

Literature Review

Student: Daniel Hess - 21971897

Supervisors: Prof. Eun-Jung Holden, Dr Daniel Wedge

Introduction

With many methods being developed for the purposes of segmentation the field is becoming dense with options. There are many applicable domains including medical, and autonomous vehicles which rely on fast and robust computer vision technologies that perform image segmentation. In the medical domain, image segmentation allows researchers or technicians to process more data from highly specialized imagery [18], while autonomous vehicles can accurately find road boundaries and vehicles [4].

This project will focus on applying segmentation for the mine pit imagery captured from an Unmanned Aerial Vehicle (UAV), specifically to accurately delineate detrital units from bedded units. This study will also look into public datasets [5] [7] that can be used as a benchmark for these techniques.

While images aren't tactile, they can still have texture. Generally, image texture is classified by non-uniform regions with associated shape or colour features, unique to the region.

The purposes of this study are to explore unsupervised techniques of segmentation with a focus on feature extraction and clustering algorithms where as many techniques involve the classification of texture. In addition to this, convolutional neural networks (CNNs) can be used as a tool to perform feature extraction, both supervised and unsupervised.

Finally, we discuss metrics which will be used to compare performance of the implemented techniques.

Feature Extraction

Gabor

Feature extraction is important for segmentation as it quantifies visual patterns in data. The idea generally being that by using certain metrics, convolutions or statistics, information can be extracted about the regions or pixels of an image, and represent this data in vector form, summarising the region. Then based on this vector space, conventional clustering techniques can be used to split the pixels into different classes or regions.

Gabor filters are used in texture analysis, which in the 2D spatial domain is comprised of a modulated Gaussian kernel. They are said to be similar to the low-level visual human responses [9] and are particularly useful for texture and edge representation [texture classification using Gabor filters].

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma y'^2}{2}\right) \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right)$$
$$x' = x \cos \theta + y \sin \theta$$
$$y' = -x \sin \theta + y \cos \theta \tag{1}$$

In equation 1 we can see that the filter has a Gaussian component coupled with a sinusoidal component. $\lambda, \theta, \psi, \sigma, \gamma$ represent sine wavelength, filter orientation, phase offset, gaussian standard deviation and aspect ratio respectively.

Xu et al [16] used gabor filters to not only extract features from grey-level images, but also from depth images which achieved good characterization of faces for the recognition process. The system benefitted from the inclusion of both intensity and depth features, compared to both individually.

Similarly, log-Gabor filters (i.e. where the kernel is Gaussian on a logarithmic scale rather than a linear scale) have also been used for feature extraction. Nava et al [13] suggest that log- Gabor filters out perform traditional gabor filters with major differences occuring and texture borders.

Auto Encoders

Neural Networks (convolutional or otherwise) dominates computer vision research for their easy setup, virutally end-to-end design and fast usage times. In this case we are interested in the use of auto-encoders as a form of feature extraction. Deep neural networks are great at finding latent variables (characteristics inferred by observable metrics). Auto-encoders reduces data to a lower dimensional form based on convolutions, and using the features generated may be useful.

Yang et al [17] propose an auto-encoder based method for unsupervised defect detection. This method uses an auto-encoder for feature extraction which it clusters over to find surface defects. This method has the benefit that it is unsupervised however achieves great performance in finding surface defects.

Raja and rani [14] use a similar approach for tumor classification from MRI images. Using various other feature extraction methods, as well as features extracted from training an auto-encoder, their method achieved higher accuracy than other existing systems. By incorporating the system with Bayesian fuzzy clustering, they were able to identify the positive regions in the MR images.

Clustering

With plenty of features extracted from the images, we need some way to select regions based on similarity on feature properties. A common method once features are extracted is to cluster these features. Provided we can effectively extract texture features from the images, clustering should provide a good framework for differentiating regions on texture.

K-means

This is one of the most popular clustering algorithms based on simplicity and fast computation. The k-means algorithm can be summarised simply:

1. Take k points (call these the means) for each cluster.
2. Assign all observations to a cluster based on euclidean distance.
3. Calculate the mean value for each given cluster and set this as the mean for that cluster.
4. Repeat steps 2-3 until observations do not change cluster.

The method has been shown to perform well for segmentation purposes based on feature extracted images when the number of clusters is known in advance. Lin et al [11] shows that given good features, decent segmentation can be achieved.

