ON OPTIMAL SOLUTIONS FOR MOBILE IMAGE RETRIEVAL APPLICATIONS

Peter Vajda, Ivan Ivanov, Lutz Goldmann, Jong-Seok Lee, Touradj Ebrahimi

Multimedia Signal Processing Group – MMSPG
Institute of Electrical Engineering – IEL
Ecole Polytechnique Fédérale de Lausanne – EPFL
CH-1015 Lausanne, Switzerland
{peter.vajda, ivan.ivanov, lutz.goldmann, jong-seok.lee, touradj.ebrahimi}@epfl.ch

1. MOTIVATION & SIGNIFICANCE

Thanks to advances in digital acquisition, processing, and storage technologies, millions of images are captured every day and shared in online social services such as Facebook[1], Flickr[2], and Picasa [3]. Furthermore, images provide an interesting way to identify or to find desired objects and locations. Image based search and retrieval is becoming increasingly popular to annotate images in large databases and for their retrieval.

With around 60% world wide penetration, mobile phones are by far the most popular electronic devices ever used. In addition to basic functionalities, modern mobile phones provide other features such as internet connection and embedded cameras. These features provide an intuitive human computer interface for web search on the go. Instead of a traditional text-based query which is quite inconvenient given the constraints of mobile phones, the query is simply formulated by capturing a photo of the object of interest. The search application will then use that photo to find similar instances of that object in a database, and provides users with associated information or services.

The limited performance and diversity of mobile devices makes a large-scale deployment of complex mobile applications rather difficult. Therefore, alternative approaches such as cloud computing [1] can be used to resolve this problem. The basic idea it to provide "convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." [2]. This centralized architecture of data, applications and computational power may be accessed and used by any mobile device with internet connection.

In many cloud applications, mobile devices are considered as thin clients in charge of data capture, rendering, and communication. While this centralized architecture maybe suitable for text, speech, and even audio-based applications, it is not efficient for image or video due to the large amount of data which has to be transmitted over a wireless channel with limited capacity. In order to reduce the information which has to be transmitted, recent work on mobile image search and retrieval (Kooaba[4], Google Goggles[5], Snaptell [6]) transfers search related computing such as feature extraction and matching from the server to the client. In a decentralized architecture the complete image search may be implemented on the mobile device.

The goal of this paper is to compare different architectures for mobile image search to derive an optimal framework given the current technologies. Furthermore, we also answer the question of feasibility and efficiency of cloud computing in this context.

2. SUMMARY & OUTLINE

The comparison of the different architectures for mobile image search is conducted in the following scenario. When a user is interested in an object and wants more information about it, he/she takes a photo with the camera of his/her mobile phone and queries the application using this image. The application matches the object with a (distributed) database from different sources, such as Wikipedia[7], Amazon[8] and eBay[9]. If a match is found, the application returns associated information such as a detailed description, product price or other pointers.

Given this mobile image search scenario the following issues have to be considered:

- In comparison to stationary devices such as servers, mobile devices are limited in terms of computational

[1] http://www.facebook.com/
[2] http://www.flickr.com/
[3] http://picasa.google.com/
[4] http://www.kooaba.com/
[5] http://www.google.com/mobile/goggles/
[6] http://www.snaptell.com/
[7] http://www.wikipedia.org/
[8] http://www.amazon.com/
[9] http://www.ebay.com/

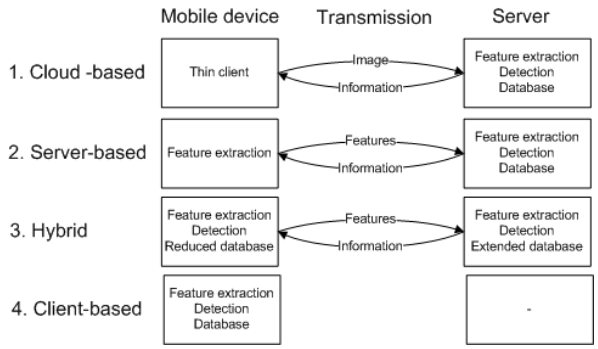

Fig. 1. Illustration of possible configurations for mobile image search with varying distribution of processing steps between mobile device and server.

power, available memory and autonomy. Therefore the amount of data which is stored and the complexity of the used algorithms have to be much lower.

- The mobility offered by these devices is mainly due to wireless transmission channels such as 2G, 3G and WiFi. While wireless transmission channels are widely available, the provided bandwidth in some locations could be limited, potentially leading to unacceptable latencies.
- While the performance of 2D object recognition has improved considerably over the last decade, 3D object recognition is still not robust enough and requires improved algorithms for feature extraction and matching.

Given an input image of the object of interest, retrieval algorithms usually consists of the following steps [3]: keypoint detection, feature extraction, feature compression, feature matching and topology verification. Considering a client-server architecture, these steps maybe distributed in different ways between the clients and servers. In this paper we compare the following alternative configurations for mobile image retrieval, as depicted in Fig. 1, covering a pure server side all the way to a pure client side configuration:

1. *Cloud-based search:* Full server side configuration, where just a thin client is used on the mobile phone. The captured image is directly transmitted to the server, where all the processing steps are performed. As a result the retrieved information is sent back to the client. Cloud application allows scalable selection of content, features and resources.
2. *Server-based search:* In order to reduce the required bandwidth for the transmission, the feature extraction and compression are performed on the mobile phone [4, 5]. The features are sent to the server, where the remaining steps are completed. The resulting information is sent back to the mobile.
3. *Hybrid search:* In a hybrid configuration retrieval is divided in two steps [3]. On the mobile phone a first retrieval is performed on a reduced dataset with recent or frequent content stored locally. An extended retrieval is then performed on the server-side using the query image as well as eventual results of local query.
4. *Client-based search:* The retrieval is performed on the mobile [4], using only locally stored datasets and less complex algorithms. This configuration is specially appealing for clients with very large storage capacity.

Considering these configurations, the current state-of-the-art in feature extraction, feature matching and topology verification, and based on a number of well motivated performance metrics, we will answer to the following questions:

1. What is the optimal architecture with respect to computational complexity, memory requirements, storage capacity, transmitted amount of data in typical mobile image retrieval applications?
2. How efficient is cloud computing for image search and retrieval on mobile phones in typical mobile image retrieval applications?
3. What techniques related to the individual steps in query are more suitable for mobile search?

Although the final article will be written as a tutorial, it will describe in more details the above mentioned configurations with an implementation example on an Android-powered mobile phone (Samsung i7500), and analyze and evaluate their respective performance in a realistic application scenario using data from Wikipedia[10], Amazon[11], and eBay[12].

3. REFERENCES

[1] A. Weiss, "Computing in the clouds," *COMPUTING*, vol. 16, 2007.

[2] "Cloud computing definition, national insitute of standards and technology," Available at: `http://csrc.nist.gov/groups/SNS/cloud-computing/index.html`.

[3] John Hong, "Computer vision for the masses," in *Workshop on Mobile Visual Search*, 2009.

[4] Bernd Girod, "Mobile visual search," in *Workshop on Mobile Visual Search*, 2009.

[5] V. Chandrasekhar, G. Takacs, D. Chen, S. Tsai, R. Grzeszczuk, and B. Girod, "Chog: Compressed histogram of gradients a low bit-rate feature descriptor," *CVPR*, vol. 0, pp. 2504–2511, 2009.

[10] `http://www.wikipedia.org/`
[11] `http://www.amazon.com/`
[12] `http://www.ebay.com/`