Book Cover Recognition Using Binary Local Feature and Color Information

Dung Phan, Soo-Hyung Kim
School of Electronics and Computer Engineering
Chonnam National University
Republic of Korea
dungptmy@gmail.com, shkim@chonnam.ac.kr

***Abstract*—In this paper, we propose an approach for book cover recognition using local invariant feature and significant bit-plane image. Local feature based object recognition has been researched over the past few decades. However, this approach ignored color information which is very useful for object recognition. It can cause wrong recognition for objects whose structures are similar. Therefore, we propose a method to combine local invariant feature and color information to make robust object recognition system. In our method, we extract and describe features using local feature detector and descriptor. Then, Random Sample Consensus (RANSAC) algorithm is used to generate homography matrix based on the local feature matches of query image and reference object image. Using the homography matrix, we can detect object location in query image, and then remove background image. Next, we calculate global color histogram from RGB color. Similarity of query image and reference object image is a linear combination of color histogram correlation and number of feature matches. As a result, the proposed method can overcome drawbacks of object recognition method using local features. By obtaining higher accuracy, the system proved that combination of local invariant feature and color information is very effective for book cover recognition.**

Keywords—book cover, object recognition, local invariant feature, significant bit-plane image, color histogram, homography, RANSAC algorithm

I. Introduction

A book cover recognition system can be applied to bookstore and library. Whenever customers or students want to get information of a book they can capture the image and search it on this system. One of common approach for book cover recognition or retrieval is content-based image retrieval (CBIR). In a typical CRIR, there are many kind of features have been used such as shape, texture, color and local invariant features. K. Iwata developed a novel library system [1] which used book cover image and borrower's face. The system can automate the administration of borrowing and returning in cooperation with the user. This system got the scanned book cover image. Gray scale mosaic feature, RGB color mosaic feature and edge feature are extracted. These features are matched by using partial area matching. Then, he improved his method by using four directional features in [2]. A book retrieval system based on near duplicate image matching is proposed deal with captured image from camera in [3]. In their system, they used the scale-invariant feature transform (SIFT) as a feature for matching. Guanghui Hu presented a book cover retrieval system using enhanced SIFT in [4]. Most of approach gave high accuracy with their own database which is not included variation such as rotation, transformation, lighting changing, etc. The traditional local feature based object recognition often missed the color information. Even though, the color information is used, it always reflects to whole image which includes background and sometimes noises that can cause wrong recognition. Moreover, in real system, we required a method which is not only obtained high accuracy but also significantly processing time. Therefore, we propose an approach for book cover recognition using local invariant feature and color feature. In our system, we extracted local features by using local feature detector and descriptor. As a state-of-art of local invariant feature, the oriented FAST and rotated BRIEF [5] (ORB) [6] is used in our system. ORB is a new feature descriptor building on the well-known FAST keypoint detector [7] and BRIEF descriptor [5] which acquire good performance and low cost. In additional, color feature is extracted from the object after detection. So, the color feature only reflects to the object completely. Thus, combination of these features is very effective for book cover object recognition.

The related work in section II presents summary of ORB feature. The proposed method using combination of ORB feature and significant bit-plane image is explained in section III. Section IV shows the experimental result with comparison to other methods. Finally, conclusion of our works and future works is discussed in section V.

II. Background

A. ORB Feature

ORB feature builds on the FAST keypoint detector and the BRIEF descriptor. Both these techniques are attractive for their good performance and low cost. ORB feature starts by detecting FAST feature points in the image. A point is classified as a corner if one can find a sufficiently large set of pixels on a circle of fixed radius around the point such that these pixels are all significantly brighter (or darker) than the central point, as shown in Fig.1. Efficient corner classification is based on a trained decision tree. ORB use FAST-9 (circular radius of 9), which has good performance. However, FAST features do not have an orientation component, which is important for description extraction. ORB feature uses a simple but effective measure of corner orientation, the intensity centroid. The intensity centroid assumes that a corner's intensity is offset from its center, and this vector may be used to compute an orientation. When the keypoints are detected, the ORB feature computes a steered BRIEF descriptor for every keypoints. The BRIEF descriptor is a bit string description of an image patch constructed form a set of binary intensity tests. Consider a smoothed image patch, p. A binary test is defined by:

