ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

Object and Scene Category Recognition Using a Combination of Dense Color SIFT Descriptors

[1]Taha H. Rassem, [2]Bee Ee Khoo, [1]Luhur Bayuaji, [2]Nasrin M. Makbol, [1]Suryanti Awang

[1]Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang (UMP), Lebuhraya Tun Razak, 26300 Gambang, Pahang, Malaysia.
[2]School of Electrical and Electronic Engineering, Universiti Sains Malaysia, 14300, Nibong Tebal, Penang, Malaysia.

ARTICLE INFO

Article history:
Received 22 February 2015
Accepted 20 March 2015
Available online 25 April 2015

Keywords:
Image category recognition;
SIFT;
Color SIFT;
Object data sets;
Scene data sets.

ABSTRACT

Background: Nowadays, many features have been proposed for image category recognition. Scale Invariant Feature Transform (SIFT) is one of the important descriptors, which is used in these systems. It is robust against rotation change, viewpoint change, scaling change, but it is partially robust against illumination change. Color SIFT descriptors are proposed to increase the illumination invariant. In this paper, the performances of different color SIFT descriptors densely extracted from the images were evaluated for object and scene recognition. RGB color SIFT, HSV color SIFT, Opponent color SIFT, Transformed-color SIFT and a new proposed color SIFT descriptor based on Ohta color space (Ohta Color SIFT) were used instead of the traditional gray SIFT. The performances of these descriptors and all their possible combinations were evaluated using challenging data sets. Caltech-04, Caltech-101, Caltech-256, Graz-02 are examples of object data sets used, whereas Oliva and Torralba data set (OT) and SUN-398 are examples of scene data sets. Using some combination of dense color SIFT descriptors, remarkable results of classification accuracy were achieved for some data sets such as Caltech-04 and Graz-02 and acceptable accuracy results for the remaining data sets as shown in experimental results.

© 2015 AENSI Publisher All rights reserved.

To Cite This Article: Taha H. Rassem, Bee Ee Khoo, Luhur Bayuaji, Nasrin M. Makbol, Suryanti Awang., Object and Scene Category Recognition Using a Combination of Dense Color SIFT Descriptors. ***Aust. J. Basic & Appl. Sci.,*** *9(12): 93-103, 2015*

INTRODUCTION

Automatically classifying the image based on its contents into the correct related category is the aim of image category recognition (Image classification). The contents of the image can be classified into two types; scene contents (e.g., outdoor, indoor, kitchen, etc.) (Fei-Fei and Perona, 2005, Amano *et al*., 2012, Mei *et al*., 2014) and object contents (e.g., cars, airplanes, motorbikes, fish, etc.) (Fei-Fei *et al*., 2004, Schulein *et al*., 2010, Mohammed *et al*., 2015). This process is not an easy task, especially for large numbers of images and categories. This is due to some challenges such as intra-class variation, inter-class variation, illumination change, scale change, rotations, viewpoint change, etc (Szeliski, 2010). Image category recognition usually includes three main steps. Firstly, deciding and selecting some points and regions in the image. This is called the detector step. There are different types of detectors as shown in Fig.1. The easiest one just selects some regions randomly. This is the random point detector. In the interest point detector, the high information regions are selected based on measuring some criteria between pixels intensity or color values. This is called the pixel interest point detectors. New interest point detectors using the image histogram such as intensity histogram or different color space histogram are proposed in (Lee and Chen, 2009, Rassem and Khoo, 2011). The random point detector and the interest point detector can be called the sparse point detectors (Lazebnik *et al*., 2006). The other type of the detectors is the dense point detector (Bosch *et al*., 2007a). It covers all locations in the image by regular patches with different scales as shown in Fig.1 (d). In recognition and classification tasks, the dense point detector achieved better performances than the interest point detector. This is because the density of the features is more helpful especially in case of low contrast images (Nowak *et al*., 2006) . Fig.1 (e) shows a new type of detector. It is called dense interest point detector. It selects all patches such as the dense sampling detector then selects the interest points from these points (Tuytelaars, 2010). Some systems have used sparse interest point detector and dense point detector

Corresponding Author: Taha H. Rassem, Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang (UMP), Lebuhraya Tun Razak, 26300 Gambang, Pahang, Malaysia.
Ph: +6095492245, Fax: +6095492144. E-mail: tahahussein@ump.edu.my

together to increase the classification accuracy (Bosch *et al*., 2007a, Boiman *et al*., 2008, Behmo *et al*., 2010). The second step is extracting the descriptors from the detected points or regions. This is the descriptor extraction step. These descriptors should have the capability to distinguish between images. The most important descriptor is the Scale Invariant Feature Transform (SIFT). The SIFT descriptor is proposed by Lowe in 1999 and until now, is playing an important role in many computer vision applications (Lowe, 1999, Lowe, 2004). This is due to its specific characteristics. It is rotation invariant, scale invariant, viewpoint invariant and partially illumination invariant. To increase SIFT illumination invariant, some researchers added some local color information to it (Bosch *et al*., 2007b, Sande *et al*., 2010). They extracted SIFT descriptor from each color channel in the color image. Different color SIFT descriptors are implemented such as HSV-SIFT (Bosch *et al*., 2007a), RGB-SIFT (Sande *et al*., 2010), Opponent SIFT (Behmo *et al*., 2010), C-SIFT (Sande *et al*., 2010), rgSIFT (Sande *et al*., 2010), Transformed-color SIFT (Sande *et al*., 2010), etc. The SIFT descriptor and the color SIFT descriptors are called the appearance descriptors (Bosch *et al*., 2007a). In addition to SIFT, there are many other descriptors in the computer vision literature such as GLOH (Gradient Location and Orientation Histogram) (Mikolajczyk and Schmid, 2005), HOG (Histogram of Oriented Gradients)(Dalal and Triggs, 2005), LBP (Local Binary Patterns) (Ojala *et al*., 2002, Maddah and Mozaffari, 2012), spin image (Lazebnik *et al*., 2005), SURF (Speeded Up Robust Feature) (Bay *et al*., 2006), region covariance (Tuzel *et al*., 2007) and self-similarity descriptor (Shechtman and Irani, 2007), geometric blur (Berg *et al*., 2005, Zhang *et al*., 2006), etc. Some researchers have combined different descriptors in their systems (Behmo *et al*., 2010, Gehler and Nowozin, 2009a, Gehler and Nowozin, 2009b, Bosch *et al*., 2007a, Boiman *et al*., 2008). Although, this needs an extensive memory and time, better classification accuracies are obtained. This is because one category may have a better response to one descriptor than another. After extracting the descriptors sparsely or densely from the images, a classifier or multiple classifiers are needed to classify these images based on their descriptors into the related categories. This is the third step. It is the learning step. In the beginning, the classification systems used some single kernel classifiers such as K-NN (Szummer and Picard, 1998, Oliva and Torralba, 2001), SVM (Zhang *et al*., 2006), etc. Appearance of new challenging data sets and the need to use many descriptors led to think about several kernel learning classifiers. These are Multiple Kernel learning (MKL) (Vedaldi *et al*., 2009), Incremental Multiple Kernel learning (IMKL) (Kembhavi *et al*., 2009), Group-Sensitive Multiple Kernel Learning (GS-MKL) (Yang *et al*., 2009a), etc.

