

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

尾関 基行† 中村 裕一 †‡ 大田 友一 †

† 筑波大学 機能工学系

〒 305-8573 つくば市 天王台 1-1-1

ozeki@image.esys.tsukuba.ac.jp

‡ 科学技術振興事業団 さきがけ研究 21

あらまし: マルチメディアコンテンツ制作のための映像取得支援システムについて述べる。まず、撮影時に注目すべき対象を分類し、それぞれに適したカメラワークをまとめ、自動撮影を実現する上で問題点を明らかにする。次に、それらの問題を解決するカメラ制御アルゴリズムを提案する。さらに、話し手の発話情報と動き情報を用いて、映像にタグを付ける手法について簡単に提案する。本研究では、以上の枠組みを用いてプロトタイプシステムを構築した。これを用いてプレゼンテーションの撮影・提示実験を行い、実際のカメラマンによるカメラ操作との比較を行うことによって、その有効性を確認した。

キーワード: 知的撮影, カメラワーク, 動作理解, マルチメディアコンテンツ, 映像生成

An Intelligent Video Production System for Recording Presentations

Motoyuki OZEKI† Yuichi NAKAMURA†‡ Yuichi OHTA†

†Institute of Engineering Mechanics and Systems, University of Tsukuba

1-1-1 Tennodai, Tsukuba

ozeki@image.esys.tsukuba.ac.jp

‡PRESTO, Japan Science and Technology Corporation

Abstract: In this paper we introduce an intelligent system for video or multimedia contents production. First, we categorized targets and purpose of shooting, and discuss the cameraworks appropriate for them. Then, we propose camera control algorithms to realize the cameraworks. Moreover, we introduce video tagging by recognizing a speaker's behaviors, such as deictic movements. In the experiments on the prototype system, our algorithms showed good performance.

key words: intelligent video production, camerawork, human behavior understanding, multimedia contents, video editing

1 はじめに

近年のマルチメディア環境の進歩に伴い、映像製作会社だけでなく、一般企業や教育機関、さらには個人レベルでも、映像を使ったわかりやすいマルチメディアコンテンツの製作に関心が集まっている。しかし、効果的な映像を取得するためには、対象を適切なカメラワークによって撮影し、編集しなければならないため、非専門家や個人が手を出すことは難しい。また、カメラマンやディレクターを雇うことには多大なコストがかかるため、多くの場合には現実的とはいえない。そのため、マルチメディアコンテンツ製作を補助するような自動化システムが必要とされている。

この問題に対し、本研究では、手元作業の撮影に関する撮影目的とカメラワークの関係を考察し、基礎となるシステムの構築を行った。また、映像への情報付加についても検討し、これらを基にして、プレゼンテーションの話し手、受け手の興味を反映した映像提示を実現することを目指す。

我々は、このような考え方に基いて実際にシステムを構築し、簡単なプレゼンテーションに対して実時間で動作させる実験を行った。

以下本稿では、本研究の基本的な考え方、カメラワークの設定とカメラ制御アルゴリズム、実験例について述べる。

2 手元作業映像の取得システム

教材映像や説明映像など、情報を正確に伝達することを目的とする映像では、伝えるべき対象を適切に捉えたショットによって、効果的に内容を伝える必要がある。このような映像コンテンツを製作するためには、撮影時のカメラ設定や情報付加（タグ付け）、映像の提示について以下のようなことを考えなければならない。

カメラワーク： 注目対象、視点、解像度、カメラ制御方法

タグ付け： 映像が説明する内容、人物の動作、発話、人物の意図、その他の対象とその状態

提示方法： 提示部分の選択、映像（カメラ）切り替え

本研究では、それぞれの要素について自動化へのアプローチを検討し、簡単な実装を行った。本研究で構築した映像取得システムの概要を図1に示す。このシステムでは、位置センサにより話し手や特定物などの位置を取得し、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、MPEG エンコーダを通して保存される。

また、教材、資料、記録として映像を利用する場合には、一通りに編集された映像を残すだけでなく、受け手が見たいものを自由に取り出せるような枠組みが望まれる。つまり、様々な目的に対応できるよう、複数のカメラからの映像を記録し、各時刻のショットがそれぞれどのような

