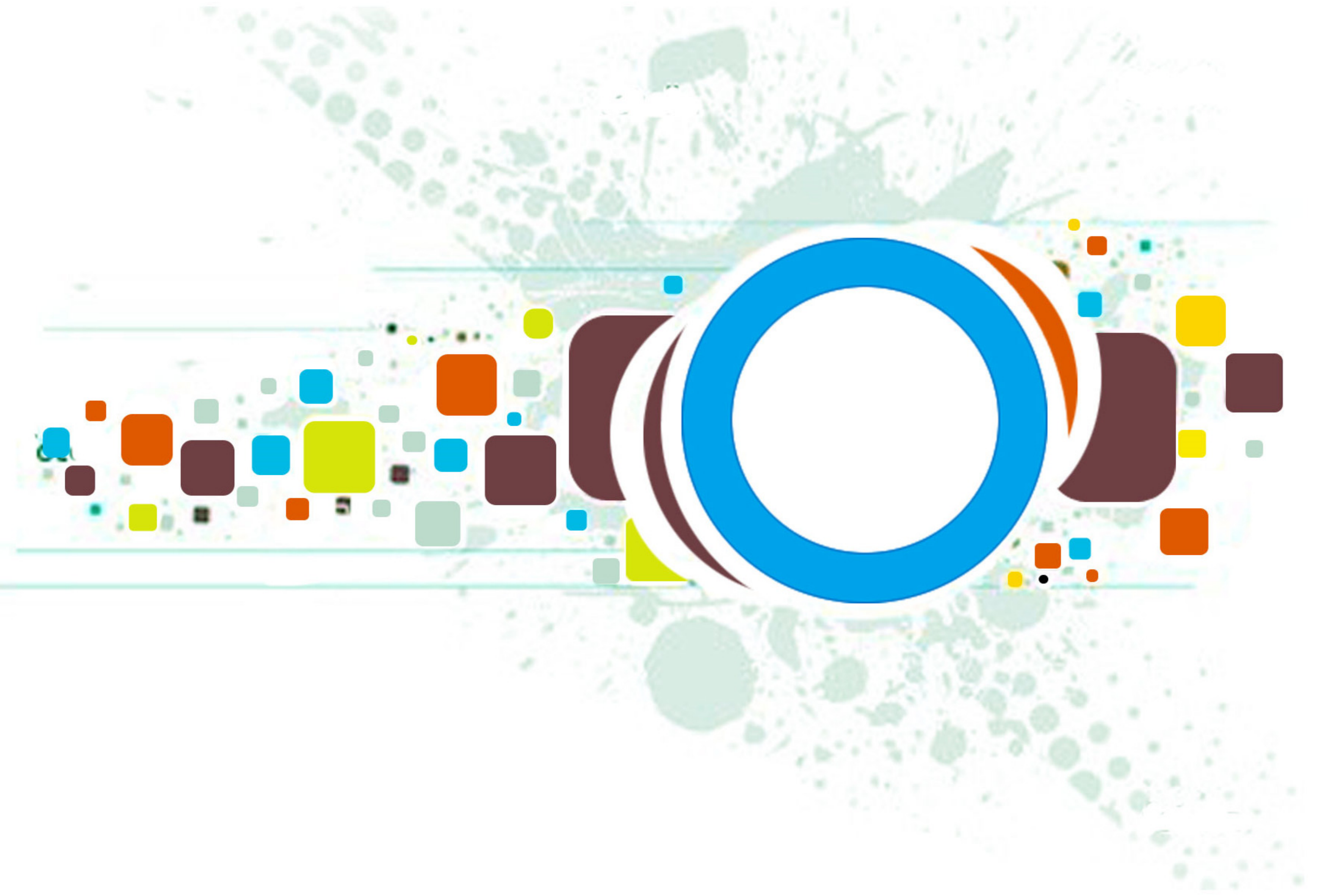
Volume 9 • Issue 3 • May / June 2015

Editor-in-Chief
Professor Hu, Yu-Chen

INTERNATIONAL JOURNAL OF IMAGE PROCESSING (IJIP)

ISSN : 1985-2304
Publication Frequency: 6 Issues Per Year



CSC PUBLISHERS
http://www.cscjournals.org

INTERNATIONAL JOURNAL OF IMAGE PROCESSING (IJIP)

VOLUME 9, ISSUE 3, 2015

EDITED BY
DR. NABEEL TAHIR

ISSN (Online): 1985-2304

International Journal of Image Processing (IJIP) is published both in traditional paper form and in Internet. This journal is published at the website http://www.cscjournals.org, maintained by Computer Science Journals (CSC Journals), Malaysia.

IJIP Journal is a part of CSC Publishers
Computer Science Journals
http://www.cscjournals.org

INTERNATIONAL JOURNAL OF IMAGE PROCESSING (IJIP)

Book: Volume 9, Issue 3, May / June 2015
Publishing Date: 30-06-2015
ISSN (Online): 1985-2304

IJIP Journal is a part of CSC Publishers
http://www.cscjournals.org

Published in Malaysia

Typesetting: Camera-ready by author, data conversation by CSC Publishing Services – CSC Journals, Malaysia

CSC Publishers, 2015

EDITORIAL PREFACE

The International Journal of Image Processing (IJIP) is an effective medium for interchange of high quality theoretical and applied research in the Image Processing domain from theoretical research to application development. This is the *Third* Issue of Volume *Nine* of IJIP. The Journal is published bi-monthly, with papers being peer reviewed to high international standards. IJIP emphasizes on efficient and effective image technologies, and provides a central for a deeper understanding in the discipline by encouraging the quantitative comparison and performance evaluation of the emerging components of image processing. IJIP comprehensively cover the system, processing and application aspects of image processing. Some of the important topics are architecture of imaging and vision systems, chemical and spectral sensitization, coding and transmission, generation and display, image processing: coding analysis and recognition, photopolymers, visual inspection etc.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Started with Volume 9, 2015, IJIP appears with more focused issues. Besides normal publications, IJIP intends to organize special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

IJIP gives an opportunity to scientists, researchers, engineers and vendors from different disciplines of image processing to share the ideas, identify problems, investigate relevant issues, share common interests, explore new approaches, and initiate possible collaborative research and system development. This journal is helpful for the researchers and R&D engineers, scientists all those persons who are involve in image processing in any shape.

Highly professional scholars give their efforts, valuable time, expertise and motivation to IJIP as Editorial board members. All submissions are evaluated by the International Editorial Board. The International Editorial Board ensures that significant developments in image processing from around the world are reflected in the IJIP publications.

IJIP editors understand that how much it is important for authors and researchers to have their work published with a minimum delay after submission of their papers. They also strongly believe that the direct communication between the editors and authors are important for the welfare, quality and wellbeing of the Journal and its readers. Therefore, all activities from paper submission to paper publication are controlled through electronic systems that include electronic submission, editorial panel and review system that ensures rapid decision with least delays in the publication processes.

To build its international reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for IJIP. We would like to remind you that the success of our journal depends directly on the number of quality articles submitted for review. Accordingly, we would like to request your participation by submitting quality manuscripts for review and encouraging your colleagues to submit quality manuscripts for review. One of the great benefits we can provide to our prospective authors is the mentoring nature of our review process. IJIP provides authors with high quality, helpful reviews that are shaped to assist authors in improving their manuscripts.

Editorial Board Members
International Journal of Image Processing (IJIP)

TABLE OF CONTENTS

Volume 9, Issue 3, May / June 2015

Pages

Stereo Correspondence Algorithms for Robotic Applications Under Ideal And Non Ideal Lighting Conditions

Deepambika V A *deepambika.va@gmail.com*
Asst Professor, ECE
LBS Inst of Technology for Women
Trivandrum, Kerala, INDIA

Abdul Rahiman M *rehman_paika@yahoo.com*
Pro Vice Chancellor
Kerala Technological University
Trivandrum, Kerala, INDIA

Abstract

The use of visual information in real time applications such as in robotic pick, navigation, obstacle avoidance etc. has been widely used in many sectors for enabling them to interact with its environment. Robotics require computationally simpler and easy to implement stereo vision algorithms that will provide reliable and accurate results under real time constraint. Stereo vision is a less expensive, passive sensing technique, for inferring the three dimensional position of objects from two or more simultaneous views of a scene and there is no interference with other sensing devices if multiple robots are present in the same environment. Stereo correspondence aims at finding matching points in the stereo image pair based on Lambertian criteria to obtain disparity. The correspondence algorithm will provide high resolution disparity maps of the scene by comparing two views of the scene under the study. By using the principle of triangulation and with the help of camera parameters, depth information can be extracted from this disparity .Since the focus is on real-time application, only the local stereo correspondence algorithms are considered. A comparative study based on error and computational costs are done between two area based algorithms. Evaluation of Sum of absolute Difference algorithm, which is less computationally expensive, suitable for ideal lightening condition and a more accurate adaptive binary support window algorithm that can handle of non-ideal lighting conditions are taken for this study. To simplify the correspondence search, rectified stereo image pairs are used as inputs.

Keywords: Stereo Vision Robotic, Correspondence Algorithm, SAD.

1. INTRODUCTION

Computer vision tries to copy the way in which the human beings perceive visual information. By means of pin hole cameras acting as eyes, computers process the information in an ‘intelligent way’ as does the human brain .The three dimensional information about the environment is essential for movement and object detection in robotic applications. One method of obtaining depth information is by using binocular stereo vision, for which two pinhole cameras are sufficient. The Greek word ‘stereos’ literally means firm or rigid. Stereo vision is a passive sensing technique which provides high resolution depth map- that is objects that are seen as 3D. In stereo vision, there is no interference with other sensor devices when multiple robots are present in the same environment. Stereo vision infer depth from two 2D images called left and right images taken from different viewpoints of a scene. Figure.1 shows an overview of stereo vision system.

Stereo correspondence aims to find matching pixels (conjugate pair) of two given input images based on Lambertian surface assumption, i.e., to find out each point in left image and the corresponding best matching point in the right image. Based on accuracy and efficiency, the correspondence algorithms can be grouped two main categories-local and global[17]. Global methods are accurate but time and computational demanding due to their iterative nature. Accuracy becomes important in applications such as precise 3D surface modeling, especially

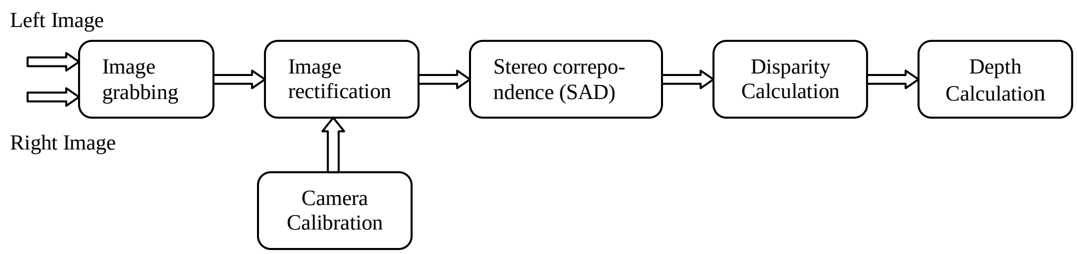


FIGURE 1: Overview of Stereo Vision System.

when dealing with object surfaces with complex reflectance behavior and poor texture. Efficiency is the main concern when the stereo system is employed in real-time applications such as robot navigation, video surveillance, and civil protection. Performance of stereo algorithms depends on the cost aggregation function used for the dissimilarity or the similarity measure. Though the feature based algorithms are faster they give only sparse disparity maps. The need of computationally less expensive stereo algorithm in combination with accuracy in real time operations are in more demanded.Also for various outdoor robotic applications, the lighting conditions are far from ideal conditions. In such conditions typical local algorithms may not give accurate results. A comparative study based on error and computational costs are done between two area based algorithms.Implementation and evaluation of Sum of absolute Difference algorithm, which is less computationally expensive ,easy to implement and suitable for ideal lightening condition and a more accurate adaptive binary support window algorithm that can handle images taken under non-ideal lighting conditions are taken for this study.

2. SUM OF ABSOLUTE DIFFERENCES (SAD) ALGORITHM

Sum of Absolute Differences (SAD) is a wide spread simple block matching algorithm. It is faster and easier for hardware implementation as it only needs a subtraction, a comparison, and a possible sign change. This area based algorithm assumes that the image window centered at a pixel have similar grey level.The correspondence problem is done by comparing blocks of pixels in each image. The dissimilarity is the sum of differences in pixel intensities, therefore the name SAD.

Computing the sum of differences between the pixel intensity values for a block around the pixel(x, y) is simple. Let f(i,j)be the intensity of the pixel at coordinates (i, j) in the reference image. For the second image ,intensity g(i, j) is used, N defines the extend of the block in either direction around the center pixel, the block has therefore width and height (2N+1). Here the rectified images are used as inputs. Therefore the 2-D stereo correspondence problem is reduced to a 1-D problem, means that the search is done along horizontal lines only as there is no vertical shift for the same pixels in the left and right images if the images are rectified[1].The variable'd' is the disparity-the displacement of corresponding matching points from one image to the other.

$$SAD(x,y,d) = \sum_{i=x-N}^{x+N} \sum_{j=y-N}^{y+N} |f_{i,j} - g_{i+d,j}| \qquad Eqn(1)$$

Each pixel in the left image is compared with every pixel on same epipolar line in right image.SAD values for all pixels are computed and the disparity is computed using winner-take-all (WTA) principle. The winner is simply the disparity associated with the smallest SAD value, in ideal case smallest SAD value will be zero.

$$disparity(x,y) = \min_{d \in D} \left(sad(x + d, y)\right) \qquad Eqn(2)$$

The Eqn (2) describes the disparity for the block around the pixel(x, y),which varies from zero to dmax, where dmax represents the highest disparity value of the stereoscopic images called disparity range. Each block from the left image is matched to a block in the right image by shifting the left block over the search area of right image [2]. Difference in index values for the SAD minimum corresponds to disparity for that area. This is continued for each block until the disparity map is completely filled and the resulting image is known as Disparity Space Image (DSI).

2.1 Performance Improvements

By applying a simple sliding window scheme, the algorithm will speed up. This works in the horizontal as well as in the vertical direction.

Horizontal sliding window:As shown in Fig.2 a window is moved from one column to next column ie., from left to the right across a horizontal scan line. The search rows are still the same, except for the pixel in the left and left and right.

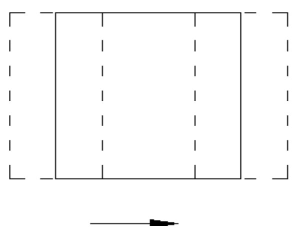


FIGURE 2: Horizontal Sliding Window: window slides across horizontal scan line.

Vertical sliding window: The sliding window scheme is also being applied when moving to the next horizontal line as shown in Fig.3 After completing a pass on one horizontal line, the window moves to the line below. The search columns are still the same, except for the pixel above and below.
The DSI obtained will be dense disparity map.

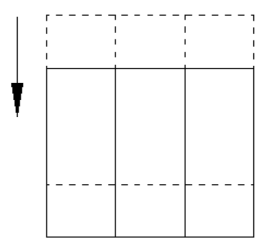


FIGURE 3: Vertical Sliding Window- window slides across vertical scan line.

2.2 Depth Computation

From the disparity values obtained, depth (Z) can be estimated by using triangulation and camera parameters [4]. The depth Z is given by Eqn.[3]

$$Z = \frac{f.b}{x_l - x_r} = \frac{f.b}{d} \qquad Eqn(3)$$

Where d is the disparity, f the focal length of the camera and b is the baseline distance between the centres of cameras.

2.3 Steps of SAD Algorithm

- Read the rectified input images.
- Select a square window from left image
- Slide this window over right image and compute SAD minimum
- Obtain the difference in index values corresponding to the SAD minimum
- Aggregate this index values (disparity) for the entire image
- Obtain depth value using triangulation.

3. ADAPTIVE BINARY SUPPORT WINDOW APPROACH

For real-time applications local and window based algorithms are suitable. They will produce decent depth maps. The size of the matching window strongly influences the performance of area-based algorithms. For high textured regions and object boundaries, small window size is good, but it fails in low textured image regions. Similarly, a large window size gives good results at low textured regions but it blurs the disparity edges. Therefore the selection of the ideal window size is important. The window size should be large enough to contain distinguished features, but small enough to keep depth discontinuities. When the local structures of image pixels are similar, it will be difficult to find their correspondences in other images without global matching. But global correspondence methods are computationally expensive and are not suitable for real-time applications. When fixed window size is used for stereo correspondence it is implicitly assumed that all the pixels with in this window are of same depth. But this is not true for object edges where depth discontinuity occurs.

Window-based stereo correspondence algorithms often exhibit a ‘foreground fattening effect’ near depth discontinuities between two objects. When this happens, samples from the far object are mistakenly measured as having the same disparity as samples on the near object. By using the local support windows, the depth ambiguity can be reduced efficiently and the discriminative power of the similarity measure is increased.

Kanade and Okutomi [5] proposed an iterative stereo matching algorithm which selects window adaptively to minimise the uncertainty in the disparity estimation. This method starts with an initial estimation of the disparity map and updates it iteratively. Hence the method is sensitive to the initial disparity estimation. Also the shape of support window is limited to a rectangle. Though the window can be expanded in chosen directions it is not suitable for pixels near arbitrarily shaped objects at depth discontinuities. This method is computationally expensive and not suitable for real time applications.

Fusiello et al. [6] presented normalised SSD (Sum of Squared Differences) algorithm with multiple window scheme using left-right consistency to compute disparity. The correlation is done with nine different windows and the disparity profile selects proper window. But this limited number of windows is not enough to cover entire range of windows of required size and shape.

Veksler [7] optimized window size by using minimum ratio cycle algorithm. This approach is the first window based algorithm using non rectangular window. Even though this correspondence algorithm performs well, it uses many user specified parameters for cost computation and is not suitable for real time systems.

Correspondence methods [8], [9] used appropriate support-weights to the pixels in a support window. But the shape and size of a local support window are fixed. Also they do not have general criteria for choosing the shape of the support window for these approaches.

The adaptive-weight algorithm proposed by Yoon [10] computes the support weight of each pixel inside a fixed sized square window. The support weights of pixels in a window are based on color similarity and geometric proximity to reference pixel. Their experiment indicates that a local based stereo matching algorithm can produce good depth maps similar to global algorithms. Though this method gives accurate results at depth discontinuities and in homogeneous regions, it is computationally expensive and is also prone to image noise.

Tombari [11] extends the adaptive-weight algorithm of Yoon's concept of colour proximity as well as information from segmentation .Though it is an accurate local stereo correspondence the computational complexity remains the same.

Leonardo De-Maeztu[12] presents a stereo correspondence using gradient similarity and locally adaptive support-weight. Here the matching measure relies on the gradient fields of neighbourhood of a pixel than the intensity of the pixel. Even though this approach improves on the accuracy of previous adaptive support-weight algorithms it needs extra computational stages by the addition of the gradient computation.

Lazaros Nalpantidis[13] proposes an algorithm for robotic application incorporates biologically and psychologically inspired features to an adaptive weighted sum of absolute differences (SAD) for stereo matching to determine the accurate depth of a scene. Though the algorithm exhibits good behaviour, its needs more computational time and it uses two user defined parameters.

Lazaros Nalpantidis & Antonios Gasteratos [14] proposes a luminosity-compensated dissimilarity measure (LCDM). Though this method process contrast differentiations it does not exhibit the same behaviour for luminosity differentiations. This stereo algorithm which is based on the proposed LCDM measure is not fast enough to achieve the real- or near real-time frame rates demanded by robotic applications.

