

議事推移モデルを用いた知的遠隔会議支援に関する検討

炭野 重雄 三村 到 亀山 達也 鈴木 敏明 伊達 哲

{ sumino, mimura, kameyama, toshiaki, akira }@crl.hitachi.co.jp

(株)日立製作所 中央研究所

議事の推移をモデル化し、そのモデルに基づく知的遠隔会議支援に関して基礎検討を行なった。モデルの導入による効果は以下の二つである。(1) 議事の推移に応じてカメラを制御して遠隔会議映像を獲得し、同映像を推移状況に適した表示レイアウトで表示することによって、遠隔会議における視野の限定を解消する。(2) 議事の推移を簡単なインタフェースで入力可能とし、議事の内容をノード、議事の推移をアーク、議事の推移のタイプをアークのラベルとして視覚表示することにより、会議進行中は一見して議事の推移を把握可能となるので進行支援に、情報を格納しておけば会議終了後はマルチメディア議事録として利用可能となる。

Consideration on Teleconference Support Method using Issue-Transition Model

Shigeo Sumino Itaru Mimura Tatsuya Kameyama Toshiaki Suzuki Akira Date

Central Research Laboratory, Hitachi, Ltd.

We propose an issue transition model which formalized conference process. We apply the model to a video teleconference. Advantages by introducing the model are as follows: (1) cancellation of restricted view on teleconferencing by adaptive camera shooting condition control with an appropriate camera-work and by display window layout control with an appropriate layout according to current issue, (2) progress support on conference and agenda-making support after conference by showing issue's contents as nodes, issue transition as arcs, and transition type as arc labels.

1 はじめに

インターネットを代表とするネットワーク環境が整備されることにより、一般大衆までもがマルチメディアの恩恵にあずかろうとしている。われわれは、2005年までに予定されている主要都市間のATM回線の整備を見越し、ビジネスユースのサービスを検討している。具体的には、大画面表示装置を用いた臨場感映像通信システムを検討しており [1]、主に以下の3技術について研究している。

- 大画面に表示するに耐えうる高精細映像の伝送 [2]
- 大画面に表示された映像に合致した音の定位 [3,4]
- 対向サイトの会議参加者との意思の疎通を促進する会議支援ソフトウェア

現在、WSをサーバ、Pen PCをクライアントとしたクライアント・サーバ構成のWS/Pen PC連携遠隔会議支援システムを構築中である。同システムの概観を図1-1に示す。

本報告では、遠隔会議で問題となる視線の不一致や視野の限定の解消、また、議論を活性化させ会議の進行を促進するような知的な会議支援の構想について述べる。具体的には、以下の項目について報告する。

- 議事の推移に応じてカメラを制御し、遠隔会議映像を獲得するAIカメラワークと、獲得した会議映像を議事の推移に応じた表示レイアウトで表示する動的表示レイアウト。
- 議事の推移を簡単なインタフェースで入力可能とし、アイコンを用いて視覚表示することにより、会議進行中は一見して議事の推移を把握可能となるので進行支援に、会議終了後は議事録として利用可能とする議事推移入力インタフェース。

