

# 動的輪郭モデルを用いた読唇母音認識システムの開発

和谷 優一<sup>†</sup> 中邨 寛<sup>†</sup> 川村 尚生<sup>‡</sup> 菅原 一孔<sup>‡</sup>

<sup>†</sup>鳥取大学 大学院 工学研究科 知能情報工学専攻

<sup>‡</sup>鳥取大学 工学部 知能情報工学科

## 1. はじめに

工場内のような高雑音環境下でロボットに音声により命令を伝達することは困難である。唇形状をカメラで撮影し母音を認識すれば雑音の影響を受けず命令を伝えることが可能となる。本稿では読唇による母音認識システムを考え、FPGA上にハードウェアとして実現する。ハードウェア化することにより高速処理とシステムの小型化に期待している。

## 2. システム構成

システム全体の構成を図1に示す。動作の流れは、まずNTSCビデオカメラを用いた映像処理装置からアナログ映像信号を取り込み、デコーダを介してFPGAボード上のメモリに格納する。格納した画像に対して動的輪郭モデルを適用し得られた唇領域を抽出した動作点の座標データを正規化し、母音認識を行う。また、唇領域抽出結果はエンコーダを介してモニタに表示する。

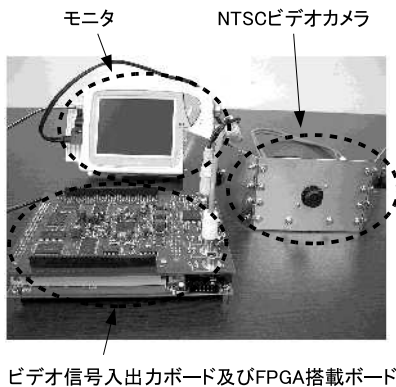


図 1: システム全体図

## 3. 唇形状抽出とその正規化

メモリに格納された画像に対して動的輪郭モデルを適用し、唇領域を抽出した動作点の座標データを得る。従来の差分フィルタ等を用いて輪郭を求める手法では、画像中、全画素に対して処理を施す必要がある。この点、動的輪郭モデルでは画像を格納しているメモリへのアクセスが動作点が移動する画素のみですむため、メ

モリへのアクセス数を減らすことができ、動作速度を大幅に向上させることが可能である。続いて得られた動作点の座標データを正規化する。正規化処理は唇形状の正規化とその大きさの正規化を行うことにする。

## 4. 母音認識実験・考察

母音認識には3階層ニューラルネットワークを用い、その学習にはバックプロパゲーション法を用いた。実験データは同一人物の各母音と未発話状態の画像50枚分ずつを使用した。学習データは実験データ以外の各母音5枚分を使用した。認識結果を表1に示す。全体の認識率は約72%であり、読唇単語認識を実現するために必要な認識率が得られたと考えられる。

表 1: 母音認識結果内訳

実際の母音	認識された母音					未発話
	あ	い	う	え	お	
あ	40	6	1	0	1	2
い	9	25	6	6	0	4
う	0	0	44	0	0	6
え	0	1	1	36	6	6
お	2	2	3	9	34	0
未発話	0	0	12	0	0	38

## 5. おわりに

本稿では、発話時の唇領域画像に動的輪郭モデルを適用し、母音認識を行う手法について述べた。母音のみで全ての単語を認識することは不可能であるが、工場のロボットに命令する等の単語数が限定される状況ならば、十分に認識は可能だと考えている。

## 参考文献

- [1] Yuusuke Sasaki, Takao Kawamura and Kazunori Sugahara. Lip Shape Extraction for Word Recognition by Using Hardware Active Contour Model. Proceedings of the 2004 International Symposium on Intelligent Multimedia, Video & Speech Processing, pp.370-373, 2004.