items m. in case of blocks, we call it as block level recall of b-Recall. **b-Recall = r(r + m)**

C. b-F-measure

Similar to the way it is defined in [1], we can refer to the b-F-measure as the contents are identified on the basis of the blocks, and it is defined as

$$\text{b-F-measure} = \frac{2 * (\text{b-Precision}) * (\text{b- Recall})}{(\text{b-Precision}) + (\text{b-Recall})}$$

7. Performance Comparison

Both the algorithms are implemented in Jdk1.5.0 on a Pentium - based Windows platform. The b-Precision and b-Recall for the text features are calculated and also performance is compared with the LH and CE [1].

The b-precision, b-recall and the b-F-measure for five different sites were calculated and are shown in the tables below. The proposed algorithm outperforms LH algorithm in all sites in all categories. This may due that the LH algorithm works only at the feature level.

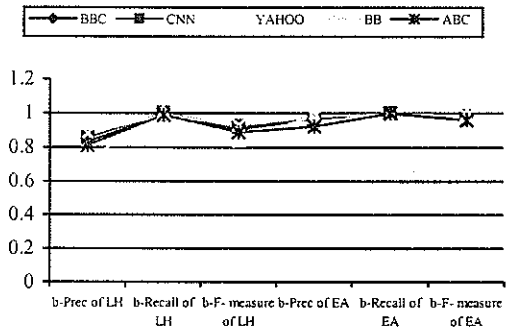


Figure 2. Performance Comparison With LH Algorithm

The proposed algorithm outperforms CE algorithm also in most web sites. On analysis, it is found out that the primary reason for this may be, CE algorithm works only on block level. When compared with both the algorithms, the proposed algorithm performs better, since this algorithm works at both feature and block level. Initially the features were considered to remove the non-content blocks, and then the remaining information is divided into blocks. Also due to string comparison there is no possibility for the redundancy. The performance comparison with LH and CE is given in Fig.2 and Fig.3.

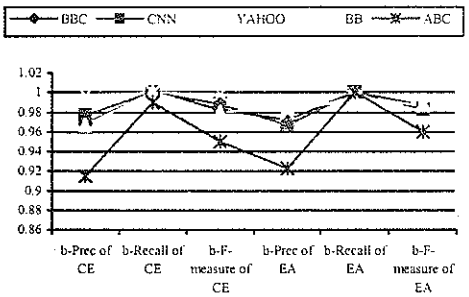


Figure 4. Performance Comparison With CE Algorithm

The snap shot images of the screen displaying the extracted text out put of our system for the given input web page is shown in Fig.4.

8. Related Works

Sandip Debnath, Prasenjit Mitra, Nirmal Pal and Lee [1] proposed an algorithm to identify the primary content of web pages by finding the inverse block document frequency based on the DOM tree structure and the minimum requirement for the content extractor is two web pages. Feature extractor is proposed only for the text features and in the block property changes from the atomic blocks they have defined. Yi and Liu [2] have proposed an algorithm for identifying non-content blocks, referred as "noisy" blocks of Web pages. Their algorithm examines several Web pages from a single Website. If an element of a Web page has the same style across various Web pages, the element is more marked as content block.

The algorithm they proposed by Lin and Ho [3] also tries to partition a Web page into blocks and identify content blocks. They used the entropy of the keywords used in a

