口領域の形状特徴と低解像度画像を特徴量とした口唇動作認識

Mouth Motion Recognition Using Shape Features and Low-resolution Images of Mouth Region

○　学　高橋真人　(中央大／JST CREST)　　髙山良裕　(中央大)　　永易武　(中央大)
正　寺林賢司　(中央大／JST CREST)　　正　梅田和昇　(中央大／JST CREST)
○Masahito TAKAHASHI, Chuo University, CREST, JST, takaha@mech.chuo-u.ac.jp
Yoshihiro TAKAYAMA, Chuo University, takayama@mech.chuo-u.ac.jp
Takeshi NAGAYASU, Chuo University, nagayasu@mech.chuo-u.ac.jp
Kennji TERABAYASHI, Chuo University, CREST, JST, terabayashi@mech.chuo-u.ac.jp
Kazunori UMEDA, Chuo University, CREST, JST, umeda@mech.chuo-u.ac.jp

An intelligent room which recognizes gestures and support operators is required in various places in recent years. In this paper, we propose a method to recognize mouth motion from images. The proposed method uses shape features and low resolution images of mouth region and recognizes mouth motion for indicating a target object in an intelligent room. DP matching is applied to low resolution images. Several experiments are performed to demonstrate the effectiveness of the proposed method.

Key Word: Intelligent Room, Mouth Motion Recognition, DP Matching, Image Processing

1. 序論

近年，画像処理や音声認識などの技術を用いて部屋全体を知能ロボット化したインテリジェントルームが病室，福祉施設，オフィス，家庭内など様々な場所で期待されている．我々は，図 1 に示すように，部屋の四隅にカメラを設置し，操作者のジェスチャを認識して家電製品の操作を行うインテリジェントルームを構築している[1]．インテリジェントルームにおける家電機器操作の方法の一つとして DP マッチングを用いた口唇動作認識を提案してきた[2]．この手法では，特徴量として低解像度画像の画素値を用いていた．本研究では，特徴量に低解像度画像の画素値だけでなく口領域形状も加えることによって認識率の向上を図る．

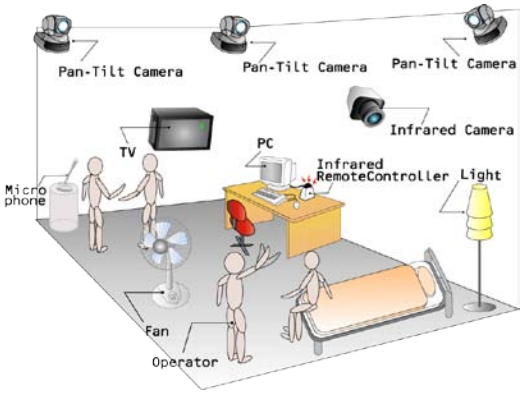
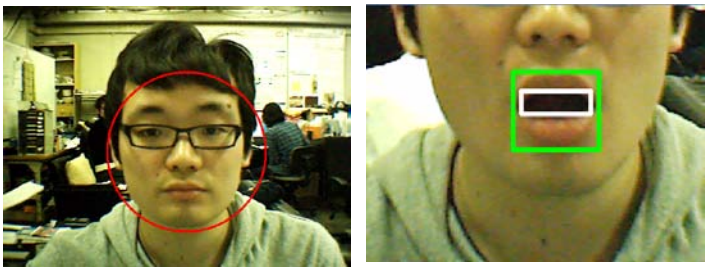


Fig.1 Conceptual figure of our intelligent room

2. 唇と開口領域の抽出

口唇動作を認識するためには口の位置を知る必要がある．

まず, OpenCV を用いて入力画像中の顔領域を検出する[3]．図 2(a)に顔領域を検出した画像の例を示す．次に，検出した顔領域の中から口の位置の検出を行う．唇の色が赤く，口を開けたときに口の中は暗くなることを利用する．具体的には，取得したカラー画像を HSI 画像に変換し， H(色相)画像と I(明度)画像を得る．そして，H 画像において閾値内の画素値領域を選択することで唇領域を抽出する．また，I 画像において閾値内の画素領域を選択することで開口領域を抽出する．抽出された唇を内包する領域（太い枠）および開口領域（細い枠）の例を図 2(b)に示す．



(a) Face region　　(b) Mouth region

Fig.2 Extraction of mouth region

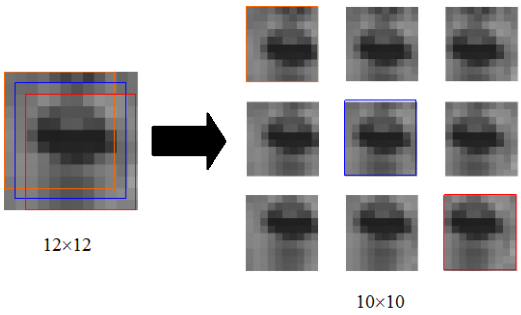


Fig.3 Nine 10×10 sub-images

3. 口唇動作の認識

口唇動作を認識するため，認識範囲を算出する．唇領域の重心を求め，検出された顔の大きさに対応して認識範囲を決定する．認識範囲の位置は口唇動作の認識中は固定する．

次に，決定された認識範囲を 12×12 に低解像度化する．低解像度化には，データ数を減らすと共に，同じ口唇動作であっても毎回発生する変化を少量なら吸収できる利点がある．低解像度化では吸収しきれない口の位置のズレが，口唇動作中に発生することが考えられるため, 12×12 の低解像度化画像から，図 3 に示すように 9 パターンの 10×10 の画像を切り出す． 9 パターンの画像のうち，事前に登録されているモデルデータとの距離が最小になる画像を選択し，DP マッチングにより口唇動作を認識する．特徴量には，低解像度画像の画素値と口領域の形状を用いる．口領域の形状として，図 4 に示す唇領域の高さ h1，幅 w1，開口領域の高さ h2，幅 w2 を利用する[4].

