

全方位画像を用いた会議記録・閲覧システム

Recording and Browsing System for Small Conference
Using Omnidirectional Images大江 展弘†
Nobuhiro Ohe平野 靖†
Yasushi Hirano梶田 将司†
Shoji Kajita間瀬 健二†
Kenji Mase

1. はじめに

会議やミーティングなどの議事記録を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るため、その有効性は高い。しかし、今日の会議では議事内容は文書で要約して記述されるため、会議の雰囲気や補足的な発言に関する情報は完全に欠落してしまう。一方、会議全体をビデオ記録してもすべてを閲覧することは時間の無駄である。そこで、人の振る舞いや視線情報を積極的に利用し、議事の内容や会議の状況認識を行い、要約や検索の手がかりを与える方法が考えられる [1]。また、カメラとマイクロホンによる会議風景の収録と半自動的な議論の構造化の仕組みを実現した事例もある [2]。

本稿では、人の興味や人と人とのインタラクションを認識する手がかりとして参加者の顔の向きを用いた会議記録・閲覧システムを提案する。顔の向き情報を用いることによって、議事録では記録や認識が困難な対象物への注目度を記録することが可能になる。本稿では、円卓会議のような1つのテーブルを囲む場での会議を対象とし、全方位画像を用いて全参加者の顔の向きを認識し、会議の記録や閲覧、検索に用いる手法を検討した。特に今回はスライドへの注目度に着目したシステムを試作した。

2. 全方位画像とパノラマ展開

円卓会議のような場において、通常のカメラでは局所的な情報しか獲得することができず、また複数台のカメラ使用は処理を難解にすることから会議の様子を取得するには適さない。そのため、本研究では会議の様子を全方位画像として取得することにする。図1は実際に会議の様子を撮影した全方位画像である。

全方位画像の座標を (x, y) 、変換後の画像の座標を (X, Y) とし、式 (1) を用いてパノラマ展開画像を作成する。ここで x_c, y_c は全方位画像の中心座標を表し、 r は全方位画像の中心から画像が映し出されている外円までの長さである。図2は式 (1) を用いて図1の画像をパノラマ展開した画像である。

$$\begin{aligned} X &= r \tan^{-1}((y_c - y)/(x_c - x)) \\ Y &= r - \sqrt{(x_c - x)^2 + (y_c - y)^2} \end{aligned} \quad (1)$$

3. 注目度判別

3.1 スライドへの注目度

会議の場では発表資料としてスライド (OHP, ホワイトボードなどを含む) を用いることが多く、スライドへ

†名古屋大学大学院情報科学研究科社会システム情報学専攻
†名古屋大学情報連携基盤センター

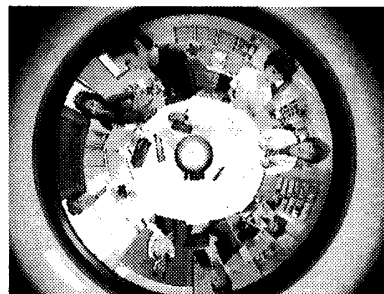


図1: 全方位画像の例



図2: パノラマ展開画像の例

の注目度の高い区間は参加者の興味が大きいと考えられる。そのような区間を自動で判別し、閲覧者に提示することで閲覧の効率が向上すると考える。スライドへの注目度を求めるには各参加者の顔の向きを求める必要がある。以下で顔の向きとスライドに注目しているかどうかを求める手法を述べる。

3.2 固有空間法による顔の向き推定

前章で得たパノラマ展開画像を用いて参加者の顔の向きを推定する。本稿では顔の向きを推定する手法として固有空間法を用いる。以下に簡単に処理の流れを示す。

ある大きさに正規化した K 枚の学習画像 $\mathbf{x}_k (k = 1, \dots, K)$ が与えられたとき、その平均画像 $\bar{\mathbf{x}}$ 、共分散行列 Σ は次式で与えられる。学習画像は様々な方向を向いている顔画像である。この共分散行列の固有方程式から固有値と固有ベクトルを求める。

$$\bar{\mathbf{x}} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k, \quad \Sigma = \frac{1}{K} \sum_{k=1}^K (\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})^T \quad (2)$$

Φ_n を Σ の上位 n 個の固有ベクトルを列ベクトルとする行列とすると、画像 \mathbf{x} の固有空間への射影 \mathbf{x}^* は次のようになる。

$$\mathbf{x}^* = \Phi_n^T (\mathbf{x} - \bar{\mathbf{x}}) \quad (3)$$

パノラマ展開画像から抽出された顔領域を入力画像とし、式 (3) を用いて射影画像を作り、内積をとって類似度とする。入力画像との類似度が最も高い学習画像における顔の向き (角度) を推定値とする。顔領域は画像中の一定閾値を越える画素数を持つ肌色領域とする。

