

平成11年度科学研究費補助金 特定領域研究 (B)

人間の行動状況理解のための時空間の非線形圧縮による

映像の情報要約機構

研究代表者	美濃 導彦 (京都大学総合情報メディアセンタ)
研究分担者	松山 隆司 (京都大学大学院情報学研究科)
	河原 達也 (京都大学大学院情報学研究科)
	亀田 能成 (京都大学総合情報メディアセンタ)
	東海 彰吾 (京都大学大学院情報学研究科)

1. はじめに

計算機技術の進歩は、計算機を計算する機械から情報メディアと変えつつある(参考文献(7)(8)). 情報メディアとしての計算機の最大の利点の一つは、表現メディアとして画像を用いるもの、すなわちビジュアルな情報を実時間で対話的に伝える視覚情報メディアである。

視覚情報メディアの中に、そのメディアが伝えるべき情報を圧縮する形態がある。これは、視覚情報メディアが実世界から視覚情報をカメラなどで簡単に取得できる状況で、本来にユーザが必要な情報は何か、知りたいことは何かをユーザとのインタラクションにより取得し、必要な視覚情報のみを効率的に取得する形態である。実世界をいかようにも観察でき、さまざまな視覚情報を取得できることを考えると、ユーザが欲しいものだけを効率的に映像化する技術は、従来の MPEG のような与えられた映像を圧縮する技術とは本質的に異なるものである。本研究では、この種の技術を映像の非線形圧縮技術と呼ぶ。

情報を圧縮するときには何らかの制約が必要である。ここでは、講義のアーカイブを作成するときには制約として講師が利用する教材を考える場合と、スタジオで映像を取得するときにはシナリオを制約として考える場合を取り上げ、それぞれの例題を通して視覚情報メディアにおける情報の非線形圧縮方式について考える。

2. 視覚情報メディアとしての映像情報要約機構

2.1. 視覚情報メディアの枠組み

視覚情報メディアは、実世界の情報を利用者に伝える枠組みとして定義される。図 1 にその枠組みを示す。コンピュータは実世界と人間の間に位置する視覚情報メディアであるが、人間も同時に実世界と対話している。この実世界は、コンピュータと対話している実世界と同一であることも考えられるし、異なる実世界であるとも考えられる。コンピュータは蓄積装置をもち、実世界から得られた視覚情報を大量に蓄積出来ると同時に、効率よく検索できる。視覚情報メディアの対話処理として考えなければならないのは、実世界との対話と人間との対話であり、視覚情報メディアがメディアとして双方と対話を行うメカニズムを情報圧縮の枠組みで考えてゆかなければならない(参考文献(1)).

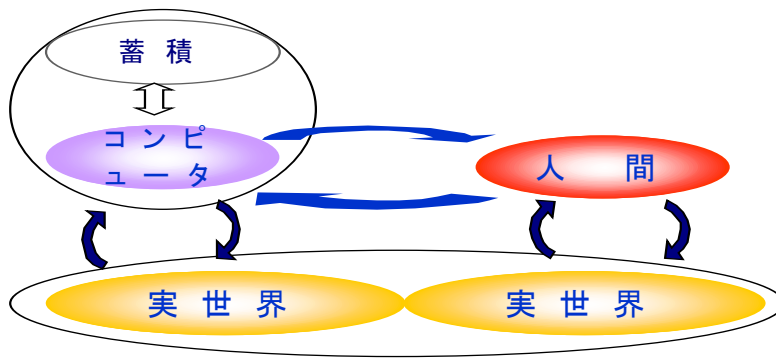


図 1 視覚情報メディアの枠組み

この図には全体で3つのループがあり、それぞれが対話処理を表している。コンピュータと表記されている視覚情報メディアは、実世界とインタラクションを行い、映像情報を取得する。取得された情報は蓄積され、後ほど利用するために何らかの構造化がなされる。ユーザが利用するときは、その要求にしたがって視覚情報メディアを介して蓄積された映像データとインタラクションを行い、必要な情報をユーザに提供する。

2.2. 映像情報の要約機構

視覚情報メディアを介した情報圧縮機構を考えるときには、情報が一方的に流れる場合と異なって、一度蓄積され、それが後で再利用される枠組みを考えなければならない。このことは、蓄積には2つの独立した処理があることを意味している。それは、実世界から情報を取得して蓄積する処理（蓄積処理）とユーザからの要求にしたがって蓄積された情報を検索し提供する処理（検索処理）である。

視覚情報の非線形圧縮処理をどちらの処理で行うかは重要である。情報を取得したときに圧縮処理を行う（蓄積処理での圧縮）のは、蓄積すべきデータ量が少なくてすむ利点がある一方で、実時間処理をしなければならないという制約がある。反対に、蓄積時は取得した全てのデータを蓄積しておき、検索時にユーザの嗜好により圧縮して提示する方法は、柔軟性があり有効であると考えられるが、蓄積すべきデータ量は膨大なものになる。しかし、オンラインでの選択処理に比べて、蓄積処理の利点は、処理そのものに時間をかけられることであり、全てのデータが蓄積されているとして、その中からどのようにして有効な情報のみを取り出し提示するかを時間軸の制約を受けることなく出来る点である。これを視覚情報メディアにおける情報圧縮の本質であると考え、実世界から得られる全ての情報を一旦蓄積し、実時間処理の制約を受けることなく、ユーザにとって意味のある圧縮を行っていく方法をとることが重要である。

ただし、カメラ制御など、実世界の状況に応じてカメラをコントロールするようなインタラクションは実時間性が必要であり、これを除外するものではない。

2.3. タスク制約の導入による非線形圧縮

実世界から得られる映像を非線形圧縮するためには制約が必要である。ここで考えてい

る圧縮は、一般的に言われている MPEG などの圧縮とは違って、原則として、実世界からユーザが望む映像を取得する処理である。ある時点でどのような映像を生成すればよいかは、実世界でそのときに何が起きているかに依存する。実世界では何らかのイベントが行われているので、そのイベントに対する制約を映像の非線形圧縮で利用することができる。

