知識流通のための知的映像取得と提示

Intelligent Video-Based Media Production and Presenting for Knowledge Distribution

中村　裕一[1]

[1]筑波大学 機能工学系　(E-mail:yuichi@image.esys.tsukuba.ac.jp)

本論文では，知識流通を支援するための映像の自動取得と，それを用いた質問応答システムについて紹介する．映像メディアを幅広い分野で知識流通の手段として用いるためには，まだまだ解決すべき課題が多い．映像メディアの取得，その利用の両面から種々の自動処理技術を開発する必要がある．我々は，これらの問題に対し，(1)教示番組の映像撮影・編集，(2)対話的に映像内容を提示するためのデータ構成やその QA 手法について研究を行っており，本論文では，その具体的な内容について紹介する．(1)に関しては，自動撮影のためのマルチカメラ環境とそれを用いてインデックスが高度に付加された映像を得る技術，(2)に関しては，映像に対する利用者の多様な質問に対して応答する QA 手法について紹介する．

キーワード：知識流通，対話的映像メディア，知的映像撮影，知的映像編集，QA

1.　はじめに

本論文では，知識流通を支援するための映像の自動取得と，それを用いた質問応答システムについて紹介する．この研究は，対話型[i]のプロセスを用いて知識を媒介するための要素技術となるものであり，その研究スコープを広くまとめると図１のようになる．この図では，人間(ユーザ)がある要求[ii]を発したときに，それに必要な返答を与えるために，種々の蓄積型メディアや実時間型情報源と人間との柔軟なインタラクションを可能にするものである．

このような枠組みで表されるように，映像は知識を表現し，伝えるための強力な手段となっている．しかし，取得が難しい，一覧性が悪い，加工が難しいといった問題も同時に抱えている．このような問題を解決するために，我々は，映像メディアの取得，編集，提示手法を一貫して扱う研究を行ってきた．具体的には，(1) プレゼンテーション，作業，個人行動等を伝えるための映像撮影・編集と動作，発話情報の利用，(2) 対話的に映像内容を提示するためのデータ構成やその QA 手法(QUEVICO)，そのインデックスを自動的に取得する研究などである．画像処理，人物の動作認識，自然言語処理を用いて映像メディアの取得と利用を自動化することによって，映像データをより手軽に扱うことを目的としている．

我々の研究では，これにより，e-Learning 等の学習環境だけではなく，実際の作業現場やそのシミュレーション環境で，その場の状況に即した情報を与えることをもめざしている．種々のアドバイス，こつ，注意点等を映像として記録しておき，これを現場でユーザと対話しながら提供することができれば，これまで先生や講師が行っていた訓練を個人的に，また随時受ける環境を構築することができる．

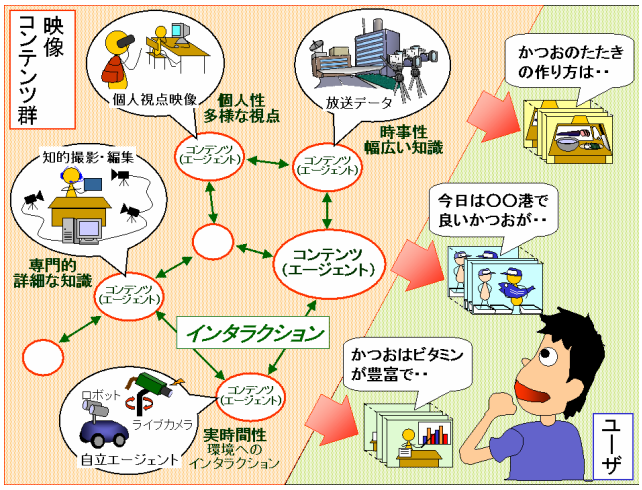

図 1: 知識流通と映像・マルチメディア (ユーザが「かつおのたたきが食べたいなぁ」と言った場面を想定)

現在のところ，まだ図１全体としての機能はできあがっていないが，その核となる要素技術の実現は着実に進んでいる．以下，本稿では 2 章で映像メディア処理の概要と問題点を，3 章で映像の自動撮影・編集を，4 章で対話的映像メディアについて順に述べる．

2.　映像メディア処理の課題: 取得と提示

我々の目的とする対話型の映像メディアは，ユーザが何らかの質問(または問いかけ)をし，それに対してシステム側が適切な映像断片を選びだし，それを適切に編集しながら提示する．そのためには，次のような機能が必要とされる．

