Introduction:

This project is used to develop a tool to Replacing the missing data in the selected regions from an image. This is also called as Inpainting method.The term inpainting is derived from the ancient art of restoring image by professional image restorers in museums etc. Digital Image Inpainting tries to imitate this process and perform the inpainting automatically. The algorithm automatically does this in a way that it looks “reasonable” to the human eye. Details that are hidden/occluded completely by the object to be removed cannot be recovered by any mathematical method. Therefore the objective for image inpainting is not to recover the original image, but to create some image that has a close resemblance with the original image.One use is in restoring photographs. In fact, the term inpainting has been derived from the art of restoring deteriorating photographs and paintings by professional restorers in museums etc. Ages ago, people were already preserving their visual works carefully. With age, photographs get damaged and scratched. Users can then use the software to remove the cracks from the photographs. Another use of image inpainting is in creating special effects by removing unwanted things from the image. Unwanted things may range from microphones, ropes, some unwanted person and logos, stamped dates and text etc. in the image. During the transmission of images over a network, there may be some parts of an image that are missing. These parts can then be reconstructed using image inpainting. There have been a few researches on how to use image inpainting for super-resolution and zooming of images.

AIM,SCOPE,PRESENT INVESTIGATION:

There has been a lot significant work carried out in the past in the field of inpainting. The algorithm at first sight may seem to be something similar to noise removal from images. Denoising is focused towards modifying individual pixels whereas inpainting aims at modifying larger regions from the image. Denoising also differs from inpainting in the way that in inpainting there is no information about the image in the region to be inpainted as opposed to noise removal where pixels may contain information about both the real data and noise.Thus specific methods are developed to answer this problem.

SCOPE:

The object of the project is to reconstruct the missing or damaged portions of the image, in order

to make it more legible and restore its unity. The whole scope of the problem can be stated as:

• Inpaint the regions from the image that have been marked by the user for inpainting. The user may mark more than one region (spatially disconnected) for inpainting.

• Given an image (I) and a region to be inpainted (L), inpainting would try to construct an image (I’) and remove the marked region in a visually plausible way (See Figure 2).

• This can be done by using information from surrounding areas and merge the inpainted region into the image so seamlessly that a typical viewer is not aware of it. The quality of the result will depend on what is missing. If the inpainting region is small and the surrounding area is without much texture, the result will be good. Texture is a measure of image coarseness, smoothness and regularity. Images with texture contain regions characterized more by variation in the intensity values than by one value to intensity. Large areas with lots of information lost are harder to reconstruct, because information in other parts of the image is not enough to get an impression of what is missing. If the human brain is not able to imagine what is missing, equations will not make it either. Details that are completely hidden/occluded completely by the object to be removed cannot be recovered by any mathematical method. Therefore the objective for image inpainting is not to recover the original image, but to create some image that has a close resemblance with the original image.

• It is not equivalent to noise reduction. We say that an image contains noise if the pixels in the image do not reflect the true intensity of the real scene. Though the two problems may sound strikingly similar, the approaches to solving these are different altogether. The main difference between noise reduction and image inpainting is that in noise, the region to be removed (i.e. noise) contains some information about the image whereas; there is no information about the image in the region to be inpainted. Also in noise, we do not have large regions of missing area.

Present Investigation:

Proposed System:

In this Proposed system, finding the optimal solution for all type of missing data replacement in images.

The occurancy of the missing data is calculated by the algorithms and mathematical calculations.

System Architure:

Data Mining Techniques:

Data mining is looking for hidden, valid, and potentially useful patterns in huge data sets. Data Mining is all about discovering unsuspected/ previously unknown relationships amongst the data.

Association

Association is one of the best-known data mining technique. In association, a pattern is discovered based on a relationship between items in the same transaction. That’s is the reason why association technique is also known as *relation technique*. The association technique is used in *market basket analysis* to identify a set of products that customers frequently purchase together.

Retailers are using association technique to research customer’s buying habits. Based on historical sale data, retailers might find out that customers always buy crisps when they buy beers, and, therefore, they can put beers and crisps next to each other to save time for the customer and increase sales.

Classification

Classification is a classic data mining technique based on machine learning. Basically, classification is used to classify each item in a set of data into one of a predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network, and statistics. In classification, we develop the software that can learn how to classify the data items into groups. For example, we can apply classification in the application that “given all records of employees who left the company, predict who will probably leave the company in a future period.” In this case, we divide the records of employees into two groups that named “leave” and “stay”. And then we can ask our data mining software to classify the employees into separate groups.

