Well-Known brands recognition by automated classifiers using local and global features

Hafsa Niaz [a,1,*], Usman Raza [b,2]

[a] Kinnaird College for Women, 93 Jail Road, G.O.R.-I, Lahore, 54000, Pakistan
[b] University of Engineering & Technology, G.T. Road, Staff Houses Engineering University, Lahore, 39161, Pakistan
[1] *hafsa.niaz.7@gmail.com* [*]; [2] *usmanahmedraza@gmail.com*
* *corresponding author*

ABSTRACT

From color and type to patterns and illustrations, brands sense to be recognizable and convey their values and personality. Here patterns and color are key elements, as they can play a vital role in brand recognition. The images used for brand classification were handpicked and collectively named as HKDataset. We have explored various feature extractors used for classification and used automated classifiers named Linear SVM to achieve higher accuracy while tuning the model parameters to achieve optimal performance. It has been observed that Support Vector Machines performs better when using GIST descriptors combined with Bag of SIFT features. We hope to apply deep learning and other sophisticated classifiers to much-expanded categories of brands in the future.

Keywords:
GIST Descriptor
K-Means Clustering
Linear SVM
SIFT features
HKDataset

I. Introduction

Advertising experts define "brand" as the "name, term, sign, symbol or design, or a combination of them projected to classify a company's products or services." Brands communicate meanings with the language of color and shape. Brands and color are un-distinguished because the color is the visual component people remember most about a brand, followed closely by shapes/symbols. So, these features can be used for recognition. However, scene classification is a very active topic in computer vision research, and much work has been done on it. However, the work specific to brand classification/recognition still has many rooms for further discoveries.

Features are the information extracted from images in terms of numerical values that are difficult to understand and correlate by humans. Two types of features are extracted from the images based on the application. They are local and global features, which are also referred to as descriptors. For this paper, we selected two approaches to classify a limited set of brands from real-life images. The two approaches being local feature classifier and global feature classifier, and both are used for object recognition.

Global image features describe an image as whole whereas local feature represents as image patches. Global feature generalizes the whole image with a single vector, and local compute with multiple distinct features like points, blobs, or edges on the image, making it more robust. Global features include contour representations, shape descriptors, and texture features. Global texture features and local features provide different information about the image because of the support over which texture varies as per the computation. Generally, for low-level applications such as object detection and classification, global features are used, and for higher-level applications such as object recognition, local features are used. The combination of global and local features improves the accuracy of recognizing the side effect of computational overheads.

For this study, we used the SIFT method to abstract local image features from the training set. The bag of words objects further defines the features by using the k-means clustering algorithm on the feature description, where each cluster defines a specific feature or a word. To achieve higher accuracy, use a gist descriptor, which creates a vector with specific parameters of the scene between indoor and outdoor images for simplicity.

II. Method

A. Literature Review

Previous work in logo recognition has focused on discovering design marks in historical documents [1] or similarity check-ups for trademark [2]. Other work attempts classification in real-world scenes. For logos in sports videos, Zhi Li [3] assumes rigid planar backgrounds and utilizes line-based intensity profiles. As local interest point descriptors have become popular, there have been approaches that use point correspondences for matching templates. Bagdanov et al. [4] Employ normalized matching using a bag of local features and cloud localization in sports videos. This method can scale to more than just a few logos. Fakhar et al. [5] use the Apriori algorithm to identify frequent spatial configurations of local features extracted on a spatial pyramid. Here mining requires a large amount of training data and is computationally costly. Kalantidis, Yannis, et al. [6] extend the bag-of-words model, incorporate Delaunay triangulation, represent triangles by signatures for each class, and use these signatures classification. However, there are limitations regarding image resolution. Romberg, Stefan, et al. [7] used HOG features extractor to recognize manually labeled logo images, complemented with non-logo images. Simultaneously, we have used non-labeled logo images handpicked from the search engines for the learning process.

