

部屋重畳型遠隔コラボレーション支援システムにおける利用者の向きの抽出 Extraction of User's Orientation for Room-Overlaying Remote Collaboration Support System

和田 理† Osamu Wada 片桐 滋† Shigeru Katagiri 大崎 美穂† Miho Ohsaki

1. はじめに

遠隔コラボレーション支援システム「t-Room」[1]は、遠隔地にある部屋どうしを仮想的に重ね合わせることで、視聴覚メディアの対称性を確保し、会議に限らない様々な共同作業の実施を支援することを目的とするシステムである。確かに、部屋を重ねる発想は、標準的なテレビ会議システムなどと比べ、部屋内の利用領域を広げることに成功した。しかし、壁面（モニリスと呼ばれる）を構成するディスプレイに対峙するように配置されたカメラを用いる映像撮影は、モニリスから離れた（部屋中心付近）領域の映像オブジェクトの、複数台のディスプレイ上の再生（多重再生）と不自然に拡大された再生（拡大再生）との問題を引き起こし、t-Room 部屋全体の利用を阻害してきた。

我々は、まず多重再生の解決を目指し、複数のカメラ映像の中から適切な1つを選択再生する方式を提案した[2]。本稿では、その選択再生における映像出力（再生）ディスプレイを指定する手法の確立を目指し、複数のカメラ撮影映像の中から利用者の正面顔のみを実時間処理によって選択する手法の実装可能性を調査する。なお、本稿では、動画再生のフレームレートである30fpsに対応する33ミリ秒以内に処理を行うことを実時間処理と定義する。

2. 関連研究

顔画像抽出処理技法の代表的な例として、Haar-like特徴を用いる手法[3]が挙げられる。Haar-like特徴を用いる顔画像抽出処理技法に関して、700MHzのIntel Pentium IIIプロセッサが搭載されたシステムにおいて、304×288ピクセルの画像に対し15fpsでの処理が可能であることが報告されている。一方、本稿で用いるシステム（以下、本システム）においては、カメラより720×1280ピクセルの画像を30fpsで取得する。即ち画像の解像度が向上し、かつフレームレートが増加している。この条件下で実時間処理による顔画像抽出を実現するために、画像サイズの縮小を行う。

近年の顔画像抽出においては、正面顔のみでなく、横顔などの正面顔以外の顔画像の抽出が求められている[4]。しかし、正面顔のみの抽出を必要とする本研究では、向きにかかわらず顔画像を抽出する手法ではなく、正面顔のみを抽出するか、顔の向き情報も合わせて抽出する手法が必要となる。

3. 実装

本システムの実験環境は、先行研究[2]と同様に、正十角形の空間の中に6面のディスプレイによって構成される壁面が配置され、ディスプレイと対峙する位置にカメラが設置されている部屋空間を利用する。本システムの設置例を図1に示す。

カメラ及びディスプレイのそれぞれは、サーバコンピュータにより制御される。サーバコンピュータには、出力先ディスプレイの選択判断を担うオブジェクト抽出サーバと、ディスプレイ上部に設置された側面カメラからの画像の取得及び処理を担うカメラサーバ、ディスプレイへの出力を担うディスプレイサーバとの3種類が存在する。各サーバ間のデータの流れを図2に示す。

基本的な処理の流れは、先行研究と同様であるが、本稿ではそれに加え、カメラサーバに顔画像抽出処理を追加する。これ



図1 実験環境。

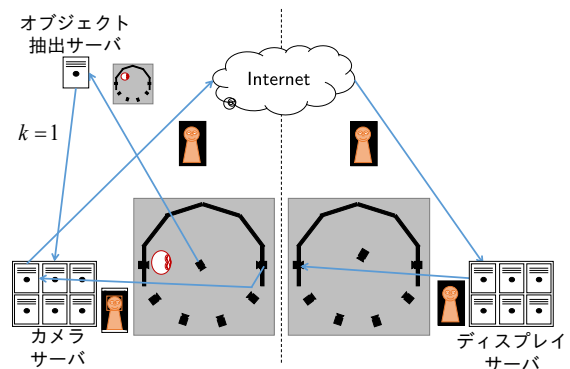


図2 本システムにおけるデータの流れ。

により、顔画像抽出処理による出力先ディスプレイ選択機構を実現する。

先に述べたように、本システムでは33ミリ秒以内に顔画像抽出処理を完了する必要がある。本稿では、オープンソースの画像処理ライブラリであるOpenCVを利用して、Haar-like特徴を用いる顔画像抽出処理を行う。

