

話者方向推定機能を持つテレビ会議用ビデオカメラの開発

Video Camera for Teleconferencing with Direction of Speaker Estimation Function

上杉 徹¹
Toru Uesugi

川村 尚生²
Takao Kawamura

清水 忠昭²
Tadaaki Shimizu

菅原 一孔²
Kazunori Sugahara

鳥取大学 大学院 工学研究科¹
Tottori University, Graduate School of Engineering

鳥取大学 工学部²
Tottori University, Faculty of Engineering

1 はじめに

本研究室で開発しているテレビ会議用のビデオカメラは、会場全体の映像を表示するために、3つのNTSCビデオカメラを用いてパノラマ映像を生成している。本研究では、音声による話者方向推定処理をした結果として選ばれた入力映像に対して動的輪郭モデルを適用し、複数の顔画像領域を抽出する。そして、話者の拡大映像とパノラマ映像を表示する処理をFPGAを用いてハードウェア実現することを試みた。

2 システムの構成

処理の流れとしては、3つのNTSCビデオカメラから送られてきた画像をデコーダを介してRGBデジタル信号に変換し、FPGAボード上のメモリに蓄積する。そしてFPGA上の各回路において、入力された映像データを基に映像処理を行い、エンコーダでアナログ信号に変換しモニタに表示する。

3 複数領域抽出手法

動的輪郭モデルは仮想的な閉曲線上にある複数の動作点に圧力、引力、反力および振動項と呼ばれる4つの力が働くことにより、閉曲線が収縮し領域を抽出する手法である。しかし、この手法は抽出対象が複数となる場合には有効ではない。そこで、閉曲線を分裂させることで画像から複数の特定領域を抽出する様子を以下に示す。

まず、図1(a)に示すように、動作点に働く引力が釣り合った時、圧力を考慮することで、着目した動作点が隣り合う2つの動作点より内側に凸となる位置へ進入し、図1(b)に示す状態となる。この時、着目した動作点間の距離 h がパラメータとして与える一定の値よりも小さくなった場合、図1(c)に示すように動的輪郭モデルの閉曲線を分裂させ、新たに動作点を生成させる。このようにして複数の目的領域の抽出が可能となる。

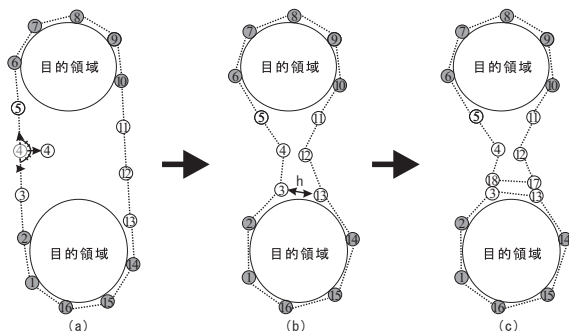


図1 分裂する動的輪郭モデル

4 話者方向推定手法

本システムでは、図2に示す2本のマイクを用いた方向推定モデルによって、3つのカメラのうち、話者を含んでいるカメラを選択している[1]。

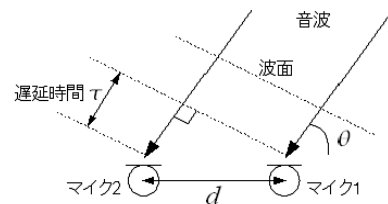


図2 方向推定モデル

5 出力映像生成処理

出力映像生成処理は以下の流れで行う。まず、音声による話者方向推定から話者を含んでいるカメラを選択し、選ばれた入力映像に対して動的輪郭モデルを適用し、複数領域を抽出する。その後、3つの入力映像からパノラマ映像を生成し、話者領域抽出結果を基に話者を拡大表示して、話者映像を生成する。最後に、話者映像とパノラマ映像から出力映像を合成し、モニタに出力する。

6 映像合成処理実験

図3に映像合成処理実験の結果を示す。分裂する動的輪郭モデルが、複数の顔領域を抽出していることが確認できる。



図3 映像合成処理実験の結果

7 おわりに

本研究では、テレビ会議用ビデオカメラの開発を目的に、分裂する動的輪郭モデルを適用し複数の顔領域を抽出し、話者の拡大映像を表示するシステムのハードウェア実現を試みた。

参考文献

- [1] 李 咏梅, 桂 竜二, 菅原 一孔, 小西 亮介: 周波数・位相推定手法に基づく複数音声の到来方向推定について, 電子情報通信学会論文誌 A, Vol.J86-A, No.3, pp.320-325 (2003)