弱い制約として講義における教材をとりあげる。講義で利用される教材は、講師がそれに沿って講義をするという前提で作られているが、実際の講義が教材に沿って行われるかどうかは全くわからない。それでも、講義そのものにはある程度定まった形態があるので、ユーザとしての受講者がのぞむ映像を生成することは可能であり、非線形圧縮処理の例題となる。また、強い制約としてスタジオ撮影におけるシナリオを考える。スタジオにおいてはシナリオに基づいて物事が進行し、大まかな情報は得られるが、映像生成にはさまざまな困難が伴う。

これらの例題を通して、視覚情報メディアにおける映像情報の非線形圧縮における制約の働きを解明してゆく。

3. 講義室における視覚情報メディアの構築と情報圧縮

講義は講師が生徒に知識を伝達する目的で行われる講師と生徒のマルチメディアコミュニケーションであり、大学のもっとも重要な活動の一つである。講義に高度に発展した情報技術を利用しようとする試みが世界的に試行され、新たな教育方法の議論とあいまってその重要性が高まってきている。

講義室で電子メディアを利用して行われる講義は、まさにマルチメディア情報が氾濫している。プロジェクタで PC の画面を投影したスクリーンは、その中に文字、線画、画像、映像などさまざまなメディア情報を含んでいる。これらは、常に講義室において提示されている。また、黒板は、文字と線画を中心とした情報を表示できる。あらかじめ用意されている教材と違って、講義の流れの中で講師が思いついた情報を提示するメディアである。したがって、どんな情報が書かれるかは全く想像できない。

講師は、身振り、手振りを使って音声で講義内容を説明する。講師の行動状況を理解することが、ユーザである受講者の欲しい映像を生成することであり、映像情報を非線形圧縮することである。

3.1. 講師の行動状況

講義を形態的に見ると、講義室のあちこちに提示されているマルチメディア情報を講師が指し示す、すなわちプレゼンテーション情報を講師が示していると考えられる。講義にはさまざまな形態があるので、まず、非線形圧縮処理の対象とする講義の形態を以下のようなものと仮定する。

1. 教材は原則として電子メディアで用意され、プロジェクタに投影される。
2. 学生は PC をもたず、講義室でプロジェクタを見ながら講義を受ける。
3. 講師は（電子）黒板に板書する。

4. ビデオは教材と切り替えてプロジェクタに投影する。

このような講義は、大学で増えつつあり、特殊なものではなくなっている。

このような形態の講義が講義室で行われた場合、講義室に存在するメディア情報として以下のようなものが考えられる。

1. プロジェクタに投影される教材
2. (電子) 黒板に提示される情報
3. 講師の指示動作
4. 講師の説明音声
5. 生徒の反応
6. 生徒の質問音声

1 の情報はすでに電子化されており、内容的には問題はない。このマルチメディア情報に対して得なければならない情報は、どのタイミングでどの内容が提示されたかという時刻情報である。2 は電子黒板を利用するかしないかでその映像化法が大きく異なる。電子黒板を利用できれば、1 の教材とほぼ同程度の処理になるが、通常の黒板の場合はさらにその情報を何らかの形で記録しなければならない。

3 の講師の指示動作が非線形圧縮された映像を取得する上で最も重要なものである。講師の行動状況は、講師の音声がどの教材を説明しているかと深い関係がある。講師の指示動作は、教材の中のどの部分を説明しているかの情報を示している。したがって、講師が講義室のどの位置にいるかを計測し、講師の映像を取得することで、講義室内のさまざまなマルチメディア情報が関連付けられることになる。4 はマイクで講師の音声をとりつづければよいが、他の情報と同期を取るために時刻情報が必要である。

5 と 6 に関しては質問しているとき以外は講師に比べて重要性が低いので、3 と 4 に準じて考えればよい。

3.2. 講義室での視覚情報取得システム

これらの情報を実世界から取得するシステムを考える(参考文献(2))。教材はすでに電子化されているので、あらかじめ計算機に与えられているが、それが切り替えられたタイミング情報を実世界から得なければならない。このためには、映像による観測をもとにした判定も考えられるが、ここではすでに開発して利用可能な音声操作プロジェクタ(参考文献(5))を利用することを考える。したがって、教材を提示するプレゼンテーションシステムは、講義の前に講師によって作成された教材をページ単位で表示するとともに、ページめくりを講師の音声コマンドによって行う。

講師のプレゼンテーション情報は講師の映像として取得する(参考文献(4))。講師は講義室内を移動するので、カメラは講師を追跡し撮影しなければならない。位置が固定された首振りカメラ1台では、講義室における講師の自由な動きを全てカバーすることは出来ない。講義室にカメラを複数台設置する。さらに、講師からプレゼンテーション情報を得るためには、講師を出来るだけ大きく撮影する必要がある。このためには、講師をズームアップして撮影しなければならず、追跡と撮影を同じカメラで行うことは困難になる。そ

ここで、カメラを大きく二種類に分離し、一方を観測カメラ、他方を撮影カメラとする。これにより、撮影カメラは観測の制約を受けることなく撮影できるので、ユーザに最もいい映像を提供できるようになると考えられる。

音声収録は講師につけたピンマイクより行う。生徒の音声はハンドマイクでしゃべらせるか、天井につけた天井マイクより収録する予定である。音声操作プロジェクタは、コマンドの前後に無音区間を要求するので、教材の切り替えのタイミングに説明が続くことを禁止する。これは、後ほど行うメディア情報間の時間同期をとる処理に大変有利なものとなる。

