LABEL PREDICTION THROUGH MULTIPLE VISUAL FEATURES

D.S.NARKHEDE[1], J.R.MANKAR[2]

[1] PG Student, Department of Computer Engineering, KKWIEER, Nashik, Maharashtra, India
[2] Assistance Prof, Department of Computer Engineering, KKWIEER, Nashik, Maharashtra, India

ABSTRACT

Multiple visual features are represented by multimedia data. Multi-feature learning aims at using the complementary structural information of visual features. The focus is on the semi-supervised learning when the label information of the training data is insufficient. Most of the existing systems face the problem of insufficient labelled data that are expensive to label by hand in real-world application. To address this problem classifier has been already proposed in the literature that select features closely similar to the query image and based on these features label prediction is done. This work aims at studying different low level feature descriptor for better feature extraction and to improve label prediction accuracy of the system by replacing Scale Invariant Feature Transform (SIFT) descriptor with Oriented FAST and rotated BRIEF (ORB) descriptor.

Keyword:- *Multi-feature learning, Multimedia understanding, Semi-supervised learning, Visual recognition, Feature extraction*

1. INTRODUCTION

Multimedia contents and images are ordinarily used to represents multiple features, multiple modalities and multiple views. For example, given a flower image, its visual contents can be constituted with some kinds of modalities such as color, shape, texture and type of flowers[2]; given video data for video concept annotation a video frame, its visual concepts can be represented by different types of low-level feature descriptors such as SIFT, HSV, HOG, etc.[3]. With multiple visual feature representation, finding how to develop the prosperous structural information about each feature in modelling is a challenging task in multimedia analysis.

At the early stage, there are three levels of information fusion: Feature level, Score level and decision level. Feature level was created feature sets from multiple feature extraction algorithms are combined into a single feature set by performing appropriate feature normalization, transformation and reduction strategy so that can improve recognition accuracy. Score level, the match scores output by multiple features is combined to produce a raw output that can be later utilized for decision-making. Fusion at score level is the most commonly quite popular approach primarily suitable to the ease of accessing and processing match scores compare to the raw data or the feature set extracted from the available data. "AND" and "OR" rule take into consideration of decision level fusions so that feature level fusion is more essential for recognition than decision level and score level fusion. Feature concatenation is diagnosed as a generic fusion approach in pattern recognition. However, it is much less useful in the multimedia content estimation because of the truth that the visual features are often independent or heterogeneous. Specifically, easy feature concatenation for high dimensional feature vectors may additionally end up inefficient and hard. One of those limitations, multi-view learning concept has been developed.

Recently much more attention has increased in the problem of learning both labelled and unlabeled images data given by predictive model so this modified version of learning problem usually is to be a semi-supervised learning , arrive in many real world applications whereas both supervised and unsupervised learning learner address to get predictive model from labelled images data while learner remove a descriptive model from unlabeled images data respectively. Most of the existing systems face the problem of insufficient labelled images data that are hard to label in real-time application. There are several methods for image classification[1][4][7][9] and visual recognition[3][5] some are based on supervised and unsupervised learning. But supervised learning have training examples with labels and unsupervised learning does not require labelled images data which is very difficult and time consuming task. In semi-supervised learning having both labelled and unlabeled images data and it is challenging task to assign label. To improve the system performance focus is on different types of feature descriptor for better feature extraction process is point of research.

Further, the idea of multi-modal joint learning is well concerned in dictionary learning and sparse representation. A number of representative works below the framework of dictionary learning were proposed for visual reputation, which include face, digit, motion, and object recognition.

2. RELATED WORK

Some methods have been developed for visual recognition, including face recognition, gender recognition, age estimation, scene categories and object recognition in computer vision community. The bag-of-features (BoF) model has been a popular image categorization, except it rejects the spatial order of local descriptors which restrictions the descriptive power of the image representation so to overcome these drawbacks, S. Lazebnik, C. Schmid, and J. Ponce [3], spatial pyramid matching (SPM) proposed in that pyramid is formed into the image space and computed features for natural scene and object recognition.

