人間の行動を知的に撮影し伝えるシステム

中村 裕一

■ 研究のねらい

コンピュータや通信技術の進歩により，映像 (動画) や映像を基にしたマルチメディアコンテンツがコミュニケーションの手段として重要な位置を占めるようになってきた．その結果，一般企業や教育機関，さらには個人のレベルでも，手軽に映像を製作し，コミュニケーションの手段とすることに関心が高まっている．しかし，映像の撮影は，世界で起こっている出来事の一部分 (時間，空間的な一部分) を知的に切り出し，編集する行為であり，真面目に取り組むとかなり難しい問題でもある．単純に撮り流したホームビデオが，他人にとって見るに耐えない代物となることからも，それがよくわかる．また，テレビ放送や映画のようなリニア[1]な映像は，受け手にとって親切な情報提示形態となっていない．長時間の映像が蓄積されていて，それを利用することが可能でも，必要な情報を探すのに長時間かかれば，誰も使いたいとは思わないだろう．

このように，映像を誰でも手軽に使えるコミュニケーション手段とするためには，映像撮影の問題を見直し，それをサポートするシステムを用意することが必要である．また，TV や映画のような形態にとらわれずに，受け手にわかりやすく，短時間で利用できる形で提供する方法を探ることが必要である．

この研究ではその一つのアプローチとして，料理や組み立て等の解説 (プレゼンテーション) 場面を題材とし，**カメラマンの機能** (人間の行動を知的に撮影する)，**ディレクターの機能** (人間の行動を認識して映像を知的に編集する)，**マルチメディア・デザイナーの機能** (映像が含んでいる情報をわかりやすく提示する) の 3 つの観点から，映像によるコミュニケーションをサポートする方法論を考え，それを検証するためのプロトタイプシステムを構築することを目的とした．

■ 研究成果

上にあげた 3 つの機能を実現するために，各々次のような手法を提案し，その検証実験を行ってきた．

カメラマンの機能 (人間の行動を知的に撮影する)

顔や手先など，撮影の主対象となる部分を複数のカメラで常に追跡して，いつでも映像として利用できる状態にする自動化撮影システムを構築した．**何をどのように伝えるかという目的**と**カメラの自動制御アルゴリズムやパラメータ**との関係を探り，不快感がなくわかりやすい映像を取得する方法を提案した．

ディレクターの機能 (人間の行動を認識して映像を知的に編集する)

人間の行動 (ここではプレゼンテーションを対象) において，重要な意味を持ち，注目する必要がある場面や部分を検出する手法を提案した．注目すべき部分は，時間的・空間的に常に変化するため，**人間の行動 (体の動きや発話等) を利用して，もっとも見せたい部分を検出する**が重要なポイントである．また，シナリオ等のあらかじめ用意された情報を用いることも有効であり，シナリオの記述と人間の言動を照合する手法も提案した．

マルチメディア・デザイナーの機能 (映像が含んでいる情報をわかりやすく提示する)

プレゼンテーションのような目的がある場合，各時点の人間の行動やその他の被写体には意味や因果関係がある．それらをうまく記述して，映像をわかりやすく提示するような映像提示手法を提案した．