This paper opted for Brand classification/recognition in real-world images by using a famous classifier, like Support Vector Machines. It was a state-of-the-art algorithm of its time, as it is also preferred in medical image processing [8] – [20] and has fewer training parameters which leads to higher stability [21]. Moreover, the Bag of Scale-Invariant Feature Transform (SIFT) feature extraction model and GIST Descriptor model will be used. As there is already much work being done to present new approaches in the field for image classification on the same old datasets, it has gained interest from [22] to create our dataset. The aim is to apply these algorithms to the new dataset and relate the results to see whether these longstanding algorithms are still the finest. As from previous research by Lou et al., the image classification-using bag of features model was more efficient than other algorithm models [23]. Some researchers [24] – [25] also reported that image classification-using bag of features and SVM as a classifier could significantly improve the accuracy of classification, as used in [26] to detect review spam on the labeled dataset of social media. Other recommend also categorizing the scenes like in Aircraft detection [27] – [28]. The texture characteristics and Gabor filtering method are used to reduce detection processing and computational complexity while refining detection effectiveness by 90%.

B. Feature Extraction Model

In this study, there are two feature extraction techniques used. The two techniques are SIFT and GIST Descriptor. The details are presented in the following description.

1. SIFT Descriptor

SIFT is a method that is consistent with scope, brightness, and angle. For our research object, this method is appropriate for extracting local features that can help describe the different brand types, which have almost the same characteristic. It mainly consists of the following steps:

a. *Firstly, the pixel will be matched with 26 pixels values, which are eight surrounding pixels in the same scope to find extreme values to locate key points in the image.*

b. *Each key point vector will be computed via the direction of the gradient of its surrounding pixels. Moreover, the most projective direction of a point is chosen as a key.*

SIFT feature vector is generated. Lowe advised characterizing each key point using 4×4 seed points in a specific calculation to boost matching constancy. Thus, 128-dimensional SIFT vectors are obtained for each key point (shown in Fig. 1).

The resulting vocabulary, which is further used to explain the individual image in a histogram, implements the clustering method and feature extraction. There are three main stages of the bag of features:

a. *Declaration of image features using SIFT*

b. *Grouping descriptor to the set of clusters (vocabulary) with vector quantization algorithm using K means clustering.*

c. *Construction of bag of features enumerates the number of features that belong to each cluster.*

Fig. 1. Features extracted from an image of Coca-Cola signage using SIFT. The circles represent the various features detected by the SIFT descriptor.

2. GIST Descriptor

The GIST feature is a global image feature, which defines specific important figures about a scene. Olive et al. firstly proposed the GIST feature. It was used in place and object recognition [29]. In conclusion, they showed that it is possible to categorize different scenes like buildings, streets, trees, sky, cars, streetlights, and persons using the GIST descriptor. Our current work shows that it is possible to use GIST for Brand recognition in typical outdoor environments (shown in Fig. 2).
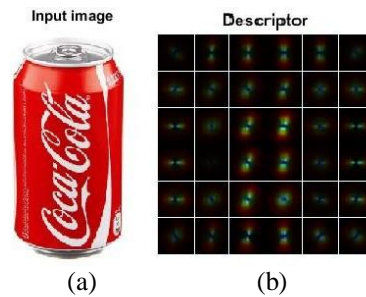

(a)   (b)

Fig. 2. The GIST descriptor for (a) the corresponding image and (b) the image shows average filter energy in each bin once.

Given an input image, a GIST descriptor is computed by following stages:

a. Convolute the image with 32 Gabor filters at four scales, eight orientations, producing 32 feature maps of the same input image size.

b. Divide each feature map into 16 regions (by a 4x4 grid) and then average each region's feature values.

c. Concatenate the 16 averaged values of all 32 feature maps, resulting in a 16x32=512 GIST descriptor.

