

Unsupervised Clustering for Image Data

Nickolas Vlahopoulos and Spiridon Kasapis

University of Michigan

Ann Arbor, MI

nickvl@umich.edu, skasapis@umich.edu

Jonathon M. Smereka

US Army DEVCOM GVSC

Warren, MI

jonathon.m.smereka.civ@army.mil

Geng Zhang

Michigan Engineering Services

Ann Arbor, MI

gengz@miengsrv.com

ABSTRACT

Clustering of image data is the process of sorting images into groups that exhibit similarities. During the clustering process, images are first reduced into feature vectors and then the statistics of their features are used for placing them into statistically similar groups. Systems that collect images during their operation (e.g. autonomous ground vehicles, satellites, etc.) create large data sets that ideally need to be sorted with minimal human effort. K-means is one of the most widely used methods for automatically sorting images. However, it is heavily influenced by initializations, the most important one being the need to know the number of clusters a priori. In order to overcome the latter shortcoming, validity indices have been used throughout the years to find the optimal number of clusters that the data should be separated in. The work presented in this paper comprises an Extension to the Variance Ratio Criterion (E-VRC) that when combined with the K-means can cluster image data of high content variance, without the need to input any information like the number of expected clusters, thus, it operates in an unsupervised manner. Comparisons with other available unsupervised methods is discussed in order to demonstrate the superior performance of the new E-VRC method and its benefits. Several publically available image datasets are used in the comparative studies. The robustness of the E-VRC method is also demonstrated by processing datasets with imbalances in their contents (i.e. many more images from certain clusters compared to the rest clusters) and by processing mixed datasets (i.e. comprised by very diverse types of images). It is concluded that the E-VRC does not dependent on initializations, does not care about the data dimensionality nor the content randomness and it is therefore a great tool for efficiently estimating the number of clusters and performing the clustering of image data.

ABOUT THE AUTHORS

Nickolas Vlahopoulos has been a Professor at the University of Michigan for 25 years and has also worked in the Industry for 7 years prior to his academic career; he has published over 100 papers and has graduated 24 PhD students.

Spiridon Kasapis completed his Bachelor's and Master's degrees in Worcester Polytechnic Institute. He obtained at the University of Michigan his second Master's degree and his PhD which was focused on Machine Learning applications in visual recognition. He has worked as an intern at the NASA Goddard Space Flight Center, where he helped characterizing the Van Allen radiation belt using the SDO satellite data. He is currently a Postdoctoral Fellow at the NASA Ames Research Center, working on Machine Learning for SEP Prediction and Detection of Solar Active Region Emergence.

Jonathon Smereka is the Ground Vehicle Robotics (GVR) Senior Technical Expert for research at the U.S. Army DEVCOM Ground Vehicle Systems Center (GVSC, formerly TARDEC) in Warren, Michigan. He received his M.S. and Ph.D. in Computer and Electrical Engineering from Carnegie Mellon University. He has worked in research and development of artificial intelligence, machine learning, image and signal processing, and computer vision with respect to biometrics, automated target recognition (ATR), and ground vehicle autonomy.

Geng Zhang is a R&D engineer at Michigan Engineering Services; he has 17 years of experience and has published 30 technical papers.

Unsupervised Clustering for Image Data

Nickolas Vlahopoulos and Spiridon Kasapis

University of Michigan

Ann Arbor, MI

nickvl@umich.edu, skasapis@umich.edu

Jonathon M. Smereka

US Army DEVCOM GVSC

Warren, MI

jonathon.m.smereka.civ@army.mil

Geng Zhang

Michigan Engineering Services

Ann Arbor, MI

gengz@miengsrv.com

INTRODUCTION

Clustering of image data is the process of sorting images into groups that exhibit similarities. During the clustering process, images are first reduced into feature vectors using a feature extractor [1]. The statistics of their features are used for placing them into statistically similar groups. Military systems that collect images during their operation (e.g. autonomous ground vehicles, satellites, etc.) create large data sets that ideally need to be sorted with minimal human effort. K-means [2] is a widely used statistical method of sorting a large number of vectors into a prescribed number of groups of similar members. The K-means method works well when there is a-priori knowledge of how many different types of members exist within the data set that is processed.

