

会話場に関係知識を提示することによる知識循環

Knowledge Circulation by Presenting Related Knowledge at Conversational Field

中沢 拓磨 角 康之 西田 豊明
Takuma Nakazawa Yasuyuki Sumi Toyoaki Nishida

京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University

The purpose of this research is to boost the knowledge circulation based on conversations which are limited to the meeting and presentation domain. We propose you to cause the knowledge circulation in short span by acquiring conversational video as knowledge and showing the conversational video obtained before at the meeting. The approach is using video of environmental camera array segmented by operation of presentation slides and handheld camera type devices.

1. はじめに

日々行われる会話シーンで起きた出来事やその会話内容について、他の人にそのまま伝えたいと思うことや、自分で後から見返したいと思うことはよくある。なぜならば会話中に起きた些細な行動や、細かな言い回しの中にも、重要な情報が含まれており、それらをうまく伝えることは知識をうまく掘り起こすことにつながるからである。このような知識の形をうまくつくりだしてゆくためには、会話をそのままとらえることができるキャプチャシステムが必要であり、その情報を探するための閲覧インタフェースが必要である。さらに潜在的には重要であるにもかかわらず、知識の獲得行動が起こらないために、零れ落ちていってしまった知識を伝達可能にするため、受動的に利用可能な知識の提示システムが必要である。

本研究ではプレゼンテーション式ミーティングの外側と内側の双方を捉えることによって、より人間の感性に近い再利用性のある会話を取り出すことが目的である。ミーティングの外側としては環境カメラアレイによる映像として捉え、ミーティングの内側としては参加者の注目の集まり方や、関心の度合いの変化を捉えてゆく。客観的な視点からだけでは得られないものを起点とすることで、知識として提示可能な重要シーンを捉えてゆくことを可能にする。

プレゼンテーション式ミーティングに注目し、再利用可能な状態へと切り出してゆくために、参加者の注目や関心の内容を発露させるものとして、カメラ付小型タブレット PC デバイスである PhotoChat [伊藤 06][角 08] を使用し、その操作とプレゼンテーションの操作、カメラ映像の関連性に注目する。ウェアラブルデバイスのような装着に時間がかかり、身体活動を制限する可能性のあるものは使用しない。

2. 関連研究

ミーティングを映像で撮り続けることにより、映像の特徴変化からシーンの切り替えを自動判別して、インデックスを付ける研究が従来からなされている [Chiu 00][Uchihashi 99]。これらは会議の外面的特徴を利用したものであるのに対し、本研究では個々の関心といった内面的特徴を利用することが狙いである。

手作業で様々なモダリティのインデックスをつけることによって、ミーティングのコーパスを作成したものととして、VACE [Chen 06] が挙げられる。これに伴うインデキシング作業のコストは高く、日常的に行うことはできない。このためミーティングに参加しながらにして、参加者全員が協調してミーティングの注目点を選び出していくことで、日常的に使えるものを目指す。

ミーティングの再利用性を目指した研究としてディスカッションメディア [友部 07] が挙げられる。ディスカッションメディアでは議論の構造を明確にすることで、高い検索性を有する議事録システムである。本研究の特徴は議論のセグメンテーションを発言など表だった行動だけでなく、心理的な要素をともなう PhotoChat デバイスの操作からも行うことで、その内容をミーティングの場に還元し、参加者が受動的に知識の流れを得られることを目的としている。

参加者の視線を自動抽出して、ミーティング映像にインデックスを付ける研究 [Stiefelhagen 99] がある。我々は同様に PhotoChat を利用することで個々の注目を抽出することで、映像を再利用可能なものとする。

3. ミーティングの記録と再利用を促すシステム

3.1 システムの概要

図 1 のように、プレゼンテーション式ミーティングが行われる場全体を俯瞰するものとして、複数の環境カメラを用いて多方向から同期的に収録する。話し手はスライドを用いてプレゼンテーションを行うが、この際スライドに現れる言語的内容とスライドの操作履歴をシステムが捕捉する。プレゼンテーションの聞き手はカメラ型キャプチャデバイスである PhotoChat を用いてミーティングに参加する。

これらの情報を関連付けてストレージに保存し、それぞれの情報を統合して観覧可能な状態にする。ミーティングの次回発表者はこれらの情報から重要な点を得ることができる。

スライドと会話映像を関連付けることで、発表の背景となる知識を運用する。発表の際にはサイドスクリーンに提示されるようにする。またスライドの再利用性に注目し、再利用されたスライドと関連付けられた会話映像コンテンツを提示されるようにする。