This work presents simple stereo correspondence algorithm using binary adaptable window as support window which will meet the real time requirements. For each fixed size window a binary mask window is generated that can take any shape within the fixed window. The mask window selects the supporting pixels in the cost aggregation phase of the SAD algorithm. This selection is performed using color similarity and spatial distance metrics. The proposed method is composed of three parts: adaptive support window computation, dissimilarity computation based on the support-window and disparity selection.

3.1 Support Weight Aggregation In The Human Visual System

When aggregating support to measure the similarity between image pixels, the support from a neighboring pixel is valid only when the neighboring pixel is from the same depth. Visual grouping is very important to form a support window. There are many visual cues used for perceptual grouping. Among them, similarity and proximity are the two main grouping concepts in classic gestalt theory. The gestalt rule of organization based on similarity (or smoothness) and proximity is one of the most important ones and has been widely used in vision research. Gestalt is a German word which can be translated as 'form, shape, or pattern' in English. The similarity principle states that people tries to group similar elements together based on features such as colour or shape. The principle of proximity states that elements that are located close to each other will be grouped together [15]. The difference between pixel colors is measured in the CIE Lab color space because it provides three-dimensional representation for the perception of color stimuli.

3.2 CIELAB

The most complete color space specified by the International Commission on Illumination is CIE L*a*b* (CIELAB). It describes all the visible colors of the human eye and was created as a device-independent model for reference. The three coordinates of CIELAB represent the lightness of the colour (L* = 0 indicates black and L* = 100 indicates white), its position between red/magenta and green .The negative values a* indicate green while positive values indicate magenta. Its position is between yellow and blue. Negative values, b* indicate blue and positive values indicate yellow. The asterisk (*) after L, a and b are part of the full name, since they represent L*, a* and b*, to distinguish them from Hunter's L, a, and b. The L*a*b* model is a three-dimensional model, and hence it can be represented properly in a three-dimensional space only.

The nonlinear relations for L*, a*, and b* are intended to imitate the nonlinear response of the eye. The uniform changes of components in the L*a*b* color space aim to correspond to uniform changes in perceived color. The relative perceptual differences between any two colors in L*a*b* can be approximated by treating each color as a point in a three-dimensional space (with three components: L*, a*, b*) and taking the Euclidean distance between them.

3.3 Adaptive Binary Support Window

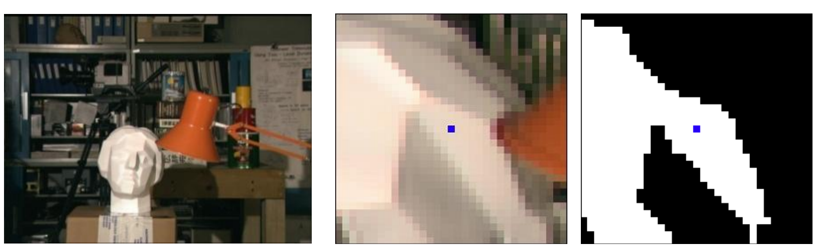
To construct the support window, with in a fixed size window we took the color similarity of the pixels in a small region of the reference image. The pixels having small Euclidean distance are grouped together to form the windows of different sizes and shapes.
The distance function can be expressed as:

$$d_{pq} = \sqrt{(Lp - Lq)^2 + \left(a_p - a_q\right)^2 + \left(b_p - b_q\right)^2} \qquad \mathrm{Eqn}(3)$$

The values of L,a,b are calculated from RGB values in the CIE Lab color space. Matching window for a pixel is obtained by placing the centre of support window on that pixel and the distance d_{pq} between the center pixel and other pixels in support window is computed. Pixels having smaller distance, which is limited to a threshold 'T' are taken for the matching process. The binary mask window is generated as follows. The support weights are chosen binary, where '0' means that this pixel will give no support to the matching window and '1'means that this pixel gives support to the matching window. There is no gradient between these two extremes.
The intensity of binary mask,

$$W = \begin{cases} 1 & \text{for } d_{pq} < T \\ 0 & \text{otherwise} \end{cases} \qquad Eqn(4)$$

where T is a threshold value. This is taken as a masking window that selects the pixels within the fixed window that will be used in the cost aggregation phase of the SAD algorithm. Only the pixel positions where the masking window is '1' (white) will be taken into account. These are the active matching regions of support window. Figure 4(b) shows such binary mask window that has been computed for a pixel of the Tsukuba image shown in Fig. 4(a). The center of the window has been highlighted using the blue color. Thus we can see that, this method constructs adaptive windows of different shapes and sizes for each pixel based on their local information.



(a) (b) (c)

FIGURE 4: An example of adaptive binary support window computation and matching. (a) Tsukuba [16]. (b) Shows the support region of the centre pixel in left image for which the adaptive binary window is to be computed. (c) Shows an adaptive binary window calculated for the centre pixel of (b).

The cost computation for dissimilarity measure and disparity aggregation are same as that of SAD algorithm as described above.

3.4 Steps of Adaptive binary Support window algorithm.

1. Read the input images.
2. Convert the colour space in RGB to CIE Lab.
3. Compute Euclidian distance from this CIE Lab.
4. Limit the Euclidian distance to a threshold value
5. Compute binary mask window based on threshold value.
6. Select a square window from left image.
7. Multiply the two windows.
8. Slide this window over right image and compute SAD minimum.
9. Obtain the difference in index values corresponding to the minimum value.
10. Aggregate this index values (disparity) for the entire image
11. Obtain depth value using triangulation.

4. RESULTS AND ANALYSIS

The two algorithms are verified using the rectified stereo image pairs from Middlebury stereo data set [16]. Each dataset of the database consists of a pair of left and right stereo images and the corresponding ground truth disparity map. The experiments are done on Matlab Platform.

4.1 Results of SAD Algorithm

For the study and evaluation of our SAD algorithms Aloe, Baby and Bowl image pairs and their ground truth data are used.Fig.5 (a) shows left view of input images ,Fig.5(b)its ground truth disparity. Figure 5(c) & (d) shows the disparity map and the corresponding depth maps obtained using fixed window size of 9x9 pixels in SAD algorithm. Here the disparity values of SAD algorithm are mapped as disparity map and the corresponding depth values are mapped as depth map.

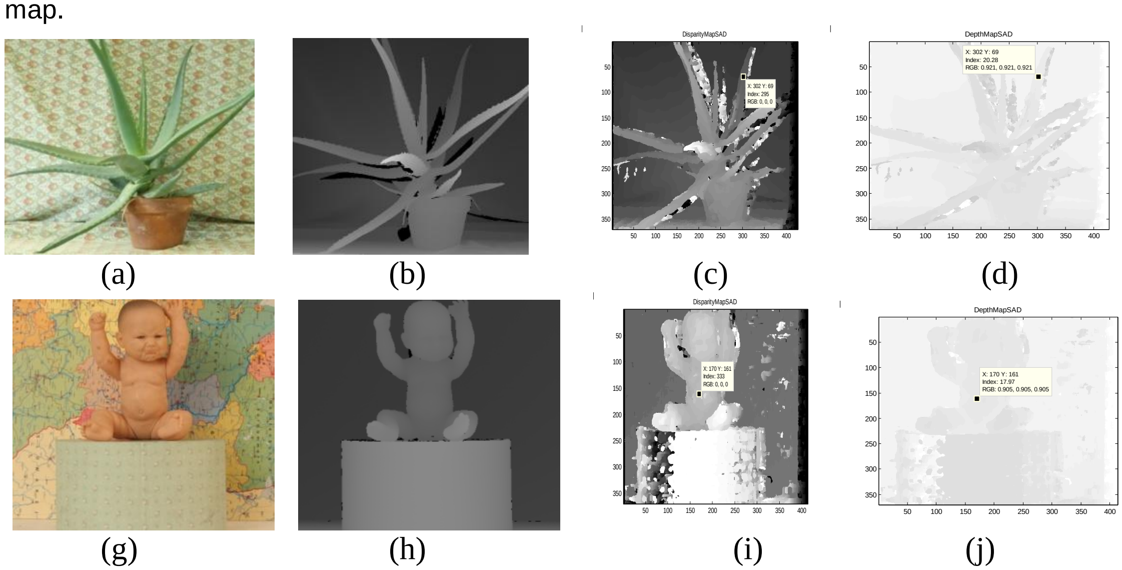
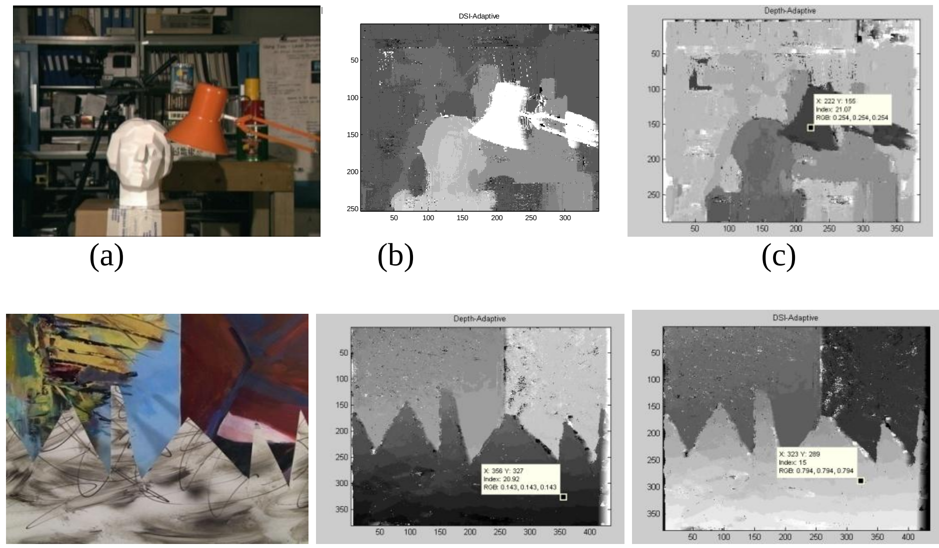


FIGURE 5: Dense disparity map &depth maps for Middlebury images and their corresponding Ground truth disparity: (a)Input image :Aloe-Left view [16] (g)Input image :Baby-Left view[16] (b) &(h) Ground truth disparities[16]. (c)& (i) Disparity map from SAD (d)&(j) corresponding depth map obtained .

4.2 Results of Adaptive Binary Support Window Algorithm

The Figure 6 (a) shows input images Tsukuba(384 x 288 image with disparity range from 0 to 15) and sawtooth (434 x 380 with disparity range from 0 to 19), These are PPM (Portable Pixel Map) images taken under bad illumination condition-texture less and scaled to a factor of 8.Fig.6 (b)&(f) are the Disparity map obtained from this algorithm. Fig.6 (c)&(g) are their corresponding depth map. The performance of this algorithm is much better since it can preserve arbitrarily shaped depth discontinuities and gives accurate results in homogeneous regions.



(a) (b) (c)

(d) (e) (f)

FIGURE 6: Results of Adaptive binary Support window algorithm. (a) & (d) input images- Tsukuba & Sawtooth, [16] (b)&(e)disparity map obtained and (c)&(f) Corresponding depth map.

4.3 Evaluation of Results

To evaluate the performance of a stereo algorithm we need a quantitative way to estimate the quality of the computed correspondences [17]. The error statistics is computed with respect to ground truth data.

1. RMS (root-mean-squared) error (measured in disparity units) between the computed depth map $d_C(x, y)$ and the ground truth map $d_T(x, y)$, i.e,

$$E = \left(\frac{1}{N} \sum_{(x,y)} |d_C(x,y) - d_T(x,y)|^2 \right)^{\frac{1}{2}} \qquad Eqn(5)$$

where N is the total number of pixels.

4.4 Evaluation of SAD Fixed Window

The algorithm shows an error of 1.4% for image Baby with window size 13x13. Fig. [7] shows Window size vs % Error for PNG images, Aloe, Baby and Bowl.

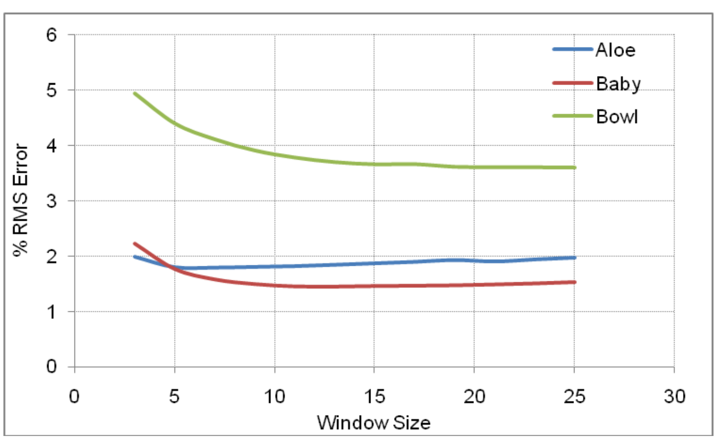


FIGURE 7: Evaluation of SAD Algorithm: Window size Vs %RMS Error for PNG images.

As the window size increases further, error increases due to edge fattening effect at depth discontinuities. For this case we go to adaptive binary approach.

4.5 Evaluation of Adaptive Binary Support Window Algorithm

The Adaptive binary support window algorithm is evaluated using the PPM images Tsukuba, Venus,and Sawtooth. Figure 8 shows the graphical representation of Window size vs % Error of our approach.

This algorithm shows an error of 0.7382% only for Tsukuba image (taken under extreme bad illumination condition) for window size 23x23.For PNG image inputs algorithm shows an error of 1.29% for Image Baby with window size 9x9.

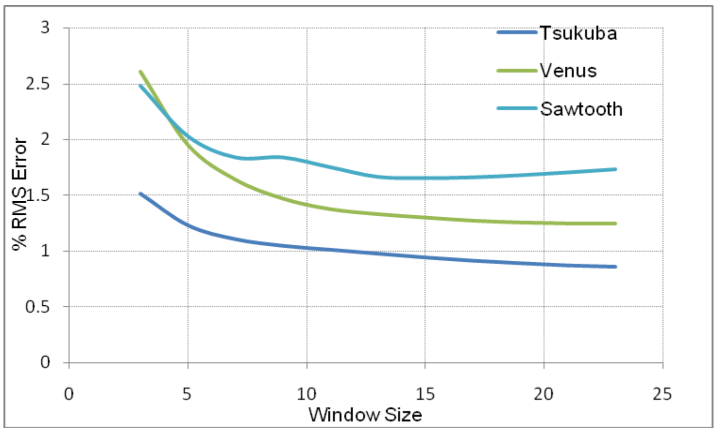


FIGURE 8: Adaptive binary support window algorithm Error evaluation. Window size vs %RMS Error for PPM images.

5. COMPARISON BETWEEN SADFIXED WINDOW AND ADAPTIVE BINARY SUPPORT WINDOW ALGORITHM

The implemented stereo algorithms focused on its real time application aspect, means that the algorithms are aimed in accuracy and computational simplicity. These two algorithms are simple and can be easily realized, if we go for a hardware implementation.

5.1 Observations in SAD Fixed Window

The algorithm is evaluated using PNG (taken under good illumination condition- good texture) and PPM (taken under bad illumination-texture less) images. Fig [9] shows the graphical representation of Window size vs % Error for PNG and PPM images. The algorithm shows an error of 1.44% for image Baby with window size 13x13.But with input image Tsukuba, error become12.7%,since this image is taken under extreme bad condition .

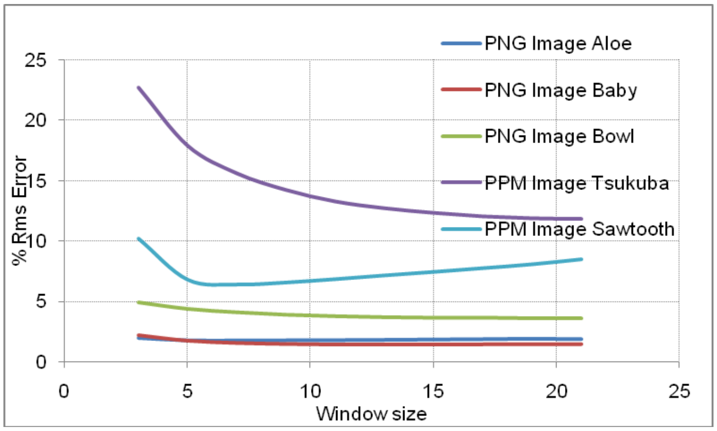
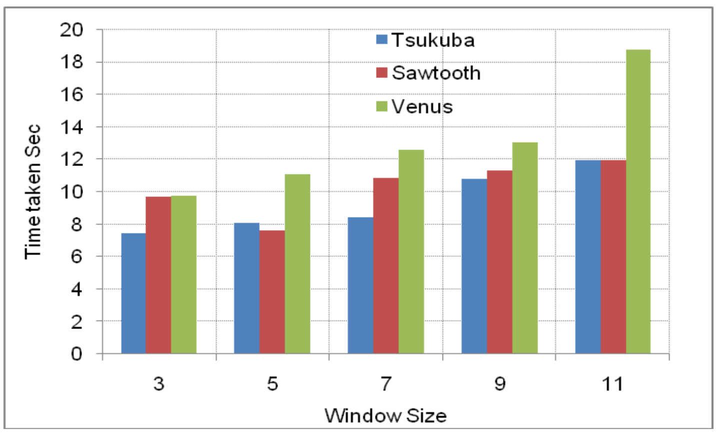
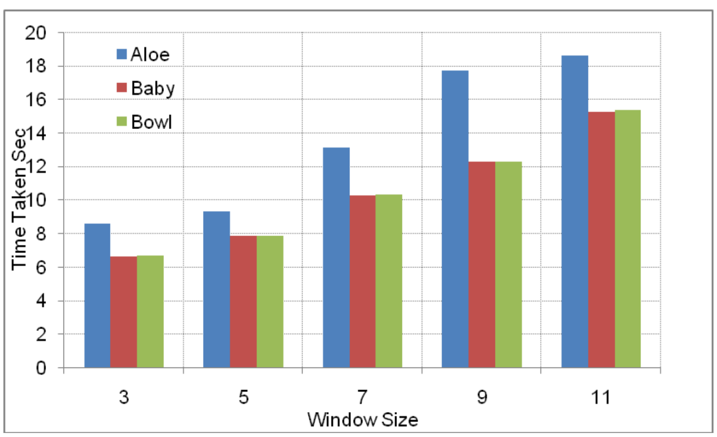


FIGURE 9: Evaluation of SAD Algorithm: Window size vs %RMS Error for PNG and PPM images.

The algorithm gives fast results, it takes around 7.4 sec for Tsukuba and 6.62 sec for Baby with smaller window sizes. The algorithm shows more error for texture less images compared to images with good texture which means that this algorithm is suitable for images having good texture. In general the algorithm gives a better performance –faster, shows less error and less computational cost for images taken under good illumination condition



(a)



(b)

FIGURE 10: Evaluation of SAD Algorithm: Window size vs. Time Taken for (a) PPM&(b) PNG images.

5.2 Observations in Adaptive Binary Support Window Algorithm

The adaptive algorithm avoids edge fattening and shows an error of 0.7382% for Tsukuba image, with window size of 23x23. If we further increase the window size this algorithm will work with a much reduced error. But with PNG images the algorithm shows more error.