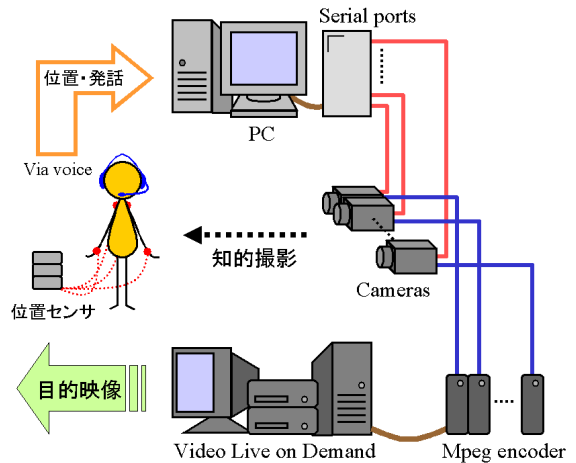


図 1: システムの概要図

情報提示に適しているかを逐次記述することが必要となる。そのために、本研究では、人物の発話と動きを利用して注目すべき部分を抽出し、映像にタグを付与する。さらに、得られたタグを基に、視聴者が見たいと思う部分を効果的に提示するようなシステムの構築を目指している。

関連する研究としては、講義映像のアーカイブ化 [2][7]、遠隔講義の伝送等 [6][9] の研究がある。遠隔講義映像取得に関する研究では、教師と黒板、教室の様子を教師と生徒の動きや位置、音声などを利用して自動的に切り替えるシステムが提案されている。

本研究では手元作業を対象とし、そのために必要となるカメラ設定および提示方法について検討した。手元作業を対象とする場合、上記のような研究と比較して、以下のような点について重点的に考慮しなければならない点が異なる。

- 手元作業の撮影では、一部分のクローズアップ映像が必要とされる。しかし、手や物体の動きが視野に対してかなり大きく速く動くために、カメラ制御が難しい。例えば、手元のアップを撮影している際に、2,3 フレームの間に手が元の視野から出てしまうことが頻繁に起こる。
- 手作業などの動作をあらかじめシナリオで細かく記述することは難しく、予測できない動作が現れることが多い。そのため、反射的に対象を追跡するカメラワークが必要となる。
- 注意を向けるべき対象として、話し手の顔、身振り、手先、物体その他、様々なものが考えられる。そのため、各映像（ショット）の持つ情報を詳細に記述し、それを基に映像提示を行う必要がある。

このように、カメラワーク（予測しにくい対象の追跡・撮影）と撮影シーンの認識（注目対象の認識）に関して、多くの新しい問題を含んでいる。これに対する、本研究のA

アプローチを以下で説明する。

3 手元作業映像取得のためのカメラワーク

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というわけではない。例えば、右手に持った注目物を撮影する場合でも、その注目物自体を撮りたいのか、それに関する操作を撮りたいのかでカメラワークは違って来る。注目物の細かい変化を撮影するためには、常に注目物を画面中央に大きく捉えておくことが望ましいが、操作を撮影するためには、できるだけカメラを固定し、視野が頻繁に動かないほうが良い。このように、各ショットの目的に応じたカメラ制御が必要である。

以下では、カメラワークに必要とされる要素について詳しく考察する。

3.1 注目対象

本研究では、注目対象を“何”の“どういう状態”を捉えるか、という観点から考察・分類を行う。ここで、“何”により撮影対象とする物体(対象物)を決定し、“どういう状態”かによりカメラ制御方法を決定する。このように、注目対象を“対象物”と“対象とする状態”に分けて考えることにより、注目対象とカメラ設定の関係を簡潔に説明することができる。以下、それぞれの分類について述べる。

撮影対象とする物体

手元作業において、話し手が視聴者に対して注目を促すものとして、操作や物体、場所などが挙げられる。また、視聴者にとって、話し手自身の挙動も重要な要素である。そこで本研究では、対象物として以下の4つを考えた。

<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

さらに、同じ<話し手>が対象物でも、話している顔、両手のジェスチャー、話し手の全身像等の自由度がある。同様なことが、4つの対象物についていえる。現在の段階では、4種類の対象物について、それぞれ3種類の撮影範囲(大・中・小)を加えた計12種類を用意している。

撮影対象とする状態

対象物の撮影で最も単純な方法は、対象物上の一点が画面の中央にくるようにカメラを制御することである。しかし、細かい動きやブレまで逐一追跡していると、視聴者に不快感を与えてしまう。本研究では、対象物のどのような状態を撮影するか、つまり、対象のどのような状態に注目するかについて、以下のように分類する。

