Modified Weighted Fuzzy C-Means Clustering Algorithm

Pallavi Khare,
Assistant Professor,
Department of Electronics and Telecommunication,
Padmashree Dr. DY. Patil Institute of Engineering and Technology,
Pimpri, Pune-18, India.

Anagha Gaikwad,
Student (BE),
Department of Electronics and Telecommunication,
Padmashree Dr. DY. Patil Institute of Engineering and Technology,
Pimpri, Pune-18, India.

Pooja Kumari,
Student (BE),
Department of Electronics and Telecommunication,
Padmashree Dr. DY. Patil Institute of Engineering and Technology,
Pimpri, Pune-18, India.

***Abstract*—Image segmentation has been a fascinating territory for examination and creating productive calculations, assuming a foremost part in high bore image translation and picture investigation. Image segmentation is defined as action of splitting an image into band of pixels. Aim of this process is to group pixels into regions. This is used in locating objects in satellite images, face recognition, iris recognition, agricultural imaging and medical imaging. Such division requests a powerful division calculation against noise. Clustering is process that make group of similar objects and is widely used technique for brain tumor detection. FCM is very delicate to noise because of the act of just power values for grouping. We proposed a new method Modified FCM algorithm based on the Distance metric for segmentation of images that have been corrupted by intensity inhomogeneities and noise. We define a new trade-off weighted fuzzy factor to adaptively control the local neighbor relationship. It is followed by an Edge detection algorithm using Canny Edge filter to overcome disconnected locales with openings or districts with a solitary pixel encountered during clustering. The edge detection technique is also to extract boundaries from the segmented image.**

Keywords—Segmentation, clustering, trade-off, distance, FCM, edge, Canny.

I. Introduction

Imaging science has long-drawn-out basically along three unique yet related lines of examination: division, enrolment and visualization. Enrolment includes discovering the change that brings distinctive images of the same item into strict spatial (and/or worldly) coinciding. Also visualization includes the presentation, control, and estimation of image information. At last, division is characterized as the procedure of parceling an image into set non-covering districts whose union is the whole image where these locales ought to in a perfect world compare to protests and their significant parts, and foundation. Most image division calculations are focused around two fundamental properties that can be separated from pixel values brokenness and closeness or a mix of them. Division of nontrivial images is a dismal issue made even inflexible by non- uniform lighting, shadows, covering items, poor differentiation in the middle of articles and foundation, thus on with some level of accomplishment to this date. Image division can be approached as three philosophical viewpoints - area, limit and edge. Image preparing strategies for quantitative examination are principally utilized as a part of computational medicinal investigation. Therefore, it gets to be conceivable to enhance the demonstrative exactness. It is the vital yet tricky capacity to precisely perceive and depict all the individual questions in an image scene.

Fundamentally we can cerebrate of a few basic ideas for segmentation. Pixel predicated strategies just use the gray estimations of the individual pixels. Area predicated strategies examine the gray values in all the more colossally enormous zones. Definitively, edge-predicated routines locate edges and after that try to tail them. The predominant demand of all these methodologies is that they are predicated just on nearby data. That being said they use this data just somewhat. Pixel predicated methods don't significantly consider the nearby neighborhood. Edge-predicated strategies search just for discontinuities, while area predicated systems break down homogeneous locales. In circumstances where we ken the geometric state of an item, model-predicated segmentation can be connected.

II. Literature Survey On Segmentation Methods Via Clustering

Clustering is a methodology which segments a given information set or information focuses into homogeneous gatherings predicated on given peculiarities such that related objects will be kept in a bunch while different objects will be in diverse clusters. Each one group is spoken to by its mean (centroid) and difference (spread) connected with the dissemination of the relating feature vectors of the information focuses in the bunch. Cluster analysis has been continuously developed and is used in many scientific disciplines such as biology, psychology, statistics, pattern recognition, economics and finance.Conventional data clustering methods can be classified into two broad categories: Hierarchical and Partitional. In hierarchical clustering, the number of clusters need not be specified a priori where only local neighbors in each step are considered. For this reason, it is difficult to handle overlapping clusters through the hierarchical clustering method. In addition, hierarchical clustering is static i.e. data points allocated to a cluster in the early stages may not be moved to a different cluster. Partitional clustering methods develop a clustering structure by optimizing a criterion function defined either locally (on a subset of the patterns) or globally (defined over all of the data). Partitional clustering can be further divided into two classes: crisp clustering and fuzzy clustering. In crisp clustering, a data point belongs to only one cluster and clusters are separated by crisp boundaries among them. In fuzzy clustering methods, data points belong to all clusters through a degree determined by the membership function.

