状況理解に基づく遠隔講義のための実時間映像化手法

亀田 能成† 石塚 健太郎‡ 美濃 導彦†

† 京都大学総合情報メディアセンター
‡ 京都大学大学院情報学研究科

概要 ：

複数の物体を状況に応じて撮影しなくてはならない遠隔講義は、分散協調視覚の研究枠組みが最も有効な事例の１つである。

我々は、講義の動的状況に基づく撮影手法を基本とし、特に動物体の観測位置に応じた講義状況を用いることで、複数台の首振りカメラ制御とその映像選択とを行う方法を提案する。また、本研究による映像生成法をカメラシステムとして構築し、音声システム、プレゼンテーションシステムとともに高速ネットワークと接続した遠隔講義システムとしてシステム化する取組みを行い、実際の遠隔講義に本手法を適用した。その結果得られた様々な知見についても考察する。

A Realtime Imaing Method for Distance Learning Based on Dynamic Situation Understanding

Yoshinari Kameda†, Kentaro Ishizuka‡, Michihiko Minoh†

† Center for Information and Multimedia Studies, Kyoto University
‡ Graduate School of Informatics, Kyoto University

Distance learning is a suitable application for cooperative distributed vision in the point that multiple active objects should be shooted according a situation of a lecture.

We propose a realtime imaging method in which we describe control of pan/tilt cameras and selection of images taken at multiple cameras in accordance with dynamic situation detected in the scene. The definition of the dynamic situations is given based on location of the moving objects. We implemented a prototype camera system, and merged it with other audio transmitting system and presentation transmitting system into our distance learning system. The system was used at actual lectures and showed the validity of our approarch. We discuss the points to be improved which were revealed during the experiments.

1　はじめに

高度に発達した情報通信技術を背景として、教育環境を改革しようという機運が高まってきている. 典型的な例は講義を遠隔の受講者に提供する遠隔講義 [1] や講義を記録するアーカイブ化 [2] である. 遠隔講義を行なうことができれば、学習者は地理的な制約なく講義を受講することができる。このためには、リアルタイムに講義の映像・音声を取得する必要があり, 大学での講義全てを人手でまかなうことは経済的に成り立たないので, 計算機による自動撮影技術 (講義映像生成手法) が望まれている.

講義室内で行われる講義をいかに映像化するかという問題は、講義がどのような形態であるかという問題とその映像をどういう形態で利用するかという問題と不可分な関係にある.

講義室内で行われる講義では、講師や生徒、教材など複数の見るべき物体が現われる。物体は講義室内の様々な場所に存在し、その中には教材のように決まった場所に提示されるものもあれば, 講師のように移動するものもある. すなわち、講義室にはマルチメディア情報を発する物体が分散しており、これらからどの部分をどのように撮影するかが本質的な問題である. 講義室のどこにどのような物体が存在するかを規定するものを講義の形態と呼ぶ。

講義室はかなり大きな空間であり, 1 台のカメラで全ての範囲をカバーすることが出来ないだけでなく, カメラそのものを移動すると講義に影響を与えることを考慮すると、固定した位置に設置した首振りカメラを複数台利用して、講義を撮影することが効果的である. 講義の映像化は、複数の物体と複数の首振カメラが分散して配置された環境で、複数のカメラが撮影対象の物体の動きに応じて、いかに協調して講義室の映像を取得するかという分散協調視覚の一つの代表的な問題となる.

一方、取得された映像の利用形態も映像化を決定する要因となる. 情報メディアの発展はさまざまな遠隔講義の形態を可能にしている. この形態は大きく 2 つに分けられる. 一方は 2 つ以上の教室間を通信回線で接続する形態, 他方は一つの講義室と個々人の計算機を通信回線で接続する形態である. これ以外に個人対個人の遠隔講義の形態も考えられるが、明らかに映像化が単純であるので、ここでは除外する.