C. Machine Learning Algorithm

Support Vector Machine (SVM) is a supervised machine-learning algorithm, which can be, used for both classification and regression challenges.
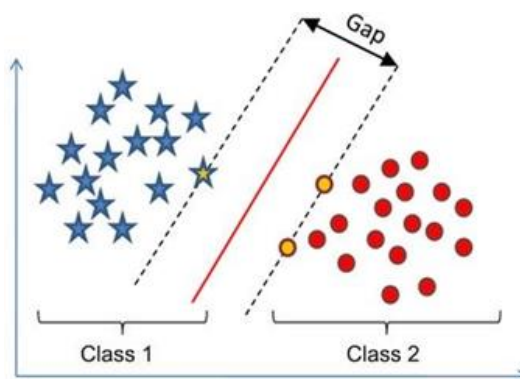

Fig. 3. The GIST descriptor for (a) the corresponding image and (b) the image shows average filter energy in each bin once.

In this algorithm, each data item is plotted as a point in n-dimensional space (where n represents the number of features). The value of each feature is the value of a distinct coordinate. Then, classification is applied by finding the hyper-plane that marks multiple classes (Fig. 3).

Local context information in the traditional SIFT and global RGB color quantization information in GIST, when combined, overcomes the shortcomings in both. By using these combined features, the brand classification task is performed with linear SVM. The influence on classification accuracy of the combined features when examined, the results are even better.

D. K-Cross Validation Technique

K-Cross validation is used to evaluate or compare learning algorithms. The data set is divided into k subsets, and the holdout method is repeated k times. Each time, one of the k subsets is used as the test set, and the other k-1 subsets are put together to form a training set. Here we have used the 5-fold Cross-validation technique where the value of k is 5, which is mostly used; since we need to find the average accuracy across all k trials, the value of k does not matter result will be the same in most cases.

E. Local Features used for classification

SIFT features were extracted by using open-source algorithms from VLFeat [30]. The SVM classifier was trained on seven different SVM's for each of the different classes for each fold. All 7 SVM classifiers are run on each test image, and the classification with the average highest score is returned for each fold.

F. Global Features used for classification

Although SIFT features are quite descriptive, many other feature representations can be useful to classify a brand. The GIST feature vector extracted from each image is normalized and used as the feature vector in place of the SIFT histogram.

III. Results and Discussion

A. Result

Intuitively, as lambda grows, the less the wrongly classified examples are allowed. The different lambda values were considered to determine a good value for lambda, and found the best results can be seen in "Table. 2" using SIFT features at lambda = 0.01, GIST features at lambda=0.0001, and combined (GIST+SIFT) features at lambda=0.001.

Table 1. Methods used for descriptors and adjusting parameters showing accuracy rate accordingly

K Cross Validation	Accuracy Testing Data		
K Value	SIFT	GIST	SIFT+GIST
1	88.6	83.3	**92.1**
2	92.1	84	**91.4**
3	88.1	81	**92.6**
4	85.5	85	**87.6**
5	86.7	82.6	**90.2**
Average %	82.2	83.13	**90.78**

Despite using suitable parameters for achieving the best SVM classification, the Accuracy rate is the highest after using combined GIST and SIFT features [31]; after this lies SIFT achieving better overall values than using the only GIST.

Table 2. Dataset

Categories	No. of trained Images per category	No of tested Images per category
Coca-cola	480	120
No Object	480	120
Michelin	480	120
Starbucks	480	120
Heineken	480	120
Agip	480	120
Gorilla	480	120

B. Performance Measure

The Precision is the number of true positives divided by the total number of elements labeled as the positive class. The recall is defined as the number of true positives divided by the total number of elements that belong to the positive class. Details can be found in Fig. 4. Fig. 4 shows the average PR curve made after making PR curves per class for each k-fold.
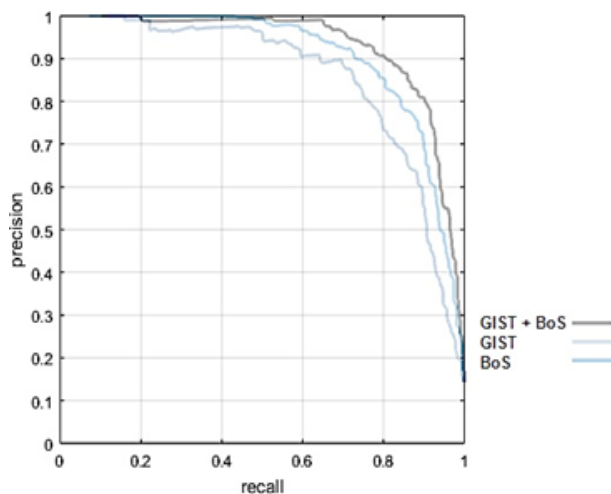

