Report

Vanshita Bansal

08-06-2020

Contents

1 Images

Images are basically information stored in the form of pixels. An image is a 2D array that stores these pixels. The images are considered to be discrete in nature. An image can give a lot of information but may require some initial processing to extract the required information as they might include noise.

1.1 Types

Depending upon the requirements there are some types of digital images:

1.1.1 Binary Images

These are simplest kind of images. The pixels attains only two values either 1 or 0 therefore need only one bit. These maybe used where the general shape or outline is required.

1.1.2 Gray Scale Images

These are monochromatic i.e. single color image. They contain gray-level information. Typically, contains 8 bits/pixel data that allows 0-255 gray level.

1.1.3 Color Images

It uses three basic colors(Red, Green, Blue) and are used to specify different colors using different composition.Usually, 24 bits/pixel image are popular.

1.1.4 Multispectral Images

These are not directly visible to human system. This may include X-ray, infrared, etc. However, the information can be made available to human using mapping to RGB components.

A continuous image can be converted to digital image using :

1)Sampling

2)Quantization

The quality of an image is basically the resolution. There are two types of resolution :

1) Spatial resolution: It is the smallest visible detail of an image.

2) Gray-level resolution: It refers to smallest visible change in gray level.

The images obtained are not used directly, some of the image enhancement techniques are applied to get a more suitable image. The two methods for image enhancement are:

1) Spatial domain methods

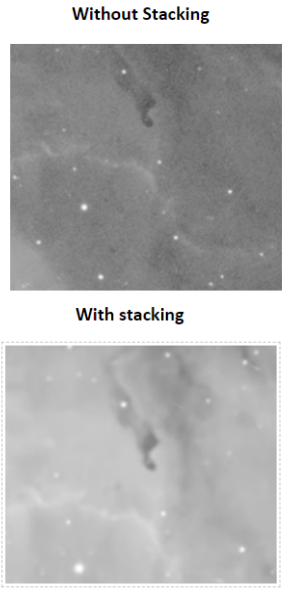
2) Frequency domain methods

1.2 Stacked Images

Generally, the images captured by the digital cameras have noise. The noise can be heat noise, read noise etc. We apply stacking to remove these noise. Stacking should not be mis-understood with improving images. Stacking uses averaging (commonly) on each of the pixels across a series of image to get the actual value, eventually removing the noise.

However, stacking more frames does not give us signal that is not present there. A point comes, when even on increasing the number of stacked images the improvement is negligible. Why? It is because most of the noise is removed. A good stack reduces so much of noise that the image signal appear stronger. Also, having more frames allows us to throw the images that are not of good quality (may be due to flash, dim light, fluctuation, blurness).

By removing the noise we mean improving the signal-to-noise ratio of the combined image. With study, we can say that signal-to-noise ratio increases by the square root of the number of the images in the stack if there are no side effects. But in practice, there are side effects. Example:



There are two methods used for stacking images:

1) Add/average that includes sigma clipped average, min/max excluded average, etc.

2) Median including clipped median, min/max exclusion, etc.

Below is the analysis of dataset on which median and sigma clipped average is applied:

Source:https://clarkvision.com/articles/image-stacking-methods/

1.2.1 Test Data

Image with embedded numbers was constructed and different images with random noise was obtained. Stacking was performed in ImagePlus selecting two methods: median and sigma clipped average.

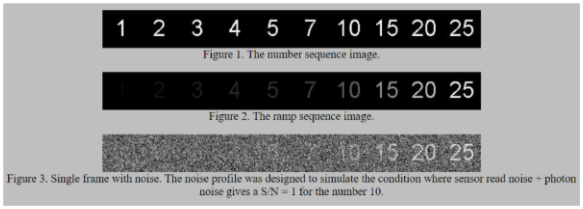


Figure 1. The number sequence image.

Figure 2. The ramp sequence image.

Figure 3. Single frame with noise. The noise profile was designed to simulate the condition where sensor read noise + photon noise gives a S/N = 1 for the number 10.

1.2.2 Results

The results of tone curve gave the best results in all the three cases. Sigma clipped average produced better results with lower noise level than median average.

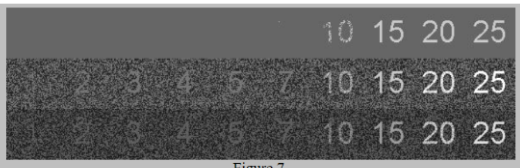


Figure 7.
Top: 100 image median combine on 14-bit data.
Middle: 100 image median combine on 16-bit data.
Bottom: 100 image median combine on tone-mapped data.

1.2.3 Conclusion

1)The averaging methods are superior to median combine.

2)As digital camera get better, the noise in the image get lower.

3)The large number of images, increase the precision.

4)Tone mapping improves image approximately upto 40x.

But all these effects highly depend upon the noise from camera, airglow and light pollution.