前者の形態では, 講義室で大画面に投影する高品質映像を伝送する必要があり、通信コストを考えると講義室からは 1 本の映像ストリームを生成することが要求される. ところが、後者の利用形態では、画質がそれほど重要でなく, 講義室から複数本の映像ストリームを提供し, 個々人が好きな映像を選択することも可能となる.

我々は, さまざまな講義形態と利用形態を想定して講義の映像化を行っている. 講義の映像化においては、講義室空間の動的状況の定義と利用が重要であると考えている。これまでも講義の自動撮影を目的とした研究が提案されているが利用する画像特徴量が一般的で講義に対して必ずしも適切でない [3] ことや、固定カメラによる映像ストリームのスイッチング [4] に限定されているなどの問題があった。

そこで、本稿では、講義の動的状況に基づく撮影手法を基本とし、特に動物体の観測位置に応じた講義状況を用いることで、複数台のカメラ制御とその映像選択とを行う方法を提案する。また、本研究による映像生成法をカメラシステムとして構築し、音声システム、プレゼンテーションシステムとともに高速ネットワークと接続した遠隔講義システムとしてシステム化する取組みを行い、実際の遠隔講義に本手法を適用した。その結果得られた様々な知見についても述べる。

以下、本稿では状況に応じた講義の自動撮影手法について、2節で手法の概要を述べ、3節では実装したシステムを用いた実験結果について述べる。

2　講義を対象とする映像生成

本節では、講義を撮影し映像化するという講義映像生成のための自動撮影手法について述べる。まず、講義映像生成における指針について議論する。その後、講義の動的状況について定義し、動物体の状態と動的状況とをどのように関連づけるかについて述べ、最後にその撮影方法及び映像選択方法について説明する。

2.1　講義の形態

講義の形態にはさまざまなものがあるが、本研究で自動撮影の対象とする講義形態を、

(1) 1 人の講師と 1 人以上の生徒によって講義が進行する
(2) 講師は HTML などで作成された電子化教材を事前に用意しそれを利用して講義を行なう
(3) 講師は補助的に黒板を用いる

(4) 講師は一定領域内を自由に移動できる
(5) 生徒は着席したまま講義を受講する
の 4 つの条件を満たすものとする。このような講義を想定した場合、講義室内に存在する動物体は講師と生徒のみであり、その結果講義室内で起こり得る全ての動的状況は講師と生徒の振る舞いによって規定できる。

2.2 講義映像生成の指針

講義の映像を取得し、視聴者に提供する場合、講義内で発生しているイベントに対してどのような映像化を行うかが問題となる。

映像化指針には、以下の 3 種類が考えられる。

1. Generic Rule Based Imaging
   いかなるイベントが発生しても対応できるよう、撮影対象の動きのみから首振カメラ操作を決定する。
2. User Based Imaging
   ユーザ参加型の映像化。ユーザの要求に応じてカメラ操作が決定される。
3. Situation Based Imaging
   イベントに応じてカメラ操作が決定される。ブロードキャスト型と呼ぶ。

我々はこれまで、ある教室で行われる講義に対して遠隔地の学生が其々の PC を用いて受講する形態を想定し、ユーザが映像化指針に直接関与する 2 番目の方式を中心に研究を行ってきた [5][6]。この方針は、映像化される事象に対して異なった見方をする視聴者が複数人いる場合に有効である。

これに対して、遠隔地の学生が一箇所に集まり受講する形態も遠隔講義としてよく実施される。このような場合、2. の映像化指針より 3. の映像化指針が有効である。そこで、後者のような形態の遠隔講義において、ブロードキャスト型映像化における動的状況の定義と、講義映像生成におけるその利用方法を述べる。

なお、講義映像生成に際し、撮影カメラの制御に関する前提として、講義に支障をきたさないようにするため、使用するカメラは固定カメラまたは首振りカメラのみとし、移動カメラは考慮しない。

