ISSN: 2752-3829 Vol. 3 No.2, (December, 2023)

Stochastic Modelling and Computational Sciences

IMAGE COMPOSITE CHARACTERISTIC-CENTRIC ENTITY DISCOVERY, IDENTIFICATION, AND LABELING

[1]Mr. Dileep Kumar, [2]Versha Chaudhari, [3]Riddhima Srivastava, [4]Tafseer Fatma and [5]Manisha Tiwari
[1]Assistant Professor MCA, [2, 3, 4,5]Research Scholar MCA
Dr. Ram Manohar Lohia Avadh University Ayodhya U.P. - 224001
[1]dileep_k_2000@yahoo.com, [2]versha03may2001@gmail.com,
[3]srivastavariddhima305@gmail.com, [4]tafseerr2000@gmail.com and
[5]manishatiwari4824@outlook.com

ABSTRACT
The identification and categorization of visual elements are crucial for efficient organization and retrieval in image-centric platforms such as Flickr, Picasa, and Facebook. This task, integral to machine learning, presents challenges and holds significant importance. Algorithms based on 'Nearest neighbor classification' have garnered attention but come with increased computational complexity during both training and testing phases. Current methods often focus on single-object descriptors, overlooking the importance of shape, color, and texture features in comprehensive object description.This research paper explores various studies concerning object mining and labeling, proposing a novel system for the discovery, identification, and labeling of composite characteristics-centric entities within images. Our approach initially employs K-Nearest Neighbors (KNN) to tag various object features during training, integrating color moments, shapes, and Gray Level Co-Occurrence Matrix (GLCM) as texture features. Subsequently, an AdaBoost classifier is utilized for object classification, resulting in the final image representation annotated with distinct object tags.Through this methodology, we strive to enhance the efficiency and accuracy of visual element identification, contributing to advancements in image search technology and database organization.

1. INTRODUCTION

In today's digital age, the proliferation of image-centric platforms such as social media networks, online repositories, and search engines has led to an exponential growth in visual data. These platforms host vast collections of images, ranging from personal snapshots to professional photographs, artistic creations, and scientific imagery. However, the sheer volume of visual content available presents a significant challenge in terms of organization, retrieval, and meaningful analysis. Effective techniques for automatically identifying, categorizing, and labeling visual elements within images have become increasingly crucial to harnessing the full potential of these repositories.

The process of discovering and labeling objects within images, known as object mining and tagging, is a fundamental task in the realm of computer vision and machine learning. Traditionally, image tagging has relied on algorithms such as Nearest Neighbor Classification (NNC) to associate images with descriptive labels based on similarity metrics. While NNC and similar methods have demonstrated effectiveness, they often overlook the complex interplay of characteristics that define objects within images.Key components of our approach include the utilization of K-Nearest Neighbors (KNN) for initial feature tagging during training, incorporating color moments, shape descriptors, and Gray Level Co-Occurrence Matrix (GLCM) features to represent object characteristics.

The significance of this research lies in its potential to improve the efficiency, accuracy, and semantic richness of image annotations, thereby facilitating more advanced search capabilities, content recommendation systems, and data-driven insights from visual data repositories.

2. LITERATURE STUDY

The literature study encompasses a wide range of research works related to object detection, recognition, and tagging in images. Various methodologies and algorithms have been proposed and explored in recent years to enhance the capabilities of automated image analysis systems.

Lai et al. [1] introduced a vision-based automated parking monitoring system utilizing geo-tagged UAV images, demonstrating the application of aerial imagery in surveillance tasks. Rajarajeswari et al. [2] focused on real-time object detection using the Yolo-V3 algorithm, showcasing advancements in object recognition techniques.

Šikić et al. [3] contributed to the field by addressing multi-class price tag detection in images, particularly relevant in retail environments for inventory management. Kaur and Singh [4] delved into medical image analysis, specifically automatic image captioning for generating diagnostic reports.