In order to overcome the shortcoming of having to specify the number of clusters in order to perform the sorting, validity indices have been used to determine in an unsupervised manner the optimal number of clusters in which the data should be separated in. Some of these are the Bayesian information criterion (BIC) [3], Akaike information criterion (AIC) [4], Dunn's index [5], Davies-Bouldin index (DB) [6], Silhouette Width (SW) [7] and gap statistic [8]. Recent advances in clustering techniques have shown that there are many ways to address the challenges of unsupervised clustering, including the need to determine the appropriate number of clusters and deal with high-dimensional and noisy data. For instance, some studies have proposed robust subspace clustering models using Mixture of Gaussians (MoG) noise modeling strategy [9], while others have proposed improved K-means algorithms that introduce ambiguity as a new constraint condition to reduce sensitivity to special data points [10]. Furthermore, there have been neural-processor-based k-means clustering techniques proposed to handle big data clustering over the network [11], and novel methods to reduce background-induced domain shift for adaptive person re-identification [12]. Additionally, some studies have proposed multi-modal 3D shape clustering methods, such as the dual contrastive learning network (DCL-Net), to discover the clustering partitions of unlabeled 3D shapes [13].

In this paper, we focus on the challenges of unsupervised clustering of images, particularly in the context of the K-means clustering method. We present a method that comprises an Extension to the Variance Ratio Criterion (E-VRC) and it is based on the Variance Ratio Criterion (VRC) method [14]. The new formulation enables the K-means to cluster image data, without the need to input any information, such as the number of true image classes. We use the term E-VRC for the new extension of the VRC method presented in this paper. Comparisons with the popular unsupervised methods discussed in the "RELATED WORK" section are presented showing the superiority of the proposed method and its unique capabilities. Finally, we demonstrate that the E-VRC is a robust method that a) is not dependent on initializations, b) does not care about the data dimensionality nor the content randomness and c) can be used as a tool for efficiently estimating the number of clusters and performing the clustering of image data.

RELATED WORK

In this paper, the performance of the new E-VRC method is compared to three other methods for unsupervised clustering: X-means [15], the U-k-means presented in [16], and a method that uses Principal Component Analysis

(PCA) for reducing the dimensionality of the feature vectors before incorporating the VRC criteria for determining the optimal number of clusters [17].

X-means [15]

The X-means uses the AIC and the BIC indices for determining the number of clusters in a dataset. This is done by incorporating a metric that measures how well the model fits the data but also by penalizing the model for having too many clusters. X-means is a hierarchical clustering method that starts with a user-defined lower-bound number of clusters K and continues to add cluster centroids where needed until a prescribed upper bound is reached. For each number of clusters the original K-means algorithm is used for conducting the sorting (Improve-Params operation). The additional centroids appear when -based on the BIC or AIC metric- the algorithm decides that a cluster needs to split (Improve-Structure operation). The algorithm oscillates between the Improve-Params and Improve-Structure operations until the upper bound K is reached. During this process, the number of clusters that achieves the best score in the X-means metric and the associated sorting is recorded and outputted as the final result.

U-k-means [16]

The unsupervised k-means (U-k-means) algorithm automatically determines the optimal number of clusters for a given set of data points without relying on any parameter selection by the user. This is achieved by optimizing a version of the k-means cost function augmented with two entropy terms. One term produces a penalty when there are too many clusters and the other produces a penalty when the clusters are imbalanced. The U-k-means algorithm is unique because, although it is based on the K-means, it does not utilize the algorithm itself at any point. Instead, the solution is achieved by equating the Lagrangian of the augmented cost function to zero. Although unique in nature, the U-k-means algorithm performs well in a very strict data domain (i.e., low number of true clusters and data dimensionality). In [16] the algorithm was tested only on low-dimensional data (2D and 3D), which were grouped into a relatively low number of true clusters (14 at most).

PCA-VRC method [17]

A high quality clustering is expected to simultaneously have the cluster centroids far apart from each other and the cluster members close to the respective centroids. The VRC indicator is comprised by two terms: the attractive term W and the repulsive term B . For any clustering W is evaluated as:

$$W = \sum_{k=1}^K \sum_{i=1}^{n_k} \|X_i - \bar{X}_k\|_2^2 \quad (1)$$

where K is the total number of clusters, n_k is the total number of members X_i in each cluster k and

$$\bar{X}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} X_i \quad (2)$$

is the centroid of each cluster k given the total number of cluster members n_k . The repulsive term B is evaluated as:

$$B = \sum_{k=1}^K n_k \|\bar{X}_k - \bar{X}\|_2^2 \quad (3)$$

where \bar{X} is the centroid when considering all members n in the entire dataset. A good clustering arrangement means that the distance W is minimized while B is maximized. Therefore, the VRC indicator is defined as [14]:

$$VRC = \frac{B(n-K)}{W(K-1)} \quad (4)$$

where n is the total number of points within the dataset and K is the number of clusters that is being considered in each iteration of the method. The authors in [14] suggest that in order to find the optimal number of clusters, clustering arrangements with $K = 2, 3, \dots, K_{max}$ are considered and the number of correct number of clusters is selected when the VRC becomes maximum. In this manner, the distance W is minimized while B is maximized globally across all

data points. Reference [17] used a PCA for reducing the dimensionality of the data before employing the VRC approach for selecting the appropriate number of clusters and considering the associated sorting as the final clustering.