2.3 動物体の位置に基づく動的状況の定義

講義室内の動的状況は、講義を構成する動物体 (講師と生徒) と静物体 (教卓、スクリーンなど) 同士の関係によって規定される。

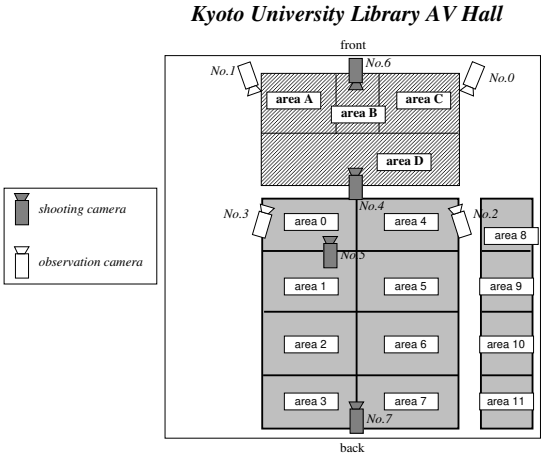

図 1: 京都大学付属図書館 AV ホールにおける領域分割

視覚的には、講義はその講義室内に存在する講義資料やそこに参加する人間によって構成される。すなわち、講義で発生する動的状況はその動的状況を発生させた動物体とその位置する領域、およびその振る舞いとそれが使用する静物体との位置関係によって規定できる。特に、動物体の存在する領域の情報はそれに対する撮影方法に大きな影響をおよぼす。また、別の見方をすれば、講義で発生する動的状況はこれらの要素を観測することで推定することができる。

ここでは、動物体である講師及び生徒それぞれに対して、動的状況の定義を述べる。

動物体である講師と静物体との関係は講師の位置と静物体との距離によって定まる。具体的には、講師の移動可能領域を静物体の位置に応じて複数領域に分割し、一つの領域には 0 または 1 つの静物体しか存在しないようにする。これにより、講師が特定の静物体が含まれる領域にいる場合は、その領域内の静物体との関係が強いと仮定する。

一方、生徒における動的状況については、生徒が座席から離れることを仮定しないので、その位置はどの生徒についても常に固定であり、講義室内の他の静物体を利用することもない。振る舞いについては、質問などで生徒が活動的になっているかどうかが考慮されるべき要因となる。

ここで例として、京都大学付属図書館 AV ホールで行われる講義について定義した動的状況を表 1に挙げる。領域は図 1に示すように分割した。

表 1: 動物体とその存在領域から規定される動的状況の例

動的状況	動物体	動物体の存在領域	領域内の静物体
左スクリーンを使って講義中	講師	area A	左スクリーン
教卓の前で講義中	講師	area B	教卓
右スクリーンを使って講義中	講師	area C	右スクリーン
講師が前縁部で講義中	講師	area D	なし
生徒が活動中	生徒	area 0 - 11	なし

図 1中斜線部が講師の移動可能領域であり、講師の移動可能領域は area A から D までの 4 つに分割されている。各領域について、area A には左スクリーン、area B には教卓、area C には右スクリーンの静物体が含まれている。area D には静物体は何も含まれていない。これにより、講師位置がどの領域に存在するかに応じて動的状況が規定される。

また、図 1中網掛け部が生徒の存在領域であり、座席を area 0 から 11 までの 12 の領域に分割している。これら各領域の生徒について生徒活性度が高ければ、その生徒らが何らかの動作をしているという動的状況である。

