IMAGE SEGMENTATION FOR PEOPLE IDENTIFICATION: AN EVALUATION OF UNSUPERVISED TECHNIQUES

Lucas Lisboa dos Santos[b], Tiago Pagano[b], Juliano Vacaro[e], Rafael Loureiro[c], Neilton Junior[c], Guilherme da Cunha[d], Erick Giovani Sperandio Nascimento[a], Ingrid Winkler[a]

[a] , Dpt. de Modelagem Computacional, Centro Universitário SENAI CIMATEC, Brasil
[b] PPG em Modelagem Computacional e Tecnologia Industrial, Centro Universitário SENAI CIMATEC, Brasil
[c] Grad. de Engenharia da Computação, Centro Universitário SENAI CIMATEC, Brasil
[d] Téc. de Desenvolvimento de Sistemas, Centro Universitário SENAI CIMATEC, Brasil
[e] HP Inc., Transformation Organization, R&D Brazil

Abstract: The evaluation of segmentation techniques is a complex activity since it depends on the target purpose. Our research is a technical evaluation of segmentation, specifically, it aims to evaluate the techniques Ant Colony Fuzzy C-means Hybrid Algorithm (AFHA), Region Splitting and Merging Fuzzy C-means Hybrid Algorithm (RFHA) with the distance between points and Kanezaki, to identify people in images from the perspective of Jaccard Index and F Measure metrics (J&F). The method was divided into four stages: the selection of the image sample, evaluation process, experiment execution, and results composed by segmented image, group, and J&F metrics. The results indicate Kanezaki has surpassed the other techniques. It is recommended future research to identify whether a correlation between quantitative and qualitative analysis exists.
Keywords: Image Segmentation; Machine Learning; Deep Learning; Segmentation Evaluation; Clustering

SEGMENTAÇÃO DE IMAGENS PARA IDENTIFICAÇÃO DE PESSOAS: UMA AVALIAÇÃO DE TÉCNICAS NÃO SUPERVISIONADAS

Resumo: A avaliação das técnicas de segmentação é uma atividade complexa, pois depende do objetivo da segmentação. Nossa pesquisa é uma avaliação de técnicas de segmentação, mais especificamente ela tem como objetivo avaliar as técnicas *Ant Colony Fuzzy C-means Hybrid Algorithm* (AFHA), *Region Splitting and Merging Fuzzy C-means Hybrid Algorithm* (RFHA) com variações na distância entre pontos e Kanezaki, para identificar pessoas em imagens sob perspectiva das métrica métricas *Jaccard Index* e *F Measure* (J&F). O método foi dividido em quatro etapas: seleção da amostra de imagens, processo de avaliação, execução do experimento e a obtenção dos resultados compostos por imagem segmentada, grupo e a métrica *J&F*. Os resultados indicam que a técnica Kanezaki superou as demais. Pesquisas futuras são recomendadas para identificar se existe correlação entre as análises quantitativa e qualitativa.
Palavras-chave: Segmentação de Imagens; Aprendizado de Máquinas; Aprendizado Profundo; Avaliação da segmentação; Agrupamento

1. INTRODUCTION

Since several applications need to recognize objects in order to support the man in daily activities, many studies in the field of computer vision have been currently investigating the recognition of objects in images. In medical research, for example, image segmentation is applied for intraoperative tracking, image alignment, and movement analysis [1]. The development of a neural system[1], called NeurReg, which verifies the movement and similarity of data, showing a specific application of image segmentation.

Segmentation techniques can be used to identify foreground objects in the image, solving common difficulties for identifying people, tracking human activities, analyzing car traffic, etc. The research of segmentation [2] evaluated the performance of five segmentation methods, adopting various metrics to identify the most accurate and efficient method for detecting moving objects (such as cars on-road) with a low computational cost. These studies depend on image segmentation techniques [3] which group pixels to simplify, extract features, and change the image, facilitating its analysis, enabling to represent objects.

The data can be classified into two groups: labeled and unlabeled. The unlabeled data consists of samples and image characteristics without any information on the natural groupings of the data. On the other hand, the labeled data uses a set marked with some meaningful labels or classes which is somehow informative [4]. To build annotated data it is necessary to manually map the image characteristics, making this a time-consuming process [5].

