Kinetik: Vehicle Detection Using Percent Of Image Conversion To Binary Method K-Means
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 4, No. 3, August 2019, Pp. 197-206

Vehicle Detection Using Percent Of Image Conversion To Binary Method K-Means

First Author, Second Author
1. Amanda Prawita Ningrum, Indonesia
2. Rajendra Nohan, Indonesia

Article Info

Keywords:
K-Means, Detect, Image, Feature extraction, Testing

Article history:
Received, 18 December 2022
Revise, 23 Desember 2022
Accepted, -

Cite:
Google Scholar

*
Amanda Prawita Ningrum
Rajendra Nohan

E-mail address:
111202113646@mhs.dinus.ac.id
111202113301@mhs.dinus.ac.id

Abstract
In this journal, we make an image vehicle detection using the K-Means algorithm. Vehicle detection is the artificial intelligence that can help us in transportation highway systems like counter vehicles passing through the road on Eid Mubarak day etc. The object in this case is divided into six classifications there are car, motorbike, van, truck, and three-wheel. On the dataset vehicle is mostly an image of a car that we get from Kaggle. To solve vehicle detection problems such as poor vehicle detection and reduced detection accuracy, we provide a new vehicle detection with a dataset at kaggle. The clustering process consists of steps in which input images are transformed into morphometrics. The next step is to classify the image data using the K-Means algorithm. The images grouped by this detection are vehicles. The first step is to determine the randomly drawn mean or center point of two image data values in the database. If there is no data transfer, the group is considered stable and group creation is completed. Seven vehicle image data are used to test this application. And the result of our experiment on vehicle detection is about ***85.7 %*** *accurate.*

1. Introduction

Vehicle detection is the artificial intelligence system capable of detecting vehicles or not vehicles which can measure the density of vehicles. Vehicle transportation consists of many types like cars, motorcycles, buses, trucks, etc. According to data obtained from the Badan Pusat Statistik (BPS), in 2018, there were 26 million users of his four-wheeled vehicle, with 16 million passenger cars, 2.5 million buses, and 7.7 million freight cars, increasing year by year. Increased has. As the automobile industry develops, companies related to these vehicles may emerge.

Digital images or pictures have become part of multimedia media which plays a role in providing information and content visually. The image as a function in dimensional space, f(x,y) where x and y are the spatial coordinates and the amplitude at a certain (x,y) coordinate is referred to as the image intensity value. Digital images can be obtained from various media or image acquisition tools with different qualities. Each image has characteristics that are used to recognize an object, this identification can be done through image processing, for example, vehicle detection images, this identification can be done through image processing. Image processing is a method for manipulating and improving image quality in various ways. The main function of digital image processing is to improve image quality so that it looks clearer and is easily recognized by humans or machines

Morphological features are obtained through the shape of objects and can be expressed through contours, areas, and transformations. Morphological features are useful for object detection or recognition. Extraction of morphological features is calculated based on the edge of the object in the image or through the moment value. The method for carrying out object edge detection uses first-order and second-order edge detection, the results of object edge detection will be used in the next process, namely for morphological feature extraction. The edge of an object in an image can be defined as the boundary between two regions or two pixels that are close together and have a sharp or high-intensity difference so that they form the edges of the object.

In the spatial domain, masks are classified. The message is hidden by adjusting the brightness of the image. Its used in color photographs. Masking functions are used to mark areas in an image where messages can be inserted, but filter functions are used to insert values into marked areas. RGB is the most common color model used in digital images. Each color in the RGB color model represents the primary spectral components red, green, and blue. The RGB color model is an incremental color model in which red, green, and blue light are combined in various ways to produce a wide range of colors. The name of the model is derived from abbreviations of the three primary colors that add up, red, green, and blue.

Because it is simple, easy to implement, capable of large data clustering, dealing with outliers, and dealing with time complexity, K-Means is the most famous and widely used clustering method in many fields. The time is linear O(nKT) when there are n documents, K clusters, and T iterations. The K-Means algorithm's strength stems from its computational efficiency and ease of use. K-means is a clustering partitioning method used to divide data into groups. K-Means can minimize the average distance of each data to its cluster through iterative partitioning. MacQueen invented this method in 1967. A subset of moving object detection is moving vehicle detection. This is a critical component of the foundation of vehicle detection monitoring. The problem of moving object detection has been extensively studied and analyzed, and cutting-edge techniques such as adaptive thresholding algorithms and clustering means have been proposed. The goal of this task preparation is to assist in the definition of vehicle detection to make it more efficient and to obtain more accurate and realistic outputs based on input.

Yuslena Sari, Andreyan Rizky Baskara, and Puguh Budi Prakoso studied Applying the Distance-Based K-Means Method to Detecting Moving Vehicles at Lambung Mangkurat University in Banjarmasin in 2022. The topic of traffic sensors is related to smart traffic. System 40 is a computer application. The K-Means method is an inheritance-solving procedure that divides a dataset into K clusters by minimizing the sum of squares of distances within each cluster. The K-Means algorithm is used in this paper to implement the Euclidean distance, Manhattan distance, Canberra distance, Chebyshev distance, and Bray-Curtis distance. The goal of this study was to compare and evaluate how these distances performed in the K-Means cluster resolution procedure .According to the Manhattan distance, the smallest MSE value of 1.328 is obtained, and the Manhattan distance also constitutes the highest PSNR value, which is 21.14. The Manhattan distance, like SSIM, returns a value close to 1 or 0.83. As the SSIM value approaches one, it indicates structural similarity. In terms of PCQI results, the Manhattan distance has the best value when compared to other distances, with a value of 0.79.

