Hybrid Feature-based Wallpaper Visual Search

Kim-Hui Yap and Zhenwei Miao
School of Electrical and Electronic Engineering,
Nanyang Technological University, Singapore 639798
Email: {ekhyap, zwmiao}@ntu.edu.sg

***Abstract*—In this paper we propose a hybrid feature-based wallpaper visual search system. As opposed to conventional techniques that use global features to perform wallpaper search, this paper proposes to integrate local and global features to support both functions of recognition (identify the product ID of the query images) and retrieval (search wallpapers that are visually similar to the query images). An adaptive SIFT is designed to extract sufficient number of local features from both the query and reference images. The combination of the sparse and dense SIFT features results in a significant improvement of the recognition rate. Global features are further incorporated in the system for the visually similar image retrieval. A new query expansion is proposed to alleviate the problems caused by cluttered background, occlusion, scale change and illumination changes. Experiments on a dataset consisting of 2,208 reference images from 218 different designs show that the proposed method can achieve a recognition rate of more than 90%.**

I. Introduction

Object visual search has been developed in different emerging applications, including landmarks, trademarks, book covers, and CD covers, etc. [1]–[6]. The developments have drawn the attention of wallpaper design companies. In the area of wallpaper design, companies need to identify a particular wallpaper design through voluminous catalogues based on user (e.g., interior designers) queries. Currently, this has to be done by manually looking through large image databases. This process is manual-intensive, cumbersome, and error-prone. In view of this, a cost-effective and efficient solution using computer vision algorithms is highly desirable.

Research on wallpaper visual search is relatively limited. A recent study [7] proposed a computational model for periodic pattern perception to characterize pattern structures. Another recent work [8] has been done to retrieve wallpaper images based on global features. The algorithms in [7] and [8] were tested only on a small scale wallpaper dataset, without addressing the issue of scalability. Moreover, the global features used in both studies cannot effectively address the variations arising out of scale change, rotation, and occlusion, which are commonly encountered in real-life scenarios.

As opposed to conventional techniques that use global features to performance wallpaper search, this paper proposes to integrate both local and global features in support of recognition and retrieval functions. Either local or global features alone cannot fully achieve satisfactory results given three distinctive types (i.e., painting, pattern, and fabric) of images in a wallpaper set. Samples of the three types are given in Fig. 1. The painting-type images, shown in Fig. 1(a), consist of unique and pictorial structures. The pattern-type images, shown in Fig. 1(b), consist of regular shape-based structures. The fabric-type images, shown in Fig. 1(c), contain subtle image features. Local feature-based algorithms, such as the Bag-of-Word (BoW) model [3], have a good performance in identifying the painting-type images, but not effective in dealing with the pattern- and fabric-type images. On the other hand, the global features, such as the Gabor feature [9], can efficiently describe these two types of images. However, global features are sensitive to the occlusion, rotation and scale changes. By combining both the local and global features, we expect to achieve a powerful visual search engine together with a high recognition rate. Specifically, this system proposes an adaptive SIFT (ASIFT) strategy to extract sufficient features from images. Dense SIFT (DSIFT) features are further incorporated [6] to cover each image in a compact manner. The system then integrates local feature descriptors [6], Scalable Vocabulary Tree (SVT) [3], and Geometric Verification (GV) [4] to identify the query images. For a query image, the global feature together with the local features are used to retrieve visually similar images from the database. Considering the global feature is often adversely affected by the cluttered

Fig. 1. Sample images in the wallpaper dataset. From top to down are the types of painting, pattern and fabric.

This work was carried out at the Rapid-Rich Object Search (ROSE) Lab at the Nanyang Technological University, Singapore. The ROSE Lab is supported by a grant from the Singapore National Research Foundation and administered by the Interactive & Digital Media Programme Office at the Media Development Authority.

978-1-4799-8391-9/15/$31.00 ©2015 IEEE

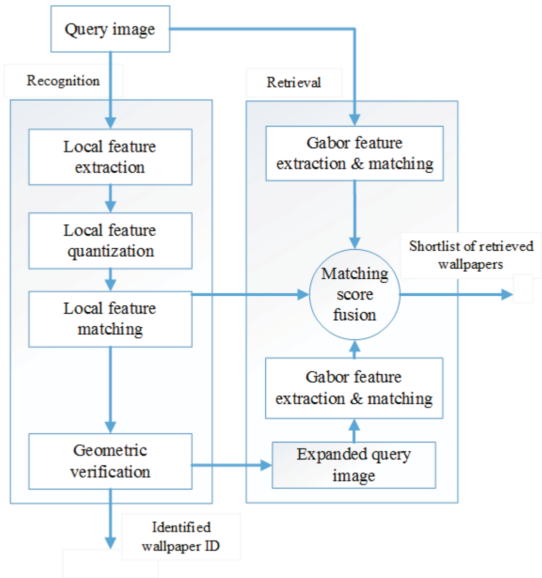


Fig. 2. Flowchart of the proposed wallpaper visual search system.

background and various variations, we propose a new query expansion algorithm to enhance the retrieval performance.