Thus, machine learning techniques can be classified into two approaches: supervised and unsupervised. The supervised approach is the process of selecting a subset of features based on some criteria to measure its importance and relevance, training a supervised model with the annotated data. On the other hand, the unsupervised approach evaluates the relevance of some characteristics, exploring the innate structures of the unlabeled data, such as their distribution, separability, and variance [4].

Image segmentation techniques can adopt both approaches, but the images may not be known by the technique and the quantity may be too large for manual annotation, making the unsupervised approach more convenient for some problems. A research gap is the usage of different metrics to evaluate the segmentation techniques, making it difficult to compare the results of these assessments. Therefore, a good image evaluation process must establish a single evaluation metric, as well as the study [6] adopts only the Jaccard Index (J) and F Measure (F), thus obtaining its average called J&F.

Advances in image segmentation [7] point out several methods, as well as ways to evaluate its results. The analysis of two segmentation techniques [7]: the Compression-based Texture Merging (CTM) and the Global and Local Saliency Analysis model (GLSA), supervised and unsupervised segmentation, respectively, under sixteen segmentation assessment metrics.

However, most research on unsupervised segmentation techniques have not analyzed the results under the J&F metric, and do not include machine learning techniques such as AFHA, RFHA, and Kanezaki. Therefore, evaluating segmentation techniques is relevant, given that a wide variety of different techniques and metrics are used, with its peculiarities, which can offer different results, under different evaluation conditions. So this study aims to establish research using only the J&F metric.

In this context, the objective of this research is to evaluate image segmentation techniques from the perspective of the Jaccard Index and F Measure (J&F) metrics to try to find groups that represent people in the images. The techniques evaluated were the Ant Colony Fuzzy C-means Hybrid Algorithm (AFHA), Region Splitting and Merging Fuzzy C-means Hybrid Algorithm (RFHA) with variations in the distance between points, and Kanezaki.

This paper is organized as follows: Section II describes the methodology, Section III presents some relevant concepts, Section IV describes the results observed and, finally, Section V presents our conclusions and further research needed.

2. METHODOLOGY

The method was divided into four steps: the selection of an image sample, development of evaluation method, development of the experiment, and the results composed by segmented image, group, and J&F metrics.

First step: it was carried out a selection of the image samples, which consisted of ten images extracted from the COCODataset© [8], with their respective segmentation masks available in the same dataset and that will be considered as the ground truth for comparison and evaluation purposes. COCODataset© was chosen because it is public and has a large number of images. It is important to note the elements to be segmented are unknown to the techniques, as they have an unsupervised approach for processing the initial dataset.

Next, we developed the evaluation method with the task to identify people in images, which consists of obtaining the most meaningful group that possibly represents people in the image generated by the segmentation techniques and comparing it with the ground truth, which is the mask of the same object or the truthful set of pixels for the object. We used the metrics Jaccard Index (J) for the region and the F Measure (F) for the limits, thus obtaining its average called J&F.

In the third stage, the experiment was carried out by segmenting ten sample images with the techniques AFHA, RFHA (Manhattan), RFHA (Mahalanobis), RFHA (Euclidean), and Kanezaki. For each segmentation technique, there were about ten executions with different hyperparameters to guarantee the best J&F for each technique without biasing the results.

One hyperparameter of the experiment which applies to all techniques is the number of groups named nClusters, which means the number of clusters with more pixels in the ground truth area, performing the joining (nClusters varied from one to five). Other customizable parameters were: dc (distance threshold) used in the AFHA, RFHA techniques and their distance variations, superPixels (amount of superpixel) and minLabels (minimum number of classes), the latter two used only in the Kanezaki technique.

Finally, in the fourth step, we calculate the overall J&F of the sample set, enabling each segmentation technique to obtain the best J&F and its hyperparameters, the segmented image, and the segmented group representing the person.

3. UNSUPERVISED IMAGE SEGMENTATION TECHNIQUES AND EVALUATION METRICS

This section presents some concepts relevant to this study. These are image segmentation techniques, distances, and quantitative metrics.

3.1. Image segmentation techniques

AFHA [9] and RFHA [10] are techniques based on optimization for the Fuzzy C-means (FCM). The third technique, Kanezaki [11], is based on convolutional neural networks (CNN) and Backpropagation. Cluster-based techniques are used to obtain the characteristics of the data structure. These have the function of identifying subgroups in data using their congruences, organizing similar data in the same group [12]. The similarity can be done according to a measure of resemblances, such as distance based on Euclidean's method or distance based on correlation, like Mahalanobis.