4. 実験

顔方向の推定可否の判断、及び実時間処理の可否の判断を目的とした、実験環境上における静止画を対象とする正面顔画像抽出処理の動作確認を行う。

本実験にはIntel Core i7-4790プロセッサ、及び8GBのメモリを搭載するサーバコンピュータを使用する。使用する画像は、t-Roomの壁面ディスプレイ上部に設置した6台の側面カメラにより撮影された、部屋空間内に無人の状態、部屋空間中心部に人物が存在し、6台各々のカメラを向いた状態、部屋空間正面のカメラと反対を向いた状態、下を向いた状態、部屋空間周辺部に人物が存在し、対峙する6台のカメラを向いた状態、部屋空間の奥のディスプレイ前に立ち、カメラと反対を向いた状態、下を向いた状態の合計17通りの画像である。画像の解像度はカメラの出力と同じ720×1280、及び縦横それぞれ $\frac{1}{4}$ に縮小した180×320を用いる。顔画像抽出処理の対象の画像の例を図3に示す。

† 同志社大学, Doshisha University



図 3 動作確認に用いる画像の例.

本システムで使用する Haar-like 特徴を用いる顔画像抽出処理の関数は、顔の存在を想定する領域を長方形の存在候補域として出力する。今回の動作確認においては、人物がカメラの方向を向いているときに存在候補域が出力される場合と、人物が存在しない場合や正面以外の方向を向いているときに存在候補域が出力されない場合とを、正しく処理が行われたものとして評価する。また、各入力画像毎に実行時間を測定し、その平均実行時間を評価する。動作確認の結果を表 1 に示す。

表 1 動作確認の結果.

解像度 [pixel]	平均実行時間 [ミリ秒]	適合率	再現率
720 × 1280	195	0.33	0.92
180 × 320	34	0.52	0.92

動作結果における適合率は式 (1)、再現率は式 (2) を用いて計算した。なお、 Tp は人物がカメラの方向を向いているときに存在候補域が出力されたフレームの数、 Fp は人物がカメラ以外の方向(斜めなど)を向いている、もしくは人物が存在しないときに存在候補域が出力されたフレームの数、 Fn は人物がカメラの方向を向いているときに存在候補域が出力されなかったフレームの数を指す。

$$\text{適合率} = \frac{Tp}{Tp + Fp} \quad (1)$$

$$\text{再現率} = \frac{Tp}{Tp + Fn} \quad (2)$$

平均実行時間に着目すると、解像度が 180 × 320 の際に目標値である 33 ミリ秒に近い値になっていることがわかる。また、顔画像抽出処理の適合率においても解像度が 180 × 320 の場合の方が良好な結果が得られることがわかる。しかしながら、いずれの解像度においても部屋空間中心部に人物が存在する際に人物の正面以外のカメラにおいて顔領域が出力される場合が存在した。この問題については 1 フレームの抽出結果でなく、複数フレームの抽出結果を利用することにより、抽出されたフレームの割合を利用する等の対策が求められる。

5. おわりに

本稿では、t-Room における映像出力先ディスプレイ選択基準の設定を目指し、Haar-like 特徴を用いる正面顔抽出機能を実装し、その動作確認を行った。動作確認の実験の結果、画像解像度が 180 × 320 であるときに 35 ミリ秒で処理が行われることを確認した。



図 4 意図しない顔画像抽出処理結果が得られた例.

動作確認実験において、撮影対象オブジェクトである人物が部屋中心部に位置する場合に、隣接する側面カメラから取得した斜め向きの顔が顔画像として抽出される場合があることがわかった。この解決のため、正面顔と斜め向きの顔を識別する機能の実装が求められる。また、今回は静止画に対する顔画像処理を行ったが、システムの動作中、人物の位置や姿勢は常に変化する。そのため、人物の位置や姿勢の変化を考慮した出力先のディスプレイを選択基準の設定が求められる。これを実現する手法として、過去のフレームにおける正面顔抽出の結果を記録し、それらを利用することが考えられる。更に、本稿における実験では除外した拡大再生の問題への対応も求められる。これを実現する手法としてカメラと人物間の距離別の顔領域の大きさの変化を利用することが考えられる。

参考文献

- [1] K. Hirata, Y. Harada, T. Takada, S. Aoyagi, Y. Shirai, N. Yamashita, and J. Yamato, "The t-room - toward the future phone", *NTT Technical Review*, Vol. 4, No. 12, pp. 26–33 (2006).
- [2] 和田 理, 片桐 滋, 大崎 美穂, "t-room における俯瞰カメラを用いた映像出力壁面選択機構の開発", 情報処理学会研究報告, Vol. IPSJ-GN-101, No. 28 (2017).
- [3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", *Proc. IEEE CVPR 2001*, pp. 511–518 (2001).
- [4] S. Lao, 山口 修, "実利用が進む顔画像処理とその応用事例: (後編) 顔画像処理の応用事例", 情報処理, Vol. 50, No. 5, pp. 436–443 (2009).