

# Rate-Efficient, Real-Time CD Cover Recognition on a Camera-Phone

Sam S. Tsai\*, David Chen\*, Jatinder Pal Singh† and Bernd Girod\*

\*Information Systems Laboratory, Stanford University, Stanford, CA 94305, U.S.A.

†Deutsche Telekom Laboratories, 128 Spear Street 4th Floor, San Francisco, CA 94105, U.S.A.

\*{ssttsai, dmchen, bgirod}@stanford.edu, †jatinder.singh@telekom.de

## ABSTRACT

Automatic CD cover recognition has interesting applications for comparison shopping and music sampling. We demonstrate a real-time CD cover recognition using a camera-phone. By snapping a picture of a CD cover with her camera-phone, a user can conveniently retrieve information related to the CD. Robust image feature extraction is applied to overcome the image distortions in the query photo. To limit the amount of data transmitted over a wireless network, we compress the query image or features extracted from the query image. On the database side, fast and reliable image matching against a database of 10,000 CD covers is accomplished using a scalable vocabulary tree.

## Keywords

scale-invariant feature, scalable vocabulary tree, mobile augmented reality, content-based image retrieval

## 1. INTRODUCTION

We present a real-time CD cover recognition system for a mobile camera-phone. A customer in a music store can snap a photo of a CD cover while shopping. The query image or a set of sparse features extracted from the query image is then uploaded to a server through a wireless connection. The server contains a large database of CD covers. Using the received query information, the server must quickly and reliably identify the matching CD cover. The server responds with the identity of the CD, and provides any other type of data the user would require of the particular CD, such as comparative prices at multiple music stores or short samples of songs on the CD.

There are three major challenges in real-time CD cover recognition using a camera-phone. The first problem is that photos of CD covers taken by mobile phones typically suffer from many geometric and photometric distortions, making the task of automatically matching a query CD cover to a clean database CD cover challenging.

These image distortions can be overcome by performing image matching using robust feature descriptors. Scale invariant feature transform (SIFT) was first proposed by Lowe et al. [1] to perform object recognition with feature descriptors that are invariant to mild rotations, scale changes, and illumination differences. Later, speeded up robust features (SURF) was proposed by Bay et al. [2] with lower complexity and comparable image matching accuracy.

The second problem is that a real-time recognition system

must meet stringent time constraints. A good recognition algorithm should scale gracefully with the size of the database. Even for a large database, the time to calculate accurate image matches must be small to provide the user a fairly interactive experience. The scalable vocabulary tree (SVT), proposed by Nistér and Stewénus [3], provides a scalable solution to the problem. It efficiently reduces the search space to a small subset of most likely database matches.

Additionally, the third problem is that network bandwidth is very limited. A 3G cellular network has a typical uplink of 50-80 kbit/sec. To keep data transmission times low, the query information for image recognition must be concise and compressed. A rate-efficient recognition scheme is needed.

In our system, only the feature descriptors and their locations are used by the server to identify the CD. A feature descriptor compression system developed by Chandrasekhar et al. [4], is applied to greatly reduce the information needed to be sent across the wireless network with minimal degradation in matching accuracy.

## 2. CD COVER RECOGNITION SYSTEM

Interaction between a mobile phone and a remote server takes place as depicted in Fig. 1. A query is initiated by sending information of the query object to the server through a wireless connection. The query may be either the image that the user has taken, or the feature descriptors extracted from the query image. Then, assuming correct image matching, the server returns the identity of the CD to the user along with the desired additional information.

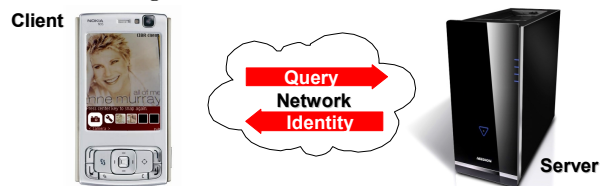


Figure 1: Camera-phone CD recognition system

### 2.1 Query Modes

To accommodate phones with different capabilities, we have implemented two different query modes. The first is the *Send Image* mode as illustrated in Fig. 2. For mobile phones with very limited computational capabilities, robust, scale-invariant feature extraction would take too long on the phone. In this case, a JPEG-compressed query image is sent to the server, and the server performs the feature extraction.

The other mode is the *Send Feature* mode, depicted in Fig. 3. Advanced mobile phones with embedded processors can perform SURF feature extraction on the device. Then, applying compression techniques, we can compress the feature

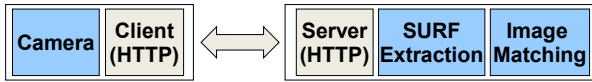


Figure 2: Building blocks of *Send Image* mode

descriptors significantly, achieving up to 20-fold data size reduction compared to floating point [4]. Upon receiving the descriptors, the server decompresses the feature descriptors and uses the received query feature set to search for the matching database CD cover.

