# Object Class Recognition using Combination of Color SIFT Descriptors

Taha H. Rassem
School of Electrical and Electronic Engineering
Universiti Sains Malaysia
Penang, Malaysia
taha.ld08@student.usm.my

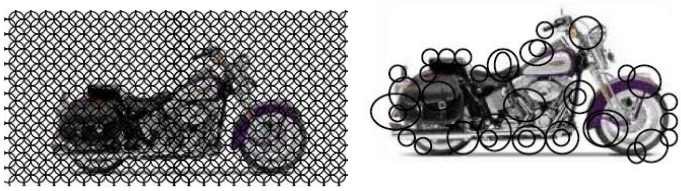
Bee Ee Khoo
School of Electrical and Electronic Engineering
Universiti Sains Malaysia
Penang, Malaysia
BeeKhoo@eng.usm.my

***Abstract*—Classifying the unknown image into the correct related class is the aim of the object class recognition systems. Two main points should be kept in mind to implement a class recognition system. Which descriptors that have a higher discriminative power that needs to be extracted from the images? Which classifier can classify these descriptors successfully? The most famous image descriptor is the Scale Invariant Feature Transform (SIFT). Although, SIFT has a high performance, it is partially an illumination invariant. Adding local color information to SIFT descriptors are then suggested to increase the illumination invariant, these descriptors can be called color SIFT descriptors. In this paper, different color SIFT descriptors were implemented to evaluate their performance in the object class recognition systems. This is due to the fact that some descriptors may have a good performance in one class and bad performance in another class at the same time. All possible combinations of these descriptors were used. Some combinations of color SIFT descriptors achieved remarkable classification accuracy. Non linear $\chi^2$-kernel support vector machine is used as a learning classifier and bag-of-features representation is used to represent the image features in this paper.**

## I. Introduction

Object class recognition or category recognition is one of the open problems in the computer vision field. Although, great works had been done but still a lot of challenges need to be solved [28]. There are two levels of image contents; scene contents such as outdoor, indoor, kitchen, etc and object contents such as cars, airplanes, motorbikes, etc. Analyzing the image and classifying this image based on its contents into the correct related class is the aim of these systems.

The accuracy of classification is affected inversely based on the number of classes. This is due to some points such as intra-class variation, inter-class variation, illumination invariant, scale invariant, rotation, viewpoint and, etc. [4]. There are three main questions in each classification system; Which points or regions have to be detected?, Which discriminative descriptors have to be extracted from these detected points or regions?, and finally, which powerful learning algorithm has the capability to differentiate between the extracted descriptors. In detector stage, different methods had been used to select an interest point such as edge detection operator, Harris-Laplace operator, Difference-of-Gaussian, etc. This can be called sparse detector at an interest point [13], [18]. Another direction is to select all points in the image uniformly. This can be called dense detectors [5], [10]. Dense and sparse detectors are shown in Fig. 1. Recently, combined multiple detectors had been used together in one system [5], [9], [10], [25]. In descriptor extraction stage, the descriptors that must be extracted either from the interest points or all points in the image should have special characteristics. These descriptors have to be robust and should be rotation invariant, scale invariant, viewpoint invariant, illumination invariant and, etc. Scale Invariant Feature Transform(SIFT) is a famous descriptor in computer vision literature [20], [21]. Although, it has a good performance, it is partially invariant to illumination changes. Adding local color information to the SIFT descriptor is suggested to increase the illumination invariant. [5], [6], [30]. Different color SIFT descriptors are implemented such as HSV-SIFT, RGB-SIFT, Opponent SIFT, C-SIFT, rgSIFT, Transformed-color SIFT and, etc. Van Sande et al. [30] studied their properties and proved these properties theoretically and then evaluated some of the color descriptors using various types of data sets. In this paper, some color SIFT descriptors were densely extracted and evaluated their effect each one alone and using all possible combinations of them. Combined different types of descriptors have a better chance to enhance the classification accuracy. This is because each class may have a good response to one descriptor than other. RGB-SIFT, Opponent SIFT, Transformed-color SIFT, and HSV-SIFT were used and evaluated using four types of object data sets, which are Caltech-101(5 classes), Caltech04 , Graz02 and Caltech101 This paper is organized as follows. First, color SIFT descriptors are described in section 2. Then, in section 3, we briefly outline the work implementation. The experimental setup and results are presented in section 4. Finally, section 5



(a) (b)

Fig. 1. Image detectors types. 1(a) Dense detector. 1(b) Sparse detector.

978-1-61284-896-9/11/$26.00 ©2011 IEEE

concludes this paper.