<状況> : シーン中での位置関係や軌跡

<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)

要求されるカメラワークは次のようになる。<状況>では、視野をできる限り固定し、対象物がシーン内でどのように振る舞うのかを把握させる必要がある。<操作>では、対象物が細かく動くような状態を素早く見つけ出し、安定した視野で捉える必要がある。<物体>では、対象物をできるだけ画面中央に置きながらスムーズに追跡し、静止した瞬間には固定した視野で捉える必要がある。

3.2 カメラ制御アルゴリズム

本研究では上記の問題に対し、カルマンフィルタによる平滑化によって、対象物のスムーズな追跡を行う。また、撮影対象とする状態に適した撮影を行うために、画面内に仮想的な枠を想定して、カメラの動作を調節する枠制御アルゴリズムを提案する。各処理で用いるパラメータを調節することによって、種々の目的に合うようにカメラワークを設定する。以下、それぞれのアルゴリズムについて詳しく説明する。

カルマンフィルタによる平滑化

追跡する対象物の細かい動きやブレをノイズと見なし、カルマンフィルタでこれを除去することによって、スムーズに追跡することを考える [1][8]。系のダイナミクスとして、対象物の動きを剛体の運動モデルで近似し、システムノイズとして加速度の入力をとる。

状態変数 x_k とシステム行列 F を下に示す。

$$x_k = \begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \end{pmatrix} \quad F = \begin{pmatrix} 1 & \Delta & \frac{1}{2}\Delta^2 \\ 0 & 1 & \Delta \\ 0 & 0 & 1 \end{pmatrix}$$

ここで、 Δ は計測のサンプリング間隔である。 x_k は状態ベクトルであり、位置・速度・加速度の要素を持つ。カルマンフィルタは、以下の式で構成される。

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - H\hat{x}_{k|k-1}) \quad (1)$$

$$\hat{x}_{k+1|k} = F\hat{x}_{k|k} \quad (2)$$

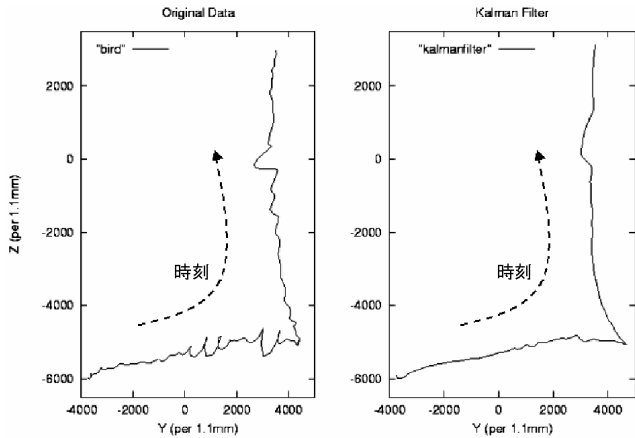
$$K_k = P_{k|k-1}H^T(I + HP_{k|k-1}H^T)^{-1} \quad (3)$$

$$P_{k|k} = P_{k|k-1} - K_kHP_{k|k-1} \quad (4)$$

$$P_{k+1|k} = FP_{k|k}F^T + \frac{\sigma_u^2}{\sigma_w^2}\Lambda \quad (5)$$

$$H = (1, 0, 0), \quad \Lambda = \text{diag}\{0, 0, 1\}$$

\hat{x}_k は対象の推定状態であり、 σ_u^2 はシステムノイズ、 σ_w^2 は観測ノイズ、 H は観測行列である。観測値 y_k が計測される毎に、式(2)~(5)の計算によって、中間推定値 $\hat{x}_{k|k-1}$ とカルマンゲイン K_k を更新し、式(1)によって推定値 $\hat{x}_{k|k}$ を計算する。 $P_{k|k-1}$ は現在の観測値が得られた時点での誤差の共分散値であり、これが最小となるように推定値が



(a) 対象物位置の軌跡 (b) 追跡目標位置の軌跡
(フィルタリングにより平滑化)

図 2: 対象物を追跡した結果

(横、縦にゆっくりと動かした手を追跡したもの)