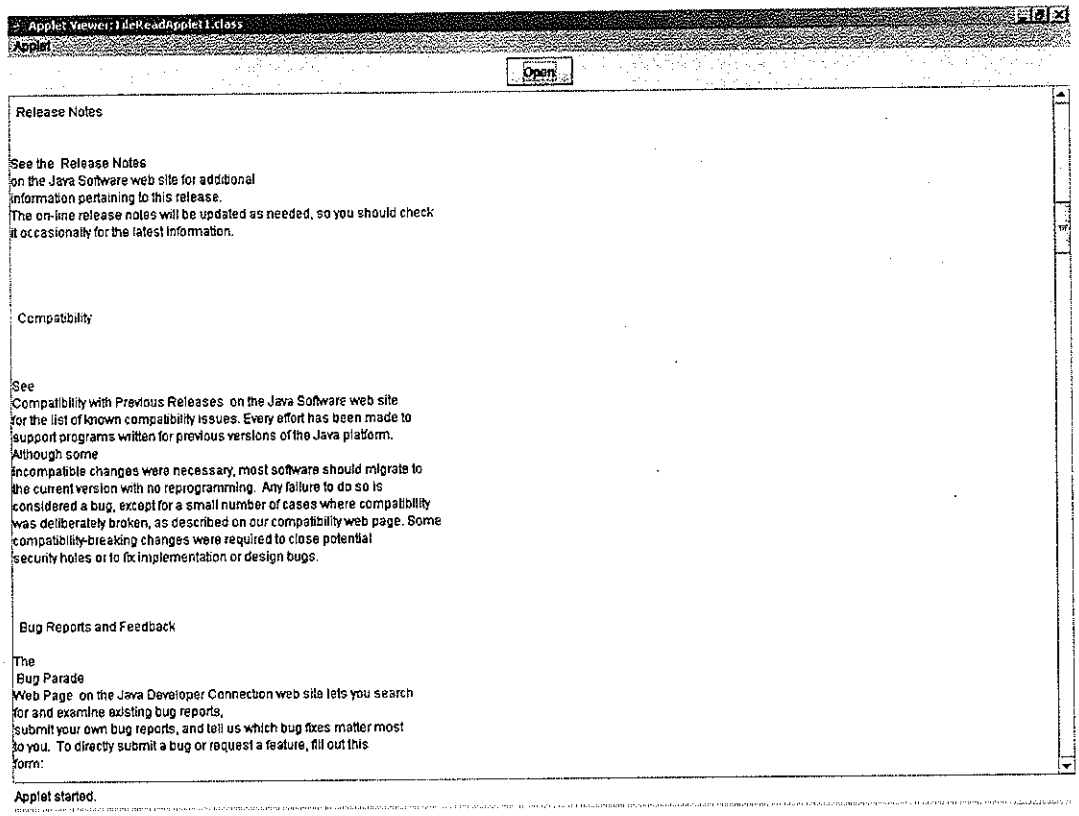


Figure 4. Extracted Text Out Put For The Given Input Web Page

block to determine whether the block is redundant. Cai, has introduced a Vision-based Page Segmentation [7] algorithm. This algorithm segments a Web page based on its visual characteristics, identifying horizontal spaces and vertical spaces delimiting blocks much as a human being would visually identify semantic blocks in a Web page. They use this algorithm to show that better page segmentation and a search algorithm based on semantic content blocks improves the performance of Web searches. Ramaswamy proposed a Shingling algorithm to identify fragments of Web Pages and use it to show that the storage requirements of Web-caching are significantly reduced. Kushmerick has proposed a feature-based method that identifies Internet advertisements in a Web page.

9. Conclusion And Future Work

All the related works in the filed studied are solely geared towards removing advertisements and they do not remove other non-content blocks. The proposed technique is capable of removing advertisements and other specific non-content blocks also. This can further be improved for the retrieval of informative image blocks of a web page. "Content-based" search will analyze the actual contents of the image. The term 'content' in this context might refer colors, shapes, textures, or any other information that can be derived from the image itself. Without the ability to examine image content, searches must rely on metadata such as captions or keywords, which may be laborious or expensive to produce.

References

[1] Sandip Debnath, Prasenjit Mitra, Nimal Pal and C. Lee Giles, *"Automatic Identification of Informative Sections of Web Pages"*, IEEE Transactions on Knowledge and Data Engineering, Vol 17, No. 9, September 2005.

[2] B. Liu, K.Zhao and L. Yi, *"Eliminating Noisy Information in Web Pages for Data Mining"*, Proc. Ninth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, PP 296-305, 2003.

[3] S.H. Lin and J.M. Ho, *"Discovering Informative Content Blocks from Web Documents"*, Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, PP 588-593, 2002.

[4] M. Karthikeyan, Krishnan Nallaperumal, K.Senthamaraikannan, K.Velu and B.Bensujin, *"A Soft Computing Model for Knowledge Mining and Trifle Management"*, International Journal of Imaging Science and Engineering (IJISE), vol. 1, No.4, GA, USA. ISSN:1934-9955, PP 132-138, Dec 2007

[5] Sven Behnke and Nicolaos B. Karayiannis, *"Competitive neural trees for pattern classification"*, in Proc. IEEE Int. Conf. Neural Networks, Washington, D.C., PP 1439-1444, June 1996.

[6] Sivanandan, Shanmugam, and Sumathi, *"Development of Soft Computing Models For Data Miming"*, IE(I) journal Vol 86, PP 22-31, May 2005

[7] D. Cai, S .Yu, J.R. Wen and W.Y. Ma,*"Vision Baseb Page Segmentation"*, Technical Report MSR-TR-2003-79 Microsoft Research Corporation, Nov 1, 2003.

