A Computer Aided Diagnosis System for Pancrea Cancer Detection Abstract

Pancrea diseases are the most common disease which causes mortality worldwide. In this study, the computed tomography images are used for the diagnosis of the pancrea diseases such as normal, small cell pancrea carcinoma, large cell pancrea carcinoma and non-small cell pancrea carcinoma by the effective extraction of the global features of the images and feature selection techniques. The images are recognized with the statistical and the shape based features. The texture based features are extracted by Gabor filtering, the feature outputs are combined by watershed segmentation and the fuzzy C means clustering. Feature selection techniques such as Information Gain, correlation based feature selection are employed with Genetic algorithm which is used as an optimal initialization of the clusters. The dataset of pancrea diseases for four classes are considered and the training and testing are done by the knn and svm.

Introduction

Most methods for automated image classification do not work directly with image data, but first extract a higherlevel description of useful features from the image. The choice of features determines a large part of the classification performance. Which features work well depends on the nature of the classification problem: for example, some problems require features that preserve and extract scale differences, whereas other problems require features that are invariant to those properties. Often, feature representations are based on standard filter banks of common feature descriptors, such as Gaussian derivatives that detect edges in the image. These predefined filter banks are not specifically optimized for a particular problem or dataset.

As an alternative to such predefined feature sets, representation learning or feature learning methods [1] learn a highlevel representation directly from the training data. Because this representation is learned from the training data, it can be optimized to give a better description of the data. Using this learned representation as the input for a classification system might give a better classification performance than using a generic set of features. Most feature learning methods use unsupervised models that are trained with unlabeled data. While this can be an advantage because it makes it easier to create a large training set, it can also lead to suboptimal results for classification, because the features that these methods learn are not necessarily useful to discriminate between classes. Unsupervised feature learning tends to learn features that model the strongest variations in the data, while classifiers need features that discriminate between classes. If the variation between samples from the same class is much stronger than the variation between classes, feature learning probably produces features that capture primarily within-class variation. If those features do not represent enough between-class variation, they might give a lower classification performance. This issue of within-class variation is relevant for many applications, including medical image analysis. For example, in disease classification, the differences between patients are often greater than the subtle differences between disease patterns. As a result, representation learners might learn features that model these between-patient differences, rather than those that improve classification.

Literature Survey

[1] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," Universite de Montr ´ eal, Tech. Rep., ´ 2012.

The success of machine learning algorithms generally depends on data representation, and we hypothesize that this is because different representations can entangle and hide more or less the different explanatory factors of variation behind the data. Although specific domain knowledge can be used to help design representations, learning with generic priors can also be used, and the quest for AI is motivating the design of more powerful representation-learning algorithms implementing such priors. This paper reviews recent work in the area of unsupervised feature learning and deep learning, covering advances in probabilistic models, autoencoders, manifold learning, and deep networks. This motivates longer term unanswered questions about the appropriate objectives for learning good representations, for computing representations (i.e., inference), and the geometrical connections between representation learning, density estimation, and manifold learning.

[2] H. Larochelle, M. Mandel, R. Pascanu, and Y. Bengio, "Learning Algorithms for the Classification Restricted Boltzmann Machine," Journal of Machine Learning Research, vol. 13, pp. 643–669, Mar. 2012.

Recent developments have demonstrated the capacity of restricted Boltzmann machines (RBM) to be powerful generative models, able to extract useful features from input data or construct deep arti- ficial neural networks. In such settings, the RBM only yields a preprocessing or an initialization for some other model, instead of acting as a complete supervised model in its own right. In this paper, we argue that RBMs can provide a self-contained framework for developing competitive classifiers. We study the Classification RBM (ClassRBM), a variant on the RBM adapted to the classification setting. We study different strategies for training the ClassRBM and show that competitive classi- fication performances can be reached when appropriately combining discriminative and generative training objectives. Since training according to the generative objective requires the computation of a generally intractable gradient, we also compare different approaches to estimating this gradient and address the issue of obtaining such a gradient for problems with very high dimensional inputs. Finally, we describe how to adapt the ClassRBM to two special cases of classification problems, namely semi-supervised and multitask learning.

[3] G. Desjardins and Y. Bengio, "Empirical Evaluation of Convolutional RBMs for Vision," Universite de Montr ´ eal, Tech. Rep., 2008.

Convolutional Neural Networks (CNN) have had great success in machine learning tasks involving vision and represent one of the early successes of deep networks. Local receptive fields and weight sharing make their architecture ideally suited for vision tasks by helping to enforce a prior based on our knowledge of natural images. This same prior could also be applied to recent developments in the field of deep networks, in order to tailor these new architectures for artificial vision. In this context, we show how the Restricted Boltzmann Machine (RBM), the building block of Deep Belief Networks (DBN), can be adapted to operate in a convolutional manner. We compare their performance to standard fully-connected RBMs on a simple visual learning task and show that the convolutional RBMs (CRBMs) converge to smaller values of the negative likelihood function. Our experiments also indicate that CRBMs are more efficient than standard RBMs trained on small image patches, with the CRBMs having faster convergence.

[4] M. Norouzi, M. Ranjbar, and G. Mori, "Stacks of Convolutional Restricted Boltzmann Machines for Shift-Invariant Feature Learning," in IEEE Conference on Computer Vision and Pattern Recognition, 2009.

In this paper we present a method for learning classspecific features for recognition. Recently a greedy layerwise procedure was proposed to initialize weights of deep belief networks, by viewing each layer as a separate Restricted Boltzmann Machine (RBM). We develop the Convolutional RBM (C-RBM), a variant of the RBM model in which weights are shared to respect the spatial structure of images. This framework learns a set of features that can generate the images of a specific object class. Our feature extraction model is a four layer hierarchy of alternating filtering and maximum subsampling. We learn feature parameters of the first and third layers viewing them as separate C-RBMs. The outputs of our feature extraction hierarchy are then fed as input to a discriminative classifier. It is experimentally demonstrated that the extracted features are effective for object detection, using them to obtain performance comparable to the state-of-the-art on handwritten digit recognition and pedestrian detection.

[5] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, “Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations,” in ICML, New York, New York, USA, 2009, pp. 609–616

There has been much interest in unsupervised learning of hierarchical generative models such as deep belief networks. Scaling such models to full-sized, high-dimensional images remains a difficult problem. To address this problem, we present the convolutional deep belief network, a hierarchical generative model which scales to realistic image sizes. This model is translation-invariant and supports efficient bottom-up and top-down probabilistic inference. Key to our approach is probabilistic max-pooling, a novel technique which shrinks the representations of higher layers in a probabilistically sound way. Our experiments show that the algorithm learns useful high-level visual features, such as object parts, from unlabeled images of objects and natural scenes. We demonstrate excellent performance on several visual recognition tasks and show that our model can perform hierarchical (bottom-up and top-down) inference over full-sized images.

Objectives

The main objective is

- To implement SVM and KNN classifier
- To achieve accuracy of minimum 80%.

