ISSUE 9 June 2020
مجلة العلوم الإنسانية والتطبيقية
Journal of Humanitarian and Applied Sciences
العدد 9 يونيو 2020

A Comparison of Three different Techniques for Object Recognition

Ahmed Almantsri [1] **Mohamed Alhamrouni [2]** **Gökhan SENGÜL [3]**

[1][2] Computer Science Dept. Faculty of Arts & Science Kasr Khiar, Elmergib University-Libya
[3] Department of Computer Engineering, Atilim University, Turkey

Correspondence to: [1] aaalmontasery@elmergib.edu.ly - [2] maalhamroni@elmergib.edu.ly

الملخص

يساعد التطور السريع في تطبيقات الحاسوب على تحسين كفاءة تقنيات معالجة الصور مثل التعرف على الأشياء من الوسائط المتعددة. خلال العقود القليلة الماضية، تم تقديم العديد من التقنيات من خلال إشراك المجالات المتعددة التخصصات المعتمدة على علوم الحاسوب كأدوات تصنيف. في هذه الورقة، استخدمنا ثلاث تقنيات مختلفة لتصنيف الصور والتعرف عليها K-Nearest Neighbors (KNN)، Support Vector Machines (SVM)، Earth Mover's Distance (EMD).

تتطلب هذه التقنيات استخراج المميزات المتعلقة بالأشياء، ولهذا الغرض قمناخوارزمية Histogram of oriented gradients (HOG). فيما يتعلق بمجموعات البيانات، فقد استخدامنا مجموعة بيانات COIL-100 كمجموعة بيانات معروفة لتجارب التعرف على الأشياء. قمنا بتقسيم مجموعة البيانات إلى سبع مجموعات فرعية. ثم قمنا باختبار ومقارنة الخوارزميات الثلاثة باستخدام هذه المجموعات الفرعية بشكل فردي. أخيرًا، قارنا النتائج ووجدنا أن SVM و EMDأكثر كفاءة على الرغم من أننا استخدمنا مجموعة فرعية كبيرة بينما يتأثر KNN وتنخفض كفاءته عندما يزداد حجم مجموعة البيانات.

Abstract

The rapid change in computer applications helps improving the efficiency of image processing techniques such as object recognition from multimedia. During the last few decades, many techniques were introduced by involving the interdisciplinary fields of computer science as a classification tool. In this paper, we used three different image classifiers techniques K- Nearest Neighbors (KNN), Support Vector Machine (SVM), and Earth Mover's Distance (EMD).

These techniques require feature extraction, such as the Histogram of Oriented Gradient (HOG) algorithm. Regarding the datasets, we used COIL-100 dataset as a well-known dataset for Object recognition experiments. We divided the dataset into seven subsets. Then, we tested and compared the three algorithms using these subsets individually. Finally, we compared the results and We found that SVM and EMD are more efficient even though we used a large subset while KNN is affected when the dataset gets larger.

keywords – Earth Mover's Distance, K-nearest Neighbors, Support Vector Machine, Object Recognition, KNN, SVM, EMD.

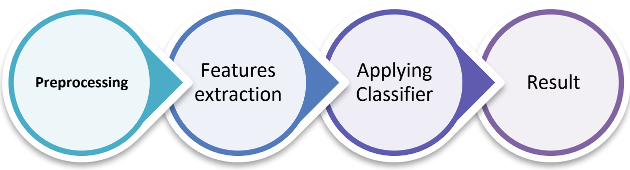
Introduction

During the last few decades, image processing concerns about utilizing advanced techniques to gain knowledge. The image file architecture depends on the coloring line by line, and this is called Raster. There are plenty of topics in image processing interest in extracting features for Recognition such as Face Recognition, Object Recognition, etc.

Object Recognition is related to computer vision as a Deep learning Technique which involves image processing methods in identifying objects from photographs. It begins with the preprocessing stage, as shown in figure 1. This stage includes importing target image and dataset, then resizing image files to reduce time consumption for classification even though it costs time for preparing input [1]. This stage ends with converting the image file into the grayscale mode or reducing the number of colors to ignore unwanted pixels.