Fuzzy c-means

Similar to kmeans is fuzzy c-means. As opposed to hard clustering, this algorithm adopts soft clustering, meaning it considers the "probability" of a particular point being within a cluster. The algorithm is as follows:

1. Choose a number of clusters
2. Assign coefficients for observations to clusters randomly.
3. Compute the centroid for each cluster (i.e. the weighted mean for each point in the cluster)
4. Compute the coefficients for each cluster and observation.
5. Repeat steps 3-4 until the change in coefficients is less than some epsilon.

The main difference to kmeans is that a specific observation isn't completely assigned to a single cluster, rather, the algorithm considers the "probability" of the point belonging to each cluster. This method allows for smoother transitions between clusters, and that centroids are weighted more heavily to points that are closer to it.

Yang et al [18] were able to segment normal and abnormal tissue in MR images by the use of fuzzy c-means clustering on intensity. Harikiri et al [6] achieved better results with the addition of other features for segmentation such as distance from center of spot region and median intensity of surround pixels.

Finally Chuang et al [1] considered a fuzzy c-means algorithm that also considers the square region around a pixel, and the clusters to which they belong. It is advantageous over normal FCM as it yields more homogenous regions and is also much less sensitive to noise.

DBSCAN

The last clustering algorithm considered here is DBSCAN. This algorithm takes a different approach to the other 2 as it considers the density of points within regions and will sometimes not assign observations to a cluster. The algorithm is as follows.

1. Find all points within epsilon size neighborhood of every observation.
2. Identify which points are core points by whether they have at least minPts neighbors.
3. Find which core points are connected to other core points and put them into their own cluster.
4. Assign each non-core point to a cluster if the cluster is within epsilon of a core point, otherwise assign it to noise.

The algorithm works well for arbitrarily shaped clusters and is robust to noise making it a well performing algorithm [10] however is very dependent on good hyperparameter selection.

The biggest drawback for DBSCAN is the ability to fine-tune the hyper-parameter epsilon. Wang et al [15] proposed a method for automatic epsilon, with selection outperforming other manually selected values.

Convolutional Neural Networks

Although mentioned earlier, this section focusses on supervised techniques used for segmentation.

Neural networks have exceptional performance in many applications. CNNs specifically have the benefit that working with images is done in the spatial domain. Deep CNNs are comprised of layers, where the first layers represent low-level features, while successive layers represent higher-level features built on the lower features.

While many CNNs have been trained to work as classifiers, work is still being done for segmentation. Long et al [12] managed to adapt architectures from previous state-of-the-art classifier systems into a segmentation system with good performance. By retaining the pretrained lower level layers and replacing the final classifier layers with a 1x1 convolution, the classifier can now predict a score for each class over the image space, up-sampling wherever necessary.

A similar approach is done by Jégou et al [8]. The architecture for this method consists of a dense down-sampling path, reducing the convolutions in dimension, a bottleneck path to learn deep features followed by an up-sampling path with the final layer being dense, and finally a 1x1 convolution soft-max layer to provide class distribution. The distinction this makes with the previous system is that it uses a deep fully connected up-sampling path before segmentation.

A different approach taken by Fujieda et al [2] was to incorporate wavelets into the CNNs. While it follows a specific architecture, the idea is that it runs concurrently wavelet transforms alongside convolutional layers, with concatenation and shortcuts incorporated.

Metrics

Different metrics exist to measure goodness-of-fit, however different metrics solve different problems. Generally, in segmentation, there are a few different metrics used to evaluate how well a method segments the data, which primarily fall into two categories: subjective and objective [3].

Objective evaluation is unsupervised, meaning it doesn't have labelled data to compare against as ground truth.

- F evaluation: Measures the average squared color error and works to regularise number of segments - has the benefit of not requiring hyper-parameters. One of the simplest objective metric however can tend to over-segment. F' is an improvement to regularisation and helps to reduce over-segmentation.
- Q evaluation: An improvement over F' , however improves bias towards under segmentation. Preference over F' is made on a data specific basis.
- E evaluation: Compares the expected entropy of a segmented region to the actual measured entropy.
- Zeboudjs contrast: Based on the internal and external contrast for the regions around a pixel.

Subjective evaluation considers a labelled ground-truth. Because we can consider segmented regions as classes, classification metrics work for evaluation on a pixel by pixel scale, however there are a few specific to segmentation.

