International Journal of Future Innovative Science and Engineering Research (IJFISER)
Volume - 2, Issue - II ISSN (Online): 2454- 1966

Research Manuscript Title

ENHANCED BOV FOR LARGE SCALE IMAGE SEARCH

S.Tamilselvan[1], M.Dhivya[2], N.Suganthi[3],

[1]Assistant professor,[2]PG scholar, [3]Senior lecturer,
[(1&2)]Department of CSE, [3]Department of ECE,
[(1&2)]AMACE, [3]Adhiparasakthi polytechnic college,

E-Mail: tmlslvn@gmail.com, dhivyame.amace@gmail.com, sugant1984@gmail.com

JUNE – 2016

www.istpublications.com

International Journal of Future Innovative Science and Engineering Research (IJFISER), Volume - 2, Issue – II,
ISSN (Online): 2454- 1966 www.istpublications.com.

ENHANCED BOV FOR LARGE SCALE IMAGE SEARCH

S.Tamilselvan[1], M.Dhivya[2], N.Suganthi[3],

[1]Assistant professor,[2]PG scholar, [3]Senior lecturer,
[(1&2)]Department of CSE, [3]Department of ECE,
[(1&2)]AMACE, [3]Adhiparasakthi polytechnic college,

E-Mail: tmlslvn@gmail.com, dhivyame.amace@gmail.com, sugant1984@gmail.com

ABSTRACT

The large scale image retrieval has attracted increasing attention in recent years, due to the large user demand and importance in underlying many useful applications such as location recognition, image compression and image reconstruction. Bag of visual Words (BovW) model is an efficient image representation popularly used in large scale image data search. Initially, this representation has been used with text search and now-a-days popularly being employed in image searches for efficient image representation. Supplementing this image representation is an image descriptor which tries to discriminate the features powerfully for better matching. SIFT and SURF is powerful image descriptors widely used in large scale image indexing and retrieval. In this research, we propose an "Enhanced BovW Model for Large Scale Image data search" which would enhance the image representation, indexing and retrieval efficiency of large scale image data search. Here, an enhanced BovW model is proposed to address the images with good spatial and contextual information among their visual phrases. Apart from the efficient representation, a suitable descriptor by extending SIFT has been devised for better discrimination of features. A distributed normalized histogram of angles of pairs of identical visual word (PIW) is computed histograms produced by each word type constitute a powerful description of intra type visual words relationship.

Index Term-Image Search, Bovw, Spatial and Contextual Modeling, SIFT.

I. INTRODUCTION

This paper concentrates on the errand of extensive scale fractional copy picture recovery. Bag of-visual words (BoWs) picture representation has been used for some interactive media and vision issues, including video occasion location, object acknowledgment picture division and vast scale picture recovery. Speaking to a picture as a visual report made out of repeatable and particular visual components that are list capable is extremely attractive. With such a representation, numerous developed strategies in Data recovery can be utilized for vision undertakings, for example, visual hunt or acknowledgment. As of late, it has been shown that BoWs picture representation is a standout amongst the most encouraging methodologies for recovery undertakings in vast scale image and video databases. In any case, exploratory

Consequences of reported works demonstrate that the generally produced visual words are still not as expressive as the content words. Generally, the exemplary visual vocabulary is made by bunching countless component descriptors. The model descriptor of every bunch is known as a visual word, which is then recorded by a whole number. In past works different quantities of visual words are created for various assignments. There are two general perceptions: 1) utilizing more visual words results as a part of better execution and 2) be that as it may, the execution will be immersed when the quantity of visual words achieves certain levels [1]. Instinctively, a bigger number of visual words demonstrate all the more fine grained apportioning of the descriptor space. Subsequently, the visual words turn out to be more discriminative in speaking to certain visual substance. The second perception is that expanding the quantity of visual words to specific levels at last soaks the execution of vision vocabulary. Naturally, isolating the element space in better scales builds the quantization mistake in visual vocabulary. This implies neighborhood highlights close in the component space may be quantized into various visual words. These perceptions unequivocally suggest the constrained expressive capacity of the great visual word. A toy example illustrating this finding is presented in Fig.1.In the figure, Filter descriptors are separated on interest focuses identified by