[1] 時間軸方向に連続したデータをそのまま時間軸に沿ってしかアクセスできない状態を言う．見たい場所に到達するために，早送りや巻き戻し等の操作が必要となる

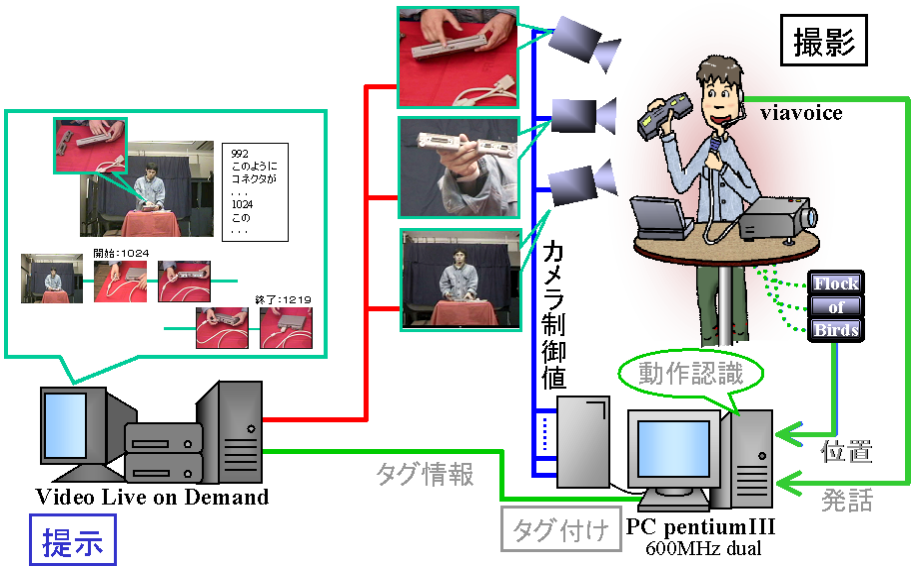

図 1: システムの概要図

図 2: 自動映像切り替えの結果 (一部)

映像中のフレーム (画像) をアイコンとし，それを概念図のように並べることで映像の構造を表現する，**ビデオアイコンダイアグラム**の手法を提案した．

このような考え方に基づいて，図 1 のようなシステムを構築した．このシステムでは，位置センサや画像処理により話し手や特定物などの位置を取得し，複数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，MPEG エンコーダを通して保存され，ランダムアクセスが可能になる．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．実際に撮影されて編集された映像の例を図 2 にあげる．静止画ではわかりにくいが，カメラの切り替えを含め，かなり自然な映像が得られている．

上記の 3 つそれぞれについて，以下で詳しく説明する．

1　人間の行動を知的に撮影する

図 1 のように，**撮影対象の位置を常に計測でき，パン/チルトカメラをコンピュータから自由に制御できる**場合を対象とし，見たい/見せたい対象に合わせたカメラワークを自動化することを目指した．

1.1 何を撮るか，どのように撮るか

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，右手に持った物体を撮影する場合でも，その物自体を撮りたいのか，それを操作している様子を撮りたいのかによってカメラワークは違ってくる．注目物自体の詳細な様子を撮影するためには，それを常に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．

そのため，この研究ではまず，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・分類し，それを基にしてカメラワークを考える．つまり，“何” により撮影対象とする物体や空間（対象物）を決定し，“どういう状態” かによってカメラの制御方法を決定する．

撮影対象とする物体・空間

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，対象物として，話し手自身を狙う**＜話し手＞**，話し手の作業を狙う**＜作業空間＞**，注目すべき物体を狙う**＜注目物体＞**，注目すべき場所を狙う**＜注目場所＞**の 4 つを設定した．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，次の 3 つに分類した．

＜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）

各々に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いて追跡し，静止した瞬間には固定した視野で捉える必要がある．

1.2 カメラ制御の方法

目的に応じてカメラワークを変えることができるように，**カルマンフィルタによる平滑化**，**枠制御アルゴリズム**を用いて，そのパラメータを調節することにした．

カルマンフィルタによる平滑化では，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．本システムでは，対象物の動きを剛体の運動モデル（等加速運動）で近似し，加速度の変化をシステムノイズとして入力する．カルマンフィルタの性質 (平滑化の度合い) は**ノイズ共分散比**に大きく左右される．そこで，このノイズ共分散比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用いる．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．

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**を用いた．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．その他，対象物が停留していると判断するための，**停留検出閾値**，**停留検出時間**，**反復許容回数**等をパラメータとして持たせている．

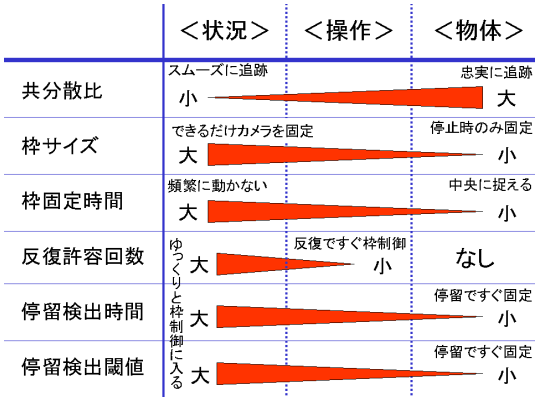

図 3: 注目対象とカメラ制御パラメータの対応図

図 4: カメラ設定の選択表 (一部分)
(左から，対象，サンプル画像，視点)