A. Calculating Distance between Clusters and Popular Distance Measures.

Centroid is the "middle" of a cluster.

$$C_m = \frac{\sum_{i=1}^{N} (t_{ip})}{N}$$

Distances are normally used to measure the similarity or dissimilarity between two data objects. The distance between the centroids of two clusters, i.e., dis (K_i, K_j) = dis(C_i , C_j).

The **Minkowski metric** favors the prime scaled feature, which dominates others. The problem can be addressed by proper normalization or other weighting schemes applied in the feature space where d is the dimensionality of the data.

$d(i,j)=\sqrt[q]{(|x_{i1}-x_{j1}|^q+|x_{i2}-x_{j2}|^q+\ldots)}$

Where i = (x_{i1}, x_{i2}, …, x_{ip}) and j = (x_{j1}, x_{j2}, …, x_{jp}) are two p-dimensional data objects, and q is a positive integer.

If q = 1, d is Manhattan distance

$d(i,j)=|x_{i1}-x_{j1}|+|x_{i2}-x_{j2}|+\ldots+|x_{ip} - x_{jp}|$

If q = 2, d is **Euclidean distance** which is defined as the distance between two points as the length of the line segment connecting them. The Euclidean distance has an intuitive appeal as it is commonly used to evaluate the proximity of objects in 2- or 3D space. It works well when a data set has "compact" or "isolated" clusters. The advantage of Euclidean distance is that it is intuitively obvious. The disadvantages are costly calculation due to the square root, and its Non-integral value.

Linear correlation among features can also distort distance measures. This distortion can be alleviated by applying a whitening transformation to the data by using the **Mahalanobis distance** measure $d_M(x, y)$ as

$d_M(x, y) = ((x - y)A^{-1} (x - y)^T)^{1/2}$

Where A is a covariance matrix.

Kernel based similarity measure capacities map information from data space to high, perhaps boundless, dimensional feature space. For a finite sample of data X, the kernel function yields a symmetric N*N positive definite matrix K, where the K (i, j) entry corresponds to the dot product between f (x i) and f (x $_j$) as measured by the kernel function. In feature space, the distance measure between any two patterns is given by:

$\| f(x_i) - f(x_j) \|^2 = \langle f(x_i), f(x_j)\rangle + \langle f(x_i), f(x_j)\rangle - 2\langle f(x_i), f(x_j)\rangle \| f(x_i) - f(x_j) \|^2 = k(i,i)\ k(i,j)\ 2k(i,j)$

B. K-means clustering

K-means (MacQueen, 1967) is one of the simplest unsupervised learning algorithms. It outlines a conceptually simple way to partition a data set into a specified number of clusters k. The algorithm aims to iteratively minimize a simple squared error objective function of the form

$$J = \sum_{j=1}^{k} \sum_{\substack{\text{all } i \\ \text{in class } j}} | x_i^j - c_j |^2 ,$$

Where c_j denotes the coordinate vector of the jth cluster and $\{x_i^j\}$ are the points assigned to the jth cluster. Minimizing J equivalently means reaching that configuration at which switching any point to a cluster other than its currently assigned one will only increase the objective function.

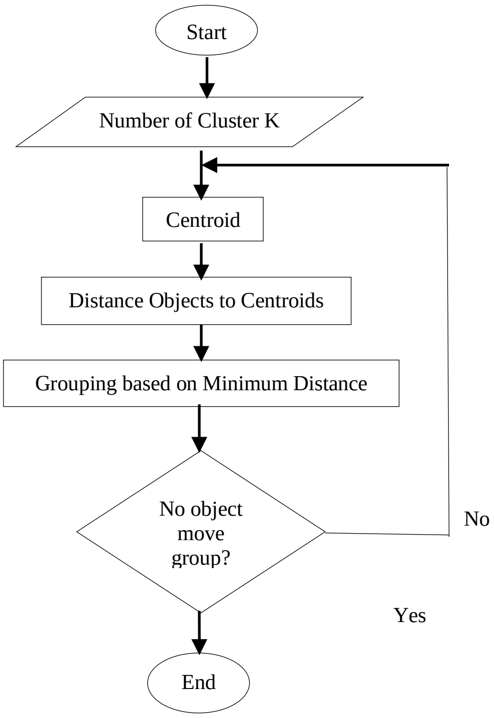

Fig. 1.Algorithm for K-means clustering

Shortcomings of K-means clustering

1) The learning calculation requires apriori assignment of the quantity of group focuses.

