

# プレゼンテーションの知的撮影システム

An Intelligent System of Recoding Presentations

尾関 基行† 中村 裕一†‡ 大田 友一†  
Motoyuki Ozeki Yuichi Nakamura Yuichi Ohta

† 筑波大学 機能工学系, ‡ 科学技術振興事業団 さきがけ研究 21  
Institute of Engineering Mechanics and Systems, University of Tsukuba  
PRESTO, Japan Science and Technology Corporation

## 1 はじめに

マルチメディア環境やネットワークの進歩に伴い、映像は従来の一方向的なものから、誰もが自分で作ったコンテンツを蓄積、配信するといった双方向的なものへと変化していくことが予想される。これにより、例えば、説明映像、教材映像、メモ映像といったコンテンツが様々な場所で製作・蓄積され、それを再利用することが望まれている。しかし、実際に利用価値のある映像を作るためには、専門的な撮影技術と編集技術、さらには映像内容を記述するメタデータの作成も必要となるため、非専門家や個人が手を出すことは難しいという現状がある。

このようなことから、個人レベルでも手軽にマルチメディアコンテンツを作ることができる自動化システムが期待されている。本研究では、この問題に対し、手元作業シーンを対象とした映像の撮影と提示の自動化を行う知的システムの構築を行っている。本稿では、我々が構築しているシステムの構成と基本的な考え方について、その概要を述べる。

## 2 システム構成

本研究では、手元作業映像の取得を目的とし、システムを次のような3つの部分に分けて、それぞれ自動化へのアプローチを検討した。

撮影: 注目対象を追跡するカメラワーク

タグ付け: 話し手の発話と動作認識を利用したタグ付け

提示: ユーザの要求に応じたインタラクティブな提示

我々の構築しているシステムの概要を図1に示す。本システムでは、位置センサにより話し手や特定物などの位置を取得し、複台の首振りカメラを制御することで自動撮影を行う。各々のカメラで撮影された映像は、MPEG1エンコーダを通して保存される。また同時に、位置センサと音声認識を併用して話し手の動作認識を行い、映像へのタグ付けを行う。さらに得られたタグを基にして、視聴者が見たいと思う部分を効果的に提示することを目指す。

以下、撮影とタグ付けの部分について手法を簡単に述べ、提示については、実際に単純なプレゼンテーション

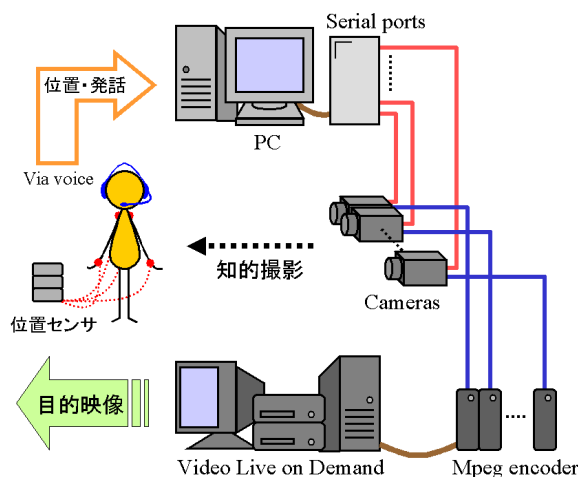


図1: システムの概要図

に適用した例を紹介する。

## 3 映像取得システムの概要

### 3.1 撮影とカメラワーク

手元作業映像取得におけるカメラワークの基本は、注目すべき対象を適切な大きさ・位置で画面に捉えることである。しかし、人や物体をただ単純に追跡すれば良いわけではなく、注目する対象に応じたカメラワークが必要となる。

これに対し本研究では、まず注目対象を「何のどういう状態に注目するか」、つまり、“撮影すべき対象物”と“撮影すべき対象の状態”に分けて考えることで、注目対象とカメラ設定の関係を簡潔にまとめた。現在の段階では、“対象物”として以下の4つを定義し、これによって追跡する対象と解像度を決定する。

話し手: 話し手自身を狙う

作業空間: 話し手の作業を狙う

注目物体: 注目すべき物体を狙う

注目場所: 注目すべき場所を狙う

“対象とする状態”については次の3つを用意し、それぞれの状態に応じたカメラ制御方法で追跡撮影を行う。

< 状況 >: シーン中での位置関係や軌跡



図 2: 自動映像切り替えの結果 (一部)

<操作>: 操作等, 細かい動きが行われている状態  
 <物体>: 対象物そのものの状態(形, 色, 静止状態等)  
 上記それぞれに適したカメラ制御を行うため, 本研究ではカルマンフィルタによる平滑化, 枠制御アルゴリズムを用いる [1]. これにより, 対象に応じて各処理で用いるパラメータを調節し, 目的に応じたカメラワークを設定することが可能である.