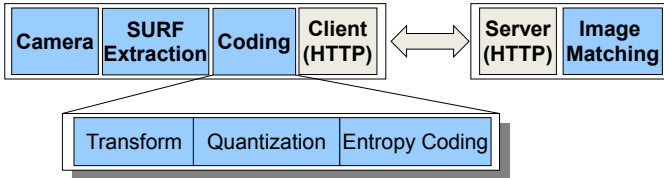


Figure 3: Building blocks of *Send Feature* mode

## 2.2 Searching through the Database

On the server side, we have implemented a database query system with 10,000 CD images, representing albums for 1000 different performers. A scalable vocabulary tree of depth 6 and branch factor 10 is built from several million SURF features.

When a query is received, the SURF feature descriptors are first extracted if *Send Image* mode is chosen. The descriptors are then classified through the SVT and a list of most probably matching CD images is generated. These most likely candidates are further processed through a ratio test and geometric consistency checking. The candidate with the most number of feature matches after all tests is reported as the best database match.

## 2.3 User Interface on the Camera-Phone

On the client side, we have developed an interface on the camera-phone. The interface has three main components: (1) a viewfinder for displaying scene contents in front of the camera (Fig.1), (2) an options menu for changing the recognition system's settings (Fig.4 (a)), and (3) a bar of recent query results (Fig. 4 (b)).

When the user takes a photo in the viewfinder, the image is compressed into JPEG format. Additionally, feature extraction is performed if *Send Feature* mode is chosen. Query data is transmitted to the server. If the server identifies the CD correctly, it will return the identity of the CD, and ask if a short audio sample should also be downloaded and played on the phone, as shown in the presentation.

## 3. PERFORMANCE AND EVALUATION

We demonstrate our system with the query database server running on a Linux system with 2GHz Xeon processor and 4G RAM. The client platform is the Nokia N95 camera-phone. Connection between the server and client can be either the 3G cellular network or the 802.11 WLAN. The user can compare the response time for operation over the two different types of network.

Overall system timing numbers are presented in Table 1. The *Send Image* and *Send Feature* modes are compared. The experiments were conducted over the 3G cellular network with a measured round trip time of one second. By compressing the feature descriptors, the query response time and quantity of transmitted data over the network is greatly reduced.

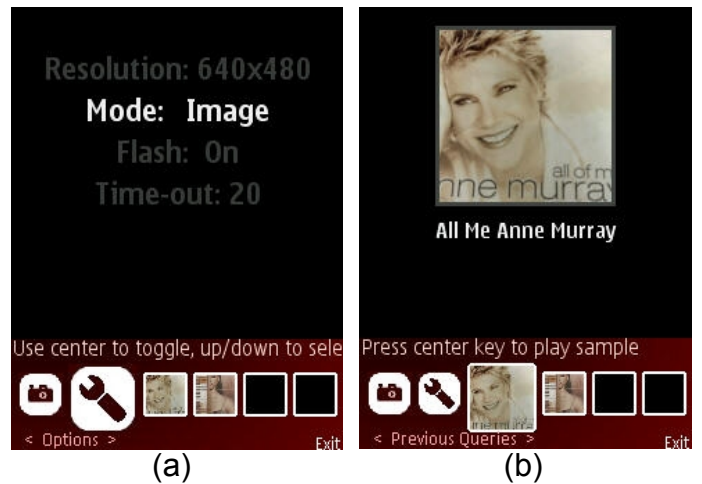


Figure 4: (a) Options menu (b) Recent queries

Mode	Send Image	Send Features
Photo Capturing and JPEG Compression	1-2 sec	1-2 sec
Feature Extraction and Compression	0 sec	8 sec
Upload to Server	4-5 sec	2-3 sec
Querying on Server	4-6 sec	3-5 sec
Overall Time	9-13 sec	14-18 sec

Table 1: Time spent on query steps

Our system achieves a recognition accuracy of over 90% for challenging query images with many geometric distortions, nonideal lighting, glare from camera flashes, and changing backgrounds, as shown in the presentation. The user can test the robustness of our system by snapping photos of CDs in different scenarios.

## 4. CONCLUSION

A real-time CD recognition system using the camera-phone is presented. Robust image feature extraction techniques are applied to overcome the matching imparities due to image distortions taken by the camera-phone. To limit the amount of data required sent over the network, feature compression techniques were applied. Fast and reliable CD recognition among a database of 10,000 CD images is achieved using a scalable vocabulary tree. The complete system has a user-friendly interface that makes CD cover recognition on the camera phone a simple and enjoyable experience.

## 5. REFERENCES

- [1] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [2] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," in *European Conference on Computer Vision*, Graz, Austria, May 2006, pp. 404–417.
- [3] D. Nistér and H. Stewénus, "Scalable Recognition with a Vocabulary Tree," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006, vol. 2, pp. 2161–2168.
- [4] V. Chandrasekhar, G. Takacs, D. Chen, S. S. Tsai, J. Singh, and B. Girod, "Transform Coding of Image Feature Descriptors," *submitted to Visual Communication and Image Processing*, Jan 2009.