

DP マッチングを用いた口唇動作認識 Mouth Motion Recognition Using DP Matching

○中西達也 (中央大学/JST CREST) 正 梅田和昇 (中央大学/JST CREST)
○Tatsuya Nakanishi, Chuo University, CREST, JST, tnakani@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. This paper proposes a mouth motion recognition technique for indicating a target object in an intelligent room. DP matching is applied to low resolution images. Effectiveness of this technique is verified by experiments.

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

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的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[1]。具体的には部屋や病室に様々なセンサやアクチュエータを取り付け、ジェスチャ認識などにより人間の意図を理解し要求に応じるといった支援を行う。

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[2]。操作対象の選択には指差し方向認識を用いているが、安定性が十分でない。本研究では対象選択のための新たな手法として、DP マッチングを用いた口唇動作認識手法を提案する。

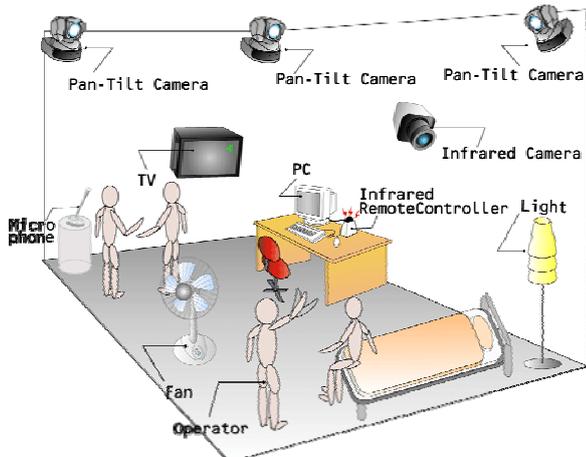


Fig.1 Conceptual figure of our intelligent room

2. 認識範囲決定手法と低解像度化

口唇動作認識を行う前に口の位置を認識する必要がある。口を開けたときに口の中は暗くなる。これを利用して口の位置の検出を行う。まず、入力されたカラー画像を HIS 画像に変換し、I (輝度)画像を得る。画素値に閾値を設定し、口を大きく開けることによって生じる、画素値が閾値以下の部分を口の中と認識する。口の中の領域の重心を求め、上下左右 60pixel を認識範囲に決定する(図 2 参照)。決定された認識範囲を 12×12 に低解像度化する(図 3 参照)。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。認識範囲の位置は口唇動作の認識中は固定する。

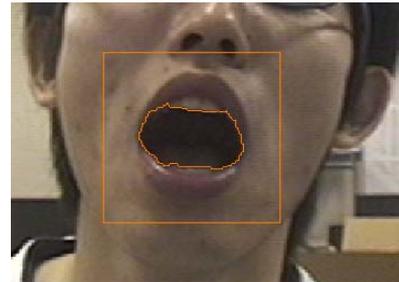


Fig.2 Region for recognition



Fig.3 Low resolution image

3. 口の位置ズレ補正手法

前節で述べた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。これに対応するため、12×12 の低解像度化画像から、図 4 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て DP マッチングを行う。

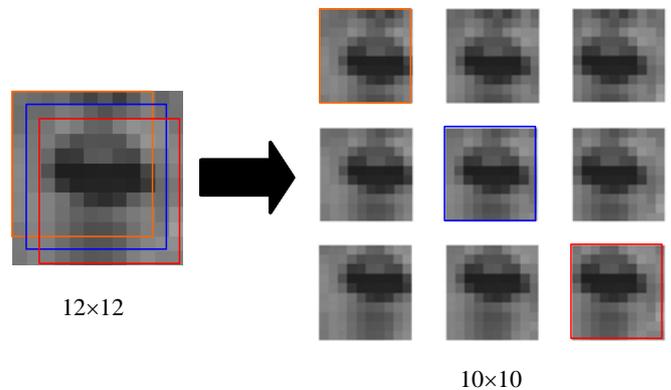


Fig.4 Nine 10×10 sub-images

4. DP マッチング

DP(Dynamic Programming)マッチングとは入力された未知のパターンデータを事前に登録したモデルデータと比較、照合し類似度を計算する手法のことである[3]。これにより入力されたジェスチャが登録されたどのジェスチャであるかを判断する。この手法は時系列データの取扱いが可能のため動画認識や音声認識などに用いられる。また入力データ数と標本データ数が異なる場合にも有効である。

