

発話者位置特定の為の顔抽出手法の検討

谷口 真一

1 はじめに

テレビ電話システムやテレビ会議システムの普及に伴い、それらのシステムにおける映像、音声を用いた協調作業の効率化に関する研究が精力的に行われている。

山下らは遠隔地間で映像に関する対称性を成立させることのできる遠隔コラボレーションシステム t-Room を構築し、映像の対称性が協調作業の作業効率を向上させるとの実験結果¹⁾を報告している。しかし、t-Room において音声の対称性は成立しておらず、それがさらなる協調作業の効率の向上を困難にしているとの指摘がある。

この問題を解決する為、映像に映った発話者の顔の抽出を行い、発話者位置を特定することで音声の対称性を向上する手法の確立を目指す。顔抽出手法には、特徴照合による探索法やテンプレートマッチング法などの種々のアプローチが存在するが、本稿では、抽出法の基本となるテンプレートマッチング法に着目し、特にその実験評価を行う為の実験システムの構築に関して報告する。

2 遠隔コラボレーションシステムの現状

t-Room とは、遠隔地間において周囲を取り巻く大型ディスプレイと多チャンネル音響通信によって構成される部屋同士を接続することで、映像と音声に関する対称性を成立させ、遠隔地間における行為やコミュニケーションに伴う相互理解を促進させるシステムである。t-Room の様子を Fig.1 に示す。



Fig.1 t-Room (出典: ²⁾より引用)

Fig.1 で確認できるように、映像に関しては、各ユーザの視界を一致させる環境の対称性、並びに遠隔地にある物体を指し示すことができる物体の対称性をほぼ確立している。つまり、映像の対称性が成立している。

音声に関しては、双方向に自由に会話できる環境を提供しており、音響環境に関する対称性を確保しているが、音源定位の制御を行っていない為、発話者位置により、音像の位置が変化してしまう。つまり、音声の対称性は成立していない。このことが遠隔地間での相互理解の妨げになる可能性があり、その為、発話者位置特定の為の顔

抽出技術が必要となる。

3 顔抽出手法の選択

3.1 処理の流れ

顔抽出とは、画像認識技術の一つである。画像認識の一般的な流れを Fig.2 に示す。

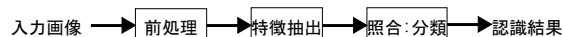


Fig.2 画像処理の流れ (出典: ³⁾より引用)

また、画像認識の主な手法には以下に述べるものがある。

- 特徴照合による探索法
- テンプレートマッチング法

以上に挙げた手法の中から本システムに導入する手法の検討を行う。その際、t-Room では動画をを用いている為、精度のみならず計算時間も配慮して検討を行う必要がある。

3.2 特徴照合による探索法

特徴照合による探索³⁾とは、細線化やヒストグラムなどの特徴抽出を経た画像を用意し、その特徴を用いて照合を行うことにより、認識を行う手法である。特長としては、画像のロバストな特徴を抽出して処理を行うことが出来る為、単純なテンプレートマッチングに比較して精度や処理時間が大幅に向上する可能性があるということが挙げられる。

3.3 テンプレートマッチング法

テンプレートマッチング³⁾とは、上記のような特徴を用いずに画像そのもののデータ表現に準じた標準パターンを用意し、入力されたパターンと 1 対 1 に重ねて比較することによって、一致度評価が高いパターンを見つけることで認識を行う方法である。特長としては、原理が簡単で比較的高速な処理が可能であるということが挙げられる。

4 テンプレートマッチング法を用いる実験システムの構築

4.1 理由

本システムを構築した理由は、テンプレートマッチング法が抽出法の基本的な考え方となっていること、また、テンプレートマッチング法が 3.3 節で述べたように原理が簡単であるという特長を持っていることから、今後、その他の顔抽出手法について検討していく際に役立つのではないかと考えたからである。

4.2 テンプレートマッチングのアルゴリズム

以下にテンプレートマッチングのアルゴリズムの詳細を示す。

1. テンプレートの作成

テンプレートマッチングにおいて重要となる標準パターンの作成を行う。今回は、t-Room のカメラで撮影された顔画像を平均顔ツール⁴⁾を用いて平均化を行ったものを標準パターンとして用いた。