EXPANDED VARIANCE RATIO CRITERION (E-VRC) METHOD

In this paper, we address the difficulties faced when the VRC is used in computer vision and image clustering (i.e. the VRC does not scale for high dimensionality and high true cluster number datasets). The VRC index is a quantity describing the degree of inter-cluster separation (Equation 3) and intra-cluster homogeneity (Equation 1). In this research we have identified two cases where the VRC index loses its ability to accurately point to the correct number of clusters: a) when the dimension of the features space is big (i.e. when image features have hundreds of elements) and b) when the image content diversity is big (i.e. when datasets include more than 15 image classes). We introduce the normalization of image features and the addition of an exponent term to Equation (4). These two changes to the VRC index calculation can remedy both aforementioned problems.

In [2] “Survey of Clustering Data Mining Techniques”, it is observed that vectors used in clustering algorithms are known to work effectively for dimensions below 16, while for dimensions above 20 their performance degrades to the level of sequential search (with some exceptions). Two solutions to the problem of high dimensionality have been proposed. The first is domain decomposition which is used when there is large data with many clusters and there is no actual dimension reduction [18]. The second is attributes transformations which are simple functions of existent attributes with popular methods being the principal components analysis (PCA) [19, 20], Singular value decomposition (SVD) [21, 22], Low-frequency Fourier harmonics in conjunction with Parseval’s theorem [23], wavelets and many other kinds of transformations [24].

Both approaches serve as “band-aids” to the core problem of the method (they can be useful in some situations but ultimately do not address the fundamental problem) and therefore are often problematic since they produce clusters with poor interpretability or lead to loss of information. The results presented here demonstrate that we can normalize the image feature vectors using the Euclidean norm ($X_{norm} = X/|X|$). This normalization enables the VRC method to solve multi-dimensional problems, as the direction of the vectors in the feature space is necessary for the solution of the problem whereas the magnitude (dimension) is not. Therefore, the first shortcoming of the VRC method is addressed by reducing dimensionality when only interested into the direction, but without losing information.

The second limitation of the VRC method is that it does not scale for datasets with a higher number of classes. This is because of the nature of Equation (4) which was devised with low values of K in mind. For datasets that do not contain many data points and that should only be split into a handful of clusters, the B and W multipliers are relatively balanced. This does not hold true when the desired number of clusters is increased. By introducing an exponent term $p < 1$ in the K term of the B multiplier, as shown in Equation (5), we restore this balance by diminishing the effect that a large K value has in reducing the numerator.

$$EV = \frac{B(n-K^p)}{W(K-1)} \quad (5)$$

Where EV is the value of the E-VRC criterion. The next section (Robustness) demonstrates how the performance of the new metric is robust over a range of p values when processing a diverse variety of datasets. The Euclidean normalization of the features in combination with adding the exponent in the VRC equation are the two key aspects of the new E-VRC algorithm presented in this paper. Algorithm 1 summarizes the operation of the E-VRC method. The range of robust values for the exponent p is determined and their effects are discussed in the Robustness Section. In all other tests presented in the paper the value remains constant $p = 0.3$. For any given image dataset, with $K_{start} = 2$ and $K_{stop} \leq n$ Algorithm 1 performs the unsupervised clustering.

Given a dataset of images, the E-VRC algorithm determines the number of clusters in the dataset and at the same time it places the images into clusters. We tested the E-VRC algorithm using two popular image datasets, ImageNet1000 (noted as ImageNet) [25] and Caltech256 (noted as Caltech) [26]. These datasets are representative of industrial applications but available in the public domain. We selected the first 50 classes from the Caltech256 dataset, as they are diverse enough, whereas we selected to use every 10 classes from ImageNet as the dataset classes are ordered based on the nature of the image contents (the classes list begins with wildlife content and continues to general imagery objects). A third “Mixed” data set is created and used in the testing. Half of the images originate from ImageNet

while the other half originates from MIT Places [27]. Figure 1 presents representative images from the three datasets used in this paper. The determination of the upper limit of clusters (K_{stop}) is not of high importance, as in other hierarchical clustering algorithms, as the performance of the E-VRC parameter in Figure 2 shows when a build-in termination criterion (if statement in Algorithm 1) is used.

