[特別講演] **コミュニケーションのための画像・映像処理**

中村 裕一
筑波大学 機能工学系
〒305-8573 つくば市 天王台 1–1–1
（yuichi@esys.tsukuba.ac.jp）
科学技術振興事業団 さきがけ研究21

あらまし: マルチメディアという言葉がすっかり定着したが，複合メディアの本質的な部分には，まだ手付かずの問題が多い．短時間でわかりやすく情報や知識を伝えるメディアの形態，また，その構築方法について，これから試行錯誤で探っていく必要がある．本稿では，複合メディアの一つである映像をテーマとし，映像を製作する処理を計算機で補助，自動化する問題について述べる．具体的には，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，蓄積，再構成等の問題について紹介し，従来の研究とこれからの方向性について考える．

キーワード: 映像によるコミュニケーション，映像生成，映像処理，マルチモーダル処理，プレゼンテーション映像の処理，知的映像撮影システム

Image and Video Analysis for Communication

Yuichi NAKAMURA
Institute of Engineering Mechanics and Systems,
University of Tsukuba
1-1-1 Tennodai, Tsukuba
PRESTO, Japan Science and Technology Corporation

Abstract: Although “multimedia” is now a very popular word, not a few open problems are left on handling and producing multimedia. We still need intensive research for realizing efficient and effective media for giving information or representing knowledge. In this paper, we focus on videos, and discuss the way for supporting or automating video production. Current research on intelligent video capturing and editing for presentation scenes or personal activities is introduced. Then, we discuss the feature and the future direction of intelligent video production.

key words: communication by images and videos, video production, video analysis, multimodal analysis, presentation video analysis, intelligent video capture

1　はじめに

近年，映像を計算機で簡単に扱うことができるようになり，種々の映像処理が提案されてきた．それに伴って，画像処理，コンピュータビジョン，マルチメディア等に関する種々の学術雑誌，国際会議で映像処理に関する研究成果が報告されるようになってきた[1]．実際に，筆者らが参加した ACM マルチメディア国際会議 ('99) の予稿集を見ても，何らかの形で映像を扱っている研究が半分以上を占めていた．また，映像の再構成を視野に入れた研究や個人のための新しいサービスの創成といった新しい研究テーマが提案されつつある．例えば，パネルディスカッションで，video portal[2]がテーマに取り上げられていた．そこでは，大量に配信されている映像を集めて，インデクスの付加や再構成を行い，個人の嗜好に合わせて提供することが目的となる．また，News on Demand を研究している Informedia Project[4] でも，これから (第 II 期) の研究ではデータを断片化し，目的に応じて再構成することを視野に入れているとのことであった．

このように，映像処理は少しづつテーマを変えながら進展してきたが，現在の研究は既存の映像の処理に大きく偏っている．つまり，経験を積んだ専門家が手間をかけて作り込んだ映像コンテンツを処理する研究が大半を占めている．

しかし，専門家以外の人でも映像やそれを基にしたマルチメディアコンテンツを作りたいという要求は大きい．また，映像の構造化やインデックスの付加など，手間のかかる問題も多い．そのため，映像製作を支援するための研究が必要となる．

このような要求に応えるためには，映像を撮影する，また，それを基にした自動的に編集を行う知的システムが必要となる．実際に，少しづつではあるが，このような研究がなされ始めている．本稿ではそのために必要となる要素技術について考え，研究例を紹介する．また，従来の画像・映像処理との違いについて，私見をまじえながら述べる．

[1] 既にいくつかの解説があるので，それを参照して頂くと流れを把握して頂けるだろう．例えば，[1][2][3] 等がある．

[2] 筆者は，portal という単語を，ユーザに対して映像コンテンツを提供する窓口，また，そのサービスを意味する単語として解釈しているが，なかなかうまく訳すことができない．良い訳をご存じの方にはぜひ教えて頂きたい．

