

審査の結果の要旨

論文提出者氏名 小河 誠己

本論文は「音声に基づく映像インデクシングに関する研究」と題し、8章よりなる。一般にビデオカメラ、デジタルカメラといった映像機器が広く利用されるようになり、個人で扱う映像も飛躍的に増えている。増大する映像データに適切なインデクシングを行うことが急務となっている。本研究では、音声データの解析を通して、ホームビデオなどの映像の分類を行っている。音声データを基本的に4つのクラスに分類することで、インデックスを与え、周囲の状況を把握するキーを与えている。音声は、複数が自然に重なり合うため、分類にあたっては、分類が層状に重なり合う層状インデクシングを提案している。2つの手法を検討するとともに、さらに詳細な分類の試みを行った。また、音声分類を利用する映像のブラウザを構築している。

第1章は、「序論」であり、本論文の背景とその構成について述べている。映像によるインデクシング、音声を用いたインデクシングについてその概略にふれ、音声を用いた手法の位置づけを述べている。

第2章は、「映像及び音声インデクシングに関する関連研究」と題し、映像インデクシングに関する関連研究をまとめている。また、音声を利用するインデクシング、ホームビデオのインデクシングについてもその動向を述べている。

第3章は、「音声の分類と特徴量に関する考察」と題し、音声解析を通じて利用できる代表的な特徴量に関して、その定義を行い、テストデータに対してそれらの統計的な振る舞いを調べている。

第4章は、「ルールに基づいたホームビデオのための音声層状インデクシング」と題する。音声を、人の発話、無音、音楽、背景音の4クラスへの分類を行う。音声は、重なり合うのが自然なため、排他的ではない重なりを許容した層状のクラス分けを提案している。各クラス毎に代表的な音声特徴量を選び、非線形関数を介した指数を定義する。音声セグメント毎にその指数を算出し、さらに閾値処理を施して、音声セグメントの暫定インデクス付与を行う。最終的に、ノイズ除去のための平滑化処理と確定処理により、音声セグメントに一つあるいは複数のインデクスを付与する。合成的に作ったテストデータにて、混合の度合いによる検出率の評価、屋外で取得したホームビデオの音声データに対しての実験評価を行った。

第5章は、「GMMを用いたホームビデオのための音声層状インデクシング」と題する。前章のアプローチとは異なり、GMM（混合ガウスモデル）の学習に基づく音声の層状インデクシングを検討した。前章と同様に、音声を4つのクラスへ分類している。GMMの混合数を変えながら、屋外で取得したホームビデオの音声に対する分類実験を行い、評価している。人が聴覚のみで分類した場合との対比を行い、適合率、再現率の評価も行っている。

第6章は、「映像のより詳細な分類」と題し、前章までの分類が4つであったのに対して、2つの観点からのより細かな解析を試みている。一つは、発話の検出に続く話者認識の試みであり、もう一つは、4つのクラスの生起パターンによる状況の分析である。前者においては、発話区間において、GMMに基づく話者識別を試みている。しかし、雑音が影響し、話者識別の精度は50%以下にとどまっている。また、後者では、発話、音楽、背景音の重なり合いの多いところを、雑踏として検出することを試みている。

第7章は、「ホームビデオのための映像ブラウザ」と題し、音声の層状インデクシングが施された映像データを一覧するためのブラウザの作成に関して述べている。

第8章は、「結論」であり本論文の成果をまとめ、残された課題について記している。

以上これを要するに、本論文では、増大する映像データのインデクシングとして、音声に基づく層状インデクシングの手法を提案し、音声を4つのクラスに層状にわけ手法を検討し、さらにより詳細な分類についても述べている。本論文で論じた音声インデクシングは、将来の映像コンテンツの管理に寄与することが期待され、電子情報工学上貢献するところが少なくない。よって本論文は博士（工学）の学位論文として合格と認められる。