Difference of Gaussian (DoG) [2]. The related pictures are spoken to as BoWs with a visual vocabulary containing 32 357 visual words, by supplanting their SIFT descriptors with the lists of the nearest visual word. One key issue of the BoW model includes visual word coordinating between images. Exact component coordinating prompts high image recovery execution. In any case, two disadvantages bargain this strategy. To start with, in quantization, a 128-D twofold SIFT highlight is quantized to a solitary number. Despite the fact that it empowers effective online recovery, the discriminative force of SIFT highlight is to a great extent lost. Highlights that lie far from each other might really fall into the same cell, along these lines delivering false positive matches. Second, the best in class frameworks depend on the SIFT descriptor, which just depicts the neighborhood slope dispersion, with uncommon portrayal of different attributes, for example, shading, of this nearby district. Subsequently, districts which are comparative in composition space however distinctive in shading space might likewise be considered as a genuine match. Both disadvantages lead to false positive matches and debilitate the picture recovery exactness.

Fig.1.Matched visual words between the same and different objects.

In this manner, it is undesirable to take visual word list as the main ticket to visual coordinating. Rather, the coordinating strategy ought to be further checked by different prompts, which ought to be productive regarding both memory and time. In spite of the fact that BoVW model is famous it has a few issues like:

1. Visual words are less enlightening force because of the effect of quantization and picture altering operations.
2. Baseline strategy for BovW needs spatial and relevant data among the visual words.
3. Relationship among the nearby fixes is not considered in the gauge technique.
4. The extraordinary format of the visual words is completely ignored, which will bring about equivocalness amid coordinating.
5. Many BovW techniques need confirmation component which falls apart recovery execution.

With a specific end goal to give an answer for every one of these issues, numerous explores have been done on BovW model and still it is under procedure. Here, certain related works of BovW models has been examined in the resulting segment with the plan of exploration targets. A generally utilized answer for rate up SIFT coordinating is to create Bag-of-visual Words (BoWs) model. The BoWs model quantizes SIFTs into minimized visual words, thus takes into account quick coordinating. With data recovery approaches like altered record files, BoWs model is likewise appropriate to expansive scale picture look errand.

II.RELATED WORK

In vast scale content based image seek application, Bag-of visual words (BovW) model on neighborhood highlights has been broadly embraced. For the most part, in those BovW-based methodologies, there are four key segments: neighborhood highlight representation, highlight quantization, record procedure, recovery scoring. In this area, we make an audit of related work in every segment.