Li et al. [5] explored object detection and localization using passive RF backscattering systems, highlighting innovations in non-visual sensing technologies. Raavi et al. [6] ventured into automated recognition of underwater objects, leveraging deep learning methods for challenging environmental conditions.

Sie and Vasisht [7] introduced RF-Annotate, an automatic RF-supervised image annotation system, bridging RF sensing with image analysis for context-aware tagging. Patil et al. [8] contributed to smart adware and product recommendation systems using object detection techniques.

Dulal et al. [9] focused on automatic cattle identification, employing YOLOv5 and augmentation techniques for improved accuracy. Parker et al. [10] addressed the live detection of foreign object debris on runways using drones and AI, showcasing applications in aviation safety.

Additionally, Rahman et al. [11] proposed Deep0Tag, a deep learning model for zero-shot image tagging, advancing capabilities in semantic understanding. Liao et al. [12] explored object modifier generation for image captioning, enhancing contextual information in automated descriptions.

Lee et al. [13] contributed to machine-assisted video tagging of elderly activities, showcasing the role of AI in healthcare monitoring. Zin et al. [14] focused on cow identification systems using ear tag recognition, illustrating applications in agricultural management.

Chowdhury et al. [15] explored object detection and classification through cascade object training, demonstrating methodologies for robust feature extraction and categorization.

These diverse studies collectively contribute to the ongoing advancements in computer vision, machine learning, and image processing domains, paving the way for enhanced automated analysis, tagging, and understanding of visual content.

3. METHODOLOGY

3.1 Feature Extraction

Scale Invariant Feature Transform (SIFT): The scale-invariant feature transform (SIFT) is a computer vision-based feature detection algorithm for detecting and describing local features in images. It supports object recognition, robotic mapping and navigation, image sewing, 3D modeling, gesture recognition, video tracking, individual identification of wildlife, and match movement. SIFT key points of objects are first extracted from a set of reference images and stored in a database. An object is recognized in a new image by individually comparing each feature of the new image with that database and searching for potential match characteristics based on the Euclidean distance of its feature vectors. From the complete set of matches, the subsets of key points corresponding to the object as well as its location, scale, and orientation in the new image are identified to filter good matches. The determination of coherent groups is done quickly using an efficient implementation of the hash table of the generalized Hough transformation. Each group of 3 or more features that correspond to an object and its position undergoes a more detailed check of the model and, later, the outliers are ignored. Finally, the probability that a particular set of features indicates the presence of an object is calculated, depending on the accuracy of the setting and the number of possible false matches. The object matches that pass all these tests can be identified as correct with high confidence [1]."

Speeded Up Robust Features (SURF):"In artificial vision, Speeded Up Robust Features (SURF) are a patented local feature detector and descriptor. It can be used for tasks such as object recognition, image recording, 3D classification or reconstruction. It is inspired in part by the SIFT descriptor (entity transformation invariable to scale). The standard version of SURF is several times faster than SIFT and, according to its authors, it is more robust for different image transformations than SIFT. To detect points of interest, SURF uses an integer approximation of the determinant of the Hesse spot detector, which can be calculated with 3 integer operations using a recalculated integral image. Its feature descriptor is based on the sum of the Haar wavelet response around the point of interest. These can also be calculated using the full image. The SURF descriptors were used to locate and recognize objects, people or faces, to reconstruct scenes in 3D, to track objects and to extract points of interest [4].

Histogram of oriented gradients (HOG):"HOG is a best feature of local texture, shape and edge information. The HOG descriptor is used in a top-down searching technique that covers the scale variation of the objects in images. It is used to increase the accuracy of localization [3]."

Gray level co-occurrence matrix (GLCM): GLCM utilized to Textural properties associate with image. Colorization will be a coloring transform in the picture or video, which will be carried on give acceptable point of interest Also clarity of the picture alternately feature. Color grayscale pictures naturally by directing, including those GLCM composition characteristic utilizing aggregate for outright contrasts. Eventually Tom's perusing matching those shade and grayscale picture composition features for each pixel block, it will be required should furnish a coloring consequence comparable to that of the format picture [7].