2　映像メディア製作の問題とは

2.1　何ができて欲しいのか

まず，映像製作の専門家以外の人達が映像メディアを利用する場面を想定してみよう．例えば，

- 映像を駆使したマルチメディア教材，説明を手軽に作る．
- 情報発信のために，WWW コンテンツとして映像を用意したり，電子メイルの代りに (電子) ビデオメイルを利用する．
- 遠隔会議，遠隔講義等において，その場の状況を映画や放送番組のように効果的に伝える．
- 自分の経験を自分や他人のために記録し，必要に応じて共有する．

等があげられる．活用場面としては，教育や会議がすぐに思いつくが，それに限らず，日常のコミュニケーションや，ちょっとした事柄や経験を伝えるような場面でも利用できる．

実際に，画像・映像を WWW コンテンツとし，手軽に情報発信・提供をしたいという要求，また，受け取りたいという要求は非常に大きい．筆者の学科にも，映像のストリーミング技術を使って，自宅で講義映像を見るのが夢だと言う学生がいる[3]．

そのため，情報発信やコミュニケーションを補助するための画像・映像メディア処理が必要とされている．つまり，映像製作の専門家でなくても，それなりに満足できる映像メディアを構築することを支援する知的システムが望まれている．後で紹介するが，上記のマルチメディア国際会議で，講義・講演映像の取得と利用方法を提案する論文が，Best Student Paper と Best Paper に選ばれたことからも，その必要性が認識されつつあることがわかる．

2.2　要素技術として何が必要か

上記の課題は，いずれも，映像によって言語や音声では伝わりにくい情報を捉え，効果的に伝えること，それをできるだけ自動化することを目的としている．そのためにはどのような要素技術が必要なのだろうか．

図 1 に，筆者らの研究 [5] が扱っている問題を簡単に示す．詳しくは 3 章で紹介するが，人間の行動，特にプレゼンテーションを人間に代わって見て，人間に

[3] 社会的・経済的な問題があるため，当人が在学中には実現する見込みはないが，技術的には実現可能な範囲にある．

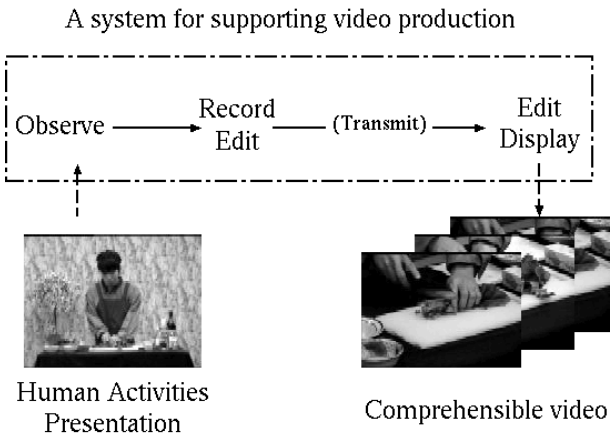

図 1: 映像によるコミュニケーション支援

代って提示・説明することがその目的である．これを簡単に説明すると次のようになる．

観測，撮影，認識: 適切な視点から適切な解像度で対象を撮影する．撮影対象，特に，人間の行動に適応した撮影が必要である．シナリオとして事前情報が与えられる場合でも，人間が機械のように正確に動くことは期待できないため，種々の認識が必須である．これはロボットの認識行動サイクル，アクティブビジョン等の考え方と類似しているが，出力が映像メディアであり，それを見るのが人間だという点が異なっている．そのため，カメラワークが重要な問題となる．

情報付加，編集，提示: 撮影されたデータを編集し，見易くわかりやすい映像データに構成する．そのためには，あらかじめシナリオや撮影時の情報等を付加データとして与えること，外部データとの関連付けを行うことも必要となる．それを基に，ユーザの目的に合わせて映像データを提示する．

このような枠組みは，筆者らの研究に限らず，上記の問題すべてに共通するものである．また，ここであげられている項目の多くは， これまでの画像計測，認識，ヒューマンインタフェース，等の技術との重なりを持つが，映像メディアを扱うために特有の問題も生じる．以下で，その具体例について紹介する．