A.Local Feature Representation

Nearby element Representation is usually figured from picture patches by either compressing the pixel slopes from various sub-districts into a high dimensional histogram such as SIFT [3] and SURF [4], mapping the high dimensional descriptor into Hamming space like LDA Hash , keeping the spatial pieces of information of edges like Edge-SIFT , or looking at the intensities of pixels at various areas to frame a progression of paired code like ORB and BRIEF .The 128-dimensional SIFT is a standout amongst the most mainstream picture neighborhood descriptors in PC vision. Filter is removed from image keypoints distinguished with Difference-of-Gaussian DoG) identifier, which figures the scales, introductions, and is ordinarily processed from image patches by either abridging the pixel slopes from various sub-areas into a high dimensional histogram such as SIFT and SURF , mapping the high dimensional descriptor into Hamming space like LDA Hash, keeping the spatial hints of edges like Edge-SIFT , or looking at the intensities of pixels at various areas to shape a progression of parallel code like ORB and BRIEF. The 128-dimensional SIFT is a standout amongst the most well-known image nearby descriptors in PC vision. Filter is extricated from image keypoints distinguished with Difference-of-Gaussian (DoG) identifier, which figures the scales, introductions, and areas of keypoints. With the scale and introduction pieces of information, picture patches encompassing keypoints could be standardized into altered introduction and size; henceforth scale and introduction invariant nearby descriptors could be separated subsequently. In light of SIFT, some other comparative descriptors like SURF, PCA-SIFT, Gradient Location and Orientation Histogram (GLOH) are proposed. A few scientists propose low-bitrate descriptors, for example, BRIEF, BRISK, CHoG, ORB, LDA Hash, and Edge-SIFT which are quick both to manufacture and match. BRIEF descriptor is proposed. Every piece of BRIEF is figured by considering indications of basic power contrast tests between sets of focuses inspected from the picture patches. In spite of the unmistakable point of preference in rate, BRIEF endures as far as unwavering quality and power as it has constrained resilience to picture bends and changes. The BRISK descriptor first proficiently identifies keypoints in the scale-space pyramid in view of an indicator propelled by FAST and AGAST. Then given an arrangement of recognized keypoints, BRISK descriptor is made as a paired concatenating so as to string the consequences of straightforward shine examination tests. The embraced identifier of BRISK gets area, scale, and introduction pieces of information for each keypoints. Consequently BRISK accomplishes introduction invariance and scale-invariance. Sphere descriptor is assembled taking into account BRIEF yet is separated with a novel corner point identifier, consequently is additionally strong to the pivot. The creators exhibit that ORB is altogether quicker than SIFT, while executes also by and large. Contrasted and SIFT, in spite of the unmistakable favorable position in velocity, these smaller descriptors show impediments in the parts of spellbinding force, vigor or all inclusive statement.

Algorithm1: for local feature Representation
Input *:number of images n*
    local of identify the feature between the images
    l(i,j)   i,j=1…..n
    start with image n
output :vector of images and identify the features.
Initialization :
    l ← 0
    local ← o
    identify the feature ← o
    e=1
for $1 \le r \le n$
do {
    choose pointer with local=l(e,i)=host{l(e,k);identify feature (k)=0 and $1 \le k \le n$}
        identify feature ← identify feature + local feature –identify feature
        e=j
        c (r) ← j
        c(n)=1
        local=local+(e,1).

B. **Highlight Quantization**

Highlight Quantization Usually, hundreds or a huge number of nearby components can be extricated from a solitary picture. To accomplish a reduced representation, high dimensional neighborhood elements are quantized to visual words, and a picture can be spoken to as a "pack" of visual words [5]. In this manner, a visual codebook should be produced previously. The most natural visual codebook era strategy is k-implies or various leveled k-implies for vast size visual codebook era. With visual codebook characterized, highlight quantization is to relegate a visual word ID to every element. The most innocent decision is to locate the nearest (the most comparative) visual expression of a given element by straight sweep, which, in any case, endures costly computational expense. Normally, surmised closest neighbor (ANN) look strategies are embraced to accelerate the seeking process, with penance of precision to some degree. A k-d tree structure is used with a best in-first change to discover inexact closest neighbors to the descriptor vector of the question. In light of the various leveled vocabulary tree, a proficient estimated closest neighbor hunt is accomplished by spreading the inquiry highlight vector from the root hub down the tree by looking at the comparing tyke hubs and picking the nearest one. A k-d timberland estimation count is proposed with diminished time flightiness. To decrease the quantization hardship, a descriptor-subordinate sensitive undertaking arrangement is proposed to portray part vector to a weighted blend of various visual words. The high dimensional SIFT descriptor space is distributed standard cross areas for the endeavor of picture portrayal with promising execution.