### 3.2 タグ付け

蓄積された映像を効果的に利用(検索・提示)するには, 映像の各部分に適切な記述を付加データとして与えることが必要となる. このような付加データには, カメラの設定, カメラの状態, 話し手の言動, 映っている対象の状態といった要素が挙げられ, これらについての記述を本研究ではタグと呼ぶ.

本研究では, 話し手の発話と動きの情報を用いて動作認識を行い, これを利用してタグ付けを行うことを考える. 動作認識については文献 [2] で詳しく述べているので, ここでは省略する.

表 1 に現時点で認識対象としている動作, およびそれに対応する指示詞と話し手の動きをまとめた. 動作の終了は手が体に近づいたこと, または画面内に仮想的に考えた枠から出たことにより検出する\*. これらの情報をタグとして映像に付与し, 映像提示に利用する.

表 1: 動作の分類と対応する発話・動き

動作の種類	発話の種類	併用する動き
例示動作	このように こうやって, 等	手が体から離れている
物体指示動作	これ系統 この + 名詞	指示・提示を行う動き
場所指示動作	ここ系統 この + 場所	指示を行う動き
調整子	ここで, これで	手が下に降りている

### 3.3 提示例

以上のシステム構成で実際の手元作業プレゼンテーションの撮影を行った. 使用したカメラはそれぞれ「話し手全身」「右手のアップ」「両手間のアップ」の3台で, 図

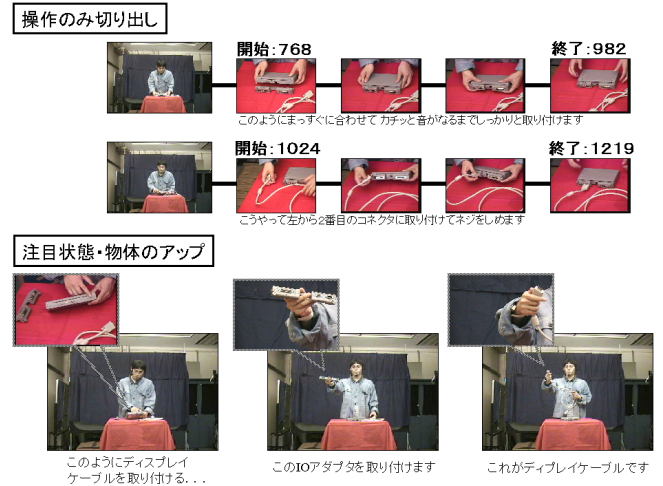


図 3: 動作認識を利用した提示の一例

のノートパソコンにディスプレイケーブルを取りつけるシーンを撮影した.

提示の一例として, 話し手の動作部分を抜き出して提示したものを図 3 に示す. 上の図は, 話し手の全身の画像とその横に対応するアップショットを動作の開始終了時刻と共に並べたものであり, 下の図は全身の映像から動作が認識された間だけ, 手元や手元作業空間のアップショットを吹き出しのように提示したものである. 図 2 は, 動作認識を利用した注目シーンの自動切り替え提示である. このような種々の提示とその自動化について研究を進めている [1][2].

## 4 まとめ

マルチメディアコンテンツ作成の為の知的撮影システムの枠組みを提案し, 簡単な実験によりその有効性を示した. 今後は, 特定物の位置やシナリオといった様々なデータを使用した映像のタグ付けとそれを有効に活用した提示について研究を進めていく予定である.

## 参考文献

- [1] 尾関基行, 中村裕一, 大田友一. プレゼンテーションの知的撮影システム 手元作業を対象とした適応的カメラワーク. PRMU2000-104, pp. 31-38, 2000.
- [2] 尾関基行, 中村裕一, 大田友一. プレゼンテーションの知的撮影システム 動作認識による映像のタグ付け. 第 6 回知能情報メディアシンポジウム, pp. 69-74, 2000.

\*ただし, 時点・区切は時間的幅を持たないため, 終了という概念は用いない.