

# 素材映像マネジメントシステム「素材バンク」の提案 ～素材映像を最大限に活用するために～

## Video Asset Processing and Management System "Video Bank" : For Maximum Usability of Video Footage

大久保 英彦† サイモン クリッピングデル† ミツ峰 秀樹† 佐野 雅規† 住吉 英樹†  
武藤 一利† 望月 貴裕† 河合 吉彦† 加藤 大一郎‡ 藤井 真人†  
Hidehiko Okubo Simon Clippingdale Hideki Mitsumine Masanori Sano Hideki Sumiyoshi  
Kazutoshi Muto Takahiro Mochizuki Yoshihiko Kawai Daiichirou Kato Mahito Fujii

### 1. はじめに

テレビで毎日放送されている多くの収録番組では、その制作のためにオンエアされる番組の時間尺をはるかに超える大量の素材映像が撮影・収録されている。番組制作者は多大な労力とさまざまなシステムを駆使し、加工や編集を経て素材映像のごく一部をオンエア番組として昇華させている。これまでも番組における素材映像の管理の重要性は制作者の間で認識されてはいたが、素材映像をシステムとして汎用的に管理することは困難な課題として残ったままである。その理由はさまざま存在するが、一つは素材映像を管理するためにメタデータ付与の必要がある点、もう一つは素材映像利用に技術的に非常に高いスキルが要求される点、が考えられる。

第一の問題点について考えると、例えば報道番組向けには、大量に蓄積された素材映像から所望の映像を見つけ出せる映像検索機能の実現が求められる。それを可能にするために、現状であれば、素材を蓄積する際に後の検索時に使うキーワードなどのメタデータ付与をおこなう作業が必要となる。映像は画像と異なり時間的な長さを持っており、それをすべて参照した上で区間を考慮したメタデータ付与を手で行うとすると、その作業はきわめて煩雑なものとなる。その煩雑な作業を膨大な量の素材映像に対して施すことは現実的には不可能である。

第二の問題点は、仮に大量の素材映像から検索して所望のイメージに近い素材映像を取り出せるシステムを構築したとしても、多くの場合はその取得した素材をそのまま利用できない点にある。特に映像合成や加工で用いられる素材映像は、周到な撮影計画のもとで撮影・収録されるものであり、そうでない素材を利用する場合には、きわめて高度で煩雑な映像加工処理を経て利用可能となる素材映像を生成する作業プロセスが必要となる。

本研究で提案する「素材バンク」はこの2つの問題点にフォーカスをあて、様々なジャンルにおける番組制作者の素材映像ハンドリングを向上させる素材映像マネジメントシステムをめざしたものである。「素材バンク」に映像素材を蓄積させると、処理時間のかかる映像解析処理が自動的に作動して検索や映像加工利用に有用なデータを生成する。つまり番組制作者は映像素材を「素材バンク」に蓄積させることで映像検索が可能になり、映像利用も容易になる。本研究では、撮影時にデータを取得するシステムとあわせて、「素材バンク」の要素技術を使用した一連の素材ハンドリング実験を行い、「素材バンク」の映像制作に対する有効性を検証した。

### 2. 「素材バンク」のシステムイメージ

「素材バンク」は、蓄積によって映像解析を行いメタデータ生成/付与をおこなう素材映像管理システムであるが、それを番組制作者が利用するための利用インターフェース、さらに撮影時に映像とともに後の処理で有用となるデータを取得するシステムをあわせてその全体を構成させている(図1)。なお、本研究でのメタデータの定義は、「映像の内容を記述したキーワードに加え、映像の特徴量を表す複合的な数値情報などを含めた付加情報全般」としている。

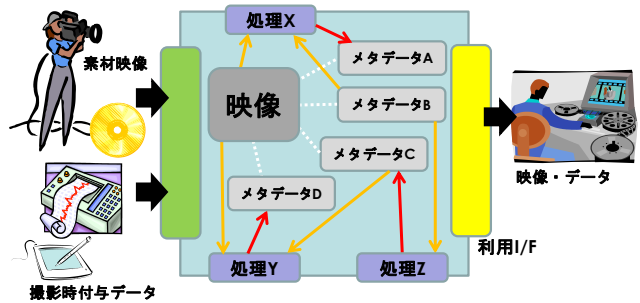


図1. 「素材バンク」での映像・メタデータの処理イメージ

### 3. 「素材バンク」を構成する要素技術

「素材バンク」では、現状で5つの映像解析処理を行う。各解析処理の結果、生成されたメタデータが素材映像に自動的に付与され、映像検索および映像加工などに利用可能となる。以下にそれらの概要を示す。

#### 3.1. 一般物体認識

映像検索のための被写体キーワード付与処理を自動的に行う。映像から代表フレームを抽出し、それを事前に学習させておいた複合的特徴ベクトル[1]に基づく特定被写体の判別器にかけることで、新たに「素材バンク」に蓄積された映像に特定の被写体が映っていた場合にその被写体名をメタデータとして映像に付与させることが可能となる。

#### 3.2. 顔検出・識別

映像検索のための人物の顔検出・人名付与処理を行う。顔検出は、肌色領域抽出の後にその抽出領域に対する顔検出処理を施すことで高速化と頑健化を図っている。人物識別については、撮影時に正面に近い顔を検出し、人物に関する可変テンプレートと照合させて[2]、照合人物に固有名をメタデータとして付与する。