Algorithm2: Build Tree, For Building the Extended K-D Tree
***Input**: Multiset Of Points S From R^d*
***Output**: K-D Tree T for S.*
***Function** Build tree(S):*
Create Root Node R with Points S and $LBR_R=R^d$
split node(R)
Return Tree T with Root R
***Function** split node (Tree Node Q with Points P and LBR_Q):*
Compute TBB_Q As Follows:
$L_i^q = Min\{ X_i ; X \in P\}, H_i^q = Max(X_i ; X \in P\}$
If** Number Of Points $||P||>L$ **Then
$M \leftarrow Arg\ Max\ (X_i^q - X_i^L)$
$\xi \leftarrow Median\ \{X_m ; X \in P\}$
$P_L \leftarrow \{ X_m < \xi ; X \in P\}$
$P_R \leftarrow \{ X_m \geq \xi ; X \in P\}$
$LBB_L \leftarrow \{ X_m < \xi ; X \in LBB_Q\}$
$LBB_R \leftarrow \{ X_m \geq \xi ; X \in LBB_Q\}$
If** Neither P_L nor P_R Is Empty **Then
Create Left Child L of Q with Points P_L and LBB_L
Create Right Child R of Q with Points P_R and LBB_R
Split node (L)
Split node(R)

C. Document Strategy

Document Strategy Inspired by the achievement of substance web crawlers, switched record structure has been adequately used for tremendous scale image look [6]. Essentially, changed report structure is a minimized representation of a small cross section, where the line and the portion mean visual word and picture, independently. In on-line recuperation, simply the photos granting general visual words to the inquiry picture ought to be checked. Along these lines, the quantity of hopeful image to be looked at is significantly lessened, accomplishing a proficient reaction. In the modified record structure, each visual word is trailed by a transformed document rundown of passages. Every passage stores the ID of the picture where the visual word shows up, and some different signs for check or similitude estimation. Case in point, Bundled Feature stores the x-request and y-request of each SIFT highlight situated in the packaged range. The geometric pieces of information, for example, highlight position, scale, and introduction, are additionally put away in the rearranged document list for geometric consistency confirmation [7] . To advance diminish the memory expense of modified record structure, a visual word vector is mapped to a low-dimensional representation by a gathering of min-hash capacities[8].Therefore, just little steady measures of information per image should be put away.

D. Recovery Scoring

Recovery scoring in the online recovery stage, in the wake of looking so as to recognize those pertinent pictures of a question up the list table, it is important to decide the applicable score of those objective images to the inquiry image. Ordinarily, the importance score is characterized by the standardized separation between the BoW vectors of the question and the database images. At the point when the codebook size is much bigger than the nearby element sum in pictures, the image vector by BoW is exceptionally inadequate and we just need to check those visual words showing up in

both pictures, which brings about extremely proficient in execution. To recognize the noteworthiness of visual words in various images, term recurrence (TF) and transformed report/imagee recurrence (IDF) are broadly connected in numerous current cutting edge calculations [6] .The exemplary TF-IDF is further improved with a matching so as to weight term packaged capabilities. The importance score is just characterized by tallying what number of sets of nearby component is matches crosswise over two pictures. Logical weighting is acquainted with enhance the great vocabulary tree approach. Insights of neighboring descriptors both on the vocabulary tree and in picture spatial space are effectively joined.

III.OUR APPROACH

We talk about our methodology as takes after. In area III-A we present our inspiration. In segment III-B we talk about our Bag of visual Word model. In segment III-C We clarify how list extensive scale picture information look taking into account SIFT descriptor.

A. Motivation

Fundamentally, the key issue of image pursuit is visual coordinating between images. At the point when pictures are spoken to by nearby elements, visual coordinating is accomplished by means of highlight coordinating between images. Instinctively, considering whether two elements from various images are a couple of substantial match, the most clear basis is to check whether the separation between them is littler than a predefined edge .In conventional Bag-of-Visual-Words based methodology, highlight coordinating is verifiably acknowledged by checking whether two components are quantized to the same visual word[9]. In any case, we as often as possible watch that, still numerous elements with substantial separation from each other are quantized to the same visual word, while numerous different components with little separation from each other are quantized to various visual words. Such wonder effortlessly causes the false positive and genuine negative. To dodge such downside, it is more desirable over check highlight coordinating by highlight separation. In addition, since ongoing reaction is a basic necessity in expansive scale picture seeks, the coordinating check ought to be performed effectively.