1.3 カメラワークを設定する

“撮影対象とする状態” とカメラ制御のためのパラメータとの関係を考えて，目的にあったカメラワークを設定する．例えば，撮影対象とする状態が**＜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．対象物を中央付近に捉えることが望ましいため，枠サイズ，枠固定時間は**＜状況＞**に比べてやや小さく設定する．それに対して，**＜物体＞**の場合は，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．このような関係を表にすると，図 3 のようになる．

このようなパラメータの値を一般のユーザが決めることは簡単ではない．そのため，図 4 のように，プレゼンテーショ映像の撮影で必要となる代表的な撮影対象と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．

1.4 有効性はどうか

実際に撮影された映像の一部を図 5 に示す．このような静止画でははっきりとはわかりにくいが，定量的な評価実験を行うことによって，カメラの無駄な動きが抑制され，見易い画面になっていることがわかった．また，各制御方法**＜物体＞，＜操作＞，＜状況＞** と**単純追跡**[2]の 4 つを比較する主観評価実験を行った結果，目的にマッチしたカメラ制御方法を「見易い」と感じることがわかった．また，ある目的に適していると評価されているカメラ制御方法が他の目的に用いると評価が悪くなることから，対象に適したカメラワークを選択することが必要であるということが実証された．

2 人間の行動を認識して映像を知的に編集する

映像の “**どこにどのような情報が含まれているか**”，また，“ **各部分がどの程度重要か**”，がわかれば，映像を効果的に提示できる．この研究では，話し手の典型的な行動のいくつかを自動認識して，このような情報を取得する．それを映像への付加情報として記録し，カメラを切り替えることによる編集やマルチメディアコンテンツへの加工に利用する．

[2]対象の位置をそのまま狙う，何も工夫していないカメラ制御方法.

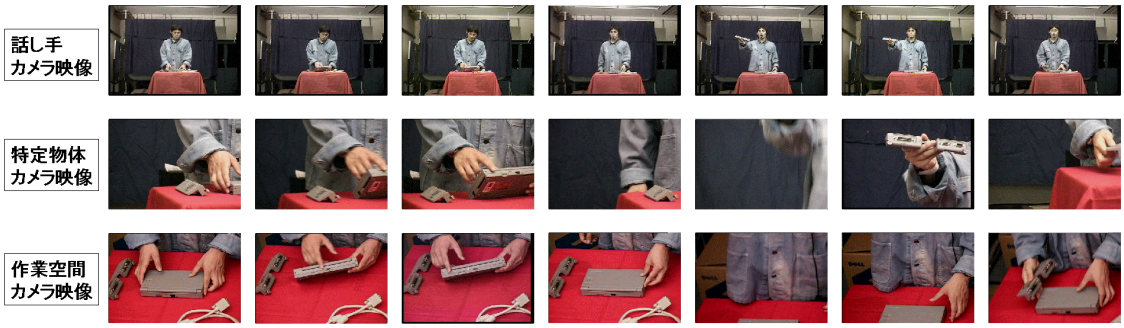

図 5: 各カメラからの映像

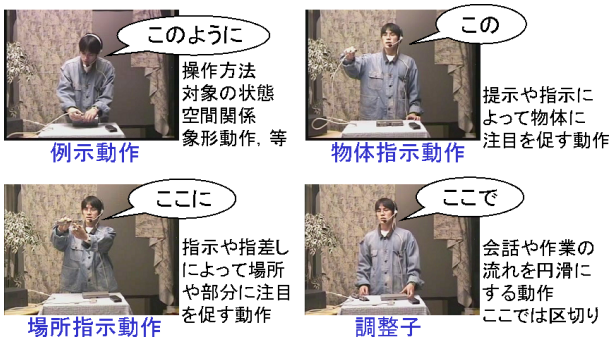

図 6: 注目を要求する動作

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

図 7: 動作の分類と対応する発話・動き

2.1 どこが重要か

話し手が見せたいと思う部分，また，受け手 (視聴者) が注目しなければならない部分として，**話し手が明確にの物体や自分の動作などに注目を要求する行動**が挙げられる．例えば,「この」と発声しながら物を提示するような動きを行っている場合，話し手は提示物体に注目を促している．このような動作には，図 6 に例を示したように，指示動作，例示動作，各種操作，注意喚起のための言動[3]等がある．以前に筆者らが構築した MMID(マルチモーダル・データベース)[4]を用いて調査したところ，発話の処理と体の動きの処理を併用することにより，かなり良い精度で指示・例示動作の抽出を行えることがわかっていた．