映像データの蓄積: 種々の要求に答えるためには，答となるデータを大量に蓄積しておく必要がある．その際

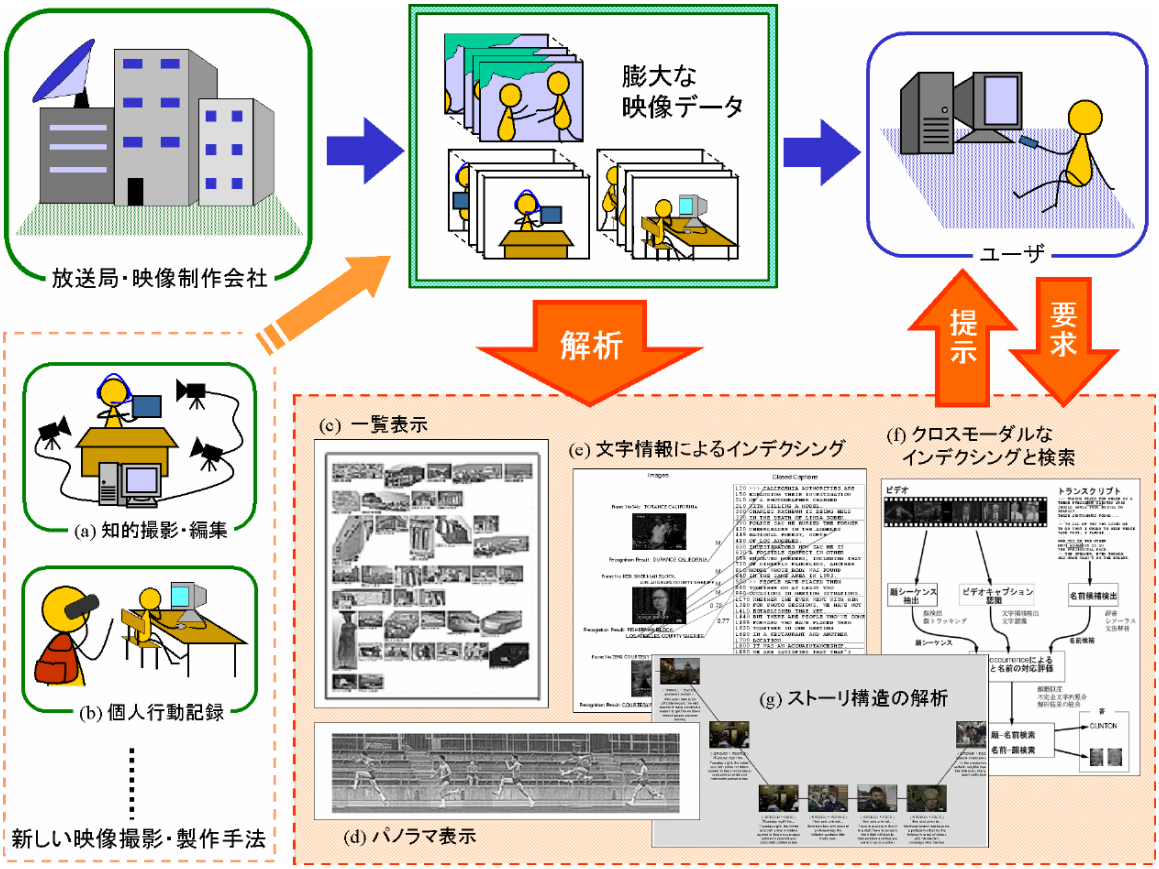

図 2:: 映像・マルチメディア処理とその目的

に，詳細なインデックスを付与すればするほど，データの利用価値が上がる．

ユーザの要求を理解する機能: ユーザの質問の意図を推定するためには，ユーザの質問を自動的に処理(自然言語処理)する必要がある．また，システムが柔軟な応答を行うためには，ユーザの置かれている状況やその文脈を認識することが必要となる．

適切なデータを検索し，編集して提示する機能: ユーザの要求に答えるためのデータを検索し，その中から適切なものを選ぶ必要がある．また，ユーザが理解しやすいように，状況に応じた編集をデータに加えて提示する必要がある．

これらの機能に関する，映像処理のこれまでの研究や実用化を分かりやすく整理すると，図２のようになる．上段が従来の映像コンテンツの流れであるが，一方向的であり，マルチメディアとしての有効利用は難しい．そのため，90 年代の初めから研究されてきたのが図２の右下部分，放映番組や資料映像等の解析である．スポーツ番組や娯楽番組のハイライトだけを見たい，ニュース番組のダイジェストを作って欲しい，長期間蓄積された映像から自分に興味のある事柄に関するものだけを集めたい等の要求に応えることによって，映像データのそれまでになかった利用方法を可能にするものである．

そのために，映像中の言語情報を解析する方法や，また，そのために，言語情報を音声認識によって得る手法，映像中のキャプションを抽出して文字認識を行う手法も提案されている[7] ([図２(e)) [8]．また，映像内の画像情報を用いた例も多い．その最も一般的な例としては，人物の顔があげられる．例えば，登場人物によるインデックスをつけるためには，顔認識を用いることが必要となる[9]．さらに，言語情報との統合によって顔画像データベースを構築し，それを検索に用いる試みも行われている[10] (図２(f)). 映像中の顔から名前, 名前から顔が検索できれば，映像中の人物情報を検索，提示するための有効な手段となる．

