Feature Repetitiveness Similarity Metrics in Visual Search

Dipu Manandhar, *Student Member, IEEE*, and Kim-Hui Yap, *Senior Member, IEEE*

***Abstract*—Repetitive patterns are significant visual cues for matching and detecting objects in images. However, information from repetitive patterns is underutilized in computer vision algorithms as they cause burstiness issue in image similarity scoring. Existing similarity metrics do not take the repetitive patterns into account, and hence cannot handle images with repetitive patterns well. In view of this, this letter presents a new feature repetitiveness similarity (FRS) metric that not only addresses the burstiness issue, but also uses the information from the repetitive patterns to enhance the retrieval performance. The proposed FRS framework detects the repetitive patterns using descriptor and geometric information of local features in the images. Unique and repetitive features are handled separately and then fused at scoring stage using the FRS metric. Experiments conducted on the benchmark Oxford and Paris datasets show that the proposed method outperforms the state-of-the-art methods by a mean average precision of 6%. This demonstrates the effectiveness of FRS metric in matching and retrieval of images with repeated patterns.**

***Index Terms*—Image retrieval, repetitive pattern, similarity metrics, visual burstiness.**

I. Introduction

REPETITIVE patterns in an image carry significant information that can characterize particular object, place, or scene. However, in computer vision, they are often considered as a challenge, which confuses image matching and hinders visual recognition. Traditionally, repetitive patterns tend to diminish the performance of state-of-the-art image recognition and retrieval algorithms [1]–[4].

Last two decades have witnessed dramatic developments in visual search and recognition. A number of local feature detectors and descriptors such as scale-invariant feature transform (SIFT)[5] have been developed to extract robust features from images. Based on these local features, the bag-of-words (BoW)[6]–[8] model is widely used for image representation in various visual search systems and applications. The BoW model quantizes the local features to discrete visual words using a single codebook [3], [6], [8]–[10] or multiple codebooks [11]. The visual codebooks are constructed using clustering algorithms such as k-means or its variants, e.g., approximate k-means (AKM) [6] and hierarchical k-means[8]. Eventually, the images are represented as BoW vectors and similarities between them are calculated using distance metrics such as histogram intersection, cosine measure, and χ^2 distance.

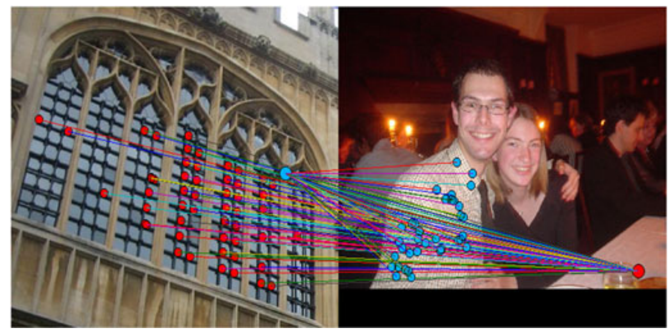


Fig. 1. Oxford query and irrelevant database image receive large similarity correspondences due to bursty false matches on repetitive patterns. Same color represent features quantized to the same visual word.

Most existing image similarity metrics do not take repetitive features into account during image matching. In fact, repetitive features violate the independent and identically distributed assumption required by the state-of-the-art algorithms in image matching and retrieval [2]–[4]. This leads to a problem popularly known as "visual burstiness," the phenomenon where certain visual elements tend to appear in bursts (see Fig. 1). All these similar features get quantized to the same visual word that overwhelm the unique ones. This distorts the image feature vector and produce biased similarity score ultimately leading to inferior performance in visual search.

For instance, let us consider conventional feature matching from the Oxford dataset in Fig. 1, a query image (left) and database image (right) form numerous matching correspondences producing a similarity score between these irrelevant objects. Clearly, a feature from the right image is matched to the repetitive features (shown in red) from the window pattern, and similarly a feature from the window is matched to bursty feature (shown in blue) from the checker pattern of the shirt. The cause of these false matchings and hence false retrieval is lack of consideration of repeated patterns. Hence, this letter presents a new feature repetitiveness similarity (FRS) metric to handle such problem of false bursty matching.