Figure 8.
Top: 100 image Sigma-Clipped average combine on 14-bit linear (input and output) data.
Upper Middle: 100 image Sigma-Clipped average combine on 16-bit linear data.
Lower Middle: 100 image Sigma-Clipped average combine on tone-mapped data.
Bottom: 100 image Sigma-Clipped average combine on 32-bit floating point output data (14-bit linear input).

2 Image Segmentation

It is the process of partitioning images into multiple segments based on common characteristics. The goal is to get more meaningful information from the data. The pixels can be grouped based on similar features such as color, shape etc. This has various application and the major one is in the field of Medical where shape detection could be helpful in detecting demantia or any other illness.

The basic type of segmentation is done using edge detection technique to obtain all the abrupt changes in pixel values. These can be used to detect objects and also shape of objects. The pixels are clustered so that these can be assigned to a group. The clustering can be done using the simple K-Means Clustering technique or any other accordingly.

Semantic Segmentation

Instance Segmentation

2.1 Semantic Segmentation

It is basically partitioning an image into parts or in simple words assigning labels to each pixels in the image. Since there are millions of pixels we cannot have label for each and that also not mean anything. Therefore, we have finite number of labels or groups that becomes part of the feature map in later stage. For each pixel, depending upon their value we assign them group to which they belong i.e. label them. A simple example can be classifying each pixel that belong to a building , car etc.

This labeling is done by finding the argmax of the feature score. The pixel labels for training images must be known beforehand for semantic segmentation. This becomes the part of the feature map.

Semantic segmentation is relatively easier compared to instance segmentation. In semantic segmentation we provide label to pixel as person , car etc., whereas in instance segmentation we need to classify them and label them as person1, person2 etc.

For semantic segmentation the image is divided into patches. The size of patch influences computations required i.e. larger patch leads to more computation. These patches are mapped to the group by feeding them to various Convolutional layers, Pooling layers and Fully connected layers. If we decrease the number of patches or size of the patch the chances of error and data loss increase.

2.1.1 How can we classify the patch?

Naive approach: Classify patch based on label of central pixel, then slide one pixel over.

Similar to how we use categorical values, we create target by one-hot encoding the class labels. (What is one-hot encoding? To work on the numerical data we need to convert our String type data using either Label encoding or one-hot encoding to numeric type. Since we have more values one-hot encoding can be done easily.)

A prediction can be made to form the map by taking the argmax of each depth-wise pixel vector.

Constructing a neural network by simply stacking a number of convolutional layers with same padding output a segmentation map. However, this can be computationally expensive. Therefore to produce a segmentation map downsampling is followed by upsampling also called Encoder/Decoder structure. Downsampling collects feature evidence from a larger area. Upsampling distributes the information of a segment back to the original pixel domain. (What is downsampling? Also called pooling, it is developing a lower resolution feature mapping using either average or max. What is Upsampling? Also called unpooling, it unsample the resolution by distributing a single value into a higher resolution (generally average).

2.2 Instance Segmentation

When we separate the instances of same class then it comes under instance segmentation. So if we have two or more objects of same category in the input

image, the segmentation map does not distinguish these as separate objects, but instance segmentation does.

The instance segmentation is done in two steps: 1) Object Detection 2) Semantic segmentation

It is achieved using Masked R-CNN.The mask is introduced for the groups in which segmentation is to be done. This is a binary mask. Masked R-CNN is combination of Faster R-CNN and FCN(fully convolutional network). ROI pooling if used can lead to data loss due to quantization of stride but ROI-align does not uses quantization.

Image Segmentation can be classified into two basic types: 1)Local segmentation 2)Global Segmentation and, The image segmentation approaches can be classfied into two types: 1) Discontinuity detection based approach: It implements edge detection as it detects the sudden changes in the pixel values and linked formed due to boundaries.

2) Similarity detection based approach: In this, image is segmented based on the similarity.The techniques that falls under this category are: thresholding techniques, region growing techniques and region splitting and merging.

2.3 Image Segmentation Methods

2.3.1 Thresholding Method

In this method we divide the pixel depending on the intensity level. This can be useful in images having lighter objects than background. We set one or more threshold values either manually or automatically. Pixels are now assigned groups based on threshold.There are three types:

1) Global thresholding: We take a single threshold value (say) T. On the basis of this T, a pixel is assigned either 1 or 0 value.

2) Variable thresholding: The threshold T can vary over the image. It may depend upon the neighbouring pixels or a function.

3) Multiple thresholding: There are more than one threshold values like T0,T1. More than one threshold values are required when we can not group or classify the pixels into just two groups and so we take more threshold and assign groups depending on the thresholds.

2.3.2 Edge Based Segmentation method

The method is based on the rapid change in the intensity values(discontinuity detection) and does not depend on single intensity value. The resulting image is a binary image. Hence it effectively works when we just wish to detect the shapes of object. Firstly, all the edges are detected and then these are connected to form objects. There are various techniques to perform this like sobel operator, canny operator etc.

2.3.3 Region Based Segmentation Method

This method segments the image into various regions having similar characteristics. There are rules defined for the characteristics of pixels and if pixels satisfy certain rule these will be grouped.

