

温度画像を用いた発声時のオンライン表情認識システムの開発

(知能情報システム学) 中野真里

1. 緒言

人とロボットが自然なコミュニケーションをとっていくためには、ロボットが人の感情を理解する必要がある。これを実現するため、さまざまな表情認識手法が研究されてきた。中でも、温度画像を用いた表情認識手法においては照明条件に左右されない表情認識が可能であり、5 表情における平均が 90 % という認識率を得ている^[1]。しかし、既報の表情認識システム^[2]は、動画画像を入力し、表情認識を行うまでにいくつかの手動操作が必要であり、実用化のための課題の一つとなっていた。

本研究では表情認識の実用化を目指し、発声時の動画画像入力から表情認識までの処理の自動化を行った。

2. 開発環境

- ・ OS : Windows XP
- ・ 使用したプログラム言語 : Visual C++6.0
- ・ DELL Optiplex GX280 CPU : Pentium 3.6GHz メインメモリ : 2GB

3. 処理概要

入力画像を AVI ファイルに保存

8 ミリビデオの温度動画画像を入力し、20 秒ずつ AVI ファイルに保存する。

静止画像を採取

保存された順に AVI ファイルを開き、その中の音声データをもとに発声直前、「ta」発声時、「ro」発声時の 3 枚の静止温度画像を採取する。AVI ファイルに保存する際、処理の関係上、連続する AVI ファイルの間に数百ミリ秒の保存できない時間が生じ、そこに発声がある場合、正しいタイミングで静止画像が採取できないため、ファイルの始まりと終わりに現れた発声時の表情については認識の対象としないこととした。

表情認識^[1]

1. 採取した静止温度画像の顔領域の位置、回転角度、サイズの補正処理を行ってから、表情の特徴が出やすい顔部品領域を抽出し、平均無表情画像との差分をとり、純粋な表情変化のみを抽出する。
2. 得られた差分画像に二次元離散的コサイン変換(2D-DCT)を施す。
3. 3 枚の画像から得た DCT 特徴量から、特徴ベクトル作成ルールに基づいて、高い確率で表情を分類できる DCT 特徴量を選び、特徴ベクトルとする。
4. 学習画像から得た特徴ベクトル空間に未知画像の特徴ベクトルを入力し、最小距離識別法による表情認識を行う。

既報の手法^[1]では AVI ファイルの保存にビデオ編集ソフトウェアを用いていた。本法では DirectShow を用いることで、プログラムによって AVI ファイルを柔軟に扱えるようにした。また、 から の処理を同時並行で行い、入力から表情認識までの時間を短くした。

4. 実験

被験者に中立的な言葉(日本名「taro」)を 5 表情(怒り、喜び、無表情、悲しみ、驚き)において発声してもらった。このときの温度動画像を入力し、前述した処理によって表情認識実験を行った。1 表情に対して 30 回の発声データを採取し、このうち 20 サンプルを学習用、10 サンプルを認識用に用いた。1 回の発声から得た静止画像は発声直前、「ta」発声時、「ro」発声時の 3 枚である。特徴ベクトルは 19 次元となった。

5. 結果

既報の手法による表情認識結果^[1]を表 1 に、本法による結果を表 2 に示す。本法では、入力した温度動画像は既報の研究^[1]と同じであるが、そこから得た AVI ファイルは異なっており、静止画像取得のタイミングも異なっている。

表 1. 既報の手法による認識率^[1]

		入力表情				
		怒り	喜び	無表情	悲しみ	驚き
認識	怒り	70	10			
	喜び		90			
	無表情			90		
	悲しみ	30		10	100	
	驚き					100

表 2. 本法での認識率

		入力表情				
		怒り	喜び	無表情	悲しみ	驚き
認識	怒り	78				
	喜び	22	88			
	無表情			100		
	悲しみ		11		100	
	驚き					100

既報の手法^[1]による認識率は平均 90 %であり、本法での認識率は平均 93 %となった。

認識用に用いた 5 表情 50 サンプルの発声データのうち、AVI ファイルの最初、もしくは最後に発声データがあり、表情認識しなかった割合は 16 %であった。なお、認識結果が出るまでの時間は 45 秒であった。このうち AVI 保存に 21 秒、静止画像採取に 12 秒、表情認識に 12 秒かかっていた。

6. 結言

既報の表情認識手法^[1]を用いて、入力から認識まで自動でできるシステムを開発した。その結果、表情認識実験では平均 93 %の認識率を得ることができた。今後はこのシステムを用いて多数の人の表情認識を行い、実用化への課題を抽出する予定である。

参考文献

- [1] 池添史隆：温度画像を用いた発声時の表情認識;京都府立大学大学院人間環境科学研究科環境情報学専攻 修士論文, 2006.