FCM is a data grouping algorithm that stands out when used in the image segmentation area. However, FCM is very sensitive to the number of clusters and their positions from the centers of the initial groups. Thus, a good result of the FCM image segmentation depends on the number of groups that you want to find and their position. Generally, these parameters can only be obtained by carrying out previous experiments, trying various combinations, or by developing optimization techniques to find these optimal numbers for their execution. In the following subsection, we will present two techniques defined as FCM optimizations.

3.1.1. FCM optimizing techniques

AFHA technique [9] has two modules: an optimizer called Ant System (AS), responsible for finding the optimal cluster number and its starting points; and Module FCM, that receives these parameters at startup, and segments the final image. The technique loads the image with an iterative optimization of FCM, where the number of centers is obtained at each algorithm interaction. The research [9] found that AFHA makes a better pre-segmentation scheme over the X-means algorithm.

RFHA [10] is a technique that combines FCM and Region Splitting Merging (RSM), following an adaptive unsupervised grouping approach for color image segmentation. Two techniques are applied to the algorithm: the histogram threshold and merging. The first is used in the formation of all possible cells to divide the image into several homogeneous regions, while the second is applied to merge nearby homogeneous regions and obtain a better startup for FCM.

A color image with representation in color channels (red, green, and blue - RGB) is composed of several homogeneous regions with different intensity ranges for each color channel, the pixels are created by those regions, according to the intensity range [10]. In the RSM module, the histogram threshold technique can successfully detect the valleys in the histogram of each color channel and be applied to the formation of all possible cells, dividing the image into homogeneous regions. The endpoints of these cells can be created with the adjacent valleys in the histogram of each color channel obtaining intensity ranges. Then the merging technique is applied, expanding the homogeneous regions obtained at the histogram threshold, thus optimizing the centers of the clusters, being used as initialization of the FCM. In research results [10], the RFHA achieved an average improvement of 12% in the cluster quality and 63% classification error reduction compared to other existing segmentation approaches.

Both AFHA and RFHA have the distance threshold as the main hyperparameter, originally measured in Euclidean distance for AFHA and Manhattan distance for RFHA.

3.2. Distances

For the segmentation of color images, the RGB color space is a commonly used approach, in which each color is represented by a triplet of red, green, and blue intensities. The color distance is used as a measure of similarity that makes pixels by region which satisfy a certain degree of color homogeneity be grouped to form a cluster [13].

These groupings are performed based on a similarity limit value. This is called threshold value or dc, which has different valid ranges for each distance metric. On the other hand, the color distance is used in the techniques RFHA and AFHA. Several metrics can be adopted to calculate these distances, so in this study, we explored the Euclidean (EUD), Manhattan (MN), and Mahalanobis (MD) distances. Under these circumstances, the functions are defined using vectors p and q, as a three-dimensional RGB data point.

The Euclidean distance metric is commonly used to compute distance in N-dimensional space vectors. For a color space with three dimensions, it is calculated according to formula 1 [14].

$$EUD(p,q) = \sqrt{\sum_{i=0}^{n} (p_i - q_i)^2} \quad (1)$$

EUD is sensitive to variations in intensity, but not very sensitive to changes in hue and saturation [15]. The valid limits for this distance in three dimensions vary from 10 to 190 to represent the homogeneity between the colors. When the dc value is greater than 190, it results in groupings with random colors [13].

The Manhattan distance is known to be of absolute value. It calculates the distance from one point to another along a path, in other words, the sum of the differences between its components. For a 3-dimensional color space, it is calculated according to the formula 2 [14].

$$MN(p,q) = \sum_{i=0}^{n} |p_i - q_i| \quad (2)$$

The valid limits of Manhattan for distances in three dimensions, are equal to the Euclidean, varying from 10 to 190 [13].

The Mahalanobis distance is based on the correlation between the variables. For a three-dimensional color space, it is defined as for formula 3 [14].

$$MD(p,q) = \sqrt{(p_i - q_i)^T V^{-1} (p_i - q_i)} \quad (3)$$

Where V is the covariance matrix. The Mahalanobis valid limit, for distances in three dimensions, varies between zero and three. However, a more reliable representation uses values between two and three [16].