## II. Color SIFT descriptors

The SIFT descriptor that is proposed by Lowe in 1999 is still a hallmark in the recognition systems because of its good characteristics [20], [21]. Based on SIFT, another extended descriptors are suggested such as Principle Component Analysis SIFT (PCA-SIFT) [22], Gradient Location-Orientation Histogram (GLOH) [16], Rotation Invariant Feature Transform(RIFT) [17], Speeded Up Robust Features (SURF) [1] and Fast SIFT(FA-SIFT) [11]. These descriptors are introduced to improve SIFT performance or reduce its complexity [22].

However, no local color information is used with them. These descriptors can be called intensity-based descriptors [30]. Another extension had been done by adding local color information to the SIFT. These descriptors can be called color-based descriptors and help to improve the illumination robustness and to make the SIFT more discriminative. In [6], the SIFT descriptors are extracted over all channels of the HSV color model. This is called HSV-SIFT. Many color descriptors are exists such as C-SIFT, rgSIFT, HueSIFT, RGB-SIFT, Opponent SIFT, HSV-SIFT and Transformed-color SIFT [2], [30]. Last four descriptors were used in this paper based on their characteristics. In the following subsections, we explain them briefly.

### A. RGB-SIFT

RGB-SIFT descriptor is obtained by computing the SIFT descriptor over all three channels of the RGB color space channels independently [30]. It is 3×128 dimension vector.

### B. HSV-SIFT

HSV-SIFT descriptor is obtained by computing the SIFT descriptor over all three channels of the HSV color space channels independently It is 3×128 dimension vector. HSV-SIFT is mostly used than other descriptors [4], [10].

### C. Opponent SIFT

Opponent-SIFT descriptor is obtained by computing the SIFT descriptor over all three channels of the Opponent color space channels independently. It is 3×128 Dimension vector. The Opponent color space channels can be described by the following equations.

$$O_1 = \frac{R-G}{\sqrt{2}} \tag{1}$$

$$O_2 = \frac{R+G-2B}{\sqrt{6}} \tag{2}$$

$$O_3 = \frac{R+G+B}{\sqrt{3}} \tag{3}$$

$O_1$ and $O_2$ represents the color information while $O_3$ represents the intensity information [13], [14]

### D. Transformed-color SIFT

The Transformed color SIFT channels can be obtained according to the following equations.

$$\acute{R} = \frac{R-\mu_R}{\sigma_R} \tag{4}$$

$$\acute{G} = \frac{G-\mu_G}{\sigma_G} \tag{5}$$

$$\acute{B} = \frac{B-\mu_B}{\sigma_B} \tag{6}$$

Where $\mu_R$ , $\mu_G$ and $\mu_B$ are the mean of R ,G and B channels respectively and $\sigma_R$,$\sigma_G$ and $\sigma_B$ are the standard deviation of each channel. The SIFT descriptors are extracted over these channels.

## III. Implementation

As shown in Fig 2, the color descriptors were extracted densely from the images, and then the training models were learned using non linear $\chi^2$-kernel SVM. Bag-of-features representation was used to represent the extracted descriptors. VLFEAT open library was used to do that [33]. Using a set of available images, the k-means clustering was used to build the visual code book. 600 words vocabulary was used for caltech-101 while 300 word vocabulary for other data sets.

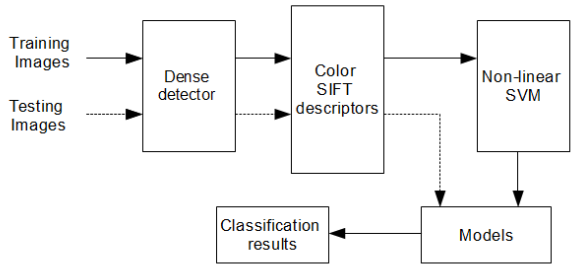



Fig. 2. System Structure

## IV. Experiments and Results

In this section, all details of the experiments setup for each data set and its results are provided. In each section, we briefly explain some pervious works that used the same data set and then compare with their results. As mentioned before, Caltech-101(5 classes), Caltech04, GRAZ02 and Caltech-101 were used in our experiments.