計算される。

カルマンフィルタを上式で表すとき、その性質は σ_u^2 と σ_w^2 の比で決まる。このノイズ共分散比 σ_u^2/σ_w^2 を追跡のスムーズさを表すカメラ制御パラメータとして定義する。ノイズ共分散比が小さいほどスムーズに追跡し、大きいほど対象物を忠実に追跡する。

実際に対象を追跡したときの軌跡を図 2 に示す。細かい動きが平滑化されることによって、スムーズに対象物を追跡していることがわかる。

枠制御アルゴリズム

画面上に仮想的な枠を想定し、対象物がその枠内にある間カメラを固定する。仮想的に配置する枠は、視野に対する割合(枠サイズ)で指定する。枠の中心は対象物の軌跡の重心とし、一定時間(枠固定時間)毎に更新する。枠から対象物が出たら枠制御モードを終了して追跡モードに入り、停留もしくは反復を検出することで再び枠制御モードに切り替える。

停留と反復については、以下のように検出する。

停留: 一定範囲内に一定時間留まっていることにより検出

反復: 追跡する対象物の速度の符号の変化を調べることによって検出

以上のアルゴリズムの流れを図 3 に示す。また、このアルゴリズムで使用するパラメータを下にまとめる。

枠サイズ: 視野に対する枠の大きさ

枠固定時間: 枠制御中の微調整の間隔

停留検出時間: 停留とみなすまでの時間

停留検出閾値: 停留とみなす対象の移動範囲

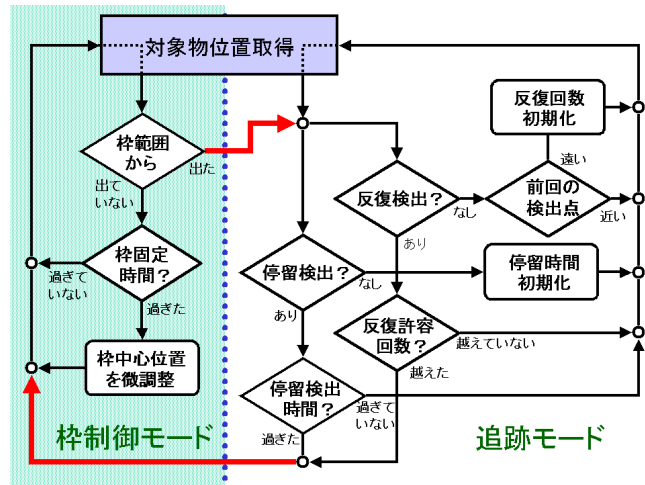


図 3: 枠制御の流れ

反復許容回数: 枠制御切り替えまでの反復の許容回数

ここで、枠サイズを大きくとるほど、対象物が大きく動いてもカメラは固定されやすくなるが、追従性が悪くなる。枠固定時間を短くするほど、対象を画面中央に置くことができるが、視野が頻繁に変化する。停留検出時間、停留検出閾値が小さいほど、少しの停留で枠制御モードに入りカメラが固定される。反復許容回数を小さくするほど、反復を含む操作に対して、素早く視野を固定する。

実際に箱を開ける作業を撮影した結果を図 4 に示す。枠制御を使った場合 (a) に、枠制御を使わない場合 (b) に比べて、安定した視野でシーンを捉えていることがわかる。