The goal of compiling this task is to assist in the definition of media detection to make it more efficient and to obtain more accurate and realistic outputs based on the input. The authors used K-Means to achieve good detection results on vehicle detection images in this study. This study employs K-Means to process seven different types of RGB images. If there is a vehicle such as a car, motorcycle, van, truck, or tricycle, the image used in this study has a rather complex background.

2. Research methods

2.1 State Of The Art

The study titled "Media Detection Using Background Subtraction and Clustering Algorithms", researched by Puguh Budi Prakoso and Yuslena Sari in 2018, investigated the comparison between clustering methods such as FCM, OTSU, and K-Means. In our research, it can be concluded that K-Means and FCM algorithms are superior to OTSU algorithms. In PSNR of the 3 methods, FCM gives higher results than other methods and MSE also shows that K-Means and FCM have less error.

A 2018 study by Jun Sang, Zhongyuan Wu, Pei Guo, Haibo Hu, Hong Xiang, Qian Zhang, and Bin Cai titled "An Improved YOLOv2 for Vehicle Detection" discusses optimizing the YOLOv2 model using the K-clustering Means algorithm to achieve an accuracy of 94.78%. However, in this study, the vehicle type and amount of data tested were still low.

The study titled "Clustering of vehicle images (motorcycles and cars)" based on the form using the K-Means algorithm, studied by Margi Cahyanti, M Ridwan Dwi Septian, Ericks Rachmat Sweden, Ravi A Salim in 2018 in Pondok China, Depok discusses the K-Means algorithm engine for clustering digital image data of cars and motorcycles based on image, color and object detection. The purpose of this study is to improve the appearance of the boundary line of an area or an object in the image. To detect, represent and display images in electronic systems, such as televisions and computers, although they have also been used in conventional photography. The conclusion of this study is that the app performs well in displaying results by design. In addition, the application can also display the image processing steps from the input image to the Prewitt image and get the pixel value of the image to save in the database. The App Bundle Process also shows some of the process steps used to achieve the end result. The final focus at the team stage is also displayed in the application that has been created.

Muhamad Octorie Feisal Tamrin of Bandung National Institute of Technology conducted the study titled "K-Means Algorithm for Public Transport Eligibility Classifier" in 2021. Many researchers have proposed clustering methods such as K-Medoids, Fuzzy MADM, and Self-Organizing Maps. The K-Means algorithm can address the shortcomings of K-Medoids, Fuzzy MADM, and Self-Organizing Map (SOM), namely complex implementation and long processing times. Furthermore, the time required to complete the learning is relatively shorter. The goal of this research is to use a genetic algorithm to determine the initial initialization of the cluster based on the K-Means and Elbow methods, the Silhouette Coefficient, and the Davies-Bouldin Index (DBI) to determine the number of clusters. As a result of the emissions test, the best cluster in the public transportation feasibility group with the carbon monoxide (CO) and hydrocarbon (HC) parameters for the system to give more optimal results. The study's conclusion is that when determining the number of clusters using the elbow method, shadow coefficient, and Davies-Bouldin index with 155 total data points, the optimal number of clusters is 3. The test parameters obtained during the process of determining the cluster center by genetic algorithm are population size of 10, probability of hybridization of 0.9, probability of mutation of 0.1, and number of iterations of 10. The amount of '0' indexed data in the clustering process using the k-means algorithm is 34, the amount of '1' indexed data is is 18, and the '2' indexed data is 103.

The study titled "Suzuki Car Purchase Pattern Based on Location and Type Using K-Means Algorithm", was carried out by Dicky Febriyandani Riyanda in 2022 at Sultan Syarif Kasim Riau National Islamic University, Pekanbaru. Discussing methods to solve this problem by applying a data mining approach, there are several popular algorithms that can be used for clustering, one of which is the k-means algorithm. The aim of this study is to determine the purchasing habits of consumers using Suzuki auto sales transaction data at PT. Buana Trada thrives. To gain insights or insights from purchase transaction data that can be used as a reference in the decision-making process of business executives. The conclusion of this study is that the K-Means algorithm is faster in the clustering process than K-Medoids and lower accuracy value than K-Medoids.

The research entitled "Background Photo Editing Application Using the K-Means Clustering Method", which was researched by Ihsan Lubis, Herlina Andriani Simamora, Sayuti Rahman, Rosyidah Siregar, Husni Lubis in 2019 at Harapan University Medan, Jl. HM. Jhoni No. 70 Medan, Indonesia, discusses the color change in the background that can work if the image has the same background. The system is done using 9 images. The more clusters used, the better the clustering results obtained. The fewer clusters used, the more difficult the clustering results will be. The success rate in changing the background color is 100% if the image has the same background color. The purpose of this study is to minimize the objective function that is set in the clustering process, which seeks to minimize variations within a cluster and maximize variations between clusters. The conclusion in this study from the results of system implementation is that the k-means method is able to cluster images and separate objects and backgrounds properly. Changing the background color is said to be successful if the background has one color and vice versa if the background has a different color it is said to be immutable. The success of changing the color of the background with an image that has the same background using the k-means method is 100%. This application can be used properly in changing the background and it is easier for users to change photo backgrounds.

