

発声時の温度顔画像からの顔部品抽出に関する研究

人間環境科学研究科 環境情報学専攻 環境情報システム学分野

胡玲琴

キーワード：温度画像，自動2値化，アフィン変換，テンプレートマッチング

1. 緒言

将来、ロボットが人間と共に平和に働き、暮らすためには、人間同士が行う程度の性能で、ロボットの頭脳としてのコンピュータに人間の感情や心理状態を理解させる必要がある。

これまでは主に可視光画像を用いた表情認識の研究が行われてきた。しかし、可視光を用いた従来の研究では、異なる照明条件で、正確な表情認識を行うことは一般には困難に思われる。

この課題を解決するため、赤外線による2次元温度分布画像を用いる手法が研究されてきた。

表情認識を行うには、表情の特徴量を測定するために、まず顔部品を正確に抽出することが必要である。発生時など動きが伴う場合には、一般に正確な顔部品抽出が困難である。この課題を解決するため、本研究では「喜び」、「悲しみ」、「驚き」、「怒り」、「無表情」の5種類の感情画像に対して、発声前と「タ」、「ロー」発声時の温度画像から顔部品抽出を行う方法を検討した。

まず、従来の方法を用いて、「タ」、「ロー」の発声時の温度顔画像からの顔部品抽出を行った。例を図1.1~1.4に示す。本研究では顔部品抽出失敗(図1.2~1.4)を減らせることを目的とする。

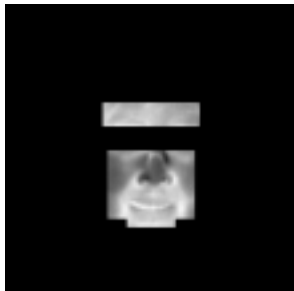


図1.1 抽出成功例

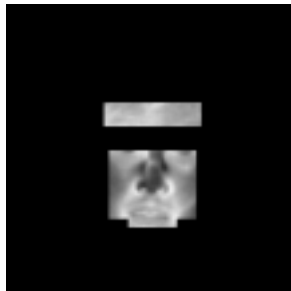


図1.2 抽出失敗例1

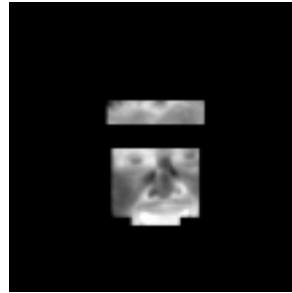


図1.3 抽出失敗例2



図1.4 抽出失敗例3

2. 温度画像における顔部品領域の抽出法の概要

本法の温度画像からの顔部品領域抽出処理フローを図2.1に示す。入力画像例を図2.2に示す。

温度画像からの顔領域の抽出には、温度画像の入力後、まず、温度画像の文字を取り除き、髪の毛領域の濃度変動を吸収するため、最大値フィルタを施し、自動閾値法で2値化により、皮膚領域候補を抽出し、その皮膚領域から微小面積の孤立点を消去した後、水平方向のフェレ径を測る。次に顔領域の重心を決定する時に及ぼす髪の毛や首の影響を軽減するため、顔領域の上部と下部から、水平方向の画素数がフェレ径の20%に達するまでの領域を削除してから、垂直方向のフェレ径と重心を測る。その重心を標準位置に合わせた後、顔幅を標準サイズに合わせる。その後テンプレートマッチングで顔領域を探索して、顔部品領域抽出する。

顔領域の重心位置を基準に、心理学における表情分析の研究及び温度画像処理による表情認識の研究を参考にして、表情変化を解析する領域として、図2.3に示す各矩形領域(眉毛領域,左頬領域,鼻頭部領域,右頬領域,口左側領域,口・顎領域,口右側領域)を抽出する。