2 議事推移モデルによる遠隔会議支援

遠隔会議で問題となる視線の不一致や視野の限定を解消するため、いかに会議の局面に応じた自然な映像を対

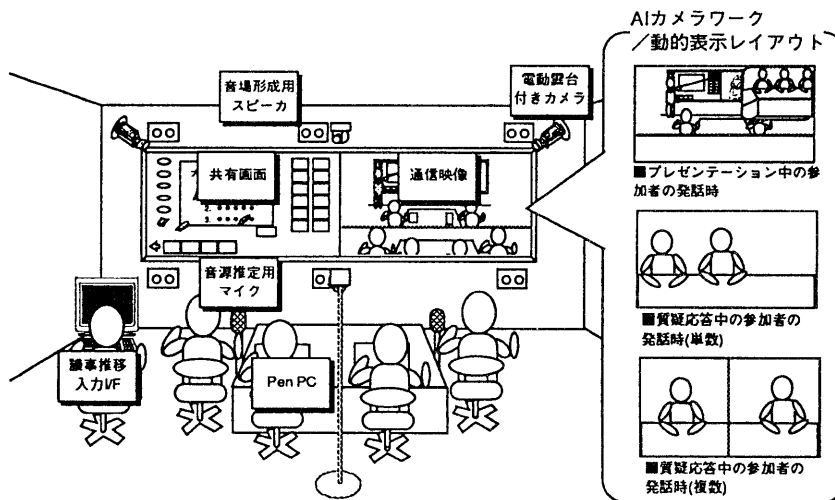


図 1-1: WS/Pen PC 連携遠隔会議支援システム概観

向サイトに伝達するか考察する。

2.1 議事推移モデルの提案

会議の大まかな進行状況は、オープニング、プレゼンテーション、質疑応答、フリーディスカッション、エンディングなど、ほぼ定型化できる。これらの局面において、例えば、対向サイトでプレゼンテーションが行なわれている場合、プレゼンタを含んだ全景の映像を表示する方が、また、フリーディスカッションが行なわれている場合、発話中の会議参加者の近景の映像を表示した方が、より会議の雰囲気や伝達可能と考える。

このような会議の進行状況に応じて映像を獲得するため、更には獲得した会議映像をその状況に適した表示レイアウトで表示するため、会議の進行状況をモデル化した議事推移モデルを提案・導入する。

しかし、現状のコンピュータ・サイエンスで会議の進行状況を自動認識するには困難が伴う。何故なら、会議参加者の発話内容を理解し、更に抽象度の高いレベルで解釈しなければ会議の進行状況を把握することは不可能だからである。よって、本報告において会議の進行状況は、書記もしくは会議参加者（ユーザ）のいずれかに入力してもらうこととする。但し、モデルを参照することによってある程度操作を限定することができるので、ユーザの負担を軽減するようなインタフェースで会議の進行状況を入力してもらうこととする。

このような書記が会議の進行状況を入力する際、会議映像、会議の音声情報なども同時に記憶しておけば、マルチメディア議事録を作成できる。また、記憶した会議

映像、音声情報のブラウザを作成すれば、会議の再現をすることも難しいことではないと考える。更には、その会議映像や音声情報などの会議の進行状況をアイコンを用いて視覚的に表示することによって、現在までの会議の流れを一見して把握可能となるので、オンラインで議事の進行支援にも利用できると考える。

以上、会議の進行状況をモデル化した議事推移モデルを提案・導入することにより、議事の推移状況に応じて対向サイトの映像を獲得し（以降、AIカメラワークと呼ぶ）、動的な表示レイアウトで会議映像を表示する（以降、動的表示レイアウトと呼ぶ）ことによって、視野の限定の解消を図る。更に、議事の推移をアイコンを用いて視覚表示することにより、会議進行中は同推移状態を参照することで進行支援に、会議終了後は議事録として利用可能となる。

2.2 議事推移モデルによる機器制御

本システムにおける議事推移モデルの位置付けを図 2-1 の機器制御概念図に示す。

各々の構成要素の役割は以下の通りである。

- (a) 遠隔会議支援ソフト：Pen PC をクライアント、WS をサーバとしたクライアント・サーバ構成の遠隔会議支援ソフト。図 1-1 に示した対向サイトの会議参加者との意思の疎通を図るための共有画面に係わり、会議資料の表示／書き込み／保存を可能とする。
- (b) 表示レイアウト制御装置：図 1-1 に示した対向サイ