2.2 Vehicle

Vehicles are transportation that is useful for transporting people or cargo. Vehicles can include coaches, bicycles, motorized vehicles such as motorbikes, cars, trucks, or buses. Currently, the development is very rapid with the presence of various types of vehicles that are scattered throughout the world. Due to the existence of various types of vehicles that have different shapes, sometimes humans cannot distinguish cars with unique shapes. And vehicle users also continue to increase by 5% annually if you look at data from Badan Pusat Statistik (BPS). The latest record in 2020 is with a total of 136 million vehicles which include 15.7 million cars, 233 thousand buses, 5 million cargo vehicles, and 115 million motorcycles.

Figure 1. Car

Figure 2. Motorcycle

Figure 3. Truck

Figure 4. Van

Figure 5. Three Wheel

In our research this time using the dataset we obtained from Kaggle which has 3000 images in a random format that have not been grouped and 953 image data that have been grouped. It is divided into 5 categories with cars, motorcycles, three-wheelers, trucks, and vans. Data vehicles are entered when we run the training menu. And when we click on the testing program, it will run the test results from the image we want to detect, not forgetting there is a clustering process using the K-Means method.

Category	.jpg	.png	.jpeg
Car	292	-	-
Motorbike	-	10	68
Three Wheel	1	2	101
Truck	2	25	323
Van	168	4	167

Table Dataset by Image format

The data we have is divided into 3 image formats namely jpg, png, and jpeg with the numbers listed in the table above. It can be seen that the jpeg format dominates the dataset that we have.

2.3 Feature Extraction

Feature Extraction is feature extraction from one form and the value obtained is then analyzed for use in the next process. The aim of feature extraction is to find important feature areas in an image according to the intrinsic characteristics and their application. Regions can be defined in global or local contexts and differentiated by shape, texture, size, intensity, statistical properties, etc.

RGB feature extraction The feature extraction technique used by RGB is to convert the RGB matrix containing pixels into vector data. Image processing operations from RGB matrix to vector data begin by storing RGB data, namely the red, green, and blue components of the image because the pixels are treated 1 RGB value sets :

$$z = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

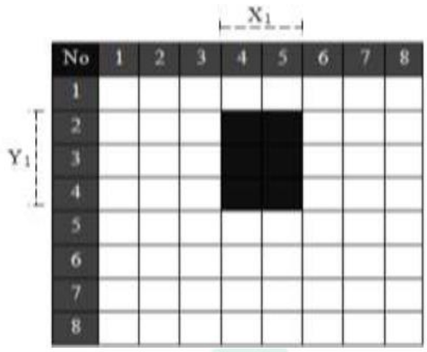
Where :

- "x1" is the intensity of red image pixels.
- "x2" is the intensity of the green image pixel.
- "x3" is the intensity of blue image pixels.

An RGB image with size XxY can be represented as 3 image components with a total XY image with a size of 3x1. The case of an image involving the y component will produce a y-dimensional vector. The description of the matrix to the vector is shown in the figure :

$$z = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ \cdots \\ x_n \end{bmatrix} \qquad z = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \qquad z = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{bmatrix}$$

This feature is carried out by extracting R, G, and B into vectors. An integral image is an image representation that has just been used efficiently. Integrated imaging facilitates calculation processes with a single and accurate scan. The pixel value is calculated from the input image passed by the vehicle feature. Each type of function in each box is a certain number of pixels, as shown below.



The variable needed to obtain vehicle characteristics is to count the number of pixels on the X and Y axes. Show that the X1 axis represents the width of a vehicle object, where the position of the point graph is at points 4 - 5 so that X1 = 2. To get the length of the vehicle object, see the Y axis For Y1, denotes the length of the vehicle object whose position is at points 2 - 4 so that Y1 = 3. Thus, one of the characteristics of the car obtained is the length of the vehicle object(Y1 =).2) and the width of the vehicle (X1 = 3).

Feature Extraction Shape Feature extraction is used by changing the spatial domain of an image into a frequency domain. The transformation is carried out using the 2D Fourier transform, especially the FFT shown in Equation 1. The result of feature extraction F is a frequency domain image, where each pixel (u,v) is obtained by transforming all the typical pixels of the image. The spatial images used are images of various types of vehicles. However, feature extraction F takes the form of complex numbers. To get the true value of the F feature, further feature extraction is performed as a spectrum (quantity/scalar) and angle (orientation). Both are obtained respectively according to Equation 2 and Equation 3.

$$F(u,v)=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} Img(m,n)e^{\left(-i2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)\right)} \quad (1)$$

$$S \rightarrow |F(u,v)|=\sqrt{\Re(u,v)^2+\Im(u,y)^2} \quad (2)$$

$$O \rightarrow \phi(u,v)=\tan^{-1}\left(\frac{\Im(u,v)}{\Re(u,v)}\right) \quad (3)$$

The final forms of the extracted features are spectral and orientation images. By changing S, O ∊ R2 → R and combining them, they find that for each line there is a point (s,o) where s∊S and o∊O. A feature vector can be represented by X∊R2.