Algorithm 1 Extended Variance Ratio Criterion (E-VRC)

```

1: Produce Image Features using ResNet34
2: Euclidean Normalization of Features
3: for  $K = [K_{start}, K_{stop}]$  do
4:   Perform K-means using  $K$  and Features
5:   Calculate E-VRC using K-means clusters
6:   Record E-VRC Value (EV)
7:   if  $EV_K \leq [EV_{K-1}, EV_{K-2}, \dots, EV_{K-10}]$  then
8:     Break For Loop
9:   end if
10: end for
11: Find  $K$  where EV is maximum ( $K_{max}$ )
12: Perform K-means using  $K_{max}$  and Features
  
```



Figure 1. Representative Images of the Datasets Used for Testing

The features X used in Equations (1) and (3) were produced using a ResNet34 feature extractor that was trained on the entirety of the ImageNet dataset. ResNet34 is a popular neural network that produces 512 dimension image vector representations. This paper focuses on performing the unsupervised clustering once the features for the images are available. Therefore, for simplicity and replicability of our results, we use the PyTorch [28] pre-trained ResNet34 model for this application. Figure 2 shows the variation in the E-VRC index for values of $K = [2, 100]$ ($K_{start} = 2$ and $K_{stop} = 100$ in Algorithm 1), although the algorithm is terminated earlier, as soon as no further increase in the E-VRC index is detected. By plotting the E-VRC index over a much longer number of clusters, a consistent decay of the E-VRC index after reaching its maximum value is demonstrated. Ten different cases are being explored, for ten different true cluster (classes) values: $K_{true} = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50$. Each class included 50 images that were randomly selected from a pool of more than 1200 images. The vertical lines indicate the maximum E-VRC value for each K_{true} case. One can observe that the E-VRC index predicts the true number of clusters in each case with very high accuracy. For the $K_{true} = 10, 15, 20, 25$ cases, the algorithm predicts the exact number of true clusters, whereas for the $K_{true} = 5, 30, 35, 40, 45, 50$ cases, it predicts a number of clusters very close to the true one. It is important to note that the results presented in Figure 2 present a favorable scenario since the feature extractor of the pre-trained ResNet34 used for generating the feature vectors of the ImageNet data processed by the E-VRC was also trained using ImageNet data. Therefore, the quality of the feature vectors used in the clustering is expected to be high in this case.

Throughout this paper we use the Normalized Mutual Information (NMI) as a metric for determining the quality of clustering. The NMI is a normalization of the Mutual Information score which scales clustering results between 0 (for clustering arrangements that present no mutual information) and 1 for perfect correlation. It contains information about how well the number of clusters is assessed and how correctly are images placed in the correct clusters. We selected this metric as it is normalized, allowing us to compare the clustering arrangements of different numbers of clusters and pictures. The NMI is a popular external measure [29-31], meaning that the true cluster labels should be known for its calculation, which is information that we do have in our test studies. Along with the

comparison between the predicted and true number of clusters (ΔK) in each run, the NMI is used to evaluate the performance of the E-VRC.

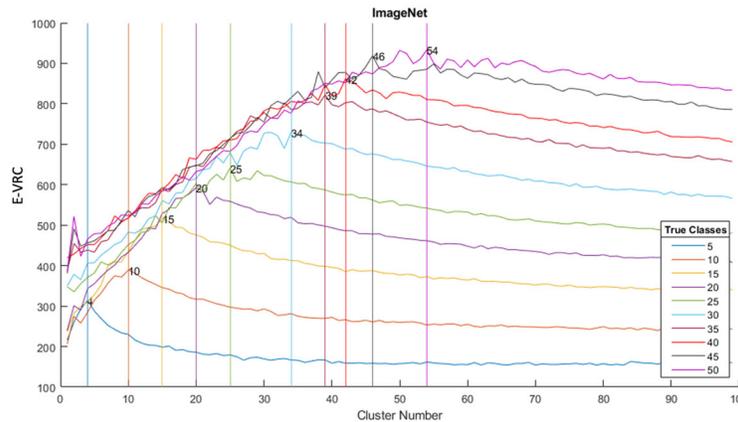


Figure 2. Variation of E-VRC Index for the Ten Different True Cluster Number Cases

Figure 3 shows in green the NMI values of the clustering generated using the E-VRC method. We compared the quality of the E-VRC method by clustering the same datasets with the three other unsupervised clustering methods discussed in the paper, namely, the PCA VRC [17], the X-means clustering algorithm [15] and the U-k-means method [16]. Each of the respective lines tracks the mean value for each case. The red line indicates the NMI accuracy obtained using the K-means method with the true number of clusters as an input. The red line indicates the best possible clustering performance. The closer an unsupervised clustering method is to the red line, the better it performs. For each determined K_{true} case, ten different runs were performed using 50 different images from each class to create the testing dataset in order to quantify the uncertainty and the noise in the data. As observed in Figure 3, the E-VRC method outperforms in almost all cases (except for $K_{true} = 5$) the other three unsupervised clustering methods while also producing results similar to the ones produced by a K-means method that uses the true number of classes as an input. The K-means method is considered to be the best case scenario, as the number of classes is not calculated by the algorithm itself, but is given as an input.