### A. Experiments using Caltech-101(5 classes)

Caltech-101(5 classes) is a subset of Caltech-101 data set. It has altogether 2624 images, including 435 faces, 800 airplanes, 123 car_side, 798 motorbike and 468 background_google. These images present relatively little clutter and variation in object pose [2]. Example images of Caltech-101(5 classes) data set are shown in Fig 3. Behmo et al. [2] 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. Our experiments had been implemented using the same setup (30 images for training and 30 images for testing). Table I shows our results while Table II shows their results [2]. Using the color SIFT descriptors, they achieved better results using Opponent SIFT ( 91.10±2.45 ) and Transformed-color SIFT ( 90.01± 3.03). Our classification accuracies are ranging between 94% (RGB-SIFT+Transformed-color SIFT) to 96.67% (HSV-SIFT with Opponent SIFT and RGB-SIFT with Opponent SIFT with HSV-SIFT) as shown in Table I. To discuss the accuracy results per class, the Opponent SIFT with HSV-SIFT, and Opponent SIFT alone are succeeded to classify four classes 100% correctly. Generally, our results are better than Behmo et al. results [2].

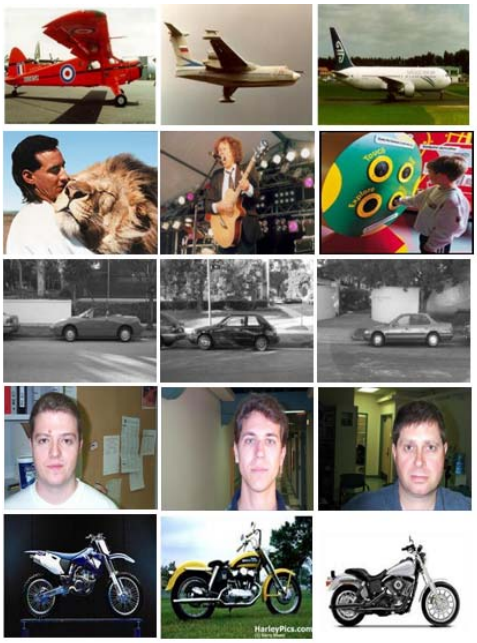


Fig. 3. Caltech101(5 classes) data set example images. Each row shows example images of one class. From top to button: airplanes, google_background, car_side, faces and motorbikes

TABLE I
CALTECH-101(5 CLASSES) RESULTS

| Descriptors | Faces | Motor bikes | Air planes | Cars side | Backg round google | Average |
|---|---|---|---|---|---|---|
| RGB-SIFT | 96.67 | 100 | 100 | 100 | 83.33 | 96 |
| Opp.SIFT | 100 | 100 | 100 | 100 | 80 | 96 |
| HSV-SIFT | 96.67 | 100 | 100 | 100 | 83.33 | 96 |
| Transf.clrSIFT | 96.67 | 100 | 96.67 | 96.67 | 83.33 | 94.67 |
| RGB-SIFT+Opp.SIFT | 96.67 | 100 | 100 | 100 | 80 | 95.33 |
| RGB-SIFT+Transf.clrSIFT | 93.33 | 100 | 100 | 96.67 | 80 | 94 |
| RGB-SIFT+HSV-SIFT | 96.67 | 100 | 100 | 100 | 80 | 95.33 |
| HSV-SIFT+Transf.clrSIFT | 96.67 | 100 | 100 | 100 | 80 | 95.33 |
| HSV-SIFT+Opp.SIFT | 100 | 100 | 100 | 100 | 83.33 | **96.67** |
| Transf.clrSIFT+Opp.SIFT | 96.67 | 100 | 100 | 96.67 | 80 | 94.67 |
| RGB-SIFT+HSV-SIFT+Opp.SIFT | 96.67 | 100 | 100 | 100 | 80 | 95.33 |
| RGB-SIFT+HSV-SIFT+Transf.clrSIFT | 96.67 | 100 | 100 | 100 | 80 | 95.33 |
| HSV-SIFT+Opp.SIFT+Transf.clrSIFT | 96.67 | 100 | 100 | 100 | 83.33 | 96 |
| RGB-SIFT+Opp.SIFT+Transf.clrSIFT | 96.67 | 100 | 100 | 100 | 86.67 | **96.67** |
| HSV-SIFT+Opp.SIFT+Transf.clrSIFT+RGB-SIFT | 96.67 | 100 | 100 | 100 | 80 | 95.33 |

TABLE II
CALTECH-101(5 CLASSES) RESULTS [2]

| class | Results |
|---|---|
| Airplanes | 95.00 ± 3.25 |
| Car-side | 94.00 ± 4.29 |
| Faces | 89.00 ± 7.16 |
| Motorbikes | 91.00 ± 5.69 |
| Background_google | 79.83 ± 10.67 |