以上のように，映像にインデックスを付け，それを基に必要な部分を再利用する枠組み自体はある程度実現されてきたが，我々の目的とする映像メディアに必要な機能が十分に実現されているわけではない．本稿では，以下の問題点に焦点を絞り，それに対処するため我々の研究について，次節以降で紹介する．

- 映像データの取得: 放送番組や既存の映像だけではコンテンツの量が不足すること，著作権のしばりを受けずに自由に使えるコンテンツが欲しいこと等

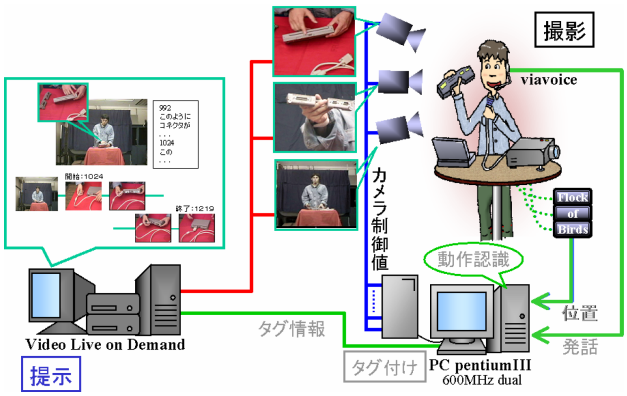

図 3: 自動的に「撮って編集する」システム

図 4: カメラ設定の選択表（一部分を抜粋．ユーザはここから自分の目的とする撮影対象を選ぶだけで良い）

から，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たいという需要は大きい．しかし，映像製作の専門家(例えば TV 局や映像製作会社)以外が，他人の視聴に耐えられるレベルの映像を撮影することが難しい．例えば，ちょっとした子供向け工作番組を撮ることを考えても，多くの人員と多大な労力が必要となる．また，検索が可能となるように映像データにインデックスを付加する処理も，まだ十分に自動処理できるとは言い難い．

- 適切なデータを検索し，編集して提示する機能: 自然言語処理技術の援用により，質問に関連するデータを検索することはある程度可能であるが，映像のように複数のモダリティを持つメディアを駆使して，できるだけわかりやすい方法でユーザに答や実例を提示する手法については，まだ十分に明らかになっていない．映像は内部に多くの情報を含んだメディアであるため，データを誤用すると大きな誤解を招くものとなる．また，適切な編集を行わなければ，非常に見づらいものとなる．

3. 映像メディアの取得: 簡単にコンテンツを作る

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，質の良い映像を撮ることは難しい問題である．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．そのため，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．

我々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)場面を題材として，図３のようなシステムを構築した[17] [20]．このシステムに，**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)の 2 つの機能を持たせることによって，手軽に映像メディアを製作する環境を実現する．

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機能．**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そのパラメータ**との関係を探り，わかりやすく不快感がない映像を取得する．

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し，それを強調するための編集を行う機能．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**ことが重要なポイントである．

我々の構築したシステムでは，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．各々のカメラに対し，図４のように，撮影の対象と目的を簡単に指定することで，その目的にあったカメラワークが決定され，カメラの pan/tilt 制御などが行われる．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．

ここで重要なのは，それぞれの時点で重要な箇所を一つだけ選択しながら撮影・記録していくのではなく，重要な情報源となる可能性のある**複数の部分を同時に追跡し，それらを映像として記録しておく**点である．これにより，種々の目的で映像を見る人の要求に応えるのが容易になる．例えば，ある時点でのある人物の表情に注目したい場合と，その時点で他の人が背後でしている行動に注目したい場合など，複数の相反する要求に応えるこ

