氏　　名	Siriwat KASAMWATTANAROTE
学位(専攻分野)	博士(情報学)
学 位 記 番 号	総研大甲第 1834 号
学位授与の日付	平成28年3月24日
学位授与の要件	複合科学研究科　情報学専攻 学位規則第6条第1項該当
学 位 論 文 題 目	Query Expansion for Visual Search Using Data Mining Approach
論 文 審 査 委 員	主　査　教授　杉本　晃宏 教授　佐藤　いまり 准教授　Duy-Dinh Le 准教授　Gene Cheung 教授　佐藤　真一　国立情報学研究所

(別紙様式 2)
(Separate Form 2)

論文内容の要旨
Summary of thesis contents

Title: Query Expansion for Visual Search Using Data Mining Approach

In the recent days of the information era, images were taken a lot by using several handy devices such as a smart phone and a digital camera. These devices create a kind of time-memory as the moments of one people life which will be used for reminding back their memory in the previous days. As the current social media and social network technologies allow people to connect and share their life and memories with friends, million images were rapidly created and increasing size of valuable image collection these days, which are being retrieved back to the devices for a looking back purpose.

In case we use any services for sharing our memories from several devices like be a PC, a mobile phone, a tablet, or even on a wearable device through the web services (e.g. Instagram, Facebook, Twitter, Flickr, etc.). Retrieving back any image is meant to load a particular image directly from the server to visualize it on your screen by using a direct URL, hash tags, or any provided API.

Additionally, images are being indexed into collections of big image databases with specific meta-data, hash tag, time stamp, geographical location, or even a specific ID. This kind of indexing is to map the unstructured data into a high-level structured space. And the retrieval will be fast and easy just like following look up tables on a server for an exact physical location of the recalling image. On the other hand, searching for an object within one or multiple images is hard and it will be even more complicated if the image was not provided with any specific meta-data. For this reason, indexing cannot be done by mapping image to high-level structured data, but rather be performed by more sophisticated techniques calling as bag-of-visual-word (BoVW) framework

Since the BoVW method is inspired by text-based information retrieval, it is quite natural for QE to be applied to image retrieval as well. One of the simplest QE techniques is called pseudo relevance feedback or blind relevance feedback in Salton et.al. The method feeds these relevant documents back to the system assuming that the top k of the retrieved ranked items are relevant and then generates the query by aggregating the features of these items. In object-based image retrieval, the standard way is to use the average of BoVW histograms of the top-k items as the refined query for the next round retrieval in order to get improved retrieval performance without an extended of user interaction. This technique works well in text-based information retrieval, since each term does not have a strict word order.

In contrast, an image has a meaning that depends on the spatial locations of its visual words. Hence, a simple application of QE to image retrieval may fail because irrelevant images may be included among the highly ranked results and also because visual words in background regions may be included in the relevant images. Therefore, the retrieved images may be irrelevant to the query.

To circumvent the problem of background clutter as in QE, takes geometric information into account in QE in order to take the advantage of geometric topology information, Total

Recall adapts spatial verification (using RANSAC or the like) between the query image (or query region) and each of the database images for an object-based image retrieval. Then the method is widely known as average query expansion (AQE). Given a ranked list of retrieved images according to the query image with query region, AQE first checks the geometric consistency between the query region and each of the retrieved images and only selects relevant visual words in verified images appearing in back-projected regions to be put into the next-round query.

AQE imposes pairwise spatial constraints on the query image and each of the database images. The method lies at the heart of recent state-of-the-art visual QE methods. But in so doing it may narrow the range of the expanded query or miss information that is in the relevant images but not in the query (e.g. it can miss objects with occlusions or small objects with low granularity or noise). However, since AQE checks only the pairwise consistency between the query and each of the highly ranked images, its performance may be affected by slight degradations in the query image.