### B. Experiments using Caltech04

Caltech04 database has altogether 5775 images, including 1074 airplanes, 1155 cars, 450 faces, 826 motorbikes, 1370 car backgrounds and 900 general backgrounds. 100 images are used for training and 50 images are used for testing for each class. Two parts of experiments were done using Caltech04. In the first part, the Caltech background images were used in the experiments as a counter class. In the latter part, GRAZ01 is used instead of Caltech background as Hegazy et al. suggested [13]. The results are shown in Table III and Table IV. For comparison, the results of some state-of the art systems are shown in Table V. Opelt et al. [25] used three types of detectors with four types of descriptors and classified Caltech-04 using adaboost. Hegazy et al. [13] used Hessian-affine interest point detector to extract SIFT and histogram of Opponent color for each image. They used adaboost for classification. Han et al. [12] used BOW to represent SIFT descriptor after using DOG detector. They model the visual word using supervised nonlinear neighborhood embedding space to be more discriminative and then used Probabilistic Neural Network for classification.

TABLE III
CALTECH04 RESULTS (CALTECH BACKGROUND)

| Descriptors | Faces | Motorbikes | Airplanes | Cars | Average |
|---|---|---|---|---|---|
| RGB-SIFT | 92 | 98 | 96 | 100 | 96.5 |
| Opp.SIFT | 100 | 96 | 98 | 100 | **98.5** |
| HSV-SIFT | 94 | 92 | 100 | 100 | 96.5 |
| Transf.clrSIFT | 96 | 98 | 98 | 98 | 97.5 |
| RGB-SIFT+Opp.SFIT | 98 | 92 | 98 | 100 | 97 |
| RGB-SIFT+Transf.clrSIFT | 94 | 98 | 98 | 98 | 97 |
| RGB-SIFT+HSV-SIFT | 92 | 92 | 98 | 100 | 95.5 |
| HSV-SIFT+Transf.clrSIFT | 92 | 96 | 98 | 98 | 96 |
| HSV-SIFT+Opp.SIFT | 98 | 96 | 98 | 100 | 98 |
| Transf.clrSFIT+Opp.SIFT | 98 | 98 | 98 | 100 | **98.5** |
| RGB-SIFT+HSV-SIFT+Opp.SIFT | 96 | 92 | 98 | 100 | 96.5 |
| RGB-SIFT+HSV-SIFT+Transf.clrSIFT | 94 | 98 | 98 | 98 | 97 |
| HSV-SIFT+Opp.SIFT+Transf.clrSIFT | 100 | 94 | 98 | 100 | 98 |
| RGB-SIFT+Opp.SIFT+Transf.clrSIFT | 98 | 96 | 98 | 100 | 98 |
| HSV-SIFT+Opp.SIFT+Transf.clrSIFT+RGB-SIFT | 96 | 94 | 98 | 100 | 97 |

As shown in Table III, the results are ranging from 95.5% using RGB-SIFT with HSV-SIFT descriptors to 98.5% using Opponent SIFT or using Transformed-color SIFT with Opponent SIFT. Opponent SIFT and HSV-SIFT and combination of HSV-SIFT with Opponent SIFT and Transformed-color are succeeded to classify two classes 100% correctly. Notable here, only the HSV-SIFT succeeded to classify all airplane images to the Airplane's class.

Hegazy et al. suggested to replace Caltech background images by GRAZ01 background images [13], [14]. This suggestion based on the GRAZ01 background images are colored while the other is not colored and the GRAZ01

images are more difficult. Using GRAZ01 background is better for us because we used color descriptors. Our results as shown in Table IV shows better results than Hegazy et al. results [13], [14]. The results are ranging from 97.5% to 99.5%. Also, Opponent SIFT achieved a better result as a single descriptor while as combined descriptors, HSV-SIFT with Transformed-color SIFT is the better. Opponent SIFT , Transformed-color SIFT and combined of HSV-SIFT with Transformed-color SIFT succeeded to classify the images to three classes 100% correctly. Out of the memory problem was faced when tried to combine all descriptors in the second part because the GRAZ01 background images size are big unlike Caltech background.