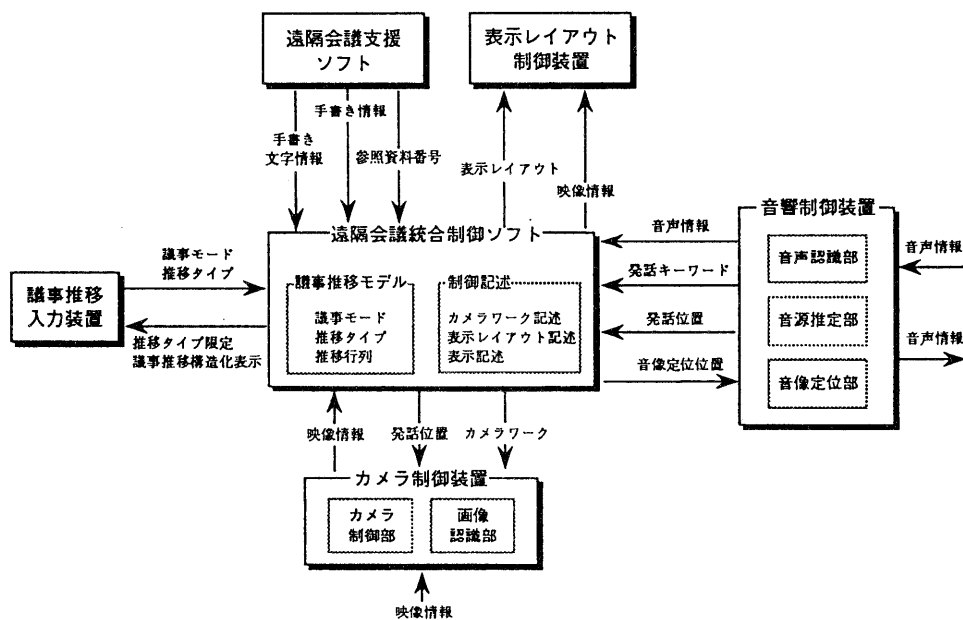


図 2-1: 議事推移モデルによる機器制御概念図

- の様子を示す通信映像に係わり、遠隔会議統合制御ソフトからの表示レイアウトに従って自サイト・対向サイトの映像を合成し、通信映像を生成する。
- (c) 音響制御装置： 会議参加者の発話を認識し発話キーワードを抽出する音声認識部、発話位置を推定する音源推定部、大画面に表示した通信映像に合致した箇所に音を定位させる音像定位部から構成される。
- (d) カメラ制御装置： 遠隔会議統合制御ソフトから得られたカメラワークに従ってカメラを制御するカメラ制御部、また、カメラで獲得した画像を認識してカメラ制御部にフィードバックさせる画像認識部から構成される。
- (e) 議事推移入力装置： 書記もしくは会議参加者のいずれかが議事の推移状況を入力するための入力装置。議事の推移状況がアイコン表示される。会議進行中は、会議参加者が同アイコン表示を参照することで現在の会議の進行具合を確認できる。会議終了後は、同アイコン表示および映像情報・音声情報を保存しておくことにより、議事録として利用できる。
- (f) 遠隔会議統合制御ソフト： 映像、音声、コンピュータデータを扱う本システム全体の制御ソフト。議事モード、推移タイプ、推移行列から成る議事推移モデルとカメラワーク記述、表示レイアウト記述、表示記述から成る制御記述とを内蔵する。

音響制御装置から得た会議参加者の発話位置と議事推移入力装置から得た議事モード・推移タイプに従って、最適なカメラワークをカメラ制御装置に送信し、通信映像を生成する。表示レイアウトの変更とともない、音響制御装置に音像の定位位置を送信し、通信映像と音の定位する位置を変更する。

同時に、カメラ制御装置から得た映像情報と音響制御装置から得た音声情報・発話キーワード、遠隔会議支援ソフトから得た参照資料番号・手書き文字情報・手書き情報をアイコン化して議事推移入力装置上に表示する。

2.3 議事推移モデル

会議の大まかな進行状況は、オープニング、プレゼンテーション、質疑応答、フリーディスカッション、エンディングなどは定型化できる。本報告ではこのような会議進行における概略の流れと個々の議論の因果関係、すなわち、どの提案に対してどのような賛成/反対/コメントがあり質疑応答はどのように行なわれたかという詳細の流れの2つのレベルで議事の推移をモデル化する。

同項目は、企画会議、ブレインストーミング、伝達のみ会議など会議の性質によっても変化すると考えられるので、予め登録可能としておく。

このように議事の推移をモデル化してシステムに内蔵することにより、会議の状況に応じた処理、例えば、状況に適したカメラワークで対向サイトの会議映像を獲得し状況に適した表示レイアウトで同映像を表示して、遠隔会議におけるヒューマンインタフェースを向上させる。

上記、議事モードと推移タイプを登録すると同時に、各議事モードに対応する推移タイプの行列、すなわち、2つの推移タイプの推移が成立するか否かの関係を記述したモード別推移行列も予め用意しておく。

