

学 位 論 文 題 名

ファジィ理論を用いた
映像信号のショット分類に関する研究

学位論文内容の要旨

本研究は、ファジィ理論を用いた映像信号のショット分類に関する基礎的な研究の成果をまとめたものである。

近年、放送のデジタル化や記録媒体の大容量化、高速通信網を介した映像配信の普及により、ユーザが保持する映像コンテンツは急速に増加している。このような状況において、蓄積されたコンテンツの中から所望の映像を効率的に閲覧するために、映像信号の検索システムが必要とされている。

映像信号の検索を行うためには、前処理として、映像信号を内容に基づく時間単位に分割し、内容を表すインデックスを付加する必要がある。この分割や分類を行う時間単位として、一般に1台のカメラで連続的に撮像された区間であるショットや、内容に関連のあるショットを統合したシーンが用いられている。そのため、映像信号より得られるビデオ信号を用いて、隣接するショット間の境界（以降、ショットカットと呼ぶ）を検出し、得られたショットカットを境界とするショットにインデックスを付加する研究が行われている。これらの研究では、ショットカットの前後で、画像の輝度値や動きベクトルが急激に変化する特徴に基づき、ショットカットを検出する。一方、隣接するシーンの境界（以降、シーンカットと呼ぶ）は、ショットカットの一部として検出される。しかしながら、ショットカット前後における輝度値や動きベクトルの変化の様子と、シーンカット前後におけるそれらの変化の様子との間に、明確な差異は存在しない。すなわち、輝度値や動きベクトルの変化において、シーンカット固有の変化は存在しない。そのため、ビデオ信号を処理しただけでは、ショットカットの中からシーンカットを検出することは困難となる。そこで、シーンカットは、ビデオ信号だけでなく、オーディオ信号も同時に切り換わることに着眼し、ビデオ信号とオーディオ信号を併せて使用することで、シーンカットの検出を行う研究が行われている。このオーディオ信号に着眼した研究では、まずオーディオ信号を分割することにより、得られるセグメントを無音や音声、音楽等のクラスに分類する。ここで、分類されたクラスが切り換わる時刻とショットカットが一致した場合、その時刻をシーンカットであると判断する。

ところで、従来のオーディオ信号を用いた映像信号の分割・分類法は、実用的なレベルには達していない。これは、オーディオ信号にフェードインやフェードアウト等の音響効果が付加された場合、オーディオ信号が切り換わる時刻（以降、オーディオカット）の検出精度が劣化することに起因する。また、背景に音楽が存在する環境下での音声や、背景に雑音の存在する環境下での音声等、複数の種類のオーディオ信号が混在する場合、オーディオ信号の分類結果が劣化する危険性がある。この精度の劣化は、それぞれのオーディオ信号の音量に様々な程度の違いが存在するにも関わらず、その差異

を考慮せずに分類を行っていることに起因する。

そこで、本論文では、音響効果や、複数の種類のオーディオ信号が存在する場合においても、高精度にオーディオ信号の分割・分類を行う手法の提案を行っている。本論文では、ファジィ理論を用いたクラスタリング手法であるファジィ c-means 法を用いることで、音響効果が付加されたオーディオ信号を処理対象とした場合においても、高精度なオーディオカットの検出を実現している。さらに、主成分分析とマハラノビス汎距離を用いてオーディオ信号の分類を行い、得られる分類結果とビデオ信号より得られるショット分割結果に対しファジィ推論を行うことで、高精度なショット分類を実現している。提案手法は、音響効果や複数の種類のオーディオ信号が存在する場合の特徴の変化に着目し、この変化の様子を、ファジィ理論を用いて定量化することにより、高精度なショット分類を可能としている。

本論文では、まず第 2 章で、ショットやシーン等の映像信号の構造について定義する。第 3 章では、既に提案されているオーディオ信号の分割・分類法について、現在主流となっている手法を取り上げ、その処理手順の説明を行う。第 4 章では、ファジィ理論の基礎となるファジィ集合について述べる。また、ファジィ理論を用いたクラスタリング手法であるファジィ c-means 法や、ファジィ推論についても述べる。第 5 章では、音響効果がオーディオカット検出に与える影響を調べ、第 3 章で説明した従来のオーディオカット検出法における問題点を指摘する。そして、この問題を解決する新たなオーディオカット検出法を、第 4 章で説明したファジィ c-means 法を用いて提案する。提案するオーディオカット検出法を用いることで、音響効果が付加されたオーディオ信号を処理対象としても、高精度にオーディオカットを検出することが可能となる。第 6 章では、ファジィ c-means 法を用いた新たなオーディオ信号の分類法を提案する。背景に音楽が存在する環境下での音声等、複数の種類のオーディオ信号が混在する場合、従来のオーディオ信号の分類法では、分類精度が劣化する問題があった。この精度の劣化は、それぞれのオーディオ信号の音量に様々な程度の違いが存在するにも関わらず、従来手法はその程度の差異を考慮せずに分類を行っていることに起因する。そこで、提案手法は、ファジィ c-means 法より得られる帰属度を用いることで、この差異を定量化し、高精度な分類を実現する。さらに、本章では、提案手法を拡張し、MPEG で符号化されたオーディオ信号から、復号化処理を施すことなく直接オーディオカットを検出する手法を提案する。現在、映像信号は一般に MPEG 等で符号化されて配信が行われるため、MPEG で符号化されたオーディオ信号への対応は、計算コストの低減、応用範囲の拡大等、様々な利点がある。第 7 章では、ビデオ信号とオーディオ信号の協調処理によるショット分類法を提案する。第 6 章で提案したオーディオ信号の分類法では、ユークリッド距離に基づくファジィ c-means 法を使用していたため、特徴量の分布形状が考慮されず、その結果、オーディオ信号の種類によっては、分類精度が劣化する問題があった。そこで、この問題を解決するために、特徴量の分布形状を考慮した距離尺度であるマハラノビス汎距離を導入し、さらにビデオ信号から得られるショット分割結果を併せて用いることで、映像信号のショット分類を実現する。提案手法は、分類の際に、ファジィ推論を導入することで、複数の種類のオーディオ信号が混在する場合においても、高精度な分類が可能となる。最後に、第 8 章において、本研究の成果について要約し、論文全体のまとめとする。