Supervised K-means results (red line) provide the best possible performance

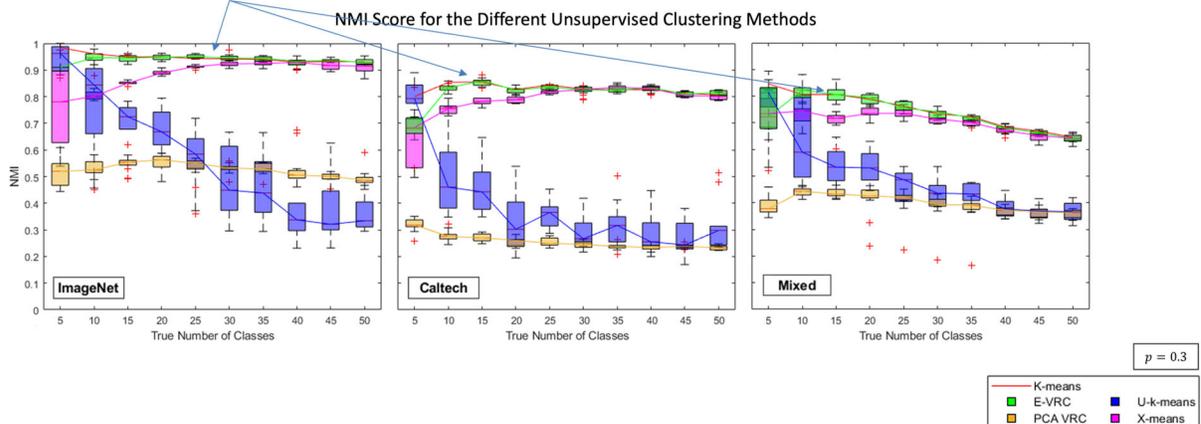


Figure 3. Box Plots for the Four Different Unsupervised Methods' NMI Results

Regardless of the true number of clusters, all methods are evaluated without changing the input parameters of the algorithm. Both X-means and U-k-means can potentially yield better results if in each case the algorithm input parameters are adjusted based on the knowledge we have regarding the dataset (i.e., the number of classes in the dataset and the number of pictures in each class). The PCA VRC method performs poorly due to the fact that significant information loss occurs when reducing a 512 long feature map to a 3D vector. The success of the E-VRC method lies on the fact that it scores well without the need to calibrate the algorithm based on the number of classes

or images per class. It is important to note that the NMI spread in the E-VRC method is significantly lower than the rest of the methods (especially U-k-means).

Figure 4 presents the respective number of clusters predicted in each run performed in Figure 3. Bubbles present the cases where different runs predicted the same number of clusters. As mentioned in the Related Work Section, U-k-means performs well when presented with low K_{true} datasets, but fails to scale for higher class variability datasets. On the contrary, X-means is a superior method for large datasets that include numerous data points and classes but tends to overestimate K when presented with smaller datasets. This can be remedied by changing algorithmic parameters such as the minimum cluster membership number or the maximum predicted K . However, parameter optimization based on prior dataset knowledge defeats the purpose of an unsupervised method. In these results the X-means performs well for a larger number of K_{true} only because the upper bound of clusters was set to be 50 in the runs conducted in this paper. As it can be observed, the X-means prediction for the number of clusters tend to gravitate towards this upper bound. The PCA VRC class prediction seems to be highly random due to the low feature quality in contrast to the E-VRC method which tends to predict correctly -with a small margin of error- the number of clusters.