TABLE IV
CALTECH-04 RESULTS USING GRAZ01 BACKGROUND LIKE HEGAZY ET AL. [13], [14] INSTEAD OF CALTECH BACKGROUND

| Descriptors | Faces | Motorbikes | Airplanes | Cars | Average |
|---|---|---|---|---|---|
| RGB-SIFT | 100 | 98 | 98 | 98 | 98.5 |
| Opp.SIFT | 100 | 98 | 100 | 100 | **99.5** |
| HSV-SIFT | 98 | 92 | 100 | 100 | 97.5 |
| Tranf.clrSIFT | 100 | 96 | 100 | 100 | 99 |
| RGB-SIFT+Opp.SIFT | 100 | 98 | 98 | 100 | 99 |
| RGB-SIFT+Tranf.clrSIFT | 100 | 98 | 96 | 100 | 98.5 |
| RGB-SIFT+HSV-SIFT | 96 | 98 | 96 | 100 | 97.5 |
| HSV-SIFT+Tranf.clrSIFT | 100 | 100 | 98 | 100 | **99.5** |
| HSV-SIFT+Opp.SIFT | 96 | 98 | 100 | 100 | 98.5 |
| Transf.clrSIFT+Opp.SIFT | 100 | 98 | 98 | 100 | 99 |
| RGB-SIFT+HSV-SIFT+Opp.SIFT | 96 | 98 | 98 | 100 | 98 |
| RGB-SIFT+HSV-SIFT+Tranf.clrSIFT | 98 | 98 | 98 | 100 | 98.5 |
| HSV-SIFT+Opp.SIFT+Tranf.clrSIFT | 98 | 98 | 100 | 100 | 99 |
| RGB-SIFT+Opp.SIFT+Tranf.clrSIFT | 100 | 98 | 98 | 100 | 99 |

TABLE V
CALTECH-04 STATE-OF-THE-ART OF RESULTS

| Reference | Faces | Motorbikes | Airplanes | Cars |
|---|---|---|---|---|
| Han et al. [12] | 93.53 | 94.41 | 96.29 | 94.35 |
| Oplet et al. [25] | 93.5 | 92.2 | 88.9 | 91.1 |
| Hegazy et al. [13] | 94 | 96 | 84 | 100 |
| Hegazy et al. [14] | 100 | 93 | 84 | 94 |
| Fergus et al. [7] | 96.4 | 92.5 | 90.2 | 88.5 |
| Thureson et al. [29] | 83.1 | 93.2 | 83.8 | 90.2 |
| Weber et al. [34] | 93.5 | 88 | - | 86.5 |

### C. Experiments using GRAZ02

The Graz02 data set is more difficult than the Caltech data set. The objects are shown on a complex cluttered background, at different scales, and with different object positions [25]. GRAZ02 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. Example images of GRAZ02 data set can be shown in Fig 4. 150 images were used for training and 75 for testing for each class after resized them to 320*480 pixels. The results are shown in Table VI. For comparison, the results of some state-of the art systems are shown in Table VII.

As shown in Table VI, Opponent SIFT is still the better single descriptor and combination of HSV-SIFT with Transformed-color SIFT and HSV-SIFT with Opponent SIFT results reaches 80% and 80.44% respectively. The highest result is achieved using the combination of all descriptors, it reaches 81.67%. To discuss the results per class,97.33% accuracy is achieved using RGB-SIFT and RGB-SIFT with HSV-SIFT for the bike's class. Table VII shows some of state-of-the art results. Although, some of the person's class accuracy reaches 86.67% as a best result. Zhang et al. [35] achieved

Fig. 4. GRAZ02 data set example images. Each row shows example images of one class. From top to button: persons, cars, bikes and counter-class(background).

TABLE VI
GRAZ-02 RESULTS

| Descriptors | Bikes | Cars | Persons | Average |
|---|---|---|---|---|
| RGB-SIFT | 97.33 | 49.33 | 78.67 | 75.11 |
| Opp.SIFT | 93.33 | 57.33 | 86.67 | 79.11 |
| HSV-SIFT | 96 | 44 | 78.67 | 72.89 |
| Tranf.clrSFIT | 89.33 | 61.33 | 85.33 | 78.66 |
| RGB-SIFT+Opp.SIFT | 96 | 52 | 78.67 | 75.56 |
| RGB-SIFT+Tranf.clrSFIT | 92 | 61 | 77.33 | 76.78 |
| RGB-SIFT+HSV-SIFT | 97.33 | 60 | 78.67 | 78.67 |
| HSV-SIFT+Tranf.clrSFIT | 94.67 | 62.67 | 82.67 | **80.00** |
| HSV-SIFT+Opp.SIFT | 93.33 | 62.67 | 85.33 | **80.44** |
| Tranf.clrSIFT+Opp.SIFT | 90.67 | 50.67 | 85.33 | 75.56 |
| RGB-SIFT+HSV-SIFT+Opp.SIFT | 93.33 | 52 | 85.33 | 76.89 |
| RGB-SIFT+HSV-SIFT+Tranf.clrSIFT | 92 | 49.33 | 85.33 | 75.55 |
| HSV-SIFT+Opp.SIFT+Tranf.clrSIFT | 93.33 | 54.67 | 84 | 77.33 |
| RGB-SIFT+Opp.SIFT+Tranf.clrSIFT | 86.67 | 49.33 | 84 | 73.33 |
| HSV-SIFT+Opp.SIFT+Tranf.clrSIFT+RGB-SIFT | 96 | 64 | 85 | **81.67** |