2) If there are two highly overlapping data then k-means will not be able to resolve that there are two clusters.
3) The learning calculation is not invariant to non-direct changes i.e. with distinctive representation of information we get diverse results (information spoke to in manifestation of Cartesian co-ordinates and polar directions will give distinctive results).
4) Euclidean separation measures can unequally weight basic components.
5) The learning algorithm provides the local optima of the squared error function.
6) Randomly winnowing of the cluster focus can't lead us to the productive result.
7) Applicable just when mean is characterized i.e. fizzles for absolute information.
8) Unable to handle strepitous information and anomalies.

C. Fuzzy C-means Clustering

Fuzzy c-means (FCM) is a scheme of clustering which allows one section of data to belong to dual or supplementary clusters. This method was developed by Dunn in 1973 and enriched by Bezdek in 1981 and it is habitually used in pattern recognition. Main objective of fuzzy c-means algorithm is to minimize:

$$J(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{c} (u_{ij})^m \|x_i - v_j\|^2$$

Where, '$\|x_i - v_j\|$' is the Euclidean distance between i^{th} data and j^{th} cluster centre.

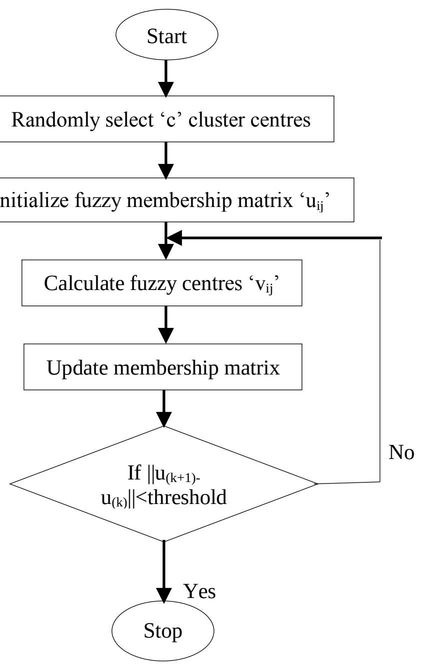

Fig. 2. Fuzzy C-means algorithm

Gains of Fuzzy c-means clustering

1) FCM gives best result for overlapped data set and is comparatively better than k-means algorithm.
2) Data point is assigned membership to each cluster center as a result of which data point may belong to more than one cluster center.

Weaknesses of Fuzzy c-means clustering

1) Apriori measurement of the number of clusters
2) Euclidean distance measures can inequitably weight underlying factors.

D. Kernelized Fuzzy C-means clustering

The basic idea of KFCM is to first map the input data into a feature space with higher dimension via a nonlinear transform and then perform FCM in that feature space. The kernel metric Fuzzy C-Means minimizes the following objective function.

$$J\phi = \sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij}^{m} \|\phi(x_j) - \phi(v_i)\|^2$$

where, u_{ij} denotes the membership of x_j in cluster i, $\phi(v_i)$ is the center of cluster i in the feature space, and ϕ is the mapping from the input space X to the feature space F. Minimization of the function has been proposed only in the case of a Gaussian kernel.

KFCM Algorithm

1) Select initial class prototype $\{V_i\}^c$
2) Update all memberships U_{ij}
3) Obtain the prototype of clusters in the forms of weighted average. Repeat step 2-3 till termination. The termination criterion is $\| V_{new} - V_{old} \| \le \varepsilon$

Where $\|.\|$ is the Euclidean norm. V is the vector of cluster centres ε is a small number that can be set by user (here ε =0.01).