3 映像の撮影と認識

3.1 観測・撮影・認識

従来から，人物の追跡やそれに基づいたカメラの制御に関する研究はあったが，スタジオのカメラシステムを自動化する応用をはっきりと明示したのは

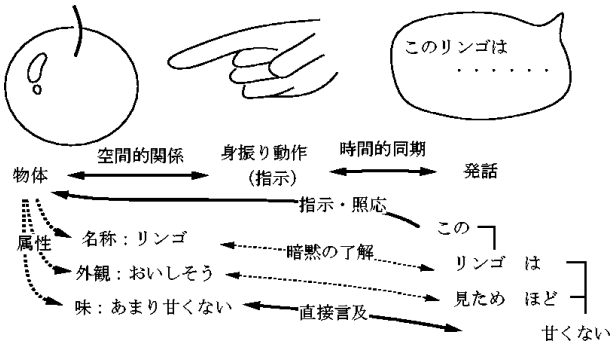

図 2: モダリティの複合的な利用

Pinhanez, Bobick らの研究が初めてである [6]．彼らのシステムでは，複数のカメラが顔，人物像全体，その他の対象を (必要に応じて追跡しながら) 撮影する．得られる複数の映像ストリームを，ディレクタの指示により切り替えることによって，映像を生成する．これは，動画像処理，顔認識をうまく組み込んだ動物体追跡・撮影システムと考えることができる．撮影システム自体の進展はその後報告されていないが，同グループでは，その後ジェスチャ認識を用いたインタラクティブシステムに関する研究が精力的に行われている (例えば，[7] 等)．

このような撮影システムにおいて特に重要な問題となるのは，撮影対象である人物の意図を推測することである．そのためには，動作，発話，周囲の状況などを観測し，これらの情報を統合的に処理しなければならない．

その問題に対して，筆者らは文献 [8] で，動作と発話を統合的に認識することによって，話し手の意図を理解する枠組みを提案している．図 2 に示されるように，人間のコミュニケーション手段である複数のモダリティ間には種々の相互関係があり，人間の意図を理解するためには，指示・照応関係，補足関係，強調関係をうまく利用することが必要であることを主張した．

しかし，ここで問題となるのが，撮影される側の人間の行動についてまだ十分に調べられていないことである．この問題に対して，筆者らは，撮影段階で複数のカメラから得られた画像，出演者の発話，動作計測データ等のマルチモーダルデータを統合的に記録するシステムの構築を行った [9]．図 3 にその概要を示す．

多視点映像: 複数台のカメラで撮影を行う．大部分

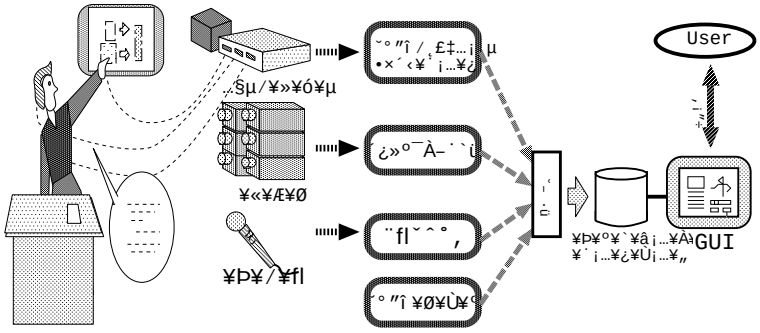

図 3: 多視点マルチモーダル映像取得システム

図 4: 多視点映像の例

表 1: 動作ラベルの語彙

置く，入れる，取り出す，持ち上げる，切る
刺す，押す，叩く，ねじる，たたむ，引っ張る
こする，振る，練る，かき回す，すくい上げる
ひっくり返す，くっつける，めくる