このモード別推移行列は、議事推移入力インタフェースに用いられ、前回入力した推移タイプを参照することによって書記が推移タイプを入力する際の選択肢を限定する役割を持つ。このようにモデルを参照することによって選択肢を限定し、書記の負担の減少を図る。また、このモード別推移行列を定義しておくことにより、推移タイプの推移に応じたより詳細な処理が可能となる。

3 議事推移モデルに基づいたカメラ・表示レイアウト制御

3.1 制御記述

AI カメラワークおよび動的表示レイアウトを行なうため、予め以下の3種類の記述ファイルを用意しておく。

■ カメラワーク記述

以下に示すようなカメラワークの設定項目の列挙ファイルを用意しておくことで、ユーザ、すなわち、システムの導入担当者、システムのメンテナンス担当者、あるいは会議参加者のいずれかが、選択肢を選択するだけで高度なカメラワークが記述可能となる。項目の列挙に際し文献 [5] を参考にした。

- (a) 被写体サイズ： 会議映像における被写体である会議参加者の大きさを設定する。選択肢： full shot, knee shot, waist shot, bust shot, close shot, extreme close shot。
- (b) 被写体人数： 会議映像における被写体である会議参加者の人数を設定する。選択肢： one shot, two shot, three shot, group shot。
- (c) 撮影地点： 自サイトと対向サイトのどちらの会議参加者を撮影するかを設定する。選択肢： this site, counter site。
- (d) 撮影方向： 各サイトにおいて前方、後方のどの方向から会議参加者を撮影するかを設定する。選択肢： front, back。
- (e) カメラ動作： 会議参加者を撮影するときのカメラの動きを設定する。選択肢： fix, panning left, panning right, zooming in, zooming out。

- (f) カメラアングル： 会議参加者を撮影するとき、カメラと水平面とがなす角度を設定する。選択肢： standing position, low angle, high angle, birds eye view。
- (g) 制限時間： 連続して撮影可能なカメラワークの制限時間を設定する。選択肢： unlimited, 10 second, 30 second, 120 second。

■ 表示レイアウト記述

通信映像を大画面表示装置に表示する際に選択可能な表示レイアウトを記述しておく。レイアウトタイプの識別子、幅、高さ、x、y 座標の最小値を格納しておく。複数のレイアウトタイプを組み合わせることで通信映像を生成する。

■ 表示記述

対向サイトでプレゼンテーションが行なわれている場合、プレゼンタを含んだ全景の映像を画面全体に表示する方が、フリーディスカッションが行なわれている場合、発話中の参加者の近景の映像を表示した方が、また、そのディスカッションが複数の会議参加者の対話でなされている時は、同参加者の映像を同時に表示した方が、より会議の雰囲気や伝達可能と考える。

このような議事の推移に応じた表示レイアウトとそのレイアウトに対応するカメラワークを可能とするため、2.3節で定義した各議事モード毎に表示レイアウトとカメラワークとの対応関係を格納しておく。この議事モード、表示レイアウト、カメラワークの対応関係の記述を表示記述と呼ぶ。

#opening
A [2310200];
#presentation-1
A [0011023];
A [0011023] + C1 [3??0001];
A [0011023] + C1 [3??0001] + C2 [3??0001];
#presentation-2
U [0011022] + D [?301022];
U [0011022] + D [?301022] + C3 [3??0002];
#question&answer
A [3??0003];
L [3??0002] + R [3??0002];
#discussion
U [2310003] + D [2300003];
#ending
U [2310003] + D [2300003];

図 3-1: 表示記述

図 3-1 に表示記述の記述例を示す。図 3-1 の 1 行目が議事モードがオープニングモードであることを、2 行目が同議事モードにおけるレイアウトタイプが“A”であり、

そのレイアウトに表示する映像を撮影する際のカメラワークが“2310200”、すなわち、被写体サイズが waist shot (2)、被写体人数が group shot (3)、撮影地点が counter site (1)、撮影方向が front (0)、カメラ動作が panning right (2)、カメラアングルが standing position (0)、制限時間が unlimited (0) であることを表している。