Feature extraction for an object can be useful in matching others with the target image. There are many algorithms in machine learning which help in understanding images as the human brain does in drawing image contents based on shapes and colors to identify them. Feature Extraction is an essential step. The feature plays a vital role in the area of image processing [2].

Matching objects requires one of the classification methods. Data Mining provides us with some highly developed algorithms which we can use for this task. Before we dig in Object recognition, we have to define which feature extraction technique that we are going to use in this comparison.



Figue 1: The steps of Object Recognition.

Image Preprocessing

The preprocessing stage is the foundation of success in obtaining the desired results in image processing. It is the initial process that guarantees the rest of the comparison process because of its great importance. It includes the operation of fetching, modifying, and configuring inputs for the methods used in the matching process. It doesn't refer to a specific procedure, but it maintains images to be ready for extracting features and apply image processing techniques.

First of all, preparing inputs starts to choose the most qualified datasets that suit the applied technique. There are many datasets available online. The dataset should include different images with different angles to get the best results in the matching process of objects.

Image Resizing

As mentioned in [3], Image resizing is the most used image processing technique to take advantage of an image structure for any project. It modifies the resolution by insertion or reduction to gain the best result. The input file may differ in size. It can be either a small or large file. Since the image file is a set of colors, then reducing the number of colors will vary the accuracy of the result.

Moreover, reducing the image size will reduce processing time while we apply the image processing technique.

Feature Extraction

Feature Description has a significant impact on computer fields such as image processing, computer vision, pattern recognition, and machine learning. There are two image file features types; global and local features. When we manipulate all image content, it is a Global Features, but if we were focusing on a specific spot on the image file, then we call it local features [4].

Image preprocessing provides many different techniques such as thresholding, binarization, normalization, etc. an image file consists of pixels full of colors. These colors represent the features that make an object recognizable. After that, feature extraction techniques are applied to get features that will be useful in classifying and recognizing objects in the image file. Feature extraction techniques are helpful in various image processing applications e.g., character recognition. As features define the behavior of an image, they show its place in terms of storage taken, efficiency in classification, and obviously in time consumption also.

Histogram oriented gradient

In 2005 Dalal and Triggs [5] invented a Histogram of Oriented Gradients (HOG). It is a technique used for extracting features from the image file to help to detect the critical sectors and find the matched object [6]. HOG became one of the most useful tools in computer vision to gain the best solution for some significant problems such as object recognition by collecting features from a dataset of objects and classify them so they can be recognized [7]. HOG starts with dividing the image file into square cells. HOG descriptors deal with object appearance of the input file and describe it by the distribution of intensity gradients, and split it into small blocks or regions that are called cells. Then it starts specifying the direction for these cells to show how they are connected, and this procedure is called the descriptor, as shown in figure 2.

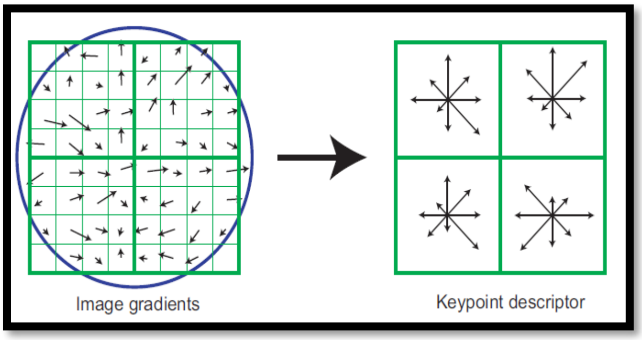


Figure 2: the mechanism of Histogram Oriented Gradient

Support Vector Machine (SVM)

In 1992, Vapnik, Boser, and Guyon introduced the Support Vector Machine (SVM). SVM is a learning method used for classification, especially for linear Classifying (see figure3). On the other hand, SVM is one of the machine learning tools used in classification and prediction. SVM is a widely used technique in different interdisciplinary fields in computer science. Computer vision employs SVM as a classification tool because it is one of the most useful learning methods in computer science. In brief, this technique aims to map feature which HOG has extracted in the input files and split them into the number of specified class for the target object that we need to match. Besides, this technique has used in various applications, such as face recognition and pattern classification [8].