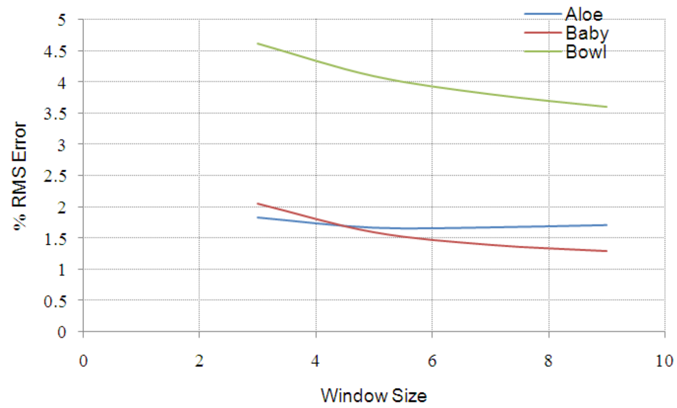


FIGURE 11: Adaptive binary support window algorithm error evaluation. Window size vs %RMS Error PNG images.

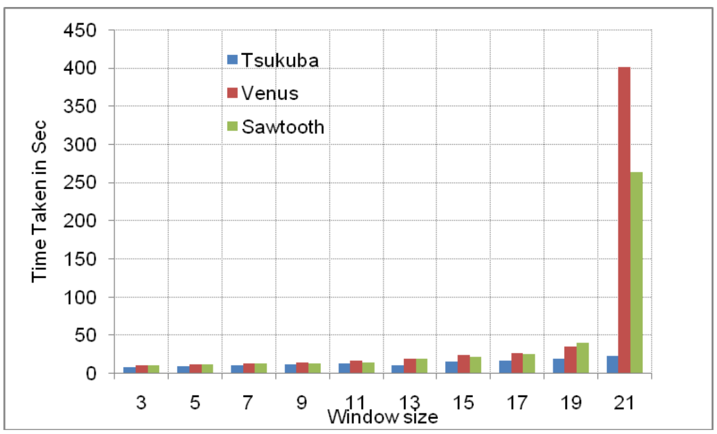


FIGURE 12: Evaluation of Adaptive binary support window Algorithm: Window size vs. Time Taken for PPM images.

Time taken for Adaptive binary support window Algorithm is 22.7 seconds for Tsukuba image with 21x21 window. The computational cost is slightly more compared to fixed window. This algorithm gives more accurate result for textureless images and make it good candidate for real-time application under bad illumination environments.

6. CONCLUSION AND FUTURE WORK

The implemented algorithms can be used for generating accurate results in different environments, suitable tor autonomous robots. The comparative study of the implemented algorithms infer that the SAD fixed window generates finer details of object's depth with moderate speed, for input images with good texture ie ,this algorithm is suitable for ideal lightning conditions .The adaptive binary window algorithm shows good accuracy for images taken under extreme bad illumination condition.

Even though this work meets its initial objective, it points towards various interdisciplinary research directions. The extension of work looks forward to an algorithm with calibration and rectification of images, which can generate more accurate results for textured and texture less images in very less time, based on FPGA implementation. From the knowledge obtained through this work more efficient stereo algorithms for various radiometric conditions, suitable to real-time applications can be developed .The above developed algorithms can be used as basis for achieving the autonomous capabilities of robotic assistance in various areas such as defense ,civil protection and interplanetary applications which makes them really interesting applied research field.

7. REFERENCES

[1] Fusiello, A., Trucco, E., & Verri, A. (2000). A compact algorithm for rectification of stereo pairs. Machine Vision and Applications, 12(1), 16-22.

[2] Hamzah, R. A., Rahim, R. A., & Noh, Z. M. (2010, July). Sum of Absolute Differences algorithm in stereo correspondence problem for stereo matching in computer vision application. In Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on (Vol. 1, pp. 652-657). IEEE.

[3] Konolige, K. (1999). Stereo geometry. SRI International.

[4] Tsai, R. Y. (1986). An efficient and accurate camera calibration technique for 3D machine vision. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1986.

[5] Kanade, T., & Okutomi, M. (1994). A stereo matching algorithm with an adaptive window: Theory and experiment. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 16(9), 920-932.

[6] Fusiello, A., Roberto, V., & Trucco, E. (1997, June). Efficient stereo with multiple windowing. In 2013 IEEE Conference on Computer Vision and Pattern Recognition (pp. 858-858). IEEE Computer Society.

[7] Veksler, O. (2002). Stereo correspondence with compact windows via minimum ratio cycle. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 24(12), 1654-1660.

[8] Prazdny, K. (1985). Detection of binocular disparities. Biological Cybernetics, 52(2), 93-99.

[9] Xu, Y., Wang, D., Feng, T., & Shum, H. Y. (2002). Stereo computation using radial adaptive windows. In Pattern Recognition, 2002. Proceedings. 16th International Conference on (Vol. 3, pp. 595-598). IEEE.

[10] Yoon, K. J., & Kweon, I. S. (2006). Adaptive support-weight approach for correspondence search. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(4), 650-656.

[11] Tombari, F., Mattoccia, S., & Di Stefano, L. (2007). Segmentation-based adaptive support for accurate stereo correspondence. In Advances in Image and Video Technology (pp. 427-438). Springer Berlin Heidelberg.

[12] De-Maeztu, L., Villanueva, A., & Cabeza, R. (2011). Stereo matching using gradient similarity and locally adaptive support-weight. Pattern Recognition Letters, 32(13), 1643-1651.

[13] Nalpantidis, L., & Gasteratos, A. Biologically and psychophysically inspired adaptive support weights algorithm for stereo correspondence. Robotics and Autonomous Systems, 58(5), 457-464.

[14] Nalpantidis, L., & Gasteratos, A. (2010). Stereo vision for robotic applications in the presence of non-ideal lighting conditions. Image and Vision Computing, 28(6), 940-951.

[15] Wertheimer, M. (1938). Laws of organization in perceptual forms.

[16] http://vision.middlebury.edu/stereo/data/

[17] Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International journal of computer vision, 47(1-3), 7-42.

[18] Wang, L., Gong, M., Gong, M., & Yang, R. (2006, June). How far can we go with local optimization in real-time stereo matching. In 3D Data Processing, Visualization, and Transmission, Third International Symposium on (pp. 129-136). IEEE.

Accelerated Joint Image Despeckling Algorithm in the Wavelet and Spatial Domains

Adib Akl adibakl@usek.edu.lb
Faculty of Engineering, Department of Telecommunications
Holy Spirit University of Kaslik (USEK)
Jounieh, P.O. Box 446, Lebanon

Charles Yaacoub charlesyaacoub@usek.edu.lb
Faculty of Engineering, Department of Telecommunications
Holy Spirit University of Kaslik (USEK)
Jounieh, P.O. Box 446, Lebanon

Abstract

Noise is one of the most widespread problems present in nearly all imaging applications. In spite of the sophistication of the recently proposed methods, most denoising algorithms have not yet attained a desirable level of applicability. This paper proposes a two-stage algorithm for speckle noise reduction jointly in the wavelet and spatial domains. At the first stage, the optimal parameter value of the spatial speckle reduction filter is estimated, based on edge pixel statistics and noise variance. Then the optimized filter is used at the second stage to additionally smooth the approximation image of the wavelet sub-band. A complexity reduction algorithm for wavelet decomposition is also proposed. The obtained results are highly encouraging in terms of image quality which paves the way towards the reinforcement of the proposed algorithm for the performance enhancement of the Block Matching and 3D Filtering algorithm tackling multiplicative speckle noise.

Keywords: Denoising, Spatial Filter, Speckle, Wavelet.

1. INTRODUCTION

Multiplicative in nature, speckle noise is a common problem found in different imaging applications such as ultrasound, sonar and radar imaging [1, 2]. Originating from the superposition of acoustical echoes coming with random phases and amplitudes during acquisition or transmission, it tends to reduce the image resolution and contrast and blur important details [1].

Despeckling can address the multiplicative nature of the noise, or transform the noisy image to the logarithmic domain where multiplicative noise becomes additive, and apply additive noise reduction techniques [2].

In the past decades, several algorithms have been proposed for image denoising. Hybrid order statistics filters for speckle reduction are proposed in [2]. A preprocessing technique is used in order to transform the noise in the logarithmic domain to a Gaussian-like noise, which allows for better filtering results using known denoising techniques [3-5]. Lee [6-8], Kuan [3, 9], and Frost [10] filters are still widely used in many applications. In general, they succeed to reduce speckle in homogenous areas. However, in heterogeneous areas, speckle is retained. Therefore, they are not able to perform a full removal of speckle without blurring any edges because they rely on local statistical data related to the filtered pixel and this data depends on the occurrence of the filter window over an area. Wavelet-based denoising techniques [4, 10, 11, 12] represent the image jointly in the spatial and frequency domains. They rely on the sparse representation of the wavelet sub-bands coefficients in order to be distinctly thresholded. The Block Matching and 3D

Filtering (BM3D) algorithm [13, 14] is a non-local denoising technique in a non-spatial domain. It combines non-local image modeling and the sparse representation of the wavelet domain.

In the following, a two-stage despeckling algorithm is proposed. It consists of jointly denoising the speckled image in the wavelet and spatial domains. At the first stage, the traditional speckle reduction filter of Kuan [3] is adapted to the specificities of the filtered image by estimating its optimal parameter value automatically, based on edge pixels statistics and noise variance. Then the resulting sub-optimal spatial filter is used at the second stage to smooth the approximation sub-band coefficients in the wavelet domain. Since wavelet decomposition is time-consuming when dealing with large sized images, a complexity reduction algorithm is also proposed. Note that the Kuan filter is used as the basis for the spatial filtering due to its versatility. However, other algorithms could be used.

The remainder of this paper is organized as follows. Existing noise reduction methods are reviewed in Section 2. The proposed algorithm is then described in Section 3. In Section 4, results are shown and discussed. Finally, conclusions are drawn and prospects are provided in Section 5.

2. EXISTING NOISE REDUCTION ALGORITHMS

In the past decades, several image denoising algorithms have been proposed. For instance, the spatial means filters (arithmetic mean, geometric mean, harmonic mean and contraharmonic mean filters [15]) smooth the local variations in an image by blurring the noise. The spatial order-statistic filters (median, max, min, midpoint and alpha-trimmed mean filters [15]) are based on ordering (ranking) the values of the pixels contained in the image area encompassed by the filter. The spatial adaptive filters (adaptive local noise reduction and adaptive median filters [15]) change their smoothing behavior based on statistical characteristics of the image inside the filter region. The spatial filters listed above have shown to successfully deal with a large panel of noisy images in situations when only additive noise is present. However, they are not able to reduce speckle noise which is non-additively combined with the underlying image.

Various nonlinear filtering techniques have also been proposed [16]. They seek to reduce the effect of speckle noise while preserving the informative structure of the underlying image. Some of the best known standard speckle noise reduction filters are the methods of Lee [6-8], Gamma [24], Kuan [3, 9] and Frost [24]. These filters use the second-order sample statistics within a minimum mean squared error estimation approach.

2.1 Lee Filter

The locally adaptive Lee multiplicative filter [7, 8] is based on a multiplicative noise image model as follows:

$$g(x,y) = f(x,y) \times n(x,y) \ , \qquad (1)$$

where g denotes the observed image, f the original image and n the multiplicative speckle noise [6].

Based on an assumption that the noise is white, with unity mean, and uncorrelated with the image f, the multiplicative Lee filter gives the best mean-squared estimate of f at each pixel $g(x,y)$ by:

$$z(x,y) = \overline{f} + k_{(x,y)}\left[g(x,y) - \overline{f} \right], \qquad (2)$$

where $z(x,y)$ is the filtered pixel, $\overline{f}$ is the mean of the original image pixels under the filtering window (i.e. the local mean) and $k_{(x,y)}$ is the gain factor given by:

$$k_{(x,y)} = \frac{\sigma_f^2}{\overline{f}^2 \sigma_n^2 + \sigma_f^2} \ , \qquad (3)$$

where σ_f^2 is the variance of the original image pixels under the filtering window (local variance), and σ_n^2 is the noise variance. The local adaptation of the filter is based on the calculation of the local statistics $\overline{f}$ and σ_f^2 from the data sample estimates $\overline{g}$ and σ_g^2 determined over a local neighborhood window.

This requires the knowledge of the Coefficient of Variation (CoV), which is the ratio of the standard deviation to the mean in homogeneous areas. If the original image is not available, the noise variance can be determined from the number of looks or the Equivalent Number of Looks (ENL). This parameter effectively controls the amount of smoothing applied to the image. The filter size greatly affects the quality of the processed image. If the filter is too small, the noise filtering algorithm is not efficient. If the filter is too large, some details of the image will be lost [17-19].

An improved version of the basic Lee filter also exists; the modified Lee filter which relies on a ratio-based edge detection algorithm used to estimate the edge strength at each pixel in the image [20].

2.2 The Gamma Maximum A Posteriori Filter

The Gamma Maximum A Posteriori (MAP) filter [21] is used primarily to filter speckled radar data while preserving high frequency features. It's based on a Bayesian analysis of the image statistics. It performs spatial filtering on each individual pixel in an image using the grey level values in a square window surrounding each pixel. The Number of Looks (NLOOK) parameter effectively controls the amount of smoothing applied to the image. It affects the speckle coefficient of variation (C_u) and the upper speckle coefficient of variation $(C_{\max})$. A small NLOOK value leads to more smoothing and a larger NLOOK value preserves more image features.

If $C_i \leq C_u$ (C_i is the image variation coefficient), the filtering window is over a homogenous area and smoothing is applied. If $C_i \geq C_{max}$, the filtering window is over an area of high local standard deviation (crossing edge-pixels), and therefore edge-pixel is replicated [24].

If $C_u < C_i < C_{max}$, the filtered pixel value based on the Gamma estimation of the contrast ratios within the appropriate filter window is given by:

$$z(x,y) = \frac{(W - NLOOK - 1)\times \overline{g} + \sqrt{D}}{2\times W}, \tag{4}$$

where:

$W = \dfrac{1+C_u^2}{C_i^2 - C_u^2}$, $D = \overline{g}^2 \times (W - NLOOK - 1)^2 + 4\times W\times NLOOK \times \overline{g} \times g(x,y)$ and $\overline{g}$ is the mean of the window pixels.

2.3 Kuan Filter

Similarly to the Lee filter, Kuan filter [3, 9, 22] is designed to smooth out speckle noise while retaining shape features in the image by applying the Minimum Mean Square Error (MMSE) criterion and it is applied to the logarithmic transformation of the noisy image. Kuan filter is used primarily to filter speckled radar data. It performs spatial filtering on each individual pixel in an image using the grey level values in a square window surrounding each pixel [24]. The resulting grey-level value for the smoothed pixel is given by:

$$z(x,y) = g(x,y) \times W + \bar{g}(1-W) \ , \tag{5}$$

where *g(x,y)* is the center pixel of the filtering window, $\bar{g}$ is the mean value of intensity within the filtering window, and *W* is a weighting function which depends on the Number of Looks (NLOOK) parameter [23] and given by:

$$W = \frac{1 - C_u^2 / C_i^2}{1 + C_u^2} \ , \tag{6}$$

where C_u and C_i are the estimated noise variation coefficient and the image variation coefficient, respectively:

$$C_u = \sqrt{\frac{1}{\mathrm{NLOOK}}}, \qquad C_i = \frac{\sigma_g}{\bar{g}}, \tag{7}$$

where σ_g is the standard deviation of intensity within the filtering window.

Same as the Lee filter case, a small NLOOK value corresponds to a high noise variation and a low *W* and therefore leads to more smoothing, while a larger NLOOK value corresponds to a low noise variation and a high *W* which preserves more image features.

Theoretically, the correct value for NLOOK should be the Equivalent Number of Looks (ENL) of the image. Depending on the application, the user may experimentally adjust the NLOOK value so as to control the effect of the filter [22].

2.4 Frost Filter

Frost method consists in adjusting the filter's parameters according to local area statistics about the target pixel [24]. When uniform regions are filtered, the filter acts as a mean filter and when high contrast regions are filtered, the filter acts as a high-pass filter with rapid decay of elements away from the filter center. Thus, large uniform areas will tend to be smoothed out and speckle removed, while high contrast edges and other objects will retain their values [24]. The frost filter can be considered as an adaptive-weighted-mean filter since it uses an adaptive filtering algorithm, which is an exponential damped convolution kernel that adapts itself to features by computing a set of weighting factors for each pixel within the filtering window as follows:

$$M_n = e^{\left(-\left(DAMP \times \left(\frac{\sigma_g}{\bar{g}} \right)^2 \right) \times T \right)} \ , \tag{8}$$

where *DAMP* is a factor that determines the extent of the exponential damping for the image, σ_g is the standard deviation of the filter window, $\bar{g}$ is the mean value within the window and *T* is the absolute value of the grey level distance between the center pixel and its surrounding pixels in the filter window [24].

The Enhanced Frost filter is an extension to the basic filter that further divides the image into homogeneous, heterogeneous and isolated point target areas [19]. It can significantly improve the ability of speckle mitigation in the vicinity of edges and small features, thus retaining more image details.

2.5 Denoising Using Wavelets

Wavelet transform [25] has been extensively studied in recent years and used for many application domains, mainly in image compression and noise reduction. It consists of a set basis functions used to analyze signals in both spatial and frequency domains simultaneously. The basic functions of the wavelet transform help to isolate signal discontinuities since high pass filtering is used to obtain detail information and low pass filtering is used to retrieve a smoother approximation of the signal, making it possible to analyze the signal at different scales [22].

Multi-resolution processing of the Discrete Wavelet Transform (DWT)

Multi-resolution processing consists in constructing a set of child wavelets from a mother wavelet using scaling and wavelet functions [22]. These two functions form a filter bank consisting of a low pass and high pass filters. The idea of multi-resolution processing through wavelet decomposition is to pass the signal through the filter bank; the signal is decomposed to detail coefficients (output of the high pass filter) and approximation coefficients (output of the low pass filter). Then, the filter outputs are down-sampled by 2 in the purpose of discarding half the samples. The decomposition is repeated to further increase the frequency resolution [25-27].

Application to Image Denoising

Considering an image corrupted by an additive noise and modeled as $g = f + n$, where f denotes the unknown, noise-free image and n the noise [28, 29], wavelet-based denoising (*Figure 1*) consists in:

1) Applying the DWT to g.
2) Thresholding the detail coefficients (Wavelet shrinkage).
3) Inverse transforming (IDWT) the result to obtain an estimation z of the original image f.

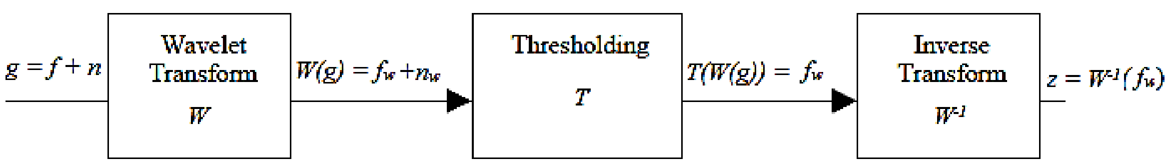


FIGURE 1: DWT Denoising Block Diagram.

When dealing with speckled images, a logarithmic transformation is applied to the noisy image before wavelet decomposition, to transform the multiplicative noise, into additive noise and after wavelet reconstruction, an exponential transformation is applied to reverse the logarithmic operation.

In wavelet domain, the output of the low pass filter consists of the high magnitude and low frequency components (approximation coefficients) and the output of the high pass filter consists of the low magnitude and high frequency components (detail coefficients). *Figure 2* shows a one level, 2D wavelet decomposition scheme [30] where L(.) represents the low pass filtering operator, and the subscript ($_L$) represents a low pass filtered output. Similarly, H(.) and the subscript ($_H$) represent high pass filtering and a high-pass filtered output, respectively. The symbol (↓2) represents a down-sampling operator by a factor of 2.