3.2 AI カメラワーク

議事モードに従ったカメラワークで会議映像を獲得する AI カメラワークの具体的な実現方法を以下に示す。

1. 議事推移入力装置から入力された議事モードに従って表示記述からカメラワークを得る。
2. カメラ制御装置は、音響制御装置から得られた発話位置情報とカメラの配置位置から複数台のカメラのうち撮影に最適なカメラを1台選定して発話者に向け、発話位置情報による概略のカメラ制御を行なう。
3. 撮影した映像がカメラワークの仕様を満足しているか否か判定するため、減点法で評価点を算出する。
4. 映像の評価点があらかじめ定めておいた閾値 T_1, T_2 ($T_1 < T_2$) に対して、 T_1 以上 T_2 未満の場合、カメラの電動雲台を制御して画角を変更し、ステップ3を繰り返す(画像認識による詳細なカメラ制御)。評価点が T_1 未満の場合、別のカメラを選択し、ステップ3から再度処理を行なう。評価点が T_2 以上の場合、その映像をカメラワークの仕様を満たしている映像として、表示レイアウト制御装置に送る。

評価点を算出する方法としては、例えば、人物の顔を認識するアルゴリズム [6] が利用できる。同アルゴリズムを用いれば、撮影した映像中の顔の領域を求めることができるので、映像中の顔領域の割合、映像中の顔領域の位置、映像中の顔領域の個数に応じて評価点を算出可能である。

上記した実現方法は、自サイトと対向サイト各々に電動雲台付きカメラが複数台、画像処理装置が1台設置されているという前提で記述したが、カメラ1台に付き画像処理装置1台が附属していれば、ステップ2以降複数台のカメラを同時に制御可能となる。

3.3 動的表示レイアウト

3.2節の AI カメラワークで獲得した会議映像を、現在の会議の状況すなわち議事モードに即した表示レイアウトで表示する動的表示レイアウトは、以下のようにして実現する。

1. 議事推移入力装置から入力された議事モードに従って表示記述から表示レイアウトを得る。

2. カメラ制御装置から得られた会議映像を処理して同表示レイアウトで大画面表示装置に表示する。
3. 音声制御装置から別の会議参加者の発話情報を得た場合、同参加者の映像を表示レイアウトに従って合成して表示する。

対向サイトのプレゼンテーションの様子を表示しているときに他の会議参加者から質問があった場合、同参加者の近景を撮影した子画面を発生し重畳表示する。このように子画面を用いることで着目しているプレゼンタの映像を確保しつつ参加者の質問の様子を把握できるようにして、視野の限定を補完する。また、質疑応答が活発になり二人の参加者の対話となった場合、同参加者の映像を同時に表示して会議の雰囲気伝達する。

4 議事推移モデルに基づいた議事進行および議事録作成支援

図4-1に議事推移入力装置に表示される議事推移入力インタフェースのイメージを示す。

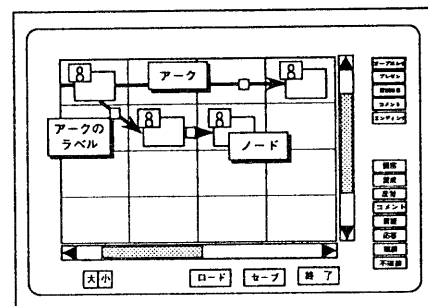


図 4-1: 議事推移入力インタフェースのイメージ

以下の情報を各ノード毎に保持し、いつ、誰が、何を発言して、何を行なったかという議事の内容をノードとして、誰が誰に対して発言したかという議事の推移をノード間を接続するアークとして、発言の因果関係すなわち推移タイプをアークのラベルとして視覚表示することによって、会議参加者に議事の推移を一見して把握可能とさせる。

- 開催会議名
- ノード番号
- ノード座標情報
- 発言者名
- 発言者映像番号
- 発言キーワード
- 発言者音声番号
- 参照資料番号
- 資料書き込み番号
- 発言時刻
- 議事モード
- 接続ノード番号
- 接続推移タイプ

また、議事の推移を入力する書記、もしくは、会議参

加者のいずれかの入力負担を減少させるため、一つ前に入力した推移タイプと2.3節で述べた推移行列とを参照することによって入力の手間を減少させる。具体的には、入力させる可能性のない推移タイプはダウンライトするなど表示に工夫して選択肢を限定し、書記の負担を減少させる。

会議進行中、議事推移入力インタフェースを会議参加者に提示すれば現在までの会議の進行状態、例えば議論不足の箇所や保留となった議論などを一見して把握できる。よって、議事推移入力インタフェースは会議進行中、議事の進行支援に用いることができる。