Problem Statement

- The identification of the diseases in the CT images were employed based on the classification of the images using classification algorithms.
- Rule-based classifiers were employed which identifies the defects by grouping the pixels that were having similar results for rules used into a group.
- The rules used in the classifier differs based on the applications for which the process is used. Linear discriminant Analysis (LDA) classifiers identifies the defects in the regions based on the difference between the features in the images.
- The artificial neural networks (ANN) classifiers classifies the images based on the similar patterns in the features extracted

Disadvantages of Existing System

- The classification of the features were compared based on the different classifiers. The feature extraction process were not compared.
- The estimation of optimal features were not employed in the existing methods.
- The textural features were more often used in the existing methods. In medical images the statistical parameters and the intensity based features were also needed to get the best features from the images

Methodology

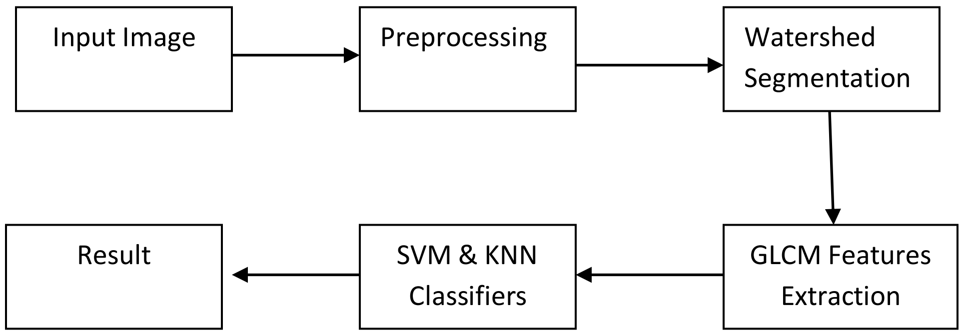
- The ROI region in the pancrea CT images were obtained and the features were extracted from those regions.
- . The extracted features were very large in number and hence the best features were selected from the extracted features.

- The selected features were then classified using different classifiers like SVM and KNN
- The performance of the process were measured based on the performance metrics like Accuracy, Sensitivity and Specificity.

Advantages

- The comparison of the classifiers helps in the study of the different types of classifiers.
- The feature selection process employed is effective for selecting the best features from the dataset.
- The efficiency of the feature selection process is due to the fisher criterion optimization included in the Genetic algorithm.
- The performance measures used proves that the proposed feature extraction method is more efficient while applying the features using SVM classifier

Block Diagram



1 Image Preprocessing

Picture improvement procedures in Picture Preparing Tool kit empower you to expand the flag to-clamor proportion and emphasize picture includes by changing the hues. Edge identification calculations let you distinguish question limits in a picture. These calculations incorporate the Sobel, Prewitt, Roberts, Watchful, and Laplacian of Gaussian strategies. The Watchful strategy can identify genuine feeble edges without being tricked by commotion.

A portion of the improvements Strategies are

- Differentiate Extending
- Commotion separating
- Histogram change

3.1.1 Adaptive Median Filtering

In our Venture We are making utilization of straight Adaptive Median Filtering. It will directly extends the first advanced estimations of the remotely detected information into new appropriation which is as appeared in figure 3.1.1

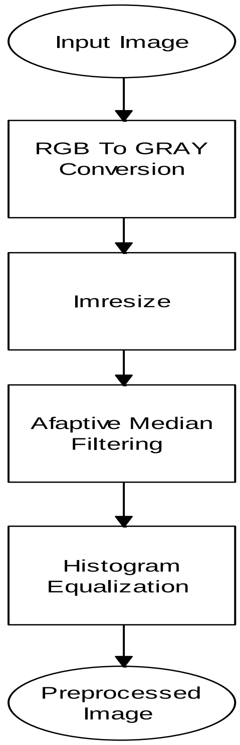


Figure 1.1.1 flowchart for Image Preprocessing

The above figures 3.1 Show the flow Chart For Preprocessing of an image. At the start the Original image is Converted from RGB To Gray and Then it is resized To required Size. Later For which adaptive Median Filter and Histogram Equalization Techniques are Applied . Finally We get the [preprocessed Output.

3.2. Image Segmentation

Image segmentation is the way toward separating a picture into different parts. This is commonly used to recognize objects or other important data in computerized pictures. The objective of division is to improve or potentially change the portrayal of a picture into something that is more significant and simpler to examine.

3.2.1. Marker Controlled Watershed Segmentation

This count considers the information picture as a topographic surface (where higher pixel regards mean higher height) and reenacts its flooding from specific seed centers or markers. A typical decision for the markers are the nearby minima of the slope of the picture, yet the technique takes a shot at a particular marker, either chose physically by the client or decided naturally by another calculation. The yield of marker control watershed division yield is as appeared in fig 3.2.1

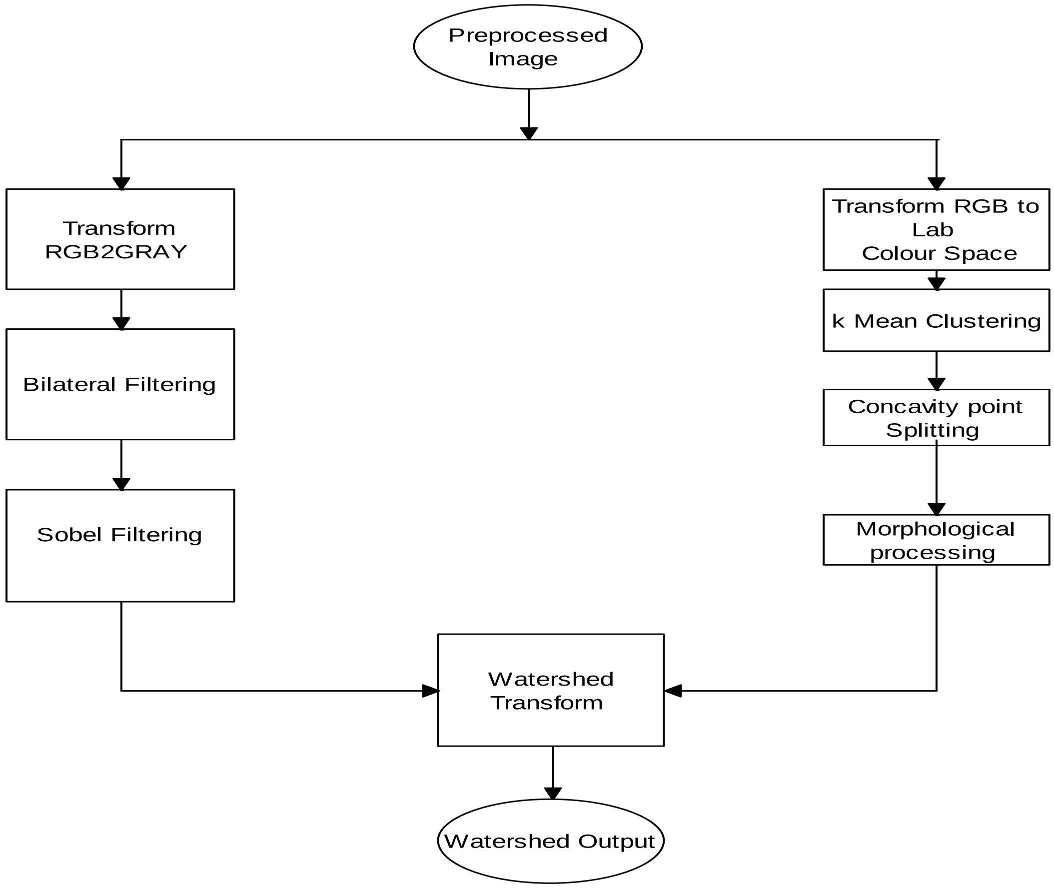


Figure 3.2.1 Flowchart of Watershed Transform

Fig.3.2.1. Shows The detailed Step By Step flow of watershed segmentation. At the beginning the preprocessed image is Given as input to the watershed segmentation. At beginning the flow starts with two part. In first part the preprocessing of image is Done as shown In figure.