As a result of filtering and down sampling (*Figure 2*), four sub-bands are obtained: scaling (approximation) coefficients (g_{LL}), horizontal detail coefficients (g_{HL}), vertical detail coefficients (g_{LH}) and diagonal detail coefficients (g_{HH}). The (HL), (LH) and (HH) sub-bands, represent the high frequency (and low magnitude) components and the (LL) sub-band represents the low frequency (and high magnitude) components. At the next level of decomposition, only the (LL) component is passed to the decomposition process.

The reconstruction process or the Inverse Discrete Wavelet Transform (IDWT) consists of assembling back the wavelet coefficients to the original image. After each inverse low and high pass filtering, an up-sampling process (zeros insertion) is required to reverse the decomposition process.

The main idea of wavelet denoising is to threshold only the high frequency components while preserving most of the features in the image by retaining the approximation sub-band.

There are two thresholding methods frequently used; hard [26, 31] and soft [32] thresholding. In hard thresholding, the input is kept if its amplitude is greater than a threshold *T*, otherwise it is forced to zero. In soft thresholding, if the absolute value of the input is less than or equal to a threshold *T*, then the output is forced to zero, otherwise, the output is a scaled version of the input.

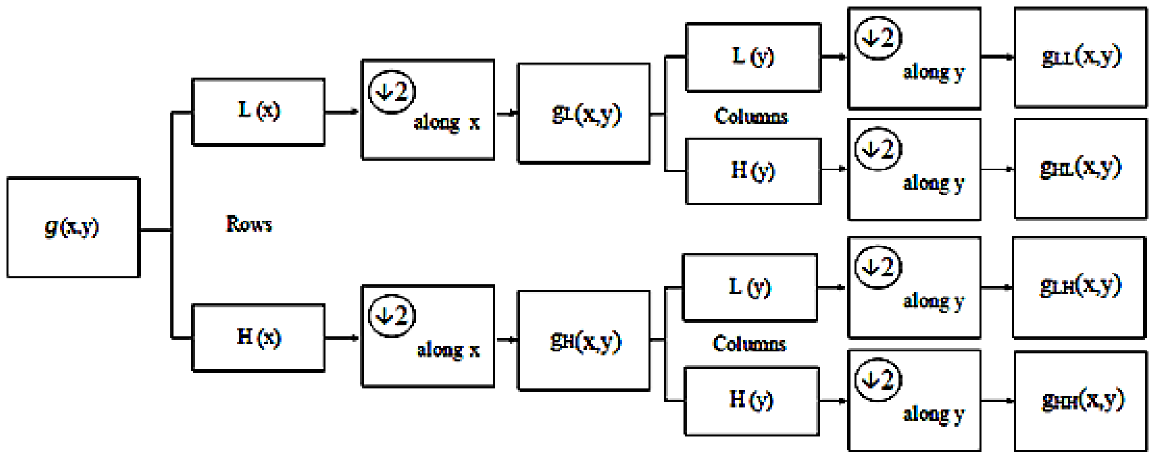


FIGURE 2: One level, 2D Wavelet Decomposition Block Diagram.

Finding a suitable threshold is an important task in wavelet shrinkage. In fact, choosing a very large threshold will shrink to zero almost all coefficients, which results in over smoothing the image, while choosing a very small threshold will yield a noisy result which is close to the input [29]. VisuShrink [25] is an approach that uses a Universal (global) threshold applied to all sub-bands and scales after decomposition [33]. However, this threshold can be unwarrantedly large because it depends on the number of pixels yielding overly smoothed images. In addition, it ignores the difference between sub-bands at different scales [34]. BayesShrink [25, 31] is a sub-band adaptive threshold selection technique that determines a specific threshold for each sub-band assuming a Generalized Gaussian Distribution (GGD) [35]. This method depends on the standard deviation of the noise-free image, and the GGD shape parameter [34]. The method consists of finding, for each sub-band, a threshold that minimizes the expected value of the mean square error (Bayesian Risk) [32, 36].

2.6 The Block Matching and 3D Filtering (BM3D) Algorithm

The BM3D algorithm is a non-local denoising technique [13, 14] in a non-spatial domain. It combines non-local modeling and the sparse representation of the wavelet domain. The algorithm is divided into three processing stages: Block matching, 3D filtering, and aggregation.

In the block matching stage, the noisy image *g* is divided into blocks $G_{x\in X}$ where *x* represents the position of each block in the whole image *X*. For each block, the patches with most resemblance are grouped in a 3D array. Therefore, groups whose elements have a high degree of similarity are constructed separately. Two patches are similar if the Euclidean distance between them is less than a given threshold [14, 37]. The block matching advantages are the induction of a high correlation in the third dimension of the 3D array and the improvement of the dispersion of all possible configurations of details present in the image [13].

3D filtering is a procedure that jointly filters a group of similar blocks by exploiting the similarities between the grouped images and inside each image in the group. It consists of three different sections: 3D transform, shrinkage and inverse 3D transform. Those three procedures are jointly applied [13] on the block. The 3D transform consists of applying a 2D wavelet transform on each patch of the block then a 1D wavelet or Discrete Cosine Transform (DCT) on the resulting patches. As a result of the sparse representation given by the 3D transform, the shrinkage process can effectively attenuate the noise by eliminating the coefficients relying under a certain chosen threshold. The inverse 3D transform consists of assembling back the thresholded coefficients to reconstruct the 3D block.

Finally, the aggregation stage consists of combining the patches in the 3D filtered groups including the reference patch. A trivial solution is to compute the weighted mean of all the estimated patches overlapping at a pixel position [13].

3. THE PROPOSED WAVELET/SPATIAL DESPECKLING ALGORITHM

In this section, a two-stage despeckling algorithm which consists in jointly denoising the corrupted image in the wavelet and spatial domains is proposed. At the first stage, an automatic estimation of the optimal Kuan filter parameter value based on edge pixels statistics and noise variance is developed. Then the resulting adaptive filter is used at the second stage to spatially smooth the approximation sub-band coefficients in the wavelet domain. Since a large number of wavelet levels makes the despeckling algorithm computationally expensive, a complexity reduction algorithm for wavelet decomposition is also proposed. Note that the proposed enhancement targets the Kuan filter due to its versatility and adaptability to speckle noise strength, while a similar study could have been performed with another spatial despeckling method to be used in the hybrid filter that will be discussed in Section 3.2; we therefore focus on the Kuan filter without loss of generality.

3.1 The Adaptive Spatial Filter

As explained earlier, Kuan filtering relies on the Number of Looks (NLOOK) parameter which significantly affects the filtering performance and is usually taken equal to the Equivalent Number of Looks (ENL) of the image. This parameter is manually adjusted, in order to control the strength of the smoothing applied to the image.

We start by implementing the additive model of the Kuan filter using a set of test images (e.g. Lena, Mandrill, LivingRoom, … [38]) corrupted by speckle noise with different distributions. By setting NLOOK = ENL, we notice that filtered images are over-blurred, that's why came our idea of analyzing the Kuan filter performance with respect to the NLOOK parameter and proposing a novel technique for selecting a suitable NLOOK value that yields near-optimal performance compared to manual parameter selection, therefore eliminating the need for several runs of the filter to reach the optimal performance. In this purpose, we define the adjusted NLOOK value as:

$$NLK\_AD = (1 + A)ENL \ , \qquad (9)$$

where *A* is the adjustment coefficient.

Figure 3 shows an example of the Lena image filtered with NLOOK=ENL (middle) and NLK_AD as defined in (9), for *A* = 0.65 (right). It can be clearly seen that the image with the adjusted NLOOK parameter (NLK_AD) is much sharper and its details are better preserved.

In the first part of this study, we aim at finding a closed form expression for the adjustment coefficient that yields near optimal performance, without the need for several runs of the filter with different NLOOK values. Therefore, we consider the effect of the filter window size, the image detail, the noise distribution and the noise variance on the optimal adjustment coefficient *A* and consequently, the NLK_AD parameter. Thus, all the listed parameters are fixed and the effect of varying each one of them is studied.

FIGURE 3: Kuan filter results with different NLOOK values. Gaussian and Uniform noise distributions are used in the top and bottom rows respectively. Each row shows (from left to right): the noisy image, the Kuan filtered image using NLOOK=ENL and the corresponding Kuan filtered image using NLK_AD with *A*=0.65. An 11×11 window with images of size 256×256 are used.

Effect of the Algorithm Parameters on the Optimal Adjustment Coefficient

A Gaussian speckle noise having a variance of 2600 is added to the Cameraman image [38] in order to find the adjustment coefficient *A* that yields the best result using different window size. The Peak Signal-to-Noise Ratio (PSNR) and the Signal to Mean Squared Error (S/MSE) variations with respect to the NLK_AD parameter are plot and ensure that, for different odd-sized filter windows, the optimal values of *A* are nearly constant.

On the other hand, test images are distorted with multiplicative speckle noise using the Uniform, Gaussian, Rayleigh, and Erlang distributions while fixing all the other parameters and the optimal value of *A* remains the same regardless of the noise distribution, for each test image.

In the purpose of studying the influence of the image itself on the filtering process, the filter is applied on different images, having different features, from [38] and [39] with all the other parameters being fixed. It can be noticed that the optimal adjustment coefficient varies with image details. For example, the Pirate image [38] has a high percentage of edges, therefore, the obtained optimal adjustment coefficient is greater than with the Lena image, which is expected since more image features have to be preserved.

To find the relationship between the optimal adjustment coefficient *A* and the amount of image detail, edge detection is performed on the set of images and the curves presenting the optimal *A* with respect to the noise variance and the number of edge pixels in each image are drawn.

The Logarithmic and Polynomial Approximations

Since the noise variance affects the filtering process, the variation of *A* with the noise variance and image detail is modeled as:

$$A = A_d + A_v \ , \tag{10}$$

where A_d and A_v are the detail and the noise variance components in *A*, respectively.

We observe that a higher noise variance corresponds to a lower value of *A*, which is expected since a lower *A* value leads to more smoothing. Additionally, a higher noise variance corresponds to higher edge percentage (P_e) values because the edge detection algorithm cannot completely differentiate between noise and small edges. Curves for different noise variances have similar behavior with respect to the percentage of edges in the image, therefore, we take as reference one of the curves (for a given variance).

The curve with the solid dark line in *Figure 4* shows the measured A_d with respect to *Pe*. It can be seen that, for *Pe* exceeding 2%, A_d varies in a logarithmic or second order polynomial fashion. The former can be expressed as:

$$A_d = \alpha \ln(P_e) + \beta \ , \qquad (11)$$

and the latter as:

$$A_d = \lambda P_e^{\,2} + \delta P_e + \mu \ , \qquad (12)$$

where α = 0.4015, β = 0.054, λ = –0.01215, δ = 0.20512 and μ= –0.01394 determined by fitting the measured results with (11) and (12) using the Least Squares method.

It can be observed in *Figure 4* that, as the image detail (*Pe*) increases, A_d increases and consequently, NLK_AD increases. Therefore, less smoothing is performed, as expected. The logarithmic and polynomial approximations mainly overlap with measured data, except for *Pe* < 2%, which proves high accuracy of the models in (11) and (12).

A similar study with respect to the noise variance shows that A_v (*Figure 5*) can be accurately modeled by a second order polynomial as:

$$A_v = a(v-2100)^2 + b(v-2100) \ , \qquad (13)$$

where $a = -2\text{x}10^{-8}$ and $b = -42\text{x}10^{-6}$ determined by curve fitting.

Note that for v = 2100, A_v = 0. In fact, A_v reflects the variation of the adjustment coefficient with respect to the noise variance compared to the reference v = 2100 which has been used to derive A_d in (11) and (12). On the other hand, A_v decreases as v increases, which is also expected since more smoothing is required (lower value of NLK_AD) when the noise increases.

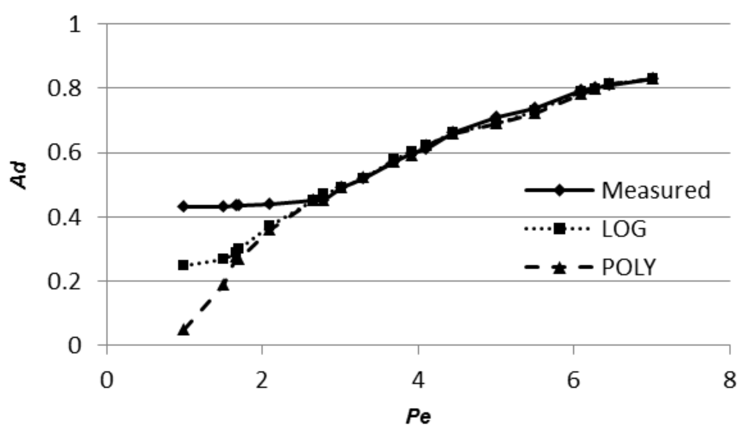


FIGURE 4: Measured and approximated values for A_d for a noise variance of 2100.

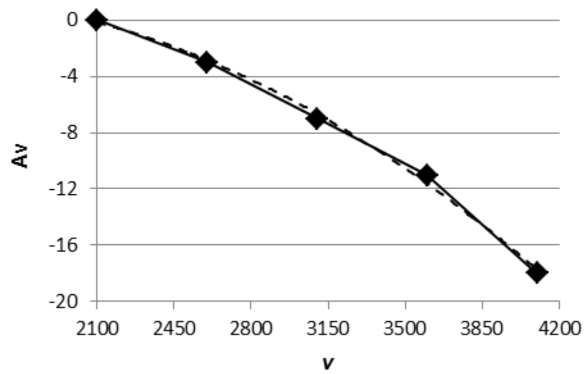


FIGURE 5: Measured (solid line) and approximated (dashed line) values for A_v.

3.2 The Hybrid Wavelet-Spatial Filter

In wavelet denoising, only detail coefficients are thresholded and approximation coefficients remain stagnant, as explained earlier. This is intuitive since the approximation component is a low-frequency image, usually assumed to be noise-free. In fact, this is only a theorectical aspect since practically, approximation coefficients also contain speckle noise as it can be clearly observed in the example shown in *Figure 6*. Wherefore, the proposed algorithm consists of smoothing the approximation coefficients in addition to detail coefficients thresholding as follows:

1) Applying the Discrete Wavelet Transform to the speckled image.
2) Thresholding the detail coefficients (Wavelet shrinkage).
3) Spatial Smoothing of the approximation sub-band coefficients.
4) Applying the Inverse Discrete Wavelet Transform to the result.

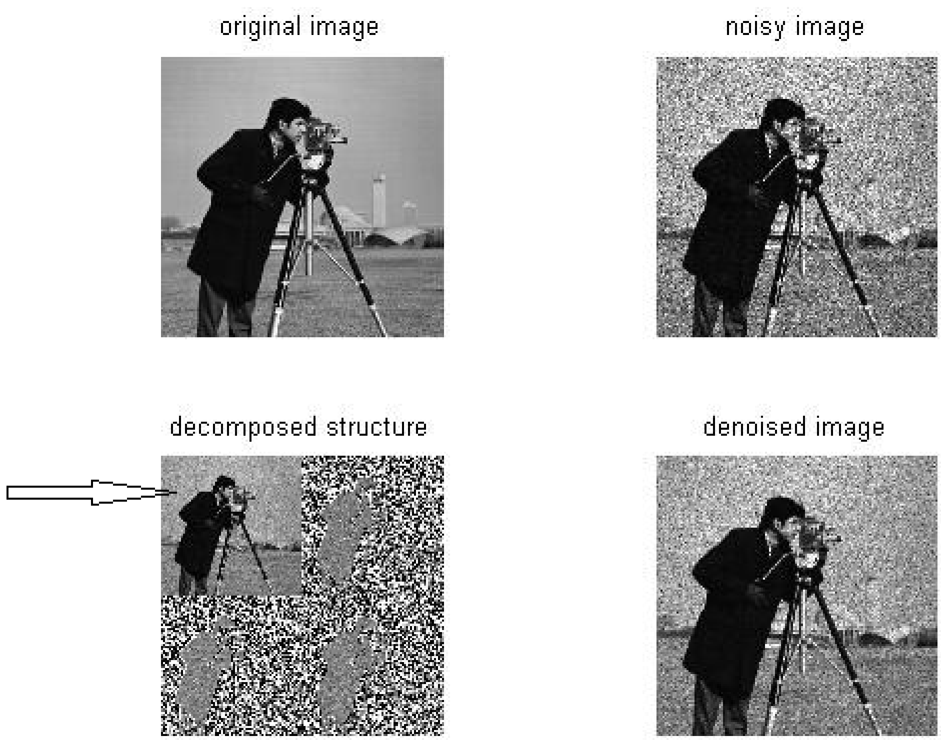


FIGURE 6: Cameraman image of size 512×512 (up-left) [38], corrupted by speckle noise of unit mean and a variance of 2400 (up-right), decomposed structure (down left) and the denoised image (down right). Note the existence of noise in the approximation image of the decomposed structure and the remaining noise in the denoised image (using sym1 wavelets with BayesShrink threshold selection).

An important issue in the approximation image spatial smoothing would be the preservation of the image characteristics. In other words, the algorithm must efficiently smooth out the noise in the approximation image without blurring its details or reducing the contrast and quality. Otherwise, the image structure and dynamics would be lost. Therefore, the approximation sub-band coefficients in the wavelet domain are smoothed using the previously developed adaptive Kuan filter which leads to a better speckle removal without image distortion, as will be seen in Section 4.

3.3 The Proposed Wavelet Acceleration Algorithm

This section deals with accelerating the wavelet filtering process used in the proposed method. The acceleration algorithm consists in finding the number of decomposition levels that yields near optimal performance, without the need for several runs of the wavelet filter with different number of decomposition levels, regardless of the image size, its edge percentage, the noise variance, the wavelet type and base, the wavelet threshold selection technique and the wavelet thresholding technique.

Relative Difference Threshold Selection

We start by applying 100 times, wavelet denoising on an image corrupted by Gaussian speckle noise using different number of decomposition levels (from 1 level to the maximum number of decomposition levels calculated as in [40]). For every run, the simulated speckle noise is regenerated and the PSNR, the Coefficient of Correlation (CoC) and the ENL metrics of the denoised image are calculated and saved.

The second step is the calculation of the relative difference between the ENL of each two consecutive decomposition levels, for the purpose of finding a reasonable relative difference threshold that allows the algorithm to stop before reaching the maximum number of decomposition levels, without loss in performance. The same is performed for different images from [38] and [39], different noise variances (2500, 2600, 3000, 3100, 3500, 4100, 4600 and 6000), different image sizes (128×128, 256×256, 512×512 and 1024×1024), different wavelet types and bases (db2, db4, db7, sym1, sym2 and sym6) and using bayesShrink threshold selection.

After analyzing the results, we noticed that a relative difference smaller than 0.2 leads to an approximately steady PSNR value. As a result, 0.2 is taken as a threshold. Therefore, our proposed algorithm consists of applying a level of wavelet decomposition and denoising, computing the relative difference between the corresponding ENL and the ENL of the previous decomposition level, and stop when the relative difference becomes less than 0.2 as shown in *Figure 7*.

By limiting the number of decomposition levels, a computational gain is expected at the expense of some loss in image quality, as will be shown in Section 4.