Benefits of KFCM clustering

1) We can acquire a directly distinguishable hyper-plane in the high-dimensional, or even in an endless feature space.
2) They can distinguish clusters with self-assertive shapes.
3) Kernel-based clustering calculations, have the capacity of managing noise and outliners.
4) There is no prerequisite for earlier information to focus the framework topological structure.
5) The kernel matrix can provide the means to estimate the number of clusters.

Downsides of KFCM

1) A precarious issue related to KFCM clustering is the selection of an "optimal" kernel for the problem at hand and on the setting of the involved parameters.
2) The kernel function in use must conform to the learning objectives in order to obtain meaningful results for un-labelled data.

E. FCM with Tradeoff Weighted Fuzzy Factor and Distance metric (KWFLICM algorithm)

We present an improved fuzzy C-means (FCM) algorithm for image segmentation by introducing a tradeoff weighted fuzzy factor and a distance metric. The tradeoff weighted fuzzy factor depends on the space distance of all neighboring pixels and their gray-level difference simultaneously. Trade-off weighted fuzzy factor is used for adaptively controlling the local spatial relationship. By using this factor, the new algorithm can accurately estimate the damping extent of neighboring pixels. Modified FCM technique includes similar steps as FCM except for the variation in the cluster updating and membership value updating criterions. The modified criterions are shown below.

$$C_i = \frac{\sum_{j=1}^{n} (u_{ij})^m y_j}{\sum_{j=1}^{n} (u_{ij})^m} \quad ; \qquad u_{ij} = \frac{1}{\sum_{k=1}^{c} (d_{ij} / d_{kj})^{2/(m-1)}}$$

Where $d_{ij} = y_j - c_i$
y=reduced dataset.

Algorithm of modified Fuzzy C means Clustering with weighted tradeoff fuzzy factor and distance metric (KWFLICM)

Step 1: Set the number of clusters c and the parameter m

$$J_m(u, v) = \sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij})^m d^2(x_j, v_i)$$

Initialize the fuzzy Cluster centroid vector $V=[v_1,\ldots,v_c]$, randomly and set $\varepsilon = 0.01$.
Step 2 : Compute u_{ij}.

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} (d_{ij} / d_{kj})^{2/(m-1)}}$$

Step 3 : Compute v_i.

$$v_i = \frac{\sum_{j=1}^{n} (u_{ij})^m x_j}{\sum_{j=1}^{n} (u_{ij})^m}$$

Step 4 : Update u_{ij}.
Step 5 : Update v_i.
Step 6 : Repeat Steps 4 and 5 until the following termination criterion is satisfied: $|V_{new} - V_{old}| < \epsilon$.

III. Block Diagram Of The Proposed Method

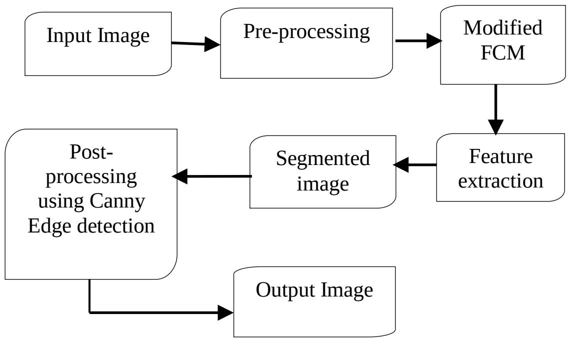

Fig. 3. Framework of the proposed system

The framework comprises of the accompanying pieces. The crude information is passed through the framework as numerical information or as waves. Material methods are connected to get the pre-processed information. Further, the information is passed through the clustering stage, which furnishes a proportional payback focuses. Characteristic extraction is then performed to get the characteristics that can out and out represent a given occasion. Next a post transforming is utilized to improve the nature of the last divided image.

A. Pre-processing

The pre-processing stage is performed to change over all traits of the information into a numeric structure that can be utilized by the clustering methodology. This is greatly valuable for lessening in measurement of the dataset utilizing standardization. In the event that the estimations of a few properties fluctuate in diverse ranges then to lessen the impact of such characteristics, all estimations of the attributes are standardized to lie in some regular reach, in the same way as [0, 1]. Pre-preparing improves the visual appearance of images and controls datasets.

B. Clustering

The clustering is a vital venture, as it is a key antecedent to the peculiarity (feature) extraction. The information for peculiarity extraction is the pre-processed information, where the marks are peeled off. Clustering is a type of unsupervised discovering that serves to discover the intrinsic structure in the information.