以上のような動的状況の定義を元に、観測に基づいて講義室内の動的状況を推定する。

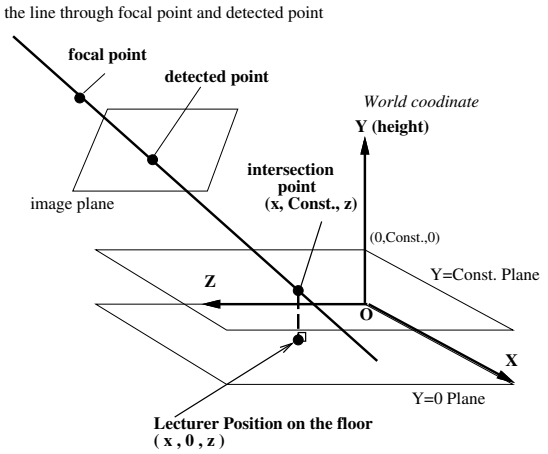

図 2: 単眼カメラによる講師位置の検出

2.4 動的状況の推定

本研究では、動物体の位置を観測することによって、物体間の関係付けを行いそこから動的状況を推定する。

動的状況を推定するための観測特徴量は、以下の 2 点である。

- 講師の位置
  講師の講義室内における位置
- 生徒の活性度
  生徒存在領域における生徒の動きの大きさの度合い

これらの値を表 1及び図 1に当てはめることで現在発生している動的状況の推定を行う。

移動する人間の位置を求める方法としてはステレオ画像処理が考えられるが、本研究では、講師が水平面内の移動しか行わないこと、および複数台のカメラを用いた際の観測範囲が大きいほど講師の移動可能領域を大きく確保できること、の 2 つの観点から、以下のような単眼カメラによる講師位置検出を行っている。図 2のような講義室の世界座標系 XYZ に対し、講師は XZ 平面に並行な面 $(Y = const.)$ でのみ運動すると仮定する。観測カメラの focal point や image plane 等のカメラパラメータを事前に取得しておくことで、image plane 上で観測した講師の位置と focal point とを結ぶ直線と、Y=const の平面との交点 (intersection point) を求めることができる。この交点の X,Z 座標が検出された講師位置となる。なお、画像上の講師の位置は、フレーム間差分により求める。

生徒の活性度としては、まず画像中での生徒の撮影領域内部をフレーム間差分処理し、フレーム間差分が発生した部分の面積を測定する。ついで、その発生した面積を生徒の撮影領域で割ることで正規化し、さらに空間中でその生徒が XZ 平面上で占有する面積で割ることで、生徒の単位床面積あたりの活性度とする。

2.5 複数カメラによる撮影と映像選択

講義室の観測による動的状況の推定結果に基づき、撮影カメラの制御と映像選択を行う。ここでは、動的状況の発生した存在領域に依存して各撮影カメラの制御を行う方法と、その映像の中から最適なものを選択する方法について述べる。

撮影カメラに静止カメラないし首振りカメラを用いるという前提と、動的状況が動物体の存在領域に強く関係付けられていることから、それぞれの動的状況を特定の撮影カメラで撮影する場合、そのパン ・チルト制御値はカメラの空間内の位置によって定めることができる。この際、静物体が利用されている動的状況の撮影については、撮影対象として動物体だけでなく静物体も映像内に入るように制御する。

１つの動的状況に対する撮影方法は、画角の問題を別に考えると、最大でその動的状況が発生している存在領域を撮影できる撮影カメラの数だけ用意できる。なお、本稿では画角については撮影対象が常に一定の大きさに映像化されるように調節するものとする。実際には、１つの動的状況に対応できる撮影カメラを、動的状況の存在領域とカメラ位置とから事前に決定しておく。

つまり、講義映像生成法とは、発生した１つ以上の動的状況に対して、どの撮影カメラを用いて撮影するべきかを選択決定する方法であると言える。以下で、その選択方法について述べる。

まず、複数発生しうる動的状況に対する選択基準を示す。この基準はブロードキャスト型映像化指針においては、映像生成目的に即してトップダウン的に決定されるべきものである。

撮影するべき動的状況の選択基準

D1: ある動的状況に対する撮影を開始してから一定時間が経過していない場合は、その動的状況を継続して撮影し、撮影カメラの変更を行わない。この一定時間のことをここでは一瞥期間と呼ぶ。