書記が議事推移入力インタフェースを操作して入力した議事の推移と、同インタフェースで用いた会議映像、音声情報などの各種会議に関する情報を同情報のリンクと共に記憶しておけば、会議終了後はマルチメディア議事録として利用できる。また、議事推移入力インタフェースのハードコピーだけでも、どのような経緯で議論が行なわれたかが一見して把握できるので、議事録の作成支援に用いることができる。

5 関連研究との比較

問題解決に至る過程や議論の内容を構造的に整理する技術として Design Rationale 技術（以降 DR と略記する）が注目されている [7]。DR の代表的な記述モデルとして IBIS (Issue Based Information System) モデルがある。文献 [7] に IBIS モデルとその改良版である gIBIS モデル、rIBIS などのサーベイが詳しく記載されている。

本報告で述べた議事推移モデルもこの DR の範疇に入る。但し、IBIS などが議論が行なわれた後の解析手法として用いられているのに対し、本モデルはオンラインの会議進行中に参照され、議事の推移が形成される点が異なる。また、DR 技術を遠隔会議支援に用い、議事の進行状況を視覚化すると共に会議映像の獲得・表示にまで利用している点が異なる。

本報告で述べた AI カメラワークと動的表示レイアウトは、テレビの討論番組でのカメラマン、スイッチャ、ディレクタの関係に例えることができる。討論番組では、(1) ディレクタが複数人のカメラマンに指令を出し、番組出席者の映像を撮影させ、(2) ディレクタがスイッチャに指令し、カメラマンの撮影した映像のうち最適な映像に切替え放映する。(3) 局面によっては、二人の番組出席者が対面するような映像を合成して放映する。

よって、本システムにおける遠隔会議統合制御ソフトがディレクタ、および、スイッチャ、カメラ制御装置がカメラマンに相当し、AI カメラワークが(1)、(2)の処理を、動的表示レイアウトが(3)の処理を行なっていることになる。これをエージェント指向で言い替えば、

カメラ制御装置がカメラマン・エージェント、遠隔会議統合制御ソフトがディレクタ・エージェントというとならえ方もできる。

6 むすび

遠隔会議で問題となる視線の不一致や視野の限定・固定を解消し、また、知的な会議支援を行なうため、議事の推移をモデル化した議事推移モデルを提案した。また、このモデルに基づいた遠隔会議支援システムに関して、以下の項目を検討した。

- 書記が入力した議事の推移に従って最適なカメラワーク／表示レイアウトを獲得し、それに従ってカメラ／表示レイアウトを制御し視野の限定・固定を解消する AI カメラワーク／動的表示レイアウトについて述べた。
- 議事の内容をノード、議事の推移をノード間を接続するアーク、発言の因果関係をアークのラベルとして視覚表示することによって、会議参加者に議事の推移を一見して把握可能とさせる議事推移入力インタフェースについて述べた。

今後の課題としては、カメラの切替え規則に従った複数カメラの協調制御、議事モード／推移タイプの項目の検証などが挙げられる。

参考文献

- [1] 三村 到 etc. 大画面高精細映像の臨場感通信への適用. 1994 年電子情報通信学会春季大会 B-912, p. 3407, 1994.
- [2] 伊達 哲 etc. 大画面高精細映像通信システム用画像符号化伝送装置の開発. 1994 年テレビジョン学会年次大会 13-1, pp. 207-208, 1994.
- [3] 鈴木 敏明 etc. 臨場感映像通信システムにおける音響処理方法の一検討. 1994 年電子情報通信学会春季大会 B-914, p. 3409, 1994.
- [4] 亀山 達也 etc. 臨場感映像通信システムにおける音響処理装置の検討. 1994 年電子情報通信学会春季大会 B-916, p. 3411, 1994.
- [5] 山岸達児. 映画・ビデオ演出の基礎技法. 教育出版センター, 1992.
- [6] Guangzheng Yang and Thomas S. Huang. Human Face Detection in a Scene. *Proceedings 1993 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 453-458, 1993.
- [7] 桑名栄二. ソフトウェア履歴利用の研究動向. 電子情報通信学会誌, Vol. 77, No. 5, pp. 531-538, 1994.