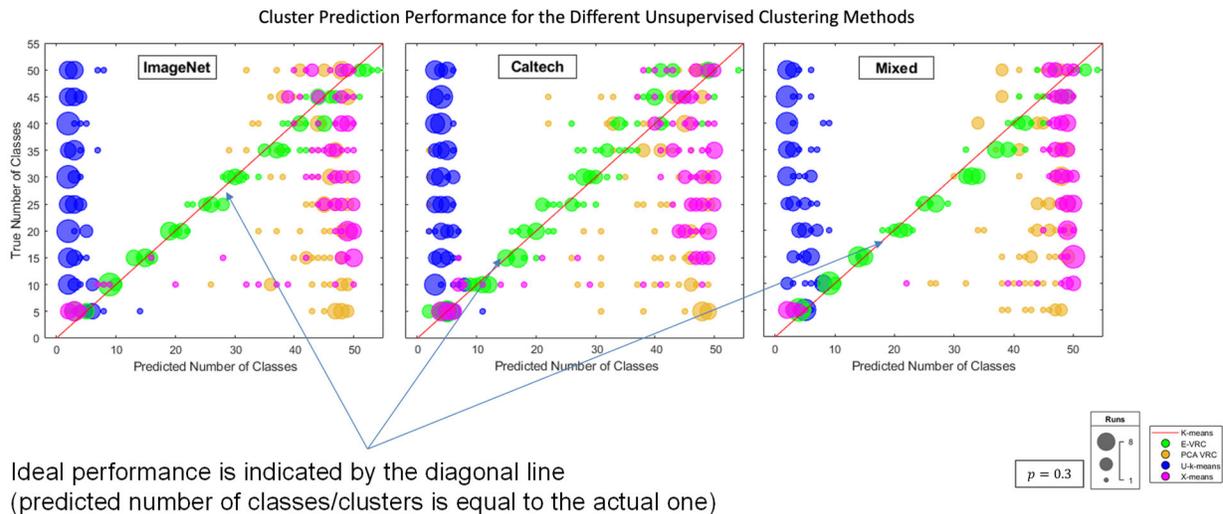


Figure 4. Bubble Plots for the Four Different Unsupervised Methods' Cluster Prediction.

ROBUSTNESS

The strength of the E-VRC method is that regardless of the size of the dataset or the type of images, the algorithm does not rely on user defined inputs. This means that both the normalization performed after the feature extraction and the selection of the p exponent can be universal in such a way that they do not need to be adjusted based on the number of classes or type of images that are processed.

Exponent Variation

To demonstrate the robustness of the method, we show that the same exponent p can be used for different datasets and that values between 0.1 and 0.5 provide stable results. This shows that the algorithm does not need to be tuned for every different case it is presented with. In all the case studies performed in the previous Section, the value of the p exponent was set to be equal to 0.3. Figure 5 explores the cluster number predictive capability of E-VRC for p values ranging from 0.1 to 1. Figure 5 shows the cluster prediction error for the different values of p . For all K_{true} cases (10 runs each with randomly selected images from each class), we calculate the predicted classes deviation $\Delta K = |K_{\text{pred}} - K_{\text{true}}|$ and present the variation using box plots. As observed, for the ImageNet dataset exponent values of $p = [0.1, 0.6]$, for the Caltech dataset values of $p = [0.1, 0.5]$, and for the Mixed dataset values $p = [0.1, 0.3]$ succeeded in predicting the correct number of classes with a relatively small margin of error. The importance of using high quality features for clustering is demonstrated since the Caltech image features -produced by a feature extractor trained on a different dataset- present a slightly lower range of exponents yielding good results, and higher ΔK variability. The

Mixed dataset comprises the most challenging one of the three since no scenery type of images were included during the training of the pre-trained feature extractor used for generating the feature vector X for the images employed in the tests. This paper focuses on performing the unsupervised clustering once the features for the images are available.

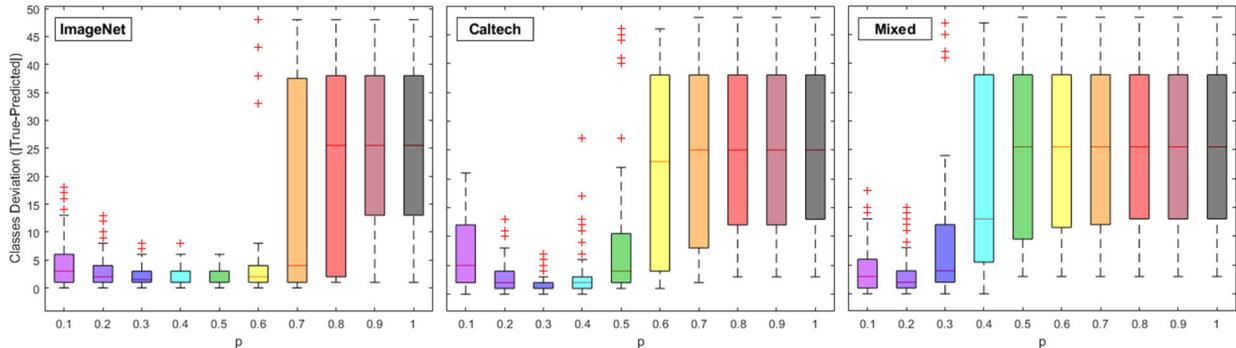


Figure 5. Box Plots for the Predicted Classes Deviation ΔK , for the Ten Different Exponent p Values Tested.

Images per Class Imbalance

For all the different cases studied above, although variability was introduced in the number of classes (i.e., different tests had different numbers of clusters), the number of images within each class was constant (50 images per class). However, in an operational setting the different classes would not contain the same number of images. To demonstrate that the E-VRC method can be used in an operational clustering setting, we introduce a certain amount of image imbalance between the different classes, showing that the method can cluster datasets that include random number of images per class.