3.3. Kanezaki

The Kanezaki technique [11] proposes to use CNN for unsupervised image segmentation, highlights the existence of similar techniques, but stands out by being innovative in unsupervised use with deep learning techniques. The technique is based on three iterative criteria never completely satisfied: pixels of a similar category must receive the same label, it is desired that pixels in the same continuous space receive the same label, and the number of unique labels must be large.

The technique uses backpropagation of the loss softmax for normalized responses of the convolutional layers. The proposed CNN assigns cluster labels to

pixels of the image and updates the convolutional filters to obtain better cluster separation. A superpixel refinement process is also introduced to achieve the spatial continuity constraint for the estimated segments. Experimental results [11] on the dataset BSDS50 demonstrated the effectiveness of the method. The hyperparameters of this technique are maxIter (Maximum number of iterations), minLabels (Minimum number of labels), superPixels (Amount of SuperPixels applying the refinement).

In short, the Kanezaki technique presents a new CNN architecture and its self-training process allows the segmentation of images in an unsupervised environment.

3.4. Evaluation metrics

The metric adopted in this investigation, J&F, was used in the Davis Challenge, which is a competition of video object segmentation, under some dataset. One of these challenge categories is the Unsupervised challenge, which evaluates algorithms that require no human interference. In a supervised evaluation algorithm, given a ground-truth mask G and the cluster segmented M, the evaluation methods must return how well M fits in G [17].

The J&F uses two complementary points of view: one based on the similarity of the region called Jaccard Index(J) and the other focused on the contour precision named F Measure(F).

Region Similarity: This is defined by the number of pixels incorrectly annotated. J is the intersection of the estimated segmentation union and the ground truth mask which was used. The J is widely used since it provides intuitive and invariable information to scale in the number of wrongly annotated pixels, defined as shown in formula 4 [17].

$$J = \left| \frac{M \cap G}{M \cup G} \right| \quad (4)$$

Contour Accuracy: In a contour-based perspective, M can be interpreted as a set of closed contours c(M), delimiting the spatial extent of the cluster segmented mask, and G can be interpreted as ground truth mask c(G). Thus, it is possible to calculate the precision and the recall (Pc and Rc), based on the contour between the contour points c(M) and c(G). F is used for making the best synthesis of the two, which can be defined as shown in formula 5 [17].

$$F = \frac{2P_c R_c}{P_c + R_c} \quad (5)$$

With J and F, is possible to calculate their average to obtain the final result called J&F. Although J&F is applied to video segmentation in the Davis Challenge, it can be used for image segmentation, since a video is a sequence of frames.

4. RESULTS AND DISCUSSION

The results in Table 1. show the best hyperparameters we applied for all techniques. The *nCluster = 1* was the best number of segments to be considered for the junction. This occurs because the metric J&F penalizes the group containing pixels outside the ground truth mask.

Table 1. Best possible configuration of Hyperparameters under an analysis of the J&F metric.

Technique	Best Hyperparameters	Best J&F	Execution Time (hours)
AFHA	dc = 10 nClusters = 1	0.334	8.941

RFHA Manhattan	dc = 130 nClusters = 1	0.409	15.306
RFHA Mahalanobis	dc = 3 nClusters = 1	0.4389	13.386
RFHA Euclidean	dc = 110 nClusters = 1	0.406	9.413
Kanezaki	maxIter = 500 minLabels = 3 superPixel = 10000 nClusters = 1	**0.560**	**0.535**

Source: Own author

In a quantitative analysis, results in table 1 indicate the Kanezaki technique surpassed the others, demonstrating the potential of the neural networks in the aspect of unsupervised segmentation for people recognition in images and surpasses traditional grouping techniques with faster execution time. When compared with RFHA Manhattan it is up to 29 times faster. We can see a result example of the experiments in figure 1, from left to right AFHA, RFHA Manhattan, RFHA Mahalanobis, RFHA Euclidean, and Kanezaki.

Figure 1. A visual example of segmentation Experiment

Source: Own author

5. CONCLUSION

This study aimed to evaluate the techniques Ant Colony Fuzzy C-means Hybrid Algorithm (AFHA), Region Splitting and Merging Fuzzy C-means Hybrid Algorithm (RFHA) with variations in their distance between points and Kanezaki for identifying people in images from the perspective of the J&F metric.