$$\tau(\mathbf{p}; \mathbf{x}, \mathbf{y}) := \begin{cases} 1 & : \mathbf{p}(\mathbf{x}) < \mathbf{p}(\mathbf{y}) \\ 0 & : \mathbf{p}(\mathbf{x}) \geq \mathbf{p}(\mathbf{y}) \end{cases}$$

where p(x) is the intensity of p at a point x. The feature is defined as a vector of n binary test:

$$f_n(\mathbf{p}) := \sum_{1 \leq i \leq n} 2^{i-1} \tau(\mathbf{p}; \mathbf{x}_i, \mathbf{y}_i)$$

To be invariant to in-plane rotation, ORB feature steer BRIEF according to the orientation of keypoint. However, steered BRIEF has significantly lower variance which will make the final descriptor redundant and harm the discriminant perfromance. ORB feature develops a learning method for choosing a good subset of binary tests with high variance. The final descriptor is about two orders of magnitude faster than SIFT, while performing as well in many situations.

B. Color Feature Extraction and Historgram Correlation

Given a color space C, the conventional color histogram H of image I is defined as equation (1).

$$H_C(I) = \{N(I, C_i) | i \in [1, \dots, n]\} \qquad (1)$$

where $N(I, C_i)$ is the number of pixels of image I that fall into cell C_i and i indicates the color levels of color space C. $H_C(I)$ represents the distribution of pixels of each color in image.

Color histogram correlation is defined as equation (3).

$$d(H_1, H_2) = \frac{\sum_I (H_1(I) - \bar{H}_1)(H_2(I) - \bar{H}_2)}{\sqrt{\sum_I (H_1(I) - \bar{H}_1)^2 (H_2(I) - \bar{H}_2)^2}} \qquad (3)$$

where

$$\bar{H}_k = \frac{1}{N} \sum_I H_1(I)$$

III. PROPOSED METHOD

We propose a method using combination of ORB feature and color information. The figure 1 shows diagram of proposed method. Feature extraction is described in the previous section. In this section, we explain about applying feature in our system.

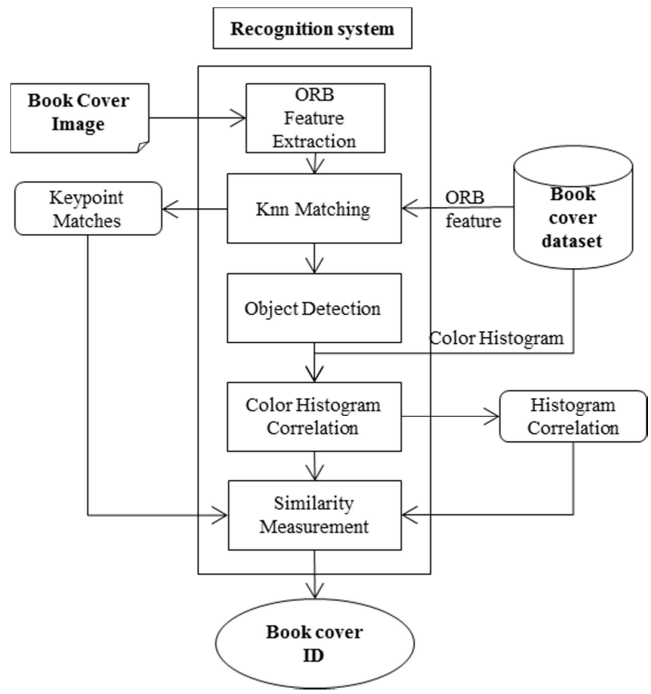

Figure 1: Flowchart of proposed System

A. Feature Matching & Object Detection

1) Feature Matching using Fast Library for Approximate nearest Neighbors(FLANN)

Feature matching is a process in which keypoint descriptors from two or more images are compared to find matches and then measure the similarity of these images. A popular method is used to match the local feature is Nearest Neighbor search between two sets of keypoints.