In second Stage the image is transformed into L*a*b Color space from RGB. For which k-mean Clustering is obtained to find Concavity point splitting. Which is ended with Morphological processing. At the end output of both preprocessing and marker extraction is used to find watershed transform of an image.

Feature Extraction

In machine learning, plan affirmation and in picture get ready, highlight extraction starts from a basic course of action of measured data and amasses induced esteems (highlights) proposed to be valuable and non-dull, empowering the following learning and theory steps, and once in a while provoking better human understandings. Incorporate extraction is related to dimensionality diminishment. Exactly when the data to a count is excessively broad, making it impossible to be in any capacity arranged and it is suspected to be abundance (e.g. a comparable estimation in both feet and meters, or the repetition of pictures presented as pixels), at that point it can be changed into a diminished plan of parts (moreover named a component vector). Choosing a subset of the fundamental segments is called incorporate assurance. The picked parts are required to contain the essential information from the data, so that the desired task can be performed by using this lessened depiction as opposed to the whole beginning data.

Morphological Processing

Dilation - grow image regions

Dilation causes objects to dilate or grow in size. The amount and the way that they grow depends upon the choice of the structuring element . Dilation makes an object larger by adding pixels around its edges.

The Dilation of an Image 'A' by a structuring element 'B' is written as **A⊕B**. To compute the Dilation, we position 'B' such that its origin is at pixel co-ordinates (x , y) and apply the rule.

$$g(x , y) = \begin{cases} 1 \text{ if 'B' hits 'A'} \\ 0 \text{ Otherwise} \end{cases}$$

Repeat for all pixel co-ordinates. Dilation creates new image showing all the location of a structuring element origin at which that structuring element HITS the Input Image. In this it adds a layer of pixel to an object, there by enlarging it. Pixels are added to both the inner and outer boundaries of regions, so Dilation will shrink the holes enclosed by a single region and make the gaps between different regions smaller. Dilation will also tend to fill in any small intrusions into a region's boundaries.

The results of Dilation are influenced not just by the size of the structuring element but by its shape also.

Dilation is a Morphological operation; it can be performed on both Binary and Grey Tone Images. It helps in extracting the outer boundaries of the given images.

For Binary Image:

Dilation operation is defined as follows,

$D(A, B) = A \oplus B$

Where,

A is the image

B is the structuring element of the order 3 * 3.

Many structuring elements are requested for Dilating the entire image.

Erosion - shrink image regions

Erosion causes objects to shrink. The amount of the way that they shrink depend upon the choice of the structuring element. Erosion makes an object smaller by removing or Eroding away the pixels on its edges.

The Erosion of an image 'A' by a structuring element 'B' is denoted as **$A \Theta B$**. To compute the Erosion, we position 'B' such that its origin is at image pixel co-ordinate (x , y) and apply the rule.

$$g(x, y) = \begin{cases} 1 \text{ if 'B' Fits 'A'}, \\ 0 \text{ otherwise.} \end{cases}$$

Repeat for all x and y or pixel co-ordinates. Erosion creates new image that marks all the locations of a Structuring elements origin at which that Structuring Element Fits the input image. The Erosion operation seems to strip away a layer of pixels from an object, shrinking it in the process. Pixels are eroded from both the inner and outer boundaries of regions. So, Erosion will enlarge the holes enclosed by a single region as well as making the gap between different regions larger. Erosion will also tend to eliminate small extrusions on a regions boundaries.

The result of erosion depends on Structuring element size with larger Structuring elements having a more pronounced effect & the result of Erosion with a large Structuring element is similar to the result obtained by iterated Erosion using a smaller structuring element of the same shape.

Erosion is the Morphological operation, it can be performed on Binary and Grey images. It helps in extracting the inner boundaries of a given image.

For Binary Images:

Erosion operation is defined as follows,

$E(A, B) = A \ominus B$

Where,

A is the image

B is the structuring element of the order 3 * 3.

Many structuring elements are required for eroding the entire image.

Opening - structured removal of image region boundary pixels

It is a powerful operator, obtained by combining Erosion and Dilation. “Opening separates the Objects”. As we know, Dilation expands an image and Erosion shrinks it. Opening generally smoothes the contour of an image, breaks narrow Isthmuses and eliminates thin Protrusions.

The Opening of an image ‘A’ by a structuring element ‘B’ is denoted as A ○ B and is defined as an Erosion followed by a Dilation, and is written as [3],

$$A \circ B = (A \ominus B) \oplus B$$

Opening operation is obtained by doing Dilation on Eroded Image. It is to smoothen the curves of the image. Opening spaces objects that are too close together, detaches objects that are touching and should not be, and enlarges holes inside objects.

Opening involves one or more Erosions followed by one Dilation.

CLOSING

It is a powerful operator, obtained by combining Erosion and Dilation. “Closing, join the Objects”. Closing also tends to smooth sections of contours but, as opposed to Opening, it generally fuses narrow breaks and long thin Gulf’s, eliminates small holes and fills gaps in the contour.

The Closing of an image ‘A’ by a structuring element ‘B’ is denoted as A ● B and defined as a Dilation followed by an Erosion; and is written as [3],

$$A \bullet B = (A \oplus B) \ominus B$$

Closing is obtained by doing Erosion on Dilated image. Closing joins broken objects and fills in unwanted holes in objects.

Closing involves one or more Dilations followed by one Erosion.

3.3.1. Gray Level Co-Occurrence Matrix (GLCM)

One of the most commonly used technique to extract textural data of Images is Gray Level Co occurrence Matrix (GLCM). The GLCM technique gives sensible surface data of a picture that can be acquired just from two pixels. Dark level co-event frameworks acquainted by Haralick [30] endeavor with portray surface by measurably inspecting how certain dim levels happen in connection to other dim levels. Assume a picture to be broke down is rectangular and has Nx lines and Ny Levels. Accept that the dim level showing up at every pixel is quantized to Ng levels.