These are of four types: 1) Region growing method: The region growing based segmentation methods segment the image into various regions based on the growing of seeds(initial pixels).The selection of initial pixel can be done manually or automatically. The region can be allowed to grow depending on either 4-connectivity or 8-connectivity. 2) Region merging method: Here, each pixel is considered as a separate group. Now each group is compared with the neighbouring group and if these are similar we merge them. 3) Region splitting method: Firstly, the entire image is considered as a single segment and then regions are splitted into regions having similar characteristics until we get groups that cannot be futher divided. 4) Region splitting and merging method: The region splitting and merging is done iteratively.

2.3.4 Clustering Based Segmentation Method

This method uses the technique of clustering pixels having similar characteristics. The clustering method applied can be Hard clustering that uses K-means clustering generally or Soft clustering that is more practical as there is noise in the data. Soft clustering implements fuzzy c-means clustering.

2.3.5 Watershed Based Method

The method uses the concept of topological surfaces. The lower surfaces are assigned water basins and the upper surfaces become the boundary. The water basins are allowed to grow until they merge with each other. The merged region serves as the separation between basins. We can use darker pixels for lower region and lighter pixel value for boundaries or vice versa.

2.3.6 Partial differential Equation Based Segmentation

The basic idea behind this method is that the segmentation problem is formulated in terms of minimizing (or at least finding critical points of) an energy functional that takes a level set function. The level set functions evolve according to an evolution partial differential equation (PDE).The PDE is derived from the minimization of the energy functional by calculating the L 2(ordinary) gradient of energy functional and using continuous gradient descent method. The second derivative is used to obtain smoothness, this could be obtained using first derivative also but we require to move around the direction as well as the gradient so we use second derivative.

2.3.7 Artificial Neural Network Based Segmentation Method

The artificial neural network try to simulate human brain. Similar to neuron we have nodes in the network. We use ANN mostly for segmentation of medical images, for separating image from background. A neural network is made of large number of connected nodes each having weights. The method has two basic steps: Feature extraction and segmentation by neural network.

The various techniques of feature extraction are: GLCM, HOG, Tamura and many more depending on requirement.

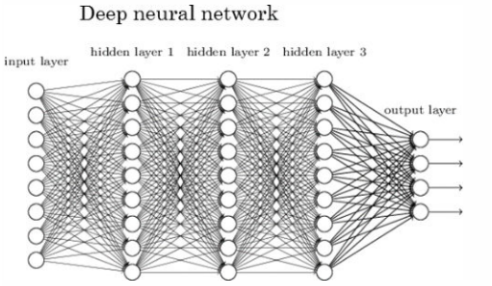
3 Comparison

COMPARISON OF VARIOUS SEGMENTATION TECHNIQUES		
Segmentation technique	Advantages	Disadvantages
Thresholding Method	no need of previous information,simplest method	highly dependent on peaks, spatial details are not considered
Edge Based Method	good for images having better contrast between objects	not suitable for wrong detected or too many edges
Region Based Method	more immune to noise, useful when it is easy to define similarity criteria	expensive method in terms of time and memory
Clustering Method	fuzzy uses partial membership therefore more useful for real problems	determining membership function is not easy
Watershed Method	results are more stable, detected boundaries are continuous	complex calculation of gradients
PDE Based Method	fastest method, best for time critical applications	more computational complexity
ANN Based Method	no need to write complex program	more wastage of time in training

4 Deep Neural Network

Neural Networks are just a simulation of human brain. We have inputs to various nodes at different stages of the network. The output at each node is calculated using different activation function. In the first level we generally use Linear function. The nodes in the network are connected with the nodes at deeper levels by assigning weights. At the hidden layers we apply either sigmoid or ReLU activation function according to the choice. The sigmoid activation function takes input and squashes the values to range -1 and 1 and other advantage is that the derivative obtained is equal to 1 minus the function,

therefore easy to compute. The rectified linear unit (ReLU) function gives 0 as the output for all the negative values and same value for positive input. The advantage of using ReLU is that it does not flatten the gradients.



The Neural network can be used for prediction. With large number of hidden layers, we can easily compute complex problems. Since we are assigning weights, how do we get to know them? These are found out using the Backpropagation method. The basic motivate is to minimize the error in prediction and therefore we calculate this error for each weight assigned and try to minimize it. For a very dense neural network this could be very expensive computationally but backpropagation does it in a very simple way. The derivatives of all the levels are calculated one by one starting from the output. Now, using chaining these are multiplied to obtain the final result which we need to minimize.

$$\frac{\partial C}{\partial x} = [\frac{\partial C}{\partial x_1}, \frac{\partial C}{\partial x_2}, \ldots, \frac{\partial C}{\partial x_m}]$$

$$\delta_j^l = \frac{\partial C}{\partial z_j^l} \qquad local\ gradient$$

There are different types of neural networks and Convolutional neural network is one of them. CNN is mostly used in image recognition as there is no need to check all the pixels applied to images one by one. CNN checks an image by blocks, starting from upper left corner and moving further. The result is passed through various convolutional layers and ultimately, we can conclude what is in the picture.