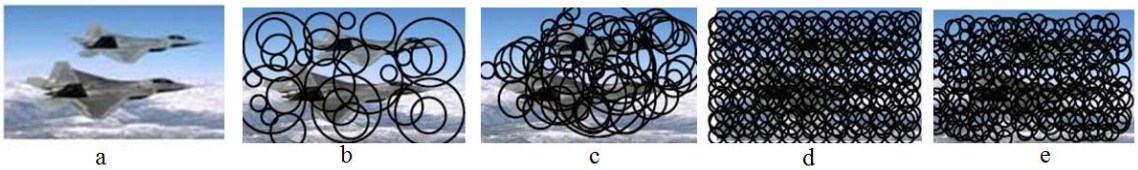

Fig. 1: Types of image detectors. 1(a) Original image. 1(b) Random point detector. 1(c) Interest point detector. 1(d) Dense point detector. 1(e) Dense interest point detector.

In this paper, the performances of the appearance descriptors were studied and tested for object and scene category recognition. Four known color SIFT descriptors were extracted densely from the images. These are HSV-SIFT, RGB-SIFT, Opponent SIFT and Transformed-color SIFT. Furthermore, a new color SIFT descriptor was proposed by adding Ohta-color information (Y. Ohta, 1980). This is called Ohta-color SIFT. The performances of these appearance descriptors and all possible combinations of them were evaluated using object data sets such as Caltech-04, Graz-02, Caltech-101 and Caltech-256 (Griffin *et al*., 2007) and scene data sets such as Oliva & Torralba (OT) (Oliva and Torralba, 2001) and SUN-398 (Scene UNderstanding) data sets (Xiao *et al*., 2010). Caltech-101, Caltech-256 and SUN-398 are the most challenging object and scene data sets. Where, Caltech-101 has 102 different object categories, Caltech-256 has 257 different object categories and the SUN-398 has 398 scene different categories. Some combinations of the appearance descriptors showed high recognition accuracy in the small object data such as Caltech-04 and Graz-02. In addition, they showed good performances with the other data sets such as OT8 and SUN-398.

This paper is organized as follows. Section 2 reviews some related work in image category recognition. The color SIFT descriptors are briefly described in Section 3. Then in Section 4, the experimental setup and results are presented. Finally, Section 5 concludes this paper.

Related Work:

The image classification or image category recognition systems have passed through several strategies. In the beginning, the researchers worked with the low level features. Different features are

suggested and extracted such as color, texture, shape, power spectrum, etc. Then, different supervised learning techniques are used for classifying the images (Vailaya *et al*., 2001). The semantic representation of the image contents is missing in this strategy.

The recent strategy used an intermediate semantic representation before performing the classification process. Some examples of semantic representation are Bag-of-Features (BoF) models, generative models that identify the semantic as a set of materials or objects that appear in the image, model the spatial correlation of basic elements such as position and spatial relationships, etc. The most recent strategy is based on extracting an invariant descriptor from a distinctive region of the image and then classifying these descriptors using supervised learning algorithms. Furthermore, it is extended to extract multiple invariant descriptors and use multiple learning algorithms in the same system (Vedaldi *et al*., 2009, Gehler and Nowozin, 2009a, Gehler and Nowozin, 2009b, Boiman *et al*., 2008).

Oliva *et al*. (2001) proposed a computational model to recognize the scene images. A very low dimensional representation of the scene called a Spatial Envelope is suggested. The dimensions of the Spatial Envelope consist of set of perceptual qualities. These are naturalness, openness, roughness, expansion, and ruggedness. These qualities or features together represent the dominant spatial structure of a scene. They used the spectral and coarsely localized information to estimate these dimensions. Using this model, a multidimensional space is generated. The scene categories that are semantically related are projected closed such as street and highways. They evaluated the system using their own data set (OT data set).

Recently, the BoF model represents an image as a histogram of its local features. (Bosch *et al*., 2008) used probabilistic Latent Semantic Analysis (pLSA) to learn the latent of unlabeled scene image based on some of labeled training scene images. The BoF model is used with different types of descriptors. They used support vector machines (SVM) and k-Nearest Neighbor (k-NN) as learning classifiers and different scene data sets for system evaluation. (Behmo *et al*., 2010) used SIFT descriptor with optimal Naive Bayes Nearest Neighbor classifier. Furthermore, they used optimal NBNN classifier with Opponent SIFT, rgSIFT, C-SIFT and Transformed-color SIFT. (Opelt *et al*., 2006) used three types of detectors with four types of descriptors and classified Caltech-04 and Graz02 using adaboost. (Hegazy and Denzler, 2008a) used Hessian-Affine interest point detector to extract SIFT and histogram of Opponent color for each image. They used adaboost for classification. They repeated the same work by replacing the GLOH descriptor instead of SIFT descriptor (Hegazy and Denzler, 2008b). (Han *et al*., 2009) used BoF to represent SIFT descriptor after using Difference of Gaussian (DOG) detector. They modeled the visual word using supervised nonlinear neighborhood embedding space to be more discriminate and then used the Probabilistic Neural Network for classification. (Lazebnik *et al*., 2006) proposed the Spatial Pyramid Matching (SPM) for scene classification. They improved the BoF image representation model by computing histograms of dense SIFT descriptor from regular image blocks. They also extended their work for object data sets. As the Caltech-101 and Caltech-256 are more challenging than other data sets, a lot of work had been done to classify these data sets. (Maji *et al*., 2008) used multi-level HOG descriptor with the intersection kernel SVM for classification. (Zhang *et al*., 2006) used SVM-KNN classifier to classify multiclass data sets. They used geometric blur descriptors from image regions. (Bosch *et al*., 2007a) used dense SIFT, dense HSV-SIFT descriptors and PHOG in their system and used Multi-way SVM as a classifier. (Yang *et al*., 2009b) and (Wang *et al*., 2010) used Sparse coding SPM (ScSPM) and Locality-constrained Linear Coding (LLC) instead of Vector Quantization (VQ) respectively to increase the discriminating power of the BoF image representation and to reduce the training complexity. Vedaldi *et al*. (Vedaldi *et al*., 2009) implemented Multiple kernel Learning (MKL) classifier to train dense SIFT, geometric blur and self similarity descriptors. To our knowledge, the best Caltech-101 and Caltech-256 classification results are achieved by (Gehler and Nowozin, 2009b) in their extension work and by (Yang *et al*., 2009a). (Gehler and Nowozin, 2009b) used V1S+ features, Local Binary Patterns, region covariance, dense SIFT, HSV-SIFT, PHOG and geometric blur. They used LP-β and LP-B to classify those descriptors. (Yang *et al*., 2009a) used GS-MKL to classify many types of descriptors. They used dense L* a* b* color SIFT, dense-SIFT, self similarity, PHOG and Gabor feature (Ralló *et al*., 2009).