[8] D. Cai, S.Yu, J.R. Wen and W.Y. Ma,*"Block Based Web Search"*, Proc. 27th Ann. Int'l ACM SIGIR Conf., PP 456-463, 2004.

[9] Kleinberg J.M. *"Authoritative sources in a hyperlinked environment"*. In Proceedings of ACM-SIAM Symposium on, Discrete Algorithms, 1998.

[10] Wang Jicheng, Huang Yuan, Wu Gangshan & Zhang *"Fuyan Web mining: knowledge discovery on the Web"*. Systems Man, and Cybernetics, 1999.

Author's Biography

Nallaperumal Krishnan received M.Sc., degree in Mathematics from Madurai Kamaraj University, Madurai, India in 1985, M.Tech degree in Computer and Information Sciences from Cochin University of Science and Technology, Kochi, India in 1988 and Ph.D., degree in Computer Science & Engineering from Manonmaniam Sundaranar University, Tirunelveli. Currently, He is heading Centre for Information Technology and Engineering of Manonmaniam Sundaranar University, Tirunelveli. His research interests include Signal and Image Processing, Remote Sensing, Visual Perception, Mathematical Morphology Fuzzy Logic, Data mining and Pattern recognition. He has authored three books, edited 18 volumes and published 25 scientific papers in Journals. He is a Senior Member of the IEEE.

K. Senthamarai Kannan is currently working as a Professor in Statistics, at the Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu. He has more than 18 years of teaching experience at post-graduate level. He has published more than 30 research papers in international and national journals and authored four books. He has visited Turkey, Singapore and Malaysia. He has been awarded TNSCST Young Scientist Fellowship and SERC Visiting Fellowship. His area of specialization is 'Stochastic

Processes and their Applications'. His other research interests include stochastic modeling in the analysis of birth intervals in human fertility, bio-informatics, data mining and precipitation analysis.

M.Karthikeyan received B.E degree in Electronics and Communication Engineering from Bharathiar University, Coimbatore, India in 1990, M.Tech., degree in Computer and Information Technology from Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India. Currently, he is doing Ph. D., in Computer and Information Technology at Faculty of Engineering, Manonmaniam Sundaranar University Tirunelveli. His research interests include Data mining and Image Processing. He is a Member of the IEEE.

T.Rajesh received the M.Tech., degree in Computer and Information Technology in the year 2007 from CITE, Manonmanium Sundaranar University, Tirunelveli, India. He is currently a Lecturer in Shirdi Sai Engineering College, Bangalore. Currently he is working in the area of data mining and its application to medical informatics. He is a member of ISTE and CSI.

Detection and Removal of Cracks in Digitized Colour Paintings

G.Wiselin Jiji[1], L.Ganesan[2],

ABSTRACT

This paper presents an automated method for detection and removal of cracks in digitized colour paintings. The algorithm starts with the extraction of crack centerlines, which are used as guidelines for the subsequent crack-filling phase. For this phase, the output of four direction differential operators are processed in order to select connected sets of candidate points to be further classified as centerline pixels using crack derived features. The final segmentation is obtained using an iterative region growing method that integrates the contents of several images resulting from crack width dependent morphological filters. The methodology has been shown to perform very well on digitized paintings suffering from cracks.

1. INTRODUCTION

Many paintings, especially old ones, suffer from breaks in the substrate, the paint, or the varnish. These patterns are usually called cracks and are caused by aging, drying, and mechanical factors. Age cracks can result from non-uniform contraction in the canvas or wood-panel support of the painting, which stress the layers of the painting. Drying cracks are usually caused by the evaporation of volatile paint components and the consequent shrinkage of the paint. Finally, mechanical cracks result from painting deformations due to external causes, e.g. vibrations and impacts. The appearance of cracks on paintings deteriorates the perceived image quality. However, one can use digital image processing techniques to detect and eliminate the cracks on digitized paintings. Such a "virtual" restoration can provide clues to art historians, museum curators and the general public on how the painting would have looked like in its initial state, i.e., without the cracks. Furthermore, it can be used as a non-destructive tool for the planning of the actual restoration. A system that is capable of tracking and interpolating cracks is presented in [1]. The user should manually select a point on each crack to be restored. A method for the detection of cracks using multi-oriented Gabor filters is presented in [2]. Different approaches for interpolating information in structured [3], [4], [5], [6], [7] have been developed. A technique that decomposes the image to textured and structured areas and uses appropriate interpolation techniques depending on the area where the missing information lies has also been proposed [8]. The results obtained by these techniques are very good. A methodology for the restoration of cracks on digitized paintings, which adapts and integrates a number of image processing and analysis tools, is proposed in this paper. The methodology is an extension of the crack removal framework presented in [9]. The technique consists of the following stages:

- Crack detection.
- Separation of the thin dark brush strokes, which have been misidentified as cracks.
- Crack filling (interpolation).