B. Bag of Visual Word Model

Bag of visual words (BoVW) is a mainstream system for picture arrangement propelled by models utilized as a part of normal dialect preparing. BoVW makes light of word course of action (spatial data in the image) and orders in light of a histogram of the recurrence of visual words. The arrangement of visual words shapes a visual vocabulary, which is built by bunching an extensive corpus of components. To get a visual vocabulary, a substantial arrangement of neighborhood elements separated from a preparation picture corpus is bunched. Along these lines the nearby element space is separated into useful locales (the visual words) and the gathering of the acquired visual words is the visual vocabulary K-means is the most regularly utilized bunching calculation. The bunching of the components is performed with a Gaussian blend model (GMM) as appeared in fig.2.

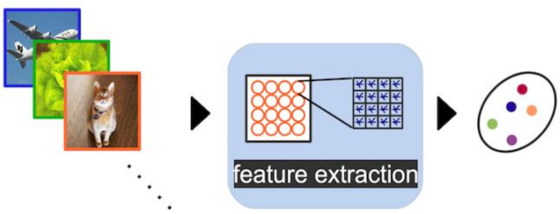

Fig.2. Visual Vocabulary Construction

The common BoVW pipeline for speaking to a picture is made out of the accompanying steps: (1) removing the nearby elements from the image, (2) encoding the neighborhood elements to the comparing visual words, (3) performing spatial binning.as appeared in fig.3.

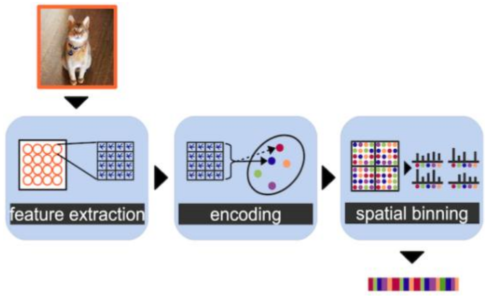

Fig.3.Image representation

In this stride neighborhood components are removed from the image. Nearby components are intended to discover neighborhood picture structures in a repeatable manner and to speak to them in hearty ways that are invariant to ordinary picture changes, for example, interpretation, turn, scaling, and relative disfigurement. Nearby elements constitute the premise of methodologies created to consequently perceive particular protests .The most prominent neighborhood highlight extraction technique is the Scale Invariant Feature Transform (SIFT), VSEM utilizes the VL Feat usage of SIFT as appeared in fig.4.

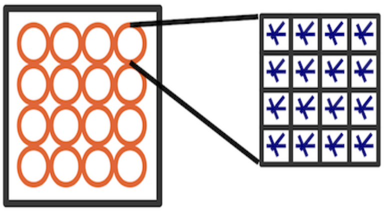

Fig.4.Feature extraction

The encoding step maps the nearby elements removed from a image to the comparing visual expressions of the already made vocabulary. The most well-known encoding system is called hard quantization (VQ), which allots every component to the closest visual word's centroid (in Euclidean separation). As of late, more powerful encoding techniques have been presented, among which the Fisher encoding has been appeared to beat all the others. VSEM utilizes both the VQ and the Fisher encoding.

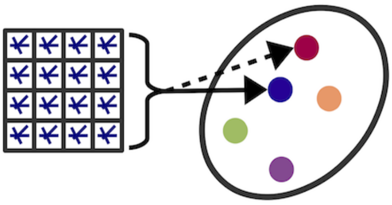

Fig.5.Encoding fig representation

A solidified method for presenting spatial data in BoVW is the utilization of spatial histograms. The primary thought is to isolate the image into a few (spatial) districts, register the encoding for every area and stack the subsequent histograms. This method is alluded to as spatial binning and it is executed in VSEM