2.4 K-Means Algorithm

K-Means clustering means that K represents the desired constant number of clusters, which in this case means that the mean homogeneity comes from a group of data defined as clusters, so K Means clustering is a descriptive or automotive data analysis method. The detection method is a data processing method with system partitions that uses unsupervised modeling. The K-Means method attempts to categorize existing data into groups with similar characteristics. Furthermore, some characteristics are out of sync with the rest of the cluster data.

The K-Means Algorithm is a problem-solving procedure that takes k input parameters and divides a set of n objects into k clusters so that the similarity between cluster elements is high while the similarity with the parts is low. Other clusters have a very small numerator. The similarity of items to clusters is determined by the close proximity of the objects using the cluster mean, or what can be considered using the cluster centroid. The fundamental idea behind K-Means is an iterative search for cluster centers. The distance of each benchmark from the cluster's center determines the cluster's center. The clustering procedure begins with the identification of the data to be clustered.$X_{ij}(i = 1, \ldots, n; j = 1, \ldots, m)$ where n is the number of variables and m is the amount of data grouped. At the start of the iteration, the center of each cluster is determined independently (optionally), $C_{ij}(k = 1, \ldots, k; j = 1, \ldots, m)$. The distance between each datum and each cluster center is then determined. To determine the distance to the data, I (X_i) in the middle of the k. cluster (C_k), named (d_{ik}), the Euclidean formula can be used, that is:

$$d_{ij} = \sqrt{\sum_{j=1}^{m}(x_{ij} - c_{kj})^2}$$

A data is a member of cluster J if its distance from the center of cluster J is the shortest when compared to the distances from the centers of the other clusters. Furthermore, the dataset belongs to each cluster. The new cluster center can be calculated by taking the mean of the data points in the cluster and applying the formula:

$$c_{kj} = \frac{\sum_{h=1}^{p} y_{hj}}{p}; y_{hj} = x_{hj} \in clusterke - k$$

The following formula is used to determine the distance between data and each centroid:

$$d(P,Q) = \sqrt{\sum_{j=1}^{p} \left(x_j(P) - x_j(Q)\right)^2}$$

Explanation:
D = document point.
P = data records
Q = midpoint or centroid data

The shortest distance between the center and the document determines the position of the document group. Another repeating formula is defined as follows

$$C\ (i) = \frac{x1 + x2 + x3 + \cdots .. + xn}{\sum x}$$

Explanation:
X_1 = Value of the 1st data record.
X_2 =Value of the 2nd data record.
X_n =Value of the N data record.
$\sum_x$ = The number of data records.

The K-Means algorithm, the most used clustering algorithm, is employed in this test. This algorithm's popularity can be attributed to its straightforward implementation and linear time complexity. This split procedure's dependency on cluster initialization makes it a drawback. This algorithm's fundamental use begins by placing each document in the best grouping possible using a distance metric. The vector of words that ought to be in the cluster's center is known as the centroid. Secondly, following the import of all papers in the cluster. Use the cluster document to recalculate the cluster's centroid. Finally, if the focus does not change, stop (to some extent). If not, go back to step 2.

2.5 Proposed Method

The research methodology used in the completion of this study includes. The study of literature at this stage is carried out by seeking information and finding knowledge related to this research, both through books, journals, the internet, and also other sources that support system design according to existing theoretical foundations. Collection of Data at this stage is done by collecting all the data, in the form of a data set to be aggregated. System Design or Modeling at this stage, an application design will be made to detect whether a vehicle is running while running in the GUI. Use the K-Means method to isolate detected histograms. Testing at this stage, a test is carried out on the system that has been built to see the output from each step that is carried out on the system.