Fig. 4. Precision and Recall Curve

C. Evaluation

For a realistic evaluation for Brand recognition, an extensive collection of photos in a real-world environment is required. Few logo datasets are available, but we feel that those datasets cannot evaluate an approach based on local and global features. They contain many images containing minimal logos. We feel that these datasets are not suitable for evaluating the recognition process rather than the visual features themselves.

Therefore, we built a new dataset, "HKDataset" [32], containing photos depicting brands. The collection is made of logo, signage, mascot, and product images of 6 different classes by downloading them from Web service with appropriate queries. There are six well-known brand classes: Coca-Cola, Starbucks, Michelin, Heineken, Agip, and Gorilla.

To facilitate achieving high precision of classifier, evaluating their sensitivity on Non-object images is very important. Therefore, these non-logo images were downloaded from Web service with the queries "building", "nature", "roads", and "urban buildings". These images are unlikely to contain brands and complete our dataset. A summary of the images per category is shown in "Table. 2".

IV. Conclusion

In this paper, we elected for Brand classification/recognition in real-world images by using famous classifier, Support Vector Machines that is a state-of-the-art algorithm of its time. For Brand recognition, we have used 5-Cross Validation Technique to check its validation on the Bag of Scale-Invariant Feature Transform (SIFT) feature extraction model and GIST Descriptor model. Firstly, we used SIFT and GIST features and then applied classification using them. However, when SIFT combined with GIST features, it resulted in a more accurate rate while tuning the parameters. It also resulted in a better Precision-Recall curve by attaining high precision and better recall. We hope to apply deep learning and any new robust algorithm on much-expanded Brands classes in the future.

References

[1] Chanda S., Prasad P.K., Hast A., Brun A., Martensson L., Pal U.Finding Logo and Seal in Historical Document Images - An Object Detection Based Approach. In: Palaiahnakote S., Sanniti di Baja G.,

Wang L., Yan W. (eds) Pattern Recognition. ACPR 2019. Lecture Notes in Computer Science, vol 12046. Springer, Cham, 2020, doi: https://doi.org/10.1007/978-3-030-41404-7_58

[2] Taileb M. Trademark Image Retrieval System Using Indexing Techniques. International Journal of Computer Science and Information Security (IJCSIS). 2019 Sep, 17 (9), available at Google Scholar

[3] Li, Z. The study of security application of LOGO recognition technology in sports video. J Image Video Proc. 2019, 46, 2019, doi: https://doi.org/10.1186/s13640-019-0441-8

[4] Andrew D. Bagdanov, Lamberto Ballan, Marco Bertini, and Alberto Del Bimbo. 2007. Trademark matching and retrieval in sports video databases. In Proceedings of the international workshop on Workshop on multimedia information retrieval (MIR '07). Association for Computing Machinery, New York, NY, USA, 79–86. doi: https://doi.org/10.1145/1290082.1290096

[5] Fakhar, B., Rashidy Kanan, H. & Behrad, A. Event detection in soccer videos using unsupervised learning of Spatio-temporal features based on pooled spatial pyramid model. Multimed Tools Appl 78, pp. 16995–17025, 2019, doi: https://doi.org/10.1007/s11042-018-7083-1