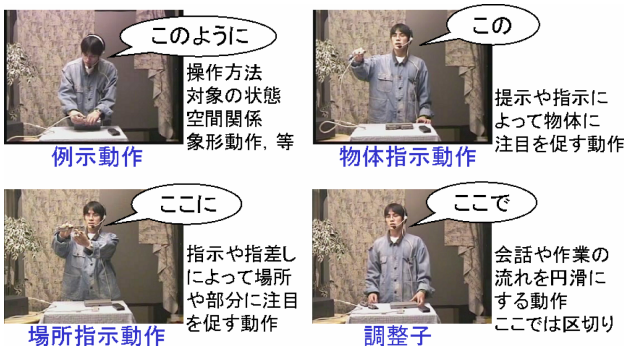

図 5: 注目を要求する動作

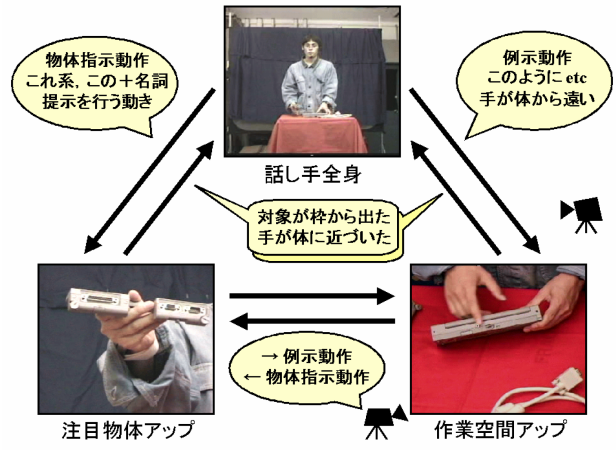

図 6: 使用したショットとその切り替え条件

とが可能となる．

このシステムでは，撮影時に，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，その時点でのシーンの状況を映像へのタグとして付加する．これが視聴者の質問に応えるためのインデックスとなる．例えば，どのような発話をしているか，また，どこに注意を向けようとしているかなど，種々の付加情報があれば，質問に応じて映像データを検索・提示するための良いインデックスとなる．

さらに，このインデックスを用いて，映像の編集を行うことができる．数秒以内の映像クリップとして提示する場合には，編集なしの映像でも十分に良い提示形態となるが，それよりも長い映像を提示する場合には，編集が必要となる[iii]．その一つの方法として，図５のように，話し手が注意を促している動作を検出し，それ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ができる．一連の映像として提示する場合の編集規則例を図６に示す．

これらのしくみを使って実際に撮影されて編集された映像の例を図７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．ここにあげたシステムは，机上作業シーンを撮影するためのシステムとなっているが，会議，講義，スポーツ等の撮影も同様の枠組みで実現できると考えられ，これからの発展が期待できる．

4. 対話的映像メディア: 教えてくれるメディア

4.1. QUEVICO の概要

料理や機械の組み立てのような作業を行うことを考えてみよう．図８のような状況で人間の先生に質問したならば，言葉だけではなく写真を見せる，図を描く，実演を行うなどしてわかりやすく教えてくれるはずである．映像のように複数のモダリティを持つメディアを駆使して，例えば，図９に示すように，できるだけわかりやすい方法でユーザに答や実例を提示することはできないだろうか．

関連する研究としては，自然言語による知的ヘルプシステムや質問応答システムに関する研究が数多く報告されてきた．しかし，動画，音声，テキスト等が混在するマルチモーダルデータの扱いには特有の問題があり，そ

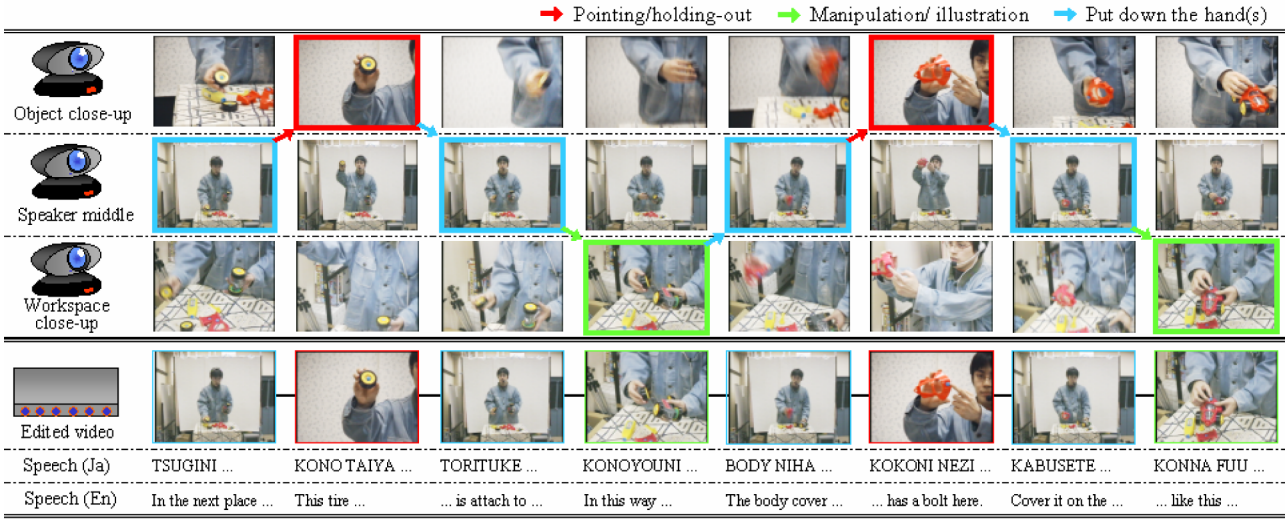

図 7: 自動撮影・編集の結果 (かなり良い映像が得られる)