Let L_x={1,2,3…N_x} be the flat spatial space, $L_{y=}${1,2,3….Ny} be the vertical spatial area, and G={0,1,2,….N_g-1} be the arrangement of Ng quantized dim levels. The set Lx×Ly is the arrangement of pixels of the picture requested by their row column assignments. At that point the picture I can be spoken to as an element of co-event framework that doles out some dark level in Lx×Ly; I: Lx×Ly→G. The dim level moves are ascertained in light of the parameters, dislodging (d) and rakish introduction (θ). By utilizing a separation of one pixel and edges quantized to 450 interims, four lattices of flat, first corner to corner, vertical, and second slanting (0, 45, 90 and 135 degrees) are utilized. At that point the un-standardized recurrence in the four primary headings is characterized by Condition 3. 3.1(a)

$$P(i,j,d,\theta) = \#$$

$$([k.l],(m,n) \in |(L_x \times L_y) \times (L_x \times L_y))$$

$$k-m=0, |l-n| = d)\; or (k-m=d, l-n=-d)$$

$$or\;\; (k-m=-d, l-n=d)\;\; or(|l-m| = d, l-n=0) \qquad --- \quad (1)$$

$$or(k-m=d, l-n=d)\;\; or(k-m=-d, l-n=-d)$$

$$I(k,l) \;\; = i, I(m,n) = j$$

where # is the quantity of components in the set, (k, l) the directions with dark level i, (m, n) the directions with dim level j. The accompanying Figure 3.3.1 represents the above meanings of a co-event network (d=1, θ= 00):

3	3	3
1	3	3
1	3	2

(a)

0º	1	2	3
1	0	0	2
2	0	0	0
3	0	1	3

(b)

45º	1	2	3
1	0	0	2
2	0	0	0
3	0	0	2

(c)

90º	1	2	3
1	1	0	1
2	0	0	1
3	0	0	3

(d)

135º	1	2	3
1	0	0	0
2	0	0	1
3	0	0	2

(e)

Figure 3.3.1(a) An example of GLCM

Despite the fact that Haralick extricated 24 parameters from co-event network, just seven are generally utilized, for example, vitality, entropy, differentiate, neighborhood homogeneity, relationship, group shade what's more, group noticeable quality as given in Conditions (2.2) to (2.8) and is put away in highlight database. In expansion, the principal arrange measurable elements (i.e., mean and standard deviation (StdDev) are utilized to portray the qualities of picture as appeared in Conditions (2.9) to (2.10) individually. The first what's more, second request factual elements are demonstrated as follows:

Vitality measures the quantity of rehashed sets and furthermore measures consistency of the standardized Network

$$Entrophy = \sum_{i,j=0}^{N-1} -\ln(P_{ij})P_{ij} \qquad (2)$$

The differentiation highlight is a distinction snapshot of the P lattice and is a standard estimation of the measure of nearby varieties show in a picture. The higher the estimation of differentiation are, the more honed the auxiliary varieties in the picture

$$Energy = \sum_{i,j=0}^{N-1} -lnP_{ij}{}^{2} \qquad (3)$$

The contrast feature is a distinction snapshot of the P lattice and is a standard estimation of the measure of neighborhood varieties show in a picture. The higher the estimation of complexity are, the more honed the auxiliary varieties in the picture.

$$Contrast = \sum_{i,j=0}^{N-1} P_{ij}\,(i-j)^{\,2} \qquad (4)$$

It quantifies the sameness of the dispersion of components in the GLCM to the GLCM corner to corner. The opposite of homogeneity results in the announcement of differentiation.

$$Local\ Homogenity = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2} \qquad (5)$$

where Pij is the pixel esteem in location (i,j) of the surface picture, N is the quantity of dark levels in the picture, μ=∑_ (i,j=0)^(N-1) iP〗 _ij is mean of the surface picture and σ=∑_(i,j=0)^(N-1) 〖P_ij 〖(i-μ)〗 ^2 〗 is difference of the surface picture.

Correlation is the measure of similarity between two images in likeness. The measures mean (m), which represents the average intensity

$$Correlation = \sum_{i,j=0}^{N-1} P_{ij} \frac{(i-\mu)\ (j-\mu)}{\sigma^2} \qquad (6)$$

$$Clusre\ Shade = \sum_{i,j=0}^{N-1} P_{ij}(i-M_x+j-M_y)^2 \qquad 3.3.1.(f)$$

$$Cluster\ prominence = \sum_{i,j=0}^{N-1} P_{ij}(i-M_x+j-M_y)^4 \qquad (7)$$

Where $M_x = \sum_{i,j=0}^{N-1} iP_{ij}$ and $M_y = \sum_{i,j=0}^{N-1} P_{ij}(i-\mu)^2$

$$mean(m) = \sum_{i=0}^{L-1} z_i P(Z_i)$$

The measures mean (m), which represents the average intensity

$$Standard\ Deviation(\sigma^2) = \sum_{i=0}^{L-1} (x_i - m)^2 P(x_i) \qquad (8)$$

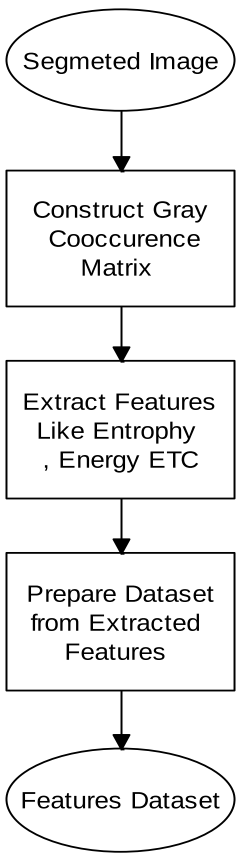


Figure 3.3.1(b) Flowchart for GLCM

Classification

Classification /Arrangement is the naming of a pixel or a gathering of pixels in light of its dark esteem. Grouping is a standout amongst the frequently utilized techniques for data extraction. In Order, generally numerous elements are utilized for an arrangement of pixels i.e., many pictures of a specific question are required. In Remote Detecting region, this system expect that the symbolism of a particular geographic territory is gathered in numerous districts of the electromagnetic range and that the pictures are in great enlistment.

3.4.1. Support Vector Machines (SVM)

• Used generally for arrangement/Classification (additionally, can be adjusted for relapse and notwithstanding for unsupervised learning applications).

• It meets exactness tantamount to Multilayer Perceptrons

Definitions of SVM and Margin

Find $f(x) = (W^T X + b)$with maximum margin, such that for points closer to the separating hyperplane, $|W^T X_i + b|$ (also called **the support vectors**)and for other points, $|W^T X_i + b| > 1$

that **w** is a vector perpendicular to the hyper plane, so we have:

$$f(\mathbf{x}) = f(\mathbf{x_p} + \frac{\mathbf{w}}{\|\mathbf{w}\|} \cdot r) = \mathbf{w^T x_p} + \mathbf{w^T} \frac{\mathbf{w}}{\|\mathbf{w}\|} r + \mathbf{b} = \|\mathbf{w}\| \cdot r \text{ (since } \mathbf{w^T x_p} + \mathbf{b} = \mathbf{0}) \tag{9}$$

Therefore : $r = \dfrac{f(\mathbf{x})}{\|\mathbf{w}\|}$

Now, solve for margin length **ρ:**

$$\boldsymbol{\rho} = \frac{f(\mathbf{x}_\oplus)}{\|\mathbf{w}\|} - \frac{f(\mathbf{x}_\ominus)}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \tag{10}$$

Hypothetical Defense:

The accompanying imbalance could be inferred:

some steady

$$h \leq \frac{R^2}{\rho^2} + 1 \tag{11}$$

margin h speaks to the VC measurement that measures how effective the learning calculation is. It is desirable over utilize the easiest conceivable calculation that gains adequately accurantly from the given information.