D2: 講師を動物体とする動的状況と、生徒を動物体とする動的状況が同時に発生した場合、後者を優先する。

D3: 生徒を動物体とする動的状況が複数発生した場合、もっとも生徒活性度の高い動的状況を優先する。

1. は、人間の映像理解に対する制約に基づく項目である。2. および 3. は講義が映像化対象であることに基づく。特に 2. については、日本で行われる講義の場合、生徒の活性度が低い傾向がみられるため、生徒を優先することとした。

撮影されるべき動的状況が決定された後、撮影カメラの選択を行う。その基準を以下に挙げる。これらの基準はブロードキャスト型映像化指針に従う限り、映像生成目的とは無関係に満たされなければならない。

撮影カメラの選択基準 :動的状況が変化した場合

Sa1: 遷移前の動的状況と遷移後の動的状況に対して、同一の撮影カメラが存在する場合はその撮影カメラからの映像選択を継続する。

Sa2: そうでない場合は、新しい動的状況を撮影するのに最も適した撮影カメラからの映像を選択する。

撮影カメラの選択基準 動的状況が変化しない場合

Sb1: 一定時間以上、同一の撮影カメラで撮影していてかつその動的状況に対して他の撮影カメラが割当てられているとき、撮影カメラからの変更を行う。この一定時間のことをここでは飽和期間と呼ぶ。

3 実験

3.1 講義映像生成の実験

前節で述べた方法に従えば、ブロードキャスト型映像化指針によって講義を撮影する手順は次のようになる。

1. 講義室内を観測し、特徴量を計測する
2. 得られた特徴量から講義室内の動的状況を推定する
3. 推定された状況を元にカメラ操作と映像選択を行う

我々は、本自動撮影手法に基づくカメラシステムを実装し、実際に遠隔講義に用いることで本手法の健全性の検証を行ってきた。

以下、各節においては、本カメラシステムによる実験結果について述べる。

3.1.1 講師位置検出の精度

本稿で提案する講義映像生成法では、講師位置の検出精度が結果に大きな影響を与える。そこで、その精度について、自動撮影カメラシステムを構築した付属図書館 AV ホール中の観測カメラで検証した。なお、以降の数値は特に断らない限り長さはメートル、時間は秒を単位として表す。

AV ホールでの世界座標系における $Y = 0.158$ 平面上での X または Z 座標軸と平行な計 9 つの直線について計測を行い、それぞれについて正確な位置からの誤差を測定した。その結果を表 2に示す。なお、該当する観測カメラの焦点中心の座標は $(5.674, 2.223, 0.212)$ であり、画像中心と焦点を結ぶ直線は $Y = 0.158$ 平面と $(1.893, 0.158, 0.936)$ で交わる。

最大誤差は観測カメラの画像中心を外れるほど大きくなる傾向が認められたが、誤差平均は 30cm 以内に抑えられており、各撮影カメラでの撮影手法の実現には支障のない精度が得られていた。

3.1.2 検出された講師位置の解析

本システムでは、検出された講師位置がどの存在領域に属するかによって撮影されるべき動的状況が決定される。そこで、講師の移動経路に何らかの傾向が認められれば動的状況の予測などに有用であると考え、検出された講師位置の解析を行った。

AV ホールにおける、同一講師による 1999 年 11 月 24 日の講義と 12 月 1 日の講義について、検出された講師位置の解析結果を示す。11 月 24 日の講義においては 4375 点、12 月 1 日の講義においては 1339 点の講師位置を検出した。床面を 0.4m 四方の正方形により分割し、正方形内にどれだけの検出点が含まれたかをもとに検出点の分布を求めた。これを図 3および図 4に示す。水平面は図 1 における教卓付近の座標を示し、各図中の長辺が教室横方向、短辺が教室奥行方向を示す。単位はメートルである。奥行方向 0.0m の位置が教室の奥にあるスクリーンのついた壁の位置を表す。また、縦軸は検出回数を示す。