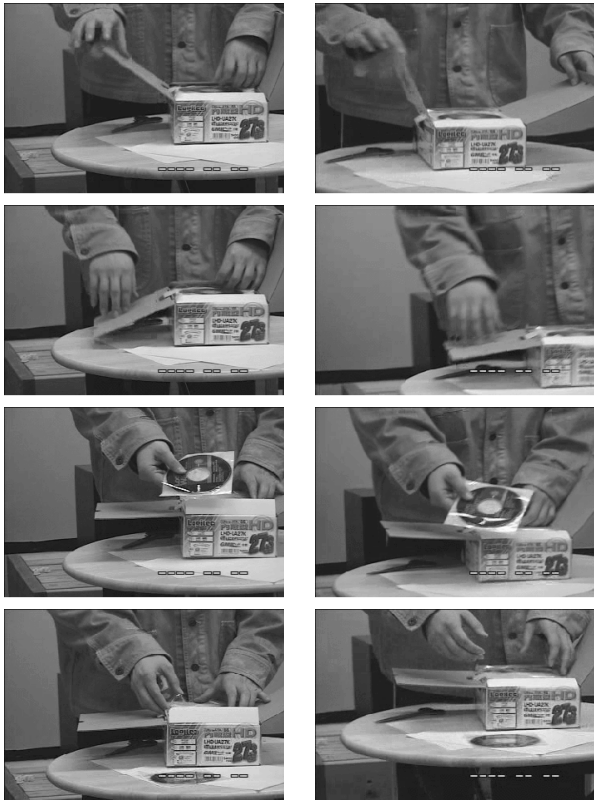
3.3 カメラ制御パラメータの設定

これまで述べてきたカメラ制御パラメータと、3.1 で述べた“撮影対象とする状態”との関係を述べる。

< 状況 > の場合には、画面内での対象物の振る舞いを把握するため、枠サイズを大きく、枠固定時間を長くとる。これにより、枠制御モードでは対象物の追従性が悪くなるため、停留検出時間、停留検出閾値、反復許容回数をそれぞれ大きく設定し、必要のない場面で枠制御モードに入らないようにする。また、激しく動く対象ではないため、カルマンフィルタの共分散比を小さくとる。

< 操作 > の場合には、対象物が細かく動く状態を素早く見つける必要があるため、反復許容回数を少なく設定する。対象物をある程度中央に捉えることが望ましいため、枠サイズ・枠固定時間はやや小さく設定する。

< 物体 > は、話し手が提示や指差しによって注目を促した物を撮影するための設定であり、できるだけ対象を画面中央に捉えるように、枠サイズ、枠固定時間を小さく設定している。提示や指差しによって対象物が停止した瞬間を捉えるために、停留検出時間および停留検出閾値を小さく設定し、カルマンフィルタの共分散比を大きくとる。



(a) 枠制御あり (b) 枠制御なし

図 4: 箱を開ける作業: 右手を対象物として追跡



図 6: カメラ設定の選択表 (一部分)
(左から, 対象ラベル, サンプル画像, 参考視点)

を選択することで, 各カメラの解像度とカメラ制御パラメータが自動的に設定される。

表 1 にそれぞれの対象ラベルに対応するカメラ制御方法と, 実際にカメラが追跡する基準点を示した。本研究では, 両手で行う作業の範囲自体を追跡する場合に, 実際にカメラが追跡する点として, 例えば両手に取り付けたセンサから両手の中点を計算し, それを追跡の基準点とする。注目物体, 注目場所に対しては, 直接物体上の点を追跡することが考えられる。しかし, 現在のシステムでは物体の一つ一つを常に追跡することが難しいため, 話し手が陽に注目を促す物体に対象を限る。そのため, 手先を追跡することにより, 指示物体や提示物体を撮影することを目的とする。¹

4 話し手の動作を用いた映像のタグ付

効果的に映像を提示するためには, 撮影している映像の各部分に付加情報を与えることが必要となる。付加情報には, 各映像部分が持つ情報, その重要性, また, ある情報を提示するための適切さ等に関するものがある。

手元作業映像に対しては, 具体的に以下のようなものがあげられる。

- カメラの役割 (注目対象, 解像度, 追跡方法等)
- 各時点でのカメラの状態 (パン・チルト角, 追跡対象の位置, その他)
- 各時点での話し手の言動, 意図
- 各時点で映っている対象とその状態

カメラの役割については, 各カメラに関連づけられた注目対象の情報を付加する。カメラの状態については, 各カメラのパン・チルト角などに加え, 追跡状況や枠制御アル

	<状況>	<操作>	<物体>
共分散比	スムーズに追跡 小		忠実に追跡 大
枠サイズ	できるだけカメラを固定 大		停止時のみ固定 小
枠固定時間	頻繁に動かない 大		中央に捉える 小
反復許容回数	ゆっくりと 大	反復ですぐ枠制御 小	なし
停留検出時間	大		停留ですぐ固定 小
停留検出閾値	大		停留ですぐ固定 小

図 5: 注目対象とカメラ制御パラメータの対応図

まとめとして, 各対象状態とカメラ制御パラメータとの関係を図 5 に示す。

3.4 カメラ設定の簡易化

上記のパラメータの値を直接決めることは, 一般のユーザにとって簡単ではない。そのため本研究では, 図 6 に示すように, プレゼンテーション映像で必要となる代表的な撮影対象を挙げ, 各々に適したカメラ制御パラメータセットを予め用意した。一覧表より目的に応じた対象ラベル