図2に講義室の機器構成を示す。図からは観測用のカメラは省かれている。撮影用のカメラは4台あり、プロジェクタは2台ある。

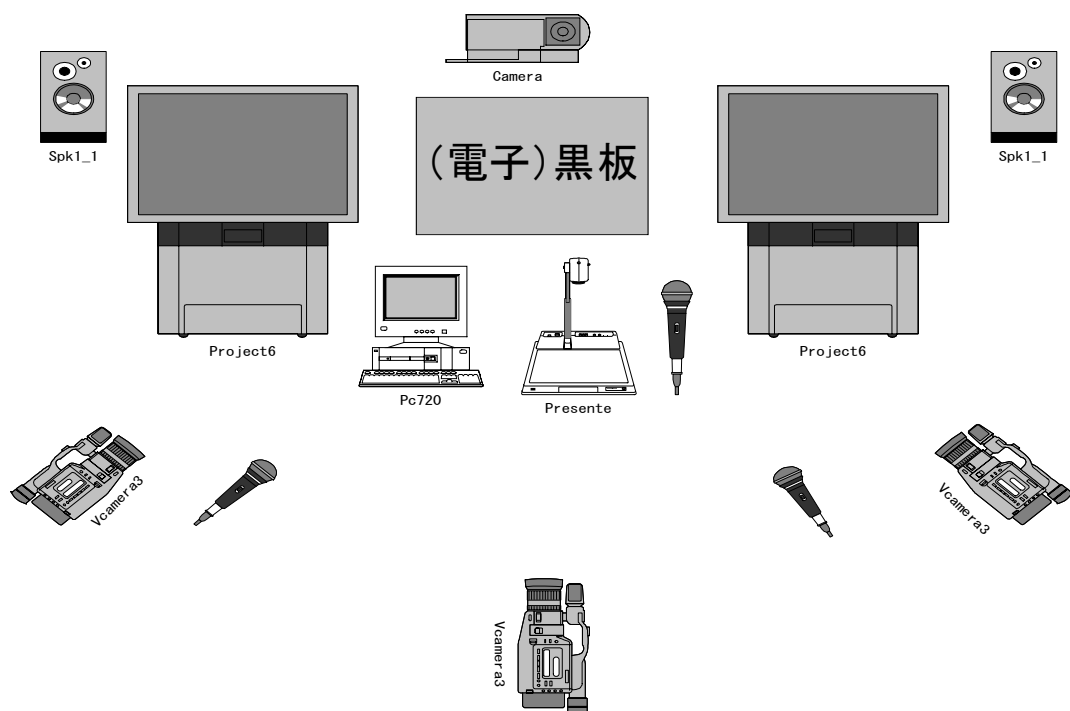


図 2 講義室の機器

3.3. メディア情報の区切り検出と時間同期

実世界には時刻があり、全てのイベントには発生の時刻が定義できる。各メディアを記録するシステムは、実世界から得た情報をそのまま記録してだけでなく、実時間で検出したさまざまなイベントをその発生時刻とともに共通のアーカイブ処理用データベースに蓄積するようにする。ここで考えるイベントは、メディア情報の区切りを検出するものとする。

プレゼンテーションシステムから得られるイベントは、教材の提示が切り替わったことである。これは音声操作プロジェクタのコマンドから検出する。ビデオが提示された場合

はスイッチャのコマンドから検出する。

カメラシステムから得られるイベントは、カメラを切り替えたタイミングである。カメラシステムは、実世界の動的状況を観測しその変化を検出してカメラ切り替えを行う。動的状況は、主として講師の位置によって識別されるので、講師が動くことによりカメラ切り替えが起こる。同時に、ユーザがあまり長く同じカメラからの映像を見ていると飽きにくるという視点から、動的状況に変化がないときでもカメラ切り替えが起こる。

音声システムのイベントは、講師がしゃべるのをやめることによって起こるとする。これは、音声の無音区間を検出することにより検出する。

これらのシステムは、それぞれ独立に検出したイベント情報をアーカイブ処理用データベースに蓄積する。このデータベースの構造を以下のように設計する。

- システム名 (who) : どのシステムがこのエントリーを作成したか。
- 対象 (what) : 対象は何か。
- 場所 (where) : 講義空間のどこで起こったか。
- 時間 (when) : イベント検出のタイムコード。