Yang et al. [4] Also projected a linear SPM uses sparse coding, spatial pooling & linear spatial pyramid matching. The Idea behind uses sparse coding for soft vector quantization, so hard and soft vector quantization problem can be solved by using Feature-Sign Search Algorithm. The goal of spatial pooling is to represent every image well manage in terms of codeword also use the histogram as for SVM classifier, but result in slow computation speed so using max-pooled features linear kernel does not work well with histogram but gives greater performance for max-pooled histogram.

Gehler et al describes a number of feature combination methods which including average kernel support vector machine (AK-SVM), product kernel support vector machine (PK-SVM), multiple kernel learning (MKL) focus on feature selection while combining features first computing average over all kernels in that distance matrices is given and goal computes one single kernel uses for SVMs but there is an ordinary fault of these methods that the computational cost is also large.

Zhang et al. [6] projected a multi-observation joint dynamic thin illustration for visual recognition, and acquire comparable performance of these works demonstrate that multi-feature joint learning incorporates a positive impact on classifier learning for visual understanding.

Semi-supervised learning has been wide deployed within the recognition task, because of these truth that training some amount of labelled information is liable to overfitting, whereas manual labeling of an outsized quantity of exactly labelled knowledge is tedious and long. In this work we concentrate on semi- supervised classification. Usually classifiers apply just labeled data (feature / label pairs) to train. Labeled instances, however are normally difficult, costly, or tedious to acquire, as they require the endeavors of experienced human annotators. Indicate while unlabeled data can be relatively easy to collect, except there has been a small number of ways to use them. Semi-supervised learning address this problem with large amount of unlabeled data, together with the labeled data, to construct better classifiers so that require less human effort and gives better accuracy.

In Laplacian graph manifold based semi-supervised learning framework Belkin et al[7] used the manifold structure of information on the unlabeled data for manifold assumption also consider assumption of consistency is given the same label when data points are closely similar or in the same cluster or manifold here local consistency refer cluster while global consistency refer manifold.

Zhou et al [8] proposed local and global consistency with graph regularization for graph based semi-supervised method.

Laplacian graph and the l2-norm regularization are used in semi-supervised feature selection algorithm (SFSS) for multimedia analysis also Laplacian graph having single view is the main method for semi-supervised learning, but it is constant with weak-extrapolating power while hessian graph has good extrapolating power in the manifold regularization.

Wang et al [10] they work in subspace sharing for action recognition based on semi-supervised multi-feature method, also include both global and local consistency for training classifier. In local consistency nearby points are likely to have same label and in global consistency points on the same structure are likely to have the same label so it gives more time to execute.

Lei Zhang [1] feature extraction process is carried out using Scale Invariant Feature Transform (SIFT). The adoption of the same grid parameters which is a spatial stride of 3 pixels with multiple resolutions. The patches are extracted and described using a local descriptor form using a SIFT. The Scale Invariant Feature Transform (SIFT) descriptor is the most widely used for feature extraction. It combines a scale invariant region detector and a descriptor based on the gradient distribution in the detected regions. The feature vector dimension is 128 which is further reduced by using PCA. There are two main difficulties which are faced while applying the Scale Invariant Feature Transform (SIFT) to the large scale database i.e. memory cost and recognition accuracy.

3. SYSTEM ARCHITECTURE

Architecture of the proposed system in Training phase and Testing phase are shown in Figure1.