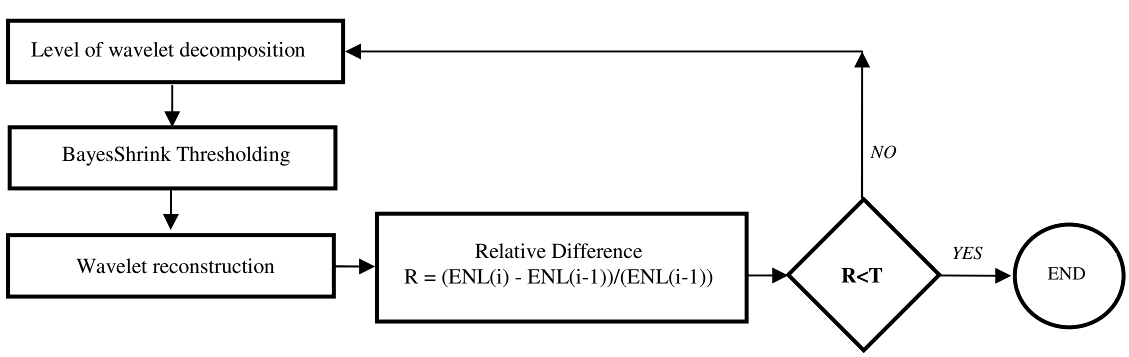


FIGURE 7: Flowchart of the Proposed Acceleration Algorithm.

4. PRACTICAL RESULTS

In this section, the proposed algorithms are put in practice. Besides the subjective qualitative evaluation based on human perception, a quantitative analysis is essential. Thus, an objective benchmark is used to study the quality-related outcomes of the filtering process. Namely, the Peak Signal-to-Noise Ratio (PSNR), the Signal-to-Mean-Squared-Error (S/MSE), the Coefficient of Correlation (CoC) and the Equivalent Number of Looks (ENL).

Peak Signal-to-Noise Ratio (PSNR)

PSNR is considered to be the least complex metric, as it defines the image quality degradation as a plain pixel by pixel error power estimate. PSNR is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation [17]. In image processing, the signal is the original image, and the noise is the error (blur) introduced by the denoising procedure.

Signal-to-Mean Squared Error (S/MSE)

In order to quantify the achieved performance improvement, the Signal-to-Mean Squared Error can also be computed, based on the original and the noisy/denoised images as follows:

$$S/MSE = 10\log_{10}\left(\frac{\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I(i,j)\right]^2}{\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[K(i,j)-I(i,j)\right]^2}\right), \quad (14)$$

where I is the original image and K the noisy or denoised image. This measure corresponds to the classical SNR in the case of additive noise [41].

Coefficient of Correlation (CoC)

In ultrasound imaging, it is important to suppress speckle noise while at the same time preserving the edges of the original image that often constitute features of interest for diagnosis. For this reason, we also considered a qualitative measure for edge preservation, the Coefficient of Correlation (CoC) metric:

$$CoC = \frac{\Gamma\left(\Delta_I - \overline{\Delta_I}, \Delta_K - \overline{\Delta_K}\right)}{\sqrt{\Gamma\left(\Delta_I - \overline{\Delta_I}, \Delta_I - \overline{\Delta_I}\right).\Gamma\left(\Delta_K - \overline{\Delta_K}, \Delta_K - \overline{\Delta_K}\right)}}, \quad (15)$$

where Δ_I and Δ_K are the high-pass filtered versions of I and K respectively, obtained with a 3×3 pixel standard approximation of the Laplacian operator [25], $\overline{\Delta_I}$ is the mean value of Δ_I and $\overline{\Delta_K}$ is the mean value of Δ_K and the operator Γ is given by:

$$\Gamma(I,K) = \sum_{i=0}^{m-1}\sum_{j=0}^{n-1} I(i,j).K(i.j), \quad (16)$$

where $m\times n$ is the size of the corresponding images I and K.

The CoC cannot exceed 1 in absolute value. It is 1 in the case of an increasing linear relationship and - 1 in the case of a decreasing linear relationship. Its value lies in between in all other cases, indicating the degree of linear dependence between the images. The absolute value of the correlation measure should be close to unity to an optimal effect of edge preservation [41].

Equivalent Numbers of Looks (ENL)

The Equivalent Number of Looks is also a good approach for either estimating the speckle noise variance in a noisy image (in general a SAR image), or evaluating the performance obtained after filtering. ENL is often calculated over a uniform region, but due to difficulties in identifying uniform areas, the image is divided into small areas of 25 x 25 pixels, the ENL is computed for each area

(17), then the average of these ENL values is taken as explained in [12]. Note that this latter method is used throughout this paper. A large ENL value usually corresponds to a better performance. The formula for the ENL calculation is given by:

$$ENL = \left(\frac{\text{mean}}{\text{standard deviation}} \right)^2 . \tag{17}$$

Note that PSNR, S/MSE, and CoC are quantitative metrics, they are based on a pixel by pixel calculation and the original image is needed for computation, whereas ENL is based on the mean and standard deviation, and it can be computed without the need of the original image [22].

4.1 The Performance of the Proposed Adaptive Spatial Filter

It is trivial that a good performing spatial filtering is necessary to get a denoised image of good quality using the proposed approach. Otherwise, the image characteristics will be lost. In this section, the NLOOK parameter estimation previously developed is exploited and compared to the results obtained with the optimal NLOOK value determined manually. Simulations were performed using different standard images from [38] and [39], corrupted by speckle noise of variance ranging from 2100 to 4100. Since many of the samples lead to an approximately same percentage (*Pe*) of edge pixels, and therefore to a similar algorithm behavior, table 1 summarizes sample results obtained for 8 different images having different *Pe* values.

It can be observed that, except for *Pe* < 2 (e.g. Lady image), both the logarithmic and polynomial estimations result in near optimal values for the adjustment parameter, and the PSNR loss does not exceed 0.33 dB and 0.16 dB with logarithmic and polynomial estimations, respectively. In fact, the behavior of the filter when dealing with few-detailed images (Lady image in table 1) is expected, since both proposed approximations fail to be accurate in cases of very low *Pe* values as shown in *Figure 4*. However, images with a percentage of edge pixels less than 2 rarely exist (in our test datasets from [38] and [39], only 11% of the images had *Pe* < 2). Furthermore, the approximation error can be largely reduced by setting A_d to a constant value (approximately 0.43) for such low values of *Pe* as implies *Figure 4*.

It is important to mention that experimental results with such high noise variances are shown to highlight the decent behavior of the proposed method in critical conditions. Moreover, when dealing with slightly speckled images, no filtering is performed (W ≈ 1 in (5)), as will be verified in section 4.2.

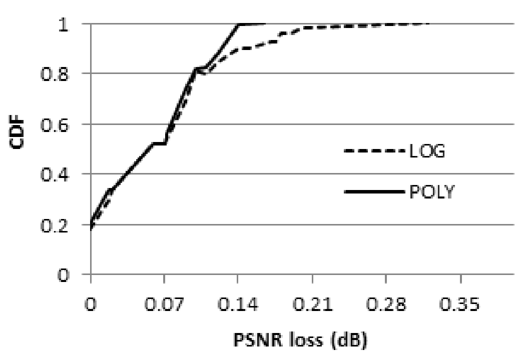
Figure 8-a shows the Cumulative Density Function (CDF) of the PSNR loss obtained with both estimations, using the whole set of test images. It can be noticed that in both methods, 50% of the filtered images undergo a loss in PSNR not exceeding 0.05 dB, whereas 85% and 89% undergo a PSNR loss that does not exceed 0.11 dB in logarithmic and polynomial approximations, respectively.

Even though values of *Pe* are more likely to occur between 2% and 8%, let us analyze the proposed models' behavior for values of *Pe* outside this interval. It can be deduced from *Figure 4* that for low values of *Pe*, the measured A_d seems to saturate at a value close to 0.43, while the logarithmic approximation reaches 0 for values of *Pe* close to 0.9% as shown in *Figure 8-b*, and becomes negative for lower values.

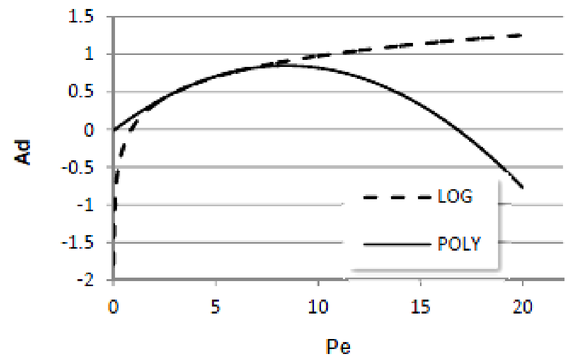
Similarly, the polynomial approximation also fails to model A_d at low values of *Pe*, but results in slightly better performance compared to the logarithmic model. This problem can be solved by using any of the two models only when *Pe* exceeds a fixed threshold (e.g. for *Pe* ≥ 2%), or by setting A_d to a constant value (e.g. A_d = 0.43) when *Pe* is below that threshold. On the other hand, the polynomial approximation reaches a peak for *Pe* ≈ 8.5%, and then decreases as *Pe* increases. Therefore, the polynomial model becomes inaccurate in this case. The same does not occur with the logarithmic approximation since it keeps increasing as *Pe* increases.

Noise variance	Image	Pe (%)	Speckled Image PSNR	Log approximation			Polynomial approximation			Optimal values		
				100xA	NLK_AD	PSNR	100xA	NLK_AD	PSNR	100xA	NLOOK	PSNR
2100	Lady	0.9719	26.1012	4.2606	12.1041	24.276	17.393	13.6457	25.695	42	16.5228	28.246
	WomanDarkHair	2.6634	20.6058	44.7361	30.1964	29.115	44.617	30.1987	29.152	44.617	30.1987	29.152
	I06	3.1488	20.5841	51.2811	28.3623	27.569	52.98	28.2943	27.963	53.321	28.5143	27.988
	Lena	4.4518	20.7322	65.3613	28.6966	27.766	65.837	28.8355	27.765	65.3613	28.6966	27.766
	Board	6.0764	19.6519	77.8522	26.12	24.224	78.376	26.349	24.237	78.376	26.349	24.237
	I13	6.2301	21.5436	79.2398	24.9994	26.599	73.987	24.9655	26.541	78.595	24.4561	26.632
	LivingRoom	7.0168	20.9285	83	28.1954	25.992	82.703	28.0957	26.012	83.63	28.4462	26.028
	Mandrill	11.597	20.5485	103.801	30.417	24.865	73.053	25.8463	24.713	103.801	30.417	24.865
2600	Lady	0.9887	25.2349	2.3487	10.611	24.451	15.098	11.9622	25.912	43	14.8931	28.838
	WomanDarkHair	3.0743	20.8759	47.8966	26.4071	28.423	47.58	26.3707	28.435	44	25.7209	28.493
	I06	3.3248	19.6843	52.0432	23.9424	27.107	50.524	23.5761	27.132	50.819	23.5999	27.281
	Lena	4.5025	19.855	63.216	24.2587	27.195	65	24.5917	27.199	63.726	24.4509	27.2
	Board	6.0455	18.7322	73	22.0966	23.658	75.598	22.5739	23.668	75.048	22.4301	23.698
	I13	6.2467	20.5233	74.9848	22.1354	26.260	75.935	22.2013	26.191	75.213	22.5847	26.261
	LivingRoom	6.5968	20.041	78.5514	23.8295	25.51	82	24.4238	25.512	78.436	23.787	25.57
	Mandrill	11.382	19.6209	100.451	26.1791	24.308	72.045	22.5894	24.322	83	23.8904	24.365
3100	Lady	0.9901	24.5231	-1.1945	9.3747	24.321	11.523	10.5359	25.819	40	13.2466	29.024
	WomanDarkHair	3.027	20.177	43.674	22.3196	27.926	43.361	22.2835	27.985	41	21.8922	28.119
	I06	3.5269	18.9928	45.997	19.767	26.501	45.624	19.5671	26.663	51.407	20.223	26.614
	Lena	4.4182	19.1324	58.8572	20.8946	26.767	62	21.2741	26.724	59.31	20.8339	26.785
	Board	6.339	18.0201	73.3508	19.8109	23.235	69	19.2277	23.228	73.601	19.7984	23.266
	I13	6.356	19.8108	75.473	19.4415	25.622	74.423	19.4445	25.992	74.337	19.5426	25.874
	LivingRoom	7.2327	19.2983	78.6463	21.3184	25.078	77.193	21.1876	25.087	76	20.8615	25.121
	Mandrill	12.095	18.9053	99.2916	23.3454	23.803	62.725	19.2675	23.891	79	21.0759	23.956
3600	Lady	0.9953	23.9122	-5.5841	8.2428	24.029	7.0173	9.2931	25.582	36	11.8257	28.94
	WomanDarkHair	3.1719	19.5572	40.9514	19.5081	27.608	40.641	19.6611	27.589	37	18.9953	27.672
	I06	3.6442	18.3526	43.3436	17.2142	26.201	43.221	16.8099	26.204	49.012	17.5012	26.254
	Lena	4.5715	18.5216	55.6266	18.2715	26.398	56.18	18.363	26.464	56.18	18.363	26.464
	Board	6.3707	17.4144	68.9511	17.4682	22.829	69.161	17.6717	22.837	65	17.0354	22.845
	I13	6.4014	19.0212	70.3316	17.9968	25.596	70.44	17.2634	25.526	72.256	17.1994	25.598
	LivingRoom	7.1453	18.7005	73.5582	18.6148	24.753	72.328	18.5073	24.736	72	18.5868	24.797
	Mandrill	11.821	18.2575	93.7692	20.7329	23.4	60.474	17.1004	23.621	73.55	18.5456	23.6257
4100	Lady	0.9971	23.4004	11.1116	7.1373	23.456	1.4499	8.1364	25.193	35	10.8747	29.183
	WomanDarkHair	3.2227	19.0674	35.9893	16.9639	27.374	35.689	16.9408	27.299	35.9893	16.9639	27.374
	I06	3.618	17.7216	38.2526	14.7722	25.921	38.221	14.7677	24.999	44.098	15.2344	25.923
	Lena	4.7283	17.9916	51.3807	16.1424	26.11	52.024	16.1452	26.118	52.024	16.1452	26.118
	Board	6.3188	16.8767	63.0227	15.4386	22.472	63.297	15.4994	22.455	57	14.8499	22.531
	I13	6.3904	18.7049	64.2214	15.9697	25.455	64.031	15.0996	25.402	60.215	15.0027	25.491
	LivingRoom	7.9094	18.1322	72.0373	16.8162	24.427	68.422	16.5037	24.457	66	16.2858	24.465
	Mandrill	12.867	17.733	91.5761	18.8708	23.045	44.941	14.2858	23.313	67	16.3813	23.372

TABLE 1: Summary of results obtained with different 512×512 speckled images using the proposed adaptive spatial filter.



(a)



(b)

FIGURE 8: (a) CDF of the loss in PSNR due to logarithmic (dotted line) and polynomial (solid line) approximations. (b) Logarithmic (dashed line) and polynomial (solid line) approximations as defined in (11) and (12).

Table 2, 3, 4 and 5 show a comparison of the Lee filter, standard (NLOOK=ENL) Kuan filter, Gamma filter, Frost filter, Enhanced Frost filter, and the proposed enhanced Kuan filter using both the logarithmic and polynomial approximations with a speckle noise of Gaussian, Uniform, Rayleigh and Exponential distributions respectively.

The same is performed for different images of different edge percentage and the results were approximately similar. It can be clearly seen that the performance of the filters is affected by the noise distribution. For the examples shown in Tables 2-5, the Lee multiplicative filter performs better than the Enhanced Frost filter with Uniformly distributed noise, and worse with Gaussian noise, the Kuan filter gives better results than the Lee multiplicative filter except for the case with NLOOK=ENL (where this latter parameter is not adjusted). Depending on the noise distribution, either of the approximations (logarithmic or polynomial) can outperform the other in Kuan filtering with adjusted NLOOK parameter.

In tables 2-5, it can also be observed that the best PSNR is obtained with either the Gamma MAP filter or the proposed (modified) Kuan filter, whereas the Frost and enhanced Frost filters preserve image features more than other filters as it can be noticed from CoC values.

	PSNR	**S/MSE**	**CoC**	**ENL**
Gaussian noise				
Noisy image	19.2482	13.5918	0.1781	13.3615
Lee multiplicative filtered image	26.9749	21.3185	0.3028	98.0283
Kuan filtered image(NLOOK=ENL)	22.8675	17.2111	0.1155	31.6124
K.filtered image (Log. approximation)	27.2248	21.5685	0.3106	106.6198
K.filtered image (Poly. approximation)	27.2389	21.5826	0.3154	110.6581
Gamma filtered image	28.1836	22.5272	0.3892	233.4135
Frost filtered image	26.3968	20.7404	0.4225	233.3297
Enhanced Frost filtered image	27.0823	21.4259	0.4394	240.8940

TABLE 2: Comparison of speckle noise filtering techniques on a 512×512 Lena image corrupted by a Gaussian speckle noise using an 11×11 window size.

	PSNR	S/MSE	CoC	ENL
Uniform noise				
Noisy image	19.1606	13.5042	0.1760	13.1826
Lee multiplicative filtered image	27.5563	21.9000	0.3348	131.8801
Kuan filtered image(NLOOK=ENL)	23.0604	17.4040	0.1210	33.0601
K.filtered image (Log. approximation)	27.7578	22.1015	0.3420	144.8634
K.filtered image (Poly. approximation)	27.7968	22.1405	0.3403	145.0305
Gamma filtered image	27.3011	22.0448	0.3034	127.2049
Frost filtered image	26.4070	20.7506	0.4214	135.5507
Enhanced Frost filtered image	27.0527	21.3963	0.4454	127.7399

TABLE 3: Comparison of speckle noise filtering techniques on a 512×512 Lena image corrupted by a Uniform speckle noise using an 11×11 window size.

	PSNR	S/MSE	CoC	ENL
Rayleigh noise				
Noisy image	19.4281	13.7718	0.1760	13.7621
Lee multiplicative filtered image	27.1461	21.4898	0.3158	106.4877
Kuan filtered image(NLOOK=ENL)	23.1365	17.4802	0.1256	34.0457
K.filtered image (Log. approximation)	27.3290	21.6726	0.3200	117.1893
K.filtered image (Poly. approximation)	27.3523	21.6960	0.3236	116.7191
Gamma filtered image	28.2304	22.5741	0.3979	230.0371
Frost filtered image	26.3868	20.7305	0.4231	240.8726
Enhanced Frost filtered image	27.0836	21.4272	0.4576	234.4497

TABLE 4: Comparison of speckle noise filtering techniques on a 512×512 Lena image corrupted by a Rayleigh speckle noise using an 11×11 window size.

	PSNR	S/MSE	CoC	ENL
Exponential noise				
Noisy image	20.0551	14.3987	0.1746	15.3187
Lee multiplicative filtered image	25.9639	20.3076	0.2597	67.2103
Kuan filtered image(NLOOK=ENL)	23.2500	17.5936	0.1368	34.4915
K.filtered image (Log. approximation)	27.2178	21.5614	0.3689	173.9484
K.filtered image (Poly. approximation)	27.1044	21.4480	0.3684	169.5444
Gamma filtered image	26.9633	21.2070	0.3179	101.6422
Frost filtered image	26.3943	20.7380	0.4299	238.9131
Enhanced Frost filtered image	27.0939	21.0376	0.4584	239.3688

TABLE 5: Comparison of speckle noise filtering techniques on a 512×512 Lena image corrupted by an Exponential speckle using an 11×11 window size.

