5T-2

口唇の動きを用いた機器操作インターフェイスシステム

山下　良博　　田村　仁†
日本工業大学情報工学部†

1.はじめに

音声認識を使ったソフトやハードが既に多数存在するが、雑音の多い場所や電気機器のノイズの影響が出る場所では認識率が下がる。また市販の音声認識ソフトを使用する際には、雑音のない静かな場所や、高性能なマイクを使って補う可能性がある。しかしカメラからの画像を使用してより発話された内容を認識出来れば、音声に依存しない処理が可能となり上記のような環境下でも認識する事が可能である。そこで、本研究では、上記のような環境での機器操作を行うのが目的である。

読唇に関する研究では、顔を正面から撮影するカメラアングル画像を用いる手法[1][3]や、顔を側面から撮影するカメラアングル画像を用いる手法[2]が提案されている。しかし、これらの多くの研究で使用されている撮影アングルは、人が読唇をする際により効果的な撮影方法を行っていない。そこで本研究では、人が読唇する際により効果的に読唇出来る撮影方法で、唇の奥行きや下の奥行きの特徴を使用して読唇処理を試みる。

3 読唇手法

3.1 対象とする単語

本研究では、機械やロボットなどの機器を操作で使用できる汎用的な 6 コマンド表 1 を対象とする。

3.2 斜めからの撮影アングル特徴

読唇は、唇の他顔全体のさまざまな特徴を使用するしかし、正面から撮影する場合口が立体的に見えず、下や唇の奥行きが分かりづらい他、頬が陰影の影響で頬の動きがわかりづらい。しかし斜めからのアングルの場合唇が立体的(図 8)に見えるので舌や唇奥行き特徴が取れる。

図 1 カメラ固定器具

図 2 撮影の様子

3.3.撮影方法

本研究では、機械などの機器操作を行うのが目的なので、対象機器への追走や目視を行う際に、壁などにカメラを設置してしまうと撮影範囲外にいってしまう可能性がある。そこで対象者にヘッドセットのようにカメラをセットする方式を用いる。人が読唇を行う際には、対象者を斜め 45 度から顔全体を撮影するアングルが最適である[4]。そこで(図 1)のようにカメラ(Logicool 社製 Qcam Pro 4000 30 万画素 CCD 最大フレームレート 30fps 解像度 320×240 視野角上下 20 度、左右 30 度　最短撮影距離 16cm)顔正面から 45 度の場所にセットし、(図 2)のように撮影を行う。

3.3 単語の推定方法

表 3 のような感情が表現されない単語レベルでは、目や眉毛はなどの特徴は必要ないため、全体画像から唇を抽出し、抽出した唇画像から面積及び唇の移動量を、連続フレーム単位での DP マッチングで比較し推定を行う。

表 1　コマンド読唇

1	起動
2	停止
3	前進
4	後退
5	左
6	右

Equipment operation interface system that used movement of lips of mouth
Yoshihiro YAMASHITA, Hitoshi TAMURA
†Department of Computer and Information Engineering, Faculty of Engineering, Nippon Institute of Technology.

4.唇の抽出方法

4.1 顔を抽出

撮影画像から CIE Lab イメージを使用して、画像図(図 3)の中心から色を取得し、取得した色の近似値を抽出し大まかな顔の部分の抽出図 4-2 を行う。

4.2 唇周辺を抽出

抽出した顔(図 4)から Y 方向 Sobel フィルタを行い、2 値化した画像図(5)から一定の範囲ないに白画素が密集した領域を切り抜く。(図 5)

4.3 唇の抽出

唇周辺画像(図 6)を CIE LCh 表色系で、色相 h だけを取り出し、色相 h の値が一定以下の場所の抽出を行う。(図 7)

4.4 唇の特徴抽出

抽出した唇から、1 フレームの前の比較した上唇と下唇の先端の垂直及び水平移動量(図 8)や、唇や口内面積を連続フレーム単位での DP マッチングで比較し単語候補順位を求める。

図 3 撮影サンプル画像

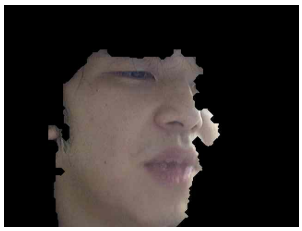

図 4 顔の抽出横

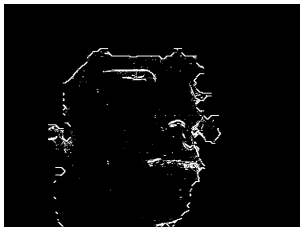

図 5　Y 方向 Sobel フィルタを適用後 二値化を適用

図 6 唇周辺を抽出

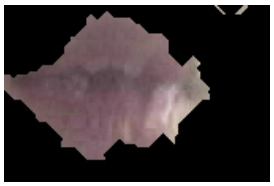

図 7 唇のみを抽出

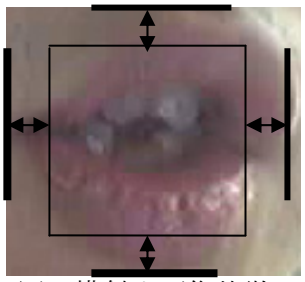

図 8 横斜め画像特徴

5.評価実験結果

実験行った場所は室内とし室内照明の光源の位置や向きを固定して、髭などの生やしてない大学 4 年生 5 人を対象とした結果、6 コマンドを人間に依存しないが認識できた。しかし固定具が重く大きすぎて不便という意見や、自分で 45 度の角度にカメラを持っていけないという意見が全員から寄せられた

また斜めからの撮影画像から得られる特徴を完全に使用していないので使用できる処理を目指す。カメラ固定器具が重く大きすぎるため、改良しなければならない。機器操作に実際に使用して評価実験が必要である。

認識させる単語を増やす際に、単語毎に唇のみの特徴で認識が可能か調べる他、感情によって唇がどの程度変化するか調べる必要がある。人が読唇する際に認識しやすい方法で行ったが、画像処理に最適なアングルなのか検討する必要がある。多くの特徴を使用できるため個人認証などに応用できる可能性がある。

6.おわりに

本研究では、人が読唇する時に最適なカメラアングルから得られる情報からの読唇処理インターフェイスを構築した。

参考文献

[1] 斎藤　剛史　小西亮介　“トラジェクトリ特徴量に基づく単語読唇”電子情報通信学会論文誌 2007/4 Vol. J90-D No.4 pp.1105-1114

[2] 井出　寿登　小越　康宏　荒木　哲郎“横顔口唇動画像における注目点追跡による読唇手法の提案”人工知能学会第 20 回全国大会(JSAI2006) 1g2-1 pp.1-2

[3] 中田　康久　安藤　護俊　“色抽出方と固有空間を用いた読唇処理”電子情報通信学会論文誌　2002/12 Vol.J85-D-II No.12 pp.1813-1822

[4] 片山　磁友　徳永　努　興口　晴久　横山　嘉徳　“聴覚障害者の受講システムにおける発話者の最適カメラアングル”電子情報通信学会総合大会後援論文集　Vol.2000 年情報・システム No.1(20000307) pp.224