2.2 重要部分を検出する

図 7 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発話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．

実際には次のような処理になる．

- 発話中の指示詞を取得し，その時刻を記録する
- 手の伸びと速度の変化の極値情報を取得し，その時刻を記録する
- 指示詞とそれに対応する動き情報が，一定時間内 (3 秒など) に起こった場合に指示 (例示) 動作が起こったとする

[3]例えば,「はい，注目！」,「よく見てください」等.
[4]映像，音声，トランスクリプト，人物動作を統合してプレゼンテーション時の行動を記録し，蓄積してきた

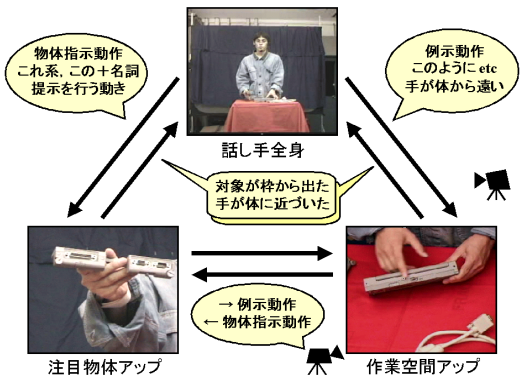

図 8: 使用したカメラと編集条件

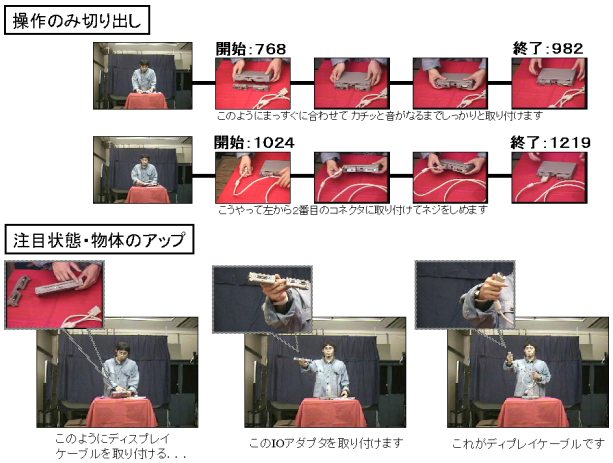

図 9: 動作認識を利用した提示の一例

図 1 のシステムを用いて動作認識実験を行った結果，システムの動作を理解している人[5]が使えば，70〜80%の認識率が得られることがわかった．認識対象としている動作も，10〜20 秒に一回程度起こり，映像を編集するために丁度よいトリガとして使えることもわかった．しかし，システムの動作を全く知らない人が使った場合には，ここで対象としている動作の出現率が低く，有効な結果とはならなかった．つまり，これからもっと幅広い動作を扱っていかなければならないことを示しており，現在はこの問題に取り組んでいる．

2.3 どのように使えるか (映像を実際に編集した例)

システム構成で実際に組み立て作業のプレゼンテーションの撮影を行った例をあげる．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，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．

既に図 2 に示したものは，カメラの自動切り替えによる編集結果である．このように，かなり自然な映像が得られることがわかっており，動作認識が精度良く行われた場合には，全自動で撮影されていることに気が付かない視聴者もいた．ただし，話し手が動作を行ってからそれが認識されるまでの遅れ時間があり，それが話し手の不安を誘う等，これから改善しなければならない点もある．

これ以外にも，図 9 のような提示にも利用可能である．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 9 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クローズ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クローズ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．

3 映像が含んでいる情報をわかりやすく提示する

映像は時間軸を持ったストリーム形のメディアであるため，冗長性が高く，一覧性が悪いことが問題となっている．例えば，料理番組を一回見ただけで手順を確実に覚えられる人は少ないだろう．そのため，録画して何度も見直したり，料理解説本を別途購入することになる．

このような問題に対して，映像の内部構造を図化する新しい手法，**ビデオアイコンダイアグラム** (Video Icon Diagram, 以下 VID と略記する) を提案した．VID は，空間的，時間的関係を始め，映像の種々の意味的構造を概念図として表現する手法である．これらの重要な関係にしたがって映像内容を整理して上手