以上を要約すると、本論文は、ファジィ理論を用いることで、高精度なショット分類を実現する手法について提案を行っている。また、本手法を実際の映像信号に適用した実験を行うことにより、その有効性及び有用性を示している。

学位論文審査の要旨

主 査 教 授 長谷川 美 紀

副 査 教 授 山 本 強

副 査 教 授 荒 木 健 治

学 位 論 文 題 名

ファジィ理論を用いた

映像信号のショット分類に関する研究

本論文は、著者が行ったファジィ理論を用いた映像信号のショット分類に関する研究をまとめたものである。

近年、放送のデジタル化や記録媒体の大容量化、高速通信網を介した映像配信の普及により、ユーザが保持する映像コンテンツは急速に増加している。このような状況において、蓄積されたコンテンツの中から所望の映像を効率的に閲覧するために、映像信号の検索システムが必要とされている。

映像信号の検索を行うためには、前処理として、映像信号を内容に基づく時間単位に分割し、内容を表すインデックスを付加する必要がある。この分割や分類を行う時間単位として、一般に一台のカメラで連続的に撮像された区間であるショットや、内容に関連のあるショットを統合したシーンが用いられる。そのため、映像信号より得られるビデオ信号を用いて、隣接するショット間の境界（ショットカット）や、隣接するシーン間の境界（シーンカット）を検出する研究が行われている。また、ショットカットやシーンカット検出の高精度化、及び映像信号の分類を目的として、映像信号より得られるオーディオ信号に着目した研究も行われている。このオーディオ信号に着目した手法では、まずオーディオ信号の種類が切り換わる時刻（オーディオカット）を検出し、得られたオーディオカットを境界とするセグメント（オーディオセグメント）を、無音や音声、音楽等のクラスに分類する。しかしながら、既存のオーディオ信号の分割・分類手法は、音響効果が施されたオーディオ信号や、複数の種類のオーディオ信号が混在する場合に、分割・分類精度が劣化する問題があった。

著者は、本論文においてこの問題を解決すべく、ファジィ理論を用いて、オーディオ信号の分割・分類結果に含まれる曖昧性を定量的に評価することで、オーディオ信号の分割・分類精度を向上させた。さらに、得られたオーディオ信号の分割・分類結

果と、ビデオ信号より得られたショットカット検出結果の両者を用い、協調処理を行うことで、映像信号のショット分類を実現した。これにより、従来の手法に比べ、高精度な映像信号の分割・分類が実現できることを示した。

以下、章を追って論点を述べる。まず論文第2章で、映像信号の構成について説明を行い、本論文で用いる用語の定義を行った。第3章では、既存のビデオ信号の分割手法について、特に代表的な手法である長坂らの分割 χ^2 検定法、及び B.T.Truong らのフェード・ディゾルブ検出法を取り上げ、その処理手順について説明を行った。第4章では、既存のオーディオ信号の分割・分類手法について、特に本論文と同様の目的のため提案された T.Zhang の手法を取り上げ、その概要の説明を行った。第5章では、L.A.Zadeh が提案したファジィ理論について、特にファジィ集合、ファジィc-means 法、ファジィ推論を取り上げ、説明を行った。第6章では、フェード等の音響効果がオーディオカット検出に与える影響について考察を行った。また、ファジィ理論を用い、音響効果の影響を受けない新たなオーディオカット検出手法を提案した。さらに、提案手法を改良し、MPEG で符号化されたオーディオ信号から、復号化処理を施すことなく、直接オーディオカットを検出する手法を提案した。第7章では、第6章で得られたオーディオカットを境界とするオーディオセグメントを処理対象とし、ファジィ理論を用いたオーディオセグメントの分類手法を提案した。第8章では、オーディオ信号とビデオ信号の協調処理による、映像信号のショット分類手法を提案した。第9章において、著者は本研究の成果について要約し、論文全体のまとめを行った。

以上を要約すると、著者は、高精度な映像信号のショット分類を実現するために、音響効果や複数の種類のオーディオ信号の混在が映像信号の分割・分類結果へ与える影響について分析を行い、映像信号の分割・分類結果に含まれる曖昧性をファジィ理論を用いて定量的に評価し、その評価結果に基づく分割・分類手法についての提案を行った。本研究を通じて、情報メディア工学・音響工学への大きな貢献をしたので、著者は、北海道大学博士（情報科学）の学位を授与される資格があるものと認める。