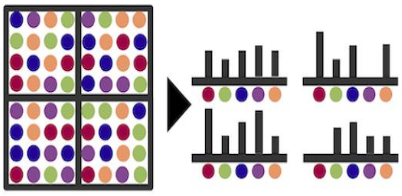

Fig .6. Spatial fig representation

In image look, the issue of highlight coordinating between images can be viewed as discovering elements closest or around closest neighbors inside of a specific reach. At the point when the element sum turns out to be substantial, say, more than one billion, it is too computationally costly to locate the closest neighbors by directly contrasting all components parallel vectors. To address this issue, modified document structure, utilized from content recovery, can be utilized for versatile indexing of extensive scale picture information set. The great execution of SIFT and SURF contrasted with different descriptors is noteworthy. It's blending of roughly limited data and the circulation of slope related elements appears to yield great unmistakable force while fighting off the impacts of restriction mistakes as far as scale or space. Utilizing relative qualities and introductions of angles lessens the impact of photometric changes. The proposed SIFT and SURF descriptor depends on comparative properties, with a many-sided quality stripped down significantly further. The initial step comprises of altering a reproducible introduction taking into account data from a roundabout area around the interest point. At that point, we build a square area adjusted to the chose introduction, and concentrate the SIFT and SURF descriptor from it. These two stages are presently clarified thus. Besides, we likewise propose an upright rendition of our descriptor that is not invariant to picture revolution and in this manner speedier to process and more qualified for applications where the camera stays pretty much level. Bag of SIFT representation and direct SVM classifier (precision of around 60-70%).You will begin by executing the modest picture representation and the closest neighbor classifier. They are straightforward, simple to execute, and run rapidly for our trial setup (under 10 seconds). Presently we are prepared to speak to our preparation and testing pictures as histograms of visual words. For every picture we will thickly test numerous SIFT descriptors. Rather than putting away several SIFT descriptors, we just tally what number of SIFT descriptors fall into every bunch in our visual word vocabulary. This is finished by finding the closest neighbor k implies centroid for each SIFT highlight. In this manner, on the off chance that we have a vocabulary of 50 visual words, and we identify 220 SIFT highlights in a image, our pack of SIFT representation will be a histogram of 50 measurements where every canister tallies how often a SIFT descriptor was relegated to that bunch and totals to 220. The histogram ought to be standardized with the goal that image size does not drastically change the Bag of highlight size.

IV.PERFORMANCE AND EVALUTION

To assess the adequacy of our technique, we directed investigations on any open benchmark data sets: the Ukbench [10] , the occasions[11] ,DupImage [12] and the MIR Flickr 1M [13] datasets.

A .Datasets

1. Ukbench: The Ukbench dataset comprises of 10200 pictures of 2550 gatherings. Every gathering contains four pictures of the same article or scene, taken from various perspectives. Each of the 10200 pictures is taken as inquiry picture. The execution is measured by the review for the main 4 competitors, eluded to as N-S score (most extreme 4).

2. Occasions: The Holidays dataset is made out of 500 inquiries from 1491 clarified individual occasion photographs. mAP (mean Average Precision) is utilized to assess the execution.

3. DupImage: The DupImage dataset contains 1104 pictures from 33 commented on gatherings. From this ground truth dataset, 108 delegate questions are chosen, and guide is utilized to assess picture recovery execution.

4. MIR Flickr 1M: This dataset contains 1 million pictures recovered from Flickr. We add this dataset to the Holidays, Ukbench, and DupImage datasets to test the adaptability of the proposed strategy.

B. Assessment

Strategies for indexing and proficient recovery with content reports are full grown, and sufficiently powerful to work with millions or billions of records without a moment's delay. Archives of content contain some circulation of words, and in this manner can be minimalistically abridged by their statement checks (known as a Bag of-word). Since the event of a given word has a tendency to be scanty crosswise over various reports, a file that maps words to the records in which they happen can take a watchword question and quickly deliver important substance. Visual vocabularies offer a straightforward however successful approach to file pictures effectively with a rearranged record. An altered record file is much the same as a list in a book, where the catchphrases are mapped to the page numbers where those words are utilized. In the visual word case, we have a table that focuses from the word number to the files of the database pictures in which that word happens. The database pictures that share its particular words. Recovery by means of the transformed document is speedier than looking each picture, accepting that not all pictures contain each word. By and by, a picture's conveyance of words is without a doubt inadequate. Since the file keeps up no data about the relative spatial design of the words per picture, normally a spatial confirmation step is performed on the pictures recovered for a given inquiry.