In this way, we need to limit h.

Accepting a straightly detachable dataset, the errand of learning coefficients w and b of bolster vector machine $f(\mathbf{x}) = (\mathbf{w}^T \mathbf{x}_i + b)$ decreases to taking care of the accompanying obliged streamlining issue: $\frac{1}{2}\|\mathbf{w}\|^2$

subject to imperatives: $y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad \forall i$

This streamlining issue can be comprehended by utilizing the Lagrangian work characterized as:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} - \sum_{i=1}^{N} \alpha_i [y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) - 1], \text{ such that } \alpha_i \geq 0, \forall i$$

where $\alpha_1, \alpha_2, \ldots \alpha_N$ are Lagrange multipliers and $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots \alpha_N]^T$.

The arrangement of the first obliged streamlining issue is controlled by the seat purpose of L(w,b, α) which must be limited concerning w and b and amplified regarding α.

Comments about Lagrange multipliers:

The arrangement of the first obliged improvement issue is dictated by the seat purpose of $L(\mathbf{w},b,\boldsymbol{\alpha})$ L(w,b, $\boldsymbol{\alpha}$) which must be limited concerning w and b and expanded as for $\boldsymbol{\alpha}$.. If $y_i(\mathbf{w}^T \mathbf{x}_i + b) > 1$, the value of α_i that maximizes L(w,b,$\boldsymbol{\alpha}$) is $\alpha_i = 0$.

If $y_i(\mathbf{w}^T \mathbf{x}_i + b) < 1$, the value of α_i that increases L(w,b,$\boldsymbol{\alpha}$) is $\alpha_i = +\infty$. However, since **w** and b are trying to decreases L(w,b,$\boldsymbol{\alpha}$), they will be altered in such a way to make $y_i(\mathbf{w}^T \mathbf{x}_i + b)$ at least equal to +1.

From this brief discussion, the so-called **Kuhn_Tucker Conditions** follow:

$$\alpha_i \{ y_{i(} W^T X_{i+b)-1} \} = 0, \forall i \qquad (12)$$

Notation:

Data points $\mathbf{x}_i$ with $\alpha_i > 0$ are called the **support vectors**

Optimality conditions:

The essential conditions for the saddle point of $L(\mathbf{w}, b, \boldsymbol{\alpha})$ are

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W_j}} = 0, \forall \mathrm{j} \quad (13)$$

$$\frac{\partial \mathrm{L}}{\partial \alpha \mathrm{W_i}} = 0, \forall \mathrm{i} \quad (14)$$

or, stated a different way, $\nabla_{\mathbf{w}} L = 0, \quad \nabla_{\boldsymbol{\alpha}} L = 0$

Solving for the essential conditions results in

$$w = \sum_{i=1}^{N} \alpha_i \, y_i x_i \quad (15)$$

$$\sum_{i=1}^{N} a_i \, y_i = 0 \quad (16)$$

By restoring $w = \sum_{i=1}^{N} \alpha_i \, y_i x_i$ into the Lagrangian function and by using $\sum_{i=1}^{N} a_i \, y_i = 0$

as a new constrain the **dual optimization problem** can be constructed as

Find $\boldsymbol{\alpha}$ that increases $\sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \ \alpha_j y_i x_j$ subject to

$\sum_{i=1}^{N} a_i \ \ y_i = 0,_{\alpha_i \geq 0},$

This is a curved quadratic programming issue, so ther is a worldwide least. There are various improvement schedules fit for taking care of this enhancement issue. The streamlining can be illuminated in O(N3) time (cubic with the extent of preparing information) and in direct time in the quantity of characteristics. (Contrast this with neural systems that are prepared in O(N) time)

Bolster Vector Machine: Last Indicator

Given the qualities α1, α2, … αN acquired by arrangement of the double issue, the last SVM indicator can be communicated from as

$$\boldsymbol{f(x) = W^T\ X_i + b = \sum_{i=1}^{N} \alpha_i\ y_i X_i^T\ X + b} \tag{17}$$

where

$$\boldsymbol{b = \frac{1}{\left|I_{support}\right|}\quad i.e\ \ \sum_{I_{support}} \left(y_i - \sum_{i=1}^{N} \alpha_i\ y_i X_i^T\ X + b\right)} \tag{18}$$

I support is the arrangement of bolster vectors.

Vital remarks:

To acquire the expectation, all information focuses from the preparation information are counseled Since αi α 0 just for the bolster vectors, just bolster vectors are utilized as a part of giving a forecast Take note of that is a scalar .

Bolster Vector Machine: Directly Nonseparable Case Up until this point, we have examined the development of bolster vector machines on straightly divisible preparing information. This is an extremely solid suspicion that is farfetched in most genuine applications.

Solution: Introducing **the slack variables** ξ_i, i=1, 2, …, N, to relax the constraint

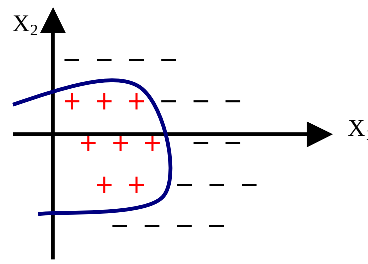
$$y_i(W^T X_i + b) \geq 1\ to\ y_i(W^T X_i + b) \geq 1 - \varepsilon_{1,}\varepsilon_1 \geq 0$$

In a perfect world, one would lean toward every slack variable to be zero and this would compare to the directly detachable case. Subsequently, the enhancement issue for development of SVM on directly nonseparable information is characterized as:

find **w** and b that minimize: $\frac{1}{2}\|w^2\| + c\sum_i \in^2$

subject to: $\mathbf{y_{i(w^T x_i + b) \geq 1 - \in_i, \in_i} \geq 0}$

Problem: Support vector machines portrayed with a straight capacity f(x) (i.e. an isolating hyperplane) have exceptionally constrained representational power. Thusly, they couldn't be extremely helpful in functional characterization issues.



Uplifting news: With a little change, SVM could take care of exceedingly nonlinear grouping issues!

Avocation: Cover's Hypothesis

Assume that informational collection D is nonlinearly divisible in the first property space. The quality space can be changed into another property space where D is directly divisible!

Caveat: Cover's Hypothesis just demonstrates the presence of the changed property space that could take care of the nonlinear issue. It doesn't give the rule to e development of the quality change!

SVM answer for grouping

Denote Φ: $\Re^M \rightarrow F$ as a mapping from the first M-dimensional credit space to the exceptionally dimensional characteristic space F.

By taking care of the accompanyingdouble issue

where C > 0 is an appropriately picked parameter. The additional term maintains each slack variable to be as close to zero as could be normal in light of the current situation.

Double issue: As in the directly detachable issue, this improvement issue can be changed over to its double issue:

find $\boldsymbol{\alpha}$ that increases subject to

$$\sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i x^t{}_j x_j \qquad \sum_{i-1}^{N} \alpha_i y_j = 0 \tag{19}$$

Take note of: The result of presenting parameter C is in obliging the scope of adequate estimations of Lagrange multipliers α_i. The most proper decision for C will rely on upon the particular informational collection accessible.