Color Sift Descriptors:

Adding local color information to SIFT descriptor is suggested to increase the illumination invariant. In (Bosch *et al*., 2007a), the SIFT descriptors are extracted over all channels of the HSV color model. Examples of color SIFT descriptors are C-SIFT, rgSIFT, HueSIFT, RGB-SIFT, Opponent SIFT, HSV-SIFT and Transformed-color SIFT (Sande *et al*., 2010). In this work, RGB-SIFT, HSV-SIFT, Opponent SIFT, Transformed-color SIFT, and the new proposed Ohta color SIFT descriptors were used as system descriptors. These descriptors were extracted densely from the image data sets. These dense SIFT descriptors are summarized in Table 1. In fact, each descriptor may have different performance in each category. Combining different descriptors together helped to solve this problem. All possible combinations of

these descriptors were used in the classification system.

Experimental Setup and Results:

Fig. 2 shows an overview of the stages of image category recognition system and BoF model. The dense color SIFT descriptors were extracted, and then the training models were learned using χ2-SVM. BoF representation was used to represent the extracted descriptors. Using a set of available images, the k-means clustering was used to build the visual code book. 300 word vocabulary was used in Caltech-04 experiments while 600 words vocabulary was used in the other data sets experiments. With each descriptor, one-level spatial histograms were computed on grids 2 x 2 (Lazebnik *et al*., 2006). In all data sets except Caltech-256 and SUN-398, the experiments were repeated five times with random training and testing images using each descriptor. The average of the results are considered as the accuracy of the classification. With Caltech-256 and SUN-398, the experiments were done only one time with each descriptor.

Table 1: Some color SIFT descriptors.

Descriptor name	Color space	Size	Conversion equation
RGB-SIFT	RGB color space	3 ×128	
HSV-SIFT	HSV color space	3 ×128	
Ohta SIFT	Ohta color space	3 ×128	$I_1=\frac{R+G+B}{3}$ $I_2=R-B$ $I_3=\frac{2G-R-B}{2}$
Opponent SIFT	Opponent color space	3 ×128	$O_1=\frac{R-G}{\sqrt{2}}$ $O_2=\frac{R+G-2B}{\sqrt{6}}$ $O_3=\frac{R+G+B}{\sqrt{3}}$
Transformed color SIFT	Transformed color space	3 ×128	$R'=\frac{R-\mu_R}{\sigma_R}$ $G'=\frac{G-\mu_G}{\sigma_G}$ $B'=\frac{B-\mu_B}{\sigma_B}$

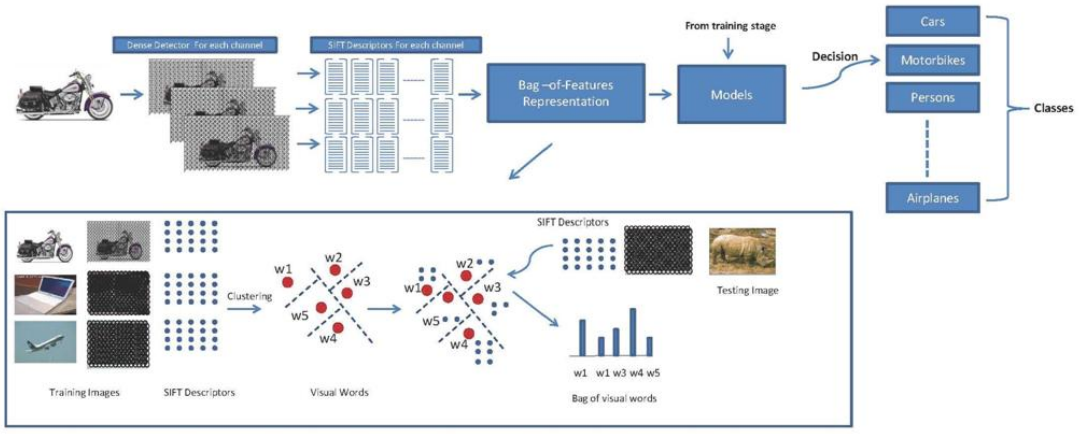

Fig. 2: Overview of image category recognition system and Bag-of-Features (BoF) stage.

Scene data set experiments:

1. Oliva and Torralba data set (OT):

OT data set is 2688 color images in total. It contains eight categories: coast category has 360 images, forest category contains 328 images, mountain category contains 374 images, open country category contains 410 images, highway category contains 260 images, inside the city category contains 308 images, tall building category contains 356 images and street category contains 292 images. These images are in the JPG format and contains average size of 265 x 265 pixels. Oliva and Torralba data set (OT) can be evaluated together as eight scene categories (OT8) and can be divided into two groups; Man-Made categories (OT4MM) which contain inside of city, highway, street and tall building categories and Natural categories (OT4N) which contain coasts, forest, mountain and open country categories. In OT8, 100 random images were selected for training and the remaining images were used for testing in each category. Using one color SIFT descriptor, the Transformed-color SIFT descriptor performed better than others. The performance using Ohta color SIFT only was the worst but on the other hand, when it is combined with other descriptors accuracy performance was improved. The best result was achieved by adding Ohta color SIFT descriptor to the combination of HSV color SIFT and Transformed color SIFT. In OT4MM, 250 images were selected for training, and the remaining images in each category were used for testing. The Transformed color SIFT achieved the best accuracy results as a single descriptor. The combination of Ohta color SIFT, RGB color SIFT and Opponent color SIFT achieved the best classification accuracy. In OT4N, 250 images were selected for training, and the remaining images in each category were used for testing. The RGB color

SIFT achieved the best accuracy result as a single descriptor. The best classification performance was achieved using the combination of Transformed color SIFT and Opponent color SIFT. The second best result was achieved using the combination of Ohta color SIFT, RGB color SIFT and HSV color SIFT. Generally, the OT four natural images' results were better than the OT four man-made images. This is because, the density of the descriptors is ideal for the low-contrast images. It is very useful for the images that have uniform regions such as sky, water, etc. On the other hand, the OT four man-made data set images have more objects and need to use some interest point detectors to distinct these objects. The OT data set classification accuracy results are shown in Table 2. For comparison, Table 3 shows some of the state-of-the art results using OT data sets.