Color Feature: Color feature are used to quantify colors presents in the different categories images. For that we computed features like Mean, Standard deviation, Skewness and Kurtosis of the object in images over different color spaces like HSV, YCbCr. This feature greater helps to identify objects.

Shape features: It's include area, perimeter and eccentricity of skin lesion images.

Area: It is represented as total number of pixel in object.

Perimeter: Distance around the boundary of the region. It is defined as number of edge pixels in segmented object image.

Major axis: It's height of the pixel of object.

Minor Axis: It's Width of the pixel of object.

3.2 Classification

Support vector machine (SVM): A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side [1]. It is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot).

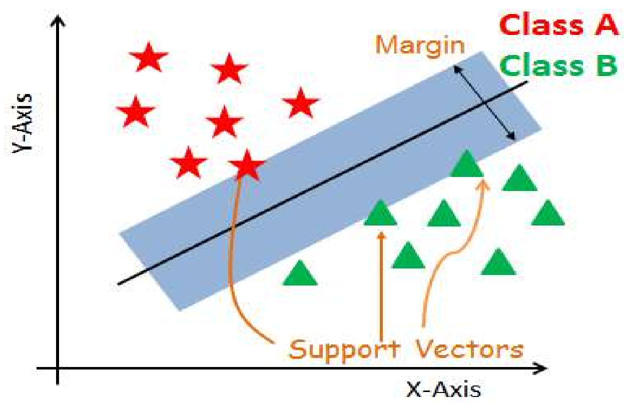

Figure 1. Support vector Machine Learmning

K-Nearest Neighbor (KNN): KNN can be used for both classification and regression predictive problems. However, it is more widely used in classification problems in the industry. To evaluate any technique, we generally look at 3 important aspects: 1. Ease to interpret output 2. Calculation time 3. Predictive Power.

KNN Algorithm is based on feature similarity: How closely out-of-sample features resemble our training set determines how we classify a given data point:

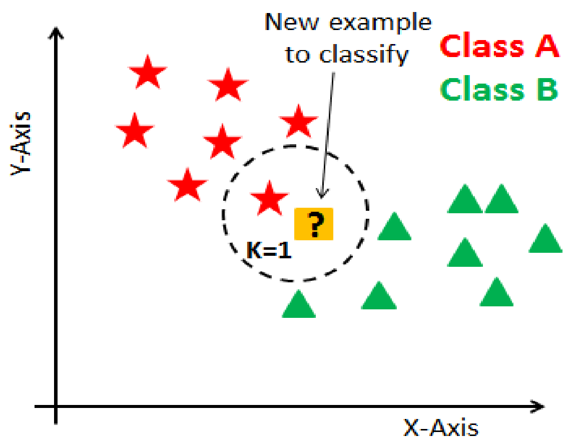

Figure 2. KNN Learning

KNN can be used for classification—the output is a class membership (predicts a class—a discrete value). An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. It can also be used for regression—output is the value for the object (predicts continuous values). This value is the average (or median) of the values of its k nearest neighbors.

Neural system (NN): A neural network consists of units (neurons), arranged in layers, which convert an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer. Generally the networks are defined to be feed-forward: a unit feeds its output to all the units on the next layer, but there is no feedback to the previous layer. Weightings are applied to the signals passing from one unit to another, and it is these weightings which are tuned in the training phase to adapt a neural network to the particular problem at hand [4].