図 8: 対話的映像メディアの利用環境の一例

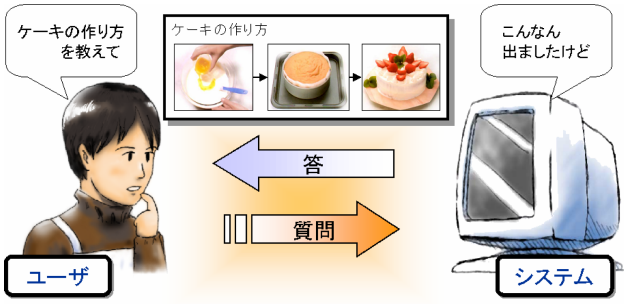

図 9: 質問に答えてくれる映像メディア

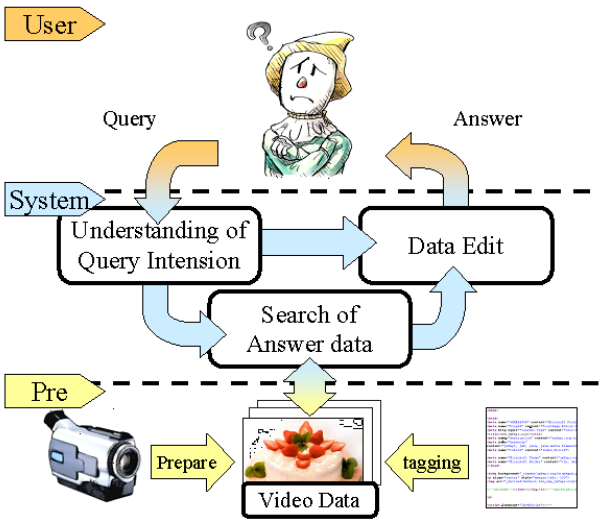

図 10: 質問応答の概略 (質問の解析・データの検索・答の編集・提示からなる)

れらの手法を単純に適用することはできない．そこで，我々は新しい枠組み *QUEVICO*[iv] を提案している．

この枠組みのポイントは以下のようになっている．

「質問と答」からのインデキシング: 作業などの映像に対して様々な質問を想定し，答となる部分をマークアップするために必要となるタグセットを設定した．これによって，質問が「要求した情報」を含むデータ断片を選択する．

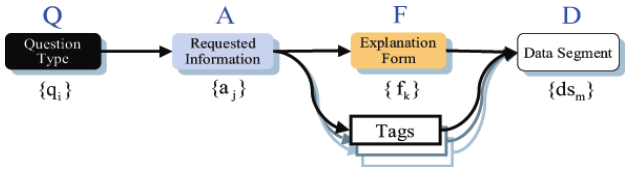

図 11: 経路モデルの概要

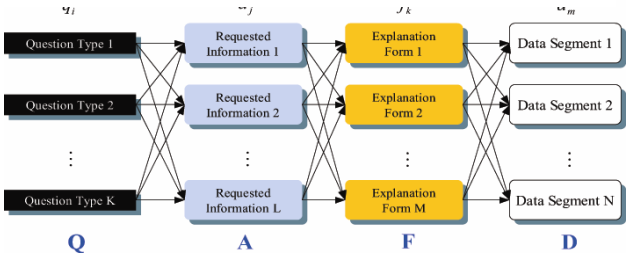

図 12: 多対多による各要素間の関係 (各矢線が各々の要素の関連性を表す．本研究ではリンクの重みと呼ぶ)

複数モダリティの効果的な利用: 複数のモダリティを用いて，質問の答として最も適切な提示方法を選択する手法，及び，十分なインデックス(タグ)が与えられていない場合でもそれなりに答える手法を提案した．

QUEVICO システムの基本的な枠組みは，図 10 に示したように，従来からの多くの質問応答システムと同様のものである．質問の解析，答となるデータの探索，答となるデータの加工と提示という経る．しかし，映像のようなマルチモーダルデータを扱うために，従来の自然言語や音声による応答システムに比べ，応答するのに使えるデータの自由度が大きくなるため，答として適しているデータを使うことの重要性が高い．そのため，我々は，上記 1 番目の項目について十分な調査を行った．

さらに，質問応答の際に，応答に適したデータを実際に探索する手法も問題となる．そのために，2 番目の項目について新しい手法を提案する．この手法は以下のような考え方に基づく．質問が行われた場合，人間はまず，質問のタイプ(Q)からその質問が要求する情報(A)が何であるのかを推測し，何が(F)その要求される情報を提供するのかを考え，それが実際に含まれているデータ断片(D)を探し出すという三段階の経路を経ることで，答となるデータ断片を求める．これをモデル化したのが図 11 の経路モデルである．十分なインデックスが(タグ)与えられれば，これで質問に答えることができる．しかし，多くの場合には完全なインデックスを付与することは難しい．そのため，複数の要求される情報や答の形態，データ断片の関連性を考え，図 12 に示されるように多対多のリンクにより経路モデルを考えることにより，「質問」と「答となるデータ断片」をつなぐ．これについては，次節で具体的に説明する．