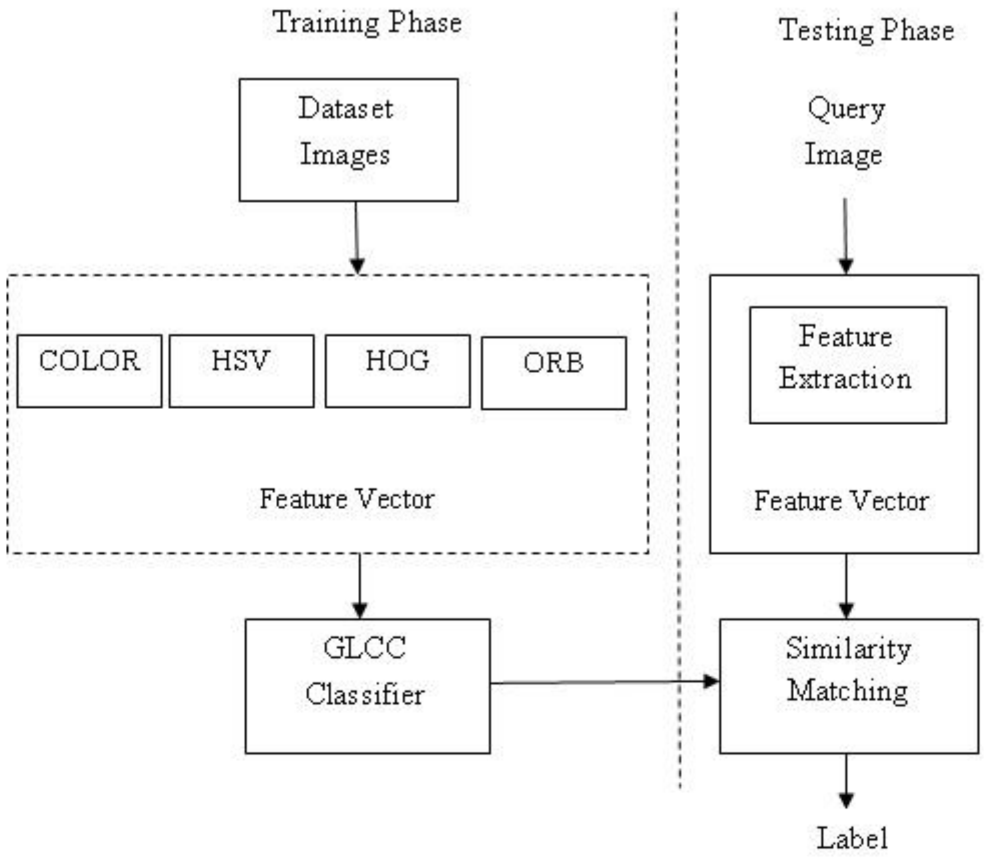

Fig -1:System Architecture

The detailed description of each block is as follows:

Feature Extration

3.1 Feature Extraction

Label prediction system extracts features as per predetermined scheme from each image present in image database and stores the feature vectors in feature database. It is offline process. Performance of label prediction depends on the feature extraction scheme used by the system.

- Color Descriptor

  Typically image having red plane, green plane and blue plane with the intensity values of respective pixel. One particular pixel will have the red intensity value, green intensity value and blue intensity value.

The three bytes of data for each pixel is split into three different parts. 256 different shades of RED, GREEN AND BLUE (one byte can store a value from 0 to 255).
One byte for the amount of RED
One byte for the amount of GREEN
One byte for the amount of BLUE
Suppose image size 198x254=50,292 pixels in images with 3 values per pixel so 50,292x3=150,876 features.

- Hue Saturation Value Color model

  After extracting RGB color values of images by using the color descriptor are given to the HSV color model. Hue is a color attribute and represents a dominant color. Saturation is an expression of the relative purity or the degree to which a pure color is diluted by white light. In the HSV model, the luminous component (brightness) is decoupled from color-carrying information (hue and saturation).

- Histogram Oriented Gradient(HOG)

  Histogram Oriented Gradient (HOG) is one of the visual descriptor generally used for object detection. The purpose of HOG in this project is to extract the HOG features for matching of images. The basic idea behind Histogram Oriented Gradient(HOG)descriptor is that appearance of the local object and shape of the object within an image is specified by the intensity gradients distribution or edge direction. In Histogram Oriented Gradient descriptor image is first divided into small overlapping regions, called cells. Then for each cell compute the histogram of gradient directions and then combine all histograms represent as descriptor.