[5]システムに関して，2，3 分の説明と 1，2 回のリハーサルを行えば，ほとんどの人がすぐに使いこなすことができた．

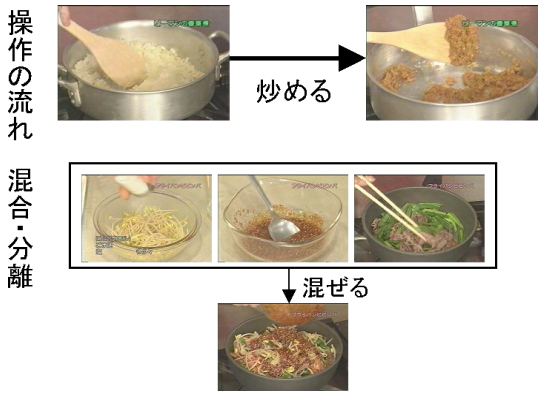

図 10: 簡単な VID の例

セリフ	動作
村山「かつおを一センチ厚に切ります」 村山「次に切身を盛り付けます」	村山：かつおを切り始める． 村山：皿に切身を盛り付けていく．

図 11: シナリオ例

に表現すれば，視聴者の理解を助け，欲しい情報を短時間で手にいれることを可能にする．

この研究では，上の 2 つの研究と異なり，シナリオやトランスクリプトがあらかじめ与えられていることを仮定する[6]．これらの情報と撮影時に行われた行動を照合することによって，映像の内部構造を認識し，図的表現を行う．

3.1 ビデオアイコンダイアグラム (VID) とは

一口に映像中の関係や構造と言っても，各々の関係を持つ実体は被写体，映像内での出来事，映像外の被参照物等，種々の場合がある．しかし，この研究ではこれら実体を細かく分類せず，全て**ビデオアイコン** (各映像断片の代表画像) で代用し，図 10 に例をあげるように，アイコン間の関係として図示する．これは，VID を見るのが人間であり，関係を持つ実体を比較的簡単に推測できるからである．さらに，このプロセスを確実にするために，VID では各アイコンに文字で注釈を添える機能を備える．

VID では，関係を表現するために関係の種類と属性の 2 つを用いる．関係の種類には**同値関係** (被写体間の同値関係や**セグメント** (映像断片のことを以後，セグメントと呼ぶ) の同時性など)，**順序関係** (セグメント間の時間的順序関係や，対象物体の変化などに代表される意味的順序関係，またイベント間の因果的順序関係など)，等の 3 つ種類を，関係の属性には，**時間** (時刻，あるいは時系列上の関係)，**空間** (実空間での物理的位置関係)，**因果** (原因・理由)，等の 6 つの種類を設定した．

3.2 シナリオから VID のためのスクリプトを用意する

VID の基となる記述はシナリオまたはトランスクリプトとし，これらに XML タグを付けて XML 文書とする．この研究で想定するシナリオ例を図 11 に示す．これを XML 記述 (**VID スクリプト**と呼ぶ) に変換する．具体的には，映像セグメントを記述するための XML 要素と，セグメント間の関係を記述するための XML 要素を用いて記述する．

図 11 のシナリオにタグを付けた例を図 12 に示す．まず`<definition>`タグで「かつお」というオブジェクトを定義し，それに $o1$ という ID と「食材」を意味する属性値を付加している．次に`<serif>`タグによりセリフを定義しており，`<sentence>`タグに ID 属性を設定することで，セリフ文に $s1$ という ID を振っている．また，`<word>`タグを用いることで単語「かつお」に A という ID を付けている．後半はシナリオ中の動作説明であり,「切る」という動作が`<action>`タグにより定義されている．`<action>`タグ中の `type` 属性は動作の種類を，また `time` 属性は対応する映像セグメントの時刻情報を示す．

現在のところ，シナリオから人手で変換してこのような記述を得ているが，自然言語処理の援用などが

[6] 上記 2 つの研究でも，シナリオやトランスクリプトが与えられていれば，より良い撮影や編集が可能であるが，それを必須とはしない．

```
<scenario>
<definition>
  <object id="o1" type="material">かつお</object>
</definition>
<serif>
  <sentence speaker="村山" id="s1" time="">
    まず、事前に焼き付けておいた
    <word ref="#o1" id="A">かつお</word>を一センチ厚に切ります
  </sentence>
</serif>
<event>
  <scene summary="かつおのたたきの調理">
    <action time="type:exist; begin:#s1" agent="村山" label="切る"
            type="cook:transform" id="a1">
      <input ref="#A" label="かつお"/> <output label="かつおの切身"/>
    </action>
  </scene>
 </event>
</scenario>
```