II. Wallpaper Visual Search System

Our proposed system incorporates the local SIFT and global Gabor features. It implements both functions of recognition and retrieval. The flowchart of the proposed wallpaper visual search system is shown in Fig. 2.

A. Wallpaper Recognition

The recognition function is built on the BoW framework [3], which is the state of art for real-time large scale image search. There are four main steps in the development of this function.

1) Feature extraction and description: We use the SIFT detector [6] to detect the interest points and the SIFT descriptor [6] to describe each point. However, the default SIFT cannot extract sufficient points on some of the images in the wallpaper database, as shown in Fig 3(a). In order to alleviate this problem, an adaptive strategy, named adaptive SIFT (ASIFT), is designed in this paper. Details of the ASIFT will be given Section III.

Some wallpaper images does not contain sufficient number of blob-like structures, for example, the images which contains only the horizontal or vertical linear structures. In these cases, both the default SIFT and the ASIFT cannot work well as they cannot extract sufficient number of interest points. Henceforth, the Dense SIFT (DSIFT) features are extracted in multiple scales and combined with the ASIFT to represent the image.

2) Feature quantization: Directly computing the similarity of the high dimension local features is time consuming and impracticable for large scale image search. The BoW model can solve this issue by treating the local features as a set of words [3]. The quantization of the local features to the words is achieved with a "codebook". In this work, the Scalable Vocabulary Tree (SVT), which is built on the hierarchical k-means clustering, is employed to organize the "codebook".

3) Image matching: After the local feature quantization, each image can be represented by a BoW histogram. We use the "Term Frequency-Inverse Document Frequency" (TF-IDF) weighting [3] to suppress the less discriminative visual words. The L_p-norm distance of the BoW histograms only relies on the nonzero bins in the histograms [3]. Therefore, we utilize the inverted file to effectively conduct the image matching [3].

4) Geometric verification: The BoW framework provided a tentative matching of the visual words in the query and reference images. However, this procedure does not use the location information of the visual words. Spurious match may occur in the matching pairs. Geometric Verification (GV) is to verify if the obtained feature correspondences are consistent with a geometric model. We apply the RANSAC-based GV on top ranked BoW images to improve the recognition rate and the identification confidence. For a query and reference image pair, if the number of matched points after GV is larger than a certain threshold (e.g. 7 in this paper), it is confident that this reference design ID category is the correct match of the query image. Thus, this verified reference design ID category will be used in the subsequent proposed query expansion, which will be elaborated in the following subsection.

B. Wallpaper Retrieval

The proposed wallpaper retrieval function includes three main steps.

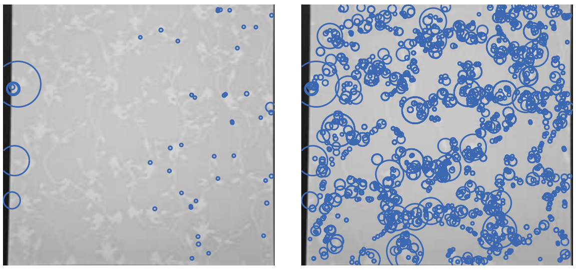
1) Query Expansion: The input query images, especially for those captured by mobile phones, often suffers from various conditions, such as illumination distortion, occlusion, orientation and scale changes. It also may contain cluttered background. This reduces its performance in retrieving visually similar image. Compared to the query image, the matched reference image obtained from the GV as aforementioned usually has high image quality. Therefore, we used the additional reference image as the expansion of the query to enhance the retrieval performance.

2) Gabor feature extraction and matching: The Gabor feature can efficiently extract the information of coarseness, periodicity, pattern complexity and inherent direction [9], which is important in psycho-physical perception of human visual system [9]. Thus, it is used as the feature for the image retrieval. However, compared to the local feature, the Gabor feature is sensitive to various changes. This sensitivity reduces the retrieval performance. In view of this, the proposed query expansion is developed. The Gabor features given in [9] is used to describe the global information of each image. The matching score is computed by the cosine distance of these Gabor features.