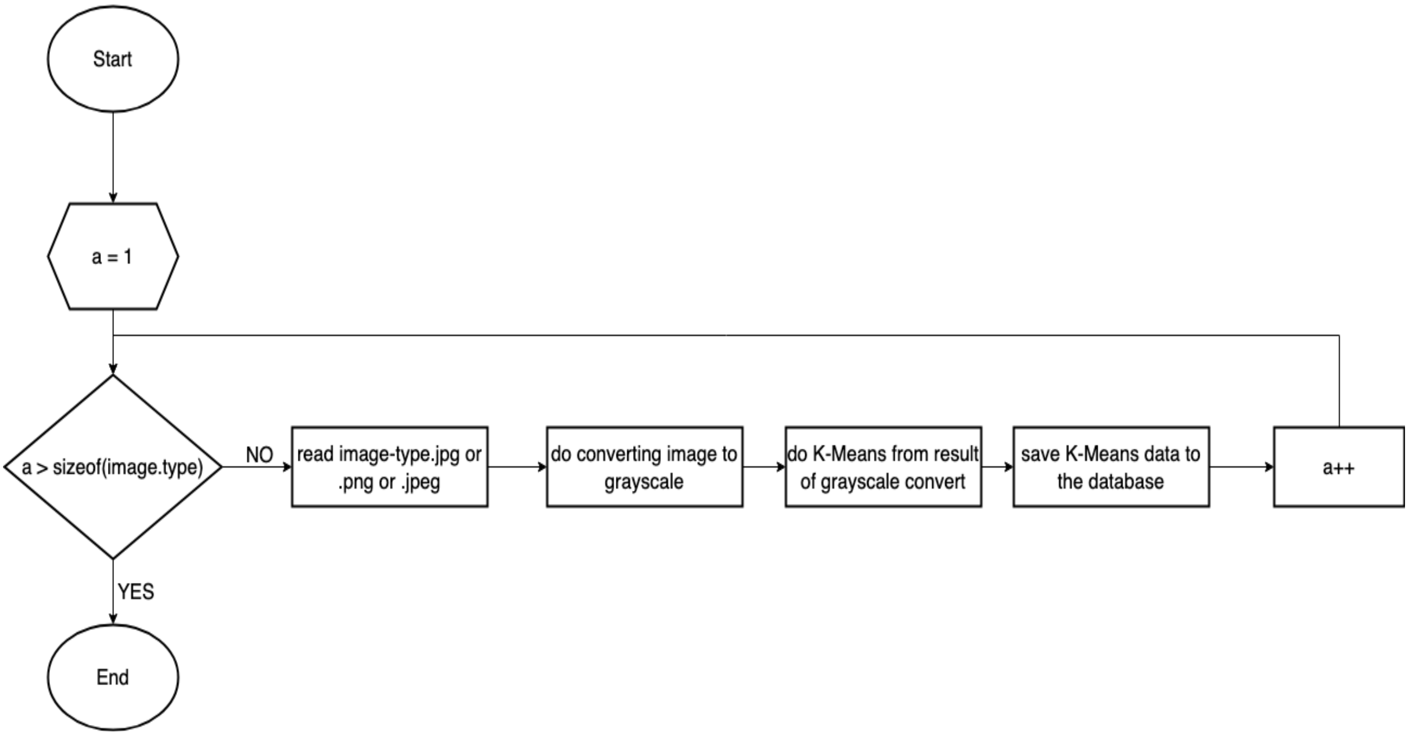
K-Means is a non-hierarchical clustering algorithm that aims to group data into clusters or groups in such a way that data with similar qualities are grouped together and data with different features are grouped together. Another cluster. The goal function defined during clustering must be minimized in order to maximize the variance between clusters while limiting variation within a cluster.

K-Means is a non-hierarchical clustering method that seeks to organize data into clusters or groups by grouping data with similar characteristics together and grouping data with dissimilar characteristics into clusters. another cluster To maximize the variance between clusters while limiting variation within a cluster, the target function established during clustering must be minimized.

$$D(i,j) = \sqrt{(x_{1i} - x_{1j})^2 + \square + (x_{ki} - x_{kj})^2}\ \square\square\square$$

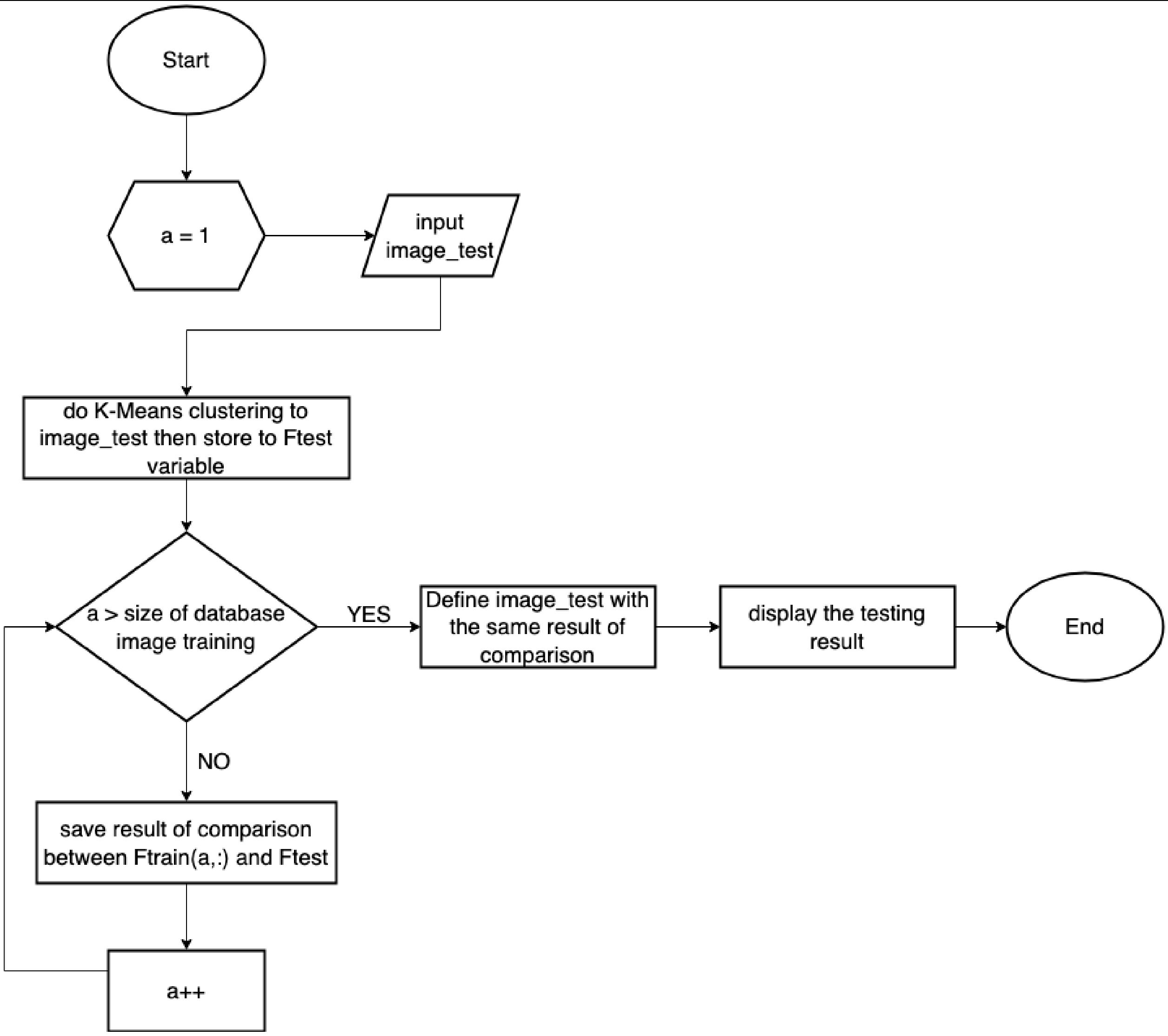
Explanation :
$D_{(i,j)}$ = The data distance from i to the center of the cluster j
$X_{(ki)}$ = Data from i on attribute data to k (Object Coordinates)
$X_{(ki)}$ = Center point of j on the kth attribute (Centroid coordinates)