We develop our proposed FRS metric based on local features extracted from images. Although robust features learned from the state-of-the-art convolutional neural networks (CNNs) via deeplearning [12] such as GoogleNet [13], ResNet [14] are used for vision tasks such as image classification [13]–[16],

Manuscript received May 5, 2017; revised July 11, 2017; accepted July 20, 2017. Date of publication July 24, 2017; date of current version August 3, 2017. This work was carried out at the Rapid-Rich Object Search (ROSE) Lab at the Nanyang Technological University, Singapore. The ROSE Lab was supported by a grant from the Singapore National Research Foundation and administered by the Interactive & Digital Media Programme Office at the Media Development Authority. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Sumohana S. Channappayya. *(Corresponding author: Dipu Manandhar.)*

The authors are with the Rapid-Rich Object Search (ROSE) Lab and School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore 639798 (e-mail: dipu002@e.ntu.edu.sg; ekhyap@ntu.edu.sg).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2017.2731426

retrieval [17], [18], object detection [19], [20], saliency detection [21], [22], our survey shows that there are no prior works that deal with repetitive pattern detection via CNNs. It is because CNNs are mostly targeted at understanding global or higher semantic image concepts learned progressively from CNN layers. Since repetitive patterns detection is an early vision that requires low-level processing, we choose the handcrafted local features (SIFT) over CNN features. As compared to CNNs, an added advantage of the local features is that they do not require large training data, costly training, and huge storage for large CNN models.

In recent years, several techniques [3], [4], [23] have been proposed to handle the adverse impact of repetitive patterns. Most of the existing methods try to downweigh the contribution of repetitive patterns in image representation. Jegou *et al.* [3] first studied the burstiness phenomenon in images, where repetitive features vote multiple times for the same image. The authors proposed to use different downweighting functions including $\log(\cdot)$, $\sqrt{(\cdot)}$ on the term frequency of the visual words. Similarly, a generalized form of inverse document frequency called Lp-norm IDF [23] has been proposed to control the overwhelming effect of large component in BoW vector. The Lp-norm IDFs are negatively correlated to the term frequency and hence the large-term frequencies are penalized during visual word weighting. Recently, Torii *et al.* proposed "Repttile" method [4] to handle the visual burstiness problem. The main idea is to detect the presence of repetitive patterns and then use soft-assignment strategy adaptively. The authors also suggested to directly truncate the larger BoW histogram bins to reduce the impact of large terms in BoW vector. Although the above-mentioned methods are proven to improve retrieval performance, it is noted that the improvements were obtained by suppressing the false matches due to overcounting of repetitive patterns. Thus, information from repetitive patterns is either lost or partially discarded. Doubek *et al.* in [24] and Schindler *et al.* [25] considered to use the repetitive patterns for image matching and retrieval. However, their methods are not scalable and hence cannot be used in large scale search and retrieval.

In view of this, this letter proposes a new method to effectively represent the repetitive features and develop an FRS metric for image scoring. The method uses repetitive pattern detection algorithm to accurately detect the repetitive patterns in images. The repetitive and unique features are handled independently such that the false matches, shown in Fig. 1, are avoided. The studies in [26] and [27] also show the benefits of feature grouping and feature selection for better image matching. In this letter, we treat the features based on their repetitiveness. An advantage of the proposed FRS method is that it can handle the burstiness issue while utilizing the information in repetitive patterns to improve the search performance.

The rest of this letter is organized as follows. Section II describes the overview of the proposed method. Section III explains our proposed FRS framework in visual search. Section IV presents the experimental results and discussion, which is followed by conclusion in Section V.