Table 6 shows a comparison between the proposed adaptive Kuan filter and the basic and enhanced versions of Lee, Kuan, and Frost filters using five different images. It can be noticed that the optimal Kuan filter yields the highest PSNR, with the drawback of performing several filter runs in order to tune the filter to yield the best possible output quality. Furthermore, by setting the NLOOK parameter to ENL, Lee and Frost filters outperform the Kuan filter in their basic and enhanced versions. Moreover, our proposed solution approaches the performance of the optimal Kuan filter, with a negligible PSNR loss, and outperforms the basic and enhanced Lee and Frost filters, for all the speckled test images used in the filtering process.

PSNR	WomanD.H.	Pirate	Lena	CameraMan	LivingRoom
Noisy image	19.55	20.02	19.42	20.1	19.9
Lee	25.39	25.91	27.14	26.01	26.2
Frost	23.69	25.08	26.38	25.9	25.89
Enhanced Lee	25.61	26.01	27.30	26.78	26.88
Enhanced Frost	25.47	25.60	27.08	26.12	26.18
Kuan	23.34	24.21	24.02	24.92	25.1
Kuan (poly. approximation)	25.78	26.57	27.35	26.95	27.21
Kuan (log. approximation)	25.88	26.59	27.33	26.89	27.1
Kuan (optimal NLOOK)	25.92	26.70	27.44	27.11	27.35

TABLE 6: PSNR obtained with different despeckling filters using images of 512×512 size and an 11×11 filtering window.

It is important to note that in the literature [13, 14, 37], the Block Matching and 3D filtering (BM3D) algorithm is tested on images corrupted by additive noise. Moreover, the method's different parameters should be adequately chosen (depending on the noise strength, image size and percentage of edge pixels etc) in order to perform well dealing with speckle noise. Those parameters are the reference block size, the searching window size, the number of similar patches in the 3D block and the shrinking threshold. Therefore, finding the optimal set of parameters dealing with speckle noise is a very difficult task and choosing a non-optimal set of parameters usually fails to outperform traditional despeckling filters. On the contrary, the proposed adaptive filer can automatically estimate the suitable ENL parameter which will be used to spatially smooth the approximation image in the wavelet domain without need for many filter runs with different parameters.

Figure 9 shows a real spine lumbar MRI image obtained from [42], filtered with Enhanced Frost, Enhanced Lee, and the proposed adaptive Kuan filter using the polynomial approximation. A visual inspection of the results shows that all the filters succeed to reduce speckle noise and result in a pleasant visual appearance, preserving edges and contours. However, compared to our proposed solution, the enhanced Lee and Frost filters show over-blurred images.

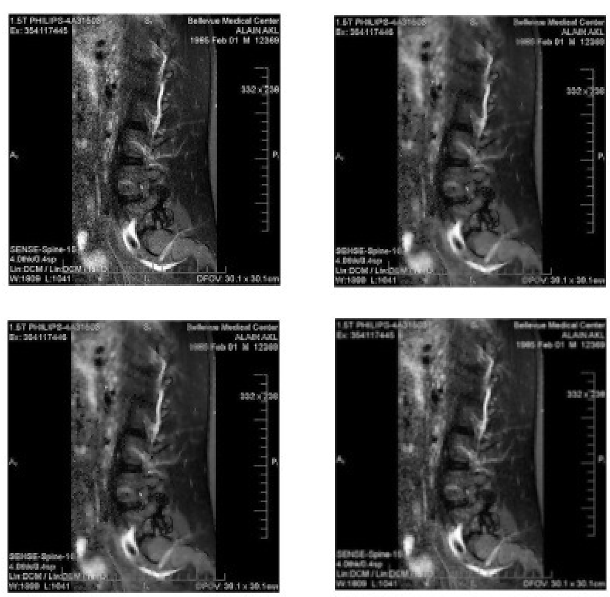


FIGURE 9: Spine lumbar MRI image [42]. Top left: original (noisy) image. Top right: Enhanced Frost filtered image. Bottom left: Enhanced Lee filtered image. Bottom right: result of filtering with our proposed solution using the polynomial approximation. A 13×13 window size is used for all the results.

Figure 10 shows a real cardiac catheterization speckled image obtained from [43], filtered with the enhanced versions of Frost and Lee and with the proposed Kuan filter (polynomial approximation). It can be observed that all the filters succeed to reduce speckle noise. However, the proposed adaptive Kuan filter outperforms the enhanced Frost and enhanced Lee filters in edge preservation. In other words, the Kuan filter with the proposed polynomial approximation yields visually better performance with the resulting image efficiently smoothed, while significantly preserving image details.

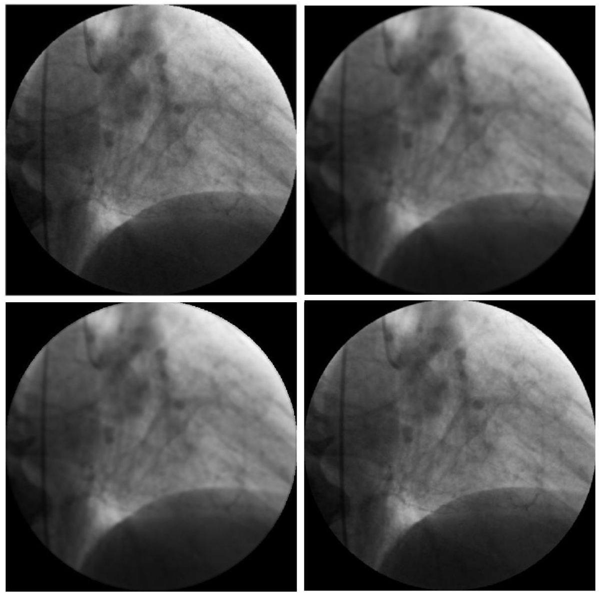


FIGURE 10: Cardiac catheterization image [43]. Top left: original (noisy) image. Top right: Enhanced Frost filtered image. Bottom left: Enhanced Lee filtered image. Bottom right: result of filtering with our proposed solution using the polynomial approximation. A 13×13 window size is used for all the results.

4.2 Performance of the Hybrid Wavelet-Spatial Despeckling Filter

This section deals with implementing the proposed hybrid wavelet-spatial domain filer. First, wavelet denoising with one decomposition level is applied on the 25 images of TID2013 database [39], each corrupted by 5 different levels of Gaussian speckle noise, then the results are compared to the same algorithm but using our enhanced Kuan filter to smooth the approximation image before wavelet reconstruction. We also apply wavelet denoising with the number of decomposition levels determined dynamically as proposed earlier, and compare the results to the same algorithm but using our enhanced Kuan filter to smooth the approximation image before wavelet reconstruction. Finally, the proposed enhanced Kuan filter is applied to the approximation image in each reconstruction level. Table 7 presents the gains (average, minimum, maximum, and standard deviation) with respect to the noisy image using different evaluation metrics (PSNR, CoC and ENL). It can be noticed that a Kuan smoothing on the approximation image with one decomposition level leads to an increase of 6.12 dB in average PSNR compared to 4.88 dB without the approximation image smoothing.

It is important to note that Kuan smoothing on the last decomposition level does not result in any significant enhancement. In fact, the smoothing strength of our adaptive Kuan filter depends on the NLK_AD parameter determined dynamically, which is very high in the approximation image of the last level, due to repetitive smoothing at every decomposition. Therefore, $W \approx 1$ (in (5)) and no filtering is performed in this case. Moreover, a Kuan smoothing filter on the approximation image of each level before reconstruction, leads to an increase of 0.2 dB in average PSNR compared to the case where approximation smoothing is performed only at the last decomposition level, without blurring the denoised image, which is obviously seen from the CoC metric.

Gaussian noise	Gains with respect to the noisy image											
	PSNR				CoC				ENL			
	AVG	MIN	MAX	STD	AVG	MIN	MAX	STD	AVG	MIN	MAX	STD
1L without Kuan smoothing	4.8808	3.0946	5.996	0.6681	0.0863	0.0143	0.0999	0.0121	21.2313	18.224	24.956	1.6711
1L with Kuan smoothing	6.1168	4.8226	6.9996	0.5514	0.0836	0.014	0.1089	0.013715	41.3668	29.2124	49.224	2.8868
Dynamic without Kuan smoothing	7.1225	5.826	7.9986	0.5538	0.1142	0.0543	0.971	0.079	60.0037	30.4054	65.2006	4.4478
Dynamic with Kuan smoothing on the last level	7.1228	5.8206	7.9991	0.5534	0.1153	0.0599	0.9174	0.0744	60.0042	30.4023	65.2505	4.4484
Dynamic with Kuan smoothing on all reconstructed levels	7.3229	5.9146	8.2906	0.556	0.1132	0.0443	0.9121	0.0742	64.7011	40.4054	72.3504	4.5485

TABLE 7: Evaluation metrics for wavelet-based denoising with and without Kuan smoothing. Sym1 wavelets and BayesShrink threshold selection technique are used.

4.3 Performance Evaluation for the Proposed Complexity Reduction Algorithm

In this section, the proposed accelerated algorithm is tested with different scenarios. *Figure 11* shows an example of the denoising results obtained using the Lena image corrupted by a Gaussian speckle of variance 3100, Haar, db4, sym4, and bior6.8 wavelets, Soft and Hard VisuShrink and BayesShrink.

The algorithm results in two wavelet decomposition levels. The first and second columns of *Figure 11* show the images obtained with one and two decomposition levels respectively. It can be seen that two decomposition levels give better performance than one decomposition level as expected. In addition, Universal Soft and Hard thresholding give more pleasant results than BayesShrink.

The same work is done with different standard test images, different noise variances and distributions, different thresholding methods, different wavelet types and families and using different decomposition levels, with images from [38, 39]. From our intensive test scenarios, we observed that a “best” denoising setup using wavelets cannot be generalized. In other words, there is no one optimal threshold selection technique and wavelet type that always give the best performance; performance depends on the wavelet thresholding algorithm, the wavelet family, the noise variance, the number of decomposition levels and the noisy image itself.

In the purpose of studying the complexity gain and PSNR loss when we stop at the estimated number of decomposition levels denoted *L_opt_algo*, we apply 10^4 times wavelet denoising on the images from TID2013 database [39] corrupted by five different levels of multiplicative Gaussian speckle noise, using BayesShrink threshold selection, ‘sym1’ wavelet type and different number of decomposition levels.

Figure 12 shows the histogram of the difference between the number of decomposition levels that yields the highest PSNR (i.e. to which we refer as “optimal”) and the number *L_opt_algo* obtained with our proposed method. It can be observed that 49% of the time, our algorithm successfully determines optimal number of levels whereas it performs 1 additional level for 11% of the time, and one and two levels less than the optimal number for 23% and 17% of the time, respectively. It is important to mention that these numbers could vary with a different dataset or simulation setup, but they give a general overview of the behavior of the proposed algorithm. As a result of erroneously estimating the optimal number of decomposition levels, some PSNR loss is incurred.

Figures 13 and *14* respectively show the histogram and the cumulative density function (CDF) of the PSNR loss due to the proposed algorithm. It can be observed that the loss does not exceed

0.04 dB and thus, can be considered as negligible. Therefore, the proposed acceleration algorithm has a good performance in terms of robustness against PSNR loss.

Figure 13 shows that there exists only 2% of no loss (loss of 0 dB), whereas *Figure 12* indicates that 49% of the images should have no loss, which could seem contradictory; in fact, referring to *Figure 14*, we notice that the PSNR loss is less than 0.01dB for 49% of the images, which can be explained to be due to numerical rounding errors in computer simulations and therefore the results in *Figures 13* and *14* are not contradictory but rather consistent with those in *Figure 12*.

FIGURE 11: Denoising using wavelets. 1st row: original image (left), speckled image (right). 2nd, 3rd and 4th row: db4 wavelet denoised images using the Universal Threshold with soft thresholding, the Universal Threshold with hard thresholding and with BayesShrink. Results obtained with one decomposition level are on the left and those obtained using two decomposition levels are on the right.

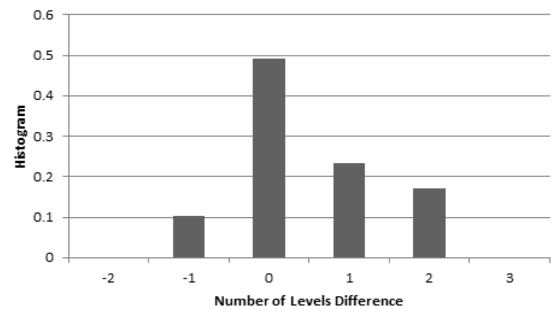


FIGURE 12: Histogram of the difference between *L_opt_algo* and the optimal number of decomposition levels.

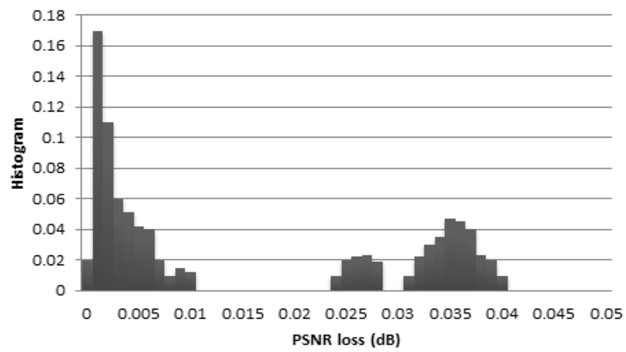


FIGURE 13: Histogram of the PSNR loss when using the acceleration algorithm.

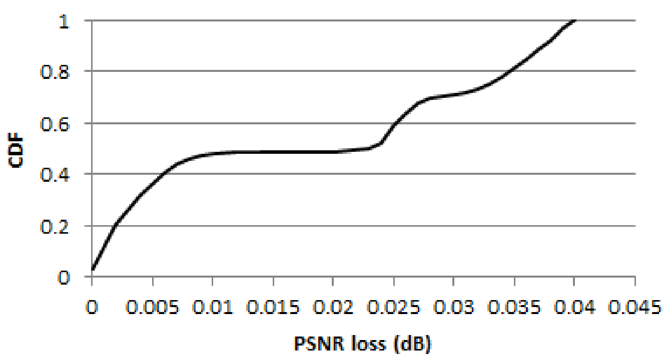


FIGURE 14: CDF of the PSNR loss when using the acceleration algorithm.

The advantage of the acceleration algorithm is the gain in complexity while keeping the PSNR loss negligible. In the purpose of analyzing this gain, the CPU time usage is computed for a number of 'sym1' wavelet decomposition levels followed by BayesShrink thresholding and the corresponding wavelet reconstruction. *Figure 15* shows the processor time usage (average of 100 runs) on a 2 GHz Intel Core 2 Duo CPU, for images from the TID2013 database [39], corrupted by 5 levels of multiplicative Gaussian speckle noise. The figure shows that one could save 0.12 microseconds per pixel by performing one decomposition level instead of two, and 0.10 microseconds per pixel for two levels instead of three. The incremental gain exponentially decreases as the number of decomposition levels increases, which is expected due to the size of the approximation image that is exponentially reduced (by a power of 2 in each dimension) for each additional level.

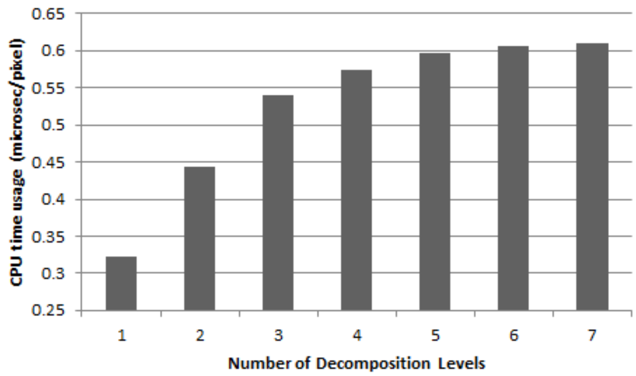


FIGURE 15: CPU time usage according to the number of decomposition levels.

5. CONCLUSION

We have proposed a hybrid wavelet-spatial denoising algorithm based on two-stage processing. An automatic parameter selection technique for the spatial Kuan filter is proposed in the first stage, which is later used to spatially smooth the approximation image of the wavelet coefficients in the second stage. An acceleration algorithm for wavelets computation is also proposed. It consists in selecting a suitable number of wavelet decomposition levels yielding near-optimal performance and eliminating the need for additional and unnecessary decomposition levels.

The procedures have been materialized on a set of different images under different conditions. In order to evaluate the quality of the results, objective metrics were used besides subjective ones identifying the method's capacity in reducing speckle noise. The obtained results quality was highly encouraging, in terms of speckle reduction and image characteristics preservation, and proved that the proposed acceleration algorithm is advantageous in complexity gain which suggests that further research in this direction could be promising, in particular in the optimal choice of the decomposition level where spatial smoothing can be performed on the approximation component.

The evaluation of repetitive runs of the spatial optimized filter is of our interest, to verify whether it allows for better noise suppression or reduction while critically preserving edges and texture. Moreover, it is known that smoothing is usually performed in a pre-processing phase before edge detection. In the proposed algorithm, edge detection was performed for the purpose of smoothing. Therefore, we propose the use of our enhanced Kuan filter recursively for the sake of edge detection. Finally, we aim at reinforcing the use of the proposed algorithm to enhance the performance of the BM3D denoising algorithm, leaving this perspective an open discussion for future considerations.

6. ACKNOWLEDGMENT

This work has been partly supported by a research grant from the Higher Center for Research at the Holy Spirit University of Kaslik (USEK), Lebanon.

7. REFERENCES

[1] A. Isar, S. Moga and D. Isar, "A new denoising system for SONAR images," *ACM Journal on Image and Video Processing*, January 2009.

[2] I. Shanthi and M.L. Valarmathi, "Speckle Noise Suppression of SAR Image using Hybrid Order Statistics Filters," *International Journal of Advanced Engineering Sciences and Technologies*, Vol. 5, No. 2, pp. 229-235, 2011.

[3] D. T. Kuan, "Adaptive Noise Smoothing Filter for Images with Signal Dependent Noise," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-7, No. 2, pp. 165-177, 1985

[4] D.H.H. Santosh, V.N.L. Kumar, D.R. Ramesh and P.S.Prasad, ''Efficient Techniques for Denoising of Speckle and Highly Corrupted Impulse Noise Images,'' *IEEE Trans. Electronics Computer Technology (ICECT)*, Vol. 3, pp. 253-257, 2011.

[5] R. Ahirwar and A. Choubey, "A Novel Wavelet-based Denoising Method of SAR Image Using Interscale Dependency," *2011 International Conference on Computational Intelligence and Communication Systems*, Gwalior, India, October 2011.

[6] J. S. Lee, "Digital image enhancement and noise filtering by using local statistics," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAM1-2, 1980.

[7] J.S. Lee and I. Jurkevich, “Segmentation of SAR images,” *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 27, No. 6, pp. 674-680, November 1989.

[8] J.S. Lee, “Speckle analysis and smoothing of synthetic aperture radar images,” *Computer Graphics and Image Processing*, Vol. 17, pp. 24-32, 1981.

[9] D. Kuan, A. Sawchuk, T. Strand and P. Chavel, “Adaptive restoration of images with speckle,” *IEEE Trans., Acoustic Speech, Signal Processing*, Vol. ASSP-35, No. 3, pp. 373-383, March 1987.

[10] A. Isar, S. Moga and D. Isar, "A new denoising system for SONAR images," *ACM Journal on Image and Video Processing*, January 2009.

[11] O.V. Michailovich and A. Tannenbaum, "Despeckling of medical ultrasound images," *IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control*, pp. 64-78, January 2006.