TABLE VII
GRAZ-02 STATE-OF-THE-ART OF RESULTS

| Reference | Bikes | Cars | Persons |
|---|---|---|---|
| Behmo et al. [2] | 78.70±4.67 | 82.05±4.88 | 76.20±5.85 |
| Hegazy et al. [13] | 74.67 | 81.33 | 81.33 |
| Hegazy et al. [14] | 80 | 78.62 | 84 |
| Zhang et al. [36] | 88.9 | 85.2 | 88.1 |
| Mutch et al. [24] | 80.5 | 70.1 | 81.7 |
| Oplet et al. [25] | 77.8 | 70.5 | 81.2 |
| Moosmann et al. [23] | 84.4 | 79.9 | - |

88.1%. Unfortunately, the opposite of what we expected, the car's class results are not comparable. 61.33% is the best of our results while 85.2% were obtained by [35].

### *D. Experiments using Caltech-101*

Caltech-101 database has altogether 9146 images, in total of 102 classes; 101 distinct object class such as faces, watches, ant and, etc. and background class. Each image is about 300x200 pixels in dimension. Same experiment setup was used such as Gheler et al. [9], [10], 30 and 15 images for training only 15 images for testing for each class. The Caltech-101 results and some of the state-of-the art results are shown in Table VIII,

TABLE VIII
CALTECH-101 RESULTS

| descriptors | 30 | 15 |
|---|---|---|
| RGB-SIFT | 69.05 | **63.14** |
| Opp.SIFT | 58.72 | 50 |
| HSV-SIFT | 63.40 | 56.47 |
| Tranf.clrSIFT | **69.11** | 61.47 |
| RGB-SIFT+Opp.SIFT | 68.36 | 61.24 |
| RGB-SIFT+Tranf.clrSIFT | 68.42 | **63.4** |
| RGB-SIFT+HSV-SIFT | 69.28 | 61.63 |
| HSV-SIFT+Tranf.clrSIFT | **70.30** | 62.16 |
| HSV-SIFT+Opp.SIFT | 65.23 | 57.32 |
| Tranf.clrSIFT+Opp.SIFT | 68.28 | 61.96 |
| RGB-SIFT+HSV-SIFT+Opp.SIFT | 67.80 | 60.78 |
| RGB-SIFT+HSV-SIFT+Tranf.clrSIFT | **69.39** | **63.53** |
| HSV-SIFT+Opp.SIFT+Tranf.clrSIFT | 69.25 | 61.63 |
| RGB-SIFT+Opp.SIFT+Tranf.clrSIFT | 69.20 | 62.81 |
| HSV-SIFT+Opp.SIFT+Tranf.clrSIFT+RGB-SIFT | **69.92** | **63.59** |

TABLE IX
CALTECH-101 SOME OF THE STATE-OF-THE ART RESULTS

| References | 30 | 15 |
|---|---|---|
| Jain et al. [15] | 69.6 | - |
| Pinto et al. [26] | 67.36 | 61.44 |
| Zhang et al. [35] | 66.23 | 59.00 |
| Frome et al. [8] | 66 | 60.30 |
| Sharma et al. [27] | 65.50 | - |
| Lee et al. [19] | 65.40 | - |
| Lazebnik et al. [18] | 64.60 | 56.40 |
| Van Gemert et al. [31] | 64.12 | - |
| Van Gemert et al. [32] | 64.10 | - |

Using 30 training images, the accuracy results are ranging from 58.72% using Opponent SIFT descriptor to 70.30% using HSV-SIFT with Transformed color SIFT descriptors. Using one descriptor only, Transformed color SIFT descriptor has achieved better result reaches to 69.11% followed by RGB-SIFT 69.05%. Notable that the combination of all descriptors has achieved the second best result reaches to 69.92%.

Using 15 training images, RGB-SIFT descriptor is the good single descriptor but the combination of all descriptors achieved a highest accuracy reaches 63.59%. Our results as mentioned in Table VIII are comparable with some of the state-of-the art results that is shown in Table IX. The best state-of-the art result is obtained by Gehler et al. [9], [10]. They achieved up to 82.1%. Comparing with their results, they used several types of descriptors and more than 48 kernels in the experiments. In our experiments, only appearance descriptor with local color information is used.