¹実際に, 特定の物体を見る, 見せる必要がある場面の多くはこのような場合であり, ここで挙げる撮影方法は, プレゼンテーションの映像取得に不可欠である。

表 1: 注目対象とカメラ設定の対応表

ラベル (撮影対象・視野角)	カメラ制御	追跡基準点
話し手・大	<状況>	腰の一点
話し手・中	<状況>	腰の一点
話し手・小	<物体>	頭の一点
作業空間・大	<操作>	両手の中点
作業空間・中	<操作>	両手の中点
作業空間・小	<操作>	右(左)手の先端
注目物体・大	<物体>	腰の一点
注目物体・中	<物体>	両手の中点
注目物体・小	<物体>	右(左)手の先端
注目場所・大	<状況>	腰の一点
注目場所・中	<物体>	両手の中点
注目場所・小	<物体>	右(左)手の先端

ゴリズムの状態を付加情報として与える。映像に映っている対象とその状態は、対象の画面上での位置、シーン内での3次元位置、状態(向き)、その他多くの項目が考えられる²。話し手の言動・意図については、指示動作、例示動作等を動き、発話から認識し、これを用いる。紙面の都合上、詳細は省略するが、これについては文献[10]等で今後報告する予定である。

5 実験

本システムでは、対象物の位置を取得するために6自由度磁気センサ Flock of Birds(Ascension Technology Corporation)を使用する。撮影用の首振りカメラとして、EVI-D30(Sony)を用いた。磁気センサの計測レートは30Hzであり、この計測値をPC(Linux)で受け取り、シリアル通信(RS232C)で首振りカメラを制御する。また、音声認識にはViaVoice(IBM)を用いた。音声認識結果と磁気センサによって得られた話し手の動き情報を用いて、指示・例示動作を抽出する。

カメラ制御アルゴリズムの評価

本システムを用いた場合の画面内での注目対象の位置およびその変化速度を調べ、文献[3][4][5]の研究で調べられている、カメラマンによるカメラ操作との比較を行った。

カメラ制御の特徴が顕著に現れるのは、撮影対象の動きが比較的速い場合である。そのため、注目物体や作業空間のアップを撮影する解像度で手先を追跡した。

図7に、カルマンフィルタによる平滑化を行った場合について、注目対象追跡時の画面内での位置分布と位置変

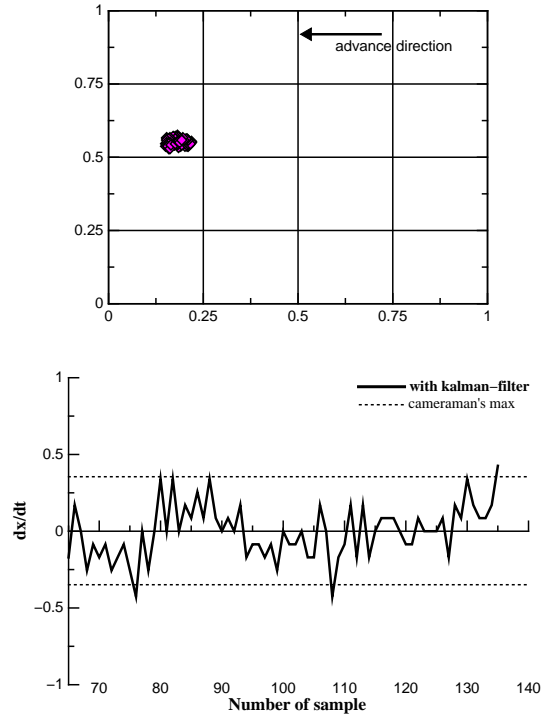


図 7: カルマンフィルタによる平滑化の場合の位置分布と変化速度(上段に正規化を行った位置分布, 下段にその水平方向の変化速度)

化速度を正規化したものを示す。ここで、位置変化速度とは、各フレームごとの注目対象の位置変化をフレーム時間(1/30s)で割ったものであり、1秒間にどれだけ注目対象の位置が変わったかを示す尺度となる。注目対象(右手)の速度はパンニング速度に換算して約4.7deg/sである。文献[5]によると、ほぼ同じ速度での実験で、一般的なカメラマンの操作では水平方向で0.28、垂直方向で0.19程度の広がりを持つことが報告されている。これに対し、本研究による手法(カルマンフィルタによる平滑化)では、水平方向で0.09、垂直方向で0.07~0.11の広がりであり、本手法でも十分に安定して追跡できているといえる。