II. Overview of the Proposed Method

The proposed FRS metric framework is shown in Fig. 2. It consists of the offline and online phases. During the offline phase, local features extracted from all database images are used for codebook generation. The FRS framework consists of two core components. First, repetitive patterns detection is performed based on the local features. Next, with the detection result, unique and repetitive features are independently quantized and separately indexed in inverted file, which stores the identity number of the database images where the visual words appear. During the query time, only visual words present in query image are checked and corresponding database images are scored efficiently.

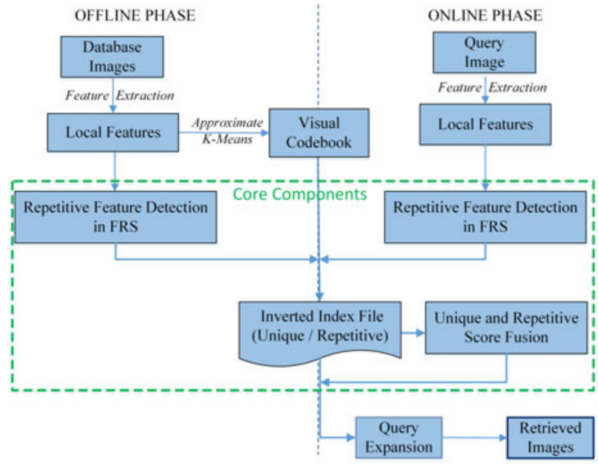


Fig. 2. Proposed framework of visual search with FRS metric.

In the online phase, local features extracted from the query image are passed through the same pipeline of repetitive pattern detection. The database images are then scored separately using the unique and repetitive features. The scores are then fused based on the proposed FRS metric, which takes feature repetitiveness into account. Finally, query expansion [28] is performed, and the system returns top retrieved images for the users.

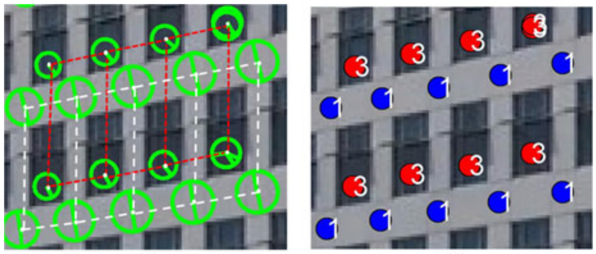
III. Proposed FRS Framework

The proposed framework consists of two core modules, namely repetitive pattern detection and repetitiveness similarity scoring, which are explained in the following sections.

A. Repetitive Pattern Detection

This section describes the repetitive pattern detection in images. It uses local geometry of the feature frames, $\mathcal{F}_i = (\mathbf{l}_i, s_i, o_i)$, i.e., location, scale and orientation, and descriptor $\mathbf{d}_i$ of the feature. We use SIFT as feature descriptor as it is robust against different imaging conditions and shown to achieve better image matching performance compared to other descriptors [29]–[31]. Although Torii *et al.* [4] has also performed coarse estimation of repetitive features, our method differs from theirs in three aspects. First, their method uses top $K(=50)$ common visual words to find feature similarity, which is prone to quantization error, whereas we directly operate on the unquantized descriptors. Second, our method makes use of orientation information, which provides more discriminative ability for correct detection of the patterns. Last but most importantly, FRS uses lattice information to recover the repeated features that are otherwise undetected by the feature detector. This will help to improve retrieval performance.

We formulate the problem of repetitive pattern detection using undirected graph $\mathcal{G}_\zeta$. The detection method is summarized in Algorithm 1. The repetitive patterns originating from similar visual elements have similar geometric properties such as similar scale, orientation, with small descriptor distance as



(a) (b)

Fig. 3. Repetitive pattern detection. (a) Sample detected local features, shown as graph with linkage established between the similar repeated features. (b) Different clusters formed by detected repetitive patterns.

(a) (b)

Fig. 4. Repetitive feature detection in FRS. Different colors and labels represent different types of features. Recovered keypoints are shown in green square. Best viewed with $2\times$ zoom.