ある一点の画素を考える。n番目の入力データの画素値 $PixelValue[n]$, m番目のモデルデータの画素値 $mPixelValue[m]$ に対し、距離 $PixelValuePD[m][n]$ ($m=0,1,2,\dots,M$ $n=0,1,2,\dots,N$)を次式で与える。

$$PixelValuePD[m][n] = \frac{|mPixelValue[m] - PixelValue[n]|}{\sqrt{(mPixelValue[m])^2 + (PixelValue[n])^2}} \quad (1)$$

n番目の入力データとm番目のモデルデータの全ての画素値の距離を求め、その平均を平均距離 $PD[m][n]$ とする。始状態(0,0)からスタートし終状態(M,N)に移動するまでの経路距離 $TotalDistance[M][N]$ が最小になる最適経路を求める。ただし $TotalDistance[m][n]$ は

$$TotalDistance[m-1][n-1]+2PD[m][n]$$

$$TotalDistance[m][n-1]+PD[m][n]$$

$$TotalDistance[m-1][n]+PD[m][n]$$

の中で一番小さい値を選択する(図5参照)。

得られた経路距離をモデルデータ数Mと入力データ数Nの和で割りマッチング距離 $Value[p]$ を求める。

$$Value[p] = \frac{TotalDistance[M][N]}{M+N} \quad (2)$$

これを登録したモデルジェスチャ全てにおいて行う。全モデルジェスチャの中で $Value[p]$ が最小となったものを入力データと同じであると判定する。

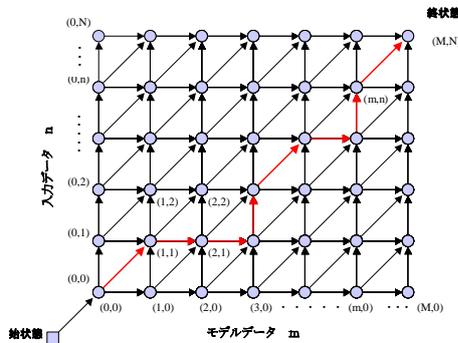


Fig.5 Optimal path of input data and model data by DP matching

5. 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率の計測実験を行った。

5.1 実験システム構成

実験システムは CCD カメラ EVI-D30 (SONY), 画像処理ボード PicPort-Color (Leutron Vision), 画像処理ソフト HALCON ver.7.0 (MVTec), PC (Pentium4 3.2GHz)から構成される(図6参照)。



Fig.6 Experimental setup

5.2 認識率測定実験

カメラを被験者の顔とほぼ同じ高さ、顔前方 400[mm]付近に設置し、蛍光灯下で3名の被験者 A, B, C に対し実験を行った。モデルデータは事前に各動作者が登録しておいた「ファン」、「テレビ」、「エアコン」、「ライト」、「ビデオ」の5単語を用いた。各口唇動作を20回ずつ行い正しく認識する確率を調べた。まず各被験者共に自身のモデルデータを登録している状態で実験を行った。結果を表1, 2, 3に示す。また、被験者 B, C のモデルデータを登録している状態で被験者 A が実験を行った。結果を表4に示す。

表1より被験者 A の口唇動作の認識率は80%前後と比較的高い値である。これに対して表2, 3の被験者 B, C の認識率は口唇動作の種類によらず認識率のばらつきがあることを示している。これは、口唇動作の個人差のために、認識し易い口唇動作が違ってくるとためと考えられる。また、表4より他人のモデルデータを用いた場合では認識率が大きく低下することが分かる。口唇動作には個人による違いがあるので、使用する人ごとにモデルデータを登録することが必要だと言える。

Table1 Result of testee A

	ファン	テレビ	エアコン	ライト	ビデオ	認識率
ファン	19	0	0	0	0	95%
テレビ	2	15	0	0	3	75%
エアコン	3	0	15	2	0	75%
ライト	4	0	0	16	0	80%
ビデオ	3	0	0	0	17	85%

Table2 Result of testee B

	ファン	テレビ	エアコン	ライト	ビデオ	認識率
ファン	17	0	1	0	2	85%
テレビ	1	5	13	0	1	25%
エアコン	0	0	12	6	2	60%
ライト	0	0	1	9	10	45%
ビデオ	0	1	0	0	19	95%

Table3 Result of testee C

	ファン	テレビ	エアコン	ライト	ビデオ	認識率
ファン	13	0	0	7	0	65%
テレビ	8	11	0	1	0	55%
エアコン	9	0	0	11	0	0%
ライト	0	0	0	20	0	100%
ビデオ	7	4	0	0	9	45%

Table4 Result of testee A using model data of testee B and C

	被験者Bのモデルデータ使用	被験者Cのモデルデータ使用
ファン	60%	5%
テレビ	25%	5%
エアコン	10%	0%
ライト	0%	60%
ビデオ	30%	35%

6. 結論と展望

本研究では口唇動作認識のシステムを構築し、認識率を実験により計測した。さらに、他人のモデルデータを用いた場合の認識率を実験により検証した。

今後は認識率の向上を目指し、インテリジェントルームでの家電の選択方法としての実装を行う。

文献

- [1] 森武俊, 野口博史, 佐藤知正, “センシングルーム ー 部屋型日常行動計測蓄積環境 第2世代ロボティックルーム ー”, 日本ロボット学会誌, Vol.23, No.6, pp.665-669, 2005.
- [2] 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.
- [3] 西村拓一, 向井理朗, 野崎俊輔, 岡隆一, “低解像度特徴を用いた複数人物によるジェスチャの単一動画像からのスポッティング認識”, 電子情報通信学会論文誌, Vol.J80-D-II, No.6, pp.1563-1570, 1997.