C. Feature extraction

It is the procedure by which certain features of interest inside an image are discovered and spoke to for further handling. It denote the move from pictorial to non-pictorial (alphanumerical, generally quantitative) information representation which can be consequently utilized as a data to various example distinguishment and arrangement procedures, which will then mark, group, or perceive the semantic substance of the image or its questions.

D. Post-processing

Image post transforming improves the nature of the completed image, by filtration and different medicines. Here calculation, for example, locale developing, pixel network or a guideline based calculation is connected to get the last portioned areas. In the proposed system we have chosen Canny edge detection technique for post-processing.

1. The Canny edge detector

The Canny edge detector is a standout amongst the most mainstream, compelling, effective and powerful edge detection operators, generally acknowledged as the best 'all-round' edge detection method advanced to date. It utilizes a set of generally extensive, situated filters at various image resolutions and consolidates the individual results into a typical edge map. The technique tries to achieve three primary objectives:

a) To lessen the amount of false edge points.
b) Achieve noble localization of edges.
c) Deliver only a solo mark on each edge.

Canny Edge Detection algorithm

1) The input image is smoothed using a Gaussian low-pass filter with a defined estimation of σ: expansive estimations of σ will suppress much of the noise at the expense of weakening potentially relevant edges.
2) The local gradient (intensity and direction) is registered for each one point in the smoothed image.
3) The edge points at the output of step 2 result in wide ridges. The algorithm thins those ridges, leaving only the pixels at the top of each ridge, in a process known as non-maximal suppression.
4) The edge pixels are then thresholded utilizing two limits T_{low} and T_{high}: edge pixels with qualities more prominent than T_{high} are viewed as solid edge pixels; edge pixels with qualities in the middle of T_{low} and T_{high} are said to be weak pixels. This methodology is known as hysteresis thresholding.
5) The algorithm performs edge linking, accumulating weak pixels that are 8-joined with the solid pixel.

IV. Applications

A. Quantitative or semi-quantitative analytic image investigation.
B. Surgical arranging.
C. Computer aided surgery.
D. Prescription Inspection and understanding of images acquired from X-beams, MRI or CAT examines, investigation of cell images, of chromosome karyotypes.
E. Agriculture: Satellite/airborne perspectives of area, for instance to decide the amount of area is being utilized for distinctive purposes, or to explore the suitability of diverse locales for diverse products, review of foods grown from the ground recognizing great and new create from old.
F. Industry: Automatic assessment of things on a generation line, examination of paper specimens.
G. Law requirement: Fingerprint examination, honing or de-smudging of velocity cam images.

V. Conclusion

Traditional FCM algorithm based pixel attributes lead to accuracy degradation. In this paper we have spoken about the execution of the three calculations FCM, K means and the Modified Fuzzy C Means weighted algorithm (KWFLICM). The aim of this paper is to propose a new Edge detection based clustering algorithm for automatic segmentation of medical images with intensity inhomogeneity. A comparative study proposes the feasibility of the proposed algorithm over FCM and K means clustering calculation. Compared with existing clustering methods, the proposed system is able to incorporate the local information more exactly. In addition, the trade-off weighted fuzzy factor and the distance measure are completely free of the empirically adjusted parameters determination, thereby allowing the automated applications.

VI. Acknowledgment

The authors are enormously obligated to the Staff of the Department of Electronics and Telecommunication Engineering, Pad. Dr. DYPIET for their technical support and all the references used in this paper.