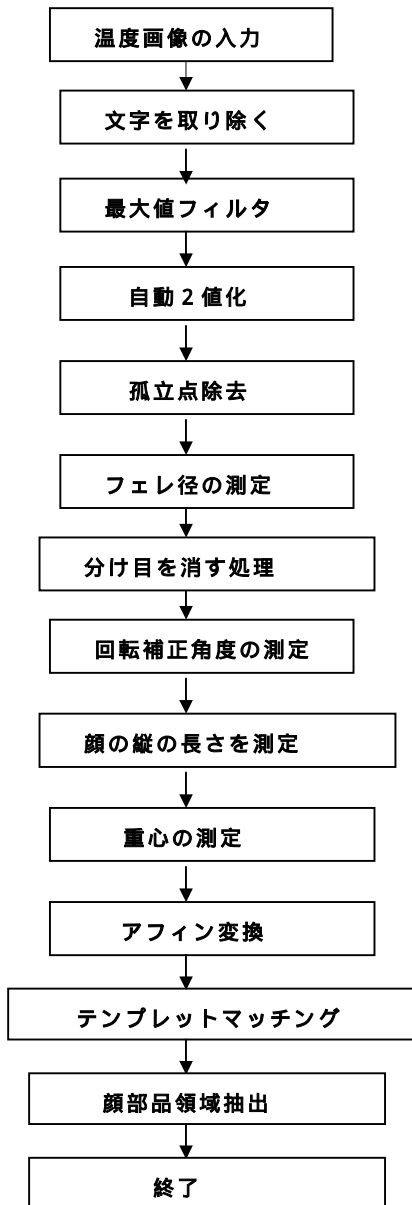


図 2.1 画像処理フロー



図 2.2 入力画像例

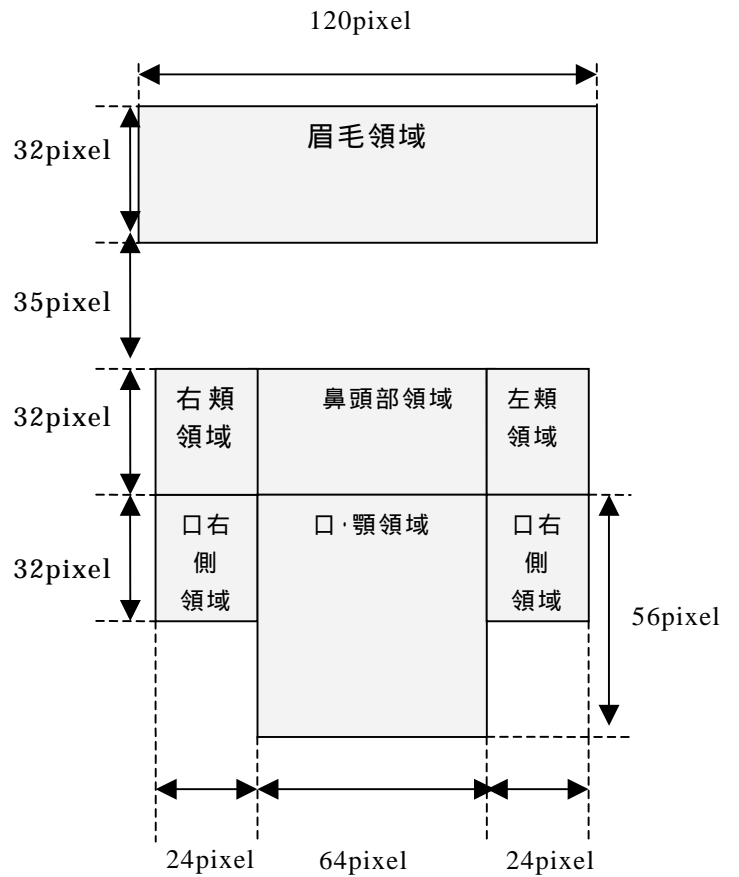


図 2.3 温度画像における顔部品領域

3. 本法の特徴

心理学での表情分析の研究を参考にして、「喜び」、「悲しみ」、「驚き」、「怒り」、「無表情」の5つを識別用の表情として選択した。全入力画像において、顔が回転しているものが多く、この回転の補正が必要である。そこで、まず、回転角度の測定方法を検討した。

3.1 回転補正角度の測定

まず、モーメント特徴量を用いて、対象領域の方向角度が次の式で計算できる。

$$\tan^2 \theta + \frac{M(2,0) - M(0,2)}{M(1,1)} \tan \theta - 1 = 0 \quad (1)$$

$$M(p,q) = \sum_{i,j} i^p j^q f_{ij} \quad (2)$$

ここで $\{f_{ij}\}$ は対象領域内で1をとり、対象領域外で0を取る2値画像である。(1)式の

θ は慣性主軸方向の角度である。計算した θ 値を元に顔領域の回転の補正角を求めることができる。この方法を種々の画像に適用したところ、顔の温度や髪型の変動が θ の値に大きな影響を与えることがわかった。そこで、新たな角度補正方法を検討した。計算方法を図 3.1 に示す。

温度画像の 2 値化後、顔上下の所定の巾以下の領域を除去し、上辺と下辺の中心座標（各々、 u_c, l_c と表記）を結び、図 3.1 のようにして、顔の角度 θ を求める。 $\theta = 90^\circ$ を回転なしとみなし、 90° からのずれ度を補正角度として、回転の補正を行う。

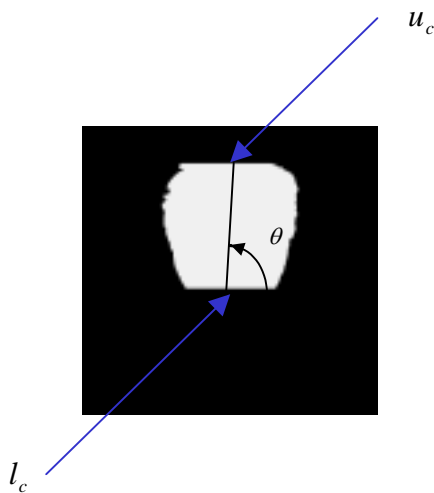


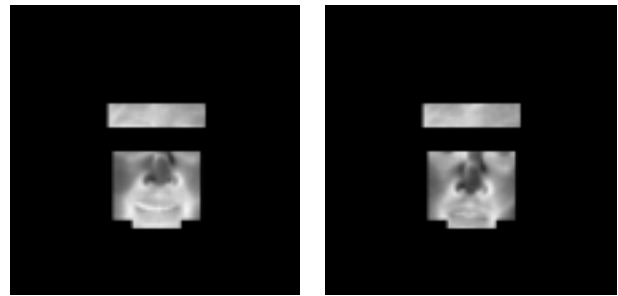
図 3.1 回転角度の計算法

この補正角度を使って、全画像に対して、回転補正を行った。そして、全画像（発声前、「タ」、「ロー」各発声時）に対して、以下のように処理結果を図 3.2 に示す、4 パターンに分類した。各パターンの内訳は、A：140 枚(31.1%)、B：60 枚(13.3%)、C：124 枚(27.6%)、D：126 枚(28.0%)であった。