図 12: VID スクリプトの例

これからの興味深い研究課題となっている．

3.3 撮影時の情報からビデオアイコンダイアグラムを作る

実際の映像データに対し，映像中の発話や動作の情報を利用して VID スクリプトとの対応をとり，スクリプト中の動作が行われた時刻などを求める．

現在は，音声認識を利用して，発話とセリフと対応づけることが可能となっており，その対応関係から映像とシナリオを対応づける．まず，映像中の発話を音声認識によってトランスクリプトに変換する．このデータと，シナリオ中の文字列対応関係が求まった後，各々の発話時刻を用いることにより，シナリオのセリフの各部分と映像フレームとの対応が求まる．これにより，映像との対応が取れた VID スクリプトが得られる．

3.4 ビデオアイコンダイアグラムの実例

料理番組や実験番組の模擬撮影を行った例をあげる．撮影には図 1 のシステムを用い，複数台のカメラで人物，テーブル，手元等を自動撮影した．このような撮影環境で「カツオのたたき」を実際に調理し，得られた VID の例を図 13 中に示す．左側の一連の流れが，説明の大まかな順序を表す．また右側の部分は，それぞれの部分に対応するシナリオ中の動作を図示したものである．シナリオには食材や中間生成物，作業などが細かく記述されており，各々に対応するビデオアイコンが配置されている．

このように簡単な例については，VID を用いて映像の内容を空間的に表現できた．また，表示する関係の属性を考慮することで，目的に応じた VID を動的に生成できることが示された．しかし，付けられているタグの種類が限定されているなどの制約から，限られた目的・視点以外には対応できないといった問題があり，今後の研究を必要としている．

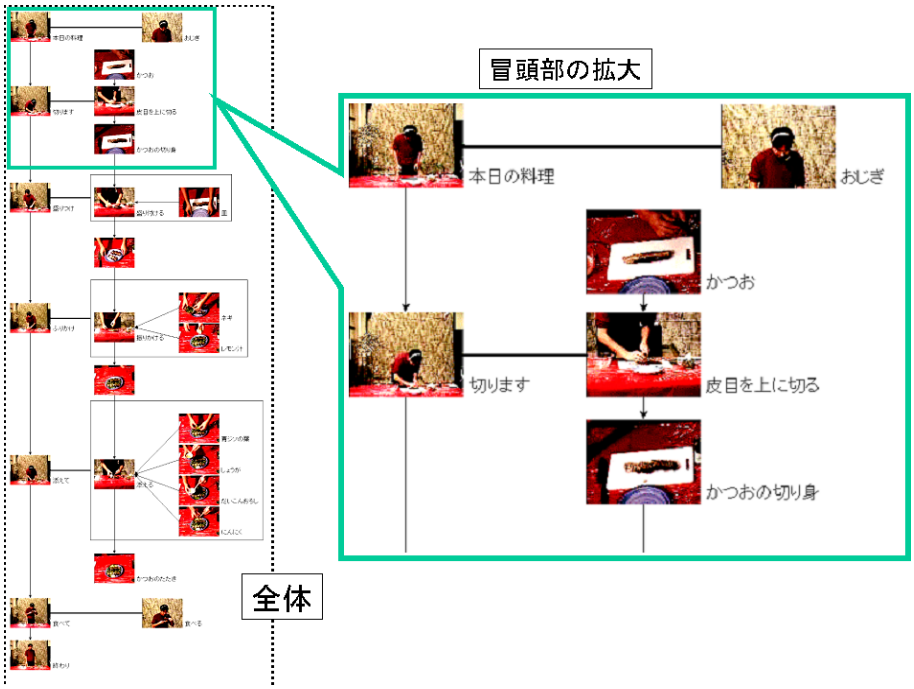

図 13: 料理映像に対するビデオアイコンダイアグラム

■ 今後の展開

「さきがけ 21」プログラムの支援を受け，3 年間最も力を注いだのは，図 1 のシステムを実際に動作させることであった．原理的には簡単な処理の組合わせではあるが，実際に総合システムとして構築し，実時間で動作させたのは世界で初めてである．心配された映像の「質」も，時には素人ビデオ以上の結果が得られるところまでいったことに満足している．一般的な問題を扱うためには，まだまだ不十分な点が多く，すぐに実用化ができるものではないが，映像コンテンツ取得を自動化するシステムの可能性を示した点で，十分な意義があった．