  Histogram Oriented Gradient(HOG) descriptors are used extract the features. The images of size 200*250 divided into 32*s32 over lapping block, results in total 154 blocks. Each block consist of 2*2 cells with size 8*8 and using 8 orientation bins results into 4928 dimensions.

- Oriented FAST and rotated BRIEF (ORB)

  ORB is basically a fusion of FAST keypoint detector and BRIEF descriptor with many modifications to enhance the performance. First it use FAST to find keypoints, then find top N points among them.

  Feature from Accelerated Segment test (FAST)

  Here identify the similarity between two images by looking at points which has a significant intensity variation with respect to its neighboring pixels.

  $$|I_x-I_p|>t$$

  where t is given threshold, I_x is the gray value of consecutive n pixels, I_p is the gray value of point P.

  Algorithmic steps:

  a) Select a pixel in the image which is to be classified as the interest point or not. its intensity be I_p.
  b) Select an appropriate threshold, t=30.
  c) Consider a circle of 16 pixels around the pixel under test as shown figure2.

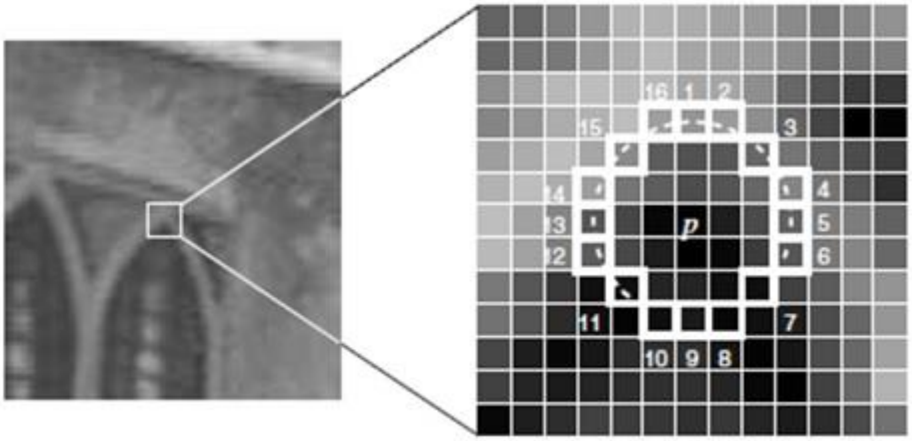

Fig -2:FAST Working

d) Now the pixel is a corner if there exists a set of N pixels in the circle (of 16 pixels) which are all brighter than I_p+ t, or all darker than I_p - t. (Shown as white dash lines in the figure 2). N was chosen to be 12.

e) A high-speed test was proposed to exclude a large number of non-corners. This test examines only the four pixels at 1, 9, 5 and 13 (First 1 and 9 are tested if they are too brighter or darker.If so, then checks 5 and 13). If is a corner, then at least three of these must all be brighter than I_p +t or darker than I_p-t.

Binary Robust Independent Elementary Features (BRIEF)

After obtaining the feature point with direction, descriptors has been built using BRIEF descriptor. BRIEF extracts descriptors around feature point by binary coding method. The image spot P is S*S around the feature point, randomly selecting n_d pairs of pixel point and defining it as tau.

Choose 256 pairs of points randomly. Now go over all the pairs and compare the intensity value of the first point in the pair with the intensity value of the second point in the pair. If the first value is larger then the second, write 1 in the string, otherwise write 0. After going over all 256 pairs, 256 characters string, composed out of 1 and 0 that encoded the local information around the keypoint.

3.2 Global Label Consistent Classifier(GLCC)

The feature vectors are created and this feature vectors are given to the GLCC which train the image dataset.

3.3 Similarity Matching

The similarity matching is performed by using the Euclidean distance between extracted image database feature and extracted query image feature then the label prediction is done to the given query images.