The results presented in Figure 6 are similar to the results presented in Figures 3 and 4. Each K_{true} case is run 10 different times using different random images from each class. The difference here is that instead of each class containing 50 images (constant), the number of images per class varied from 10 to 50. Here, in every K_{true} case, 20% of the classes included 10 images per class; 20% of the classes had 20 images per class; 20% of the classes had 30 images per class; 20% of classes had 40 images per class; and the remaining 20% of classes had 50 images per class. Exponent p retains the value of 0.3 in all results presented in Figure 6.

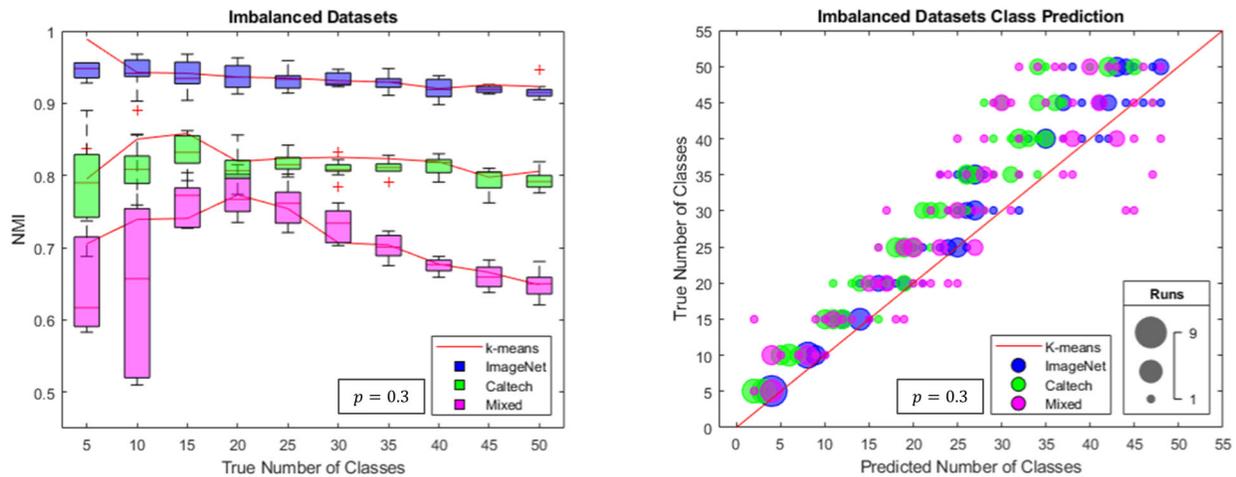


Figure 6. Box Plots and Bubble Plots for Highly Imbalanced Datasets

As observed in Figure 6, when faced with an imbalanced dataset, the E-VRC method produces results that are similar to those presented in Figures 2 and 3. Although a very small drop in the NMI index can be detected, this behavior is expected as the algorithm is presented with a more challenging clustering problem. However, a similar drop in the

NMI index is observed when the supervised (i.e. known number of true clusters) K-means clustering method is used (red lines in Figure 6).

CONCLUSIONS

This paper presents a new method for unsupervised clustering. An algorithm that can efficiently cluster groups of images without any prior knowledge about them is useful in many applications such as autonomous vehicle navigation or the creation of deep learning image datasets for training CNNs. Further, it can be used for sorting raw data collected in the field so that they can be easily used when creating synthetic environments for simulations. The results obtained by the new method exhibit better performance when compared to three other unsupervised clustering methods. The robustness of the E-VRC metric in assessing the number of clusters in a dataset is demonstrated. The E-VRC performs consistently well when processing different datasets that contain both a balanced and an imbalanced number of images per class. In military ground vehicle applications new unsupervised clustering capability allows for the cross-correlation of the sorted images with other relevant data to identify significant events and plan for the appropriate action through control algorithms embedded in the vehicle. Other than its direct use in clustering image datasets, extensions of this work could lead to new, more efficient and capable unsupervised networks for image feature extraction. Therefore, this work makes important contributions in unsupervised clustering.

ACKNOWLEDGEMENTS

We acknowledge the technical and financial support of the Automotive Research Center (ARC) in the University of Michigan in accordance with the Cooperative Agreement W56HZV-19-2-0001 of the U.S. Army DEVCOM Ground Vehicle Systems Center (GVSC) in Warren, MI, USA. We also thank Dr. Sinaga and Dr. Yang [16] for providing us with the U-k-means code, which we use to compare the E-VRC results.

