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, 学位の種別: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)	論文要旨
Type(English)	Summary

（博士課程）

Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	計算工学　専攻	申請学位（専攻分野）： Academic Degree Requested	博士（　工学　） Doctor of
学生氏名： Student's Name	井上　中順	指導教員（主）： Academic Advisor(main)	篠田　浩一　教授
		指導教員（副）： Academic Advisor(sub)	

要旨（英文 800 語程度）

Thesis Summary（approx.800 English Words ）

The dissertation "Efficient and Effective Semantic Indexing for Large-Scale Video Resources" consists of 7 chapters.

Chapter 1 [Introduction] presents the background and the motivation for video semantic indexing. Semantic indexing is aiming to assign semantic concepts such as objects, events, and scenes to a video segment. A brief explanation of statistical approaches to semantic indexing is given along with a description of the difficulties in bridging the semantic gap between low-level features such as RGB features and high-level semantic concepts.

Chapter 2 [Semantic Indexing] gives an overview of recent approaches to semantic indexing. We show the most common framework for semantic indexing, which consists of three steps: 1) low-level feature extraction, 2) video representation, and 3) detection. For the first step, low-level feature extraction methods such as scale invariant feature transform (SIFT) and histogram of oriented gradients (HOG) are reviewed. For the second step, a number of studies for bag-of-visual-words representation and its extensions such as a kernel codebook and Fisher vectors are described. For the third step, discriminative learning methods such as support vector machines (SVMs) are reviewed.

Chapter 3 [Multi-Modal Semantic Indexing] describes our multi-modal semantic indexing system, which uses Gaussian-mixture-models (GMMs) with audio and visual features. Our system, which consists of the following three steps, extends the bag-of-visual-words system to a probabilistic framework. In the first step, visual and audio low-level features are extracted from a video shot. For visual features, SIFT features are extracted from each local region detected by a Harris-affine detector, a Hessian-affine detector, and a dense sampling detector. For audio features, Mel-frequency cepstral coefficient (MFCC) features, which describe the short-time spectral shape, are extracted from audio frames. In the second step, a video shot is represented by a GMM supervector, which concatenates GMM parameters as a single vector. Here, GMM parameters are estimated by using Maximum A Posteriori (MAP) adaptation. Finally, detection scores are calculated by using Support Vector Machines (SVMs). This chapter also presents experiments on the TRECVID Semantic Indexing dataset. It provides 800 hours of Internet video with annotations for 346 types of semantic concepts such as "airplane flying", "car", "cityscape", "nighttime" "singing", and "dancing". We detect the 346 semantic concepts from the dataset to show the effectiveness of our multi-modal system.

Chapter 4 [Tree-structured Gaussian Mixture Model] describes a fast MAP adaptation technique for improving the speed of the GMM parameter estimation. With this technique, a tree-structured GMM is constructed to calculate probabilities only for important Gaussian components and to skip the calculation for the others. This chapter provides our algorithm to construct the tree-structured GMM and its application to the semantic indexing system in Chapter 3. This chapter also provides our experiments on the TRECVID dataset. It shows that the calculation time for GMM parameter estimation using MAP adaptation is reduced by 76.2%, while high detection performance was maintained.

Chapter 5 [q-Gaussian Mixture Model] describes a q-Gaussian mixture model (q-GMM), which is a mixture of q-Gaussian distributions. First, the definition of a q-Gaussian distribution derived from Tsallis statistics is provided. Since it has a parameter q to control the tail-heaviness of a Gaussian distribution, a long-tailed distribution obtained for $q > 1$ is expected to effectively represent complexly correlated data, and hence, to improve robustness against outliers. Second, a semantic indexing system using a q-GMM is described. Here, we propose a q-GMM kernel to improve the modeling accuracy of our semantic indexing system in Chapter 3. Finally, this chapter provides experiments on the PASCAL VOC 2007 classification challenge dataset, which provides 10,000 images of 20 objects. We show the effectiveness of the q-GMM in our semantic indexing framework.

Chapter 6 [Neighbor-To-Neighbor Search] describes Neighbor-to-Neighbor (NTN) search to further improve the speed of the system when densely sampled image descriptors are used as low-level features. Based on the fact that image features extracted from two adjacent points are usually similar to each other, this search algorithm effectively skips some calculations for such similar image features. First, this chapter provides an NTN search algorithm in a simple framework, vector quantization (NTN-VQ). Second, it provides an extension of the NTN search to a GMM. Finally, it describes our experiments on the PASCAL VOC 2007 classification challenge dataset. Our experiments show that NTN-VQ reduces the computational cost by 77.4% for VQ, and NTN-GMM reduces it by 89.3% for a GMM, without any significant degradation in classification performance.

Finally, in Chapter 7 [Conclusion and Future Work], the conclusions of this study and our future work focusing on applications of our semantic indexing method to event detection, video summarization, and object localization are described.

備考 ： 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).