しかし、カメラマンが注目対象の進行方向に対し、前方側に比較的大きなスペースをとるのに対し、本手法では前方のスペースが小さくなるのがわかる。これはカメラの制御の遅れによるものであり、遅れが少なくなるに従って画面中央(0.5,0.5)に近い位置で捉えられることになる。この問題に対しては、予測制御を用いることを検討中である。

位置変化速度については、一般的なカメラマンでは水平方向に最大で0.35程度であるのに対し、やや大きめの値(最大0.40程度)をとっているが、大きく逸脱することはない。

次に、枠制御アルゴリズムを使用した場合としなかった場合(カルマンフィルタのみ)について、図4の箱を開ける操作を対象に実験を行った。その結果を図8に示す。

²ただし、これについてはまだ実装しておらず、今後の課題となっている。

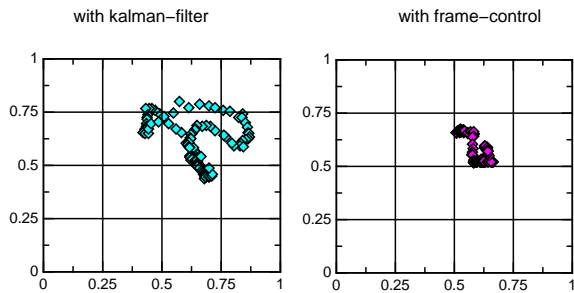


図 8: 枠制御を行わない場合と行った場合の位置分布と変化速度: 上段にそれぞれの位置分布, 下段にその水平方向の変化速度)

表 2: 撮影例で使用したカメラ設定

カメラ	対象ラベル	制御方法	追跡基準点
カメラ 1	話し手・中	< 状況 >	腰の一点
カメラ 2	注目物体・小	< 物体 >	右手の先端
カメラ 3	作業空間・中	< 操作 >	両手の中点

この図では, 図 7 とは異なり, シーン中の静止点の視野内での位置変化をプロットした. 位置分布の広がりから, 枠制御を使用した場合の方が, 安定した視野でシーンを捉えていることがわかる. また, 枠制御を使用しない場合は, 水平方向の位置変化速度が, カメラマンの最大値である 0.35 を大きく逸脱している. これに対し, 枠制御を行った場合には, 最大でも 0.40 程度となり, 見やすい映像が取得できている.

以上のように, 画面内での対象物の広がり, 位置変化速度について, 本研究でのカメラ制御アルゴリズムが概ね満足できるものであることが確認できた.

プレゼンテーションの撮影例

実際の手元作業プレゼンテーションの撮影を行った. 使用したカメラは図 6 の 3 台であり, それぞれのカメラ設定を表 2 に示す. プレゼンテーションは, ノートパソコンにディスプレイケーブルを取りつける説明とした. 各カメラの映像を図 9 に示す.

Count	動作[指示詞内容]	注目対象	フレーム: 開始 - 終了
0th	操作・状態[このように]	作業空間	225 - 345
1th	注目・名付[この]	注目物体	497 - 706
2th	操作・状態[このよう]	作業空間	858 - 1037
3th	注目・名付[これ]	注目物体	1217 - 1308
4th	操作・状態[こうやって]	作業空間	1488 - 1668
⋮	⋮	⋮	⋮

図 10: タグ付けの例

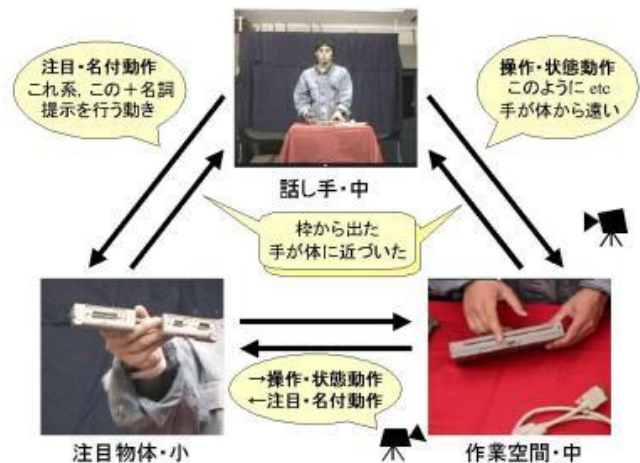


図 11: 切り替え条件

「話し手・中」のカメラでは, 話し手の全体的な状況や動きが画面内に収まる解像度に設定されており, また, 話し手が多少動いてもカメラを固定する枠制御によって, 話し手の状況が把握しやすい映像となっている.