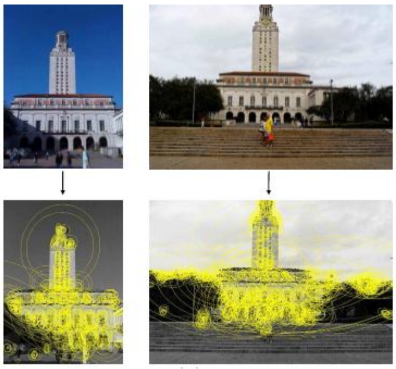

Fig .7.The meager particular focuses are likely ideal.

Fig.8.A thick multi-scale extraction

The regular content portrayal of a "Bag of-words "can be mapped over to the visual area: the picture's observational appropriation of words is caught with a histogram numbering how often every word in the visual vocabulary happens. This representation is that it deciphers a (for the most part expansive) arrangement of high-dimensional neighborhood descriptors into a solitary meager vector of altered dimensionality over all pictures. This thusly permits one to utilize numerous machine learning calculations that naturally expect the information space is vectorial whether for directed characterization, highlight determination, or unsupervised picture grouping.

VI. CONCLUSION

In this paper, a novel scalar quantization scheme is proposed on Bag of visual word (BovW) model for large scale image search. A SIFT descriptor can be easily adapted to the classic inverted file structure for indexing. BovW model address the images with good spatial and contextual information among their visual phrases a hash based indexing scheme has been employed with fusion based verification algorithm to provide better verification during retrieval. The benchmark dataset is used to measure and store the performance of this model.

REFERENCES

[1] B. T. Mark J. Huiskes and M. S. Lew, "New trends and ideas in visual concept detection: The MIR flickr retrieval evaluation initiative," in Proc.ACM Multimedia Inform. Retr. (MIR), 2010, pp. 527–536.

[2] D. Liu, G. Hua, P. Viola, and T. Chen, "Integrated feature selection and higher-order spatial feature extraction for object categorization," in Proc. CVPR, 2008, pp. 1–8.

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

[4] D. Nistér and H. Stewénius, "Scalable recognition with a vocabulary tree," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2006, pp. 161–2168.

[5] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in Proc. Comput. Vis. Pattern Recognit. (CVPR), vol. 2.2006, pp. 2161–2168.

[6] H. Jégou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in Proc. 10th EurConf. Comput. Vis. ECCV, 2008, pp. 304–317.

[7] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust Features," in Proc. Eur. Conf. Comput. Vis., May 2006.

[8] 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.

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

[10] Q. Tian, S. Zhang, W. Zhou, R. Ji, B. Ni, and N. Sebe, "Building Descriptive and discriminative visual codebook for large-scale image applications," Multimedia Tools Appl., vol. 51, no. 2, pp. 441–477, 2011.

[11] W. Zhou, Q. Tian, Y. Lu, L. Yang, and H. Li, "Latent visual context learning for web image applications," Pattern Recognit., vol. 44,nos. 10–11, pp. 2263–2273, Nov. 2011.

[12] Wengang Zhou, Houqiang Li, Richang Hong, Yijuan Lu, Mand Qi Tian, "BSIFT: Toward Data-Independent Codebook for Large Scale Image Search," IEEE,vol .24,no.3,march.2015.

[13] W. Zhou,H. Li,Y. Lu,and Q. Tian, "SIFT match verification by geometric coding for large-scale partial-duplicate web image search,"ACMTrans,Multimediacomput.,commun.,Appl.,vol.9,no.1,p.4,2013.