結果を基に，相関の強い高次特徴を検出している．

IBM Research は，cross feature ensemble learning (CFEL) という手法を利用し，高精度な高次特徴抽出を行っている．CFEL とは，視覚的特徴量とテキスト情報のそれぞれの特徴量に基づいて SVM で構築された高次特徴抽出器に基づき，アノテーション情報が付与されていない映像ファイルに対してそれぞれの SVM を適用して得られた高次特徴抽出結果を疑似学習データとして利用して得られる pseudo モデルも併せて構築する方式である．評価時には，構築された複数の高次特徴抽出器の結果を，さまざまな方法で統合している．

◆ Search（検索）

［タスク概要］

Search タスクの目的は，与えられたクエリを満たすショットを効率的に検索するシステムの開発である．タスクの目的自体は，FE タスクと類似しているが，本タスクでは，検索実験に人間が介在していることが前提となっている点が，FE タスクと大きく異なっている．すなわち，検索精度とともに，検索システムのインタフェースも評価の対象となっているのが，本タスクの特徴といえる．また，検索課題にあたる Topic も，FE タスクのそれよりも複雑であるほか，Topic 自体がテキスト文に加え，正解ショットのサンプル（映像と静止画）によって提示されている点も本タスクの特徴である．**図 -3** に，TRECVID 2004 の Search タスクにおける Topic の例（Topic No.125）を示す．図 -3 の例では，Topic の内容を表すテキスト文とともに，Topic に対する正解の例として，静止画が 1 点，映像（ショット）が 2 点示されている．

本タスクでは，Manual と Interactive の 2 つの実験条件が設定されている（**図 -4** 参照）．Manual 実験では，与えられた検索課題（Topic）を基に，システム利用者が検索システムに入力するクエリを作成し，得られた検索結果の精度の評価を行う．一方，Interactive 実験では，Manual 実験と同様，与えられた Topic を基に，システム利用者がクエリを作成し，検索システムに入力する．そして，その結果得られた検索結果などを基に，システム利用者がクエリを修正し，再検索を行うことができる．この再検索処理は，1 Topic あたりの制限時間（15 分）以内であれば，何度でも行うことができる．検索の精度評価は，最終的な検索結果に対して行われる．また，

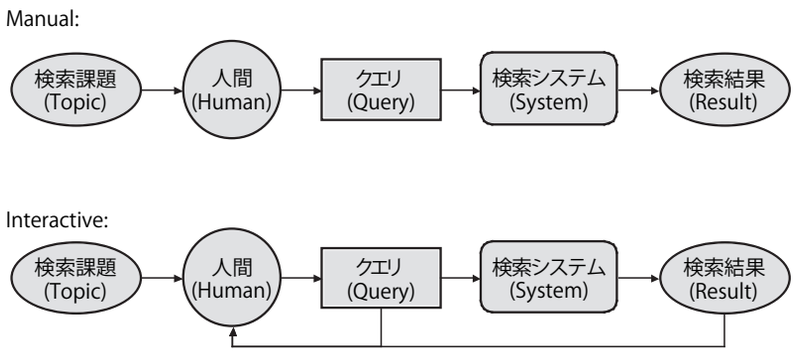

図 -4　TRECVID Search タスク実験条件概要

本タスクでは，上記の 2 条件のほか，検索時に一切人間が介在しない Full automatic 条件での実験も，非公式な実験として実施された．

本タスクの評価基準は，FE タスクと同様，再現率・適合率・平均適合率である．また，評価のための正解データも，FE タスク同様，プーリング方式によって構築されている．

［主な手法］

従来，Search タスクでは，音声認識結果から得られるテキスト情報を中心とした検索手法の有効性が確認されているが，TRECVID 2004 の参加者の多くは，テキスト情報を単独で利用した検索方式と，テキスト情報に視覚的特徴を組み合わせた検索方式を比較した実験を行い，ほとんどの場合，テキスト情報と視覚的特徴を組み合わせた検索方式の優位性が報告されている．