[12] R. Ahirwar and A. Choubey, "A Novel Wavelet-based Denoising Method of SAR Image Using Interscale Dependency," *2011 International Conference on Computational Intelligence and Communication Systems*, Gwalior, India, October 2011.

[13] DabovKostadin, “Image and video Restoration with nonlocal transform-domain filtering,” *Ph.D. dissertation, Image/video restoration and enhancement*. Finland: Tampere University of technology, 2010.

[14] A. Buades, B. Coll, and J. M. Morel, “A non-local algorithm for image denoising,” *IEEE Trans. Comp. Vision. Pattern Reco.,* Vol.2, pp. 60-65, June 2005.

[15] Gonzales. Rafael C. and Woods. Richard E., “Digital Image Processing,” *Third Edition, Pearson Prentice Hall*, pp. 311-334, 346-356, 463-476, 488-509, 2008.

[16] C. Ju and C. Moloney, “An Edge-Enhanced Modified Lee filter for the Smoothing of SAR Image Speckle Noise,” *IEEE Workshop on Nonlinear Signal Image Processing*, Mackinac Island, MI, 1997.

[17] Himani Goyal, “Qualitative and Quantitative Evaluation of Image Denoising Techniques,” *International Journal of Computer Applications*, Vol. 8, No. 14, pp. 31-34, October 2010.

[18] D.M. Smith, “Speckle reduction and segmentation of Synthetic Aperture Radar images,” *International Journal of Remote Sensing*, Vol. 17, No. 11, pp. 2043-2057, 1996.

[19] L. V. Patil, K. K. Rane, A. P. Salvi, and S. S. Shitole, "Evaluation of Statistical Digital Image Filters," *European Journal of Scientific Research*, June 2011.

[20] S.S. Ganugapati and C.R. Moloney, "A ratio based edge detector for speckled images based on maximum strength edge pruning," *Proceedings of ICIP- 95, Washington, D.C.*, Vol. 2, pp. 165-168, October 1995.

[21] A. Lopes, E. Nezry, R. Touzi and H. Laur., "Maximum a posteriori speckle filtering and first order texture models in SAR images," *Proceeding of Geoscience and Remote Sensing Symposium '90*, College Park, MD, Vol. 3, pp. 2409-2412, May 1990.

[22] Hong Sern Tan, "Denoising of Noise Speckle in Radar Image," *B.E. Thesis*, School of Information Technology and Electrical Engineering, The University of Queensland, October 2001.

[23] L. Klaine, B. Vozel, and K. Chehdi. "Adaptive filtering of multiplicative noise by a new differential method," *Proceeding of the Sixth International Symposium on Signal Processing and Its Applications'* 2003, July 2003.

[24] S. Aybike and S. B. Hasan, "Techniques of speckle noise reduction in ultrasound images," *Proceeding of the 19th Signal Processing and Communications Applications Conference*, pp. 375-378, 2011.

[25] R. Rangarajan, R. Venkataramanan and S. Shah, "Image Denoising Using Wavelets: Wavelets & time frequency," *Report*, December 16, 2002.
Available at: http://www-personal.engin.umich.edu/~rvenkata/551_code/Report.pdf

[26] Sani M. Isa, Ary Noviyanto and Aniati Murni Arymurthy, "Optimal Selection of Wavelet Thresholding Algorithm for ECG Signal Denoising," *IEEE Trans. on Advanced Computer Science and Information System*, pp. 365-370, December 2011.

[27] I. Daubechies, "Ten lectures on wavelets," *The Conference Board of the Mathematical Sciences*, *Series in Applied Mathematics*, Vol. 61, pp. 194-202, 1994.

[28] I. Daubechies, "Ten lectures on wavelets," *The Conference Board of the Mathematical Sciences, Series in Applied Mathematics*, Vol. 61, pp. 198-202 and 254-256, 1994.

[29] I. Daubechies, "Ten lectures on wavelets," *The Conference Board of the Mathematical Sciences, Series in Applied Mathematics*, Vol. 61, pp. 271-280.

[30] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans., Pattern Anal. Machine Intell.,* Vol. 11, pp. 674–693, July 1989.

[31] S. Grace Chang, Bin Yu and Martin Vetterli, "Adaptive Wavelet Thresholding for Image Denoising and Compression," *IEEE Trans on image processing*, Vol. 9, pp. 1532-1546, September 2000.

[32] David L. Donoho, "De-Noising by Soft-Thresholding," *IEEE Trans. on information theory*, Vol. 41, No. 3, pp. 613-627, May 1995.

[33] Donoho, D.L., Johnstone, "Spatial adaptation by wavelet shrinkage," *Biometrika, 81*, pp. 425-455, 1994.

[34] Shi Guangmin and Li Fudong, "Image Denoising with Optimized Subband Threshold," *Proceeding of the Fifth International Conference on Computational Intelligence and Multimedia Applications*, 2003.

[35] Yunyi Yan, Baolong Guo, Zhanlong Yang and Xiang Fu, “Image Noise Removal via Wavelet Transform and r/K-PSO,” *Fourth International Conference on Natural Computation*, pp. 544-547, 2008.

[36] Aleksandra Pizurica and Wilfried Philips, “Estimating the Probability of the Presence of a Signal of Interest in Multiresolution Single- and Multiband Image Denoising,” *IEEE Trans. on image processing*, Vol. 15, No. 3, March 2006.

[37] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3D transform-domain collaborative filtering,” *IEEE Trans. Image Process.*, Vol. 16(8), pp. 2080-2095, August 2007.

[38] ImageProcessingPlace.com, Image Databases.

[39] TAMPERE IMAGE DATABASE 2013 TID2013, version 1.0.

[40] M. Misiti, Y. Misiti, G. Oppenheim, and J.M. Poggi, Matlab Wavelet Toolbox User’s guide version 3, 2005.

[41] A. Achim, A. Bezerianos and P. Tsakalides, “Ultrasound Image Denoising ViaMaximum A Posteriori Estimation Of Wavelet Coefficients,” *Proceedings of the 23rd annual international conference of the IEEE*, Vol. 3, pp. 2553-2556, 2001.

[42] M.R 12369 R_11_1200, Bellevue Medical Center, Mansourieh, El Metn, Lebanon.

[43] Interventional Cardiology Service, C07-372, N.D. Maritime hospital, Byblos, Lebanon.

A Comparative Study of Content Based Image Retrieval Trends and Approaches

Satish Tunga *satish.tunga@msrit.edu*
Dept. of Telecommunication Engineering, MSRIT, Bengaluru,
& Research Scholar, Jain Univ., Bengaluru, India

D. Jayadevappa *devappa22@gmail.com*
Dept. of Electronics & Inst. Engineering,
JSS Academy of Technical Education, Bengaluru,
VTU, Belgavi, India

C. Gururaj *gururaj.c@gmail.com*
Dept. of Telecommunication Engineering, BMSCE, Bengaluru,
& Research Scholar, Jain Univ., Bengaluru, India

Abstract

Content Based Image Retrieval (CBIR) is an important step in addressing image storage and management problems. Latest image technology improvements along with the Internet growth have led to a huge amount of digital multimedia during the recent decades. Various methods, algorithms and systems have been proposed to solve these problems. Such studies revealed the indexing and retrieval concepts, which have further evolved to Content-Based Image Retrieval. CBIR systems often analyze image content via the so-called low-level features for indexing and retrieval, such as color, texture and shape. In order to achieve significantly higher semantic performance, recent systems seek to combine low-level with high-level features that contain perceptual information for human. Purpose of this review is to identify the set of methods that have been used for CBR and also to discuss some of the key contributions in the current decade related to image retrieval and main challenges involved in the adaptation of existing image retrieval techniques to build useful systems that can handle real-world data. By making use of various CBIR approaches accurate, repeatable, quantitative data must be efficiently extracted in order to improve the retrieval accuracy of content-based image retrieval systems. In this paper, various approaches of CBIR and available algorithms are reviewed. Comparative results of various techniques are presented and their advantages, disadvantages and limitations are discussed. .

Keywords: Content-based Image Retrieval, Semantics, Feature Extraction.

1. INTRODUCTION

Content Based Image Retrieval (CBIR) is a challenging task. Current research works attempt to obtain and use the semantics of image to perform better *retrieval.* Image database management and retrieval has been an active research area since the 1970s [1]. The term Content-based image retrieval was originated in 1992, when it was used by T. Kato to describe experiments into automatic retrieval of images from a database, based on the colors and shapes present. Since then, this term has been used to describe the process of retrieving desired images from a large collection on the basis of syntactical image features. With the rapid increase in computer speed and decrease in memory cost, image databases containing thousands or even millions of images are used in many application areas [2] such as medicine, satellite imaging, and biometric databases, where it is important to maintain a high degree of precision. With the growth in the number of images, manual annotation becomes infeasible both time and cost-wise. Content-based image retrieval (CBIR) is a powerful tool since it searches the image database by utilizing visual cues alone. CBIR systems extract features from the raw images themselves and calculate an association measure (similarity or dissimilarity) between a query image and database images based on these features. CBIR is becoming very popular because of the high demand for searching image databases of ever-growing size. Since speed and precision are important, we need to develop a system for retrieving images that is both efficient and effective.

Image retrieval has been an extremely active research area over the last 10 years, but first review articles on access methods in image databases appeared already in the early 80s [3]. CBIR systems retrieve images from that database which are similar to the query image. Primarily research in Content Based Image Retrieval has always focused on systems utilizing color and texture features [4]. There has also been some work done using some local color and texture features. These account for Region Based Image Retrieval (RBIR) [5]. There are three important feature components for content based image retrieval [6]. The most common are color [7, 8], texture [9, 10]and shape [11, 12] or combinations of these. These features are combined to achieve higher retrieval efficiency [8]. In content-based image retrieval (CBIR), the image databases are indexed with descriptors derived from the visual content of the images. Most of the CBIR systems are concerned with approximate queries where the aim is to find images visually similar to a specified target image. In most cases the aim of CBIR systems is to replicate human perception of image similarity [13].The outline of the content based image retrieval system is shown in Figure 1.

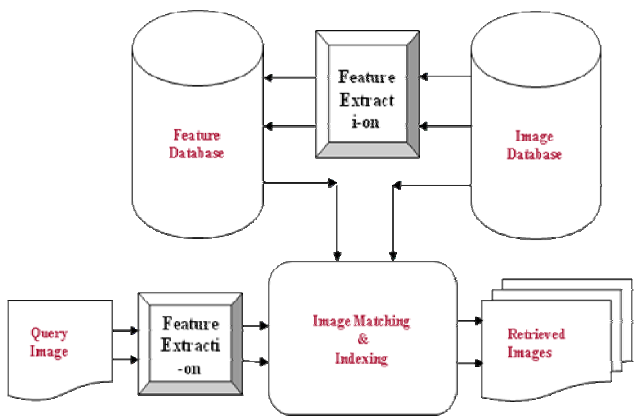


FIGURE 1: CBIR System and Its Various Components.

The process of CBIR consists of the following stages:

Image acquisition: It is the process of acquiring a digital image.

Image Database: It consists of the collection of n number of images depends on the user range and choice.

Image preprocessing: To improve the image in ways that increases the chances for success of the other processes. The image is first processed in order to extract the features, which describe its contents. The processing involves filtering, normalization, segmentation, and object identification. Like, image segmentation is the process of dividing an image into multiple parts. The output of this stage is a set of significant regions and objects.

Feature Extraction: Features such as shape, texture, color, etc. are used to describe the content of the image. The features further can be classified as low-level and high-level features. In this step visual information is extracts from the image and saves them as features vectors in a features database .For each pixel, the image description is found in the form of feature value (or a set of value called a feature vector) by using the feature extraction .These feature vectors are used to compare the query with the other images and retrieval.

Similarity Matching: The information about each image is stored in its feature vectors for computation process and these feature vectors are matched with the feature vectors of query image (the image to be search in the image database whether the same image is present or not or how many are similar kind images are exist or not) which helps in measuring the similarity. This step involves the matching of the above stated features to yield a result that is visually similar with the use of similarity measure method called as Distance method. Here is different distances method available such as Euclidean distance, City Block Distance, Canberra Distance.

Retrieved images: It searches the previously maintained information to find the matched images from database. The output will be the similar images having same or very closest features [14] as that of the query image.

User interface: This governs the display of the outcomes, their ranking, and the type of user interaction with possibility of refining the search through some automatic or manual preferences scheme [15].

1.1. *Challenges of CBIR Systems*

There could be many challenges faced by a CBIR system such as:

- The issue related to the Semantic gap where it means the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation. The user wants to seek semantic similarity, but the database can only provide similarity by data processing.
- The expectation of users for huge amount of objects to search among.
- Sometimes incompleteness query specification seems to be a challenge.
- Incomplete image description is also a source of challenge to an efficient CBIR system.

1.2. Search Options for CBIR Systems

Since the early 1990s, content-based image retrieval has become a very active research area. Both commercial and research image retrieval systems, have been built. Most image retrieval systems support one or more of the following options:

- Random browsing
- Search by example
- Search by sketch
- Search by text (including key word or speech)
- Navigation with customized image categories.

Today, there is the provision of a rich set of search options, but in practical applications which involves actual users still need systematic studies to explore the trade-offs among the different options mentioned above. Here, we will select a few representative systems and highlight their distinct characteristics. [15]

1.3. Real World Requirements

Building real-world systems involve regular user feedback during the development process, as required in any other software development life cycle. Not many image retrieval systems are deployed for public usage, save for Google Images or Yahoo! Images (which are based primarily on surrounding meta-data rather than content). There are, however, a number of propositions for real-world implementation. For brevity of space we are unable to discuss them in details, but it is interesting to note that CBIR has been applied to fields as diverse as Botany, Astronomy, Mineralogy, and Remote sensing [16, 17, 18, 19]. With so much interest in the field at the moment, there is a good chance that CBIR based real-world systems will diversify and expand further. Implementation of an IRM-based [20] publicly available similarity search tool on an on-line database of over 800,000 airline-related images [21] etc. Screen-shots can be seen in Fig. 2 and Fig. 3 respectively Based on our experience with implementing CBIR systems on real- world data for public usage, we list here some of the issues that we found to be critical for real-world deployment.

Performance: The most critical issue is the quality of retrieval and how relevant it is to the domain-specific user community. Most of the current effort is concentrated on improving performance in terms of their precision and recall.
Semantic learning: To tackle the problem of semantic gap faced by CBIR, learning image semantics from training data and developing retrieval mechanisms to efficiently leverage semantic estimation are important directions.
Volume of Data: Public image databases tend to grow into unwieldy proportions. The software system must be able to efficiently handle indexing and retrieval at such scale.
Heterogeneity: If the images originate from diverse sources, parameters such as quality, resolution and color depth are likely to vary. This in turn causes variations in color and texture features extracted. The systems can be made more robust by suitably tackling these variations.
Concurrent Usage: In on-line image retrieval systems, it is likely to have multiple concurrent users. While most systems have high resource requirements for feature extraction, indexing etc., they must be efficiently designed so as not to exhaust the host server resources. Alternatively, a large amount of resources must be allocated.
Multi-modal features: The presence of reliable meta- data such as audio or text captions associated with the images can help understand the image content better, and hence leverage the retrieval performance. On the other hand, ambiguous captions such as “wood” may actually add to the confusion, in which case the multi-modal features together may be able to resolve the ambiguity.
User-interface: As discussed before, a greater effort is needed to design intuitive interfaces for image retrieval such that people are actually able to use the tool to their benefit.
Operating Speed: Time is critical in on-line systems as the response time needs to be low for good interactivity. Implementation should ideally be done using efficient algorithms, especially for large databases. For computationally complex tasks, off-line processing and caching the results in parts is one possible way out. System Evaluation: Like any other software system, image retrieval systems are also required to be evaluated to test the feasibility of investing in a new version or a different product. The design of a CBIR benchmark requires careful design in order to capture the inherent subjectivity in image retrieval. One such proposal can be found in [22].

The machine learning algorithm predicts the category of the query image which is nothing but the semantic concept of the query image. Hence instead of finding similarity between the query image and all the images in database, it is found between the query image and only the images belonging to the query image category. Also when the entire database is searched, the retrieval result contains images of various categories.

1.4. Existing CBIR Systems

Some of the existing CBIR systems [16] are as follows:

1. QBIC or Query by Image Content It is the first commercial content based retrieval system. This system allows users to graphically pose and refine queries based on multiple visual properties such as color, texture and shape. It supports queries based on input images, user-constructed sketches, and selected colour and texture patterns.
2. VisualSEEK and WebSEEK VisualSEEk is visual feature search engine and WebSEEk is a World Wide Web oriented text/image search engine, both of which are developed at Columbia University.
3. Virage Virage is content based image search engine developed at Virage Inc.It supports color and spatial location matching as well as texture matching.
4. NeTra This system uses color, shape, spatial layout and texture matching, as well as image segmentation.
5. MARS or Multimedia Analysis and Retrieval System, this system makes use of colour, spatial layout, texture and shape matching.
6. Viper or Visual Information Processing for Enhanced Retrieval .This system retrieves images based on color and texture matching.
7. The img (Anaktisi) is a CBIR system on the web based on various descriptors which includes powerful color and texture features. The img (Anaktisi) provides different ways to search and retrieve them.

FIGURE 2: (a) Searching Image on COIL-20 Database (b) Searching Aircraft Images on Airlines.net database

2. CURRENT CBIR TECHNIQUES

Existing general-purpose CBIR systems roughly fall into three categories depending on the approach to extract features i.e., histogram, color layout, and region-based search. There are also systems that combine retrieval results from individual algorithms by a weighted sum matching metric [23], or other merging schemes [24].

2.1. Global Feature Based CBIR Systems

Some of the existing CBIR systems extract features from the whole image not from certain regions in it; these features are referred to as Global features. Histogram search algorithms [25] characterize an image by its color distribution or histogram. Many distances have been used to define the similarity of two color histogram representations. Euclidean distance and its variations are the most commonly used. The drawback of a global histogram representation is that information about object location, shape and texture is discarded. Color histogram search is sensitive to intensity variations, color distortions, and cropping. Color histogram search is sensitive to intensity variations, color distortions, and cropping. In simple color layout indexing [25], images are partitioned into blocks and the average color of each block is stored. Thus, the color layout is essentially a low resolution representation of the original image. However, as with pixel representation, although information such as shape is preserved in the color layout representation, the retrieval system cannot perceive it directly. Color layout search is sensitive to shifting, cropping, scaling, and rotation because images are described by a set of local properties [24]. Image retrieval using color features often gives disappointing results, because in many cases, images with similar colors do not have similar content. This is due to the fact that global color features often fails to capture color distributions or textures within the image.

D. Zhang [26] proposed a method combining both color and texture features to improve retrieval performance. By computing both the color and texture features from the images, the database images are indexed using both types of features. During the retrieval process, given a query image, images in the database are firstly ranked using color and features. Then, in a second step, a number of top ranked images are selected and re-ranked according to their texture features. Two alternatives are provided to the user, one is the retrieval based on color features, and the other is retrieval based on combined features. When the retrieval based on color fails, the user will use the other alternative which is the combined retrieval. Since the texture features are extracted globally from the image.

2.2. Region Based CBIR Systems

Region-based retrieval systems attempt to overcome the deficiencies of global feature based search by representing images at the object-level. A region-based retrieval system applies image segmentation to decompose an image into regions, which correspond to objects if the decomposition is ideal [27]. The object-level representation is intended to be close to the perception of the human visual system (HVS). Since the retrieval system has identified what objects are in the image, it is easier for the system to recognize similar objects at different locations and with different orientations and sizes. Region- based retrieval systems include the Natra system [28], and the Blobworld system [29].