Using the current cluster membership, recalculate the cluster center. The proprietary cluster's center is the uniform average of all data/objects. Rep with a new cluster center for each object. The clustering is resolved when the center of the cluster no longer changes. Alternatively, return to step 3 until the center of the cluster no longer changes.



The Flow Of The Training Algorithm Using K-Means

First, define a with the value 1 for initialization to for loop. And then decide to make for loop and finished on the size of the image with type format like .png, .jpg, or .jpeg. On every loop there is some process the first is read image, the second is to convert the image to grayscale, then calculate K-Means from the grayscale image and the last is to save the data to the dataset. Don't forget to increment a value to avoid an infinity loop.



The Flow Of The Testing Algorithm Using K-Means

For the testing make initialization with define a with 1 to make looping. After that input, the image to be tested don't forget to calculate image_test and store to Ftest variable. And then decide to make looping and finished the size of the dataset image training. On every loop, there is a process to save comparison between Ftrain and Ftest. Don't forget to increment a value to avoid an infinity loop. Lastly after out from looping there is some process like defining the image_test with the result of comparison and then displaying it.

3. Results and Discussion

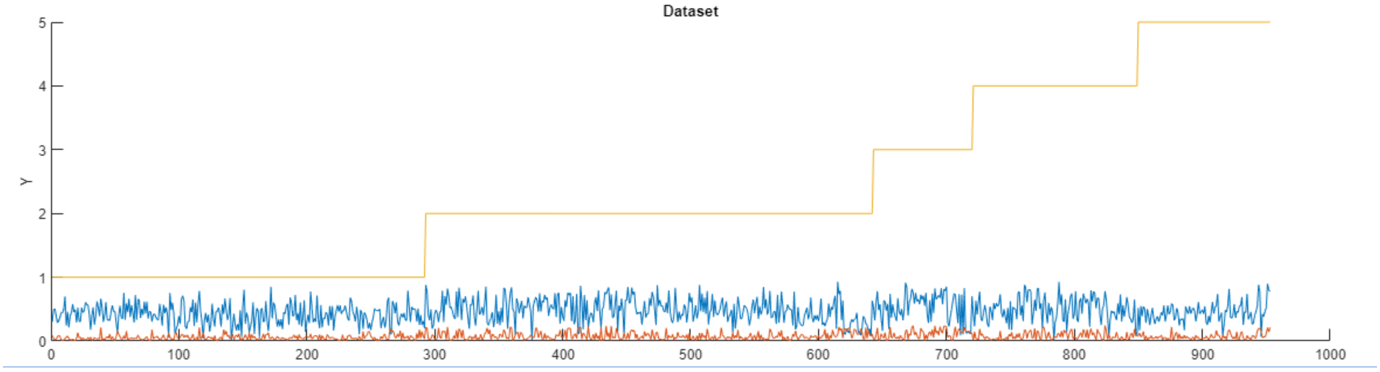
At this testing stage, there is a grouping of the images obtained from the dataset into 5 types of conditions, namely car, van, motorcycle, truck, and three-wheel. This grouping of images is made automatically so that it saves time in the testing phase which uses looping in it or what is commonly known as Data Clustering. At this stage, the data collected in the database will be processed using the K-Means algorithm.

In this process displays randomly selected data to determine the initial centroid, the selected centroid data is displayed in the table, then each image data will be calculated using the Euclidean Distance formula to the initial centroid. The calculated data is the edge data for each image stored in the database. The calculation begins with the 0th iteration until the last iteration provided that there is no group movement. Then the image grouping is processed and saved to the dataset to train the computer. In the iteration, we perform various stages of image processing before saving it to the dataset.