は可動カメラであり，ホストコンピュータからズーム，パン，ティルトを制御される．これによって，撮影対象を常に追跡しながら撮影を行うことができる．この多視点画像の例を図 4 に示す．これにより，カメラは注目しなければならない可能性のある要素を集中的に撮影できる．なお，撮影対象の追跡には，磁気センサで計測された身体各部の位置を利用している．

動作計測データ: 話し手の上半身の動作を磁気センサを利用して計測する．この磁気センサは受信器の位置と方位角の 6 自由度をビデオレート又はそれ以上の周波数で測定できる．

発話文: 発話をテキストとして入力する．音声で記録する共に，テキストとして手動で書き起こして利用する[4]．各発話文には，開始，終了時の映像のフレーム番号を付加する．

動作ラベル (正解データ): 話し手の動作について，人間が判断した結果 (正解データ) を動作の開始時刻，終了時刻とともに記録する．本研究ではプレゼンテーションや作業を対象としているので，動作の語彙は表 1 の 19 語と「指示」に限定した．

各データはすべて映像フレームに対して同期がとられている．これによって得られたデータを蓄積することによって，人間の行動や発話の形態，相互関係を網羅的に調べたり，認識システムを構築する際のテストベッドとすることができる．

[4]実際の認識実験を行う際には音声認識エンジンを利用する．

筆者らのグループで，実際に蓄積したデータを用いて，簡単な動作認識とそれを利用した映像提示を行った例を文献 [10, 5] に報告した．これらの研究では注目すべき対象を注目要素と定義し，プレゼンテーションの話し手の言動から注目要素を推定する．具体的には，指示・例示動作，その他の動作を抽出し，その部分を重点的に見せる． まだシステムが十分に完成していない段階であるが，指示・例示動作を比較的良好に抽出することができること，それを基に映像の切り替え，要約を行うことができることを報告している．また，図 5 のように，目的に応じて重要フレームを抜き出し，視点を選ぶことによって映像の要約とすることを提案している．

以上で述べてきたような撮影システムの将来的な形は，注視対象，撮影方法，撮影時間，その他の条件を目的に応じて自動的に決定し，それを画像，音声と同時に記録していく総合的なシステムになるだろう．

また，新しい映像表現のために，機械システムや画像処理技術を駆使することも試みられている [11]．ここでは，放送番組レベルの高い完成度が必要とされるため，必ずしも自動化が主眼になっていないが，これからの映像生成研究には大いに採り入れられるべきものである．

3.2 個人視点映像の取得と処理

個人的な行動をビデオ等で記録し，個人的な記憶の補助として使う研究が進められてきた [12][13]．このような考え方で個人の見聞きした情報を映像を用いて記録すると，経験を蓄積したり，伝えるための手段とすることができる．