Table 2: Results of OT data set.

Descriptors	OT8	OT4N	OT4MM	Descriptors	OT8	OT4N	OT4MM
RGB[1]	83.16	91.47	91.47	RGB+HSV+Transf.	84.23 91.59 84.12	91.59	84.12
Opp[2]	80.94	91.47	82.73	HSV+Opp.+Transf.	84.61	93.04	84.65
HSV[3]	81.07	90.28	81.35	RGB+Opp.+Transf.	84.55	92.66	84.25
Transf.[4]	84.40	90.83	84.30	Ohta+HSV+Transf.	85.75	92.49	83.21
Ohta[5]	61.86	75.98	65.12	RGB+HSV+Opp.	84.10	92.71	84.86
RGB+ Opp.	84.37	92.91	84.97	Ohta+RGB+HSV	84.46	92.85	83.88
RGB+Transf.	83.65	91.42	83.43	Ohta+RGB+Opp.	85.43	92.54	84.69
RGB+HSV	83.27	92.29	83.96	Ohta+RGB+Transf.	84.96	92.53	85.57
HSV+Transf.	84.41	91.42	82.49	Ohta+HSV+Opp.	83.49	92.26	82.17
HSV+Opp.	83.11	92.32	82.64	Ohta+Opp.+Transf.	85.18	92.32	85.04
Transf. + Opp.	84.08	93.12	84.60	HSV+Opp.+Transf.+RGB	84.87	92.54	83.97
Ohta + RGB	84.45	92.40	84.54	Ohta+RGB+HSV+Opp.	85.02	92.73	85.42
Ohta + Opp.	82.27	91.25	83.84	Ohta+Transf.+Opp.+HSV	85.15	93.08	85.67
Ohta + HSV	82.20	90.71	79.10	Ohta+RGB+Transf.+Opp.	84.98	92.10	85.48
Ohta + Transf.	84.15	92.47	84.99	Ohta+RGB+Transf.+HSV	84.72	92.31	84.22
				Ohta+RGB+Transf.+HSV+Opp...	85.26	93.17	85.03

[1]RGB SIFT. [2]Opponent SIFT. [3]HSV-SIFT. [4]Transformed color SIFT. [5]Ohta SIFT.

Table 3: Results of some state of the art using OT data set.

References	OT8	OT4N	OT4MM
Oliva *et al*. (2001)	83.7	89	89
Bosch *et al*.(2007)	87.8	93.90	94.8
Our best result	85.75	93.17	85.67

2. SUN 398 data set:

SUN data set is a new scene data set and contains 908 scene categories and more than 130,519 JPG color images (Xiao *et al*., 2010). Now, It is considered as the largest scene data set. (Xiao *et al*., 2010) selected only 398 categories from the 908 scene categories and used them in their experiments. The largest category of SUN-398 is living room category that contains 2361 images whereas the small categories contain only 100 images. As reported in (Xiao *et al*., 2010), 50 images were used for training and 50 images for testing in each category. While the worst results were obtained using the Ohta SIFT color SIFT, the best result was achieved using a combination of the Ohat color SIFT, the HSV-SIFT and the Transformed color SIFT descriptors. Table 4 shows the results of SUN-398 classification. More than 12 descriptors are combined together to classify SUN-398 data set (Xiao *et al*., 2010). These are GIST, HOG, dense SIFT, LBP, sparse SIFT histograms, SSIM, tiny images, line features, texton histograms, color histograms, geometric probability map, and geometry specific histograms. The best single descriptor was HOG which achieves around 27.2% while 38% of classification accuracy is achieved using all descriptors. The combination of the Ohta color SIFT, the HSV-SIFT and the Transformed color SIFT descriptors achieved the second best results for SUN-398 classification as shown in Fig. 2. It is around 28.77%.

Table 4: Results of SUN-398 data set.

Descriptors	50	Descriptors	50
RGB	25.98	RGB+HSV+Transf.	27.42
Opp	25.59	HSV+Opp.+Transf.	27.75
HSV	24.91	RGB+Opp.+Transf.	27.83
Transf.	25.70	Ohta+HSV+Transf.	28.77
Ohta	17.11	RGB+HSV+Opp.	28.48
RGB+ Opp.	27.74	Ohta+RGB+HSV	28.50
RGB+Transf.	25.86	Ohta+RGB+Opp.	28.33
RGB+HSV	27.12	Ohta+RGB+Transf.	28.36
HSV+Transf.	27.38	Ohta+HSV+Opp.	28.19
HSV+Opp.	26.92	Ohta+Opp.+Transf.	28.49
Transf. + Opp.	27.99	HSV+Opp.+Transf.+RGB	28.41
Ohta + RGB	28.38	Ohta+RGB+HSV+Opp.	28.15
Ohta + Opp.	26.15	Ohta+Transf.+Opp.+HSV	28.33
Ohta + HSV	26.10	Ohta+RGB+Transf.+Opp.	28.76
Ohta + Transf.	28.52	Ohta+RGB+Transf.+HSV	28.41
		Ohta+RGB+Transf.+HSV+Opp...	28.65

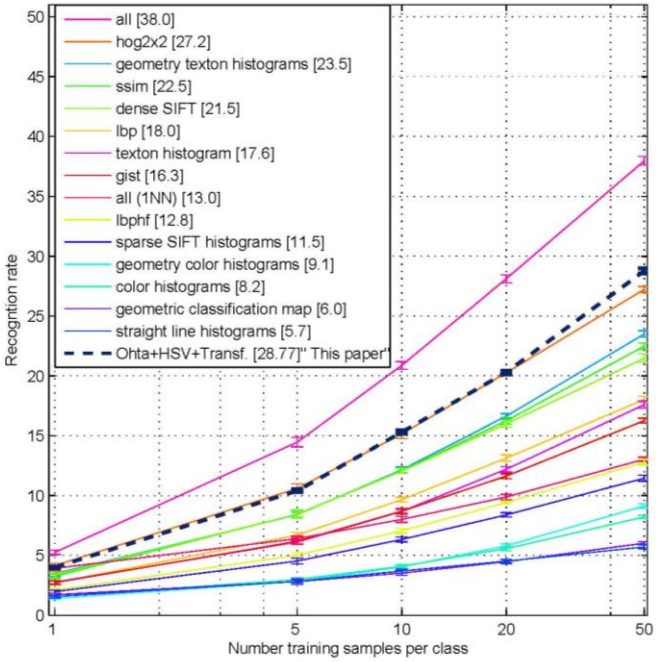

Fig. 3: Our SUN-398 classification accuracy results with Xiao *et al*. (2010) results.