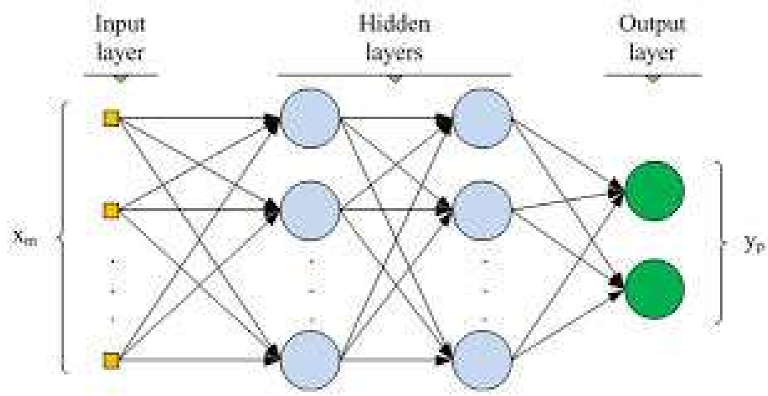

Figure 3. NN Classification

In supervised learning, the classification step often happens in the last layer, and takes the key features of the sample as input from the previous layers. There are different classification functions, depending on the use case. The most common one uses the Softmax function - where for each sample, the result is the probability distribution over the classes. So if there are 1000 possible classes, the Softmax function will return a 1000 dimension vector, where the i'th value would be the probability that the sample is of the i'th class. The total sum of all probabilisties will add up to 1.

4. Proposed Approach

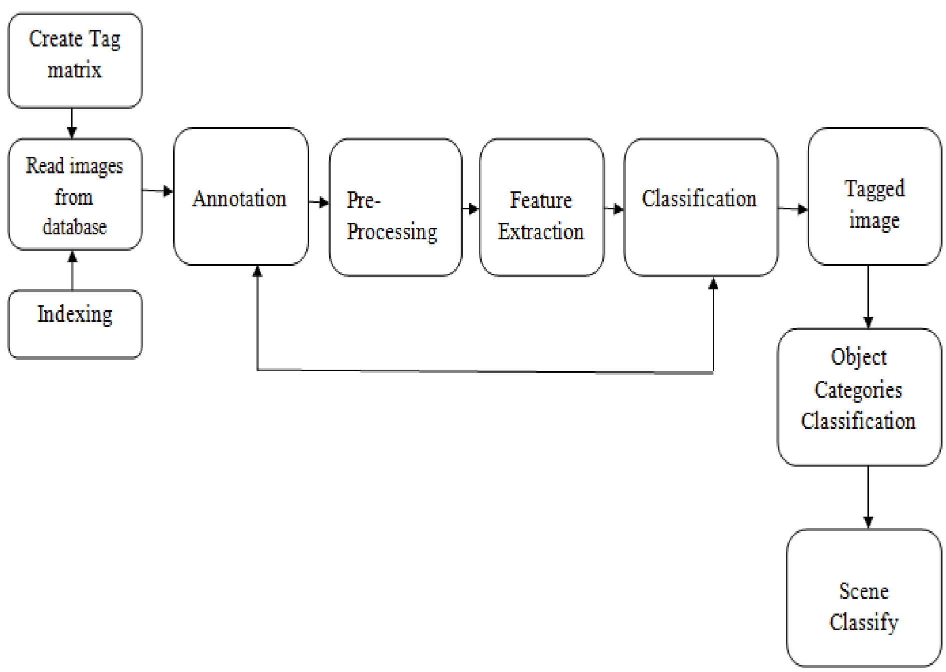

Figure 4. Proposed Approach Block Diagram

First, we created tag matrix then we taking one data/image it involved different objects and read the images from database. And creating indexing or data.

Input: Image I
Output: Tagged image I'
Proposed_Sys (I)
Begin
1.annotation=kNN(datasets)
//FV – Feature Vector
1. FV1=Color_moment(I)
2. rgbTogray(I)
3. FV2 = Shape(I)
//Shape extracts features from image I and stores in FV,
4. FV3=GLCM(I)
where FV = {FV1, FV2, FV3, ….., FVk} and k is variable
// SFV – Strong Feature Vector
SFV = ADABOOST(FV)
// ADABOOST takes out strong features of image I from FV and stores in SFV. Thus, SFV⊂ FV
5 = SVM(SFV)
End

5. RESULTS AND ANALYSIS

We have led examinations Previously, Matlab-2016a. Those examinations would ledutilizing three types of images: road scene dataset, basic auto dataset also complex office dataset [1]. Each dataset holds pictures hosting fluctuating level of impediment.

