論文 / 著書情報
Article / Book Information

題目(和文)	大規模映像資源のための高速・高性能なセマンティックインデクシング
Title(English)	Efficient and Effective Semantic Indexing for Large-Scale Video Resources
著者(和文)	井上中順
Author(English)	Nakamasa Inoue
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第9552号, 授与年月日:2014年3月26日, 学位の種類:課程博士, 審査員:篠田 浩一,佐藤 泰介,徳永 健伸,村田 剛志,杉山 将
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9552号, Conferred date:2014/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

(博士課程)

論文審査の要旨及び審査員

報告番号	甲第	号	学位申請者氏名	井上 中順	
論文審査 審査員		氏名	職名	氏名	職名
	主査	篠田 浩一	教授	杉山 将	准教授
	審査員	佐藤 泰介	教授		
		徳永 健伸	教授		
		村田 剛志	准教授		

論文審査の要旨 (2000 字程度)

本論文は「Efficient and Effective Semantic Indexing for Large-Scale Video Resources (大規模映像資源のための高速・高性能なセマンティックインデクシング)」と題し、英文7章より構成されている。

第1章「Introduction」では、本研究の背景、目的、概要について述べている。セマンティックインデクシングは映像から物体・イベント・シーンなどの意味的コンセプトの検出を目的としていることを述べ、従来の統計的アプローチの概要とともに、セマンティックギャップに起因する問題の難しさについて述べている。

第2章「Semantic Indexing」では、近年のセマンティックインデクシングに対するアプローチについて概観している。まず、従来の様々な手法に共通する枠組みが、1) 局所特徴抽出、2) 映像表現抽出、3) 検出、の3段階から成ることを述べ、それぞれについて代表的な手法を説明している。第1段階に関しては、Scale Invariant Feature Transform (SIFT) や Histogram of Oriented Gradients (HOG) などの特徴量に関して説明している。第2段階に関しては、Bag-of-Visual-Words 法およびその発展的手法について述べている。第3段階に関しては、サポートベクトルマシンなどの統計的学習手法について概観している。

第3章「Multi-Modal Semantic Indexing」では、画像・音響特徴量と混合ガウス分布を用いたセマンティックインデクシングシステムについて述べている。具体的には、本システムの第1段階では、画像と音響に関する局所特徴量の抽出、第2段階では、混合ガウス分布のパラメータ推定、第3段階では、サポートベクトルマシンを用いた検出スコアの算出が行われることを説明している。さらに、本システムの評価実験についても述べており、800時間のインターネット映像からなる TRECVID データセットにおいて、364種類の意味的コンセプトの検出を行うことで、本システムの有用性を示している。

第4章「Tree-structured Gaussian Mixture Models」では、混合ガウス分布のパラメータ推定を高速化するための高速 MAP 適応について述べている。本手法は、木構造混合ガウス分布を構成し、重要な混合成分に関してのみ計算を行うことで、計算量の削減を実現することを述べ、木構造を構成するアルゴリズムについて説明している。さらに、本手法を第3章のシステムに組み込んだ、TRECVID データセットにおける評価実験について述べている。実験では、高速 MAP 適用により、検出精度を高く維持した状態で、計算量が 76.2%削減されたことを述べている。

第5章「q-Gaussian Mixture Models」では、q-ガウス分布の混合分布である、q-混合ガウス分布について述べている。まず、Tsallis 統計から導出される q-ガウス分布の定義について述べている。次に、q-ガウス分布は、分布の裾の長さを調整するパラメータ q を有し、 $q > 1$ の場合に得られる裾の長い分布を用いることで、外れ値に対してより頑健なモデル化が行われることを述べている。さらに、q-混合ガウス分布を用いたカーネルについて述べ、3章のシステムに組み込む方法を説明している。最後に、20物体に関する1万枚の画像からなる PASCAL VOC 2007 物体認識データセットにおいて、評価実験を行い q-混合ガウス分布の有用性を示している。

第6章「Neighbor-To-Neighbor Search」では、Neighbor-to-Neighbor (NTN) 探索という高速化手法について述べている。ここでは、Dense sampling に基づいた画像特徴抽出において、2つの隣接した点から抽出された特徴量がしばしば類似しているという事実を指摘し、NTN 探索がそのような類似した特徴量に関して計算を省略することで高速化を実現するということを述べている。次に、ベクトル量子化と混合ガウス分布それぞれに関する NTN 探索のアルゴリズムについて説明している。最後に、PASCAL VOC 2007 データセットにおいて、評価実験を行い認識精度の劣化なしに、ベクトル量子化に関しては 77.4%、混合ガウス分布に関しては 89.3%の計算量が削減されたことを述べている。

第7章「Conclusion and Future Work」では、本論文の結論および貢献について述べ、今後の課題についてまとめている。

以上を要するに、本論文は、映像検索における重要な技術であるセマンティックインデクシングにおいて、高性能であり、かつ、高速なアルゴリズムを提案するものであり、工学上貢献するところが大きい。よって我々は、本論文が博士(工学)の学位論文として十分価値あるものと認める。