Object data set experiments:

1. Graz-02 data set:

The Graz-02 data set is more challenging than the Caltech data set. The objects are shown on a complex cluttered background, at different scales, and with different object positions (Opelt *et al*., 2006). The Graz-02 contains 311 images of the category person, 365 images of the category bike, 420 images of category car and 380 of the category no-bikes-no-person-no-car category (counter class). Each image is about 640 × 480 pixels in dimension. It is an extension from Graz-01. In this data set, 150 images were used for training and 75 for testing for each category after resizing them to 320 × 480 pixels. The results are shown in Table 5. In terms of the average of the three categories, Transformed color SIFT descriptor achieved the best as a single descriptor, while the combination of Opponent SIFT and Transformed color SIFT achieved the highest accuracy result. Table 6 shows the state-of-the art results using Graz-02.

2. Caltech-04 data set:

The Caltech-04 data set has altogether 5775 images, including 1074 airplanes, 1155 cars, 450 faces, 826 motorbikes, 1370 car backgrounds and 900 general backgrounds. In each category, 100 images were used for training and 50 images for testing. As shown in Table 7, Opponent color SIFT performed better than others as a single descriptor. The combination of Ohta color SIFT, RGB SIFT and Transformed color SIFT achieved the highest classification accuracy. This combination classified three types of images to their related categories 100% correctly. Table 8 shows the state-of-the art results using Caltech-04.

Table 5: Results of Graz-02 data set.

Descriptors	Bikes	Cars	Persons	Average	Descriptors	Bikes	Cars	Persons	Average
RGB	91.11	66.22	85.33	80.89	RGB+HSV+Transf.	92.89	72.44	84.44	83.26
Opp	93.78	64	90.22	82.67	HSV+Opp.+Transf.	93.78	65.78	88.89	82.82
HSV	94.22	66.67	87.56	83.70	RGB+Opp.+Transf.	94.22	71.11	90.67	85.33
Transf.	92.45	73.33	86.22	84	Ohta+HSV+Transf.	94.67	69.34	87.11	83.71
Ohta	84.44	46.67	78.67	69.93	RGB+HSV+Opp.	96	71.55	85.78	84.44
RGB+ Opp.	91.55	70.67	87.55	83.62	Ohta+RGB+HSV	94.66	59.56	88	80.74
RGB+Transf.	93.33	76.44	80.55	83.44	Ohta+RGB+Opp.	96	70.67	87.11	84.59
RGB+HSV	91.56	72	86.67	83.41	Ohta+RGB+Transf.	94.67	65.33	85.78	81.93
HSV+Transf.	92.89	66.66	88	82.52	Ohta+HSV+Opp.	93.33	68.89	89.78	84
HSV+Opp.	94.22	73.33	87.56	85.04	Ohta+Opp.+Transf.	94.22	70.67	89.78	84.89
Transf. + Opp.	94.67	76.45	88.45	86.52	HSV+Opp.+Transf.+RGB	93.78	66.22	87.55	82.52
Ohta + RGB	92.44	69.78	83.55	81.92	Ohta+RGB+HSV+Opp.	95.56	63.11	89.34	82.93
Ohta + Opp.	94.22	70.67	92	85.63	Ohta+Transf.+Opp.+HSV	95.55	65.33	88.89	83.26
Ohta + HSV	93.33	66.22	89.33	82.96	Ohta+RGB+Transf.+Opp.	93.33	68.44	89.78	83.85
Ohta + Transf.	93.78	69.78	83.56	82.37	Ohta+RGB+Transf.+HSV	96.44	63.11	89.78	83.11
					Ohta+RGB+Transf.+HSV+Opp.	94.22	72.44	88	84.89

Table 6: Results of some state of the art using Graz-02.

References	Bikes	Cars	Persons
Behmo *et al*. (Behmo *et al*., 2010)	78.70	82.05	76.20
Hegazy *et al*.(Hegazy and Denzler, 2008a)	74.67	81.33	81.33

Hegazy *et al*.(Hegazy and Denzler, 2008b)	80	78.62	84
Zhang *et al*.(Zhang *et al*., 2010)	88.9	85.2	88.1
Mutch *et al*.(Mutch and Lowe, 2008)	80.5	70.1	81.7
Oplet *et al*.(Opelt *et al*., 2006)	77.8	70.5	81.2
Moosmann *et al*.(Moosmann *et al*., 2007)	84.4	79.9	--
Our best result(Transf.+Opp.)	94.67	76.45	88.45

3. Caltech-101 data set:

The Caltech-101 data set has altogether 9146 images, in a total of 102 classes; 101 distinct object classes such as faces, watches, ant, etc., and background class. Each image is about 300 × 200 pixels in dimension (Fei-Fei *et al*., 2007). The smallest category size is 31 images. The inter-class variability is the challenge of Caltech-101 in addition to the number of the categories. In each category, 30 and 15 images were used for training while 50 images were used for testing. The best result was obtained using the combination of all color SIFT descriptors. As a single descriptor, Transformed color SIFT achieved better results than others. Table 9 shows the results of Caltech-101 using different color SIFT while Table 10 shows the results of some state-of-the art using Caltech-101.

Table 7: Results of Caltech-04.