[1]Department of Computer Science & Engineering, Sivanthi Aditanar College of Engineering Tiruchendur. e-mail-id:jijivevin@yahoo.co.in

[2]Department of Computer Science & Engineering, A.C.College of Engineering & Technology, Karaikudi.

A certain degree of user interaction, most notably in the crack detection stage, is required for optimal results. However, all processing steps can be executed in real time and thus the user can instantly observe the effect of parameter tuning on the image under study and select in an intuitive way the values that achieve the optimal visual result. Needless to say that only subjective optimality criterion can be used in this case since no ground truth data are available. The opinion of restoration experts that inspected the virtually restored images was very positive.

2. Overview Of Proposed Method

The method herein presented can be schematically described by the functional block diagram in Figure 1, where we identify three main processing phases: 1) preprocessing, for background normalization and thin crack enhancement; 2) crack centerline detection, for defining a set of connected segments in the central part of the cracks; and 3) crack segmentation, for finally labeling the pixels belonging to the crack. These phases are further subdivided in several steps, as follows:

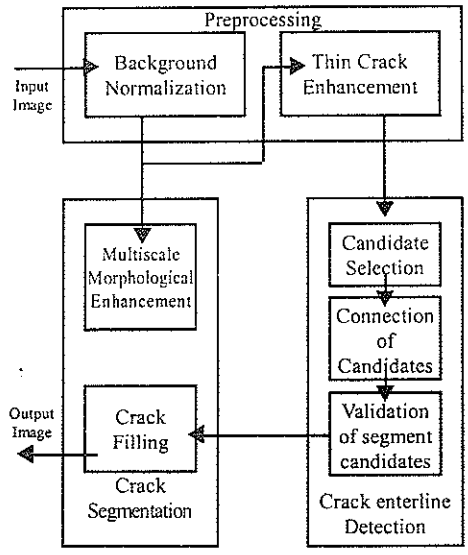


Figure 1 : Crack Detection & Removal Functional Diagram

Preprocessing Phase

1) *Background normalization,* which normalizes the input image by subtracting an estimate of the background obtained by filtering with a arithmetic mean kernel. 2) *Thin crack enhancement* by processing with a set of line detection filters, corresponding to the six orientations 0, 30, 60, 90,120 and 150; for each pixel, the highest filter response is kept and added to the normalized image.

Crack Centerline Detection Phase:

- Selection of Crack centerline candidates, using directional information provided from a set of six directional Difference of Offset Gaussians filters.
- Connection of the candidate points obtained in the previous step, by a region growing process guided by some image statistics.
- Validation of centerline segment candidates based on the characteristics of the line segments; this operation is applied in each one of the six directions and finally combined, resulting in the map of the detected crack centerlines.

Crack Filling Phase:

- Crack filling by a region growing process using as initial seeds the pixels within the centerlines obtained in the crack centerline detection phase; the growing is successively applied to the four scales and, in each growing region step, the seed image is the result of the previous aggregation. Each one of these phases is detailed and illustrated in the next sections.

3. Detection Of Crack Centerlines

The green channel is considered in our work as the natural basis for crack segmentation because it normally presents a higher contrast between cracks and

background. Crack centerlines are herein defined as connected sets of pixels which correspond to intensity maxima computed from the intensity profiles of the crack cross sections. Geometric Property derives from local intensity properties, the crack cross profile-taking advantage of the fact that the maximum local intensity usually occurs at the crack central pixels. To locate candidate pixels belonging to the central part of a crack, the original methodology developed by the authors for the detection of perifoveolar capillaries [10] was further adapted and extended. The underlying idea of this approach is that, the response of directional differential operators, using kernels adapted to the local crack direction, has opposite signs on the two hillsides of an ideal crack cross profile; we will, therefore, explore this fact by considering the occurrence of specific combinations of filter response signs. To carry out the initial selection of the most likely centerline segments, the magnitude of the filter response is kept on the positions that verify one of the established sign conditions; this newly generated image is then segmented using region growing in order to retain just those points where restrictive intensity and connectivity conditions meet. For each one of these segments, we compute the mean intensity, the maximum intensity, and the length of the pixel set; these features are further combined for final validation of the segments belonging to crack centerlines. The detailed processing operations involved are described in the following subsections.