Let $\mathbf{I}_e$ be the matched reference image after the GV and $\mathbf{I}_q$ be the query image. Let G_e and G_q be the Gabor feature matching scores of each database image with respective to $\mathbf{I}_e$ and $\mathbf{I}_q$. The combined score G of the Gabor feature matching scores with respective to these two images in retrieval can be modeled as a linear combination of G_e and G_q, i.e.,

$$G = \frac{aG_q + bG_e}{a+b}. \tag{1}$$

In case if no matched reference images are identified during the GV process, only the Gabor matching score with respective



(a) (b)

Fig. 3. Interest points detected by (a) the default SIFT detector, and (b) the SIFT detector with the adaptive contrast threshold selection strategy. The circles represents the scale of each interest points.

to the query image will be used. Then,

$$G = G_q. \quad (2)$$

3) Matching score fusion: The global features can effectively handle the pattern- and fabric-types wallpapers but not the painting-type ones. Meanwhile, the local features are useful in dealing with the painting-type images. Therefore, we fuse the matching score from both the local SIFT feature and global Gabor feature. The similarity S between the query and retrieved image is defined as an average of the matching score of the local and global features, as

$$S = (G + V)/2, \quad (3)$$

where V is the matching score from the local feature with the BoW framework as shown in Fig. 2.

III. Adaptive SIFT Detector

The SIFT detector is widely used in computer vision because of its advantages in dealing with the scaling, rotation, occlusion, and cluttered background [6], [10], [11]. It is composed of four parts: (1) scale space generation and extreme detection, (2) unreliable points removal, (3) orientation assignment and (4) descriptor extraction.

SIFT detector employs two thresholds to remove unstable points: 1) a fixed threshold λ to remove the low contrast features which may sensitive to noise and (2) a threshold r to remove the un-blob like structures, such as the edge, ridge and straight line structures. However, the fixed contrast threshold λ may lead to some limitations in some cases. Excessive interest points may be detected in some images, while scarce in others, which leads to a poor search performance. For example, in Fig. 3(a), only few interest points are detected by the SIFT detector due to the low contrast of the image intensity. This leads to incorrect identification of the image. Therefore, in order to detect sufficient interest points, we proposed an adaptive threshold selection strategy for the SIFT detector, named adaptive SIFT (ASIFT). This simple yet effective method enables the low contrast interest point detection, a situation which is frequently encountered in query images taken by mobile phones.

The ASIFT is given in Algorithm 1. The threshold λ to remove the low contrast interest point is adjusted based on the number of detected points n. If n is smaller than a threshold N, λ will be decreased by half to include more low contrast points. A predefined maximum number T is used to limit the number of iterations. An example is shown in Fig. 3(b). It reveals that the proposed adaptive SIFT detector detects more points in the unique interest regions, which increases the discriminative capability of the local features. Experiments in Section IV shows that the performance increases with the proposed ASIFT.

Algorithm 1: Framework for the adaptive SIFT detector and descriptor

Input: image $\mathbf{I}$, $\lambda = \lambda_0$, $t = 0$; **Output**: Interest points $\mathbf{D}$ and relavant descriptors $\mathbf{F}$;
Generate the multi-scale image sets and the DoG image sets;
do
    Extract DoG extrema and refine the locations;
    Remove the low contrast points by λ and the poor definedly peaks by r, update the number of detected point n;
    $\lambda = \lambda/2$, $t = t + 1$;
while $n < N$ *and* $t < T$;
Assign orientations to the interest points $\mathbf{D}$;
Extract the descriptors $\mathbf{F}$ of the points in $\mathbf{D}$;

IV. Experiments

A. Wallpaper Dataset

Table I shows the composition of the wallpaper dataset. The reference image set contains 2,208 images which are provided by a wallpaper design company. The test image set contains 3,720 images.

TABLE I. Reference and Test Images in the Wallpaper Datasets.

Data type	No. of category	Total reference images	Total test images
Painting	77	522	1,014
Pattern	41	246	697
Fabric	100	1,440	2,009
Total	218	2,208	3,720

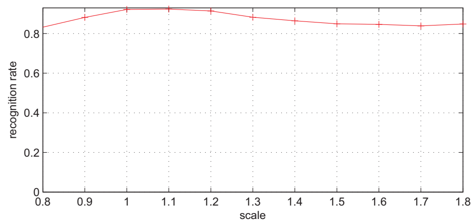
B. Recognition Rate

For this experiment, the hierarchical k-means clustering is used to train the SVT. Branch of 10 and depth of 6 are set for the SVT. A total of 1,000,000 words are trained for the quantization. The SIFT descriptor is used to describe the interest points extracted from images. While keeping the original ratio of the width to the length, the images are normalized to the standard size with the longest dimension as 640 pixels. For the adaptive SIFT detector, the number of interest points threshold N is set to be 1,000. The maximum number of iterations T is set as 4.

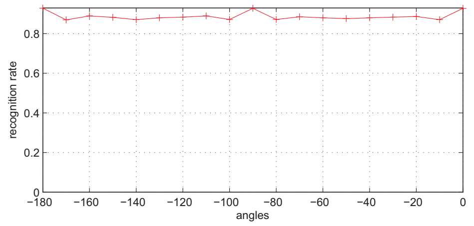
TABLE II. Recognition Rate on the Wallpaper Datasets.