Given two image I_1 and I_2, the n-dimensional vector space is formed as equation (2)

$$\forall k_i \in K_1 : l_i = \min_{l \in K_2} |d(k_i) - d(l)| \qquad (2)$$

In our system, we applied Local Sensitive Hashing (LSH) method integrated to FLANN [8] to speed up FLANN. The FLANN method uses priority queue to expand searching according to the distance of each k-means domain from the query. The LSH method [9] relies on projection functions that map similar data points into the same buckets that can be efficiently accessed in hamming space.

2) Object Detection by estimating perspective transfomration using Random Sample Consensus Algorithm (RANSAC)

RANSAC is a powerful estimation method used to find the perspective transformation between two images. Iteratively, RANSAC picks a random subset of matches from the putative match list and fits a model (homography) to them. For fitting an affine homography, a minimum of four element subset is required. This usually termed a minimal subset. The model is tested against all other correspondences in the putative match list. Correspondences that fit the model are considered as hypothetical inliers, otherwise they are considered as hypothetical outliers. After a fixed number of iterations, the model with highest number of hypothetical inliers is selected. Then we transform the keypoints from object in dataset and get the object in query image. The figure 2 shows an example of detecting object in query image.

Once perspective transformation matrix is obtained, we extract the region of object in the query image by building a convex polygon around the keypoint of query image which is in match list. The color histogram is calculated from this area. Similarity between color information is the correlation between two color histogram of region of query object and object in database.

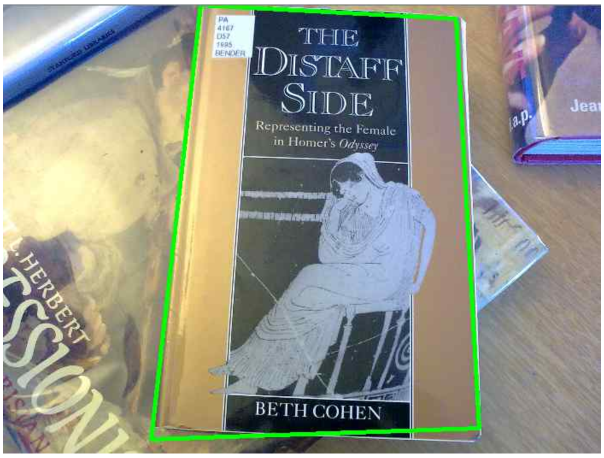

Figure 2: Example of Object Detection

B. Object Recognition based on similiraty of template matching

An object is recognized if it can be detected in query image. That means we can estimate at least a perspective transformation between query image and object image in dataset. Then choose the most similarity object image from dataset. The similarity measurement is formed as equation (4)

$$s(I_1, J_1) = \frac{num\_matches\,(I_1, J_1)}{\max\_of\_num\_matches} + d(H(I_1), H(J_1)) \quad (4)$$

where

$$\max\_of\_num\_matches = \max_{J \in V}(num\_matches\,(I_1, J_i))$$

I indicates a query image, J indicates an image in dataset and V is set of images in dataset.

IV. EXPERIMENTAL RESULT

Our system has been implemented in Visual Studio 2010 environment. For evaluation, we experimented with the Stanford Mobile Visual Search dataset which has several key characteristics: widely varying lighting condition, perspective distortion, typical foreground and background clutter collected from heterogeneous low and high-end camera phones. The database includes 505 images of 101 book cover images. For each book cover, there are 4 images taken from 4 different devices (IPhone, Cannon, Droid, 5800) and 1 reference image. The figure 2 shows examples of dataset with many challenges.

Table 1 shows the comparison between using SIFT feature and ORB feature. Although the SIFT feature gave high accuracy, processing time is not good for a real-time system. With the ORB feature, we can reduce the processing time. In comparison table, we try with many case of using ORB to get the best result for our system. However, with the limitation of number of dataset, we evaluate our system relatively on this dataset.