A. Preprocessing Phase

Median filter is used to normalize the input image. As small cracks are very thin structures and usually present low local contrast, their segmentation is a difficult task. To improve the discrimination between these thin cracks and the background noise, the normalized image is processed with a set of line detection filters [11], corresponding to the four orientations 0, 45, 90, and 135. The set of convolution kernels used in this operation is shown in Figure 2. For each pixel, the highest filter response is kept and added to the normalized image. This resulting image is the base of all subsequent operations used for the detection of crack centerlines.

$$\frac{1}{6}\begin{pmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{pmatrix}; \frac{1}{6}\begin{pmatrix} -1 & -1 & 2 \\ -1 & 2 & -1 \\ 2 & -1 & -1 \end{pmatrix};$$

$$\frac{1}{6}\begin{pmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{pmatrix}; \frac{1}{6}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$$

Figure : 2 Line Detector Filters Used For Thin Crack Enhancement

B. Detection Of Centerline Segment Candidates

i) Initial Selection of Candidate Points:

The first operation is to extract crack centerline pixels. When a first-order derivative filter is applied orthogonally to the main orientation of the crack, derivative values with opposite signs are obtained on the two crack hillsides, which simply mean that there will be positive values on one side of the crack cross section and negative values on the other.

Cracks can occur in any direction, the selection of a set of directional filters whose responses can be combined to cover the whole range of possible orientations is required. We tested two (0^0 and 90^0), four (0^0, 45^0, 90^0, and 135^0), and six (0^0, 30^0, 60^0, 90^0, 120^0 and 150^0)directions of same mask and concluded that the solution based on four directions is an interesting trade-off between detection accuracy and computation time. The particular kernels used in this work are first-

order derivative filters, known as difference of offset Gaussians filters (DoOG filters), with prevailing responses to horizontal (0°), vertical (90°), and diagonal (45°, 135°) directions. The DoOG filters have De monstrated greater immunity to noise because they depend on larger kernel derivative filters.

$$\begin{pmatrix} -1 & -2 & 0 & 2 & 1 \\ -2 & -4 & 0 & 4 & 2 \\ -1 & -2 & 0 & 2 & 1 \end{pmatrix}$$

Figure : 3 Kernel Of Doog Filter Used In My Work

The particular kernel used for detecting vertical centerline candidates is the row gradient filter presented in Fig 3; the other three kernels are just rotated versions of this filter. The centerline candidates that are retained after the analysis of image rows .The intensity of each pixel is representative of the filter response. Finally, the adequacy of our methodology to locate points in the central part of the cracks can be confirmed. The kernels used for processing the images, and the distinctive directions that are searched for the occurrence of derivative sign combinations.

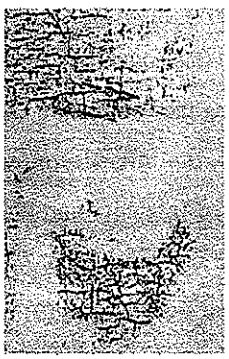
ii) Connection of Candidate Points

Each one of the four images resulting from the sequence of operations described before is processed independently in order to produce a set of four connected crack segments with a common main orientation. From each image containing the selected set of candidate points in one specific direction, an initial collection of centerline segments is generated by a region growing process that is started with a set of seed points, verifying restricted value conditions, which are afterwards extended by aggregating other neighboring pixels with lower filter responses. Both seed and aggregation thresholds are defined based on statistics derived from the histogram of the image containing the set of candidate points; the values of the seeds are above a limit depending on the mean and standard deviation of this distribution, while aggregation threshold is the histogram mode. The threshold value, T_{seed}, for seed selection is evaluated by the function in equation (1).

$$T_{seed} = \mu - \alpha\sigma \qquad (1)$$

In this equation, the value of the coefficient is equal for all the images of the database, while and , respectively, the mean and standard deviation of the distribution are dependent on the properties of each particular image. A correct choice of this threshold value is critical for the elimination of noisy segments, usually found in the background.The result of this processing sequence is a set of connected segments,as Shown in the Fig 4(b).

(a)Input Image



(B) Detected Crack Centerlines

Figure 4.crack Centerline Segments