連絡先: 中沢拓磨, 京都大学大学院情報学研究科, 京都市左京区吉田本町 工学部 10 号館 214 号室

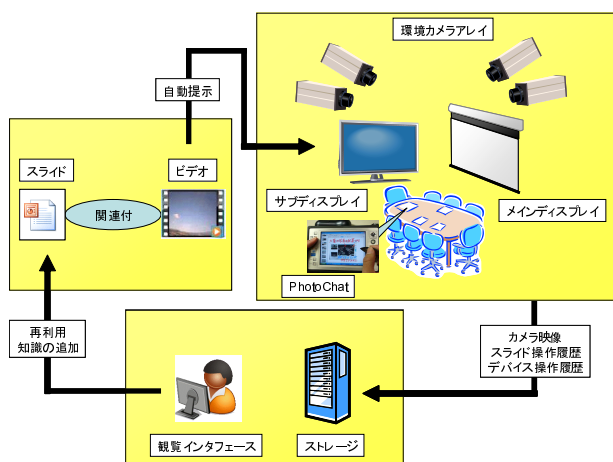


図 1: システムの構成

3.2 カメラの接続方式

カメラは室内のいたるところに設置可能である必要があるが、カメラの接続方式が IEEE1394 や USB であると、ケーブル長が数メートルまでに制限される。そのためカメラの設置箇所のすぐ隣に制御用のコンピュータを配置する必要があるが、この場合それぞれのコンピュータはマスターの制御コンピュータに Ethernet などですらに接続する必要が出てくる。またカメラの制御に十分な性能を持つ程度のコンピュータを、自由に配置されるカメラのすぐ隣に設置することは困難である。そのためカメラはあらかじめ Ethernet によるネットワーク機能を持つものを選択し、制御用コンピュータとの接続はこのネットワーク機能を介して TCP/IP によって行った。制御用コンピュータは離れた位置に一括して設置されるため、カメラの設置の自由度を阻害しない。

制御用コンピュータはクラスタ構成のコンピュータを用いている。撮影の際は複数カメラを同時に作動させるため、複数の動画ストリームが同時に取り扱われることになるが、安定性を確保するため、図 2 のようにコンピュータのノード 1 つに対してカメラ 1 つを割り当て、互いのストリームが干渉しないように構成している。

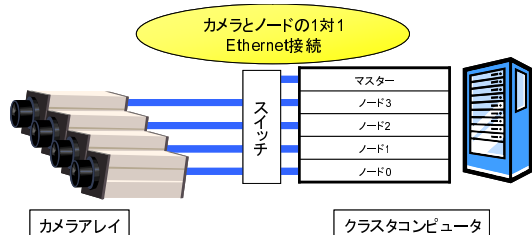


図 2: 複数のネットワークカメラによる映像記録

3.3 カメラの同期方式

撮影システムの負荷によってフレームレートが変動することが起きやすいが、これを補正するためには 1 フレームごとにすべてのフレームの調整を行う必要がある。それはフレームの変動による映像の時間とのずれは撮影時間がたつにつれて蓄積するため、部分だけを修正することはできないからである。本実装におけるシステムでは、ネットワークカメラ、制御

コンピュータの内部クロックを、それぞれ制御マスターに対して NTP で同期させている。エンコード方式を Motion JPEG とし、1 フレーム単位に時刻情報を持たせることで時刻同期可能とした。実際には Ethernet 伝送による不安定性により、カメラから出力される映像のフレームレートは変動するが、各フレームの持つ時刻情報は制御コンピュータに送られた時刻ではなく、フレームが生成された時刻であるため補正が可能である。カメラの仕様から最大フレームレートが規定されるため、変動フレームレートの映像を固定フレームレートの映像に変換するには以下のアルゴリズムによって行えばよい。

1. 最大フレームレートを R とすると、フレームレート R の固定フレームレート映像に変換する
2. 以下を繰り返す
 - (a) 新たに生成するフレームに対応する時刻以前の未処理のフレームが存在すればそれを挿入する
 - (b) なければ NULL フレームを挿入する

映像の撮影開始シグナルと撮影終了シグナルの同期発行も実装されており、ファイル単位での比較も容易に行える。

3.4 カメラの操作方式

簡単に実装するには、カメラを起動する際にコンソールからコマンドを実行するような方式をとればよいが、カメラシステムの利用者の利便性を考慮し、Web インタフェースのボタンにより撮影開始と撮影終了を可能にしている。これにより、LAN 内の任意の PC から操作が可能となり、操作者の位置の自由度も改善されている。

