

K-42 遠隔サイト間での注目の共有支援のための映像撮影・伝送システム An intelligent video conference system for supporting attention sharing

尾形 涼
Ryo Ogata

尾関 基行
Motoyuki Ozeki

中村 裕一
Yuichi Nakamura

大田 友一
Yuichi Ohta

1. はじめに

本稿では、対話シーンを知的撮影・編集、得られた映像を伝送して、遠隔地の人間と注目の共有を支援する自動化システムを紹介する。このシステムの目的は、遠隔地間でのコミュニケーションのための撮影、編集だけでなく、議事録を映像アーカイブとして取得すること、将来的には、より高度な映像コンテンツの自動取得を行なうことである。

2. 遠隔サイト間でのコミュニケーション

遠隔地の相手と円滑なコミュニケーションを行なう場合、その相手の周辺の様子、その相手が注目しているもの、その相手が把持しているものに関する様々な情報が必要である。このような情報を伝えるためには、シーンでの人物の表情や注目物体をわかりやすく撮影したり、お互いの視線に近いカメラワークで撮影するといったことを考えなければいけない。また注目を共有するためには、お互いの注意が正しい対象に向けられるように映像の編集を行なう必要がある。

そのために本研究では映画やTVで用いられる技法を参考に、必要となるショットを考察し、それらを切替えることで、効果的な映像を生成することを目指す。構築したシステムの概要を図1に示す。具体的には、複数のカメラによる撮影、発話や動作*を利用した自動編集を行う。

映像はMPEGに符号化して、イーサネットを利用して相手のサイトに伝送される。汎用的なネットワークを通信回線として利用するので、汎用性が高く安価なシステムを構築することができる。このシステムは両方のサイトから自由に制御し、実時間での実験・オフラインでの実験の両方を行なうことができる。

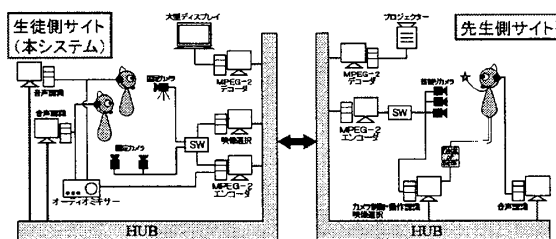


図1: 実験を行ったコミュニケーションシステムの構成図

*今回は未実装

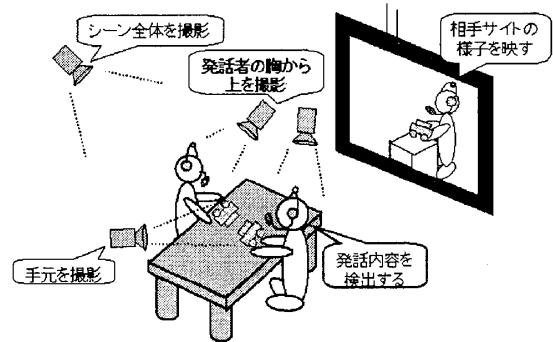


図2: カメラの配置

3. 対話シーン撮影の設定

映画やTVの撮影・編集の基本テクニック[1]を参考にしてカメラを配置した。その理由、狙った効果を次に示す。

1. 各人物の胸から上を映したショット

- 発話者を映すことでその人物が話していることを明確にする
- 話者の表情は状況理解のための情報として有用である

2. 作業現場である机上のアップショット

- 物理的に細かい作業は多くの場合、人物の手元で発生する
- アップで映すことで相手に視覚的に理解させる

3. シーン全体が映るショット

- 状況を理解するためには、対話者の周辺の情報も重要である
- 注目対象を映すのに適切なショットが上の2つでは得られない場合はこのショットで代用とする

これらのショットの中から一つを選択し、相手に提示映像として送信する。実際のカメラ配置は図2の様に設定した。

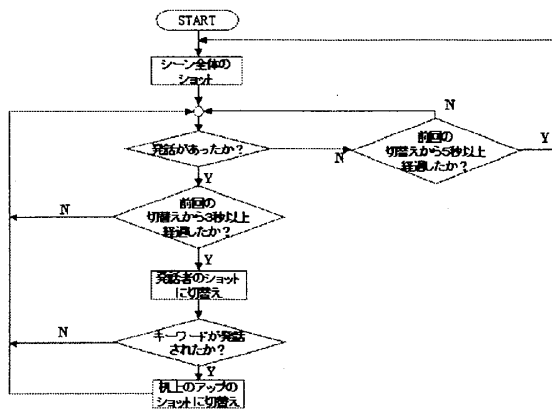


図 3: カメラ切替のアルゴリズム

4. 映像の自動編集例

今回は音声認識だけを用いた簡単なアルゴリズムを設定し、システムの動作を確認した。音声認識には Vi-aVoice を使用し、それによって得られた発話データと図 3 に示す映像提示アルゴリズムに従って、切替え命令をスイッチャーに送信する。

ショットの切替の規則は以下の通り。

- シーン全体を映すショットをマスター・ショットとし、基本的にはこのショットを送信する
- 発話を検出したら発話した人物のショットを最低でも 3 秒間送信する[†]
- 発話データの中に「この」、「このような」などの手元作業を示す言葉が含まれていた場合は、手元のショットを送信する
- 誰の発話も検出されずに 5 秒が経過した場合、その時点の重要な注目物体はないと判断して、シーン全体のショットを送信する

実験例として遠隔地間の約 2 分 30 秒程の対話を示す。内容としては、先生側のサイトで作業についての説明が行われ、それを見ながら生徒側のサイトで二人の生徒が会話をしながら作業を行うという場面を想定した。ただし、先生側では我々が以前から構築している尾関らのシステム [3] により撮影及び編集を行ない、生徒側には本稿で述べてきたシステムを用いる。

実験の結果、得られた映像の一部を図 4 に示す。左側に本システムにより撮影され相手に伝送される映像、右側に相手サイトから伝送された映像を示す。図からわかるように、会話の内容に沿って状況を理解するのに適切だと思われるショットに随時切り替わっている。

[†]3 秒というのは連続で複数人の発話を検出した場合にショットの切り替えが連続で起こり、見ている者の目が疲れるのを防ぐためである [2]。



図 4: コミュニケーション映像の一部

上記の例には示されていないが、これまでの実験では、発話内容を完全に認識できず、正しい映像に切り替わらないことも多くあった。今後は、音声認識ソフトの認識精度も考慮に入れた研究が必要である。

5. まとめ

遠隔コミュニケーションにおいて注目共有の支援を行なうための自動撮影・編集を行なうシステムを構築した。今回は 4 つのカメラで撮影した映像を、随時切替えながら提示する簡単な撮影・編集システムを構築し、実際の作業場面を想定した実験を行った。今後は、より高度なカメラワークの実装や、映像編集方法について考察し、システムに取り入れる予定である。

参考文献

- [1] Daniel Arijon 著, 岩本憲児, 出口丈人 訳, 「映画の文法」, 紀伊国屋書店, 1980.
- [2] 諏訪三千男, 菅野順吉ほか. 映像編集の秘訣. 玄光社, 1999.
- [3] Motoyuki Ozeki, Yuichi Nakamura, Yuichi Ohta. Camerawork For Intelligent Video Production - Capturing Desktop Manipulations. IEEE International Conference on Multimedia and Expo, pp.41-44. 2001.