Lowlands（CWI Amsterdam と University of Twente の共同チーム）は，ショットの映像全体から抽出される視覚的特徴，ショットのキーフレーム（静止画）から得られる視覚的特徴，および音声認識結果から得られる言語的な特徴を表す確率モデルを 3 種類構築し，それぞれのモデルを単独で利用した検索と，複数のモデルを組み合わせた検索の比較実験を実施し，複数モデルによる検索の有効性について報告している．

また，シンガポール国立大は，まず Topic をあらかじめ定められたカテゴリに分類し，次に，Topic が分類されたカテゴリごとに有効な特徴量を選択もしくは重み付けしてから検索を行う手法により，高い検索精度を達成している．

TRECVID 2005 に向けて

TRECVID 2003 および 2004 の実験データに含まれる「ABC World News Tonight」「CNN Headline News」の各番組の映像ファイルは，当初から 2 年計画で利用されることになっており，TRECVID 2005 では新たな映像ファイルが実験データとして用意されている．具体的には，CNN, NBC などの英語ニュース番組（約 74 時間分）に加え，アラビア語のニュース番組（約 43 時間分）と中国語のニュース番組（約 52 時間分）が追加された．

TRECVID 2005 で実施されるタスクは，前年から継続される SBD タスク，FE タスク，Search に加え，新たに Low-level feature extraction タスクが加わる．Low-level feature extraction とは，ズームイン・アウト，パン，チルトなどのカメラモーションが出現しているショットを抽出するタスクである．同タスクは，SBD タスクと FE タスクや Search タスクの難易度にかなりのギャップがあるとの参加者の指摘を受け，新たに開催が決まったタスクである．

また，前述の通り，これまでの FE タスクや Search タスクでは，音声認識結果などから得られるテキスト情報に基づく検索手法が主流になっており，映像検索技術の発展という TRECVID の主旨に沿った成果が得られていない．そのため，TRECVID 2005 では，実験的な試みとして，「BBC Rushes」という実験データが用意されている．このデータは，英国の放送局 BBC が番組制作時に撮影した映像の素材を集めたものであり，ナレーションや字幕スーパーなどといった効果がほとんど施されていない，生の素材に近い映像データである．「BBC Rushes」はほぼ未加工な映像データゆえ，音声認識処理を実行しても有益なテキスト情報を得ることは難しいと想像されるが，逆に，映像データから得られる映像・音響特徴量のみに基づく映像情報検索研究への取り組みを促進させる実験データとなり得る．ただし，TRECVID 2005 では，BBC Rushes を利用したタスクはまだ設定されておらず，当面は同データによって設定できるタスクについての議論が進められることとなる．

以上に述べた TRECVID の説明からも明らかな通り，

TRECVIDでは実験データの整備からタスクの定義に至るまで，参加者間の議論によって決定される内容が多い．映像情報検索分野の主たる研究者の意見によって構築されるTRECVIDのテストコレクションは，映像情報検索性能評価のためのデータとしてデファクトスタンダードになっており，今後も注目する必要がある．

明日の映像情報検索技術とその評価

前述の通り，映像情報検索分野においては，TRECVIDの実験データが1つのデファクトスタンダードになりつつある．しかし，今後の映像情報検索技術発展のためには，まだ解決すべき課題は多い．以下，いくつかの課題について述べる．

(1) ユーザのニーズに即した映像情報検索課題の策定

TRECVIDのFEタスクやSearchタスクで設定されている検索課題は，必ずしもユーザのニーズに即していないと考えられる．たとえば，番組制作者などであればまだしも，大量の映像データの中から「船」が写っているショットを検索する一般ユーザはほとんどいないだろう．映像情報検索技術に対するニーズは，確たるものはなく，検索対象映像データやユーザ自身の利用シーンなどに大きく依存するが，実用的なニーズを視野に入れたタスクを設定する必要がある．