Issue: Bolster vector machines spoken to with a straight capacity f(x) (i.e. an isolating hyperplane) have extremely constrained depictional power. In that capacity, they couldn't be extremely helpful in useful characterization issues.

Uplifting news: With a slight adjustment, SVM could take care of exceedingly nonlinear characterization issues!!

Support: Cover's Hypothesis

Assume that informational index D is nonlinearly distinct in the first property space. The quality space can be changed into another characteristic space where D is straightly distinct!

Proviso: Cover's Hypothesis just demonstrates the presence of the changed property space that could take care of the nonlinear issue. It doesn't give the rule to the development of the quality change!

SVM answer for order:

Indicate Φ: $\Re^M \rightarrow F$ as a mapping from the first M-dimensional credit space to the very dimensional characteristic space F.

By taking care of the accompanying double issue

find $\boldsymbol{\alpha}$ that maximizes $$\sum_i \alpha_i - \tfrac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) \quad (20)$$

subject to $$\sum_{i-1}^{N} \alpha_i y_j = 0 \quad (21)$$

the resulting SVM is of the form

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{w}^{\boldsymbol{T}} \emptyset(\boldsymbol{x}_{\boldsymbol{i}}) + \boldsymbol{b} = \sum_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{N}} \boldsymbol{\alpha}_{\boldsymbol{i}} \boldsymbol{y}_{\boldsymbol{i}} \emptyset(\boldsymbol{x}_{\boldsymbol{i}})^{\boldsymbol{T}} \emptyset(\boldsymbol{x}) + \boldsymbol{b} \quad (22)$$

Viable Problem: Although SVM are effective in managing profoundly dimensional characteristic spaces, the way that the SVM preparing scales directly with the quantity of properties, and considering restricted memory space could to a great extent confine the decision of mapping Φ. Arrangement: Kernel Trick

It permits registering scalar items (e.g.) in the first property space. It takes after from Mercer's Theorem:

There is a class of mappings Φ that has the following property:

$$\emptyset(\boldsymbol{x_i})^{\boldsymbol{T}}\emptyset(\boldsymbol{y}) = \boldsymbol{K}(\boldsymbol{x},\boldsymbol{y}) \tag{23}$$

where K is a corresponding kernel function.

Examples of kernel function:

• Gaussian Kernel: $K(\mathbf{x},\mathbf{y}) = e^{\frac{\|\mathbf{x}-\mathbf{y}\|^2}{A}}$, A is a constant

• Polynomial Kernel: $K(\mathbf{x},\mathbf{y}) = (\mathbf{x}^T\mathbf{y}+1)^B$, B is a constant

By introducing the kernel trick:

The dual problem: find α that maximizes

$$\sum_{\boldsymbol{i}} \boldsymbol{\alpha_i} - \frac{\boldsymbol{1}}{\boldsymbol{2}}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{j}} \boldsymbol{\alpha_i\alpha_j y_i x^t{}_j x_j K}(\ \boldsymbol{x_i},\boldsymbol{y_i}\ ) \tag{24}$$

subject to $\sum_{i-1}^{N} \alpha_i y_j = 0$

The resulting SVM is:

$$f(\mathbf{x}) = \mathbf{w}^T\Phi(\mathbf{x}_i) + b = \sum_{i=1}^{N}\alpha_i y_i K(\mathbf{x}_i,\mathbf{x}) + b$$

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{w^T}\emptyset(\boldsymbol{x_i}) + \boldsymbol{b} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{\alpha}\ \ \boldsymbol{i\,y_i K}(\boldsymbol{x_i},\boldsymbol{x}) + \boldsymbol{b} \tag{25}$$

Some down to earth issues with SVM

•**Modeling decisions:** When utilizing a portion of the accessible SVM programming bundles or tool compartments a client ought to pick (1) bit work (e.g. Gaussian bit) and its parameter(s) (e.g. consistent A), (2) steady C identified with the slack factors. A few decisions ought to be analyzed utilizing approval set keeping in mind the end goal to locate the best SVM.

SVM preparing does not weigh well with the measure of the preparation information (i.e. scaling as O(N3)). There are a few arrangements that bestow accelerate of the first SVM calculation:

chunking; begin with a subgroup of D, construct SVM, apply it on all information, include "dangerous" information focuses into the preparation information, evacuate "pleasant" focuses, rehash).

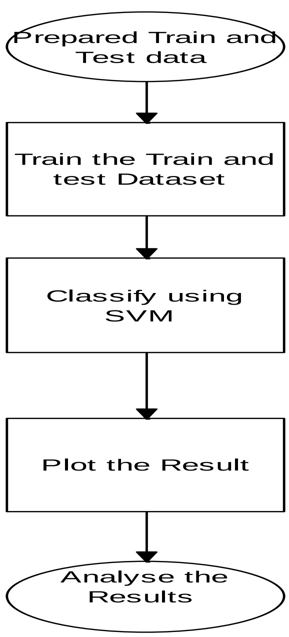


Figure 3.4.1 Flowchart for SVM Classifier

The figures 3.4.1 show the flowchart for the SVM Classifier. From the Results of GLCM Features Prepare the test and train data for SVM Classifier . Later train the prepared trained data and test data into the classifier for classification as shown in flowchart. At the end plot and analyze the result.

k-nearest neighbor algorithm

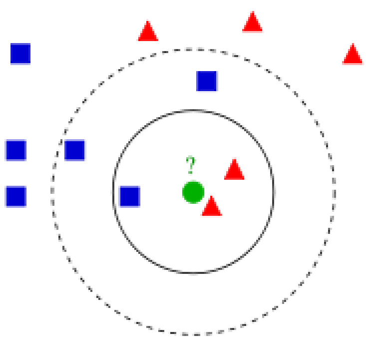
Overview

The *k*-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms. An object is classified by a majority vote of its neighbors, with the object being assigned the class most common amongst its *k* nearest neighbors. *k* is a positive integer, typically small. If $k = 1$, then the object is simply assigned the class of its nearest neighbor. In binary (two class) classification problems, it is helpful to choose *k* to be an odd number as this avoids difficulties with tied votes.

The same method can be used for regression, by simply assigning the property value for the object to be the average of the values of its *k* nearest neighbors. It can be useful to weight the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones.

The neighbors are taken from a set of objects for which the correct classification (or, in the case of regression, the value of the property) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. In order to identify neighbors, the objects are represented by position vectors in a multidimensional feature space. It is usual to use the Euclidean distance, though other distance measures, such as the Manhattan distance could in principle be used instead. The *k*-nearest neighbor algorithm is sensitive to the local structure of the data.

Algorithm



Example of *k*-NN classification. The test sample (green circle) should be classified either to the first class of blue squares or to the second class of red triangles. If $k = 3$ it is classified to the second class because there are 2 triangles and only 1 square inside the inner circle. If $k = 5$ it is classified to first class (3 squares vs. 2 triangles inside the outer circle).