このような場合，利用者の周辺を常に撮影しなければならないため，身体装着 (ほとんどは頭に装着) して人間と共にカメラを移動させることになる．その

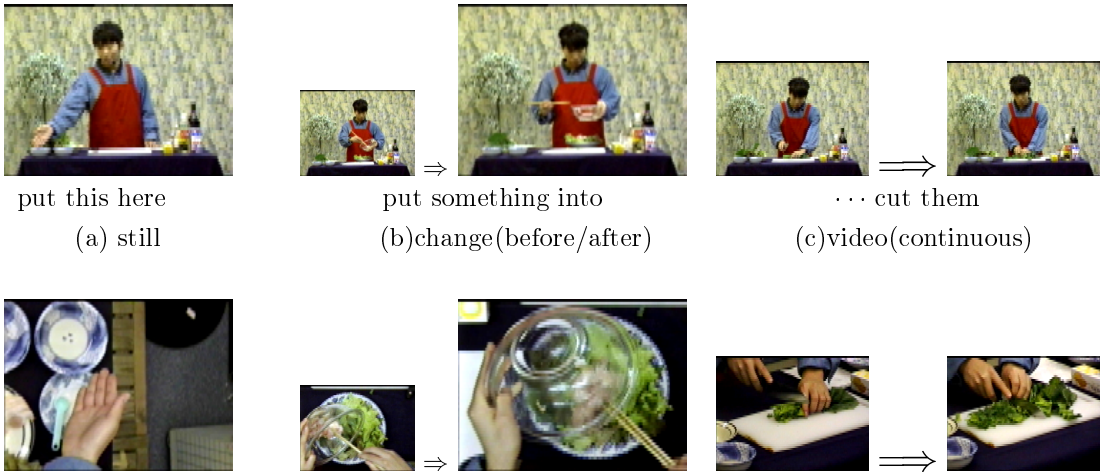

図 5: 重要フレームと視点の選択: 上段が時刻だけの選択．下段は視点 (カメラ) を選択した場合

際に問題となるのは，得られる映像データが長時間のデータになること，身体動作のために映像が見辛いものになることであり，そのため，映像として長時間提示できるものにはならない．したがって，構造化や要約を行うことが必須となる．

そのために，筆者らは，利用者が何かに注目したシーンを核として個人行動記録・要約を実現する手法を提案してきた [15][16]．まず，図 6 のような注目動作を考え，このような動作が起ったシーンを注目シーンとする．

対象を注視する動作 (積極的な注目): 図 6(a) のように人が静止したまま静止している対象物をじっと見る場合や，図 6(b) のように動いている物体を首を振りながら，または体を動かしながら視界に入れ続けるよう追いかける等の動作．

長時間の視点の停留 (停留シーン): 同じ場所で比較的長時間，同じ出来事を見続ける．この場合は，図 6(a) に体の揺れが複合的に加わったものになる．

これらの注目シーンを，画像中の見かけの動きを検出することによって検出する．検出された注目シーンを用いて生成した要約例を図 7 に示す．左一列に停留シーンが並べられており，縦方向が時間の経過を表す．この実験例では，部屋からキッチンへ移動し，そこで一連の料理を行う過程が構造化され，作業の概略が分かりやすく提示されている．また，このような映像を用いることによって，個人の経験から必要な情報を検索する効率が良くなるという結果が得られている．

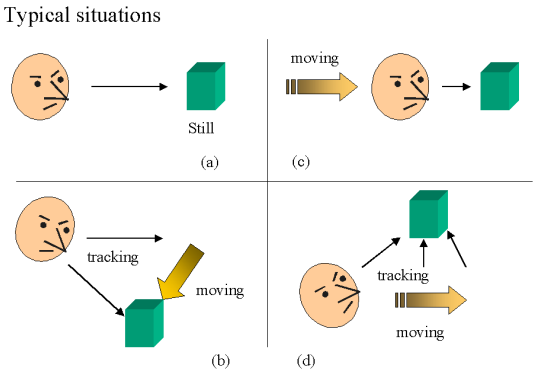

図 6: 人の注目動作 [15]

以上のように，個人視点からの映像を要約し，短時間の映像の集まりとして提示することによって，経験の蓄積や共有のために映像を用いることが容易になる．

また，将来的には，環境に用意されたカメラやスタジオに用意された高機能なカメラからの映像との併用が面白い研究課題になると考えられる．

4 映像の編集と提示

4.1 カメラの選択，要約

まず，マルチメディア国際会議で Best Student Paper[17] と Best Paper[18] に選ばれた講義・講演に関する映像生成研究について紹介する．

Mukhopadhyay らの研究 [17] では，2 つのカメラで教室を撮影する．一つのカメラ (overview) は広い視野で講義の概観を撮影し，もう一つのカメラ (tracking)

図 7: 注目シーンによる要約例: 前半では，本を見るという冗長な場面は簡潔にまとめられ，後半では，作業内容の詳細が提示されている．

は講師を追跡して講師の表情や動作を撮影する．この 2 つを切り替えることによって 1 つの映像を生成し，この映像と講義スライドとを併せて提示する教材を作る．この 2 つのカメラからの映像ストリームを切り替える方法は以下のようになっている．