(2) 国内の映像データへの対応

TRECVIDでは，米国で放送されたニュース番組を中心として実験データを構築しているが，番組が制作される国によって，演出効果などが大きく異なることは明らかである．たとえば，日本のニュース番組では必ずといっていいほど複数のキャスターが番組を進行しているが，そのような場面はTRECVIDの実験データではほとんど見られない．それゆえ，国内のテレビ番組に有効な映像情報検索技術の開発のためには，国内で制作された映像データを中心とした実験データの構築が必須である．しかし，日本国内では，残念ながら国立情報学研究所が頒布している「映像処理評価用映像メディアDB」[6]以外には，映像情報検索技術評価のためのデータコレクション構築に向けた動きは見られない．また，「映像処理評価用映像メディアDB」は，収録されている映像データの量（2時間弱）がTRECVIDの実験データと比べると圧倒的に少ないほか，TRECVIDのタスクにあたる実験の枠組みが整っていないなどの問題があり，映像情報検索技術を評価するための実験データとしてはまだ不十分であるといわざるを得ない．

(3) テレビ番組以外の映像データへの対応

TRECVIDに限らず，多くの既存研究では，テレビで放送される番組を評価対象のデータとして利用している．テレビ番組も重要な映像データではあるが，その一方，家庭のホームビデオや携帯端末などで撮影されたパーソナル映像データや，監視カメラなどの映像データについても，映像情報検索のニーズはある．しかしながら，このような映像データを対象とした，良質な実験データはまだ存在しないのが現状である．

以上の課題を勘案すると，今後の映像情報検索技術の健全な発展のためには，TRECVID以外のデータコレクションの構築が必要であることは明らかであろう．特に，国内で制作されている映像データに有効な映像情報検索技術の発展のためには，多くのデータを保有しているテレビ放送局からの映像データの提供および使用許諾など，多方面からの協力体制が必要不可欠であるといえる．国外に目を向けると，フランスのINA[☆2]やMoving Image Archive[☆3]など，テレビ放送映像が含まれた映像アーカイブはいくつか存在するものの，収録されている映像の画質にばらつきがあるなど，映像情報検索研究の実験には必ずしも適していない．国内の映像データを対象とした良質なデータコレクションが構築できれば，我が国における映像情報検索研究の発展につながり，世界各国に先行した研究成果につながると期待される．

日本国内で着々と整備される映像放送や配信などのサービスの利便性をさらに向上させるためには，映像情報検索技術のさらなる発展が不可欠である．今後，映像情報検索技術の研究を促進させるためのデータコレクション構築に関する取り組みが国内でも本格的に開始されることを祈念する．

参考文献

1) Hanjalic, A. : Generic Approach to Highlight Extraction from a Sports Video, IEEE ICIP 2003, Vol.1, pp.1-4(Sep. 2003) など．
2) 新田，馬場口：放送型スポーツ映像の意味内容獲得のためのストーリー分割法，電子情報通信学会論文誌 (D-II)，Vol.J86-D-II, No.8, pp.1222-1233(Aug. 2003) など．
3) Huang, Q., Liu, Z. et al.: Automated Generation of News Content Hierarchy by Integrating Audio, Video, and Text Information, IEEE ICASSP 99, Vol.6, pp.3025-3028(Mar. 1999) など．
4) Boykin, S. et al.: Improving Broadcast News Segmentation Processing, Proceedings of IEEE Multimedia Systems, pp.744-749(1999) など．
5) Sugano, M., Furuya, M., Nakajima, Y. and Yanagihara, H. : Shot Classification and Scene Segmentation Based on MPEG Compressed Movie Analysis, IEEE PCM 2004, Vol.I, pp.271-279(Nov. -Dec. 2004) など．
6) 馬場口，栄藤，佐藤，安達，阿久津，有木，越後，柴田，全，中村，美濃，松山：映像処理評価用映像データベースについて，電子情報通信学会技術研究報告，PRMU2002-30(June 2002).

（平成17年7月14日受付）

☆2 http://www.ina.fr/index.en.html
☆3 http://www.archive.org/details/movies