

全方位画像からのインタラクション認識を用いた会議記録の編集・閲覧システム

大江 展弘[†] 平野 靖[‡] 梶田 将司[‡] 間瀬 健二[‡]
名古屋大学大学院情報科学研究科[†] 名古屋大学情報連携基盤センター[‡]

1. はじめに

会議の議事内容を記録し後日閲覧することは、記憶の想起や新たな発想に結びつくと考えられ、その有効性は高い。ビデオカメラ等を用いて収録された会議記録は、通常の議事録や会議資料の文書情報では不可能な、ノンバーバル情報の記録や会議のディテールを保持できるという利点がある。しかしその一方で、会議全体のビデオ記録をすべて閲覧する時間の確保や、ユーザが期待するシーンを発見する際に手間がかかるといった問題がある。そこで効率のよい会議記録閲覧を可能とするため、記憶の想起を助けるキューを検索条件として会議ビデオ記録を閲覧する研究がある [1]。

本稿は、全方位画像を用いるとともに、文献 [1] では扱っていない会議参加者のインタラクション行動情報を用いた会議記録の検索・閲覧手法を検討する。本稿では、1つのテーブルを囲む場での比較的少人数の会議を対象とし、インタラクション行動として会議参加者の注目行動を扱う。注目対象物をスライド、他の参加者、テーブルに置かれているノートや資料とし、会議参加者の注目行動を検索の条件とする会議記録の編集・閲覧システムを提案する。

2. 全方位画像を用いた会議記録

本稿では、会議記録を全方位画像として収録する。これによりテーブルを囲む会議参加者を一枚の画像で取得することが可能となる。

全方位画像を用いた会議記録の例として文献 [2] がある。これは会議記録を全方位カメラと4個のマイクロフォンで行い、発話者の特定や会議参加者の位置情報を閲覧シーン選択の手がかりとしている。本稿は、文献 [2] では認識していない会議参加者のインタラクション行動を会議記録の編集・閲覧の手がかりとして用いる点を特長とする。また、会議参加者の顔方向を推定することを考慮し、全方位画像をパノラマ展開処理した画像を用いる。

3. 会議参加者の注目行動

3.1 会議参加者のインタラクション行動

本研究では、会議記録を閲覧する際の検索の手がかりとして会議参加者のインタラクション行動に着目する。インタラクション行動認識の手がかりとしては参加者のジェスチャ、注目行動、表情変化等が考えられる。本稿では顔方向を推定することで認識可能と思われる注目行動を判別する。注目行動の判別により、その時点での参加者の興味を集中している対象物が特定でき、検索・閲覧の際に利用できると考えられる。提案するシステムで

表 1: 顔方向推定結果

	被 1	被 2	被 3	被 4
実験 1 水平方向誤差	11.4	16.3	13.9	7.7
実験 1 垂直方向誤差	5.8	10.3	8.8	11.3
実験 2 水平方向誤差	25.3	17.9	17.6	6.8
実験 2 垂直方向誤差	6.9	10.7	9.5	10.9

は、スライド、他の参加者、参加者の正面下方スペース（ノート、資料等）を注目対象物とする。

3.2 注目行動判別

注目行動判別のため、全会議参加者の顔方向を推定する。顔方向は濃淡顔領域画像を用いた固有空間法による処理から推定する [3]。本稿では、水平方向に 10 度間隔で -120 度 \sim 120 度の 25 段階、垂直方向に 10 度間隔で -30 度 \sim 20 度の 6 段階の方向を向いた顔画像をあらかじめ撮影し、学習画像とする。なお、水平、垂直方向とも、水平・正面の顔の向きを 0 度とする。椅子は自由に回転できるようにして撮影した。

顔領域の重心位置から求める参加者位置、および手動入力されたスライド位置等の位置関係から注目していると考えられる角度範囲を決定し、推定顔方向を用いて注目判別を行う。

3.3 顔方向推定の精度評価

顔方向推定の精度評価実験を行った。表 1 は被験者 4 人に対して行った顔方向推定の精度を示している。実験 1 では、注目点の方向に椅子を回転・固定し、被験者が注目点に対し常に正面を向いて注視するようにした。実験 2 では、被験者に椅子を自由に回転してもらい、自然な姿勢で注目点の注視を行った。被験者の注目する方向はともに学習画像収録時と同様である。また、被験者は学習画像と同じ人物で実験した。

誤差の傾向として、外側に向くほど精度の劣化が見られた。また実験 1 に比べ実験 2 の精度が落ちる傾向があるのは、注目点を注視する際、水平、垂直方向への回転の他に、被験者の水平・正面方向を軸とする回転も出てくるため、誤判別が生じたと考えられる。