また，このように実証的なシステムを構築したことから，他の研究との協力が行えるようになった．科学研究費・特定領域研究「視覚情報メディアのためのパターン認識・理解」では，複合コミュニティ空間 (現実世界と仮想世界を合成した場で複数の人がコミュニケーションを行う空間を呼ぶ) をサポートするシステムとして利用されている．また，本年度から 5ヶ年計画の科学研究費・学術創成研究「人間同士の自然なコミュニケーションを支援する知能メディア技術」において，様々な用途に用いることのできる映像コンテンツを獲得するためのプラットホームとして活用される予定である．

今後は，扱う対象を少しづつ広げること，人間の自然なさりげないコミュニケーションを扱えるようにすること，マルチメディアコンテンツとしての利便性を高めること等，種々の点から高度なシステムへ向けてのアプローチをする予定である．

さらに，このような実際的な研究とともに，映像を情報としてどのように表現し，扱うかという点について考察していきたい．映像がコミュニケーション手段として日常に深く浸透しているにもかかわらず，その情報学的な表現や扱いにはまだ十分なアプローチがなされていないのが現実である．映像がこれからますます大量に流通・蓄積されていくことを考えると，映像の本質を「情報と知」の面から探ることが必須であろう．

最後に，暖かい助言を頂いた領域総括の安西先生，領域アドバイザーの先生方，様々なご援助を頂いた科学技術振興事業団の皆様に感謝の意を表します.

■ 成果リスト

- M.Ozeki, Y.Nakamura, Y.Ohta Camerawork for Intelligent Video Production — Capturing Desktop Manipulations, Proc. Int. Conf. on Multimedia and Expo, TA1.5, 2001
- M.Murayama, Y.Nakamura, Y. Ohta, Diagram Generation From Tagged Texts Toward Document Navigation, Proc. Int. Conf. on Multimedia and Expo, FP0.4, 2001
- 村山正司，伊津野英克，中村裕一，大田友一ビデオアイコンダイアグラムによる映像内容の構造表現, 信学技報 PRMU2001-45, pp.47–54, 2001
- 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 —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
- 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— , 信学技報 PRMU2000-104, pp.31-38, 2000
- Y. Nakamura, J. Ohde, Y. Ohta, Structuring Personal Activity Records based on Attention — Analyzing Videos from Head — mounted Camera, Proc. 15th International Conference on Pattern Recognition Track4, pp.220-223, 2000
- 大出純哉，中村裕一，大田友一，映像による個人行動記録・要約システムとその評価 —注目シーン検出と要約の評価— , MIRU2000 論文集, pp.I-499-504, 2000
- Y. Nakamura, J. Ohde, Y. Ohta, Structuring Personal Experiences — Analyzing Views from a Head — mounted Camera, Proc. International Conference on Multimedia and Exposition 2000,TP10-5, 2000
- 村山正司，中村裕一，大田友一概念図の自動生成による文書内容の可視化—タグ付き文書からの自動変換—, 第 5 回知能情報メディアシンポジウム, pp.117-124, 1999
- Y. Nakamura, Multimodal Approach toward Intelligent Video Production, International Workshop on Multimedia Intelligent Storage and Retrieval Management, pp.1-8, 1999
- 村山正司，中村裕一，大田友一，概念図の自動生成によるタグ付文書の可視化, 信学技報，思考と言語研究会, 1999
- 村山正司，中村裕一，大田友一，知識ナビゲーションのための概念図の自動生成, 情処研報 AI99-33, pp.29-36, 1999
- 大出純哉，中村裕一，大田友一，ビデオ映像による個人行動記録システムにおける注目シーンの検出, 第 5 回画像センシングシンポジウム，pp.179-184, 1999
- 中村裕一, 村山正司, 大田友一，図的メディアと言語メディアの統合による知識の解析と提示, 第 4 回知能情報メディアシンポジウム, pp.31-38, 1998