

氏名	張 鶴
学位の種類	博士(工学)
学位記番号	富理工博甲第66号
学位授与年月日	平成26年3月21日
専攻名	数理・ヒューマンシステム科学専攻
学位授与の要件	富山大学学位規則第3条第3項該当
学位論文題目	Content-Based Product Image Retrieval and Classification (内容に基づく商品画像検索と分類)
論文審査委員 (主査)	菊島 浩二 唐 政 清水 正明

## 【学位論文内容の要旨】

Content-based image retrieval (CBIR) is an application of computer vision techniques to the image retrieval problem (searching for digital images in large databases). The "content-based" means that the search analyzes the contents of the image rather than the metadata such as keywords, tags, or descriptions which are associated with the image. The term "content" in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself.

Electronic commerce emerged in the 1970s. With the development of internet, online shopping becomes more and more popular. Product images become the main form of commodity exhibition, and is also the decisive factor for consumers to understand and purchase goods. But so far, most of the shopping sites still rely on the establishment of keyword indexing. The retrieval methods based on the keywords can not retrieve accurately and unable to describe the visual content well, which can not meet the needs of the users. Consequently, content-based product image retrieval and classification has become a hot topic of current research.

In the chapter 1, the concepts of CBIR and text based image retrieval are analyzed combining with E-commerce applications. The retrieval methods based on the keywords can not retrieve accurately and unable to describe the visual content well, which can not meet the needs of users. The general framework of CBIR and some typical image retrieval systems are presented. In addition, we give a brief outline of the performance evaluation of retrieval methods.

Content-based image representation is the basis of image retrieval and classification. Chapter 2 explores the image representation methods from the aspects of low-level

representation and intermediate semantic representation. The low-level representation can be divided into two categories, global representation and local representation; the intermediate semantic representations include local semantic concept, in which the BOW model on the basis of local semantic concept model is one of the recent mainstream methods for image retrieval and image classification.

In chapter 3, we proposed a combination algorithm of color descriptor, LBP texture descriptor and HOG shape descriptor for product image retrieval. The color histogram has the advantages of simple, invariant of image rotation, scale and translation; however, the color histogram lost the space distribution information. HOG is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization. LBP analyzes the fix window features with structure methods and extract global feature with statistic methods. The LBP and HOG features provide texture and shape information of an object within an image, respectively. In addition, they are robust to noise, so it is beneficial to combine the three features together. The product image retrieval experiments indicate the combination can boost the performance of retrieval system significantly.

Chapter 4 applies visual attention model to the product image retrieval algorithm. HVS (human visual system) model is used by visual experts to deal with biological and psychological processes that are not yet fully understood. The study on HVS shows that, when people observed images, the brain based on visual attention mechanism can quickly respond to the area of interest and draw visual attention to the part of the image. Consequently, it's certainly reasonable to establish VAM (Visual Attention Model) through simulation of the human visual system to get the most attractive part and represent it with a gray scale image. This chapter analyzes the existing visual attention

model features and apply visual attention model to the product image retrieval algorithm. Firstly, the saliency map is generated based on visual attention model, on which the saliency part is extracted with dynamic threshold method. Then, the edge information of saliency part is obtained through Canny operator. Image retrieval is implemented through combining the color histogram of saliency map and gradient direction histogram of saliency edge. The proposed algorithm highlights the perception of object areas, inhibits background effects and improves retrieval performance.

For the efficiency of product retrieve, it is of necessity to classify the numerous products into some categories automatically. Each category can be classified into many sub-categories. Chapter 5 implements product classification depending on the visual characteristics and stacked auto-encoder classifier. A stacked auto-encoder consists of multiple layers of sparse auto-encoders, the outputs of each layer is the inputs of the successive layer. An auto-encoder attempts to learn appropriate features to represent its raw input, and higher layers tend to learn higher-order features, which construct a hierarchical grouping of the input.

In chapter 6, SVM (Support Vector Machine) and PHOG (Pyramid Histograms of Orientation Gradients) descriptor are employed to implement product classification. Support vector machines can efficiently perform a non-linear classification using the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. PHOG is an excellent image global shape descriptor, which consists of a histogram of orientation gradients over sub-region of each image at each resolution level. In this thesis, we adopt SVM classifier combined with PHOG descriptors to implement product-image classification.

Experimental results showed the effectiveness of the proposed algorithms. Future

study will focus on the selection more efficient complimentary descriptors set (such as appearance descriptor, color descriptor, texture) and design more effective complicated combination and learning algorithms (kernel based SVM, saliency analysis and deep auto-encoder) to improve the overall performance of product retrieval and classification system.

## 【論文審査の結果の要旨】

当学位論文審査委員会は、標記の博士学位申請論文を詳細に査読し、また論文発表会を平成26年2月7日(金)に公開で開催し、詳細な質疑を行って論文の審査を行った。以下に審査結果の要旨を記す。

内容に基づく画像検索(CBIR)は画像のデータベースから、視覚的に指定されたクエリのイメージの類似画像を取得するシステムである。「内容に基づく」とは、キーワード、タグ、また画像に関連する説明文書などの検索ではなく、画像の内容を解析していることを示す。ここで言う「画像の内容」とは、色、形、テクスチャ、または画像自体から得られる他の情報を指す。

電子商取引は1970年代に初めて現れ、近年、インターネットの発展に伴い、オンラインショッピングはますます流行している。商品展示は商品画像がメインになり、また、消費者が商品を理解し、購入するための決定的な要素でもある。しかし、今までのオンラインショッピングサイトのほとんどは、依然として、キーワードのインデックス付けに依存している。キーワードに基づいた検索方法では、ユーザーの要求する条件を満たすことができないため、学位申請論文では画像の内容に基づく商品画像検索と分類を研究課題とし、まず、色記述子、LBPテクスチャ記述子とHOG形状記述子の組み合わせたアルゴリズムを提案し、さらに、視覚特性を用いた画像検索モデルを加え、その有効性を示した。最後に、商品画像分類を実現するために、PHOG記述子と組み合わせたSVMのエンコーダを採用し、画像検索システムに実装した。実験結果により、提案したアルゴリズムの有効性を示した。

学位申請論文は6章で構成される。各章の概要を以下に示す。

第1章では、論文の概要、内容に基づく画像検索と電子商取引アプリケーション

ンにおける典型的な画像検索システムを提示し、各種検索手法の性能評価を述べた。

第2章では、画像の表現方法について述べた。

第3章では、色記述子、LBPテクスチャ記述子とHOG形状記述子の組み合わせたアルゴリズムを提案し、検索システムの性能が高まることを示した。

第4章では、視覚的注意のモデル(VAM)を商品画像検索アルゴリズムに適用し、対象領域の知覚を強調し、バックグラウンド効果を阻害することによって、検索性能をより向上させることができることを示した。

第5章では、人間の視覚特性を用いたモデルを提案した。

第6章では、商品画像分類を実現するために、PHOG記述子と組み合わせたSVMのエンコーダを採用し、画像検索システムに実装した。実験結果により、提案したアルゴリズムの有効性を示した。

学位申請論文で提案したアルゴリズムは非常に有効で、工学的応用のみならず、学術的にも価値が高い。

よって、当博士論文審査委員会は本博士学位申請論文が博士の学位を授与することに十分に値するものと認め、合格と判断した。