The training examples are vectors in a multidimensional feature space. The space is partitioned into regions by locations and labels of the training samples. A point in the space is assigned to the class *c* if it is the most frequent class label among the *k* nearest training samples. Usually Euclidean distance is used.

The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the actual classification phase, the test sample (whose class is not known) is represented as a vector in the feature space. Distances from the new vector to all stored vectors are computed and *k* closest samples are selected. There are a number of ways to classify the new vector to a particular class, one of the most used technique is to predict the new vector to the most common class amongst the K nearest neighbors. A major drawback to use this technique to classify a new vector to a class is that the classes with the more frequent examples tend to dominate the prediction of the new vector, as they tend to come up in the K nearest neighbors when the neighbors are computed due to their large number. One of the ways to overcome this problem is to take into account the distance of each K nearest neighbors with the new vector that is to be classified and predict the class of the new vector based on these distances.

Domain Explanation

Image Processing is a technique to enhance raw images received from cameras/sensors placed on satellites, space probes and aircrafts or pictures taken in normal day-today life for various applications. Various techniques have been developed in Image Processing during the last four to five decades. Most of the techniques are developed for enhancing images obtained from unmanned space crafts, space probes and military reconnaissance flights. Image Processing systems are becoming popular due to easy availability of powerful personnel computers, large size memory devices, graphics software etc. Image Processing is used in various applications such as:

- Remote Sensing
- Medical Imaging
- Non-destructive Evaluation
- Forensic Studies
- Textiles
- Material Science
- Military etc

The common steps in image processing are image scanning, storing, enhancing and interpretation.

1.1Methods of Image Processing

There are two methods available in Image Processing.

1.1.1Analog Image Processing

Analog Image Processing refers to the alteration of image through electrical means. The most common example is the television image. The television signal is a voltage level which varies in amplitude to represent brightness through the image. By electrically varying the signal, the displayed image appearance is altered. The brightness and contrast controls on a TV set serve to adjust the amplitude and reference of the video signal, resulting in the brightening, darkening and alteration of the brightness range of the displayed image.

1.1.2Digital Image Processing

In this case, digital computers are used to process the image. The image will be converted to digital form using a scanner – digitizer and then process it. It is defined as the subjecting numerical representations of objects to a series of operations in order to obtain a desired result. It starts with one image and produces a modified version of the same. It is therefore a process that takes an image into another. The term digital image processing generally refers to processing of a two-dimensional picture by a digital computer. In a broader context, it implies digital processing of any two-dimensional data. A digital image is an array of real numbers represented by a finite number of bits. The principle advantage of Digital Image Processing methods is its versatility, repeatability and the preservation of original data precision.

1.2 Image Processing Techniques

The various Image Processing techniques are:

- Image representation
- Image preprocessing
- Image enhancement
- Image restoration
- Image analysis
- Image reconstruction
- Image data compression

1.2.1 Image Representation

An image defined in the "real world" is considered to be a function of two real variables, for example, f(x, y) with f as the amplitude (e.g. brightness) of the image at the real coordinate position (x, y).

1.2.2 Image Preprocessing

Scaling

The theme of the technique of magnification is to have a closer view by magnifying or zooming the interested part in the imagery. By reduction, we can bring the unmanageable size of data to a manageable limit. For resampling an image Nearest Neighborhood, Linear, or cubic convolution techniques are used.

I. Magnification

This is usually done to improve the scale of display for visual interpretation or sometimes to match the scale of one image to another. To magnify an image by a factor of 2, each pixel of the original image is replaced by a block of 2x2 pixels, all with the same brightness value as the original pixel.

II. Reduction

To reduce a digital image to the original data, every mth row and mth column of the original imagery is selected and displayed. Another way of accomplishing the same is by taking the average in 'm x m' block and displaying this average after proper rounding of the resultant value.

1.2.3 Image Enhancement Techniques

Sometimes images obtained from satellites and conventional and digital cameras lack in contrast and brightness because of the limitations of imaging sub systems and illumination conditions while capturing image. Images may have different types of noise. In image enhancement, the goal is to accentuate certain image features for subsequent analysis or for image display. Examples include contrast and edge enhancement, pseudo-coloring, noise filtering, sharpening, and magnifying. Image enhancement is useful in feature extraction, image analysis and an image display. The enhancement process itself does not increase the inherent information content in the data. It simply emphasizes certain specified image characteristics. Enhancement algorithms are generally interactive and application dependent.

Some of the enhancements Techniques are

- Contrast Stretching
- Noise filtering
- Histogram modification

1.2.4 Image Analysis

Image analysis is concerned with making quantitative measurements from an image to produce a description of it. In the simplest form, this task could be reading a label on a grocery item, sorting different parts on an assembly line, or measuring the size and orientation of blood cells in a medical image. More advanced image analysis systems measure quantitative information and use it to make a sophisticated decision, such as controlling the arm of a robot to move an object after identifying it or navigating an aircraft with the aid of images acquired along its trajectory. Image analysis techniques require extraction of certain features that aid in the identification of the object. Segmentation techniques are used to isolate the desired object from the scene so that measurements can be made on it subsequently. Quantitative measurements of object features allow classification and description of the image.

1.2.5 Image Segmentation

Image segmentation is the process that subdivides an image into its constituent parts or objects. The level to which this subdivision is carried out depends on the problem being solved, i.e., the segmentation should stop when the objects of interest in an application have been isolated e.g., in autonomous air-to ground target acquisition, suppose our interest lies in identifying vehicles on a road, the first step is to segment the road from the image and then to segment the contents of the road down to potential vehicles. Image thresholding techniques are used for image segmentation.

1.2.6 Classification

Classification is the labeling of a pixel or a group of pixels based on its grey value. Classification is one of the most often used methods of information extraction. In Classification, usually multiple features are used for a set of pixels i.e., many images of a particular object are needed. In Remote Sensing area, this procedure assumes that the imagery of a specific geographic area is collected in multiple regions of the electromagnetic spectrum and that the images are in good registration.

1.2.7 Image Restoration

Image restoration refers to removal or minimization of degradations in an image. This includes de-blurring of images degraded by the limitations of a sensor or its environment, noise filtering, and correction of geometric distortion or non-linearity due to sensors. Image is restored to its original quality by inverting the physical degradation phenomenon such as defocus, linear motion, atmospheric degradation and additive noise.

1.2.8 Image Reconstruction from Projections

Image reconstruction from projections is a special class of image restoration problems where a two- (or higher) dimensional object is reconstructed from several one-dimensional projections. Each projection is obtained by projecting a parallel X-ray (or other penetrating radiation) beam through the object. Planar projections are thus obtained by viewing the object from many different angles. Reconstruction algorithms derive an image of a thin axial slice of the object, giving an inside view otherwise unobtainable without performing extensive surgery. Such techniques are important in medical imaging (CT scanners), astronomy, radar imaging, geological exploration, and nondestructive testing of assemblies.

1.2.9 Image Compression

Compression is a very essential tool for archiving image data, image data transfer on the network etc. They are various techniques available for lossy and lossless compressions. One of most popular compression techniques, JPEG (Joint Photographic Experts Group) uses Discrete Cosine Transformation (DCT) based compression technique. Currently wavelet based compression techniques are used for higher compression ratios with minimal loss of data.