4. Experimental Setup

4.1 OXFORD FLOWER 17 Dataset

Oxford Flower 17 dataset having 1360 flower .JPEG images. The dataset includes 17 categories each category contains 80 images. Size of all images is either 666 *500 or 512 *500. Figure.9.1 shows sample images from Oxford flower 17 dataset in 5 categories viz Daffodil, Snowdrop, Bluebell, Crocus and Tulip. Table 1 shows all 17 categories of images present in Oxford flower 17 dataset.

Fig -3: OXFORD FLOWER 17 Dataset

Table- 1: Categories of Oxford Flower 17 Dataset

No.	Category
1	Daffodill
2	Snowdrop
3	Lilyvalley
4	Bluebell
5	Daisy
6	Tigerlily
7	Crocus
8	Iris
9	Tulip
10	Fritillary
11	Sunflower
12	Coltsfoot
13	Dandelion
14	Cowslip
15	Buttercup
16	Windflower
17	Pansy

4.2 Performance Evaluation

For comparison of system with existing methods classification accuracy are important parameters. Classification accuracy is calculated as follows :
A=t/n*100
A=Classification Accuracy
t = The number of samples correctly classified
n = The total no of samples
Here we are using 10-fold cross validation for accuracy calculation.

4.3 Results

To test the proposed system different experiments were performed using OXFORD Flower 17 dataset. In order to assess the performance of the proposed system, an image set containing 1360 images jpg format is used. From these images 10% of images used as testing images and remaining images used for training. We have tested our system in terms of accuracy in correctly classified images from OXFORD Flower 17 dataset.

Table 2 and Table 3 can be observed that existing system does not classify images but proposed system can classify them.

Table- 2: Analysis of Correctly classified images for (Class Daffodil)

Sr.No.	Image No	Existing System	Proposed System
1	Image_01	No	Yes
2	Image_02	Yes	Yes
3	Image_03	Yes	Yes
4	Image_04	Yes	Yes
5	Image_05	Yes	Yes
6	Image_06	No	Yes
7	Image_07	Yes	Yes
8	Image_08	Yes	Yes
9	Image_09	No	Yes
10	Image_10	Yes	Yes

Table- 3: Analysis of Correctly classified images for (Class Snowdrop)

Sr.No.	Image No	Existing System	Proposed System
1	Image_11	Yes	Yes
2	Image_12	No	No
3	Image_13	Yes	Yes
4	Image_14	Yes	Yes
5	Image_15	Yes	Yes
6	Image_16	No	Yes
7	Image_17	Yes	Yes
8	Image_18	Yes	No
9	Image_19	No	Yes
10	Image_20	No	Yes

Table 4 and Table 5 can be observed that existing system as well as proposed system does not classify images and vice-varsa. But proposed system is better for classification of images.

Table- 4: Analysis of Correctly classified images for (Class Bluebell)

Sr.No.	Image No	Existing System	Proposed System
1	Image_21	Yes	Yes
2	Image_22	No	No
3	Image_23	Yes	Yes
4	Image_24	Yes	Yes
5	Image_25	Yes	Yes
6	Image_26	Yes	Yes
7	Image_27	No	Yes
8	Image_28	Yes	Yes
9	Image_29	No	No
10	Image_30	Yes	Yes

Table- 5: Analysis of Correctly classified images for (Class Iris)

Sr.No.	Image No	Existing System	Proposed System
1	Image_31	Yes	Yes
2	Image_32	No	Yes
3	Image_33	Yes	Yes
4	Image_34	Yes	Yes
5	Image_35	Yes	Yes
6	Image_36	Yes	Yes
7	Image_37	No	Yes
8	Image_38	Yes	Yes
9	Image_39	No	No
10	Image_40	Yes	Yes

Table 6 shows Summary of label prediction in terms of classes. OXFORD FLOWER 17 dataset having 1360 images, it consists of 17 classes with 80 images per class. Class Daffodil, Iris and Buttercup have more correctly classified images out of 80 images and class Lilyvalley, Fritillary and Dandelion shows same results for classification of images.