図 13: 料理映像「かつおのたたきの調理」

表 1: 質問タイプとして要求される情報の例(抜粋)

質問タイプ	要求された情報
～とはどのような作業ですか	説明, 用法
誰が～しているのですか	動作主, 方法, 程度
何を～するのですか	対象, 入力
～するには何が必要ですか	入力, 道具
～したらどうなりますか	出力, 終点
何を使えばいいのですか	入力, 道具, 方法, 程度
どこで～しているのですか	場所, 始点, 終点, 方法
～で使うものはどこにありますか	始点, 場所
どこに～すればいいですか	終点
いつ～しますか	時間

これらを実現するために, 図 13 に示したような多視点の映像データが複数台のカメラによって撮影され, QUEVICO で定めたタグセットによりマークアップされて，未編集のまま蓄えられる．システムは，ユーザとの対話を通じて提供すべき情報を推定し, 図 14 に示すように返答する．例えば，ユーザがかつおを切り身にする作業において「どの程度切るのですか」と質問した場合，システムは「程度」を説明する映像断片と「1cm 程度の厚さにスライスする」という言語的な説明をユーザに提示する．現在の仕様では，表 1 を含む 30 種類程度の質問に対して，複数のモダリティを有効に利用してユーザに答えることができる．

4.2. 質問に答えるプロセス

ユーザの質問に答えるためには，既に述べたように，数段階のプロセスを経る必要がある．システムがユーザの質問に答えるプロセスの概要を図 15 に示す．まず，ユーザの質問を解析し「状況」,「トピック」,「質問パターン」を特定する．次に，インデキシングされ蓄積されたデータを探索し答を含むデータ断片を得る．そして最後に，そのデータ断片を編集することで，適切な形で答を提示する．

本研究では，作業を対象としているので，状況とは現在行っているタスクであり，トピックとは質問の話題に

図 14: QUEVICO の応答例

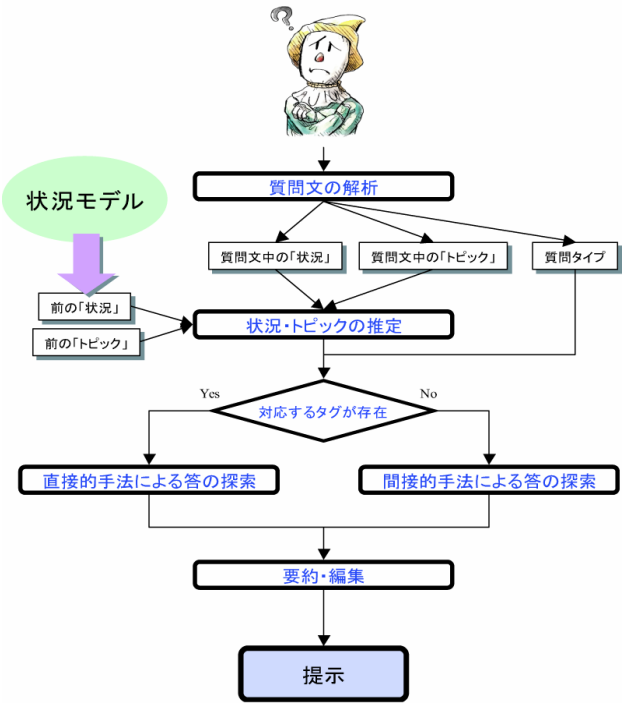

図 15: 質問に答えるプロセス

されているタスクや物体として定義できる．また，質問パターンは質問が要求している複数の情報群の重みとして表現する．例えば，「どの程度～すればよいですか」という質問に対する質問パターンは，各要求された情報との関係が (agent, degree, instrument, …) = (0.1, 1.0, 0.6, …) のように設定されている．

システムは，図 15 に示された手順により，質問の答となるデータ断片を探索する．この探索には，前節で述べた経路モデルを用いる．この時，要求された情報に直接対応するタグがデータ中に付与されているか否かにより，直接的手法と間接的手法を使い分ける．

直接的手法は，図 12 の下部の経路を用いる探索手法である．これは，「質問と答」によるマルチモーダルデータへのインデキシングに基づく手法であり，ユーザの要求に対応するタグを通じて答を含むデータ断片を得る．直接的手法は，本研究の中心をなすものであり，的確な答を得る唯一の方法である．そのため，蓄えられるマルチモーダルデータには十分なタグが付けられていることを前提とする．

間接的手法は，図 12 の上部の経路を用いるものであり，このようなマルチモーダルデータの特徴を利用し，答を含む可能性のあるデータ断片を探すことでユーザの質問に答える．具体的には，付加されたメタデータとその属性（データタイプ，モダリティ，意味的な役割）を元にしたスコアにより答を含む可能性の高いデータ断片を推定する．