また、ある講師位置の検出点から次の講師位置の検出点までの時間を計測し、講師位置検出に基づく講師の講義室内での滞在時間分布を図 5および図 6に示す。なお、図中の縦軸は秒数を示す。

これらの解析結果から、講義ごとに滞在位置の嗜好があることがわかる。このことから、その講師の好む滞在位置に撮影カメラを予め向けておくことや、嗜好する位置以外の移動に対し、撮影カメラを向けないでおき映像選択することを避けること、ひいては各動的状況に対する撮影カメラの割り付けの最適化など、個々の講師に適応する処理が可能であると考えられる。

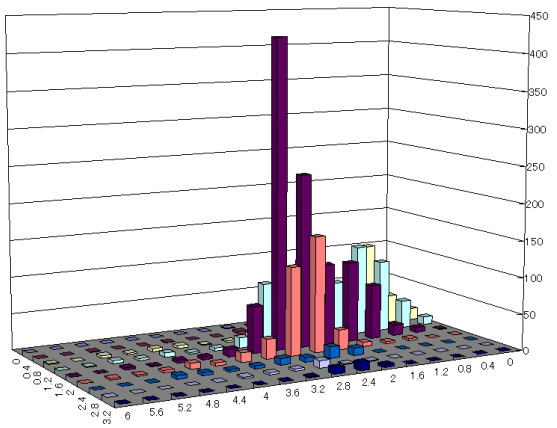

図 3: 11 月 24 日の講義において検出された講師位置点数分布

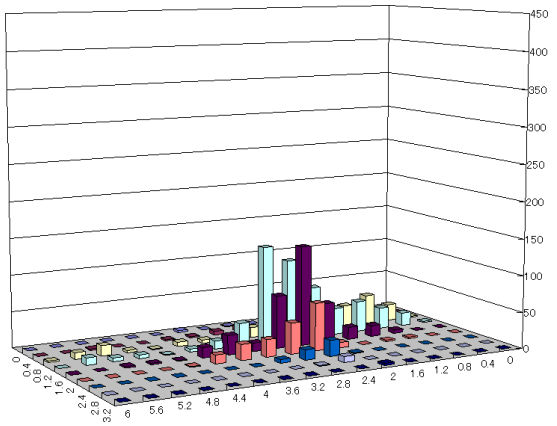

図 4: 12 月 1 日の講義において検出された講師位置点数分布

3.2 遠隔講義への適用

遠隔講義においては、本手法に基づくカメラシステムだけではなく、音声や教材を呈示・伝送するシステムが必要である。そこで、我々は、それらのシステムを映像・教材伝送を行う高速ネットワークと結合し、遠隔講義システムを構築した。

図 7にその概要を示す。この遠隔講義システムは、本手法に基づく Video Imaing だけではなく、音声を取得する Audio Capturing、教材の呈示・指示を電子的に実現する Teaching Material Paging/Pointing、及び手書文字・図形を電子黒板で伝送する Hand writing の 4 部分から構成される。

特に教材および手書文字・図形については、そ

表 2: 講師位置検出の精度 (単位 :m)

直線	最大誤差	最小誤差	平均誤差	誤差の分散
x=0.32	0.47	0.00	0.19	0.02
x=0.82	0.34	0.00	0.16	0.01
x=1.32	0.42	0.00	0.19	0.01
x=1.82	0.54	0.00	0.24	0.03
z=0.33	0.34	0.00	0.14	0.01
z=0.83	0.27	0.00	0.11	0.01
z=1.33	0.26	0.00	0.12	0.01
z=1.83	0.39	0.00	0.19	0.01
z=2.33	0.54	0.02	0.30	0.02