[6] Yannis Kalantidis, Lluis Garcia Pueyo, Michele Trevisiol, Roelof van Zwol, and Yannis Avrithis. Scalable triangulation-based logo recognition. In Proceedings of the 1st ACM International Conference on Multimedia Retrieval (ICMR '11). Association for Computing Machinery, New York, NY, USA, Article 20, pp. 1–7, 2011, doi: https://doi.org/10.1145/1991996.1992016

[7] Stefan Romberg, Lluis Garcia Pueyo, Rainer Lienhart, and Roelof van Zwol. Scalable logo recognition in real-world images. In Proceedings of the 1st ACM International Conference on Multimedia Retrieval (ICMR '11). Association for Computing Machinery, New York, NY, USA, Article 25, pp. 1–8, 2011, doi: https://doi.org/10.1145/1991996.1992021

[8] Murugan, A., Nair, S.H. & Kumar, K.P.S. Detection of Skin Cancer Using SVM, Random Forest and kNN Classifiers. J Med Syst 43, 269, 2019, doi: https://doi.org/10.1007/s10916-019-1400-8

[9] C. Venkatesan, P. Karthigaikumar, A. Paul, S. Satheeskumaran and R. Kumar, "ECG Signal Preprocessing and SVM Classifier-Based Abnormality Detection in Remote Healthcare Applications," in IEEE Access, vol. 6, pp. 9767-9773, 2018, doi: 10.1109/ACCESS.2018.2794346

[10] Yadav, S.S., Jadhav, S.M. Deep convolutional neural network based medical image classification for disease diagnosis. J Big Data 6, 113, 2019, doi: https://doi.org/10.1186/s40537-019-0276-2

[11] Harimoorthy, K., Thangavelu, M. Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system. J Ambient Intell Human Comput (2020). doi: https://doi.org/10.1007/s12652-019-01652-0

[12] Izonin, I., Trostianchyn, A., Duriagina, Z., Tkachenko, R., Tepla, T., Lotoshynska, N.: The combined use of the Wiener polynomial and SVM for material classification task in medical implants production. Int. J. Intell. Syst. Appl. (IJISA) 10(9), pp. 40–47, 2018, doi: https://doi.org/10.5815/ijisa.2018.09.05

[13] A. K. Sari and F. M. Widya Prasetya, "Linear SVM for Classifying Breast Cancer Data Encrypted Using Homomorphic Cryptosystem," 2019 5th International Conference on Science and Technology (ICST), Yogyakarta, Indonesia, 2019, pp. 1-6, doi: 10.1109/ICST47872.2019.9166454.

[14] Jorge D. Mello-Román, Julio C. Mello-Román, Santiago Gómez-Guerrero, Miguel García-Torres, "Predictive Models for the Medical Diagnosis of Dengue: A Case Study in Paraguay", Computational and Mathematical Methods in Medicine, vol. 2019, Article ID 7307803, 7 pages, 2019, doi: https://doi.org/10.1155/2019/7307803

[15] M. Szymkowski, E. Saeed, M. Omieljanowicz, A. Omieljanowicz, K. Saeed and Z. Mariak, "A Novelty Approach to Retina Diagnosing Using Biometric Techniques With SVM and Clustering Algorithms," in IEEE Access, vol. 8, pp. 125849-125862, 2020, doi: 10.1109/ACCESS.2020.3007656.

[16] Chang, W.; Liu, Y.; Xiao, Y.; Yuan, X.; Xu, X.; Zhang, S.; Zhou, S. A Machine-Learning-Based Prediction Method for Hypertension Outcomes Based on Medical Data. Diagnostics 2019, 9, 178, doi: https://doi.org/10.3390/diagnostics9040178

[17] Khalaf M. et al. An Application of Using Support Vector Machine Based on Classification Technique for Predicting Medical Data Sets. In: Huang DS., Jo KH., Huang ZK. (eds) Intelligent Computing Theories and Application. ICIC 2019. Lecture Notes in Computer Science, vol 11644. Springer, Cham. 2019, doi: https://doi.org/10.1007/978-3-030-26969-2_55

[18] Blum, D., Liepelt-Scarfone, I., Berg, D. et al. Controls-based denoising, a new approach for medical image analysis, improves prediction of conversion to Alzheimer's disease with FDG-PET. Eur J Nucl Med Mol Imaging 46, pp. 2370–2379, 2019. https://doi.org/10.1007/s00259-019-04400-w

[19] L. Ali et al., "An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure," in IEEE Access, vol. 7, pp. 54007-54014, 2019, doi: 10.1109/ACCESS.2019.2909969.