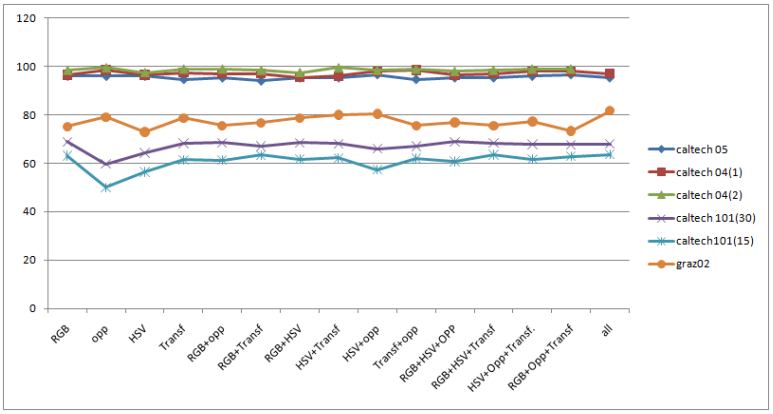



Fig. 5. The average results of all descriptors for all data sets

## V. CONCLUSION

Color SIFT descriptors as an extension of SIFT appearance descriptor assisted to increase the classification results of the object recognition system. Opponent SIFT appears having a higher performance in small data sets such as Caltech-101(5 classes), Caltech04 and GRAZ02 as a single color descriptor. Furthermore, combined Opponent SIFT with Transformed color SIFT and combined Opponent SIFT with HSV-SIFT achieved good results in Caltech-04 and Caltech-101(5 classes) respectively. In Caltech-101, the worst performance is obtained using Opponent SIFT (for 30 and 15 training images) while the best performance of all is obtained using HSV-SIFT and Transformed color SIFT combination for 30 training images and using all descriptors for 15 training images. Figure 5 summarized the effect of all descriptors for all data sets based on classification results. Generally speaking, densely color SIFT descriptors can give a good performance in the object recognition system with small data sets that have little clutter and variation in object pose. In addition, an acceptable performance with another data set that has complex clutter, different object position and scales, and large number of classes. The latter data sets such as Caltech 101, GRAZ02 and others that are mentioned in the literature need to represented with different types of descriptors such as appearance descriptors (SIFT ,color SIFT and, etc.) , shape descriptors such as PHOG and, etc., texture descriptors such as LBP, and, etc. Furthermore, extract the descriptors densely from the images that have a complex background and clutter is not enough. Interest points have to be detected using a powerful detector in addition to the dense detector. Good examples of these systems are Gehler et al. system [10], Boiman et al. system [3], Bosch et al. [6] and, etc. Combined some of these descriptors with all types of color SIFT descriptors with different types of detectors will be our future work.

## ACKNOWLEDGMENT

This work is supported by USM graduate assistant fellowship.

## REFERENCES


[1] H. Bay, T. Tuytelaars, and L. J. V. Gool, "Surf: Speeded up robust features." in *ECCV (1)*, ser. Lecture Notes in Computer Science, A. Leonardis, H. Bischof, and A. Pinz, Eds., vol. 3951. Springer, 2006, pp. 404–417.

[2] R. Behmo, P. Marcombes, A. Dalalyan, and V. Prinet, "Towards optimal naive bayes nearest neighbor," in *European Conference on Computer Vision (ECCV 2010)*, 2010, pp. IV: 171–184.

[3] O. Boiman, E. Shechtman, and M. Irani, "In defense of nearest-neighbor based image classification," in *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008.*, 2008, pp. 1 –8.

[4] A. Bosch, "Image classification for a largre number of object categories," PhD Thesis, Departament d'Electrnica, Informtica i Automtica. Universitat de Girona, 2007,.

[5] A. Bosch, A. Zisserman, and X. M., "Representing shape with a spatial pyramid kernel," in *Proceedings of the 6th ACM international conference on Image and video retrieval*, ser. CIVR '07, 2007, pp. 401–408.

[6] A. Bosch, A. Zisserman, and X. Muoz, "Image classification using random forests and ferns," in *IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007.*, 2007, pp. 1 –8.

[7] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. 2003*, vol. 2, 2003, pp. II–264 – II–271 vol.2.

[8] A. Frome, Y. Singer, and J. Malik, "Image retrieval and classification using local distance functions." in *NIPS*, B. Schlkopf, J. Platt, and T. Hoffman, Eds. MIT Press, 2006, pp. 417–424.