Hardware Specification

Processor	:	Intel Dual Core and Above
RAM	:	2 GB and Above
Hard Disk	:	150 GB and Above
Camera	:	Web Camera

3.4 Software Specification

Operating System	:	Windows 7 and Above
MATLAB Version:		MATLAB8.1 R2013a

3.5 Software Description

3.5.1 MATLAB

MATLAB (**mat**rix **lab**oratory) is a multi-paradigm numerical computing environment and fourth-generation programming language. Developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, Fortran and Python.

Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine, allowing access to symbolic computing capabilities. An additional package, Simulink, adds graphical multi-domain simulation and Model-Based Design for dynamic and embedded systems.

In 2004, MATLAB had around one million users across industry and academia. MATLAB users come from various backgrounds of engineering, science, and economics. MATLAB is widely used in academic and research institutions as well as industrial enterprises.

It is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. Typical uses include:

- Math and computation
- Algorithm development
- Modeling, simulation, and prototyping
- Data analysis, exploration, and visualization
- Scientific and engineering graphics
- Application development, including Graphical User Interface building

MATLAB is an interactive system whose basic data element is an array that does not require dimensioning. This allows you to solve many technical computing problems, especially those with matrix and vector formulations, in a fraction of the time it would take to write a program in a scalar noninteractive language such as C or Fortran.

The name MATLAB stands for matrix laboratory. MATLAB was originally written to provide easy access to matrix software developed by the LINPACK and EISPACK projects, which together represent the state-of-the-art in software for matrix computation.

MATLAB features a family of application-specific solutions called toolboxes. Very important to most users of MATLAB, toolboxes allow you to *learn* and *apply* specialized technology. Toolboxes are comprehensive collections of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems. Areas in which toolboxes are available include signal processing, control systems, neural networks, fuzzy logic, wavelets, simulation, and many others.

MATLAB has evolved over a period of years with input from many users. In university environments, it is the standard instructional tool for introductory and advanced courses in mathematics, engineering, and science. In industry, MATLAB is the tool of choice for high-productivity research, development, and analysis.

3.5.2 MATLAB System

The MATLAB system consists of five main parts:

The MATLAB language

This is a high-level matrix/array language with control flow statements, functions, data structures, input/output, and object-oriented programming features. It allows both "programming in the small" to rapidly create quick and dirty throw-away programs, and "programming in the large" to create complete large and complex application programs.

The MATLAB working environment

This is the set of tools and facilities that you work with as the MATLAB user or programmer. It includes facilities for managing the variables in your workspace and importing and exporting data. It also includes tools for developing, managing, debugging, and profiling M-files, MATLAB's applications.

Handle Graphics

This is the MATLAB graphics system. It includes high-level commands for two-dimensional and three-dimensional data visualization, image processing, animation, and presentation graphics. It also includes low-level commands that allow you to fully customize the appearance of graphics as well as to build complete Graphical User Interfaces on your MATLAB applications.

The MATLAB mathematical function library

This is a vast collection of computational algorithms ranging from elementary functions like sum, sine, cosine, and complex arithmetic, to more sophisticated functions like matrix inverse, matrix eigenvalues, Bessel functions, and fast Fourier transforms.

The MATLAB Application Program Interface (API)

This is a library that allows you to write C and Fortran programs that interact with MATLAB. It include facilities for calling routines from MATLAB (dynamic linking), calling MATLAB as a computational engine, and for reading and writing MAT-files.

System Requirements

SOFTWARE REQUIREMENTS:

OS : Windows

Software : Mat lab

HARDWARE REQUIREMENTS:

Processor : Intel Pentium.

RAM : 2GB

Detailed Work Plan

To collect Dataset

Working on preprocessing.

Working on filtering techniques.

Working on Segmentation.

Working on Classifier.

Conclusion

The system consists of pre-processing, segmentation, feature extraction and final classification. The proposed marker controlled watershed segmentation technique separates the touching objects in the image. It provides best identification of the main edge of the image and also avoids over segmentation. Two different classifiers are used for the classification purpose. Using both SVM and KNN classifier increases accuracy of detection and reduces false detection. The proposed technique gives very promising results comparing with other used techniques.

REFERENCES

[1] Y. Bengio, A. Courville, and P. Vincent, “Representation Learning: A Review and New Perspectives,” Universite de Montr ´ eal, Tech. Rep., ´ 2012.

[2] H. Larochelle, M. Mandel, R. Pascanu, and Y. Bengio, “Learning Algorithms for the Classification Restricted Boltzmann Machine,” Journal of Machine Learning Research, vol. 13, pp. 643–669, Mar. 2012.

[3] G. Desjardins and Y. Bengio, “Empirical Evaluation of Convolutional RBMs for Vision,” Universite de Montr ´ eal, Tech. Rep., 2008. ´

[4] M. Norouzi, M. Ranjbar, and G. Mori, “Stacks of Convolutional Restricted Boltzmann Machines for Shift-Invariant Feature Learning,” in IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[5] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, “Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations,” in ICML, New York, New York, USA, 2009, pp. 609–616.

[6] ——, “Unsupervised learning of hierarchical representations with convolutional deep belief networks,” Communications of the ACM, vol. 54, no. 10, pp. 95–103, Oct. 2011.

[7] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[8] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Muller, “Building a reference multimedia database for interstitial ¨ pancrea diseases,” Computerized Medical Imaging and Graphics, vol. 36, no. 3, pp. 227–38, Apr. 2012.

[9] A. Depeursinge, A. Foncubierta-Rodr´ıguez, D. Van de Ville, and H. Muller, “Pancrea Texture Classification Using Locally-Oriented Riesz ¨ Components,” in MICCAI, 2011, pp. 231–238.

[10] ——, “Multiscale Pancrea Texture Signature Learning Using the Riesz Transform,” in MICCAI, 2012, pp. 517–524.

[11] A. Depeursinge, D. Van de Ville, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Muller, “Near-affine-invariant texture learning for pancrea ¨ tissue analysis using isotropic wavelet frames.” IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society, vol. 16, no. 4, pp. 665–75, Jul. 2012.

[12] A. Depeursinge, T. Zrimec, S. Busayarat, and H. Muller, “3D pancrea ¨ image retrieval using localized features,” in SPIE Medical Imaging, 2011.

[13] Y. Song, W. Cai, Y. Zhou, and D. D. Feng, “Feature-Based Image Patch Approximation for Pancrea Tissue Classification,” IEEE Transactions on Medical Imaging, vol. 32, no. 4, pp. 797–808, Apr. 2013.

[14] Y. Song, W. Cai, H. Huang, Y. Zhou, D. Feng, Y. Wang, M. Fulham, and M. Chen, “Large Margin Local Estimate with Applications to Medical Image Classification,” IEEE Transactions on Medical Imaging, vol. 34, no. 6, pp. 1362–1377, Jun. 2015.

[15] A. Foncubierta-Rodr´ıguez, A. Depeursinge, and H. Muller, “Using ¨ Multiscale Visual Words for Pancrea Texture Classification and Retrieval,” in Medical Content-Based Retrieval for Clinical Decision Support, 2012, pp. 69–79