Descriptors	Faces	Motor bikes	Airplanes	Cars	Average	Descriptors	Faces	Motor bikes	Airplanes	Cars	Average
RGB	92	98	96	100	96.5	RGB+HSV+Transf.	94	98	98	98	97
Opp	100	96	98	100	98.5	HSV+Opp.+Transf.	100	94	98	100	98
HSV	94	92	100	100	96.5	RGB+Opp.+Transf.	98	96	98	100	98
Transf.	96	98	98	98	97.5	Ohta+HSV+Transf.	96	96	98	100	97.5
Ohta	96	86	88	74	86	RGB+HSV+Opp.	96	92	98	100	96.5
RGB+ Opp.	98	92	98	100	97	Ohta+RGB+HSV	94	94	98	100	96.5
RGB+Transf.	94	98	98	98	97	Ohta+RGB+Opp.	96	94	98	100	97
RGB+HSV	92	92	98	100	95.5	Ohta+RGB+Transf.	96	100	100	100	99
HSV+Transf.	92	96	98	98	96	Ohta+HSV+Opp.	98	94	100	100	98
HSV+Opp.	98	96	98	100	98	Ohta+Opp.+Transf.	98	96	100	100	98.5
Transf. + Opp.	98	98	98	100	98.5	HSV+Opp.+Transf.+RGB	96	94	98	100	97
Ohta + RGB	94	96	98	100	97	Ohta+RGB+HSV+Opp.	94	92	98	100	96
Ohta + Opp.	96	94	100	100	97.5	Ohta+Transf.+Opp.+HSV	100	96	98	100	98.5
Ohta + HSV	96	92	100	98	96.5	Ohta+RGB+Transf.+Opp.	98	96	98	100	98
Ohta + Transf.	96	98	100	100	98.5	Ohta+RGB+Transf.+HSV	94	96	98	100	97
						Ohta+RGB+Transf.+HSV+Opp	96	96	98	100	97.5

Table 8: Results of some state of the art using Caltech-04.

References	Faces	Motorbikes	Airplanes	Cars
Han *et al*. (2009)	93.53	94.41	96.29	94.35
Oplet *et al*. (2006)	93.5	92.2	88.9	91.1
Hegazy *et al*.(2008)a	94	96	84	100
Hegazy *et al*.(2008) b	100	93	84	94
Fergus *et al*. (2003)	96.4	92.5	90.2	88.5
Thureson *et al*. (2003)	83.1	93.2	83.8	90.2
(Weber *et al*., 2000)	93.5	88	-	86.5
Our best result(Ohta+RGB+Transf.)	96	100	100	100

4. Caltech-256 data set:

The Caltech-256 data set has altogether 30,607 images, in a total of 257 classes; 256 distinct object classes such as faces, watches, ant, etc., and background class. Each image is about 300 × 200 pixels in dimension (Griffin *et al*., 2007). Caltech-256 is more challenging than Caltech-101. The smallest category size is 80 images. Caltech-256 shows more inter-class variability and intra-class variability than Caltech-101 in addition to more category numbers. Using this data set, the system is trained using 30 images and tested using 25 images from each category. RGB-SIFT descriptor performed better than other single descriptors. The highest accuracy results were obtained using the combination of Ohta color SIFT, RGB-SIFT, Transformed color SIFT and Opponent SIFT. Table 11 shows the results of Caltech-101 using different color SIFT while Table 12 shows the results of some of the state of the art using Caltech-256.

Table 9: Results of Caltech-101 data set.

Descriptors	30	15	Descriptors	30	15
RGB	68.59	62.64	RGB+HSV+Transf.	69.01	63.44
Opp	59.50	53.27	HSV+Opp.+Transf.	68.86	61.79
HSV	63.60	57.45	RGB+Opp.+Transf.	68.95	63.88
Transf.	69.26	63.27	Ohta+HSV+Transf.	69.35	63.31
Ohta	36.87	31.20	RGB+HSV+Opp.	68.40	62.09
RGB+ Opp.	68.76	62.70	Ohta+RGB+HSV	68.78	63.46
RGB+Transf.	68.89	62.48	Ohta+RGB+Opp.	69.10	63.25
RGB+HSV	68.99	62.333	Ohta+RGB+Transf.	70.13	64.34
HSV+Transf.	69.43	62.59	Ohta+HSV+Opp.	65.33	57.12
HSV+Opp.	65.25	58.21	Ohta+Opp.+Transf.	68.91	63.09
Transf. + Opp.	68.45	63.40	HSV+Opp.+Transf.+RGB	69.17	63.25

Ohta + RGB	69.96	63.88	Ohta+RGB+HSV+Opp.	69.49	63.59
Ohta + Opp.	59.19	53/09	Ohta+Transf.+Opp.+HSV	69.51	63.55
Ohta + HSV	64.64	57.47	Ohta+RGB+Transf.+Opp.	70.38	65.03
Ohta + Transf.	69.51	63.42	Ohta+RGB+Transf.+HSV	70.26	64.71
			Ohta+RGB+Transf.+HSV+Opp...	70.36	64.07

Table 10: Results of some state of the art using Caltech-101.

References	30	15
Our best results (Ohta+RGB+Transf.+Opp.)	70.38	65.03
Ma *et al*. (Ma and Grimson, 2008)	70.38	-
Jain *et al.* (Jain *et al*., 2008)	69.6	-
Pinto *et al*.(Pinto *et al*., 2008)	67.36	61.44
Zhang *et al*.(Zhang *et al*., 2006)	66.23	59.44
Frome *et al*.(Frome *et al*., 2006)	66	60.30
Sharma *et al*.(Sharma *et al*., 2008)	65.50	-
Lee *et al*.(Lee *et al*., 2009)	65.40	-
Lazebnik *et al*.(Lazebnik *et al*., 2006)	64.60	56.40
Van Gemert *et al*. (van Gemert *et al*., 2010)	64.12	-
Van Gemert *et al*. (van Gemert *et al*., 2008)	64.10	-

Conclusions:

In most computer vision applications to enhance the performance of the SIFT, color SIFT descriptors are proposed as an extension of the traditional gray SIFT. In this paper, four known color SIFT descriptors with a new proposed color SIFT were densely extracted and used for the image category recognition task. These are HSV-SIFT, Opponent color SIFT, Transformed color SIFT, RGB-SIFT and new Ohta color SIFT. The performance of these descriptors and all their possible combinations were evaluated using different scene and object data sets. While Opponent color SIFT descriptor achieved higher classification accuracy as a single descriptor in Caltech-04 data set, the Transformed color SIFT and the RGB-SIFT descriptors were the best as single descriptors in the remaining data sets. The combination of four descriptors (Ohat SIFT, RGB-SIFT, Transformed color SIFT and Opponent color SIFT) performed the best classification accuracy results in the challenging data sets such as SUN-398, Caltech-101 and Caltech-256. Furthermore, the combination of Ohta SIFT, HSV-SIFT and Transformed color SIFT achieved the best results in SUN-398 and OT8 data sets. In OT4NN, the highest classification accuracy was obtained using the combination of all color SIFT descriptors. It was very close to the state of the art results of OT4NN data set that was obtained by Bosch *et al* (Bosch *et al*., 2008). Generally speaking, densely color SIFT descriptors showed the best performance with Caltech-04 and Graz-02. Moreover, they appeared an acceptable performance with the remaining data sets that had complex clutters, different object positions and scales, and a large number of classes. Comparing with (Xiao *et al*., 2010) experiments on the SUN-398 data set, our best results of this data set were the second highest ranking results. Although, the results of the classification accuracy with Caltech-101 and Caltech-256 were below the best state of the art result that achieved by (Yang *et al*., 2009a) (84.3%) and (Gehler and Nowozin, 2009b) (50.8%) respectively, the obtained results outperformed some of the state of the art as shown in Table 10 and Table 12. This proves that the appearance descriptors alone are incapable of distinguishing and classifying the data sets with a large number of categories. For large scale data sets, it is needed to use different types of descriptors such as appearance descriptors (e.g. SIFT, color SIFT, etc.), shape descriptors (e.g. PHOG), texture descriptors (e.g. LBP), etc. Furthermore, as the new data sets have more challenges such as complex backgrounds, cluttering, intra-class variation, inter-class variation, etc., the dense detector is not enough. So, in addition to the dense detector, powerful interest point detectors should be used to select the high distinctive points. Use of different descriptors with different detectors explains the reasons for achieving the best results of the state of the art in (Yang *et al*., 2009a, Gehler and Nowozin, 2009b).

Table 11: Results of Caltech-256.

Descriptors	30	Descriptors	30
RGB	33.15	RGB+HSV+Transf.	33.85
Opp	28.39	HSV+Opp.+Transf.	34.24
HSV	30.27	RGB+Opp.+Transf.	35.05
Transf.	32.54	Ohta+HSV+Transf.	35.18
Ohta	18.05	RGB+HSV+Opp.	34.09
RGB+ Opp.	34.52	Ohta+RGB+HSV	34.65
RGB+Transf.	32.92	Ohta+RGB+Opp.	34.91
RGB+HSV	33.81	Ohta+RGB+Transf.	34.94
HSV+Transf.	34.09	Ohta+HSV+Opp.	32.12
HSV+Opp.	31.98	Ohta+Opp.+Transf.	35.11
Transf. + Opp.	34.01	HSV+Opp.+Transf.+RGB	34.12

Ohta + RGB	34.71	Ohta+RGB+HSV+Opp.	34.89
Ohta + Opp.	29.18	Ohta+Transf.+Opp.+HSV	35.08
Ohta + HSV	30.66	Ohta+RGB+Transf.+Opp.	35.46
Ohta + Transf.	34.63	Ohta+RGB+Transf.+HSV	35.19
		Ohta+RGB+Transf.+HSV+Opp...	35.05

Table 12: Results of some state of the art using Caltech-256.

References	OT4MM
Our best results(Ohta+RGB+Transf.+Opp.)	35.46
Griffin *et al*.(Griffin *et al*., 2007)	34.20
Yang *et al*.(Yang *et al*., 2009b)	34.02
Van Gemert *et al*.(van Gemert *et al*., 2008)	27.17

ACKNOWLEDGMENT

The authors gratefully acknowledge for the support from Fundamental Research Grant Scheme (RDU130114), awarded to Universiti Malaysia Pahang by Ministry Of Education Malaysia.

REFERENCES

Amano, K., D.H. Foster, M.S. Mould, J.P. Oakley, 2012. Visual search in natural scenes explained by local color properties. J. Opt. Soc. Am. A, 29: A194-A199.

Bay, H., T. Tuytelaars, L. Van Gool, 2006. SURF: Speeded-up robust features. Proceedings of the European Conference on Computer Vision. ECCV.

Behmo, R., P. Marcombes, A. Dalalyan, V. Prinet, 2010. Towards optimal naive bayes nearest neighbor. Proceedings of the European Conference on Computer Vision. ECCV.

Berg, A., T. Berg, J. Malik, 2005. Shape matching and object recognition using low distortion correspondences. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE.

Boiman, O., E. Shechtman, M. Irani, 2008. In defense of Nearest-Neighbor based image classification. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, pp: 1-8.

Bosch, A., A. Zisserman, X. Munoz, 2007a. Image Classification using random forests and ferns. Eleventh International Conference on Computer Vision, IEEE, pp: 1-8.

Bosch, A., A. Zisserman, X. Munoz, 2007b. Representing shape with a spatial pyramid kernel. ACM international conference on Image and video retrieval. ACM.

Bosch, A., A. Zisserman, X. Muoz, 2008. Scene classification using a hybrid generative/discriminative approach. IEEE Trans. Pattern Anal. Mach. Intell., 30: 712-727.

Dalal, N., B. Triggs, 2005. Histograms of oriented gradients for human detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE.

Fei-Fei, L., R. Fergus, P. Perona, 2004. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. Computer Vision and Pattern Recognition Workshop. IEEE, 178-178.

Fei-Fei, L., R. Fergus, P. Perona, 2007. Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. Comput. Vis. Image Und., 106: 59-70.

Fei-Fei, L., P. Perona, 2005. A Bayesian Hierarchical Model for Learning Natural Scene Categories. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition IEEE.

Frome, A., Y. Singer, J. Malik, 2006. Image retrieval and classification using local distance functions. Neural Information Processing Systems(NIPS). MIT Press.

Gehler, P., S. Nowozin, 2009a. On feature combination for multiclass object classification. Twelfth International Conference on Computer Vision. IEEE.

Gehler, P., S. Nowozin, 2009b. On feature combination for multiclass object detection-Extension work, http://www.vision.ee.ethz.ch/~pgehler/projects/iccv09/index.html.

Griffin, G., A. Holub, P. Perona, 2007. Caltech-256 object category dataset. In: TECH. REP. 7694, C. I. O. T. (ed.) Tech. Rep. 7694, California Institute of Technology. Tech. Rep. 7694, California Institute of Technology.

Han, X.H., Y.W. Chen, X. Ruan, 2009. Object class recognition with supervised nonlinear neighborhood embedding of visual words. Proceedings of the First International Conference on Internet Multimedia Computing and Service. ACM.

Hegazy, D., J. Denzler, 2008a. Boosting colored local features for generic object recognition. Pattern Recognition and Image Analysis, 18: 323-327.

Hegazy, D., J. Denzler, 2008b. Generic object recognition using boosted combined features. Proceedings of the 2nd international conference on Robot vision. Springer-Verlag.

Jain, P., B. Kulis, K. Grauman, 2008. Fast image search for learned metrics. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE.