2. 画像読み込み

対象が含まれている可能性のある画像と対象のテンプレート画像を読み込む。

3. 特徴抽出

それぞれの画像に対して特徴の抽出を行う。今回はモノクロ特徴抽出のみを行った。

4. 評価値の算出

画像間の一致度を評価する為に、注目画像に対する一致度評価手法を使用する。今回代表的な一致度評価手法として、輝度差の二乗和 (SSD) を用いる手法、正規化相互相関関数 (NCC) を用いる手法を使用した。

5. 決定

一致度評価手法を用いて算出された評価値と一致度の閾値を比較し、評価値が一致度の閾値以上であれば、現在位置を標準パターンがある場所とみなす。評価値が一致度の閾値を満たしていない場合は、テンプレートを移動させ、4. の処理に戻る。

4.3 システム仕様

本システムは、t-Room のソースコードが C++言語、Qt を用いていることから、C++言語、Qt 環境上で動作するシステムとした。本システムの実行例を Fig.3 に示す。



Fig.3 実行例 (出典：自作)

5 一致度評価関数

5.1 輝度差の二乗和

輝度差の二乗和 (Sum of Squared Difference:SSD) は、入力パターンと標準パターンの輝度差を二乗した値を足

し合わせた値を評価値とする。この評価値は小さい値ほど一致度が高いことを意味する。SSD の特徴として、輝度差を用いている為、入力パターンが全体的な輝度の変化を起こすと評価値が大幅に増加してしまうという点が挙げられる。入力パターンを $I(X, Y)$ 、標準パターンを $T(Wt, Ht)$ とすると SSD は式 (1) で表わせる。

$$R_{SSD}(X, Y) = \sum_{j=0}^{Ht-1} \sum_{i=0}^{Wt-1} (I(X+i, Y+j) - T(i, j))^2 \quad (1)$$

5.2 正規化相互相関関数

正規化相互相関関数 (Normalized Cross-Correlation:NCC) は、入力パターンと標準パターンの輝度値の相関係数を求め、その相関係数の値を評価値とする。この評価値は 1 に近い値程一致度が高いことを意味する。NCC の特徴として、パターン毎に輝度値を輝度の平均値で引くことで正規化を行っている為、入力パターンが全体的な輝度の変化を起こしても評価値に影響を与えにくいという点が挙げられる。入力パターンを $I(X, Y)$ 、標準パターンを $T(Wt, Ht)$ とすると NCC は式 (2) で表わせる。

$$R_{NCC}(X, Y) = \frac{\sum_{j=0}^{Ht-1} \sum_{i=0}^{Wt-1} (I(X+i, Y+j) - \bar{I})(T(i, j) - \bar{T})}{\sqrt{\sum_{j=0}^{Ht-1} \sum_{i=0}^{Wt-1} (I(X+i, Y+j) - \bar{I})^2 \cdot \sum_{j=0}^{Ht-1} \sum_{i=0}^{Wt-1} (T(i, j) - \bar{T})^2}} \quad (2)$$

6 現状及び今後の課題

一致度評価関数を用いたテンプレートマッチング実験システムの構築を行ったが、評価関数の優劣に着目したシステム評価は行っていない。その理由としては、標準パターンに用いる平均顔を作る為の顔画像データが十分に集まっていないことが挙げられる。今後の課題としては、顔画像の収集、構築した実験システムの評価が挙げられる。また、他の手法を用いて、結果の比較を行う必要がある。

参考文献

- 1) 山下直美, 平田圭二: 遠隔コラボレーションシステムにおける物体の共有, The 20th Annual Conference of the Japanese Society for Artificial Intelligence 2006, 2C2-3
- 2) Mirai no Denwa t-Room
<http://www.mirainodenwa.com/index.html>
- 3) 田村秀行 著: コンピュータ画像処理, オーム社, 2002.
- 4) 平均顔合成ツール, 東京大学工学部電子情報科原島苗村研究室
<http://www.hc.ic.i.u-tokyo.ac.jp/project/face/HeikinTool/>