間接的手法では，以下の三種類の情報を元に答を含むデータ断片を選択する．

表 2: メディアタイプの一覧

メディアタイプ	説明
映像（シーン）	シーン空間全体を捉えた視点の映像
映像（動作主）	動作主を捕らえた視点の映像
映像（動作）	動作主が行う動作空間を捉えた視点の映像
映像（対象物体）	動作の対象となる物体を捉えた視点の映像
発話文	動作主が行う発話文
発話文中の単語	動作主が行う発話文中の単語あるいはフレーズ
音	シーン中で発生する音

メディアタイプ: 本研究では，表 2 に示すメディアタイプを用意している．このメディアタイプを，図 12 に示した多対多の経路モデルに基づき選択する．本研究では，多対多の経路モデルを 2 階層のニューラルネットワー　クとみなし，人手による初期値設定とデータを元にした学習により質問の答を含む可能性が高いメディアタイプを求めることができる．

タグの可換性: ある質問の答が別の質問の答を含んでいるということはあり得ることである．例えば，方法を要求する「どのように切ればいいのですか」という質問と程度を要求する「どの程度切ればいいのですか」という質問の二つの質問があった場合，対応する映像断片を見せることでどちらの答とも成り得る．このような場合，本研究ではタグが可換であると言う．このようなタグの可換性を計る手段として，メディアタイプの選択で要求された情報と情報を提供する形態間の重みを利用する．具体的には，図 12 の経路モデルにおいて，要求された情報に対する各情報を提供する形態間の重みを表すリンクの値から，相関値を用いる．

データ断片の長さ: もし，得られたデータ断片に完全に答が含まれていたとしてもそれが数時間もの長さを持っているのでは良い答であるとは言えない．本研究では，タグを付けられたデータ断片からデータタイプ毎に平均 μ と分散 σ を求め，その長さが平均から外れるほどスコアが悪くなるように設定している．

答となるデータ断片は直接的手法と間接的手法を用いることで次に示すプロセスにより得ることができる．

1.ユーザの質問を受け付け，トピックと質問タイプを推定する．
2.トピックに質問に対する適切なタグが付けられている場合には直接的手法で答が見つかる．その場合には，

得られたデータ断片を答とする.

3. 2 の手法が適用できない場合，間接的手法により，モダリティタイプ，タグの可換性，データ断片の長さに対するスコアを導出する．最も高いスコアを得たデータ断片を答とする.

以上のような方法で, 図 14 に示したような応答が得られる．ユーザの現在置かれている状況に合わせて答を提示する部分はまだまだ不十分であるが，映像データをもとに質問に答えるメカニズムとして，実際のデータに対して動作するものを提案したという点で，意義が大きい．今後，自然言語処理部分を強化して，自然な応答ができるようにすること，ユーザの置かれている状況を画像処理などによってセンシングすること，より高度なタグ付けの自動化を行うこと，等の課題があり，これから手がけていく予定である.

5. おわりに

本稿では，知識流通を支援するための映像の自動取得と，それを用いた質問応答システムについて紹介した．具体的には，映像を手軽に利用する環境やそのための要素技術として，映像の知的撮影・編集システム，対話的映像メディアについて述べたが，これらが，実際のプレゼンテーションシーンや机上作業の記録映像に対して動作することを示した.

本稿で提案したモデルは，幅広い範囲で利用可能なものであると考えられるが，まだその詳細については十分に整理されていないのが現状であり，これからの事例蓄積, 問題の整理等が必要である．また，ユーザとの対話・会話のためには，ユーザの置かれている状況の認識が不可欠であり, 今後, 画像や音声の自動認識技術を用いて，より状況に適した対話・会話が可能となるシステムのモデルを探る予定である．そのためには多様な技術が必要であるため，他の分野との横断的な研究協力を進めていく.

さらに，ここでは紹介しなかったモダリティ変換(例えば, 文章や映像の図的表現[23] [24]や, 個人行動記録(例えば, [15]，遠隔通信，映像編集等の問題も大いに知識流通のために活用できる可能性がある．上述の研究課題と合わせ, これらの可能性を探っていく予定である.

参考文献

1) 中村, 向川 (1998)「画像・映像の知的生成と編集 - CV 技術を用いた新しい画像・映像処理」 松山, 久野, 井宮編『コンピュータビジョン：技術評論と将来展望』(第 17 章), 新日本コミュニケーションズ.

2) 有木 (1999)「メディア解析から見たパターン認識」『信学技報』 (PRMU99-171).

3) 中村, 外村 (1999) 「見たい部分を簡単に短時間で--- 気の利いた映像メディア技術を目指して---」『信学誌』(Vol.~82, No.~4).

4) 外村, 谷口, 阿久津 (1994)「PaperVideo:紙を用いた新しい映像インタフェース」 IEICE,IE94-59.

5) Taniguchi. Y.,Akutsu. A.,Tonomura. Y.(1997). PanoramaExcerpts:Extracting and Packing Panoramas for Video Browsing, *ACM Multimedia97*.