3.5 カメラの運用実態

このカメラアレイシステムは、精細コミュニケーションセンサルームである IMADE Room [来嶋 07] に設置されている。IMADE Room を用いた 4 種類の実験によってカメラアレイシステムは運用された。解像度は 640×480 であり、最大フレームレートは 30fps、ビットレートは 5Mbps 程度である。カメラは 4 台運用され、1 台あたりの総収録時間数は約 40 時間である。すなわち 1 台あたり実データだけで 90GB 程度のディスク容量を使用している。この他にテスト撮影や、変換作業領域などを合わせると、200GB 程度使用することになる。そのため持続的な収録を可能にするためには、定期的に高圧縮率なコーデックによってエンコードを行い、大容量ストレージに順次退避してゆく必要がある。

4. ミーティング映像の再利用性確保

プレゼンテーション式ミーティングの映像をただ漫然と撮影しているだけでは、それを再利用するためには映像全体を見返す必要があり、たぶんに手間と時間を要してしまう。このためなんらかの目印となるものを映像に付与することが必要である。またカメラを複数台用意したとき、映像の利用者がすべての視点のカメラを確認しなくとも、必要な方向のカメラを選択し、利用者に提示することが望まれる。

これらの手がかりとなる目印としてプレゼンテーションスライドの操作や発話タイミングなどが考えられるが、ここでは会話の場に PhotoChat を用いることで、参加者全員がその場で協動的にミーティングへの目印をつけていくことにより実現する。

4.1 時間方向での会話の切り出しと意味づけ

PhotoChat はカメラ付小型タブレット PC デバイスであり、このデバイス上で写真と手書き文字によって他のユーザと会話が可能である。以前より議事録作成ツールの可能性が指摘されているが、本格的な利用を目指すには客観的なカメラ映像との同時運用が望ましい。



図 3: PhotoChat の議事録利用

図 3 のように、PhotoChat では議事録としての運用可能性が示されている。その例としては発言中の話者の写真を撮り、その写真に会話ふきだしを付与しそこに会話内容を記述するものがある。また話し手が一方的にプレゼンテーションを行っている場であっても、聞き手側はそれをさえぎることなくスライドを写真に取り、その上でメモや質問、内容へのつっこみなどの会話を行うことができる。このような聞き手側の行動はカメラやマイクで捉えることができず、聞き手の内面の発露が PhotoChat 上で起こっていると考えられる。この聞き手の内面の変化は、カメラなどに映る外界の状況からもなんらかの影響を受けていると考えられる。そのため PhotoChat の操作履歴からカメラ映像に対して目印を付与したうえで、その利用方法を検討する。

PhotoChat の特徴としては、利用者が撮影対象に気がつきその重要性を計った上でカメラを構えて撮影するため、実際の注目時刻に対して撮影時刻が遅れる可能性がある。さらにその上の手書き文字が実際の発話内容を忠実に再現するものだとすれば、手書き文字に対応する発話はその時刻以前に行われたことになる。

別の可能性として、PhotoChat の利用者は随時その撮影・書き込み内容を確認するのであるが、特定の写真に対して全体の注目や書き込みが集中する時間については、なんらかの重要な部分が発話についても起こっていると推測できる。たとえば、スライドの特定の部分について深く言及が起きたり、時間的にさかのぼる内容に言及されたとき、利用者はそのことを確認しようとしたり、思い出そうとして、それらが表現された写真、あるいは書き込みに注目しようとすると考えられる。そのため利用者の動きからミーティング全体のダイナミクスを取得して、映像の時間ごとの利用価値を判別することに役立てることができる。

4.2 カメラで捕らえるべき範囲と映像選択

図 4 のようなカメラの設置を構成する。

まずプレゼンテーション式ミーティングを撮影の対象としていることから、スクリーン方向のカメラとスクリーンから全体

を見渡すカメラを用意する。

次に人物に向かう方向の映像が必要である。この際撮影範囲は同じ方向に座る人物が全員含まれるようにする。カメラ台数が増えれば各個人を詳細に捕らえるカメラを設置し、表情や視線などが解析可能な十分な映像を捉える。

