



Title	階層ベイズモデルの導入による映像の構造に注目したシーン分割の高精度化に関する研究 [論文内容及び審査の要旨]
Author(s)	宋, 妍
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第11764号
Issue Date	2015-03-25
Doc URL	https://hdl.handle.net/2115/58918
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	doctoral thesis
File Information	Yan_Song_review.pdf, 審査の要旨



学位論文審査の要旨

博士の専攻分野の名称 博士 (情報科学) 氏名 宋 妍

審査担当者 主 査 教 授 長谷山 美紀

 副 査 教 授 山本 強

 副 査 教 授 荒木 健治

学位論文題名

階層ベイズモデルの導入による映像の構造に注目したシーン分割の高精度化に関する研究
(A Study on Accurate Scene Segmentation Using Video Structures Based on Bayesian Hierarchical Model)

本論文は、階層ベイズモデルの導入による映像の構造に注目したシーン分割の高精度化に関する研究の成果をまとめたものである。

近年、高速・大容量な通信回線の普及に伴い、インターネットによる動画配信サービスの利用者数が増加している。特に、YouTube や Dailymotion などに代表される動画配信サービスの拡大は著しく、大量のユーザが膨大な量の映像を視聴している。そのため、蓄積された映像の中からユーザが望む映像を検索するシステムの必要性が高まっている。

一般に、映像は意味内容の異なる場面により構成されることから、意味内容の切れ目に注目して予め映像を分割することで、効率的な検索を行うことが可能となる。そのため、映像を同一意味内容のまとまりであるシーンに分割する技術が、映像解析の基礎技術として広く研究されている。従来より、シーンは同一もしくは類似する内容を表すショットが連続する区間として定義され、ショットは1台のカメラにより切れ間なく撮像される区間であり、映像のシーン分割は、隣接するショットの意味が同一もしくは類似であるか否かを判定し統合することで行われる。ところで、既存のシーン分割手法には、代表的なものとして2つの手法が挙げられる。1つは映像の構造に注目した手法であり、もう1つは階層ベイズモデルに基づく手法である。前者の手法は、編集映像に現れる、類似するショットの繰り返し構造を見つけることで、シーン分割を行う。この手法では、映像の構造を解析する際、注目する繰り返し構造を示す最長の時間幅を設定する必要がある。しかしながら、最長時間幅は映像の種類によっても編集者によっても異なるため、一意に決定することが困難であるという問題が存在した。一方、階層ベイズモデルに基づくシーン分割手法では、シーン数やシーン境界の位置に関する事後確率の最大化により、最適なシーン境界を検出する。しかしながら、この手法は、その事後確率をショット間の類似度のみに注目して定義しているため、同一シーン中の隣接するショット間で類似度が低い組み合わせが存在すると誤ってシーン境界と推定する場合がある。つまり、映像の構造を考慮せずにショット間の類似度のみに注目しているため、シーン分割の精度が低下するという問題が存在した。

上に示す通り、既存の代表的な2種類のシーン分割手法には、各々に問題が存在した。この解決策として、映像の構造に注目した手法においてシーンの最長時間幅を一つの値に設定するのではなく、複数の値に変化させてシーン境界の候補を求め、それらの中から階層ベイズモデルに基づく手法によって最適なシーン境界を選択することが考えられる。これにより、各々の手法の問題が解決可能と

なることから、映像の構造だけでなく、階層ベイズモデルに基づく手法との連携により、より高精度なシーン分割手法の実現が期待できる。

そこで、本論文では、映像の構造と階層ベイズモデルの両者を利用することで、映像の構造に注目する手法が抱える繰り返し構造の最長時間幅を一意に決定することが困難であるという問題を解決し、高精度に映像のシーン分割を実現する手法を提案している。具体的に提案手法では、最長時間幅を異なる複数の値に設定して得られたシーン分割結果をシーン境界候補とし、その中から適切なシーン境界を選択することで問題を解決する。このとき、提案手法では、各シーン境界候補が真のシーン境界となる事後確率を階層ベイズモデルに基づいて定義し、これをマルコフ連鎖モンテカルロ法により最適化することで、適切なシーン境界の選択を可能とする。さらに、本論文では、階層ベイズモデルにおける事後確率の算出において、映像を構成する画像信号と音響信号の両方の特徴量を用いることで高精度化を試みる。異なる信号の特徴量を統合して用いるためには、比較可能な数値への変換が必要である。本論文では、異なるショットの全ての組み合わせで類似度を算出して分布を求め比較可能とすることで、事後確率を算出する。

本論文では、第2章において、従来より提案されているショット分割法について説明し、あわせて用語の定義を行う。第3章では、従来の映像のシーン分割法手法の概要を述べるとともに、解決すべき問題点を明らかにする。第4章では、映像の構造と階層ベイズモデルの両者を用いた高精度なシーン分割手法を提案し、実験によりその有効性を示す。また、第5章では、画像と音響に関する特徴量の類似度に関して、その分布を考慮しながら比較可能とする定式化を導入し、シーン分割精度の向上が可能となることを明らかにする。最後に第6章において、本論文のまとめを示す。

これを要するに、著者は、シーン分割の従来手法に存在する問題点を分析し、映像の構造に注目する手法と階層ベイズモデルに基づく手法の両者を用いることにより、高精度な映像のシーン分割を実現する手法を提案している。また、本提案手法を実際の放送映像に適用した実験を行い、その有効性を示しており、本研究を通じて、情報科学の分野に貢献するところ大なるものがある。

よって著者は、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。