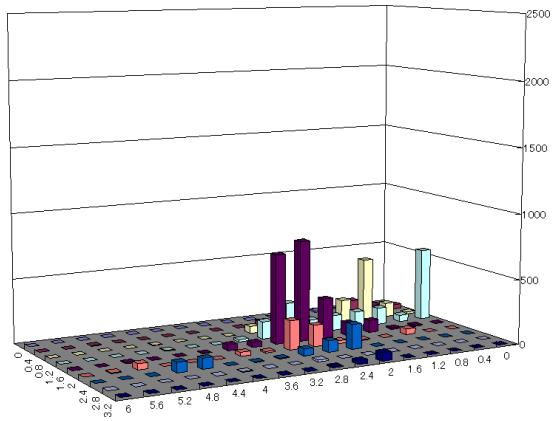

図 5: 11 月 24 日の講義において検出された講師滞在時間の分布

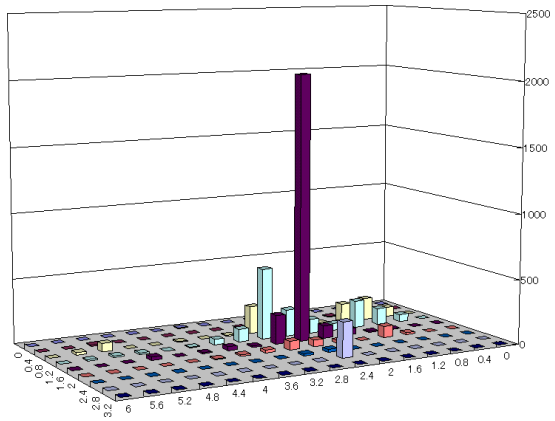

図 6: 12 月 1 日の講義において検出された講師滞在時間の分布

の内容を判別可能なほど高精細にビデオカメラで撮影して伝送することは、ビデオカメラの解像度および必要となる通信帯域の観点からみて非経済的である。そこで、本遠隔講義システムではそれらの情報は HTML に基づくプレゼンテーションシステムや電子黒板によって別途伝送する。これにより、講義映像生成においては、講師がどの教材を用いて講義を行ってどのような身振りをしているかを撮影することが目標となる。図 8に遠隔講義システムのスクリーン設計を示す。遠隔地には 2 面スクリーンが用意され、片面に本カメラシステムによる講義映像が表示され、もう 1 面には講義資料が呈示される。なお、どちらからの講義にも対応できるよう、カメラシステムを含む遠隔講義システムは双方の講義室に設置した。