講義室での視覚情報収録システムを図3と図4に示す。

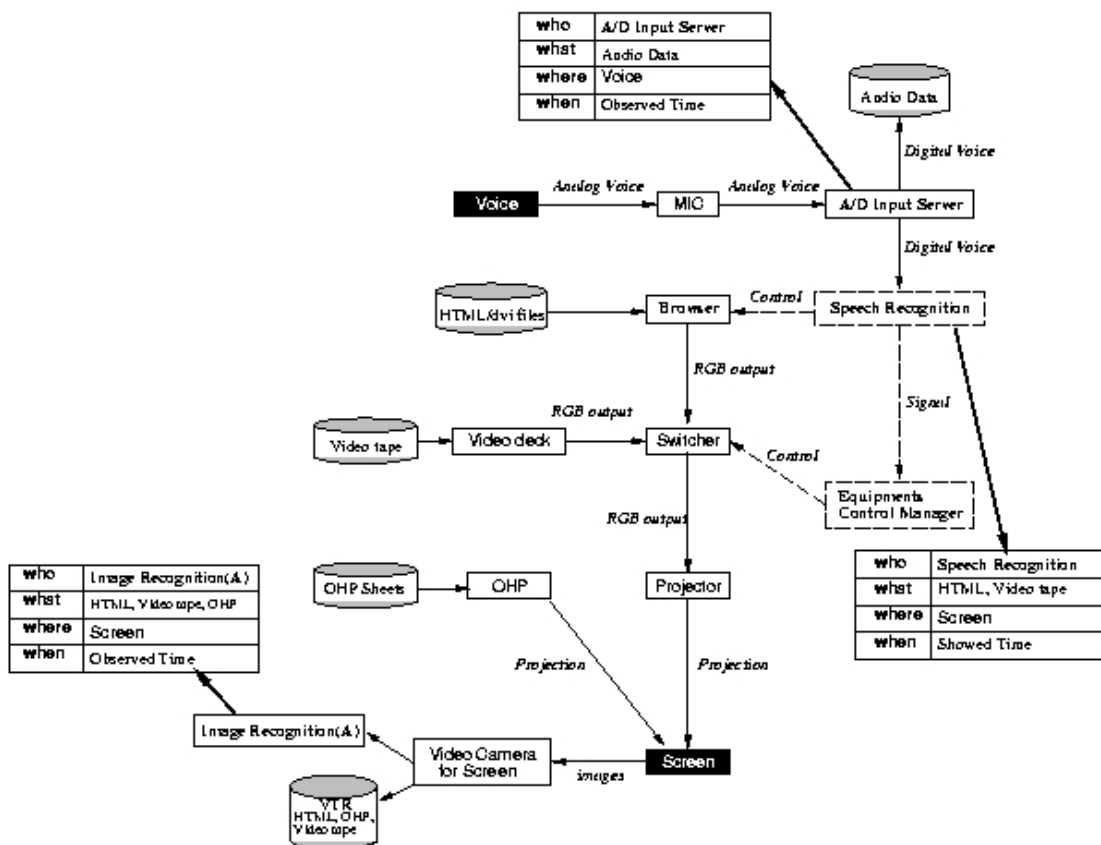


図3 講義映像蓄積システム(1)

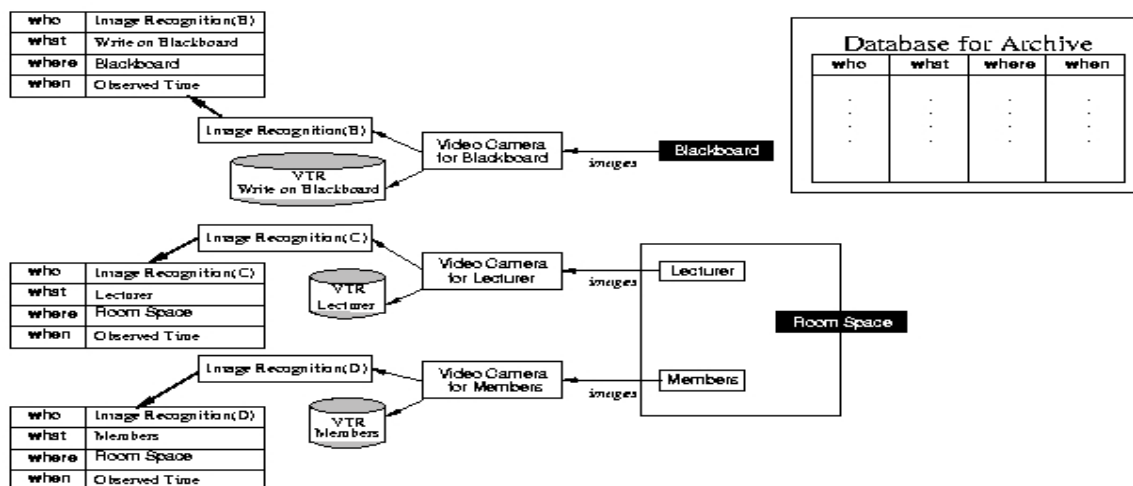


図 4 講義映像蓄積システム(2)

3.4. 区切り情報の収集と解析

実際の講義を収録し、区切り情報の収集と解析を行った。対象として、京都大学工学部の講義 2 回分、京都大学大学院情報学研究科の講義を 1 回分収録した。講義は各 90 分で、学部の講義 2 回の講師は同一人物であり、大学院の講義の講師は学部の講義の講師とは異なる。

表 1 講義収録により得られた区切り情報の数

講義種類	教材	音声	映像		
			講師	生徒	黒板
学部 1	8 8	1 7 4 1	7 0	1 0 9	7 7
学部 2	4 6	2 6 4 9	1 2 0	1 1 9	1 1 9
大学院	2 8	2 4 1 5	1 8 4	2 0 0	9 5

これらの講義収録によって得られた区切り情報の数を表 1 に示す。ある時間での区切り情報の分布を図 5 に示す。図 5 中左から講義開始からの経過秒数、教材、音声、映像（講師映像、生徒映像、黒板映像）の区切り情報が示されている。各時間帯において、空白または | は区切り情報が得られなかったことを表し、- は一つ、= は二つ以上の区切り情報が得られたことを示す。音声においては、空白は無音状態、| は有音状態を意味する。

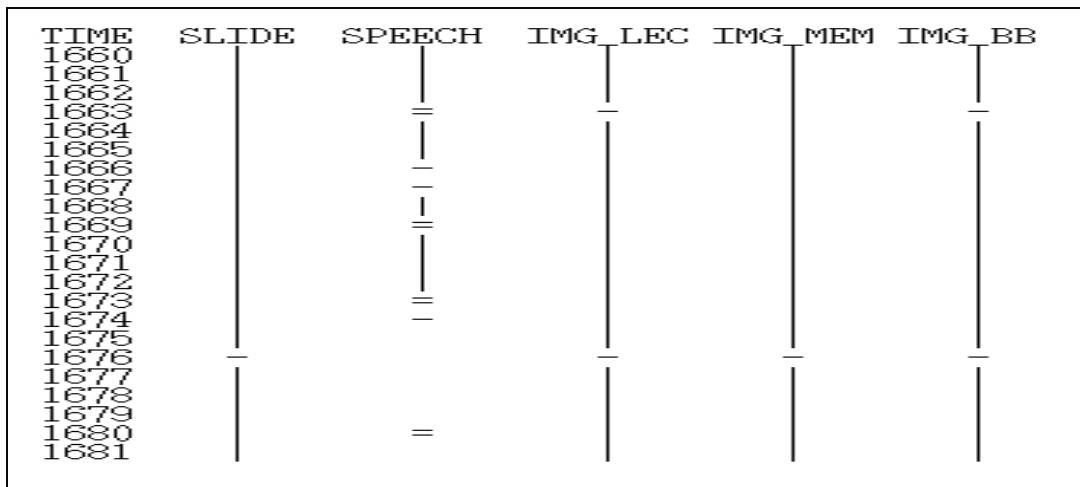


図 5 区切り情報の分布例

得られた区切り情報に対し，異種メディア間の区切り情報がどの程度近接していたかを検証した。

まず，音声操作プロジェクタを用いる前提により，テキストメディアの区切り情報は音声メディアの区切り情報の一部と完全に合致する。

次に，テキストメディアの区切り情報が得られた時点の前後 1.0 秒以内に映像メディアの区切りが 1 つ以上ある確率を調べた。その結果，上記 3 講義に関して 0%から 7.1%の低い確率であり，このことからテキストメディアと映像メディアに関する同期は時間情報からはほとんど取ることができない。