1. OHP スライドの交換時刻を画像から検出する．
2. スライドの交換をトリガにして，映像を切り替える．スライド交換時は overview カメラで 8 秒間 (交換前 3 秒，交換後 5 秒)，それ以外は tracking カメラの映像とする．
3. 見易くするために，短すぎるカットは除去する．
4. 退屈させないために，長すぎるカットは分割し，もう一方のカメラからの映像を挿入する．

上記のルールを適用することによって，概ね良い結果が得られたと彼らは報告している．このような簡単な手法でも，実際にシステムを構築し，本当に動かしてみせたところが評価できる．

He らの研究 [18] では，講義映像の要約の問題を扱っている． 通常，一連の講義は長時間にわたるため，学生側には要約が欲しいところである．また，教師側にとっても，内容をチェックしたり差し換えたりするのに便利である．そのために，種々の方法で要約を生成し，要約の良さを評価することがこの研究の目的である．

要約の生成手法は以下の 4 通り試された．

- スライド交換時刻を用いて要約 (S)
- 音声のピッチだけを用いて要約 (P)
- スライド交換時刻，音声のピッチ，講義映像に対するユーザのアクセス回数を用いて要約 (SPU)
- 講師が自分で要約 (A)

各々の要約の長さをほぼ同じに設定した場合の，各々の方法の一致度，及び，要約の良さを評価している．実験結果では，要約の良さが，$A >> SPU \simeq U > S$ となったことを報告している． 要約の良さをいくつかの点から評価するという試みは大いに評価できるが，講師が自分で要約したもの以外は，あまり大きな差が出ていない．これからのより詳細な検討が必要であろう．

日本では，前節で述べた筆者らの研究，亀田，美濃らの研究 [19][20]，大西らの研究 [21] がある．亀田らは遠隔講義のための講義映像の撮影方法．映像切り替え方法を提案している．この研究では，数台のカメラで教室を撮影し，遠隔地へ講義映像を配信することを目的とする．可動カメラによって講師を追跡すること，講師の位置，黒板の情報を元に映像切り替えを行う手法が提案されている．また，最近の発表では講師の指示動作を用いることが提案されている [20]．これらのシステムは実際の講義中に実時間で動作させることを想定しており，今後の展開を大いに期待できる．

4.2 映像への情報付加

映像の自動編集を行うためには，映像の構造を明記するデータや，ショット，シーンなどに対する付加情報を映像に与えるのが有効である．

ここで，映像の素材データにどのような付加情報が必要と考えられているかについて，少し紹介しておこう．Davenport らはユーザと対話的に映像を提示する Interactive storytelling を実現しようとしている [22][23]．そのために各ショット (映像の最小単位) に付加すべき記述を大きく以下のように分類している．

Perspective: 誰の視点 (立場) で撮影されたか

Camera and sound recorder: どのように撮影されたか (例えばカメラの位置やその他のパラメータ)

Content: 映像中で何が起こっているか

Context: どのような前後関係や背景があるか

また，柴田らは映像部品の記述モデルを提案している [24]．それには，被写体，被写体の動作，被写体の状態，関係，カメラワーク，音声に関する項目が含まれる． 例えば，この記述モデルを用いた素材データを用いて，映像の構造を自動的に作り出すこと等が応用として考えられる．

また，最近では，素材をデータベースに収録し，それを再構成する形で映像を製作するを集めて要求に合わせて組み直す手法 [25]，素材を階層的に管理して，編集する手法 [26] 等が提案されている．

前節までで述べてきた映像の自動撮影に関しても，上記の研究で提案されている映像の整理法を用いれば，より手軽に映像編集が可能になるであろう．

5 映像生成の位置づけと評価

5.1 編集済映像の処理との違い

最初に述べたように，従来から既存映像の解析が行われてきたが，これからも以下のような理由で，編集済の映像に対する研究が行われ続けるだろう．

- 膨大な既存映像データの蓄積があり，それを再利用可能な形に変換する必要がある．
- 映像の送り手側 (放送局等) は，営業上の問題，著作権上の問題から，全てのデータを伝送することができるわけではない．したがって，受け手側で映像を解析し，受け手にとって利用しやすいデータに変換する必要がある．