† 日本放送協会 (NHK) 放送技術研究所

‡ 一般財団法人 NHKエンジニアリングシステム

### 3.3. 映像内容検索

映像検索の際にキーワードが付けられていない、あるいは検索意図をテキストで表現するのが難しいケースでの映像検索を可能とするために、複数カットにまたがる“場面”(シーン)単位で特徴を表すメタデータを算出し付与する[3]。これは、各シーンを構成する要素画像片(ブロック領域画像)の種類とそれぞれの出現比率を数値化したものである。これにより、キーワードを介さない類似シーン検索が可能となるほか、自身の持ち込んだ画像群をクエリーとして類似シーンを検索することもできるようにしている。

### 3.4. 撮影カメラパラメータ算出

映像合成に必須のメタデータともいえるカメラパラメータを素材映像に付与する。そのために、撮影時にカメラ姿勢やレンズ情報をリアルタイムで取得しデータ化するハイブリッドセンサー[4]を開発し、「素材バンク」に素材映像と同期して蓄積する。さらに「素材バンク」蓄積の後は、頑健な特徴点抽出とそれに基づくバンドル調整法によるカメラパラメータ推定[5]を行う。これはセンサーデータの精度向上に利用したり、逆に特徴点抽出が難しい部分のカメラパラメータ算出にセンサーデータを利用するという形で2つを融合することで、その頑健な算出が可能となる。算出されたカメラパラメータはメタデータとして利用インターフェースから取り出し、CG合成などに利用できる。

### 3.5. 映像領域分割

さまざまな映像合成に利用される、映像の指定被写体部分領域抽出を目的とした素材映像の時空間領域分割[6]を行う。領域分割結果として得られる各種領域情報が映像にメタデータとして付与される。これは利用インターフェースを介してグラフカット処理による合成素材用被写体領域抽出に適用させることが可能となる。

## 4. 実験

本研究では「素材バンク」の運用を想定し、素材映像取得から映像利用までの一連の処理を実行させる素材ハンドリング実験を行った。現状では素材映像を質および量の面で十分に準備できていないため、用途によって素材映像を分け、2つの利用ケースについて実験を行った。

#### 4.1. ケース#1: 素材映像検索

アーカイブされた番組映像を素材映像とみなし、その映像を基に被写体名キーワード自動付与処理とシーン特徴量算出処理を施した。それに基づいてキーワードおよびクエリー画像群による映像検索を行った。本実験で利用した映像は自然番組を中心とした224番組、約4500シーン分である。検索対象の被写体キーワードは、「山」や「海」などの一般的な対象を数百種類準備し、その学習データには正例・負例それぞれ300サンプル程度使用した。

#### 4.2. ケース#2: CG合成

素材映像撮影時にハイブリッドセンサーから取得したカメラ姿勢データと、素材映像の特徴点解析に基づくバンドル調整を融合したカメラパラメータ算出を行い、それに基づいたCG合成実験を行った。ここでは素材映像撮影時にあえて特徴点を取得しにくいフラットな背景を撮影するようなカメラモーションを演出することで、2つの処理の融合によるカメラパラメータ算出の効果を確認した。

## 5. 結果と考察

素材映像のキーワードによる検索実験について、図2にその検索結果例を示す。キーワード検索については、テスト映像として使用した番組映像に含まれていた「山」や「海」をキーワードとした場合については良好な検索結果となった。またクエリー画像検索においても、視覚的に類似する候補をピックアップすることができた。



図2. キーワード「海」によるシーン検索結果例

CG合成実験結果例を図3に示す。前述の演出によるテスト素材映像と、それに対する自動処理で生成したカメラパラメータを使用して生成させたCG素材を合成させた結果、精度よくCG合成が可能であることを確認した。



図3. カメラパラメータを適用したCG合成例  
(上部の“吹き出し”がCG素材)

## 6. まとめ

本研究で提案する「素材バンク」の考え方は番組制作時の素材映像マネージメントに対する一つの有効なアプローチとなりうるものである。今後は番組制作のニーズに応える機能拡充を図るとともに、メタデータ管理の具体的なフォーマット設計を行い、実際の番組制作に利用できるシステム構築を目指していく。

### 文 献

- [1] 河合吉彦, 藤井真人, “エッジ共起を考慮した回転・スケール不変な局所画像特徴と一般物体認識への適用”, 電子情報通信学会技術研究報告, vol.111, no.349, CS2011-78, IE2011-102, 2011, p.139-144
- [2] サイモンクリピンゲル, 藤井真人, “人物不特定可変テンプレートをを用いた顔画像認識の高速化”, 第11回情報科学技術フォーラム講演論文集, no.3, H-008, 2012, p.131-132
- [3] 望月貴裕, 佐野雅規, 藤井真人, “多重スケール画像片ワードヒストグラムを用いた映像検索”, 電子情報通信学会技術研究報告, vol.112, no.385, PRMU2012-89, MVE2012-54, 2013, p.75-80
- [4] 加藤大一郎, 武藤一利, 岡本浩幸, Moro Alessandro, 関敏一, 水上慎太郎, “放送用カメラの動きを計測するハイブリッドセンサー”, 第30回日本ロボット学会学術講演会予稿集, RSJ2012AC4I3-1, 2012
- [5] 三ッ峰秀樹, 武藤一利, 藤井真人, 加藤大一郎, “映像加工を考慮した頑健なカメラパラメータ取得法”, 画像電子学会第41回年次大会 予稿集, 2013, R6-1
- [6] 大久保英彦, 藤井真人, “分割領域の隣接状態変化を利用した動画領域分割手法の検討”, 第264回画像電子学会研究会講演予稿, no.264, 12-03-11, 2013, p.73-78