Data type	Painting	Pattern	Fabric	Overall
Baseline SIFT+BoW	78.9%	63.3%	55.9%	60.5%
Proposed DSIFT+ASIFT+BoW	**97.6%**	**91.8%**	**89.1%**	**90.5%**

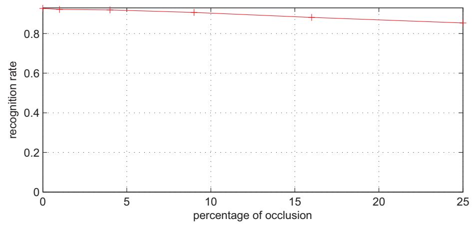
The recognition results are shown in Table II. The standard SIFT with the BoW algorithm [3] is used as the baseline. It shows that the proposed method improves the overall performance of the system by 20% when compared with the baseline. The proposed method can achieve significant improvemnt in



(a)



(b)



(c)

Fig. 4. Recognition against (a) scale change, (b) rotation and (c) percentage of occlusion.

all three different types of wallpapers, i.e, painting, pattern and fabric types.

C. Performance under Different Variations

The stability of the system under different scale change, rotation, and occlusion is evaluated on the painting- and pattern-type images. The results are shown in Fig. 4. For painting and pattern types, the system is tolerant of:

- A scale change of 0.8-1.8X on captured images will not affect the recognition rate by more than 9%.
- A rotation up to 180 degrees on captured images will not affect the recognition rate by more than 6%.
- An occlusion up to of 25% on captured images will not affect the recognition rate by more than 8%.

D. Retrieval Performance

As the relevance/similarity of the retrieved wallpapers is subject to individual perception, we include three examples in Fig. 5 to illustrate the retrieved results. From top to the bottom on the left column are the query images consisting of painting, pattern and fabric type, respectively. The top 3 retrieved wallpaper designs for each query are shown on the right. It can be seen that the proposed method is effective in retrieving visually similar wallpapers.

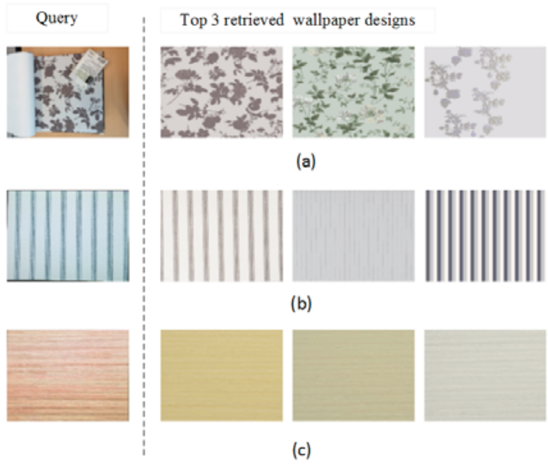


Fig. 5. Wallpaper retrieval results by the proposed hybrid method. Left column shows the query images. From top to bottom are query images of the painting, pattern and fabric type, respectively. Right column shows the corresponding retrieved results.

V. Conclusions

In this work we propose a hybrid feature-based wallpaper visual search system. It implements both the recognition and retrieval. This system is robust to scale and rotation variations. Besides the wallpapers, the developed visual search engine can also be deployed in other related domains, including floor tiles, curtains, etc.

References

[1] K.-H. Yap, T. Chen, Z. Li, and K. Wu, "A comparative study of mobile-based landmark recognition techniques," *IEEE Intelligent Systems*, vol. 25, no. 1, pp. 48–57, 2010.

[2] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[3] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, vol. 2, pp. 2161–2168.

[4] B. Girod, V. Chandrasekhar, D. M. Chen, N. M. Cheung, R. Grzeszczuk, Y. Reznik, G. Takacs, S. S. Tsai, and R. Vedantham, "Mobile visual search," *IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 61–76, 2011.

[5] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, 2003.

[6] D. G. Lowe, "Distinctive image features from scale-invariant key-points," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[7] Y. X. Liu, R. T. Collins, and Y. Tsin, "A computational model for periodic pattern perception based on frieze and wallpaper groups," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 3, pp. 354–371, 2004.

[8] Q. Yu, *Development of a perception oriented texture-based image retrieval system for wallpapers*, Thesis, 2008.

[9] J. Han and K. K. Ma, "Rotation-invariant and scale-invariant gabor features for texture image retrieval," *Image and Vision Computing*, vol. 25, no. 9, pp. 1474–1481, 2007.

[10] Z. W. Miao and X. D. Jiang, "Interest point detection using rank order LoG filter," *Pattern Recognition*, vol. 46, no. 11, pp. 2890–2901, November 2013.

[11] Z. W. Miao and X. D. Jiang, "A vote of confidence based interest point detector," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 1459–1463.