また，映像メディアの区切り情報が得られた時点の前後 1.0 秒以内に音声メディアの区切りが 1 つ以上ある確率についても調べた。その結果，41.3%から 60.0%の確率を得ることができたが，これは両メディアの区切り情報をランダム分布させた場合の平均確率 49.4%から 64.1%と大差なく，偶然性が高い。

以上の結果から，講義形態に与えた制約によりテキストメディアと音声メディアは区切りの同期が取れているが，それらと全く別の処理を行っている映像メディアは区切りの同期を取ることが難しく，現状では適切な構造化が困難と言える。従って，映像メディアの区切り情報生成には，他メディアの情報を利用し同期を取るような，アーカイブ化用の撮影方法が必要とされる(参考文献(6))。

3.5. 本節のまとめ

講師の行動状況を理解することにより，講義の情報を取得し受講者に有用な情報を提示するシステムを構築するために，講義情報を取得する手法を考案し，実際の講義の情報を収集する実験を行った。現在のところ，さまざまな問題が検出されており，改善が必要であるが，蓄積された情報をどのように提示してゆくかの方式とも関連があり，視覚情報メディアとしてきっちりと設計していかなければならない。

4. スタジオにおける視覚情報メディアの構築と情報圧縮

スタジオにおいて作成される視覚情報としては、映画やドラマなどのストーリーを持つものがあげられる。これらの映像では、スタジオで撮影されるシーンの状況があらかじめ台本などの形で決まっており、その事前知識に基づいた映像の撮影や加工が行われる。その際、そのシーンの状況が受け手にとって分かりやすく見飽きない映像になっていることが望まれ、さらに、送り手の意図を伝える映像であることも重要な要素である。

これらの要素を映像に付加するためには、単にシーンの状態を撮影し映像とするだけでなく、撮影の方法や見せ方に何らかの工夫が必要となる。このための一つのアプローチとして、シーンを複数のカメラを動かしながら多角的に撮影し、それらの映像を切り替えて提示することによって映像に変化をもたせ、受け手を映像に引き付けることがあげられる。従来の映像作品でもこの方法がよく利用されている。

このような、複数のカメラの操作や切り替えなど、広い意味でのカメラワークを付加してシーンを映像化することが、送り手、受け手の両者にとって必要であり、これによって初めて有用な視覚情報が構築されると考えられる。このことは、このように撮影された映像が、単にシーンの状況を提示するだけでなく、受け手による理解を助け、かつ、送り手の意図を伝えるメディアとして、より高い情報が映像の中に圧縮されたメディアとして位置付けることができるということを示している。

本研究では、このようなカメラワーク込みの映像を自動的に撮影する自動スタジオシステムについて考える。このシステムは、計算機から首振りやズームなどの制御が可能な首振りカメラをシーン内の複数の位置に設置し、それらを制御しながら撮影や映像の切り替えを行い、最終的な映像シーケンスを生成するものである。ここでは、シーンの状況があらかじめわかっていると、この事前知識に基づいて、カメラの配置や制御を立案し、所望の映像シーケンスとしてシーンを撮影し、映像の切り替えや加工を行うシステムの構築を目指す。

しかし、シーンを適切なカメラワーク込みの映像として可視化するためには、カメラを事前知識にのみ基づいて制御するだけでは不十分である。これは、カメラワークがある固定されたものとしては実行できず、シーンの状況に応じて適切に制御されるべきものであるからである。事前知識として得られる情報も、実際のシーンの状況とは必ずしも一致せず、幾何学的、時間的なずれが生じる。これらのずれに応じて、動的なカメラ制御の修正を行いながら求められる映像としてシーンの状況を撮影し、最終的な映像を作り出すことができるシステムが必要である。

本文では、スタジオにおける視覚情報生成において、まず、シナリオに基づくシーン状況の映像化システムについて、シーン状況の事前知識（シナリオ）と最終的な映像シーケンスの構成についての事前知識（ストーリーボード）を入力として、複数のカメラの適切な制御と切り替えを自動的かつ、適応的に行うための枠組みについて説明し、次に、個々のコンポーネントについて行った予備的実験について述べる（参考文献(9)）。

4.1. シナリオに基づくシーンの映像化システム

スタジオ内で進行するシーンの状況の動的変化を複数のカメラを用いて撮影し、ひとつの映像シーケンスとして映像化するシステムを考えると、以下の事項を決定する必要がある。

- カメラ配置：シーンを何台のカメラを用いて撮影するのか、また、そのシーン内のどのような位置にカメラを設置せればよいのか。
- 動的カメラ制御：求められた映像シーケンスに沿った映像を撮影するために、カメラの視点位置や首振りなどを、シーンの状況に応じてどのように制御するのか。
- 映像シーケンスとしての構成：カメラの切り替えをどのように決定し、最終的なひとつの連続した映像シーケンスとして構成するか。

ここでは、スタジオ内で撮影されるシーン状況、および、その状況をどのような映像として撮影したいかは事前知識として得られている場合を想定する。つまり、システムへの入力として以下の2点を考える。

- シナリオ：3次元的なシーンの状況の記述であり、シーンの幾何学的な情報や、その中での撮影対象の位置や向き、および、その時間変化などが定められたものとする。さらに、会話や身振りなど対象の動作についてもイベントという単位で記述し、それぞれのイベントの時間的關係（前後関係、同時関係など）や意味的關係（因果関係、階層的関係）を記述する。
- ストーリーボード：シナリオに記述されたシーンの状況をどのようなショットの組み合わせによって最終的な映像シーケンスとして構成するかを記述したもので、各ショットをどのような画像として撮影するかを定める2次元的な記述である。映像シーケンスを構成するショットの順に、そのショットがシナリオに含まれるどのイベント（群）を映像化するのか、さらに、そのショットにおける対象捉え方を規定する。具体的には、対象の画面内での大きさ（クローズアップ、バーストショットなど）や画面内での位置（中央、右、左など）、対象の捉え方としてのカメラ制御のタイプ（パニング、チルティングなど）などである。