本遠隔講義システムを実際にアメリカ合衆国の

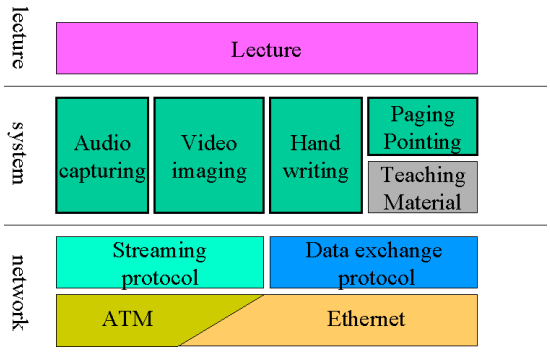

図 7: 遠隔講義システム

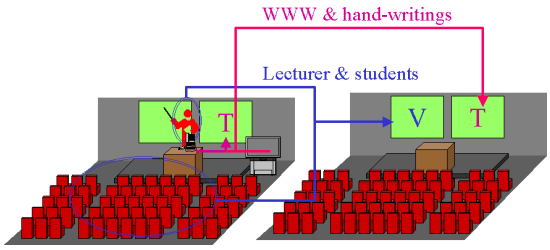

図 8: 遠隔講義システムのスクリーン設計

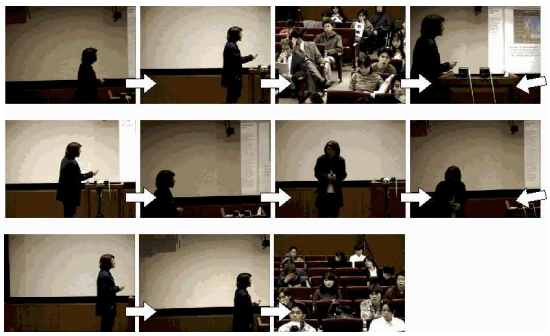

図 9: 遠隔講義の実施例

大学との遠隔講義に 1 学期間利用した。現在は受講者によるアンケート結果を集計中であるが、聞取り調査では概ね良い評価を得ている。図 9に本システムを用いて実際に行われた遠隔授業の一例を示す。

遠隔講義に対して実際にカメラシステムを適用してきた結果、次のような知見を得た。

1: 映像生成は講義空間で起こるイベントに依存するが，画像処理による観測だけでは決定できない．特に講義では講師の発話内の指示語によって撮影されるべき対象が指定されることがあった。
2: 実利用においては映像を撮りつづける必要があり，動的状況として定義されていないイベントにも対応しなければならない．
3: カメラを頻繁に動かすと、例えそれが撮影対象を追跡するためであっても、生成された映像の評価が落ちる傾向にある．
4: 遠隔講義システムにおいては、映像を生成するカメラシステムはプレゼンテーションシステムと協調しなければならない．

これらの知見に基づいて、これまでに開発してきた映像生成法を改良し、実利用出来るシステムを構築していくことが今後の 1 つの課題である．

4 おわりに

本研究では，多様な情報提供が可能な遠隔講義システムの実現を目指して，ブロードキャスト型映像化指針に基づく映像生成法を提案した。この方法では、室内の動物体の存在領域に応じた講義室空間の動的状況に基づいて、複数の撮影カメラを適切に制御し最適な映像を選択することを実現している。

実験結果の解析から講師により滞在位置に嗜好性があることが判明したので、これを利用した講義状況の先見処理が行えるかどうか考察を進めていく予定である。また、講師の音声など他の特徴量を利用することでよりきめ細かい動的状況を検出できるかどうかも検討する。

また、本提案手法を実際にカメラシステムとして構築し、遠隔講義への組み込みを行った。このシステムを利用して実際に遠隔講義を行うことで、その有用性を確認した。遠隔講義システムとして、これまでの分散協調視覚の枠組みを音声や教材との協調というマルチメディア処理へ拡張することも今後の課題である。

謝辞　本研究の一部は科学研究費基盤研究 B-09558034, 学術振興会未来開拓プロジェクト「分散協調視覚」(JSPS-RFTF 96P00501) の支援を受けている.

参考文献

[1] 宇井修, 中山実, 清水康敬, “衛星通信講座における講義形態と学習者評価の関係,” 信学論, J80-D-II, No.4, pp.892-899, 1997.

[2] D.Aboud, G., Atkeson, C., Feinstein, A., Goolam-abbas, Y., Hmelo, C., Register, S., Sawhney, N. N., Tani, M., “Classroom 2000: Enhancing Classroom Interaction and Review,” GVU Technical Report GIT-GVU-96-21, 1996.

[3] 大西正輝, 泉正夫, 福永邦雄, “情報発生量の分布に基づく遠隔講義撮影の自動化,” 信学論, J82-D-II, No.10, pp.1590-1597, 1999.

[4] 大野直樹, 先山卓朗, 椋木雅之, 池田克夫, “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論文集, pp. 31-38 1999.

[5] Y. Kameda, H. Miyazaki, and M. Minoh, “A Live Video Imaging for Multiple Users,” Proc. of International Conference on Multimedia Computing and Systems (ICMCS’99), Vol.2, pp.897-902, 1999.

[6] 宮崎英明, 亀田 能成, 美濃 導彦, “複数のカメラを用いた複数ユーザに対する講義の実時間映像化法,” 信学論, J82-D-II, No.10, pp.1598-1605, 1999.