Figure 5. Surf+SVM Result

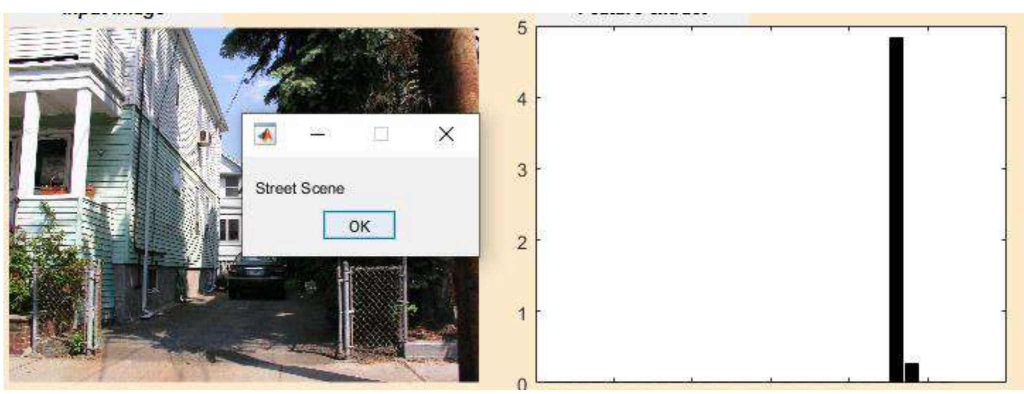

Figure 6. Color+Texture+SVM Result

In this section, first we present the performance for street scene images under high occlusion. Subsequently, we present the performance considering different sizes of images.

Table I: Analysis

Method	Confusion Matrix			Precision	Recall	Accuracy
Surf	0.9583	0	0.0417	0.9036	0.9583	68.99%
	0.0789	0.1579	0.7632	0.8716	0.1579	
	0.0233	0.0233	0.9535	0.5423	0.9535	
Color+Texture	5	0	0	0.8330	1.0000	92.31%
	1	3	0	1.0000	0.7500	
	0	0	4	1.0000	1.000	

6. CONCLUSION

In this study, we have explored and compared two distinct methods for object mining and labeling: SURF-based feature extraction and color+texture-based feature extraction. Our evaluation of these methods has been based on several performance metrics, including the confusion matrix, precision, recall, and overall accuracy.The results obtained from our experiments reveal interesting insights into the strengths and limitations of each approach. The SURF-based method demonstrates high precision and recall values, particularly for certain classes, indicating its effectiveness in specific scenarios. However, its overall accuracy remains slightly lower compared to the color+texture-based method.On the other hand, the color+texture-based method exhibits remarkable accuracy, surpassing 90% in our evaluation. This suggests that the combined use of color and texture features provides a robust framework for object identification and labeling within images.One notable observation is the disparity in recall values between the two methods. While the SURF-based approach shows relatively low recall for certain classes, the color+texture-based method achieves consistently high recall across all classes. This indicates that the color+texture-based method is more adept at capturing diverse object characteristics and improving overall recognition performance.

Overall, our findings underscore the importance of considering multiple feature extraction techniques and their impact on object mining and labeling tasks. The choice of method should align with the specific requirements of the application, balancing factors such as precision, recall, and overall accuracy.

Future research directions may involve further refinement of feature extraction algorithms, exploring hybrid approaches that combine the strengths of different methods, and conducting extensive testing across diverse datasets to validate the generalizability of our conclusions.By advancing our understanding of composite characteristic-centric entity discovery, identification, and labeling, we contribute to the broader field of computer vision, paving the way for enhanced image analysis capabilities and intelligent content management systems.

REFERENCES

1. C. Y. Lai, H. S. Jo, C. H. Lai, and M. Dunn, “Vision-Based Automated Parking Monitoring Using Geo-Tagged UAV Images,” in 2024 International Conference on Green Energy, Computing and Sustainable Technology (GECOST), 2024, pp. 136–140. doi: 10.1109/GECOST60902.2024.10474933.