Algorithm 1: Repetitive Pattern Detection and Clustering.

```
Input: N_p feature frames, F = (l_i, s_i, o_i) and descriptor d_i
Output: Clusters C, each representing a type of pattern
Initialize Graph G_ζ with                          ▷ Disconnected Graph
        N_p vertices i.e. features
        E: Edges, e_ij ∈ E := ∅
Normalize descriptors d_i := d_i / ||d_i||          ▷ L_2 Normalization
for all features F do
    Calculate constraint functions on ζ:
        Φ_ij = d_i · d_j;  θ_ij = o_i − o_j
        ρ_ij = s_i / s_j;  Δ_ij = ||l_i − l_j||_2
end for
for i, j = 1 : N_p do
    if ( all constraints on ζ are satisfied ) then
        e_ij = e_ji := TRUE                           ▷ Create Edge
    end if
end for
Compute the connected components C of G_ζ
                                                  ▷ Clusters of Features
```

shown in Fig. 3(a). Features (nodes in graph) are considered similar and edges e_{ij} (linkages shown in dotted lines) are created between them if they satisfy a set of similarity constraints (ζ) defined as follows.

1) Cosine similarity of descriptors $(\Phi) \geq 0.9$.
2) Difference angle in orientations $(|\theta|) \leq 0.5\,\text{rad}$.
3) Relative ratio of scales $(1/1.3) \leq \rho \leq 1.3$.
4) Spatial proximity between features $(\Delta) \leq 10\,\text{s}$ (scale).

The graph is built when all the linkage between the similar features have been established. Next, the connected components of the graph are collected and the repeated features (in Fig. 3(b), labeled as 1 and 3) are grouped into different clusters, defined by $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_{Nc}\}$. Further, to estimate the underlying lattice structure for each detected repeated features, Hough transform is used to estimate the existence of specific patterns. The undetected features in the patterns are also recovered. Fig. 4 shows images with final output detected repetitive patterns detection in different colors and labels.

B. Repetitiveness Similarity Scoring

The detected repetitive patterns contain useful information for image matching. Specifically, the cardinality of cluster is a good indication of feature repetitiveness. A cluster with large number of features refers to highly repeated structures, whereas singleton cluster or cluster with very few members mean their features are unique in appearance.

The visual information of an object can be modeled as combination of its various visual attributes [32], [33]. To identify an image object, unique features may sufficiently characterize it. However, if the image has a large number of repeated structures, they become an important information that should be used to complement the unique features. Under this situation, the existing method downweighs or discards repetitive features leading to information loss. To overcome this, the proposed FRS metric handles repetitive features and nonrepetitive ones separately, i.e., we quantize and match these features independently such that bursty false matches between unique and repetitive features are completely avoided.

Formally, let I be an image with detected repetitive feature clusters $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_{Nc}\}$, where Nc is the total number of cluster detected. Let $\mathrm{n}(\mathbf{c}_j)$ denote the cardinality of the cluster $\mathbf{c}_j$. Each cluster $\mathbf{c}_j$ is identified as unique ($\mathcal{C}_{\mathcal{U}}$) or repetitive ($\mathcal{C}_{\mathcal{R}}$) based on the rule

$$\mathbf{c}_j \in \begin{cases} \mathcal{C}_{\mathcal{U}} & \text{if } \mathrm{n}(c_j) < T \\ \mathcal{C}_{\mathcal{R}} & \text{if } \mathrm{n}(c_j) \geq T \end{cases} \tag{1}$$

where T is a positive threshold empirically set to 5. All features $F_{\mathcal{U}} = \{f : f \in \mathcal{C}_{\mathcal{U}}\}$ are nonrepetitive, and hence are called unique features, whereas features $F_{\mathcal{R}} = \{f : f \in \mathcal{C}_{\mathcal{R}}\}$ are called repetitive features. These are separately quantized applying function ($\mathscr{Q}$) in (2) using a codebook with visual words: $\{v_1, v_2, \ldots v_i \ldots v_z\}$. Therefore, two BoW representations: unique $\mathbf{I}_{\mathcal{U}}$ and $\mathbf{I}_{\mathcal{R}}$ are obtained for each image.