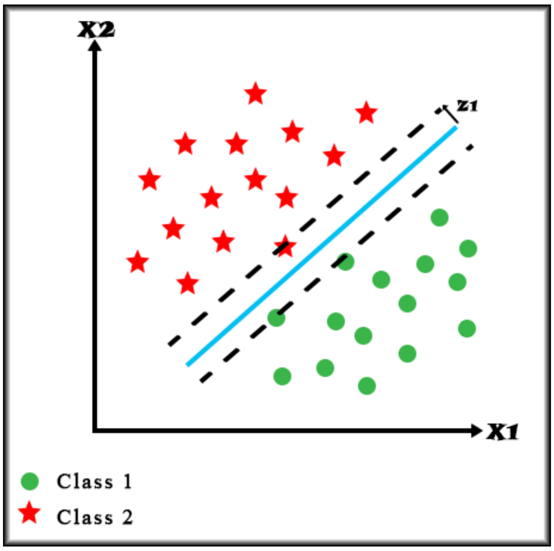


Figure 3: the classification of Support vector machine

Multi-Class SVM problem

Multi-Class SVM is a developed version of SVM. Researchers introduced many tools simulating SVM process to be multi-dimension classification which means classify for more than two classes; such as One-vs.-One approach [9], One-vs.-All [10].

One-vs.-All approach aims to classify each class against the rest of them, and this represents a binary classification for the classic SVM as the following: Class_1 vs. not

Class_1, Class_2 vs. not Class_2;......until last Class, in this case, Class_1 represents the positive samples, and the all other classes together represent the negative samples.

In the second case, Class_2 has trained up with not Class_2. Each case, Class_2 represents the positive samples, and all other classes together represent the negative samples, including Class_1, and the classifier continues up to the last class. Finally, the classifier selects a suitable class related to each tested sample.

K-Nearest Neighbor (KNN)

KNN classifier is a well-known classifier depends on calculating the distance to the centroid to determine which class does the object belongs to, based on this, KNN is a supervised learning method. Because of the simplicity of KNN, machine learning applications involve KNN as a classification method, regression, and pattern recognition [11].

Overall, KNN is easy to implement and highly efficient for applying classification techniques to solve computer vision problems such as Object recognition. The functionality of KNN depends on the extracted features by HOG and classifying them based on the specified centroids called classes, which represents training vectors (Figure 4).

Moreover, the KNN algorithm classifies the objects by three steps [12];

- Computes the distance between all training vectors and test vector.
- Chooses K closest vectors.
- Computes the average of closest vectors distances.

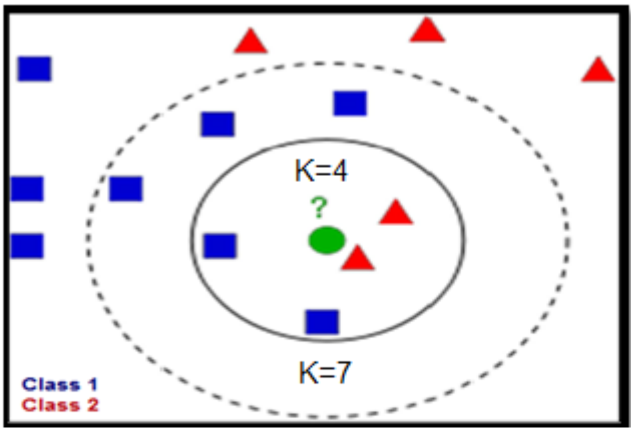


Figure 4 K-nearest neighbor method

In other words, in KNN, the output considered is membership of a class. Any new object is classified depending on the vote's number of its neighbors. In k nearest neighbors, if K=1, this means the object belongs to the class of that single nearest neighbor. However, in KNN, there is no particular way to select K; only we try to choose the best one. KNN doesn't need the training examples; it uses the training; therefore, it is a lazy learning algorithm. Moreover, it uses a training set directly to train and then classify an input when inputs and k value are given.