The first stage is to read the testing images that we take from a folder on the computer for further processing. The image from the computer was converted into a grayscale image for processing as raw material for the K-Means function we created. After conversion, the image is processed into the K-Means function earlier. It is the result of the output that we then save to the dataset. And in the test we enter the image that we will test for then the K-Means modeling process is carried out and we test our image data with the previously stored dataset. And in the final stage obtain results in the form of detection results.

3.1 Data Training

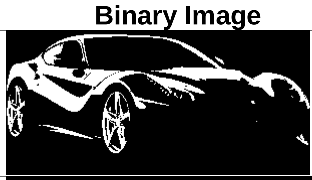
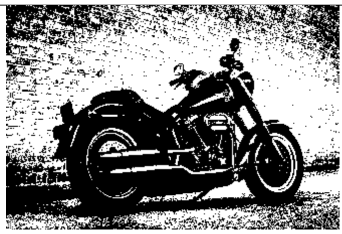
The following is the shape of the 2D distribution plot of the train data that we have where there are 953 images divided into 5 categories, namely cars, motorcycles, three-wheelers, trucks, and vans. Because the data is entered by looping, it can be seen that the datasets underclass tend to increase and flatten according to what we did in the data training process. The division of these categories can be seen by looking at the yellow color which is divided into 5 categories according to what we input. Categories that are included in the dataset include the 1st entry car, the 4th van, the 2nd truck, the 3rd motorcycle, and the 5th three wheel.



Display For Data Distribution

3.2 Testing

After experimenting 7 times on vehicle detection using K-Means in our research. Here we show our experiments:

No.	Original	Binary Image	Variable	Result
1.			M = 0.1511 S = 0.0917	Detect Car (**TRUE**)
2.			M = 0.3268 S = 0.2002	Detect Truck (**TRUE**)
3.			M = 0.4901 S = 0.0129	Detect Motorcycle (**TRUE**)

4			M = 0.3052 S = 0.0722	Detect Van (**TRUE**)
5			M = 0.8969 S = 0.2110	Detect Three Wheel (**TRUE**)
6			M = 0.6995 S = 0.0686	Detect Truck (**FALSE**)
7			M = 0.8087 S = 0.1447	Detect Van (**TRUE**)
ACCURACY				**85.7%**

From the 7 experiments that we did, we obtained an accuracy of 85.7% with some wrong results, including in experiment number 6. The experiment did not produce data that was in accordance with what we expected. Accuracy is not optimal because the dataset we have is still not diverse, especially in the distribution of images in each type of .jpg, .jpeg and .png images.

4. Conclusion

After testing the detection of vehicles by grouping images of cars, motorbikes, vans, trucks, and three-wheels using the K-Means algorithm. Using the data tested, there are 7 vehicle data, consisting of 2 car images, 2 van images, and 1 motorcycle, truck, and three wheels each. By obtaining the result that the vehicle detection is running well by displaying the design synchronous effect. In addition, the application can also display image processing stages starting from the input image to becoming an image and obtaining image pixel values to be stored in the database. The application grouping process also displays several process stages used to get the final result.

And at the end of the grouping stage, it is also displayed in the application that has been built. From the data tested, 7 vehicle image data were a car, motorbike, van, truck, and three-wheel and none of the images had an asynchronous effect from grouping. Grouping errors in this image are more caused by the stages before the image processing that is not right. It is necessary to add a stage before the image process similar to image normalization so that there is no hue that is too majority in each image before the image grouping process is carried out.

In this research, a classification process has been carried out to determine an object which is a vehicle of car, motorbike, van, truck, and three-wheel. The method used is the K-Means algorithm. This study uses the K-Means algorithm because it is an algorithm that is very widely used for clustering data. And very easy to implement. In this study, the K-Means algorithm works to determine whether the cluster image is a vehicle image or not. For the results after the training and testing data process, we get a training accuracy of 85.7%. Based on the accuracy that has been obtained both from training and testing, it still gets less than optimal accuracy. Because there is still a lack of data preprocessing techniques and parameters given to the K-Means algorithm used. This can be proven in the testing table attached to chapter 3.4, where the fruit images are still classified as vehicles, while in one image the car is not detected as a vehicle.

The research that has been done, still requires some adjustments, both from the preprocessing technique, the features used for training, and the parameters used for the K-Means algorithm. From the preprocessing technique used, it can be considered to add dilation and transformation morphological processes to the binary image so that the object under study can be recognized properly so that the K-Means algorithm process can be optimized. For the feature extraction used, you might consider adding the LBP extraction feature so that the K-Means algorithm can learn more data and can cluster data more accurately. In future studies, it is hoped that there will be an increase, both in terms of accuracy, the techniques used, or the parameters are given. So that the accuracy obtained can be further improved and maximized from the research that has been carried out.

5. Reference

[1] Alamsyah, D. (2018). Maximum Likelihood Classification with Fast Fourier Transform Feature Extraction for Car Recognition. *Jurnal Teknologi Dan Sistem Komputer*, *6*(1), 32–36.

[2] Sumaryanti, L. (2019). Ekstraksi Fitur Morfologi Menggunakan Metode Deteksi Tepi Pada Citra Digital. *MJRICT : Musamus Journal Of Research Information and Communication Technology*, *1*(2), 41–47.