[20] Remya Ajai A.S., Gopalan S. Analysis of Active Contours Without Edge-Based Segmentation Technique for Brain Tumor Classification Using SVM and KNN Classifiers. In: Jayakumari J., Karagiannidis G., Ma M., Hossain S. (eds) Advances in Communication Systems and Networks. Lecture Notes in Electrical Engineering, vol 656. Springer, Singapore, 2020. doi: https://doi.org/10.1007/978-981-15-3992-3_1

[21] Raczko, E., and Zagajewski, B. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. Eur. J. Remote Sens. 2019, (50), pp. 144–154, doi: https://doi.org/10.3390/rs12030516

[22] Tous, R., Gomez, M., Poveda, J. et al., "Automated duration of brand-related social media images with deep learning", in Multimedia Tools and Applications, 2018, 77, (20), pp 27123–27142, doi: https://doi.org/10.1007/s11042-018-5910-z

[23] X. Lou, D. Huang, L. Fan, A. Xu, "An Image Classification Algorithm Based on Bag of Visual Words and Multi-kernel Learning" in Journal of Multimedia, 2014, 9, (2), pp 269–277. available at Google Scholar

[24] K. Roy and S. V. R. G, "ART based clustering of bag-of-features for image classification," 2012 5th International Congress on Image and Signal Processing, Chongqing, 2012, pp. 841-846, doi: 10.1109/CISP.2012.6470016

[25] W. Lin, Y. Wu, W. Hung, and C. Tang, "A Study of Real-Time Hand Gesture Recognition Using SIFT on Binary Images," in Advances in Intelligent Systems and Applications, 2013, 2, pp. 235-246, doi: https://doi.org/10.1007/978-3-642-35473-1_24

[26] Rout, J.K., Singh, S., Jena, S.K. et al., "Deceptive review detection using labeled and unlabeled data", in Multimedia Tools and Applications, 2017, 76, (3), pp 3187-3211, doi: https://doi.org/10.1007/s11042-016-3819-y

[27] Lingling Yu, Qingxiang Yang and Limin Dong, "Aircraft target detection using multimodal satellite-based data", in Signal Processing, 2019, Vol. 155, doi: https://doi.org/10.1016/j.sigpro.2018.09.006

[28] J. James, J. J. Ford and T. L. Molloy, "Below Horizon Aircraft Detection Using Deep Learning for Vision-Based Sense and Avoid," 2019 International Conference on Unmanned Aircraft Systems (ICUAS), Atlanta, GA, USA, 2019, pp. 965-970, doi: 10.1109/ICUAS.2019.8798096.

[29] A. Oliva, and A. Torralba, "Building the Gist of a Scene: The Role of Global Image Features in Recognition," in Visual Perception for Progress in Brain Research, 2006, Vol. 155, doi: https://doi.org/10.1016/S0079-6123(06)55002-2

[30] 'VLFeat Open Source Computer Vision Library.', http://www.vlfeat.org/ , accessed September 24, 2017

[31] Min Zeng, Fuhao Zhang, Fang-Xiang Wu, Yaohang Li, Jianxin Wang, Min Li, Protein–protein interaction site prediction through combining local and global features with deep neural networks, Bioinformatics, Volume 36, Issue 4, 15 February 2020, Pages 1114–1120, https://doi.org/10.1093/bioinformatics/btz699

[32] 'HKDataset', https://www.kaggle.com/hafsaniaz/hkdataset/ created on October 25, 2020.