

温度画像処理を用いた発声時の表情認識の高精度化

(情報処理) 池添史隆

1. 緒言

将来、人間とロボットが平和に共存していくためには、コンピュータに人間の感情を理解させる必要がある。既報の温度画像を用いた発声時の表情認識では、認識率を上げる必要があると報告されている[1]。誤認識を起こす原因の1つに、顔部品領域の抽出が正確に行えない事があげられる。

顔部品領域の抽出は、全ての画像に対して行う基本的な処理である。全ての画像を同じ条件下で処理するため、位置、大きさ、回転角度を補正した後、顔部品領域を抽出している。しかし、発声時は顔の動きが生じやすい。そのため、適切な補正が行われないうまま、領域を抽出してしまいがち。

そこで本研究では、顔部品領域の抽出が正確に行えなかった場合は、抽出失敗として認識させる事により、表情認識の高精度化を図った。

2. 開発環境

- ・ プログラム言語は、Visual C++ 6.0 を使用。
- ・ 温度画像は 256 階調のビットマップ形式を用いた。

3. 画像処理手順

以下に表情別特徴量の学習から、未知画像の表情認識に至るまでの処理手順を示す。

(学習処理)

「発声前」と「タ」、「ロー」発声時の、平均無表情データを作成する。

- (1) 回転角度・倍率・位置等の補正を行った後、部品領域を抽出して合計濃度を求める。
- (2) それらの平均濃度を求める。

表情別の特徴量を作成する。

- (1) 各表情の画像に対し、平均無表情データとの差分をとり、得られた差分画像に対して離散的コサイン変換(DCT)を行う。
- (2) 量子化における、境界点を求める。
- (3) 各画像・各領域の0-1データを作成する。
- (4) 作成された0-1データから、特徴量抽出のルールに基づいて特徴量を選び出す。

それぞれの0-1データを、ニューラルネットワークに学習させる。

(認識処理)

認識対象となる、未知画像についても (1),(3),(4)の処理を行う。

未知画像の特徴量から、ニューラルネットワークによる表情認識を行う。

4 . 実験

4 - 1 . 条件

学習には、4種類の表情(怒り、喜び、中立、悲しみ)の温度画像を用いた。「発声前」と「タ・ロー」発声時の画像1枚ずつを1組として、各表情につき20組の画像を学習に用いた。認識には、各表情につき10組の画像を用いた。ただし、前頁にある処理手順の学習段階における(2)は、正確な顔部品領域の抽出が前提条件であるため、抽出失敗分の画像は使用していない。ここで、それぞれの表情において、顔部品領域の抽出が正確に行えなかった画像を「抽出失敗」のパターンとした。そして、最大8種類でのパターン認識を行い、顔部品領域の抽出失敗分を含めない、4パターンでの認識結果との比較を行った。



図1 . 抽出された顔部品領域



図2 . 抽出失敗例

5 . 結果と考察

まず顔部品領域の抽出失敗分を含まない、4パターンで認識をおこなった。その認識率は、怒り - 20%、喜び - 90%、中立 - 0%、悲しみ - 100%であった。

次に、各表情における顔部品領域の抽出失敗分を加えた8パターンでの認識では、中立の認識率が0%から100%へ向上した。しかし、悲しみの表情に40%ほどの認識率の低下がみられた。他、数パターンの認識を行ったが、全ての表情の認識率を向上させるには至らなかった。これは各表情における抽出失敗分において、その特徴量に類似していたものがあつたものと考えられる。そこで、各表情での抽出失敗分を学習させた場合と、学習させなかった場合での認識結果を統合して認識率を得る事とした。

6 . 結言

顔部品抽出の失敗分をパターンとして学習を行い、その認識結果を基本の4パターンでの結果と統合する事により、平均の認識率は52%から77%となった。今後の課題としては、情報統合を行わずに、より高精度な認識が行える特徴量を得る事があげられる。

参考文献

[1] Y.Yoshitomi, S.Kim, T.Kawano and T.Kitazoe, "Effect of Sensor Fusion for Recognition of Emotional States Using Voice, Face Image and Thermal Image of Face", Proc. of 9th IEEE International Workshop on Robot and Human Interactive Communication, (2000), 178-183.