さらに、シナリオ（3次元的記述）とストーリーボード（2次元的記述）の2つの事前知識を関係付け、適切なカメラワークによってシーン状況を映像化するために、以下のデータベースを利用する。

- カメラワークのノウハウ：既存の映像作品などで使用されているカメラワークの技術を、その用途や効果によって分類してデータベース化したものである。映画などで使用されているカメラワークには、受け手のシーン状況理解を混乱させないような工夫がなされている（参考文献(10)）。これらの情報を利用して、シナリオとストーリーボードの関係付けを自動化、もしくは、ストーリーボードを決定する際の

支援やストーリーボードの妥当性チェック機構として利用する。

これらの入力に基づいて、適切な撮影を行うためのカメラワークプランニングを行う。ここでは、撮影前に行うオフラインプランニングと、撮影時に行うオンラインプランニングとの2段階のプランニングによってシーンの状況に適応的な撮影を実現する。

- オフラインプランニング：与えられた事前知識から、カメラの台数や配置、それらの操作や映像の切り替えのスケジュールを決定する。このプランニングは、撮影前に行い、このプランニング結果に基づいて具体的にカメラを配置する。
- オンラインプランニング：オフラインのプランニング結果に基づいて撮影を行う際には、実際の状況と事前知識との間に幾何学的、時間的なずれが生じる。このずれに対してシーンの状況に適応的にカメラワークを修正しながら撮影を行う。具体的には、撮影の際に与えられたストーリーボードの指定に沿うように、カメラの視線方向を調整し、さらに、カメラの切り替えのタイミングを切り替え前後2台のカメラの画像やカメラパラメータがある条件をなるべく満たすように決定しながら撮影を行う。

このようなシステムの構成を図6に示す。ハードウェア構成としては、シーン内に設置するカメラ群が通信ネットワークで相互に接続され、カメラの切り替えやその前後での適切なカメラ制御を行うために複数のカメラが協調的に動作するものを考える。さらに、システム全体を統括するコントローラ・スイッチャを設置し、事前知識に基づいたストーリー進行を、カメラ群からの映像やカメラパラメータ情報に基づいて判断し、適切な切り替えの実行と、各カメラのスケジューリング管理を行う。

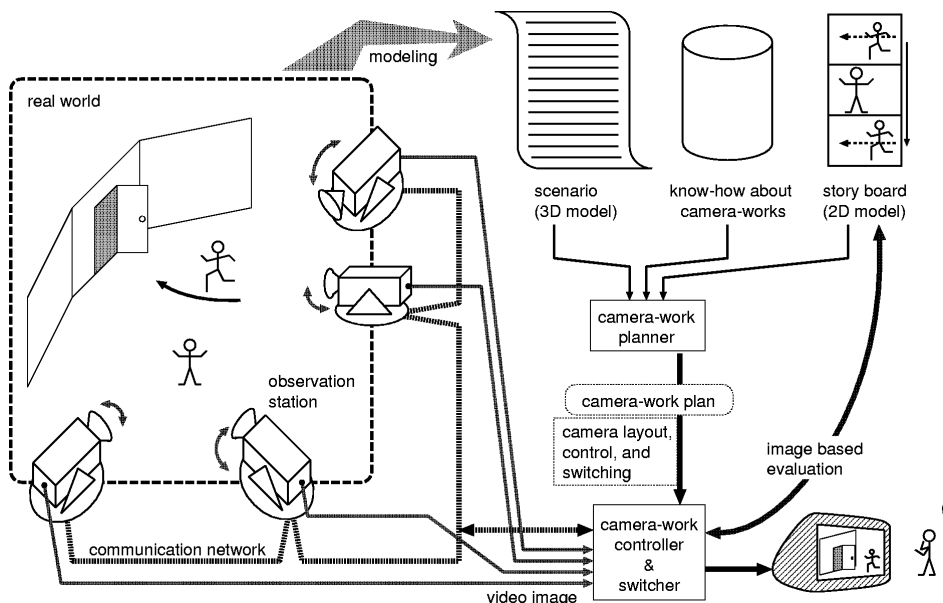


図 6 シナリオに基づく撮影システムの構成

本研究では、このような基本的フレームワークに従って、シーンの映像化を行うシステムを実現することを目指す。われわれのグループでは、カメラの制御に関して、単一カメラ、あるいは、複数のカメラの協調によってシーンの状況を取得する方法についての研究が並行して進められており、本研究ではその成果を利用することを前提とする（参考文献(11)(12)(13)）。

4.2. プロトタイプシステム

ここでは、前述の枠組みでおこなった予備実験について述べる。一つは、静的な2次元シーン状況におけるカメラ配置のプランニング実験、もう一つは、複数（2台）のカメラの切り替えを動的シーン状況に合わせて決定するシミュレーション実験である。

4.2.1. 静的シーンでのカメラ配置

ここでは、2次元平面的な世界を考え、その中で対象（円形で向きを持つ）の配置が静的な場合に、カメラの配置を決定するものである。具体的には、図7に示すような3種類の評価関数、(a)カメラから対象のどの部分が観測できるか、(b)画面の中央に捕らえられているか、(c)画面上での対象の大きさが適当かを考え、それらの総合評価値が最も高くなる位置をカメラ位置とするものである。