Kembhavi, A., B. Siddiquie, R. Miezianko, S. Mccloskey, L. Davis, 2009. Incremental multiple

kernel learning for object recognition. Twelfth International Conference on Computer Vision. IEEE.

Lazebnik, S., C. Schmid, J. Ponce, 2005. A sparse texture representation using local affine regions. IEEE Trans. Pattern Anal. Mach. Intell., 27: 1265-1278.

Lazebnik, S., C. Schmid, J. Ponce, 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE.

Lee, H., R. Grosse, R. Ranganath, A.Y. Ng, 2009. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. International Conference on Machine Learning.

Lee, W.T., H.T. Chen, 2009. Histogram-based interest point detectors. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, , 20-25 June 2009. IEEE, pp: 1590-1596.

Lowe, D., 1999. Object Recognition from Local Scale-Invariant Features. Seventh International Conference on Computer Vision,. IEEE.

Lowe, D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vision, 60, 91-110.

Ma, X., W. Grimson, 2008. Learning coupled conditional random field for image decomposition with application on object categorization. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE.

Maddah, M., S. Mozaffari, 2012. Face verification using local binary pattern-unconstrained minimum average correlation energy correlation filters. J. Opt. Soc. Am. A, 29: 1717-1721.

Maji, S., A.C. Berg, J. Malik, 2008. Classification using intersection kernel support vector machines is efficient. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,, 23-28 June 2008. IEEE.

Mei, K., P. Dong, H. Lei, J. Fan, 2014. A distributed approach for large-scale classifier training and image classification. Neurocomputing, 144: 304-317.

Mikolajczyk, K., C. Schmid, 2005. A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell., 27: 1615-1630.

Mohammed, M.M., A. Badr, M.B. Abdelhalim, 2015. Image classification and retrieval using optimized Pulse-Coupled Neural Network. Expert Systems with Applications, 42: 4927-4936.

Moosmann, F., B. Triggs, F. Jurie, 2007. Fast discriminative visual codebooks using randomized clustering forests. Neural Information Processing Systems (NIPS). MIT Press.

Mutch, J., D.G. Lowe, 2008. Object class recognition and localization using sparse features with limited receptive fields. Int. J. Comput. Vision, 80: 45-57.

Nowak, E., F. Jurie, B. Triggs, 2006. Sampling Strategies for bag-of-features image classification. Proceedings of the European Conference on Computer Vision European Computer Vision. ECCV.

Ojala, T., M. Pietik, T. Mäenpää, 2002. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Trans. Pattern Anal. Mach. Intell., 24: 971-987.

Oliva, A., A. Torralba, 2001. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. Int. J. Comput. Vision, 42: 145-175.

Opelt, A., A. Pinz, M. Fussenegger, P. Auer, 2006. Generic object recognition with boosting. IEEE Trans. Pattern Anal. Mach. Intell., 28: 416-431.

Pinto, N., D. Cox, J. Dicarlo, 2008. Why is real-world visual object recognition hard? PLoS computational biology, 4: e27.

Ralló, M., M.S. Millán, J. Escofet, 2009. Unsupervised novelty detection using gabor filters for defect segmentation in textures. J. Opt. Soc. Am. A, 26: 1967-1976.

Rassem, T.H., B.E. Khoo, 2011. New color image histogram-based detectors. Proceedings of the Second international conference on Visual informatics: sustaining research and innovations Springer-Verlag.

Sande, K.V.D., T. Gevers, C. Snoek, 2010. Evaluating color descriptors for object and scene recognition. IEEE Trans. Pattern Anal. Mach. Intell., 32: 1582-1596.

Schulein, R., C.M. Do, B. Javidi, 2010. Distortion-tolerant 3D recognition of underwater objects using neural networks. J. Opt. Soc. Am. A, 27: 461-468.

Sharma, G., S. Chaudhury, J.B. Srivastava, 2008. Bag-of-features kernel eigen spaces for classification. International Conference on Pattern Recognition.

Shechtman, E., M. Irani, 2007. Matching local self-similarities across images and videos. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA. IEEE.

Szeliski, R., 2010. Computer Vision: Algorithms and Applications, Springer.

Szummer, M., R. Picard, 1998. Indoor-outdoor image classification. IEEE International Workshop on Content-Based Access of Image and Video Database. IEEE.

Tuytelaars, T., 2010. Dense interest points. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 13-18 June 2010. IEEE, pp: 2281-2288.

Tuzel, O., F. Porikli, P. Meer, 2007. Human detection via classification on riemannian manifolds. Proceedings of the IEEE Computer Society

Conference on Computer Vision and Pattern Recognition. IEEE.

Vailaya, A., M.A.T. Figueiredo, A.K. Jain, Z. Hong-Jiang, 2001. Image classification for content-based indexing. IEEE Trans. Image Process., 10: 117-130.

Van Gemert, J., J. Geusebroek, C. Veenman, A. Smeulders, 2008. Kernel codebooks for scene categorization. Proceedings of the European Conference on Computer Vision. ECCV.

Van Gemert, J., C. Veenman, A. Smeulders, J. Geusebroek, 2010. Visual word ambiguity. IEEE Trans. Pattern Anal. Mach. Intell., 32: 1271-1283.

Vedaldi, A., V. Gulshan, M. Varma, A. Zisserman, 2009. Multiple kernels for object detection. Twelfth International Conference on Computer Vision. IEEE.

Wang, J., J. Yang, K. Yu, F.L.T. Huang, Y. Gong, 2010. Locality-constrained linear coding for image classification. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE.

Weber, M., M. Welling, P. Perona, 2000. Unsupervised learning of models for recognition. Proceedings of the European Conference on Computer Vision. ECCV.

Xiao, J., J. Hays, K. Ehinger, A. Oliva, A. Torralba, 2010. SUN database: Large-scale scene recognition from abbey to zoo. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 3485-3492.

Ohta, Y., T.K. and T. Sakai, 1980. Color information for region segmentation. Computer Graphics and Image Processing,, 13: 222-241.

Yang, J., Y. Li, Y. Tian, L. Duan, W. Gao, 2009a. Group-sensitive multiple kernel learning for object categorization. Twelfth International Conference on Computer Vision. IEEE.

Yang, J., K. Yu, Y. Gong, T. Huang, 2009b. Linear spatial pyramid matching using sparse coding for image classification. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE.

Zhang, H., A.C. Berg, M. Maire, J. Malik, 2006. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition,. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE.

Zhang, Z., Z.N. Li, M.S. Drew, 2010. Learning image similarities via probabilistic feature matching. International Conference on Image Processing, 26-29 Sept. 2010. IEEE, 1857-1860.