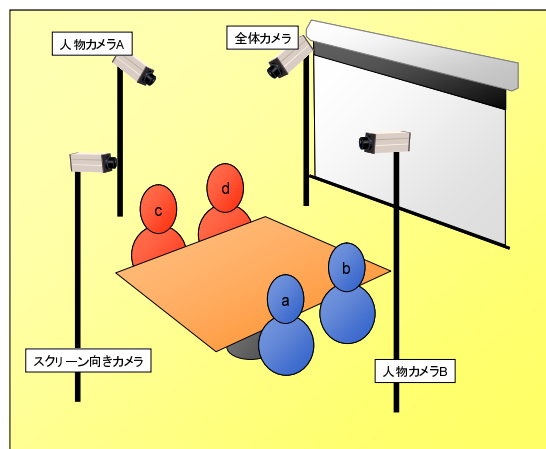


図 4: 映像選択

図 4 のようなカメラの設置状況の下で、映像選択をすることを考える。参加者はそれぞれ手元に PhotoChat デバイスを持ち、これを用いてミーティングのシーンを切り取ってゆくのであるが、積極的な発話姿勢にある人は、PhotoChat を同時に使うことは難しい。すなわち積極的な発話姿勢では、目配せをしジェスチャーを伴っていると考えられるので、PhotoChat を見ることも持つこともできない状況である。このため、PhotoChat の利用頻度の低いところに発話の中心があると考えられる。たとえば人物 a が積極的に発言しているとするれば、PhotoChat のアクティブユーザは b, c, d であり、それに対応して人物カメラ A の映像を選択する。

また参加者が撮影する対象の多くは、スライドの画面、他の参加者、手元の資料と考えられる。このうちスライドの画面については、スライド切り替わりの瞬間を他の方法によって捉えるほうがよい。それ以外の場合、撮影対象は撮影者と違う側にいると考えると、撮影者の背後に設置しているカメラの映像を選択することが考えられる。たとえば人物 a が撮影を行ったならば、その対象は c, d であると推測し、人物カメラ B の映像を選択する。

5. まとめ

本稿では、プレゼンテーション式ミーティングドメインにおける知識の提示に向けて会話を収録する環境を提案した。今後、実験協力者を用いて運用実験を行い、客観的評価を得る必要がある。また提示知識の妥当性を検証し、提示の優先順位やタイミングについてさらに検討を進める必要がある。

参考文献

[Chen 06] Chen, L., Rose, R. T., Qiao, Y., Kimbara, I., Parrill, F., Welji, H., Han, T., Tu, J., Huang, Z., Harper, M., Quek, F., Xiong, Y., McNeill, D., Tuttle, R., and Huang, T.: VACE multimodal meeting corpus, in *Second International Workshop on Machine Learning for*

Multimodal Interaction (MLMI 2005), Vol. 3869 of *Lecture Notes in Computer Science*, pp. 40–51, Springer (2006)

[Chiu 00] Chiu, P., Kapuskar, A., Reitmeier, S., and Wilcox, L.: Room with a rear view: Meeting capture in a multimedia conference room, *IEEE Multimedia*, Vol. 7, No. 4, pp. 48–54 (2000)

[Stiefelhagen 99] Stiefelhagen, R., Yang, J., and Waibel, A.: Modeling focus of attention for meeting indexing, in *ACM Multimedia '99*, pp. 3–10ACM (1999)

[Uchihashi 99] Uchihashi, S., Foote, J., Girgensohn, A., and Boreczky, J.: Video Manga: Generating semantically meaningful video summaries, in *Proceedings of Multimedia '99*, pp. 383–392ACM (1999)

[伊藤 06] 伊藤 惇, 角 康之, 久保田 秀和, 西田 豊明: Photo Chat: 互いの視点画像に「書き込む」ことによるコミュニケーション支援, 第 20 回人工知能学会全国大会, 人工知能学会 (2006)

[角 08] 角 康之, 伊藤 惇, 西田 豊明: PhotoChat: 写真と書き込みの共有によるコミュニケーション支援システム, 情報処理学会論文誌, Vol. 49, No. 6 (2008)

[友部 07] 友部 博教, 土田 貴裕, 大平 茂輝, 長尾 確: ディスカッションメディア: 会議コンテンツの構造化と効率的な閲覧システム, 第 21 回人工知能学会全国大会, 人工知能学会 (2007)

[來嶋 07] 來嶋 宏幸, 坊農 真弓, 角 康之, 西田 豊明: マルチモーダルインタラクション分析のためのコーパス環境構築, 情報処理学会研究報告. ヒューマンコンピュータインタラクション研究会報告, Vol. 2007, No. 99, pp. 63–70 (2007)