図 1: インタラクション検出モジュールインターフェース

Editing and Browsing System of Meeting Record Using the Interaction Recognition from Omnidirectional Images

[†]Nobuhiro Ohe [‡]Yasushi Hirano [‡]Shoji Kajita [‡]Kenji Mase
[†]Graduate School of Information Science, Nagoya University
[‡]Information Technology Center, Nagoya University



図 2: 閲覧モジュールインタフェース

4. 会議記録の編集・閲覧システム

筆者らは会議参加者の注目行動を検索の手がかりとした、会議記録の編集・閲覧システムを開発している。本システムは会議参加者の注目行動に関するメタデータファイルを作成するインタラクション検出モジュールと、注目行動によるシーン検索・閲覧が可能な閲覧モジュールに分かれている。以下でそれぞれのモジュールの処理の流れを述べる。

4.1 インタラクション検出モジュール

会議記録画像を読み込み、初期フレームで手動によって会議参加者、スライド位置、テーブルタイプと大きさを設定する。設定後、追跡処理を開始することで顔領域を自動追跡し、顔方向と位置を求めると同時にそれぞれの注目対象物に対する注目判定を行う。全追跡が終了した後、得られた参加者の顔方向、位置と注目情報をファイルに出力する。図 1 はインタラクション検出モジュールの自動追跡処理中の画面である。画面下部は初期設定を行う部分である。また上部では顔方向の推定結果を出力しており、中央部のパノラマ展開画像から抽出した顔領域と推定角度に対応する学習画像を表示している。

4.2 会議記録閲覧モジュール

会議閲覧モジュールでは、インタラクション検出モジュールで出力したファイルを読み込む(図 2)。その後、以下で述べる検索条件を図 2(a) の下部に入力し、検索することで、条件に一致するシーンが提示される。なお、検索条件として下記の 4 種類を扱うことができる。

- 検索条件
- スライドへの注目人数と注目時間
 - 下方向への注目人数と注目時間
 - 各参加者への注目人数と注目時間
 - 各参加者のスライドや下方向への注目時間

図 2(b) は検索条件 a において注目人数 2 人、注目時間 10 秒を検索条件として入力したときの条件一致シーン提示画面の例であり、図 2(c) は条件が一致した複数のシーンの中からユーザが選択したシーンを再生した画面である。

また本システムではスライドや下向きへの注目人数、および各参加者の顔方向や位置情報の時間変化グラフを閲覧することができる。このグラフを利用することは、検索条件を入力する際の補助となる。

4.3 考察

インタラクション検出モジュールでは、初期処理で各情報を手動入力する必要があるが、操作にそれほどの負荷はかからず、使いやすく感じられた。しかし顔領域を抽出する際、参加者の手が顔領域に重なることで、顔領域抽出や顔方向推定が失敗するケースが度々見られた。このため、さらに別の処理を加え、顔領域を適切に追跡する必要がある。閲覧モジュールでは、判別結果に従った検索結果が得られることを確認した。ただし、検索結果が主観判別とは異なる場合があり、特に検索条件 b, c を用いた検索でこのケースは多かった。これは、検索条件が各参加者や下方向の場合、注目判定の際に用いる角度範囲がスライドの場合よりも小さいためと思われる。

本システムでは現在のところ画像、音声情報を会議データとして記録しているが、処理には画像情報のみを用いている。そこで、例えば複数マイクを用いた自動発話者特定や、発話者のピッチやパワーといった、音声処理により得られる情報を同時に扱うことで、マルチモーダルな検索・閲覧が可能となると考える。

5. おわりに

本研究では、スライドや他の参加者、ノートや資料への注目行動を会議におけるインタラクション行動とし、それを検索の手がかりとした会議記録の編集・閲覧システムを試作した。今後の課題として、顔方向推定の精度の向上が挙げられる。精度の向上により、会議参加者同士の向き合いなども条件に入れることができ、音声情報などと組み合わせることにより、会議の雰囲気などマルチモーダルな情報を利用した検索が可能になるとと思われる。また本システムは現在会議記録の検索機能のみであるが要約機能も加えることを検討している。

謝辞

本研究は総務省 SCOPE(No.:041306003) の支援により行われた。

参考文献

- [1] Alejandro Jaimes, Kengo Omura, Takeshi Nagamine, and Kazutaka Hirata, "Memory Cues for Meeting Video Retrieval", CARPE'04, pp.74-85, 2004.
- [2] Dar-Shyang Lee, Berna Erol, Jamey Graham, et al, "Portable Meeting Recorder", ACM Multimedia 2002, pp.493-502, 2002.
- [3] 渡辺 彰裕, 斎藤 英雄 "固有空間法を用いた濃淡画像からの顔の向き推定法", 信学技報 PRMU, Vol.97, pp.111-116, 1997.