Next, the score between the query image (Q) and database image (I) is computed as follows. Let w_i and q_i be the ith visual word occurrence for Q and I, respectively, and $\mathrm{idf}(i)$ be associated inverse document frequency. The score between query and database image is computed using similarity score (S) in (3) for both $\mathbf{I}_{\mathcal{U}}$ and $\mathbf{I}_{\mathcal{R}}$ representations

$$\mathscr{Q}(f) = \arg\min_i \|f - v_i\|_2^2 \tag{2}$$

$$S_k(\mathbf{I}, \mathbf{Q}) = \frac{\sum_{i=1}^{z} w_i \cdot q_i \cdot \mathrm{idf}(i)^2}{\|\mathbf{I}_k\|\,\|\mathbf{Q}_k\|}, \quad k \in \{\mathcal{U}, \mathcal{R}\} \tag{3}$$

$$\text{where,} \quad \|\mathbf{I}\| = (\Sigma_i\, w_i^2)^{\frac{1}{2}} \;\; \text{and} \;\; \|\mathbf{Q}\| = (\Sigma_i\, q_i^2)^{\frac{1}{2}}.$$

Let $S_{\mathcal{U}}$ and $S_{\mathcal{R}}$ be the scores for the unique and repetitive representations. The final similarity score is computed using (4), which takes the unique as well as repetitive features, and the

TABLE I
mAP Comparision of Different Methods

Method	Oxford Dataset	Paris Dataset
TF-IDF [6]	63.7	69.7
Burstiness weighting [3]	68.2	71.2
Repttile method [4]	71.2	74.2
Neural codes [36]	54.5	–
Local CNN features [37]	64.9	69.4
NetVLAD [38]	70.8	78.3
Proposed FRS	**77.3**	**81.1**

degree of repetitiveness α into account

$$\Omega = \beta S_{\mathcal{U}} + \lambda \cdot \alpha S_{\mathcal{R}} \tag{4}$$

$$\text{where, } \alpha = \frac{\mathrm{n}(\mathcal{C}_{\mathcal{R}})}{\mathrm{n}(\mathcal{C}_{\mathcal{R}}) + \mathrm{n}(\mathcal{C}_{\mathcal{U}})}, \text{ and } \beta = \frac{\mathrm{n}(\mathcal{C}_{\mathcal{U}})}{\mathrm{n}(\mathcal{C}_{\mathcal{R}}) + \mathrm{n}(\mathcal{C}_{\mathcal{U}})}. \tag{5}$$

Here, $\beta = 1 - \alpha$ is a measure of feature uniqueness. λ is a regularization parameter experimentally set to 0.9 to calibrate the scores $S_{\mathcal{U}}$ and $S_{\mathcal{R}}$. The joint score Ω in (4) fuses the information of unique and repetitive features while alleviating the problem of bursty false matching.

Note that the degree of repetitiveness α and hence β in (5) are not design parameters, but rather their values are determined based on the outcome of repetitive pattern detection. In the absence of repetitive patterns, the repetitive term (α) becomes zero and hence all features will be treated equally.

IV. Experiments

The proposed FRS framework is evaluated for image retrieval using two standard benchmark datasets: Oxford dataset and Paris dataset, which contain 5062 and 6412 images, respectively. Both contain 55 query images of distinct Oxford and Paris landmarks. For performance evaluation, average precision (AP) is used to calculate the area under precision–recall curve. The APs for all the query images are averaged to give mean average precision (mAP) for the particular dataset.

Local features are extracted using difference of Gaussians with affine estimation, which are described by 128D SIFT descriptor [5], and further converted to *Root*SIFT [34]. Visual coodebook with 1M visual words is learnt from respective database using AKM based on FLANN library [35]. Repetitive as well as unique features from database images are indexed in inverted file systems. We use squared idf in visual word weighting.