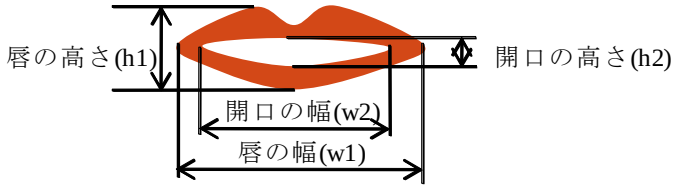


Fig.4 Shape features of mouth region

4. 実験

提案した手法の有用性を検証するため口唇動作認識の実験を行い，認識率を評価した．

4.1 実験システム構成

実験システムは CCD カメラ UCAM-E130(ELECOM)，画像処理ボード PicPort-Color (Leutron Vision), 画像処理ソフト (OpenCV)，画像処理ソフト HALCON ver.7.0(MVTec)，PC (Pentium4 3.2GHz)から構成される(図 5 参照).

Fig.5 Experimental setup

4.2 認識実験

カメラを被験者の顔とほぼ同じ高さ，顔前方 300[mm]付近に設置し，蛍光灯下で 2 名の被験者 A，B に対し実験を行った．モデルデータは事前に被験者 A が登録しておいた「Fan」,「TV」,「Light」,「Up」,「Down」の 5 つの単語を用いた．各口唇動作を 100 回ずつ行い正しく認識する確率を調べた．表 1 に示す各特徴量を用いた認識結果を表 2，3 に示す．

表 2，3 から，画素値のみを特徴量とした G0 において，被験者 A の認識率は 59.6%，被験者 B は 35.6%である．モデルデータを登録した被験者 A とモデルデータを登録していない被験者 B の認識率に大きな差があることがわかる．これは，唇の色に個人差があることが原因であると考えられる．また，被験者 A，B 共に口唇動作の種類により，認識率にばらつきがある．

唇領域のアスペクト比を特徴量とした G1 は，被験者 A の認識率が 77.0%，被験者 B の認識率は 61.0%と高く，被験者 A，B の間の認識率の差も小さい．

開口領域のアスペクト比を特徴量とした G2 は，被験者 A，B の間の認識率の差は最も小さいが，被験者 A の認識率は 43.4%，被験者 B の認識率は 42.0%と認識率自体は低い．これは，開口領域は歯や舌が含まれてしまうことがあるため，唇領域に比べ唇の抽出が正確に行えていないことなどが考えられる．

本研究で提案した口領域の形状特徴と低解像度化を特徴量とした G4 は，被験者 A の認識率は 81%，被験者 B の認識率は 62%前後と最も高く，被験者 A と B の認識率の差も小さい．この理由として，G0，G1，G2 の特徴量は認識しやすい単語がそれぞれ異なり，それらを組み合わせた特徴量である G4 はそれらの長所を効果的に利用できた結果だと考えられる．

「Fan」の単語は他の 4 つに比べ比較的認識率が高い．これは「Fan」が他の 4 つの単語と比べ発音時間が短く，そのため他の 4 つの単語との違いがはっきりと表れた結果だと考えられる．

「Light」の単語は被験者 A，B の認識率の差が最も大きい．これは「Light」の単語は他の 4 つの単語に比べ口を開くことが多い単語であるため，喋り方による個人差が顕著に表れた結果だと考えられる．

Table 1 Features of using mouth motion recognition

特徴量	使用する値
G0	画素値
G1	b1/h1(唇の幅÷唇の高さ)
G2	b2/h2(開口の幅÷開口の高さ)
G3	G1 と G2 を一つの特徴量としたもの
G4	G0 と G1 と G2 を一つの特徴量としたもの

Table 2 Result of testee A[%]

認識単語 / 特徴量	Fan	TV	Light	Up	Down	平均
G0	79	71	31	48	69	59.6
G1	84	68	72	70	91	77.0
G2	70	13	55	26	53	43.4
G3	76	55	49	75	48	60.6
G4	92	81	70	81	79	80.6

Table 3 Result of testee B[%]

認識単語 / 特徴量	Fan	TV	Light	Up	Down	平均
G0	48	19	2	8	71	35.6
G1	82	66	54	41	62	61.0
G2	66	17	34	78	15	42.0
G3	73	44	53	83	24	55.4
G4	81	75	64	38	56	62.8

5. 結論と展望

口領域の形状特徴と低解像度画像を特徴量とする口唇動作認識を提案し，実験により本手法の有効性を確認した．今後の展望としては，より正確な唇と開口領域の抽出が挙げられる．口領域形状の抽出に色情報だけでなくエッジ抽出や動的輪郭法などを用いて正確に抽出することで，認識率を向上させることができると考えられる．認識率向上には，認識手法を DP マッチングから HMM に変更することなども挙げられる．

文　献

[1] 入江耕太，若村直弘，梅田和昇，“ジェスチャ認識に基づくインテリジェントルームの構築”，日本機械学会論文集 C 編, Vol.73, No.725, pp.258-265, 2007.
[2] 中西達也，寺林賢司，梅田和昇，“DP マッチングを用いた口唇動作認識”，電気学会論文誌 C，Vol.129，No.5，pp.940-946，2009.
[3] 怡土順一，OpenCV：“物体検出.”
http://opencv.jp/sample/object_detection.html .
[4] 齊藤 剛史，小西 亮介，“トラジェクトリ特徴量に基づく単語読唇”，信学論，Vol.J90-D，No.4，pp.1105-1114，2007.