3.3 スライド注目判定

前節で求めた顔の向きと推定値を用いて、スライドを注目しているかどうかを特定する。今回は大型ディスプレイの設置してある位置をスライドの位置とし事前に手動で入力しておく。また各参加者の位置をパノラマ展開画像における抽出顔領域の重心座標から推定する。そして、参加者、スライド、カメラの位置関係からスライドを注目していると考えられる角度の範囲を決定し、範囲内に推定角度が含まれるとき注目していると決定する。

4. 予備実験

4.1 学習画像の取得

スライドへの注目度判定の精度を確認するために予備実験を行った。今回用いた学習画像は、会議参加者本人のものとした。角度は、横方向では左側面を -90 度、右側面を 90 度に設定し、 15 度間隔の 13 段階、縦方向では上向きの最大を 30 度、下向きの最大を -60 度に設定し、 30 度間隔の 4 段階とした。また、固有空間の次元数は累積寄与率が 85% 以上となるよう決定した。

4.2 実験方法と結果

3人の被験者にタイミングに合わせてスライドや正面、他の被験者を見るよう教示し、その様子を撮影した。取得した全方位画像に対し3章で述べた手法を用いてスライドへの注目の有無を求め、正しい結果と比較した。

その結果、参加者のスライド注目の正しい判別結果の割合は平均で約 87% だった。誤認識の原因としては学習画像の枚数の不十分さや顔領域の不適切な抽出による推定角度の誤りが考えられる。

5. 試作システムの作成

5.1 会議記録の編集・閲覧

会議参加者のスライドへの注目度は、長時間の会議記録をする手がかりとなる。そこで、会議記録の検索を可能とするためのデータファイルを作成する編集システムと、条件に合うシーンを検索できる閲覧システムを開発している。

- 会議編集システム

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を決定した後、顔領域を自動追跡し、3章で述べた手法を用いて顔の向きと位置を求める。全追跡が終了した後、得られた参加者の顔の向きと位置の情報をファイルに保存する。こうして特定参加者に注目した閲覧が可能になる。

- 会議閲覧システム

編集システムで得られたファイルの読み込み後、各時間でのスライドへの注目度を決定する。その後、検索条件としてスライドへの注目人数、それが続いた時間を入力し検索することで、条件に一致したそれぞれのシーンが提示され閲覧が可能となる。

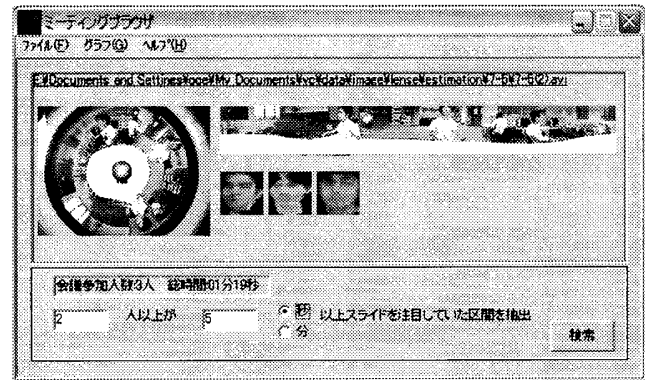


図 3: 検索条件入力画面の例

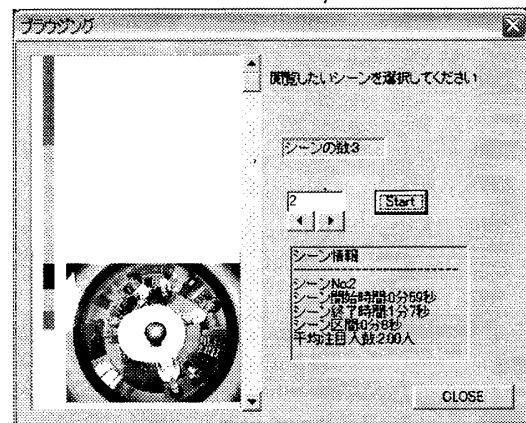


図 4: 検索結果の例

図 3, 図 4 はそれぞれ前章の実験のデータを用いた検索条件入力画面、検索結果画面の例であり、判別結果に従った検索結果が得られることを確認した。

6. おわりに

本研究では、会議参加者の顔の向きを用いて、スライドへの注目度の情報を付与することで、それを手がかりとした検索が可能な会議記録・閲覧システムを試作した。今後の課題として、顔の向き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、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や要約が可能になると思われる。

謝辞

本研究の一部は、文部科学省 21 世紀 COE プログラム「社会情報基盤のための音声映像の知的統合」による。

参考文献

- [1] 角康之, 間瀬健二, “インタラクション・コーパス構築の試みとしてのミーティング・キャプチャ”, ヒューマンインタフェースシンポジウム 2002, pp.241-244, 2002.
- [2] K.Nagao, T.Shimizu, and K.Kaji, “Discussion Mining: Knowledge Discovery from Semantically Annotated Discussion Records”, In Proceedings of the International Workshop “From Semantic Web to Semantic World”, 2003