Using the J&F metric, Kanezaki technique considerably surpassed the others and, among the techniques based on Fuzzy C-means, the RFHA technique with the Mahalanobis distance was slightly better. This paper indicated the Kanezaki technique outperformed the other techniques for the task of identifying people using the J&F metric. Further research is needed under a qualitative approach to identify if a correlation between quantitative and qualitative analysis exists based on J&F.

Acknowledgments

I am grateful to CIMATEC, HP and the brasilian Informatics Law that made possible the existence of the HP Vialab Research Group, which I am part of.

6. REFERENCES

[1] ZHU, W. et al. Neurreg: Neural registration and its application to image segmentation. In:The **IEEE Winter Conference on Applications of Computer Vision**. [S.l.: s.n.], 2020. p. 3617–3626

[2] AGRAWAL, S.; NATU, P. Segmentation of moving objects using numerous background subtraction methods for surveillance applications. **International Journal of Innovative Technology and Exploring Engineering(IJITEE)**, v. 9, n. 3, p. 2553–2563, 2020.

[3] GONZALES, R. C.; WOODS, R. E. **Digital image processing**. [S.l.]:Prentice hall New Jersey, 2002.

[4] Ang, J. C. et al. Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection. **IEEE/ACM Transactions onComputational Biology and Bioinformatics**, v. 13, n. 5, p. 971–989, 2016.

[5] JING, L.; TIAN, Y. Self-supervised visual feature learning with deep neural networks: A survey. **IEEE Transactions on Pattern Analysis andMachine Intelligence**, IEEE, 2020.

[6] KHOREVA, A.; ROHRBACH, A.; SCHIELE, B. Video object segmentation with language referring expressions. In: SPRINGER. **AsianConference on Computer Vision**. [S.l.], 2018. p. 123–141.

[7] WANG, Z.; WANG, E.; ZHU, Y. Image segmentation evaluation: a survey of methods. **Artificial Intelligence Review**, Springer, p. 1–38, 2020.

[8] LIN, T.-Y. et al. Microsoft coco: Common objects in context. In:SPRINGER. **European conference on computer vision.** [S.l.], 2014. p.740–755.

[9] YU, Z. et al. An adaptive unsupervised approach toward pixel clustering and color image segmentation. **Pattern Recognition**, Elsevier. 43, n. 5, p. 1889–1906, 2010.

[10] TAN, K. S.; ISA, N. A. M.; LIM, W. H. Color image segmentation using adaptive unsupervised clustering approach. **Applied Soft Computing**, Elsevier, v. 13, n. 4, p. 2017–2036, 2013.

[11] KANEZAKI, A. Unsupervised image segmentation by backpropagation.In: IEEE. 2018 **IEEE international conference on acoustics, speech and signal processing (ICASSP).** [S.l.], 2018. p. 1543–1547.

[12] GAN, G.; MA, C.; WU, J. **Data clustering: theory, algorithms, and applications**. [S.l.]: Siam, 2007.

[13] LOO, P. K.; TAN, C. L. Adaptive region growing color segmentation for text using irregular pyramid. In: SPRINGER. **International Workshop On Document Analysis Systems.** [S.l.], 2004. p. 264–275.

[14] WALTERS-WILLIAMS, Janett; LI, Yan. Comparative study of distance functions for nearest neighbors. In: **Advanced techniques in computing sciences and software engineering**. Springer, Dordrecht, 2010. p. 79-84.

[15] WESOLOWSKI, S.; DONY, R. D.; JERNIGAN, M. Global color image segmentation strategies: Euclidean distance vs. vector angle. In: IEEE. **Neural Networks for Signal Processing IX: Proceedings of the 1999IEEE Signal Processing Society Workshop (Cat. No. 98TH8468).** [S.l.], 1999. p. 419–428.

[16] GALLEGO, G. et al. On the mahalanobis distance classification criterion for multidimensional normal distributions. **IEEE Transactions onSignal Processing**, IEEE, v. 61, n. 17, p. 4387–4396, 2013.

[17] PERAZZI, F. et al. A benchmark dataset and evaluation methodology for video object segmentation. In**: Computer Vision and PatternRecognition.** [S.l.: s.n.], 2016