A Review on Label Prediction through Multiple Visual Features

Ms. Dhanashri S. Narkhede[1] Prof. J. R. Mankar[2]
[1]M.E Student [2]Assistant Professor
[1,2]Department of Computer Engineering
[1,2]K. K. W. I. E. E. R., 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 focusing on computational time of the system by replacing SIFT descriptor by ORB feature descriptor.

Key words: Multi-feature learning, multimedia understanding, Semi-supervised learning, visual recognition, visual features

I. Introduction

Multimedia contents and images are ordinarily used to represents multiple modalities, multiple views and multiple features. 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 modeling 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.

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.

In this work a multiple visual features are jointly learned with effective knowledge and feature structure sharing for robust visual classification and in the low level features we are using ORB which is fast descriptor as compare to SIFT.

In the below sections we are going to discuss about related work done for the proposed research area. We refer some existing research paper for completing this task. It is given as follow:

II. Related work

The existing methods of multiple visual features for images are divided into four categories.

A. Visual recognition

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 [4], 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. [5] 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 doesn't work well with histogram but gives greater performance for max-pooled histogram.

In [6], 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 SVM$_s$ but there is an ordinary fault of these methods that the computational cost is also large.

Zhang et al. [7] 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 sturdy classifier learning for visual understanding.

B. Graph-Based Semi-supervised Learning

Semi-supervised learning has been wide deployed within the recognition task, because of these truth that training some amount of labeled information is liable to overfitting, whereas manual labeling of an outsized quantity of exactly labeled 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[8] 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 [9] proposed local and global consistency with graph regularization for graph based semi-supervised method.

Ma et al [10] Laplacian graph and the l_2-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.

C. Multi-View Learning

Wang et al [11] 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 but it gives more time to execute.

D. Feature Extraction

SIFT [12] uses a feature descriptor with 128 floating point numbers and also Consider thousands of such features so it takes lots of memory and more time for matching whereas BRIEF descriptor which gives the shortcut to find binary descriptors with less memory, faster matching, still higher recognition rate.

SURF descriptor [13] that are calculate fastly and match whereas conserving the discriminative power of SIFT also SIFT, SURF depends on local gradient histograms however it uses integral images to speed up the computation also need different parameter settings are possible however, since a vector of 64 dimensions previously yields good recognition performance, that version has develop into a de facto standard but the drawback is that it is mathematically complicated and computationally heavy. SIFT is based on the Histogram of Gradients. That is, the gradients of each pixel in the patch need to be computed and these computations cost time. It is not effective for low powered devices and also doesn't work well with lighting changes and blur images.

III. Problem Formulation

To design and develop a system for Label Prediction through Multiple Visual Features

IV. System Architecture

In this system there are two phases: testing and training phase. Initially in the training phase the training images from dataset is loaded into the system, after that the features are extracted by low level feature descriptor such as color, HSV, HOG and ORB etc and generate feature vector. In the next step training data of m features and training parameters are given as input to the GLCC classifier. At the testing phase same process is done for query image. Input the query image and apply feature extraction process and generate feature vector. Training image data and extracted features are given as input to the GLCC Classifier that select features closely similar to the query image and based on these interpretation label prediction is done.

Figure 1 represents proposed system architecture. Processing of proposed work is takes places as following way:

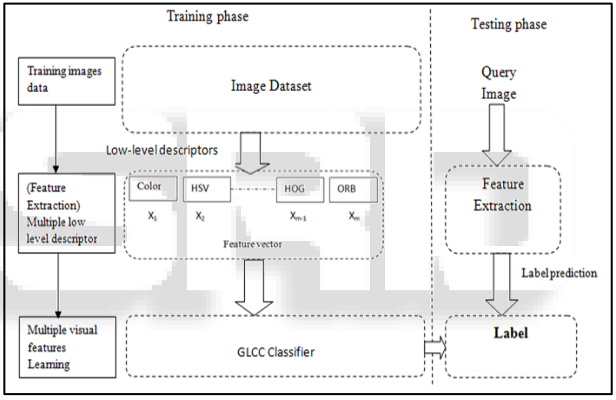

Fig. 1: System architecture

A. Image Dataset

Image dataset contains n number of images. I={I_1,I_2,...............,In}

B. Feature Extraction

Features are extracted from the images by using low-level feature descriptors such as Color, HSV,HOG and ORB. Color feature descriptor extract features of image for every pixel result in RGB value i.e RED, GREEN AND BLUE and ranges from 0 to 255 also construct color feature vector. HSV (Hue Saturation Value) Hue define color sensation of the light and works circular. Saturation indicates purity of color and Value contain color with maximum value i.e. 255can be any color with its maximum brightness and also construct HSV feature vector. HOG (Histograms of Oriented Gradients) image can be divided into blocks. Compute the gradient vector at each pixel according to x-direction and y-direction also compute magnitude and angle of a vector and put them into 9-bin histogram then normalizing gradient vectors after normalizing histogram and last normalize block. ORB(Oriented FAST and Rotated BRIEF) used to extract features from image and construct ORB feature vector.

C. GLCC Classifier

After extraction of the features classifier select features closely similar to the query image and based on this similar features classifier assign label to image.

V. Conclusion

In this survey paper, several existing techniques have studied and analysed in section II. Traditional feature extraction methods work effectively and efficiently to extract image features. Some feature descriptors such as, SIFT, SURF, Color, HSV and HOG efficiently extract image features. This work aims at studying different low-level feature descriptor for better feature extraction and focusing on computational time of the system by replacing SIFT descriptor by ORB feature descriptor. ORB is built on the well-known FAST keypoint detector and the recently-developed BRIEF descriptor it is faster than SIFT. Given a query image, feature vectors are constructed and label prediction is done.

Acknowledgment

I am deeply expressing my sincere gratitude to my project guide Prof. J. R. Mankar who always supported and guided me with valuable comments. I express my immense pleasure and thankfulness to all faculty members of the Department of Computer Engineering of K. K. W. I. E. E. R., Nashik

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

[5] 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.

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

[7] 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.

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

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

[10] 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.

[11] 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.

[12] “Distinctive Image Features from Scale-Invariant Keypoints,” International Journal of Computer Vision, vol. 20, no. 2, pp. 91–110, 2004.

[13] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, “SURF: Speeded Up Robust Features,” Computer Vision and Image Understanding, vol. 10, no. 3, pp. 346–359, 2008.