(a)
(b)
(c)
(d)
(e)

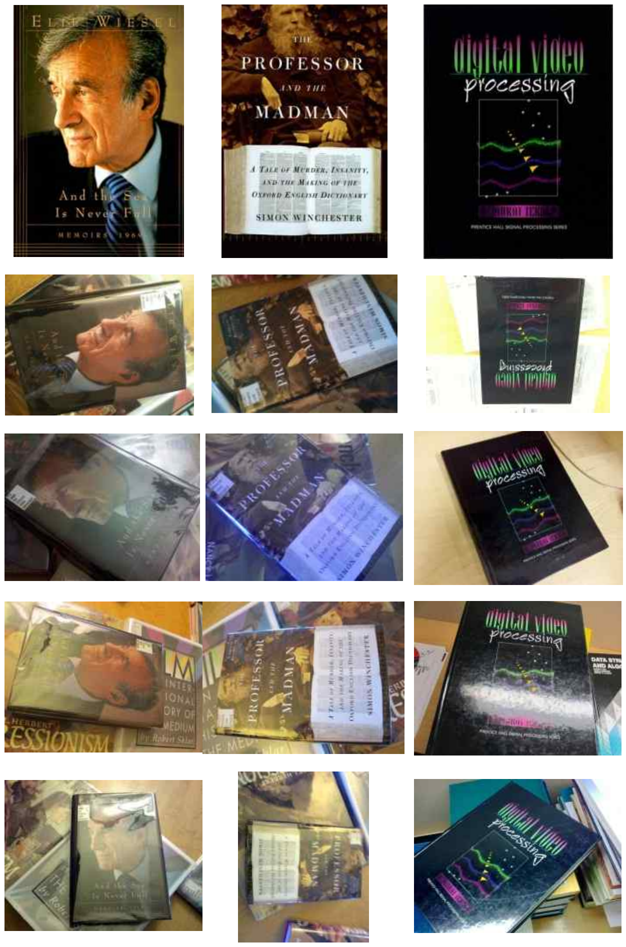

Figure 3: Examples of dataset with many challenges: (a) Reference Image; (b),(c),(d),(e) Images captured from many devices.

Table 1: Result Comparison between different kind of using features

Method	Accuracy	Time/Query
SIFT	95%	8s
ORB(500)	89%	0.542s
ORB(500) + Color	91%	0.646s
ORB(1000)	95%	1.5s
ORB(1000) + Color	95%	1.65s
ORB (1500)	96%	3s
ORB(1500) + Color	96%	3.026s

Note: ORB(N) means using N ORB features for recognition.

V. CONCLUSION

A book cover recognition system is proposed in this paper. In this system, we combine invariant local feature and global feature, color histogram to recognize a book cover object in the image. Invariant local feature has been researched and became robust over few decades. Besides, color information is one of important feature for object recognition that always reflects a lot of information of object. Then, combination of local feature and global color information is become more robust which is show in the experimental result. However, dataset is used to evaluate the system is not large. So, it is not make an evaluation exactly .In additional, a real-time application like recognition system always requires an optimization on processing time. So, we are going to improve the system to meet requirement of a real-system by optimizing processing time and evaluating with large dataset.

ACKNOWLEDGMENT

REFERENCES

[1] K. Iwata, J. Yamamatsu, K. Yamamoto, K. Kato and M. Ishida, "A Novel Library System", Proc. Vision Interface 2000, pp.3 18-322, 2000.

[2] K. Iwata, K. Yamamoto, M. Yasuda, K.Kato, M. Ishida, K. Murata, "Book cover identification by using four directional features field for a small-scale library system", Proc. 6th International Conference on Document Analysis and Recognition, 2001.

[3] Xun Tang "Book Retrieval Based on Near-Duplicate Image Matching", 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012).

[4] Guanghui Hu, Yong Zhao "Book Cover Retrieval Using Enhanced SIFT", 2012 IEEE fifth International Conference on Advanced Computational Intelligence (ICACI) October 18-20, 2012.

[5] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features", Proc. European Conference on Computer Vision, 2010.

[6] Rublee and, V. Rabaud, "ORB: an efficient alternative to SIFT or SURF", International Conference on Computer Vision, 2011, pp. 2564 -2571.

[7] Rosten and T. Drummond., "Machine learning for high-speed corner detection", In European Conference on Com-puter Vision, volume 1, 2006.

[8] Marius Muja and David G. Lowe, "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration", in International Conference on Computer Vision Theory and Applications (VISAPP'09), 2009.

[9] Qin Lv, William Josephson, Zhe Wang, Moses Charikar, Kai Li., "Multi-Probe LSH: Efficient Indexing for High-Dimensional Similarity Search", Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB). Vienna, Austria. September 2007.