そこで，既存映像の処理と，本稿で述べてきた映像生成の研究との主な違いをあげることにより，新しい研究分野としての位置づけをはっきりさせておこう．まず，従来の映像処理では解けなかった問題が，十分な事前情報と種々のセンサによって，解ける問題に変換できることがあげられる．

計測，認識: 撮影環境，シナリオ，その他の事前情報を最大限に利用できる．従来の映像処理では，シーンの再構成や認識が ill-posed な問題となっていたが，これを回避できる．また，種々のセンサを用いることにより，映像以外の情報を取得できる．

さらに，構造解析処理の目的も異なってくる．

構造化の目的: 従来の映像処理では，編集済の映像を断片に分割し，各断片の役割，編集者の意図を推定することが行われてきた．映像生成の研究では，このような処理が必要なくなるが，逆に，シナリオや人物の行動からその意図を抽出することによって，映像の構造化を計る必要がある．

また，“入っていない情報は処理することができない”という問題を回避できる．

データの多様性と編集・要約の自由度: 多数のカメラで撮影した映像，その他のセンサ出力を並行して記録することにより，提示に必要となるデータを網羅的に取得できる．つまり，多くの素材を並行して残しておくことができるため，自由度の高い編集が可能になる．それに対して編集済の映像では，一つを残し他が捨てられているため，受け手側の処理に適さない部分が多くなる[5]．そのため，加工の自由度が小さい．

しかし，以下のような欠点もあり，人間の介在を想定したハイブリッドなシステムを構築することが必要だろう．

映像の質: 見て面白く楽しい映像を作ることが難しい．現在のところ，専門家が知恵と手間をかけて作った映像にはかなわない．また，魅力的な俳優に出演してもらうことも難しい．そのため，娯楽性や芸術性が重視される用途には使いにくい．

5.2 新しい評価基準が必要

本稿で紹介してきた研究分野は，その可能性が少し見えてきたといった段階であり，まだまだ種々の実現方法を模索していく必要がある．しかし，映像生成の自由度が非常に大きいため，同時に何らかの評価を行っていかなければ，研究の存在意義自体が疑われることになるだろう．

その評価として重要なものには以下のようなものがあげられる．

- 理解しやすさ (概要の把握しやすさ，細部のわかりやすさ，記憶に残る度合い，その他)
- 生理的評価 (見やすさ，疲労感，その他)
- 感性的評価 (面白い，美しい，その他)

これらの項目を実際に定量的に評価するためには

[5] 例えば，会議の出席者 A に関する情報が欲しいのに，映像中では延々と人物 B が映っている場合など．

多くの工夫が必要となるが，従来の心理的評価手法に加えて，以下のものが有効であろう．

問題の正答率，解答時間: 映像の視聴者に問題を出し，その正答率が良いもの，または解答時間が短かったものを，良い映像 (またはその加工品) とする．

個人による生成結果との一致度: 被験者 (専門家，素人，その他) に自分の最も良いと考える生成物を作ってもらい，それとの一致度を評価する．

この他にも種々の評価方法が確立されることが望まれる．また，評価の信頼性を高めるためには，研究者間でデータセットを共有し，それを用いた評価実験をする必要がある[6]．

6 おわりに

本稿では，コミュニケーションのために映像メディアを製作する問題，また，それを加工する問題について概説した．映像メディアを生成するためには，動画像処理だけでなく，種々の認識処理，メディア処理が必要になることを述べ，いくつかの特徴的な研究を紹介した．また，この分野はまだあまり認知されておらず，他研究分野との差異も十分に認識されていないため，5 章では従来の研究との差を明確化しようと試みた．本稿には多くの私見が入っているため，細部について多くの読者の賛同を得られる自信はないが，おおよその問題意識は共有して頂けることを期待している．映像生成を新しい問題として十分に認知して頂くとともに，本稿が映像生成に対する興味をもって頂けるきっかけとなれば幸いである．

参考文献

[1] 有木康雄. メディア解析から見たパターン認識. 信学技報 PRMU99-171, 1999.