[3] Sari, Y., Rizky Baskara, A., & Prakoso, P. B. (2022). *PENERAPAN METODE K-MEANS BERBASIS JARAK UNTUK DETEKSI KENDARAAN BERGERAK*. *9*(6).

[4] Setiawan, M. R., Sari, Y. A., & Adikara, P. P. (2019). *Klasifikasi Citra Makanan Menggunakan K-Nearest Neighbor dengan Fitur Bentuk Simple Morphological Shape Descriptors dan Fitur Warna Grayscale Histogram* (Vol. 3, Issue 3).

[5] Ro'isa, F., & Suartana, I. M. (2019). Implementasi Steganografi dengan Menggunakan Metode Masking and Filtering untuk Menyisipkan Gambar ke dalam Citra Digital. In *Journal of Informatics and Computer Science* (Vol. 01).

[6] Umar, R., & Riadi, I. (2018). *Sistem Identifikasi Keaslian Uang Kertas Rupiah Menggunakan Metode K-Means Clustering Identification System of Rupiah Currency using K-Means Clustering Method*. *17*(2), 179–185.

[7] Lubis, I., Simamora, H. A., Rahman, S., Siregar, R., & Lubis, H. (2019). Aplikasi Edit Foto Background Dengan Menggunakan Metode K-Means Clustering. In *Jurnal Sistem Informasi*.

[8] Sri Wisna, J. H., Matulatan, T., Hayaty, N., Informatika, J., Teknik, F., Maritim Raja Ali Haji Jl Politeknik Senggarang, U., & Author, C. (2020). *Jurnal Sustainable: Jurnal Hasil Penelitian dan Industri Terapan*. *09*(01), 8–14.

[9] Yunita, F. (2018). PENERAPAN DATA MINING MENGGUNKAN ALGORITMA K-MEANS CLUSTRING PADA PENERIMAAN MAHASISWA BARU (STUDI KASUS : UNIVERSITAS ISLAM INDRAGIRI). In *Jurnal SISTEMASI* (Vol. 7).

[10] Raihan, A., Dian Nathasia, N., Teknologi Komunikasi dan Informatika, F., Nasional Ps Minggu, U., Selatan, J., & Khusus Ibukota Jakarta, D. (2022). *ALAT PENGHITUNG JUMLAH KENDARAAN OTOMATIS PADA AREA PARKIR APARTEMEN BERBASIS INTERNET OF THING MENGGUNAKAN ARDUINO UNO*.

[11] Cahyanti, M., Ridwan, M., Septian, D., Swedia, R., Salim, R. A., Informasi, J. S., Komputer, I., & Gunadarma, U. (2020). *PENGELOMPOKKAN CITRA KENDARAAN (MOTOR DAN MOBIL) BERDASARKAN BENTUK MENGGUNAKAN ALGORITMA K-MEANS*.

[12] Yousif, B. B., Ata, M. M., Fawzy, N., & Obaya, M. (2020). Toward an Optimized Neutrosophic k-Means with Genetic Algorithm for Automatic Vehicle License Plate Recognition (ONKM-AVLPR). *IEEE Access*, *8*, 49285–49312. https://doi.org/10.1109/ACCESS.2020.2979185

[13] Chen, L., Ye, F., Ruan, Y., Fan, H., & Chen, Q. (2018). An algorithm for highway vehicle detection based on convolutional neural network. *Eurasip Journal on Image and Video Processing*, *2018*(1). https://doi.org/10.1186/s13640-018-0350-2

[14] Prakoso, P. B., & Sari, Y. (2019). Vehicle detection using background subtraction and clustering algorithms. *Telkomnika (Telecommunication Computing Electronics and Control)*, *17*(3), 1393–1398. https://doi.org/10.12928/TELKOMNIKA.V17I3.10144

[15] Li, X., & Xing, L. (2019). Use of Unmanned Aerial Vehicles for Livestock Monitoring based on Streaming K-Means Clustering. *IFAC-PapersOnLine*, *52*(30), 324–329. https://doi.org/10.1016/j.ifacol.2019.12.560

[16] Zong Chen, J. I. (2021). Automatic Vehicle License Plate Detection using K-Means Clustering Algorithm and CNN. *Journal of Electrical Engineering and Automation*, *3*(1), 15–23. https://doi.org/10.36548/jeea.2021.1.002

[17] Sang, J., Wu, Z., Guo, P., Hu, H., Xiang, H., Zhang, Q., & Cai, B. (2018). An improved YOLOv2 for vehicle detection. *Sensors (Switzerland)*, *18*(12). https://doi.org/10.3390/s18124272

[18] Mahto, P., Garg, P., Seth, P., & Panda, J. (2020). Refining Yolov4 for vehicle detection. *International Journal of Advanced Research in Engineering and Technology*, *11*(5), 409–419. https://doi.org/10.34218/IJARET.11.5.2020.043

[19] Ding, X., & Yang, R. (2019). Vehicle and Parking Space Detection Based on Improved YOLO Network Model. *Journal of Physics: Conference Series*, *1325*(1). https://doi.org/10.1088/1742-6596/1325/1/012084

[20] Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance