論 文 の 内 容 の 要 旨

論文題目　Extended Inverted Index and Fisher Kernel Approaches for Binary Local Feature-based Image Retrieval
（バイナリ局所特徴による画像検索のための拡張転置インデックスおよびフィッシャーカーネル）

氏　　名　　内田　祐介

With an increasingly wide-spread use of mobile devices such as Android phones or iPhones, mobile visual search (MVS) has become one of the major applications of image retrieval and recognition technology. With MVS, we can recognize the surrounding world with mobile devices using its built-in camera as an input to image recognition or retrieval systems. The recognition or retrieval results might be effectively shown using augmented reality technology on mobile devices for instance. Thus, mobile devices are now one of the best platforms for image retrieval systems.

While some research focuses on server-client systems in the context of MVS, we assume the situation that visual search is performed directly on the mobile device. We call the latter type of MVS "local MVS". Local MVS does not require any server and it works without a network, realizing faster recognition. In this thesis, we aim at developing a practical MVS system focusing on recent binary local features for efficiency. Although the performance of mobile devices has been improved, it is not sufficient to use non-binary features that requires a heavy processing load. Surprisingly, there are few studies focusing on the image retrieval using binary features. In many studies, binary features are used for feature-level matching between image pairs, not for image retrieval. One reason why binary feature-based image retrieval is not well-studied is that binary features are considered to be not enough robust to apply them to an image retrieval problem. However, we believe that binary features are very important for image retrieval on mobile devices and that there is a large room for improvement in the accuracy of binary feature-based image retrieval if we consider the recent significant advances in image representations for non-binary features. Thus, in this thesis, we explore the potential of binary features in the area of image retrieval and establish the

basis of binary feature-based image retrieval that can be used to real applications. To this end, we propose and evaluate three approaches in order to achieve a practical binary feature-based image retrieval system.

First, we propose the application of the Fisher vector representation to binary features aiming at improving the accuracy of binary feature-based image retrieval by considering underlying distribution of binary features. Main contribution of this approach is to model binary features using the Bernoulli mixture model (BMM) and derive the closed-form approximation of the Fisher vector of BMM. To the best of my knowledge, this is the first time to model binary features with BMM and apply the Fisher vector approach. We show that, by modeling binary features with BMM, it becomes possible to evaluate how informative different binary features are. Experimental results show that the proposed Fisher vector outperforms the BoVW method on various types of objects.

Second, we propose a substring extraction method that extracts informative bits from original binary vector and stores in the inverted index in order to improve the bag-of-visual words framework. These substrings are used to refine visual word-based matching. This is the first time to bring the idea of the Hamming embedding method to binary features. The advantage of this approach is its practicability. The developed system is very simple but effective, achieving good trade-offs between search precision, memory requirement, and speed. In addition, a modified version of the local naive Bayes nearest neighbor scoring method is proposed in the context of image retrieval, which considers the density of binary features in scoring each feature matching. The proposed system can retrieve the database with one million images in 87 [ms] and its accuracy significantly outperforms that of the state-of-the-art local MVS system.

Finally, we propose to integrate the advantages of the above two approaches. Starting with general match kernel, we show that the Fisher kernel-based similarity measurement can be implemented using the extended inverted index structure. Using the assumption that posterior probability is peaky, the Fisher kernel is linked with the BoVW framework, resulting in two proposed method, namely BMM-VW and BMM-FK. BMM-VW is a variant of BoVW, where VWs are defined by the BMM components. BMM-FK is the modified version of the second approach, where more appropriate similarity measurement is used. In order to ensure real-time applications, the method called randomized BMM trees is also proposed, which significantly accelerates the

calculation of the quantization in BMM-VW and BMM-FK. In experiments, it is shown that the BMM-FK significantly outperforms the two previous approaches and the conventional state-of-the-art system in terms of the image retrieval accuracy.

We have developed real applications based on the above approaches, where a stand-alone system, a server-client system, and a hybrid system are used as a backend. Through these practical applications, it has been proven that our developed systems have sufficient potential for practical usages.