[9] P. Gehler and S. N., "Extension work - on feature combination for multiclass object detection- 2009."

[10] P. Gehler and S. Nowozin, "On feature combination for multiclass object classification," in *IEEE 12th International Conference on Computer Vision, 2009*, 2009, pp. 221 –228.

[11] M. Grabner, H. Grabner, and H. Bischof, "Fast approximated sift." in *ACCV (1)*, ser. Lecture Notes in Computer Science, P. J. Narayanan, S. K. Nayar, and H.-Y. Shum, Eds., vol. 3851. Springer, 2006, pp. 918–927.

[12] X.-h. Han, Y.-W. Chen, and X. Ruan, "Object class recognition with supervised nonlinear neighborhood embedding of visual words," in *Proceedings of the First International Conference on Internet Multimedia Computing and Service*, ser. ICIMCS '09, 2009, pp. 25–28.

[13] D. Hegazy and J. Denzler, "Boosting colored local features for generic object recognition," *Pattern Recognition and Image Analysis*, vol. 18, pp. 323–327.

[14] ——, "Generic object recognition using boosted combined features," in *Proceedings of the 2nd international conference on Robot vision*, ser. RobVis'08, 2008, pp. 355–366.

[15] P. Jain, B. Kulis, and K. Grauman, "Fast image search for learned metrics," in *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008.*, 2008, pp. 1 –8.

[16] Y. Ke and R. Sukthankar, "Pca-sift: A more distinctive representation for local image descriptors." in *CVPR (2)*, 2004, pp. 506–513.

[17] S. Lazebnik, C. Schmid, and J. P., "Semi-local affine parts for object recognition," in *British Machine Vision Conference*, vol. volume 2, 2004, pp. 779–788.

[18] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006*, 2006.

[19] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09, 2009, pp. 609–616.

[20] D. Lowe, "Distinctive image features from scale-invariant key-points," *Intl. Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[21] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the International Conference on Computer Vision*, vol. 2, 1999, pp. 1150–1157.

[22] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[23] F. Moosmann, B. Triggs, and F. Jurie, "Fast discriminative visual codebooks using randomized clustering forests," in *Neural Information Processing Systems (NIPS)*, nov 2006.

[24] J. Mutch and D. G. Lowe, "Object class recognition and localization using sparse features with limited receptive fields," *Int. J. Comput. Vision*, vol. 80, pp. 45–57, October 2008.

[25] A. Opelt, A. Pinz, M. Fussenegger, and P. Auer, "Generic object recognition with boosting," *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, vol. 28, no. 3, pp. 416 –431, 2006.

[26] N. Pinto, D. D. Cox, and J. J. DiCarlo, "Why is real-world visual object recognition hard?" *PLoS Comput Biol*, vol. 4, 01 2008.

[27] G. Sharma, S. Chaudhury, and J. Srivastava, "Bag-of-features kernel eigen spaces for classification," in *19th International Conference on Pattern Recognition, 2008. ICPR 2008.*, 2008, pp. 1 –4.

[28] R. Szeliski, *Computer Vision: Algorithms and Applications*, 1st ed. Springer, 2010.

[29] J. Thuresson and S. Carlsson, "Appearance based qualitative image description for object class recognition," 2004, pp. Vol II: 518–529.

[30] K. van de Sande, T. Gevers, and C. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, vol. 32, no. 9, pp. 1582 –1596, 2010.

[31] J. van Gemert, J.-M. Geusebroek, C. J. Veenman, and A. W. Smeulders, "Kernel codebooks for scene categorization," in *Proceedings of the 10th European Conference on Computer Vision: Part III*, 2008, pp. 696–709.

[32] J. van Gemert, C. Veenman, A. Smeulders, and J.-M. Geusebroek, "Visual word ambiguity," *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, vol. 32, no. 7, pp. 1271 –1283, 2010.

[33] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," 2008.

[34] M. Weber, M. Welling, and P. Perona, "Unsupervised learning of models for recognition," in *Proceedings of the 6th European Conference on Computer Vision-Part I*, ser. ECCV '00, 2000, pp. 18–32.

[35] H. Zhang, A. Berg, M. Maire, and J. Malik, "Svm-knn: Discriminative nearest neighbor classification for visual category recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006*, 2006.

[36] Z. Zhang, Z.-N. Li, and M. S. Drew, "Learning image similarities via probabilistic feature matching," in *17th IEEE International Conference on Image Processing (ICIP), 2010*, 2010, pp. 1857 –1860.