Table- 6: Summary of label prediction in terms of classes

Sr. No.	Class	Correctly classified Images (out of80 images)	
		Existing System	Proposed System
1	Daffodil	69	73
2	Snowdrop	53	70
3	Bluebell	67	70
4	Iris	72	74
5	Tulip	70	71
6	Sunflower	69	72
7	Coltsfoot	70	71
8	Buttercup	71	74
9	Windflower	71	72
10	Lilyvalley	67	67
11	Fritillary	72	72
12	Dandelion	72	72

Table 7 and Table 8 shows the accuracy of label prediction for existing system and proposed system in terms of predicted label images using ORB feature descriptor and SIFT descriptor. It can be observed that label prediction accuracy of proposed system is better than existing system.

Table- 7: Accuracy of label prediction for the Existing System

Sr.No.	Correctly Classified images (out of 136 images)	Accuracy(%)
1	119	87.5%
2	117	86.02%
3	118	86.76%
4	118	86.76%
5	119	87.5%
6	120	88.23%
7	119	87.5%
8	118	86.76%
9	119	87.5%
10	120	88.23%
	Average Accuracy(%)	87.2%

Table- 8: Accuracy of label prediction for the proposed System

	Correctly Classified images (out of 136 images)	Accuracy(%)
1	119	87.5%
2	117	86.02%
3	120	88.23%
4	119	87.5%
5	122	89.70%
6	125	91.91%
7	119	87.5%
8	118	86.76%
9	119	87.5%
10	122	89.70%
	Average Accuracy(%)	88.2%

Classification accuracy of proposed system is compared with Existing system is shown in figure 4. The classification accuracy is tested on Flower 17 dataset. The accuracy of Flower 17 dataset is increased with proposed system by using ORB.