Table I presents mAPs obtained for the proposed FRS framework and comparison with other state-of-the-art methods. For the Oxford dataset, the proposed FRS framework achieves a mAP of 77%, which outperforms Philbin's TF-IDF [6] method by 13.6%, burstiness weighting [3] by 9.1%, and Torii's Repttile method [4] by 6.1%. Similarly, for the Paris dataset, the proposed method achieves a mAP of 81% with surpasses Philbin's TF-IDF by 11% [6], the burstiness weighting [3] by 9.9%, and Torii's Repttile method [4] by 6.9%.

We also compare our method with deep learning based approaches [36]–[38], where global features extracted from CNN networks are used for image retrieval. As shown in Table I, our method outperforms deep learning methods. This demonstrates the importance of repetitive pattern detection for image matching, which is not considered in current deep learning based approaches [36]–[38].

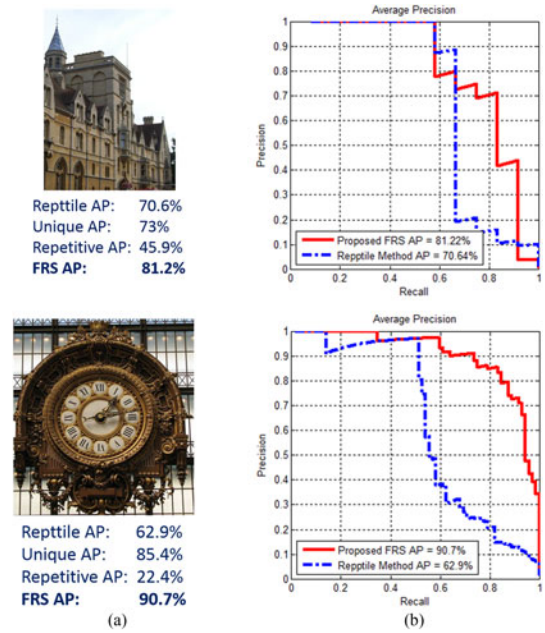


Fig. 5. Example queries images from the Oxford (top row) and Paris (bottom row) datasets with AP. (a) Query images. (b) Precision–recall curve and AP.

It takes 1.8 s to query an image with approximately 1500 detected features using the following settings: MATLAB implementation and VLFeat library, a computer with i5-4570 CPU and 32-GB memory. The computational time can be further reduced with implementation using compiling language such as C with code optimization.

Fig. 5 shows retrieval performances for sample query images using the proposed FRS metric and the Repttile method. The top row shows an Oxford query and its precision–recall curve where our FRS method covers more area under the curve than the Repttile method. For this query, the Repttile method [4] achieves an AP of 70.6%, whereas only using the unique features achieves an AP of 73%. The FRS metric uses information of repetitive patterns and complements it with unique features to increase the AP to 81.2%. Hence, this further verifies that the proposed method utilizes the repetitive patterns to enhance the retrieval performance.

Overall, the existing state-of-the-art methods either do not consider or downweigh/limit the contribution of repetitive features, which results in visual information loss. In contrast, the proposed FRS framework detects the repetitive patterns and offers an effective representation via proposed FRS metric and hence achieves better retrieval performance. This clearly demonstrates the effectiveness of the proposed method.

V. Conclusion

This letter proposes a new FRS metric in visual search. The proposed method makes effective use of the repetitive patterns in the images to improve the image matching and search performance. A new algorithm for accurate detection of repetitive pattern in images is introduced. Based on the detection, the repetitive features and unique features are handled separately, which help in addressing the burstiness issue in visual search. The repetitive patterns are not nuisance but can be used to complement the unique features in visual recognition. The experimental results clearly show the effectiveness of the proposed method.

References

[1] J. Knopp, J. Sivic, and T. Pajdla, "Avoiding confusing features in place recognition," in *European Conference on Computer Vision*. New York, NY, USA: Springer, 2010, pp. 748–761.