In the above figure, there are two sets, blue squares represent the first Class, and the red triangles represent the second Class, and these Classes are represented on a feature space. Features can be imagined as a space like a space that includes all the data, for example, two-dimensional space. So, these data have two features, which are two coordinates x and y, and these data can be represented in our space. For example, the data represented by a green circle is new, and we want to add it to the combined red & blue set, this step is called the classification. In order to do classification, there is a method to check the closest neighbor; in the figure, the red triangle is the nearest neighbor for the new data (green circle), therefore; the new data is added to Class 2. In this example, the classification process relies on the nearest neighbor; therefore, this technique is called the Nearest Neighbor.

Moreover, if we let K=4, this means four closest objects. Thus, there are two red objects and two blue objects; in this case, the new data will be added to red objects. But we let K=7, it means there are two red objects and five blue objects; thus, we must add the new data to the blue objects (Class 1). In the KNN algorithm, it is good to give importance to k neighbors; at the same time, we should give equal importance for the all. For example, if K=4, this means there are two blue objects (Class 1) and two red objects (Class 2), but the new data (green circle) is nearer to red objects than others, therefore, it is added to red objects (Class 2).

Earth Mover's Distance (EMD)

The Earth mover's distance (EMD) helps to find the distance among a set of distributions and map them based on the shortest distance. Mathematically, it is called Wasserstein distance [13]. EMD is a method to calculate the distance between a variety of data types, and some call them weights. In contrast, others call them mass based on the solution, for example matching histograms for extracting features of an object or semantic approaches such as word similarity.

Let’s say that we have two histograms A and B. the main purpose is to calculate the distance between them by generating the distance matrix D(i, j).

i represents the bin number in histogram(A).

j represents the bin number in histogram(B).

EMD computes the flow matrix F(i,j) to find the weight of transportation from i-th bin to j-th bin of both histograms (A,B). Based on this, it finds the flow matrix which helps to choose the minimum distance for every transportation between histogram A and B.

$$\mathrm{EMD(A,B)} = \min_{F} \sum_{i,j} F(i,j)D(i,j)$$

EMD generates the Work matrix for the result of the computed Flow matrix and the distance between two bins as an amount of work [14].

ISSUE 9 June 2020
مجلة العلوم الإنسانية والتطبيقية
Journal of Humanitarian and Applied Sciences
العدد 9 يونيو 2020

Therefore, EMD is a probability distribution to find the minimum amount of work required to convert one distribution into the other as shown in figure 5.

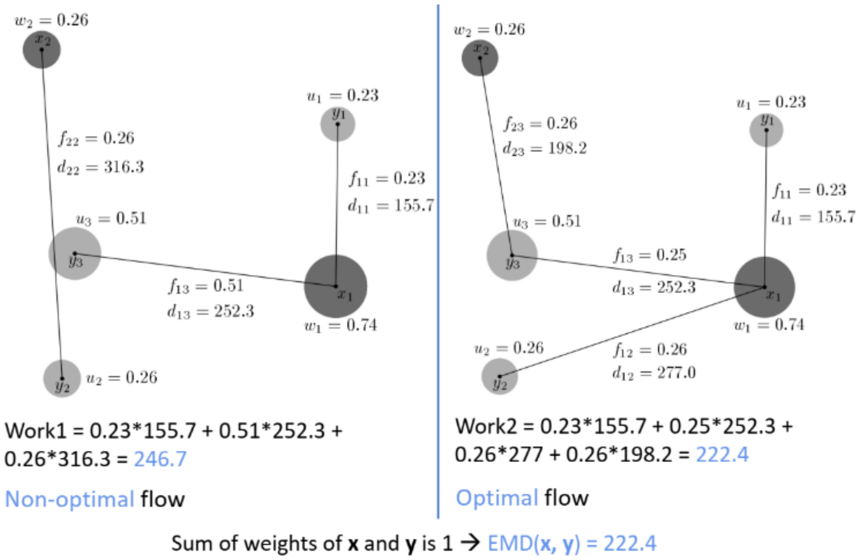


Figure 5: the probability distribution of calculating the work matrix of EMD (Cohen, 1999)

Experimental Result

The dataset for this experiment is the Columbia Object Image Library (COIL-100) [15]. It consists of a 900-object image. Each object includes ten images with different angles. It helps to measure the ability to recognize objects in different shapes. We divided this dataset into seven subsets that contains ten images for each object. These subsets are (20, 50, 100, 250, 500, 750, 900); for example, the subset of 20 has two objects, which means ten images for each object.