2. P. Rajarajeswari, P. Ramakrishna, S. Sayyed, S. Shaik, and J. Vemuri, “Detecting Tags for Real-Time Objects in Images Using the Yolo-V3 Algorithm,” in 2023 4th International Conference on Smart Electronics and Communication (ICOSEC), 2023, pp. 1471–1477. doi: 10.1109/ICOSEC58147.2023.10276319.
3. F. Šikić, B. Filipović, Z. Kalafatić, M. Subašić, and S. Lončarić, “Multi-Class Price Tag Detection in Images of Supermarket Shelves,” in 2023 International Symposium on Image and Signal Processing and Analysis (ISPA), 2023, pp. 1–6. doi: 10.1109/ISPA58351.2023.10279709.
4. R. Kaur and G. Singh, “Automatic Image Captioning for Medical Diagnosis Report Generation,” in 2023 11th International Conference on Intelligent Systems and Embedded Design (ISED), 2023, pp. 1–6. doi: 10.1109/ISED59382.2023.10444545.
5. W. Li et al., “Object Detection and Localization by Passive RF Backscattering System,” in 2023 International Conference on Microwave and Millimeter Wave Technology (ICMMT), 2023, pp. 1–3. doi: 10.1109/ICMMT58241.2023.10276390.
6. S. Raavi, P. B. Chandu, and T. S., “Automated Recognition of Underwater Objects using Deep Learning,” in 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), 2023, pp. 1055–1059. doi: 10.1109/ICOEI56765.2023.10125839.
7. E. Sie and D. Vasisht, “RF-Annotate: Automatic RF-Supervised Image Annotation of Common Objects in Context,” in 2022 International Conference on Robotics and Automation (ICRA), 2022, pp. 2590–2596. doi: 10.1109/ICRA46639.2022.9812072.
8. S. Patil, S. Deshmukh, P. Yadav, and S. Chaudhari, “Smart Adware and Product Recommendation using Object Detection,” in 2022 IEEE International Conference on Data Science and Information System (ICDSIS), 2022, pp. 1–6. doi: 10.1109/ICDSIS55133.2022.9915851.
9. R. Dulal et al., “Automatic Cattle Identification using YOLOv5 and Mosaic Augmentation: A Comparative Analysis,” in 2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2022, pp. 1–8. doi: 10.1109/DICTA56598.2022.10034585.
10. A. Parker, F. Gonzalez, and P. Trotter, “Live Detection of Foreign Object Debris on Runways Detection using Drones and AI,” in 2022 IEEE Aerospace Conference (AERO), 2022, pp. 1–13. doi: 10.1109/AERO53065.2022.9843697.
11. S. Rahman, S. Khan, and N. Barnes, “Deep0Tag: Deep Multiple Instance Learning for Zero-Shot Image Tagging,” IEEE Transactions on Multimedia, vol. 22, no. 1, pp. 242–255, 2020, doi: 10.1109/TMM.2019.2924511.
12. L. Liao, Y. Song, and Y. Zhang, “Object Modifier Generation for Image Captioning,” in 2020 Chinese Automation Congress (CAC), 2020, pp. 52–57. doi: 10.1109/CAC51589.2020.9327182.
13. C. Lee, H. Choi, S. Muralidharan, H. Ko, B. Yoo, and G. J. Kim, “Machine Assisted Video Tagging of Elderly Activities in K-Log Centre,” in 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), 2020, pp. 237–242. doi: 10.1109/MFI49285.2020.9235269.

14. T. T. Zin et al., “Cow Identification System using Ear Tag Recognition,” in 2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech), 2020, pp. 65–66. doi: 10.1109/LifeTech48969.2020.1570625232.
15. A. M. Chowdhury, J. Jabin, E. T. Efaz, M. E. Adnan, and A. B. Habib, “Object detection and classification by cascade object training,” in 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2020, pp. 1–5. doi: 10.1109/IEMTRONICS51293.2020.9216377.