- Probabilistic random index: Essentially a cluster-invariant evaluation of accuracy. Does not assign a specific class to a cluster and just measures whether classes appear together within a cluster.
- Variation of Information: A combination of the entropy for both a cluster and the ground truth, as well as considering the mutual information between clusters. VOI measures the loss of information between clusters.
- Global Consistency Error: Consider a metric to evaluate the per pixel error of a segmentation. The GCE of a partition, is the scaled sum of error of all pixels in a region.
- Boundary Displacement Error: Consider the error on a boundary pixel, the error is defined as the distance between that boundary pixel, and the closest boundary pixel in the other ground-truth.

References

- [1] CHUANG, K.-S., TZENG, H.-L., CHEN, S., WU, J., AND CHEN, T.-J. Fuzzy c-means clustering with spatial information for image segmentation. *computerized medical imaging and graphics* 30, 1 (2006), 9–15.
- [2] FUJIEDA, S., TAKAYAMA, K., AND HACHISUKA, T. Wavelet convolutional neural networks. *arXiv preprint arXiv:1805.08620* (2018).
- [3] GARCIA-LAMONT, F., CERVANTES, J., LÓPEZ, A., AND RODRIGUEZ, L. Segmentation of images by color features: A survey. *Neurocomputing* 292 (2018), 1–27.

- [4] HA, Q., WATANABE, K., KARASAWA, T., USHIKU, Y., AND HARADA, T. Mfnet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2017), IEEE, pp. 5108–5115.
- [5] HALLEY, F., ET AL. *Perceptually relevant browsing environments for large texture databases*. PhD thesis, Citeseer, 2012.
- [6] HARIKIRAN, J., LAKSHMI, P., AND KUMAR, R. K. Multiple feature fuzzy c-means clustering algorithm for segmentation of microarray images. *International Journal of Electrical and Computer Engineering* 5, 5 (2015).
- [7] HOSSAIN, S., AND SERIKAWA, S. Texture databases—a comprehensive survey. *Pattern Recognition Letters* 34, 15 (2013), 2007–2022.
- [8] JÉGOU, S., DROZDZAL, M., VAZQUEZ, D., ROMERO, A., AND BENGIO, Y. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (2017), pp. 11–19.
- [9] JONES, J. P., AND PALMER, L. A. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of neurophysiology* 58, 6 (1987), 1233–1258.
- [10] KURUMALLA, S., AND RAO, P. S. K-nearest neighbor based dbscan clustering algorithm for image segmentation. *Journal of Theoretical and Applied Information Technology* 92, 2 (2016), 395.
- [11] LIN, W.-T., LIN, C.-H., WU, T.-H., AND CHAN, Y.-K. Image segmentation using the k-means algorithm for texture features. *World Academy of Science, Engineering and Technology* 65 (2010), 612–615.
- [12] LONG, J., SHELHAMER, E., AND DARRELL, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 3431–3440.
- [13] NAVA, R., ESCALANTE-RAMÍREZ, B., AND CRISTÓBAL, G. A comparison study of gabor and log-gabor wavelets for texture segmentation. In *2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA)* (2011), IEEE, pp. 189–194.
- [14] SIVA RAJA, P., AND RANI, A. V. Brain tumor classification using a hybrid deep autoencoder with bayesian fuzzy clustering-based segmentation approach. *Biocybernetics and Biomedical Engineering* 40, 1 (2020).
- [15] WANG, C., JI, M., WANG, J., WEN, W., LI, T., AND SUN, Y. An improved dbscan method for lidar data segmentation with automatic eps estimation. *Sensors* 19, 1 (2019), 172.
- [16] XU, C., LI, S., TAN, T., AND QUAN, L. Automatic 3d face recognition from depth and intensity gabor features. *Pattern recognition* 42, 9 (2009), 1895–1905.
- [17] YANG, H., CHEN, Y., SONG, K., AND YIN, Z. Multiscale feature-clustering-based fully convolutional autoencoder for fast accurate visual inspection of texture surface defects. *IEEE Transactions on Automation Science and Engineering* 16, 3 (2019), 1450–1467.
- [18] YANG, M.-S., HU, Y.-J., LIN, K. C.-R., AND LIN, C. C.-L. Segmentation techniques for tissue differentiation in mri of ophthalmology using fuzzy clustering algorithms. *Magnetic Resonance Imaging* 20, 2 (2002), 173–179.