We started the experiment in this order KNN, SVM, and EMD, by choosing some random samples of objects so we can verify the outcomes of this comparison. Table 1 shows all the collected data from this experiment.

For KNN and SVM, we only matched the features of the target object to the remaining objects of the chosen subset. Since they are classifiers, the training sets represent the target image and testing sets representing the rest of the subset.

Table 1: shows the results of all three methods based on all subsets of objects.

	Dataset Size	20	50	100	250	500	750	900
Method	KNN	**100**	**100**	**100**	**91.11**	**83.56**	**78.07**	**79.38**
	SVM	**94.44**	**95.56**	**96.67**	**97.78**	**97.11**	**97.04**	**96.3**
	EMD	**100**	**93.3**	**95.55**	**91.08**	**91**	**97.76**	**91.08**

For KNN, SVM, in each subset, we used one image for each object as a test set, and the remaining as a training set. The output is a matrix equal to the number of objects in the training set. Every index has the class number, which represents the matching classes. Overall, we count each class in this matrix for every object.

The result illustrates the capability of every method to find the matched objects to the target. KNN provided high matching rates for a lower number of objects subsets, while SVM gave a steady result for all subsets, as shown in figure 5. On the other hand, EMD was able to recognize objects more accurately than KNN, with large subsets. The density of colors affects the result by identifying other items even though it shows the right number of matched objects.

This test was on the highest number of objects subset. We did test five times for every subset with a different object in both color and shape to guarantee the best result out of these methods. EMD matched all nine correct objects with two unmatched objects with a total of 11 objects, as shown in figure 6.

The unmatched images for the shewing gum were classified as the target image because we can tell it is very close in shape and color, especially after resizing.The feature extractions will track the path of the general shape of the object, which means the possibility of similarities between the objects, as in Figure 1, the similarity between the toy car and the chewing gum.

Overall, the result is containing all nine images with two incorrect objects.

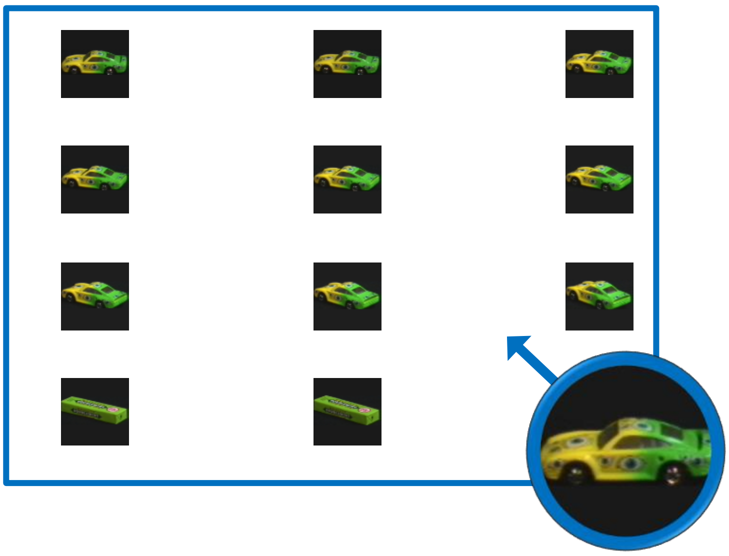


Figure 6: The experiment using EMD

The stable result for SVM proves the ability of classification even with a high number of objects, while KNN seems to be affected with lager subsets of objects.