VII. References

[1] S. Krinidis and V. Chatzis, "A robust fuzzy local information C-means clustering algorithm," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1328–1337, May 2010.
[2] Mr.Tara.Saikumar and Dr. M. Sundarana Reddy, "Optimized Kernel Fuzzy C-Means Clustering for Bio-Images Using Level Set Method" , International Journal of Electronics Communication and Computer Engineering, Volume 4, Issue (6), ISSN 2249–071X, National Conference on Recent Trends in Computer Science and Technology (NCRTCST)-2013.
[3] Sayana Sivanand," Adaptive Local Threshold Algorithm and Kernel Fuzzy C-Means Clustering Method for Image Segmentation", International Journal of Latest Trends in Engineering and Technology (IJLTET), Vol. 2 Issue 3, ISSN: 2278-621X, May 2013.
[4] Chiheb Eddine Ben N'Cir[1], Nadia Essoussi[1], Patrice Bertrand[2], "Overlapping clustering based on kernel similarity metric", Larodec, ISG University of Tunis Ceremade[1], Université Paris-Dauphine[2].
[5] Xiaojun LOU, Junying LI and Haitao LIU, "Improved Fuzzy C-means Clustering Algorithm Based on Cluster Density", Journal of Computational Information Systems 8: 2 pp. 727-737, 2012.
[6] Mahdi'eh Motamedi and Hasan Naderi, "Data clustering using Kernel based algorithm", International Journal of Information Technology, Control and Automation (IJITCA) Vol.4, No.3, July 2014.
[7] Srinivasa K G, Venugopal K R and L M Patnaik, "Feature Extraction using Fuzzy C - Means Clustering for Data Mining Systems", International Journal of Computer Science and Network Security, VOL.6 No.3A, March 2006.
[8] Maryam Rastgarpour and Jamshid Shanbehzadeh, "A New Kernel-Based Fuzzy Level Set Method for Automated Segmentation of Medical Images in the Presence of Intensity Inhomogeneity" , Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine Volume 2014, Article ID 978373.
[9] Naouel Baili, "Fuzzy Clustering with Multiple Kernels".
[10] P.Sivasangareswari and K.Sathish Kumar, "Fuzzy C-Means Clustering With Local Information and Kernel Metric for Image Segmentation" , International Journal of Advanced Research in Computer Science &

Technology, Vol. 2 Issue Special 1, ISSN : 2347 - 8446 (Online), ISSN : 2347 - 9817 (Print), Jan-March 2014.

[11] Hesam Izakian, Ajith Abraham," Fuzzy C-means and fuzzy swarm for fuzzy clustering problem". Expert Systems with Applications 38, 1835–1838, 2011.

[12] James C. Bezdek, Robert Ehrlich and William Full, "FCM: The Fuzzy C-means Clustering Algorithm", Computers & Geosciences Vol. 10, No. 2-3, pp. 191-203, 1984.

[13] G.J.Abisha Shiji and Ms. Berakhah .F. Stanley, "Fuzzy Clustering With Spatial Information For Image Segmentation Using Kernel Metric", International Journal of Engineering Research and Applications (IJERA), ISSN: 2248-9622 International Conference on Humming Bird, 01st March 2014.

[14] Liu Yang, "Distance Metric Learning: A Comprehensive Survey", 2006.

[15] Saritha A K, Ameera P.M, "Image segmentation based on kernel fuzzy C means clustering using edge detection method on noisy images", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 2, ISSN: 2278 – 1323, February 2013.

[16] Deepika Singh and Anjana Gosain, "Robust new Distance Kernelized Approach to Distributed Clustering", International Journal of Engineering Science & Advanced Technology, Volume-3, Issue-4, 165-169, ISSN: 2250-3676, Aug-Sep 2013.

[17] D. Napoleon and K. Ragul, "An Efficient Multi-Kernel Fuzzy C-Means Algorithm to Analyze and Segment the fastidious Tumor region in Medical Images", International Journal of Modern Embedded System (IJMES) Volume No.-2, Issue No.-4, ISSN: 2320-9003(Online) August, 2014.

[18] K. M. Bataineh, M. Naji and M. Saqer, "A Comparison Study between Various Fuzzy Clustering Algorithms", Jordan Journal of Mechanical and Industrial Engineering, Volume 5, Number 4, ISSN 1995-6665, Aug. 2011.

[19] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

[20] "Medical image analysis", Atam P. Dhawan, Second edition, published by John Wiley & Sons, Inc., Hoboken, New Jersey, ISBN 978-0-470-622056, 2011.

[21] "Digital Image Processing", Bernd Jähne, 5th revised and extended edition, Springer, ISBN 3-540-67754-2, 2002.

[22] "Fundamentals of Digital Image Processing-A Practical Approach with Examples in Matlab" by Chris Solomon, Toby Breckon, Wiley Blackwell publishing, ISBN 978 0 470 84472 4, 2011.

[23] "Practical Image and Video Processing Using MATLAB", Oge Marques, published by John Wiley & Sons, Inc., Hoboken, New Jersey, ISBN: 9781118093481, 2011.