次に図 3.2 の B, C に例示した下方向へのずれを補正する手段として、テンプレートマッチングを検討した。

対象物テンプレートが画像パターンとして表され、探索する画像領域の対象部分との類似度を調べることによって一致する位置を求める方法がテンプレートマッチングである。

- A. 正確に抽出が行われている画像例 B. 下にわずかにずれ、かつ、わずかに顔が回転している画像例



- C. 下に大きくずれ、かつ、わずかに顔が回転している画像例 D. 顔が大きく回転している画像例



図 3.2 処理結果の分類

このテンプレートマッチングの代表的な方法のひとつに残差逐次検定法 (SSDA: Sequential Similarity Detection Algorithm) がある。本研究では SSDA 方法を用いた。上下方向の位置補正を目的としたため、1次元のテンプレートを作成し、テンプレートマッチングを行った。各座標におけるテンプレートとの残差が最小になるような対象領域の平行移動量 (k_m) を求めて、対象領域を $-k_m$ だけ平行移動することにより上下方向の位置補正を行う。対象領域の移動量 k を 0~50 とした、残差 $s[k]$ は、

$$s[k] = \sum_j (s[j][k] - a[j])^2 \quad (3)$$

で求まる。ここで、 $s[j][k]$ は k だけ平行移動した対象画像、 $a[j]$ はテンプレート画像の濃度の各々横方向の和である。テンプレート画像には、「無表情」画像の濃度平均画

像を用いた。

画像は大きさ 480×480 画素で、BMP 形式ファイルである。

テンプレートマッチングを施した結果、顔部品抽出結果が以下ようになった。

A . 正確な抽出画像 : 200 枚(44.4%)(図 3.3)

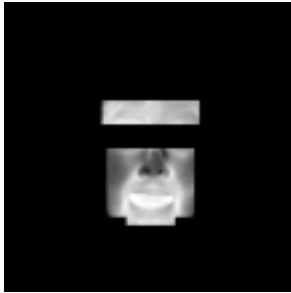


図 3.3 正確な抽出画像例

B . 少し回転のずれがある画像 : 124 枚 (27.6%)(図 3.4)

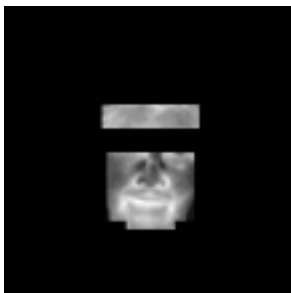


図 3.4 少し回転のずれがある画像例

D . 大きな回転のずれがある画像 : 126 枚 (28.0%)(図 3.5)

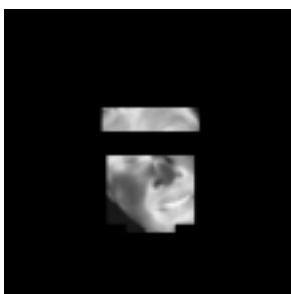


図 3.5 大きな回転のずれがある画像例

4 . 結論

発声時の温度顔画像から顔部品抽出を行う方法を検討した。実験により、今回提案

した回転補正とテンプレートマッチングを使うことにより、顔部品抽出正確率が向上した。今後は、回転補正の精度の向上策の検討を行い、本法を表情認識法の精度向上に活用する。

参考文献

- 1) Y.Yoshitomi, N.Miyawaki, S.Tomita and S.Kimura,“Facial Expression Recognition Using Thermal Image Processing and Neural Network”,Proc. of 6th IEEE International Workshop on Robot and Human Communication,(1997),380-385.
- 2) Y.Yoshitomi,M.Murakawa and S.Tomita,“Face Identification Using Sensor Fusion of Thermal Image and Visible Ray Image”,Proc. of 7th IEEE International Workshop on Robot and Human communication,(1998),449-455.
- 3) 吉富康成,「赤外線画像を用いた人間情報解析」,システム/制御/情報,42(1998),535-543.
- 4) 河野貴子, 吉富康成, 富田重幸,「温度画像と可視光画像の情報統合による表情認識」,情報処理学会九州支部研究会報告,(1999),55-62.
- 5) 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.
- 6) T.Ikeda, Y.Yoshitomi, T.Kitazoe and S.Kim, “Integration System for Emotional States Using Visible and Infrared Face Image and Voice”, Proc. of 4th International Symposium on Artificial Life and Robotics, 1(2001),54-57.