REFERENCES

1. Linda G Shapiro, George C Stockman, et al. Computer vision. Vol. 3. Prentice Hall New Jersey, 2001.
2. Pelleg, Dan; Moore, Andrew, "Accelerating exact k -means algorithms with geometric reasoning," Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '99, San Diego, California, United States: ACM Press: 277–281, 1999.
3. Robert E Kass and Adrian E Raftery. "Bayes factors". In: Journal of the american statistical association 90.430 (1995), pp. 773–795.
4. Hamparsum Bozdogan. "Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions". In: Psychometrika 52.3 (1987), pp. 345–370.
5. Joseph C Dunn. "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters". In:(1973).
6. David L Davies and Donald W Bouldin. "A cluster separation measure". In: IEEE transactions on pattern analysis and machine intelligence 2 (1979), pp. 224–227.
7. Peter J Rousseeuw. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis". In: Journal of computational and applied mathematics 20 (1987), pp. 53–65.
8. Robert Tibshirani, Guenther Walther, and Trevor Hastie. "Estimating the number of clusters in a data set via the gap statistic". In: Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63.2 (2001), pp. 411–423.
9. Jing Yao et al. "Robust subspace clustering via penalized mixture of Gaussians". In: Neurocomputing 278 (2018), pp. 4–11.
10. Xinyu Geng et al. "An improved K-Means algorithm based on fuzzy metrics". In: IEEE Access 8 (2020), pp. 217416–217424.
11. Fouad H Awad and Murtadha M Hamad. "Improved k-means clustering algorithm for big data based on distributed smartphoneneural engine processor". In: Electronics 11.6 (2022), p. 883.
12. Jianjun Lei et al. "Reducing Background Induced Domain Shift for Adaptive Person Re-Identification". In: IEEE Transactions on Industrial Informatics (2022).

13. Guoting Lin et al. “Multi-Modal 3D Shape Clustering with Dual Contrastive Learning”. In: Applied Sciences 12.15 (2022), p. 7384.
14. Tadeusz Calinski and Jerzy Harabasz. “A dendrite method for cluster analysis”. In: Communications in Statistics-theory and Methods 3.1 (1974), pp. 1–27.
15. Dan Pelleg, Andrew W Moore, et al. “X-means: Extending k-means with efficient estimation of the number of clusters.” In: Icml. Vol. 1. 2000, pp. 727–734.
16. Kristina P Sinaga and Miin-Shen Yang. “Unsupervised K-means clustering algorithm”. In: IEEE access 8 (2020), pp. 80716–80727.
17. Thomas Kaster, Volker Wendt, and Gerhard Sagerer. “Comparing clustering methods for database categorization in image retrieval”. In: Joint Pattern Recognition Symposium. Springer. 2003, pp. 228–235.
18. Andrew McCallum, Kamal Nigam, and Lyle H Ungar. “Efficient clustering of high-dimensional data sets with application to reference matching”. In: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. 2000, pp. 169–178.
19. KV Mardia. “Multi-dimensional multivariate Gaussian Markov random fields with application to image processing”. In: Journal of Multivariate Analysis 24.2 (1988), pp. 265–284.
20. Rene Vidal, Yi Ma, and S Shankar Sastry. “Principal component analysis”. In: Generalized principal component analysis. Springer, 2016, pp. 25–62.
21. Michael W Berry and Murray Browne. Understanding search engines: mathematical modeling and text retrieval. SIAM, 2005.
22. Michael W Berry, Susan T Dumais, and Gavin W O’Brien. “Using linear algebra for intelligent information retrieval”. In: SIAM review 37.4 (1995), pp. 573–595.
23. Rakesh Agrawal, Christos Faloutsos, and Arun Swami. “Efficient similarity search in sequence databases”. In: International conference on foundations of data organization and algorithms. Springer. 1993, pp. 69–84.
24. Eamonn Keogh et al. “Locally adaptive dimensionality reduction for indexing large time series databases”. In: Proceedings of the 2001 ACM SIGMOD international conference on Management of data. 2001, pp. 151–162.
25. Jia Deng et al. “Imagenet: A large-scale hierarchical image database”. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE. 2009, pp. 248–255.
26. Gregory Griffin, Alex Holub, and Pietro Perona. “Caltech-256 object category dataset”. In: (2007).
27. Zhou B, Lapedriza A, Khosla A, et al. Places: a 10 million-image database for scene recognition. IEEE Trans Pattern Anal Mach Intell 2017; 40(6): 1452–1464.
28. <https://pytorch.org/vision/stable/models.html>
29. Tarald O Kvalseth. “On normalized mutual information: measure derivations and properties”. In: Entropy 19.11 (2017), p. 631.
30. Pablo A Estevez et al. “Normalized mutual information feature selection”. In: IEEE Transactions on neural networks 20.2 (2009), pp. 189–201.
31. Aaron F McDaid, Derek Greene, and Neil Hurley. “Normalized mutual information to evaluate overlapping community finding algorithms”. In: arXiv preprint arXiv:1110.2515 (2011).