Clustering

Clustering is a data mining technique that makes a meaningful or useful cluster of objects which have similar characteristics using the automatic technique. The clustering technique defines the classes and puts objects in each class, while in the classification techniques, objects are assigned into predefined classes. To make the concept clearer, we can take book management in the library as an example. In a library, there is a wide range of books on various topics available. The challenge is how to keep those books in a way that readers can take several books on a particular topic without hassle. By using the clustering technique, we can keep books that have some kinds of similarities in one cluster or one shelf and label it with a meaningful name. If readers want to grab books in that topic, they would only have to go to that shelf instead of looking for the entire library.

Prediction

The prediction, as its name implied, is one of a data mining techniques that discovers the relationship between independent variables and relationship between dependent and independent variables*.* For instance*,* the prediction analysis technique can be used in the sale to predict profit for the future if we consider the sale is an independent variable, profit could be a dependent variable. Then based on the historical sale and profit data, we can draw a fitted regression curve that is used for profit prediction.

Sequential Patterns

Sequential patterns analysis is one of data mining technique that seeks to discover or identify similar patterns, regular events or trends in transaction data over a business period.

In sales, with historical transaction data, businesses can identify a set of items that customers buy together different times in a year. Then businesses can use this information to recommend customers buy it with better deals based on their purchasing frequency in the past.

Decision trees

The A decision tree is one of the most commonly used data mining techniques because its model is easy to understand for users. In decision tree technique, the root of the decision tree is a simple question or condition that has multiple answers. Each answer then leads to a set of questions or conditions that help us determine the data so that we can make the final decision based on it. For example, We use the following decision tree to determine whether or not to play tennis:

It is a multi-disciplinary skill that uses machine learning, statistics, AI and database technology.

The insights derived via Data Mining can be used for marketing, fraud detection, and scientific discovery, etc.

Mean Shift Algorithm

Meanshift is a clustering algorithm that assigns the datapoints to the clusters iteratively by shifting points towards the mode. The [mode](https://en.wikipedia.org/wiki/Mode_(statistics)) can be understood as the highest density of datapoints (in the region, in the context of the Meanshift). As such, it is also known as the mode-seeking algorithm. Meanshift algorithm has applications in the field of image processing and computer vision.

Given a set of datapoints, the algorithm iteratively assign each datapoint towards the closest cluster centroid. The direction to the closest cluster centroid is determined by where most of the points nearby are at. So each iteration each data point will move closer to where the most points are at, which is or will lead to the cluster center. When the algorithm stops, each point is assigned to a cluster.

Iteration:

1. Initial state. The red and blue datapoints overlap completely in the first iteration before the Meanshift algorithm starts.

2. End of iteration 1. All the red datapoints move closer to clusters. Looks like there will be 4 clusters.

3. End of iteration 2. The clusters of upper right and lower left seems to have reached convergence just using two iterations. The center and lower right clusters looks like they are merging, since the two centroids are very close.

4. End of iteration 3. No change in the upper right and lower left centroids. The other two centroids’ have pulled each other together as the datapoints affect each clusters. **This is a signature of Meanshift, the number of clusters are not pre-determined**.

5. End of iteration 4. All the clusters should have converged.

6. End of iteration 5. All the clusters indeed have no movement. The algorithm stops here since no change is detected for all red datapoints.

**The Meanshift Algorithm**

I will have to touch lightly on the mathematics for this part. I hate math so I will try my best to explain in terms for non-math people.

You will need a few things before you start to run Meanshift on a set of datapoints *X*:

1. A function *N(x)* to determine what are the neighbours of a point *x* ∈ *X*. The neighbouring points are the points within a certain distance. The distance metric is usually [Euclidean Distance](https://en.wikipedia.org/wiki/Euclidean_distance).

2. A kernel *K(d)* to use in Meanshift. K is usually a [Gaussian Kernel](https://en.wikipedia.org/wiki/Radial_basis_function_kernel), and d is the distance between two datapoints.

Now, with the above, this is the Meanshift algorithm for a set of datapoints *X*:

1. For each datapoint *x* ∈ *X*, find the neighbouring points *N(x)* of *x*.

2. For each datapoint *x* ∈ *X*, calculate the *mean shift* *m(x)* from this equation:

3. For each datapoint *x* ∈ *X*, update *x* ← *m(x)*.

4. Repeat 1. for *n\_iteations* or until the points are almost not moving or not moving.

The most important piece is calculating the mean shift *m(x)*. The formula in step 2. looks daunting but let’s break it down. Notice the red red encircled parts are essentially the same:

Let’s replace that with *W*i, so the formula becomes this:

Look at the general formula for [weighted average](https://en.wikipedia.org/wiki/Weighted_arithmetic_mean) in Wikipedia gives us this:

Which is just the same thing! Essentially, the meanshift is just calculating **the** *weighted average* of the affected points w.r.t. to *x*. From this perspective, the formula is less mystifying, right?

To summarize: The algorithm finds a set of nearby points that affect a datapoint, then shift it towards where most of the points are, and the closest points have more influence than the further points. Repeat this for all datapoints until nothing changes.