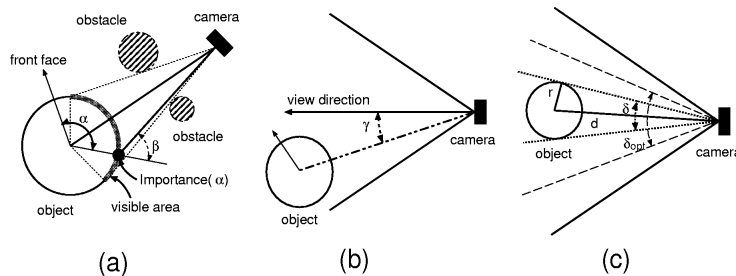


図7 カメラ配置の評価関数

実験の結果を図8に示す。(a)のような2つの対象の配置を1台のカメラで撮影する場合、それぞれの対象を撮影する場合の最適な配置は、(c)(d)のようになり、両方の対象を同時に捉える場合には、(b)のような配置となる。なお、濃淡で示しているのは、シーン内の各店での評価値の分布であり、明るい部分ほど評価値が高い。

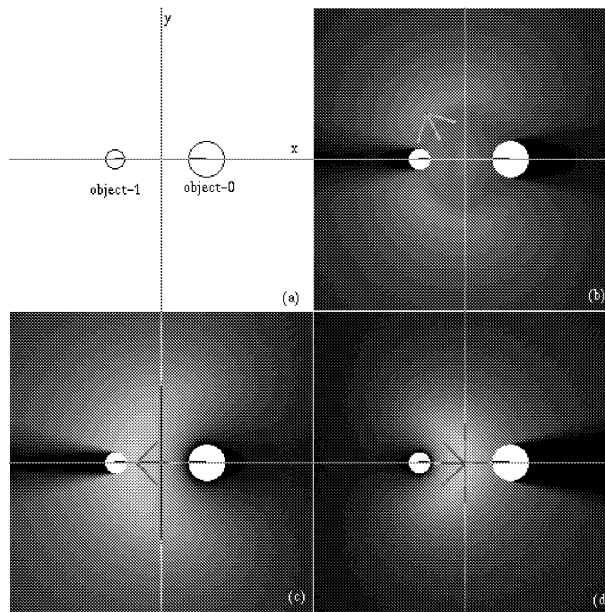


図 8 静的シーンでのカメラ位置算出結果

4.2.2. 動的シーンでの滑らかなカメラ切り替え

次に動的なシーンでのカメラの切り替えタイミングをシーンの状況に合わせて決定する実験について述べる。ここでは、図 9 に示すような状況を考える。撮影には、2 台のカメラを使い、等速直線運動する対象を捉える。それぞれのカメラは対象を中央に捉えながらパン回転で対象を追跡するものとする。この時、カメラの切り替えの前後で等速運動していることを表現するため、切り替えの時点で対象の背景に対する見かけの速度が同じになるように切り替えることを考える。

具体的にシナリオとして与えられた経路に基づいて決めたカメラレイアウトを図 10 に示す。シナリオに基づいたオフラインのプランニングでは、経路の中間地点に到達したときに切り替わりが発生する。この時のカメラのパン速度変化を図 11 左に示す。

次に実際の撮影の状況として、シナリオに記述された経路にずれが付加されたものを考える。図 11 右に示すように、オフラインの計画通りの時刻でカメラを切り替えた場合、切り替えの前後でカメラのパン速度が大きくかわることになり、見かけ上は対象の速度が変化するように観測される。そこで、2 台のカメラのパン速度が同一になる瞬間に切り替えを行うようにすると、切り替えの前後での速度は一致させることができる。この方法で切り替えた場合、対象が等速に動いている様子を正しく伝えることができる映像とすることができ、オンラインのプランニングによる動的な切り替え制御が有効に働いたといえる。