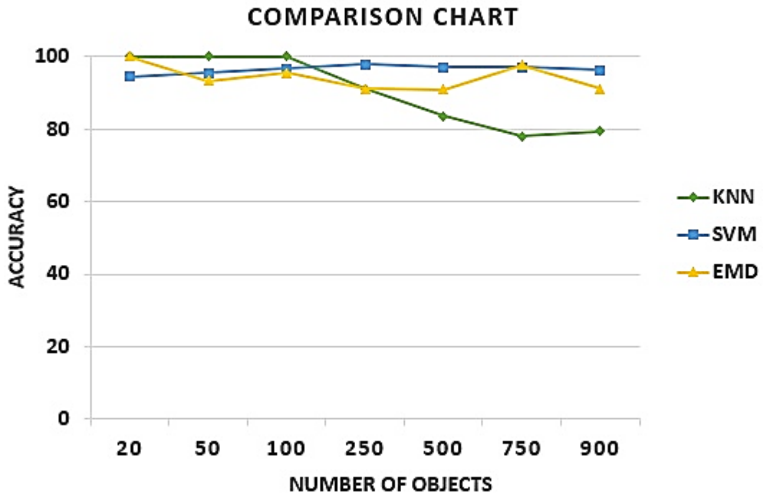


Figure 7: the chart of all results

Conclusion

In this paper, we have introduced a comparison between three image processing methods in object recognition. We applied them on the "Columbia Object Image Library (COIL-20)" since it provides objects and multiple images for each object in different positions. We noticed that three of them were able to extract the matched object with a small error percentage. We started with KNN and SVM since they both use the all-vs-all approach by using the training set as target objects and testing set as the remaining of the datasets. Based on the result, we found SVM and EMD are more efficient and useful for object recognition than KNN when it comes to a large number of datasets.

The richness of colors and type of background may affect the efficiency of classification. Moreover, the assumption of extracting objects from different angles shows unstable results.

For future lines of questioning, combining EMD with KNN or SVM as a double layer of classification may grant a more qualified object recognition method.

References

[1] J. d. Vries, Object Recognition: A Shape-Based Approach using Artificial Neural, The Netherlands: University of Groningen, 2006.

[2] G. K. a. P. Bhatia, "A Detialed Review of Feature Extraction in Image Processing Systems," in *2014 Fourth International Conference on Advanced Computing & Communication Technologies*, Rhtak, 2014.

[3] N. Z. T.-Y. L. F. W. Y. K. X. Z. Weiming Dong, "Summarization-Based Image Resizing by Intelligent Object Carving," *IEEE TRANSACTION ON VISUALIZATION AND COMPUTER GRAPHICS,* vol. 20, no. 1, pp. 111-124, JANUARY 2014.

[4] M. H. Ali Ismail Awad, Image Feature Detectors and Descriptors; Foundations and Applications, Switzerland: Springer International Publishing (Verlag), February 2016.

[5] N. a. B. T. Dalal, "Histograms of Oriented Gradients for Human Detection," in *IEEE Computer Society Conference on Computer Vision and Patterns Recognition (CVPR'05)*, 2005.

[6] C. H., *An ANN Approach to Identify Whether Drive is Wearing Safety Belts,* 2013.

[7] P. F. G. P. .. K. P. a. J. D. Mary, "Human Identification Using Periocular Biometrics," *Journal of Science, Engineering and Technology Research (IJSETR),* vol. 2, 2013.

[8] V. Jakkula, *Tutorial on Support Vector Machine (SVM),* School of EECS, Wachington State University, 2006.

[9] J. C. N. C. J. S.-T. Platt, *Large Margin DAGs for Multiclass Classification,* NIPS, 1999.

[10] R. ,. A. K. Rifkin, "In Defense of one-vs-all classification," *Journal of machine learning research,* no. 5, pp. 101-141, Jan 2004.

[11] G. Şengül, "Classification of parasite egg cells using gray level coccurence matrix and KNN.," *Biomedical Research,* vol. 27, no. 3, pp. 829-834, 2016.

[12] L. R. ,. B. Sudha, "Gait based gender identification using statistical pattern classifiers," *International Journal of Computer Applications,* vol. 40, no. 8, 2012.

[13] F. N. S. S. Sylvain Boltz, "EARTH MOVER DISTANCE ON SUPERPIXELS," in *Proceedings / ICIP ... International Conference on Image Processing*, 2010.

[14] S. Cohen, *Dissertation: Finding Color and Shape Patterns in Images,* Stanford University, 1999.

[15] A. N. S. K. N. H. M. S, "Columbia Object Image Library (COIL-100)," Technical Report CUCS-006-96, February 1996.