「注目物体・小」のカメラでは, 提示したものがはっきりと見える解像度で, 対象を常に追跡して提示の瞬間に固定することにより, 必要なショットをタイミング良く取得できていることがわかる.

「作業空間・中」のカメラでは, カメラ視点を上にして, 正面にある「話し手・中」のカメラからは見づらい操作を分かりやすく捉えている. また, 作業が丁度入程度の解像度であり, カメラを固定してわかりやすく操作を捉えることができている.

映像提示例

この実験例では, 話し手の動作認識により, 図 10 に示すようなタグを映像に順次付加している. これを利用して, 図 11 の方法で単純に映像を切り替えた例を図 12 に示す. 注目物体用カメラは「注目・名付」動作³に, 作業状態用カメラは「操作・状態」動作⁴にそれぞれ対応して

³本稿では詳しく説明していないが, 「この」や「これが」などの指示詞と提示を行う動きが同時に行われた場合に, 注目を促すような動作が行われたことを表す.

⁴「このように」や「こうやって」などの指示詞と手が体から離れている状態が同時に起こった場合に操作や対象の状態を示す動作が行



図 9: 各カメラの映像および各種情報



図 12: 切り替えの効果

おり、各動作をトリガとして「話し手・中」のカメラからそれぞれのカメラに切り替わる。この切り替え処理はスイッチャーを使用して自動で行っており、動作が行われてから 2 秒程度の遅れで映像が切り替わる。図 9 の最上段の映像に比べ、提示物体や作業の様子がわかりやすくなっていることがわかる。

なお、ここに示したカメラ切り替えは提示方法の一例であり、タグ付された映像を利用して、様々な提示（例えば、ある動作だけ抜き出すなど）が可能であると考えられる。今後、種々の効果的な利用方法を探っていく予定である。

6 まとめ

映像コンテンツ、マルチメディアコンテンツ製作補助のために、複数カメラを使用した知的撮影システムの枠組みを提案した。そのために、カメラワークを重点的に検討し、注目対象に応じたカメラ制御を行うアルゴリズムを提案した。現在は、まだ必要とされる機能の一部が実装された段階であるが、いくつかの実験によって提案手法の有効性を確認した。

今後の課題として、まず、未実装の部分を完成させることがあげられる。それに加え、注目対象の分類がまだ不十分であり、この点についての十分な検討が必要である。また、対象の大きさやスピードによってズーム値を変化させる方法などを検討する必要もある。これらの研究とともに、発話内容や特定物の位置など、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を行って

われたことを表す。

いく予定である。

参考文献

- [1] 有本卓. カルマン・フィルタ. 産業図書, 1988.
- [2] L. He, et al. Auto-summarization of audio-video presentations. *Proc.ACM Multimedia*, pp. 489–498, 1999.
- [3] 加藤大一郎, 山田光穂ほか. スタジオ番組における放送カメラマンのカメラワークと視線の動きの分析. *テレビジョン学会誌*, Vol. 49, No. 8, pp. 1023–1031, 1995.
- [4] 加藤大一郎, 山田光穂ほか. 被写体を追尾撮影時の放送カメラマンのカメラワーク分析. *テレビジョン学会誌*, Vol. 50, No. 12, pp. 1941–1948, 1996.
- [5] 加藤大一郎, 石川秋男ほか. カメラワーク分析と映像の主観評価実験. *映像情報メディア学会誌*, Vol. 53, No. 9, pp. 1315–1324, sep 1999.
- [6] 宮崎英明, 亀田能成. 複数カメラを用いた講義映像の実時間作成法. *MIRU'98*, pp. 123–128, 1998.
- [7] S. Mukhopadhyay and B. Smith". Passive capture and structuring of lectures. *Proc.ACM Multimedia*, pp. 477–487, 1999.
- [8] 西山清. パソコンで解くカルマンフィルタ. 丸善, 1993.
- [9] 大野直樹, 池田克夫ほか. 遠隔講義における講義状況に応じた送信映像選択. 第 5 回知能情報メディアシンポジウム, pp. 31–38, 1999.
- [10] 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.