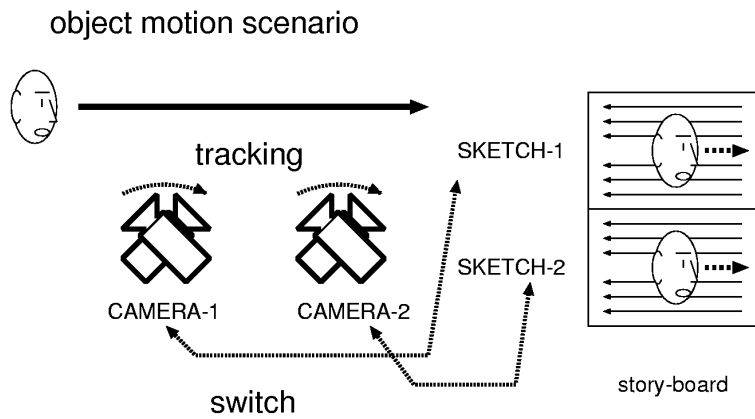


図 9 想定するカメラワーク

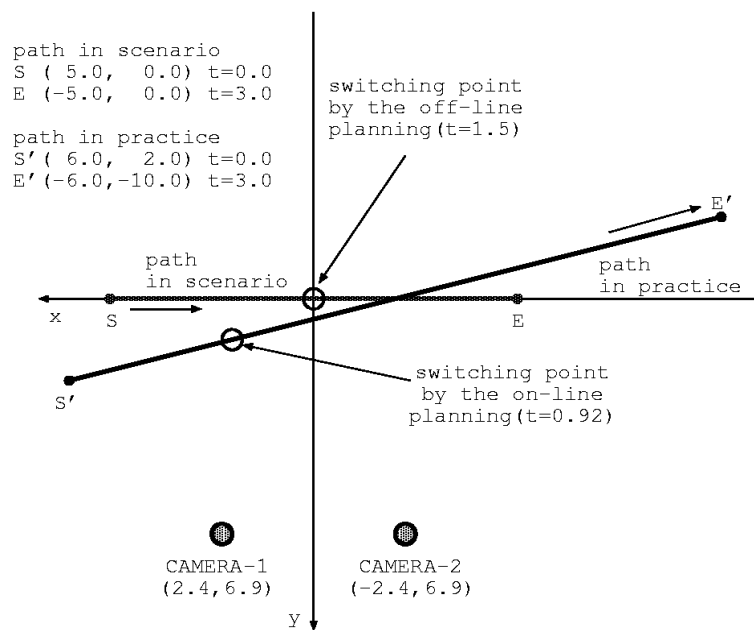


図 10 対象の軌跡とカメラ配置, および切り替えのタイミング

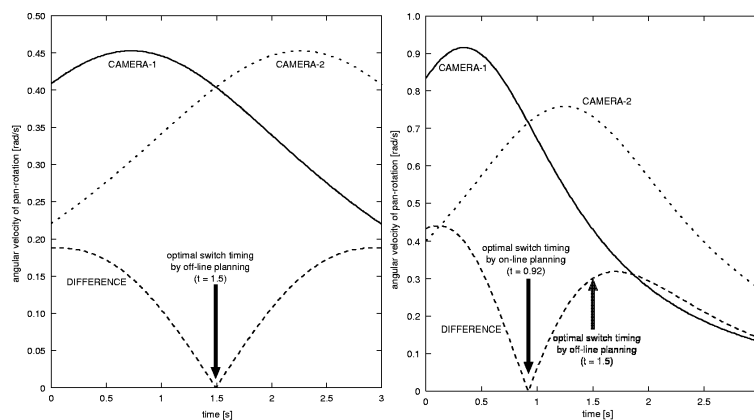


図 11 切り替え前後2台のカメラのパン速度変化 (左: オフライン, 右: オンライン)

4.3. 本節のまとめ

本研究では、スタジオにおける視覚情報メディアの構築のためのシステムの枠組みとその要素の検討、および、予備的な実験について述べた。現在、オフライン、オンライン両者のプランニングについて検討をすすめており、オフラインでは、より動的なシナリオに対してカメラの配置などを決定する方法について、オンラインでは、実際のカメラを制御しながら、シーンの状況を観測し、検出される様々イベントに基づいてカメラワークの制御が行われるシステムを構築しつつある。このなかで、シナリオやストーリーボードの具体的な記述についても、幾何学的だけでなく、時間順序やそのまとまりとしてのイベント群の関係記述を含めて、より汎用性の高い記述形態について検討を行っている（参考文献(14)(15)）。

本研究で扱う視覚情報メディアでは、映像に映っているものそれ自身だけでなく、対象の捉え方や、映像の切り替わりが持つ情報に注目していると言え、従来の映像メディアと物理的なレベルでは同等であるものの、そこに含まれる情報はより大きなものとなっているといえる。つまり、撮影や加工の方法を工夫することによって、より高次元の情報圧縮がなされているとみなすことができ、映像で伝えるべき内容に沿った情報圧縮の一形態として有効である考えられる。

5. おわりに

視覚情報メディアにおいて実世界から得られる映像の情報圧縮を考える場合、情報を発生する実世界とのインタラクションとそれを利用するユーザとのインタラクションを同時に考える必要がある。これが、視覚情報メディアの新しい立場であり、双方のインタラクションを通して情報を圧縮してゆく枠組みが望まれている。

この立場から、本研究では、実世界でのカメラ制御の問題とユーザに提示すべき映像生成の問題を、独立した二つのフィールドで検証することにより、その背後にある原理原則を解明する方向で研究を進めてゆく。同時に構築したシステムが、実際に利用できるようなものにしてゆくことも重要な課題と考えている。

参考文献

- (1) 角所考, 美濃導彦, “コミュニケーションのための視覚情報処理”, 画像電子学会誌, Vol. 28, No. 5, pp.495-502, 1999.
- (2) 美濃導彦他.講義の自動アーカイビングシステムの構築. 第 58 回情処全大 4U-01, 1999
- (3) 石塚健太郎他, 区切り情報を用いたマルチメディア講義情報の自動構造化. 第 58 回情処全大 4U-04, 1999
- (4) 亀田能成他.講師追跡によるカメラ映像の自動切り替え. 第 58 回情処全大 2V-04, 1999.
- (5) 河原達也, 石塚健太郎, 堂下修司:発話検証に基づく音声操作プロジェクタとそ

- れによる講演の自動ハイパーテキスト化.情報処理学会論文誌, Vol.40, No.4, pp.1491--1498, 1999.
- (6) 飯塚重善他, マルチメディア講義情報を用いた WBT 教材作成, 第 58 回情処全大 2X-09, 1999
 - (7) 田村秀行, 池田克夫 編, “知能情報メディア”, 総研出版, 東京, 1995.
 - (8) 池田克夫, 田村秀行, 全 炳東, “知能情報メディア — マルチメディアの進化形 —”, 信学誌, Vol. 79, No. 8, pp. 788-792, 1996.
 - (9) 東海彰吾, 松山隆司, “動的シーンの可視化のためのシナリオに基づく強調的カメラワークの生成”, 第 4 回知能情報メディアシンポジウム論文集, pp.9-16, 1998
 - (10) ダニエルアリホン (著), 岩本憲児, 出口丈人 (訳), “映画の文法”, 紀伊国屋書店, 1980.
 - (11) 松山隆司, “分散協調視覚—視覚・行動・コミュニケーション機能の統合による知能の創発”, 画像の認識・理解シンポジウム (MIRU'98) 論文集 I, pp.343-352, 1998.
 - (12) 松山隆司, 和田俊和, 物部祐亮, “視点固定型パン・チルト・ズームカメラを用いた実時間対象検出・追跡”, 情報処理学会論文誌, Vol.40, No.8, pp.3169-3178, 1999.
 - (13) 松山隆司, 和田俊和, 丸山昌之, “能動視覚エージェントによる移動対象の協調的追跡”, 画像の認識・理解シンポジウム (MIRU'98) 論文集 I, pp.365-370, 1998.
 - (14) 間瀬健二, クラウディオ S.ピンヤネス, アーロン F.ボビック, “インタラクティブシステム設計のための時区間表現によるスクリプト記述法”, 情報処理学会論文誌, Vol.39, No.5, pp.1403-1413, 1998.
 - (15) J.F.Allen, “Towards a General Theory of Action and Time”, Artificial Intelligence, Vol.23, pp.123-154, 1984.