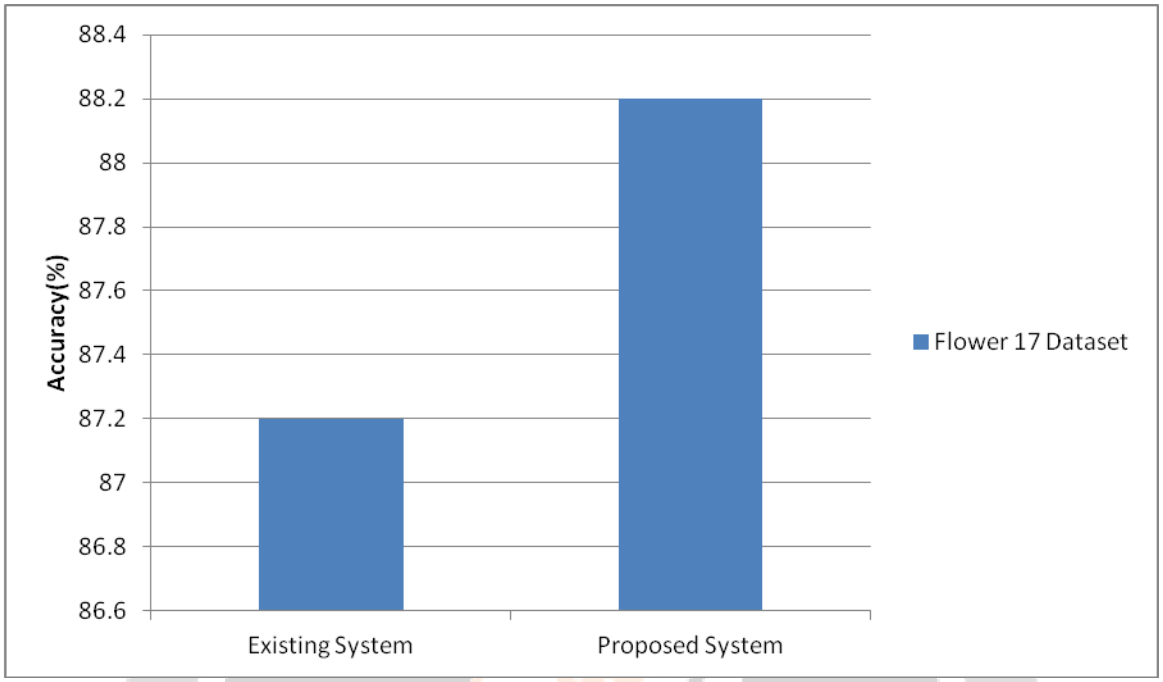

Fig -4: Graph for classification Accuracy

4. CONCLUSIONS

Visual classification using multiple visual features is a challenging task. It is important to note that selection of features for image classification may affect the overall classification accuracy as well as performance of the system. Proposed approach addresses same problem effectively. A novel approach to implement classifier which will select features closely similar to the query image and perform features label prediction has been presented in this report. In the proposed approach system feature extraction techniques are used to get features that will improve overall classification accuracy. Proposed system utilize rotated BRIEF(ORB) descriptor based features which perform better than the earlier method for image classification which usages SIFT as major feature for classification. The label prediction accuracy of proposed system is better than the existing system.

5. ACKNOWLEDGEMENT

The authors would like to thank Prof. Dr. K. N. Nandurkar, Principal and Prof. Dr. S. S. Sane, HOD and vice Principal of K.K.W.I.E.E.R., Nashik for their kind support and suggestions. We would also like to express special thanks to all staff members of Computer Engineering Department and our colleagues who knowingly or unknowingly helped us to complete this course of work.

6. REFERENCES

[1] Zhang, Lei, and David Zhang. "Visual Understanding via Multi-Feature Shared Learning with Global Consistency", IEEE Transactions on Multimedia 18.2 (2016): 247-259.

[2] Chaku Gamit, Prof. Prashant B. Swadas, Prof. Nilesh B. Prajapati "Literature Review on Flower Classification", IJERT, Vol. 4 Issue 02, February-2015.

[3] Lazebnik, Svetlana, Cordelia Schmid, and Jean Ponce. "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories", 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Vol. 2. IEEE, 2006.

[4] Yang, Jianchao, et al. "Linear spatial pyramid matching using sparse coding for image classification",Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.

[5] P. Gehler and S. Nowozin," On feature combination for multiclass objective classification", in Proc. ICCV, pp. 221-228, 2009.

[6] H. Zhang, N.M. Nasrabadi, Y. Zhang, and T.S. Huang, "Multi-Observation Visual Recognition via Joint Dynamic Sparse Representation," in ICCV, pp. 595-602, 2011.

[7] M. Belkin and P. Niyogi, "Semi-supervised learning on manifolds, Machine Learning," vol. 56, pp. 209-239, 2004.

[8] D. Zhou, O. Bousquet, T.N. Lal, J. Weston, and B. Scholkopf, "Learning with local and global consistency," in Proc. NIPS, 2004.

[9] Z. Ma, F. Nie, Y. Yang, J.R.R. Uijlings, N. Sebe, A.G. Hauptmann, "Discriminating Joint Feature Analysis for Multimedia Data Understanding", IEEE Trans. Multimedia, vol. 14, no. 6, pp. 1662-1672, 2012.

[10] S. Wang, Z. Ma, Y. Yang, X. Li, C. Pang, A.G. Hauptmann, "Semi-Supervised Multiple Feature Analysis for Action Recognition",IEEE Trans. Multimedia, vol. 16, no. 2, pp. 289-298, Feb. 2014.

[11] Y. Yang et al., "Multi-feature fusion via hierarchical regression for multimedia analysis", IEEE Trans. Multimedia, vol. 15, no. 3, pp. 572581, Apr. 2013.

[12] M. Hassaballah, Aly Amin Abdelmgeid and Hammam A.Alshazly,"Image Features Detection, Description and Matching", Studies in Computational Intelligence 630, DOI 10.1007/978-3-319-28854-3-2