[2] 中村, 向川. 第 17 章, “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」. 新日本コミュニケーションズ, 1998.

[3] 中村, 外村. 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, 1999.

[4] H. Wactlar, et al. Intelligent Access to Digital Video: The Informedia Project. *IEEE Computer*, Vol. 29, No. 5, 1996.

[5] Y. Nakamura. Multimodal approach toward intelligent video production. *International Workshop on Multimedia Intelligent Storage and Retrieval Management* (`http://www.info.uqam.ca/~misrm/`), 1999.

[6] C. Pinhanez and A. Bobick. Intelligent studios: Using computer vision to control tv cameras. *Proc. of the IJCAI'95 Workshop on Entertainment and AI/Alife,Montreal (also MIT Perceptual Computing TR-324)*, 1995.

[7] A. Wilson and A. Bobick. Parametric hidden markov models for gesture recognition. *IEEE Trans. PAMI*, Vol. 21, No. 9, 1999.

[8] 中村, 西谷, 大田. プレゼンテーション映像における話者の行動理解. 信学技報 PRU95-143, 1995.

[9] Y. Nakamura, et al. MMID: Multimodal Multi-view Integrated Database for Human Behavior Understanding. *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, 1998.

[10] 木村ほか. 発話情報と動作情報を用いたプレゼンテーション映像の要約. 信学技報 PRMU97-197, 1998.

[11] 井上誠喜. 放送における映像処理 ～新しい映像表現を目指して～. 信学技報 PRMU99-71, 1999.

[12] 飯島ほか. 日常生活映像から検出された人物像によるエピソード想起. 信学技法, PRMU97-196, 1998.

[13] 石川ほか. エピソード映像の時空間的階層呈示による記憶想起. 信学技法, PRMU98-186, 1999.

[14] T. Jebara, et al. “dypers: Dynamic personal enhanced reality system”. *MIT Media Laboratory, Perceptual Computing Technical Report ♯463*.

[15] 大出, 中村, 大田. 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, 1999.

[16] 大出, 中村, 大田. ビデオ映像による個人行動記録・要約システム. 第 5 回知能情報メディアシンポジウム, 1999.

[17] S. Mukhopadhyay and B. Smith. Passive capture and structuring of lectures. *Proc. ACM Multimedia*, 1999.

[18] L. He, et al. Auto-summarization of audio-video presentations. *Proc. ACM Multimedia*, 1999.

[19] 亀田ほか. 講師追跡によるカメラ映像の自動切り替え. 情処全大, Vol. 58, 2V-04, , 1999.

[20] 大野ほか. 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, 1999.

[21] 大西, 松本, 福永. 情報発生量による遠隔講義映像の自動生成とその評価. 信学技報, PRMU98-176, 1999.

[22] W. Mackay and G. Davenport. Virtual video editing in interactive multimedia applications. *Communication of the ACM*, Vol. 32, No. 7, 1989.

[23] G. Davenport, T. Smith, and N. Pincever. Cinematic primitives for multimedia. *IEEE Computer Graphics and Applications*, Vol. 11, No. 4, 1991.

[24] 柴田. 映像の内容記述モデルとその映像構造化への応用. 信学論, Vol. J78-D-II, No. 5, 1995.

[25] 住吉英樹. 放送番組製作におけるデータベースとその生成技術. 情報処理, Vol. 41, No. 1, 2000.

[26] 上田博唯. 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, 1999.

[6] 実際に，PRMU 研究会 VDB-WG(馬場口委員長) では，映像処理のための共通テストベッド整備について検討し始めている．