6) Wactlar. H., Kanade. T., Smith. M., and Stevens. S. (1996). Intelligent Access to Digital Video *The Informedia Project. IEEE Computer* (Vol.~29, No.~5)

7) 佐藤, 金出 (1999)「文字認識と異種情報の対応関係に基づいたニュース放送からの情報抽出」『情処論』 (Vol.~49, No.~12)

8) 有木ほか (1996)「ニュース映像中の記事に対する音声・文字・映像を用いた索引付けと分類」 『信学技報』 (PRMU96-97)

9) 佐藤 (1999) 「ドラマ映像における登場人物のアノテーションシステム」 第 5 回知能情報メディアシンポジウム

10) Sato. S., Nakamura. Y., and Kanade. T. (1997). Name-it: Naming and detecting faces in video by the integration of image and natural language processing. *IJCAI*.

11) Nakamura. Y. and Kanade. T. (1997). Semantic analysis for video contents extraction --- spotting by association in news video. *ACM Multimedia*.

12) Smith. M. and Kanade. T. (1997). Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques. *IEEE CVPR*,.

13) 上田 (1999)「コンピュータを駆使した最新の放送番組製作技術」. 『情報処理』 (Vol.~40, No.~11).

14) Informedia Experience on Demand. ``http://www.informedia.cs.cmu.edu/eod/''

15) Kubota. S, Nakamura. Y, Ohta. Y (2002) Detecting Scenes of Attention from Personal View Records -- Motion estimation improvements and cooperative use of a surveillance camera, *Proc. IAPR Workshop on Machine Vision and Applications* (pp.209-213).

16) 中村 (2000) 「コミュニケーションのための画像・映像処理」 『信学技報』, PRMU99-252.

17) Ozeki. M, Nakamura. Y, Ohta.Y (2001) Camerawork for Intelligent Video Production ----Capturing Desktop Manipulations, *Proc. Int. Conf. on Multimedia and Expo* (pp.41--44, CD--ROM TA1.5).

18) Ozeki. M, Itoh. M, Nakamura. Y, and Ohta. Y, (2002) ``Tracking

hands and objects for an intelligent video production system," *Proc. Int. Conf. on Pattern Recognition (* pp.1011—1014).

19) Izuno. H, Nakamura. Y, Ohta.Y (2002) QUEVICO: A Framework for Video-based Interactive Media, *Int'l Workshop on Intelligent Media Technology for Communicative Reality* (pp.6-11).

20) Ozeki. M, Nakamura. Y, and Ohta.Y . (2002) ``Human behavior recognition for an intelligent video production system", *IEEE Proc. Pacific-Rim Conference on Multimedia*, (pp.1153—1160).

21) 西田ほか (2003.3) 「料理教示発話の構造解析」言語処理学会 第 9 回年次大会,

22) 西田豊明 (研究代表) (2003.3)「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」科学研究補助金, 学術創成研究, 研究成果報告書

23) 村山, 中村, 大田 (2003) 「DocScape: 文章の概覧性を高めるための概念図の生成と利用」『情処論』 (Vol.44, No.4)

24) 村山, 伊津野, 中村, 大田 (2001) 「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)

---

i) ここで用いる「対話型」と「会話型」の厳密な区別は定義されていないが, ここで扱う対話には多人数での会話が含まれていないため,「対話型」と表記することにしている.

ii) 「かつおのたたきが食べたいな」のようなちょっとした発話から知識の検索要求まで, 種々のものが考えられる.

iii) 人間の注意力が持続するのは短くて 2, 3 秒, 長くて, 15 秒程度と言われており, 映像に注意を向け続けさせておくためには, カメラワークや編集の助けが必要となる.

iv) QUEstion-based Video COmposiotion: この名は, 古事記に現れる久延毘古神を典拠とする. 久延毘古神は, 案山子の姿をしており世の様々な出来事に熟知しているとされる.

Intelligent Video-Based Media Processing for Knowledge Distribution

Yuichi NAKAMURA [1]

[1] Institute of Engineering Mechanics and Systems, University of Tsukuba (yuichi@image.esys.tsukuba.ac.jp)

This paper introduces a new support for knowledge distribution by our intelligent video production system and question-answering scheme for video-based interactive media. For realizing easy-and-effective use of video-based interactive media for knowledge distribution, we have various topics to tackle intensively. We need to approach by both media contents acquisition and supports for flexible use. This paper presents, for this purpose, (1) an intelligent video production system, and (2) a question-answering scheme for video-based interactive media. This paper describes, as for (1), the multi-camera system that automatically track targets and capture videos with useful indices obtained by human behavior recognition; as for (2), the question-answering technique for dealing with various questions by the users.

Key Words: *knowledge distribution, video-based interactive media, intelligent video capture, intelligent video editing, question answering*