We propose a new method called Query Bootstrapping (QB). The key idea of this variant of QE is to use the consistency among highly ranked images, instead of using only the pairwise consistency between the query and each of the ranked images. Doing so relaxes the over-dependency on a query that affects AQE, and thus QB may be more robust to the degradation and/or variation in the query images. We regard frequently co-occurring visual words in highly ranked images as relevant. We use frequent item sets mining (FIM) to efficiently find co-occurring visual words in highly ranked images, and we also use LO-RANSAC to check the geometric consistency of the highly ranked images and remove those that do not pass the check. FIM outputs frequent patterns, each of which is composed of a set of visual words that co-occur frequently in the top-k highly ranked images. We then use the visual words appearing in the frequent patterns to formulate the next-round query by averaging together BoVW histograms of the original query and highly ranked (optionally geometrically verified) images with only the visual words in the frequent patterns. To do this, we propose tf-fi-idf as an extension of tf-idf that takes into account frequent patterns (fi). This method requires the parameters to be carefully designed, namely, the support as the fraction of co-occurrences in the top-k highly ranked images, and the top k as the number of highly ranked images to be fed to QB. There have been a number of previous attempts at using FIM for image retrieval; however, very few of them have dealt with automatic optimization of such parameters. Here, we devised an adaptive support selection algorithm that returns both the minimum support (minsup) and maximum support (maxsup) in order to find the optimal faction of frequent patterns out of the top-k images. Moreover, we also created an algorithm that selects a suitable inlier threshold for the LO-RANSAC geometric consistency verification, which can be used to indirectly determine the value of k of the top-k highly ranked images. We tested our approach on standard benchmark datasets (Oxford 5k, Oxford 105k, Paris 6k, Oxford 1M, and Paris 1M) and found that it outperforms a BoVW baseline, yields a significant performance improvement over AQE, and preserves higher robustness to query degradations.

(別紙様式 3)
(Separate Form 3)

博士論文の審査結果の要旨
Summary of the results of the doctoral thesis screening

Query Expansion for Visual Search Using Data Mining Approach
Siriwat KASAMWATTANAROTE

本論文は、Query Expansion for Visual Search Using Data Mining Approach (画像検索のための問い合わせ拡張におけるデータマイニング技術を用いたアプローチ)と題し、画像を問い合わせとした物体検索の高精度化に関する技術について述べている。画像を問い合わせとした物体検索技術は、大量の画像データベースを効果的に活用するためにきわめて重要な情報である。問い合わせに合致した画像検索する技術の高精度化は様々な研究者らにより検討されている重要な研究課題である。本論文は、データマイニング技術に基づく画像検索のための問い合わせ拡張技術による画像検索の高精度化について検討しており、初期検索結果内の頻出共通局所特徴をデータマイニング技術により発見して問い合わせ拡張を行う提案手法である Query Bootstrapping (QB)、データマイニング技術に必要なパラメータであるサポートの自動決定手法、幾何整合性を考慮した QB の拡張と幾何整合性判定に必要なパラメータの自動決定技術について検討を行い、英文にてまとめている。

第一章 Introduction(序論)では、本研究の動機、画像検索の構成、その関連技術ならびに課題、本論文の貢献についてまとめている。

第二章 Literature review (関連研究)では、本研究に深く関連する研究について広範囲に調査している。

第三章 Query Bootstrapping: A Visual Mining based Query Expansion (Query Bootstrapping: ビジュアルマイニングに基づく問い合わせ拡張)では、本論文の中核をなす QB 技術について、説明している。特に、頻出アイテムマイニング技術を用い、初期検索結果内の頻出共通局所特徴を発見する技術、それに基づいて問い合わせ拡張を行う技術について説明している。

第四章 Query Bootstrapping extended (Query Bootstrapping 拡張)では、QB 技術の拡張として、頻出マイニング技術のためのパラメータの自動決定技術、ならびに幾何整合性を考慮した QB の拡張と幾何整合性判定に必要なパラメータの自動決定技術について検討している。

第五章 Speed-up mining process (マイニング処理の高速化)では、本手法で最も計算時間のかかる頻出アイテムマイニング処理を、データ行列の転置により高速化する手法について検討している。

第六章 Experimental setup, evaluations, and discussion (実験、評価と議論)では、様々な視点に基づく実験について報告しており、事例を交えて提案手法の利害得失を明らかにしている。

第七章 Conclusions (結論)にて本論文の成果をまとめている。

本論文で提案している画像検索手法は、ベンチマークデータに基づく評価によると、既存手法を上回る高い検索性能を達成していることが示されている。様々なデータに基づく多面的な実験により提案手法の利害得失を明らかにしており、本論文の信頼性も高いと考えられる。画像検索の高精度化は社会的要請も高く、本論文で得られた知見は学術的な意義も高い。本論文の根幹部分は、査読付き論文誌である電子情報通信学会論文誌にてすでに公表済みであり、学術的にも評価されている。このように、本論文の画像検索技術に関連

(別紙様式 3)
(Separate Form 3)

する学術的・社会的貢献は少なくないと考えられる。

以上に基づき審査した結果、本論文は学位を授与するのに十分なレベルであるものと判定した。