[2] R. G. Cinbis, J. Verbeek, and C. Schmid, "Image categorization using Fisher kernels of non-iid image models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2184–2191.

[3] H. Jegou, M. Douze, and C. Schmid, "On the burstiness of visual elements," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1169–1176.

[4] A. Torii, J. Sivic, M. Okutomi, and T. Pajdla, "Visual place recognition with repetitive structures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 11, pp. 2346–2359, Nov. 2015.

[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[6] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[7] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1470–1477.

[8] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2006, vol. 2, pp. 2161–2168.

[9] H. Jegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 304–317.

[10] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Lost in quantization: Improving particular object retrieval in large scale image databases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[11] Z. Wu, Q. Ke, J. Sun, and H.-Y. Shum, "A multi-sample, multi-tree approach to bag-of-words image representation for image retrieval," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 1992–1999.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[13] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[17] A. Salvador, X. Giro-i Nieto, F. Marques, and S. Satoh, "Faster R-CNN features for instance search," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 394–401.

[18] G. Tolias, R. Sicre, and H. Jégou, "Particular object retrieval with integral max-pooling of CNN activations," arXiv:1511.05879, 2015.

[19] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[20] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.

[21] J. Kuen, Z. Wang, and G. Wang, "Recurrent attentional networks for saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3668–3677.

[22] G. Li and Y. Yu, "Visual saliency based on multiscale deep features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5455–5463.

[23] L. Zheng, S. Wang, and Q. Tian, "Lp-norm IDF for scalable image retrieval," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3604–3617, Aug. 2014.

[24] P. Doubek, J. Matas, M. Perdoch, and O. Chum, "Image matching and retrieval by repetitive patterns," in *Proc. Int. Conf. Pattern Recognit.*, 2010, pp. 3195–3198.

[25] G. Schindler, P. Krishnamurthy, R. Lublinerman, Y. Liu, and F. Dellaert, "Detecting and matching repeated patterns for automatic geo-tagging in urban environments," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–7.

[26] P. Turcot and D. G. Lowe, "Better matching with fewer features: The selection of useful features in large database recognition problems," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2009, pp. 2109–2116.

[27] X. Liu, B. Lang, Y. Xu, and B. Cheng, "Feature grouping and local soft match for mobile visual search," *Pattern Recognit. Lett.*, vol. 33, no. 3, pp. 239–246, 2012.

[28] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman, "Total recall: Automatic query expansion with a generative feature model for object retrieval," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–8.

[29] L. Juan and O. Gwun, "A comparison of SIFT, PCA-SIFT and SURF," *Int. J. Image Process.*, vol. 3, no. 4, pp. 143–152, 2009.

[30] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[31] N. Y. Khan, B. McCane, and G. Wyvill, "Sift and surf performance evaluation against various image deformations on benchmark dataset," in *Proc. 2011 Int. Conf. Digit. Image Comput. Techn. Appl.*, 2011, pp. 501–506.

[32] A. Tversky, "Features of similarity," *Psychol. Rev.*, vol. 84, no. 4, pp. 327–352, 1977.

[33] R. Pramod and S. Arun, "Object attributes combine additively in visual search," *J. Vis.*, vol. 16, no. 5, pp. 8–8, 2016.

[34] R. Arandjelovi and A. Zisserman, "Three things everyone should know to improve object retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2911–2918.

[35] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *Proc. VISAPP Int. Conf. Comput. Vis. Theory Appl.*, 2009, vol. 2, pp. 331–340.

[36] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, "Neural codes for image retrieval," in *European Conference on Computer Vision*. New York, NY, USA: Springer, 2014, pp. 584–599.

[37] J. Yue-Hei Ng, F. Yang, and L. S. Davis, "Exploiting local features from deep networks for image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 53–61.

[38] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN architecture for weakly supervised place recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 5297–5307.