The Natra and the Blobworld systems compare images based on individual regions. The motivation is to shift part of the comparison task to the users. To query an image, a user is provided with the segmented regions of the image and is required to select the regions to be matched and also attributes, e.g., color and texture, of the regions to be used for evaluating similarity. Such querying systems provide more control to the user. However, the user's semantic understanding of an image is at a higher level than the region representation.

Natsev et al. considered the similarity model WALRUS [30], which is a robust model for scaling and translation of objects within an image. Each image is first decomposed into regions. The similarity measure between two images is then defined as the fraction of the area of the two images covered by matching regions. However, WALRUS focuses on the development of a fast and effective segmentation method instead of an image-to-image similarity measure. Consequently, region matching should be necessary before image matching. The authors proposed a greedy heuristic for computing the similar region pair set with the maximum area. The basic idea is to iteratively choose the best pair of matching regions that maximizes the area covered by the regions. In [31], the mean shift algorithm is used for segmentation of images and interested regions are indexed using cluster-based tree to increase the efficiency of the retrieval process. However, this system uses only color as image signature, which is sensitive to shifting, cropping, scaling, and rotation. Region based image retrieval [32] uses low-level features including color, texture, and edge density. For color, the histogram of image regions are computed, for texture co- occurrence matrix based entropy, energy, etc, are calculated, and for edge density it is Edge Histogram Descriptor (EHD) that is found. To decrease the retrieval time of images, an idea is developed based on greedy strategy to reduce the computational complexity.

Li and Wang et al [33], proposed the Integrated Region Matching (IRM) algorithm, which allows matching a region of one image to several regions of another image to measure the similarity between images i.e. the region mapping between any two images is a many-to-many relationship. As a result, the similarity between two images is defined as the weighted sum of distances in the feature space, between all regions from different images. Compared with retrieval systems based on individual regions, such as Blobworld, the IRM approach decreases the impact of inaccurate segmentation by smoothing over the imprecision in distances.

Fuzzy Club [34] addresses the issue of effective and efficient content based image retrieval by presenting an indexing and retrieval system that integrates color, texture, and shape information for the indexing and retrieval, and applies these region features obtained through unsupervised segmentation, as opposed to applying them to the whole image domain. Fuzzy Club emphasizes improving on a color feature “inaccuracy” problem in the region based literature that is color histogram bins are not independent. Fuzzy Club first segments an image into regions of 4x4 blocks and extracts color and texture features on each block. The k-means algorithm is used to cluster similar pixels together to form a region. The Lab color space is used to extract color features and Haar wavelet transform is used to extract three texture features. A secondary clustering is performed to reduce query processing time. Regions with similar features are grouped together in the same class.

3. FEATURE EXTRACTION

Feature extraction is a means of extracting compact but semantically valuable information from images. This information is used as a signature for the image. Similar images should have similar signatures. If we look at the image shown in Figure 4, the white color and the texture of the building are characteristic properties. In a similar way, the sky can be described by its blue color. Furthermore, we can take the size of the objects in the image into account. Representation of images needs to consider which features are most useful for representing the contents of images and which approaches can effectively code the attributes of the images. Feature extraction of the image in the database is typically conducted off-line so computation complexity is not a significant issue. This section introduces three features: texture, shape, and color, which are used most often to extract the features of an image.

3.1. Colour

One of the most important features visually recognized by humans in images is color. Humans tend to distinguish images based mostly on color features. Because of this, color features are the most widely used in CBIR systems and the most studied in literature. Color is a powerful descriptor that simplifies object identification, and is one of the most frequently used visual features for content-based image retrieval. To extract the color features from the content of an image, a proper color space and an effective color descriptor have to be determined. The purpose of a color space is to facilitate the specification of colors. Each color in the color space is a single point represented in a coordinate system. Several color spaces, such as RGB, HSV, CIE L*a*b, and CIE L*u*v, have been developed for different purposes [35]. Although there is no agreement on which color space is the best for CBIR, an appropriate color system is required to ensure perceptual uniformity.

FIGURE. 3. (a) Image Properties (b) Sample Image (c) Colour Histogram of Sample Image

Therefore, the RGB color space, a widely used system for representing color images, is not suitable for CBIR because it is a perceptually non-uniform and device-dependent system [36]. After selecting a color space, an effective color descriptor should be developed in order to represent the color of the global or regional areas. Several color descriptors have been developed from various representation schemes, such as color histograms [37], color moments [38], color edge [39], color texture [40], and color correlograms [41].

3.2. Colour Histogram

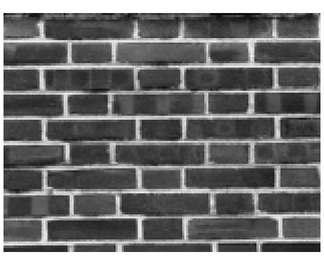
The most commonly used method to represent color feature of an image is the color histogram. A color histogram is a type of bar graph, where the height of each bar represents an amount of particular color of the color space being used in the image [35]. The bars in a color histogram are named as bins and they represent the x-axis. The number of bins depends on the number of colors there are in an image. The number of pixels in each bin denotes y-axis, which shows how many pixels in an image are of a particular color. The color histogram can not only easily characterize the global and regional distribution of colors in an image, but also be invariant to rotation about the view axis.

In color histograms, quantization is a process where number of bins is reduced by taking colors that are similar to each other and placing them in the same bin. Quantizing reduces the space required to store the histogram information and time to compare the histograms. Obviously, quantization reduces the information regarding the content of images; this is the tradeoff between space, processing time, and accuracy in results [41]. Color histograms are classified into two types, global color histogram (GCH) and local color histogram (LCH). A GCH takes color histogram of whole image and thus represents information regarding the whole image, without concerning color distribution of regions in the image. In the contrary, an LCH divides an image into fixed blocks or regions, and takes the color histogram of each of those blocks. LCH contains more information about an image, but when comparing images, it is computationally expensive. GCH is known as a traditional method for retrieving color based images. Since it does not include color distribution of the regions, when two GCHs are compared, one might not always get a proper result when viewed in terms of similarity of images [42]. An example of a color histogram in the HSV color space can be seen with the image in Figure 6.

3.3. Texture

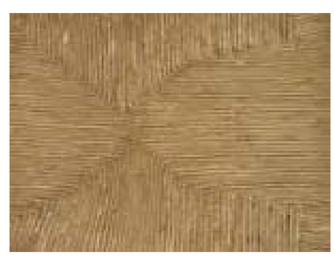
Texture definitions are based on texture analysis methods and the features extracted from the image. However, texture can be thought of as repeated patterns of pixels over a spatial domain, of which the addition of noise to the patterns and their repetition frequencies results in textures that can appear to be random and unstructured. Texture properties are the visual patterns in an image that have properties of homogeneity that do not result from the presence of only a single color or intensity. The different texture properties as perceived by the human eye are, for example, regularity, directionality as shown in figures (a) to (d).

In real world scenes, texture perception can be far more complicated. The various brightness intensities give rise to a blend of the different human perception of texture as shown in figures (e) & (f). Image textures have useful applications in image processing and computer vision. They include: recognition of image regions using texture properties, known as texture classification, recognition of texture boundaries using texture properties, known as texture segmentation, texture synthesis, and generation of texture images from known texture models. Since there is no accepted mathematical definition for texture, many different methods for computing texture features have been proposed over the years. Unfortunately, there is still no single method that works best with all types of textures. According to Manjunath and Ma [44], the commonly used methods for texture feature description are statistical, model-based, and transform-based methods. The texture feature description categories are explained below.

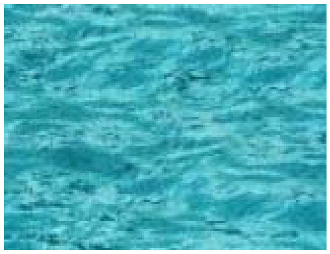


(a)Regular

(b) Irregular



(c) Directional



(d) Non Directional

(e) Complex texture 1

(f) Complex texture 2

FIGURE. 4. (a) to (d): Simple Textures Images; (e) to (f): Complex Textures Images

(i) Statistical Methods:

Statistical methods analyze the spatial distribution of grey values by computing local features at each point in the image, and deriving a set of statistics from the distribution of the local features. They include co- occurrence matrix representation, statistical moments, gray level differences, autocorrelation function, and grey level run lengths. The most commonly used statistical method is the Gray-level Co-occurrence Matrix (GLCM) [45]. It is a two-dimensional matrix of joint probabilities between pairs of pixels, separated by a distance, d, in a given direction, r. It is popular in texture description and is based on the repeated occurrence of some gray level configuration in the texture; this configuration varies rapidly with distance in fine textures and slowly in coarse textures. Haralick [45] defined 14 statistical features from gray-level co-occurrence matrix for texture classification, such as energy, entropy, contrast, maximum probability, autocorrelation, and inverse difference moment. Gray-level co-occurrence matrix method of representing texture features has found useful applications in recognizing fabric defects and in rock texture classification and retrieval [46].

(ii) Model Based Approaches:

Model-based texture methods try to capture the process that generated the texture. By using the model-based features, some part of the image model is assumed and an estimation algorithm is used to set the parameters of the model to yield the best fit [47]. To describe a random field, assume the image is modeled as a function f (r, ω), where r is the position vector representing the pixel location in the 2-D space and ω is a random parameter. Once a specific texture ω is selected, f (r, ω) is an image, which is a function over the 2-D grid indexed by r. Function f (r, ω) is called as a random field. There are currently three major model based methods: Markov random fields by Dubes and Jain [48], fractals by Pentland [49], and the multi-resolution autoregressive features introduced by Mao and Jain [50].

(iii) Transform Domain Features:

The word transform refers to a mathematical representation of an image. There are several texture classifications using transform domain features in the past, such as discrete Fourier transform, discrete wavelet transforms, and Gabor wavelets. On the other hand, wavelet analysis breaks up a signal into shifted and scaled versions of the original wavelet (mother wavelet), which refers to decomposition of a signal into a family of basis functions obtained through translation and dilation of a special function. Moments of wavelet coefficients in various frequency bands have been shown to be effective for representing texture [46]. Gabor filter (or Gabor wavelet) has been shown to be very efficient. Manjunath and Ma [44] have shown that image retrieval using Gabor features outperforms that using other transform features.

3.4. Shape

One of the common used features in CBIR systems is the shape. Shape of an object is the characteristic surface configuration as represented by the outline or contour. Shape recognition is one of the modes through which human perception of the environment is executed. It is important in CBIR because it corresponds to region of interests in images. Shape feature representations are categorized according to the techniques used. They are boundary-based and region-based [50]. In region based techniques, all the pixels within a shape are taken into account to obtain the shape representation. Common region based methods use moment descriptors to describe shape [51]. Region moment representations interpret a normalized gray level image function as a probability density of a 2-D random variable. The first seven invariant moments, derived from the second and third order normalized central moments, are given by Hu [52]. Comparing with region based shape representation; contour based shape representation is more popular. Contour based shape representation only exploits shape boundary information. Simple contour-based shape descriptors include area, perimeter, compactness, eccentricity, elongation, and orientation. Complex boundary-based descriptors include Fourier descriptors, grid descriptors, and chain codes [42]. In our proposed system, we do not consider shape features during similarity distance computation. Including shape feature in the proposed system is one of our future works.

3.5. Shape Features

Shape information can be 2D or 3D in nature, depending on the application. The three shape descriptors are: Region Shape, Contour Shape and Shape 3D. 2D shape descriptors, the Region Shape and Contour Shape descriptors are intended for shape matching. They do not provide enough information to reconstruct the shape nor to define its position in an image. These two shape descriptors have been defined because of the two major interpretations of shape similarity, which are contour-based and region-based. Region Shape and Contour Shape descriptors as well as the Shape 3D descriptor are described in more detail below.

(i) Region Shape

The shape of an object may consist of a single region or a set of regions as well as some holes in the object. Since the Region Shape descriptor, based on the moment invariants [53], makes use of all pixels constituting the shape within a frame, it can describe any shape. The shape considered does not have to be a simple shape with a single connected region, but it can also be a complex shape consisting of holes in the object or several disjoint regions. The advantages of the Region Shape descriptor are that in addition to its ability to describe diverse shapes efficiently it is also robust to minor deformations along the boundary of the object. The feature extraction and matching processes are straightforward. Since they have low order of computational complexities they are suitable for shape tracking in the video sequences [54].

(ii) Contour Shape

The Contour Shape descriptor captures characteristics of a shape based on its contour. It relies on the so-called Curvature Scale-Space (CSS) [55] representation, which captures perceptually meaningful features of the shape. The descriptor essentially represents the points of high curvature along the contour (position of the point and value of the curvature). This representation has a number of important properties, namely, it captures characteristic features of the shape, enabling efficient similarity-based retrieval. It is also robust to non-rigid motion [53, 54].

3.6. Similarity Measure

The similarity between two images is defined by a similarity measure. Selection of similarity metrics has a direct impact on the performance of content-based image retrieval. The kind of feature vectors selected determines the kind of measurement that will be used to compare their similarity [23]. If the features extracted from the images are presented as multi-dimensional points, the distances between corresponding multi-dimensional points can be calculated. Euclidean distance is the most common metric used to measure the distance between two points in multi-dimensional space [56]. For other kinds of features such as color histogram, Euclidean distance may not be an ideal similarity metric or may not be compatible with the human perceived similarity. Histogram intersection was proposed by Swain and Bllard [57] to find known objects within images using color histograms.

3.7. Indexing Structures

When manipulating massive databases, a good indexing is a necessity. Processing every single item in a database when performing queries is extremely inefficient and slow. When working with text-based documents, creating good indexes is not very difficult. Next, in-depth processing only needs to be done with these documents. When searching for images, however, this approach is much more difficult. Raw image data is non-indexable as such, so the feature vectors must be used as the basis of the index. Popular multi-dimensional indexing methods include the R-tree and the R*-tree algorithms [23]. The Self Organizing Map (SOM) is also one of the indexing structures [58]. The SOM is trained to match the shape of the data in the feature space. After the training, the closest node in the SOM is calculated for every image in the database. When a query is done, the first thing to be done is to calculate the closest SOM node, also known as the best matching unit (BMU), to the query image's feature vector. When this is done, we know which images in the database are the closest to the query image.

4. IMAGE REGION MOMENTS

Image moments and their functions have been utilized as features in many image processing applications, viz., pattern recognition, image classification, target identification, and shape analysis. Moments of an image are treated as region-based shape descriptors. Among region-based descriptors, moments are very popular. These include invariant moments, Zernike moments and Legendre moments.

4.1. Invariant Moments

Invariant moments or geometric moments are the simplest moment functions with the basis $\varphi_{pq}(x,y) = x^p y^q$. The geometric moment function m_{pq} of order $(p+q)$ is defined by

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y);\ p,q = 0,1,2,.. \tag{1}$$

The geometric central moments that are invariant to translation are defined by

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y);\ p,q = 0,1,2,.. \tag{2}$$

The seven invariant moments are given by the following equations

$$I_1 = \eta_{20} + \eta_{02} \tag{3}$$

$$I_2 = (\eta_{20} + \eta_{02})^2 + (2\eta_{11})^2 \tag{4}$$

$$I_3 = (\eta_{20} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{02})^2 \tag{5}$$

$$I_4 = (\eta_{20} + 3\eta_{12})^2 + (\eta_{21} + \eta_{02})^2 \tag{6}$$

$$\begin{aligned} I_5 &= (\eta_{20} - 3\eta_{12})(\eta_{20} + \eta_{12})[(\eta_{20} + \eta_{12})^2 \\ &-3(\eta_{21} + \eta_{02})^2] + (3\eta_{21} - \eta_{02})(\eta_{21} + \eta_{02}) \\ &[3(\eta_{20} + \eta_{12})^2 - (\eta_{21} + \eta_{02})^2] \end{aligned} \tag{7}$$

$$\begin{aligned} I_6 &= (\eta_{20} - \eta_{02})[(\eta_{20} + \eta_{12})^2 - (\eta_{21} + \eta_{02})^2] \\ &+4\eta_{11}(\eta_{20} + \eta_{12})(\eta_{21} + \eta_{02}) \end{aligned} \tag{8}$$

$$\begin{aligned} I_7 &= (3\eta_{21} - \eta_{02})(\eta_{20} + \eta_{12})[(\eta_{20} + \eta_{12})^2 - \\ &3(\eta_{21} + \eta_{02})^2] - (\eta_{20} - 3\eta_{12})(\eta_{21} + \eta_{02}) \\ &[3(\eta_{20} + \eta_{12})^2 - (\eta_{21} + \eta_{02})^2] \end{aligned} \tag{9}$$

$n_{p,q} = \dfrac{\mu_{pq}}{y}, \gamma = 1 + \dfrac{p+q}{z}$, for $p+q = 2,3,...$

Invariant moments are invariant to translation, rotation and scaling [59, 60].

4.2. Zernike Moments

Zernike Moments (ZM) is orthogonal moments and can be used to represent shape content of an image with minimum amount of information redundancy [69, 70]. Orthogonal moments allow for accurate reconstruction of the image, and makes optimal utilization of shape information. Zernike Moments (ZM) are widely used in CBIR as shape descriptors [18, 19]. Z Zernike moments are derived from the orthogonal Zernike polynomials. Hence, it is an orthogonal moment. The Zernike moments are given by

$V_{nm}(x,y) = V_{nm}(r\cos\theta, r\sin\theta) = R_{mn}(\gamma)\exp(jm\theta)$ (10)

$R_{mn}(\gamma)$ is the orthogonal radial polynomial, and is given by

$$R_{mn}(\gamma) = \sum_{z=0}^{(n-|m|)/2} (-1)^2 \frac{(n-s)!}{s!x\frac{(n-2s+1|m|)!(n-2s-|m|}{2}} \gamma^{n.2s} \quad (11)$$

$n = 0,1,2,...0 \le |m \le n;| \; n-|m|$ is even.

The Zernike moments for the image $f(x,y)$ are defined by equation

$$Z_{nm} = \frac{n+1}{\pi}\sum_{r}\sum_{\theta} f(r\cos\theta, r\sin\theta)$$
$$R_m(\gamma)\exp(jm\theta)r \le 1 \quad (12)$$

Zernike moments are invariant to rotation, translation and scaling. Also they are robust to noise and minor variations in shape. But the computational complexity of Zernike moments is high [59,61].

4.3. Legendre Moments

Legendre moments use Legendre polynomials as the kernel function. The two-dimensional Legendre moments of order $(p+q)$ for an image $f(x,y)$ are defined by equation

$$L_{pq} = \frac{(2p+1)(2q+1)}{4}\int_{-1}^{1}\int_{-1}^{1} P_p(x)XP_q(y)f(x,y)dxdy \;\; x,y \in [-1,1] \quad (13)$$

where the Legendre polynomial $P_p(x)$ of order $'p'$ is given by the equation

$$P_p(x) = \sum_{k=0}^{p}\left((-1)^{\frac{p-k}{z}} \frac{1}{zp} \frac{(p+k)x^k}{\left[\frac{p-k}{z}\right]!\left[\frac{p+k}{z}\right][k]} \right) \quad (14)$$

The Legendre moments described in equation (11) can be expressed in discrete form by

$$L_{pq} = \lambda_{pq}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_p(x_i)P_q(y_j)f(i,j) \quad (15)$$

Where $\lambda_{pq} = \frac{(2p+1)(2q+1)}{N^2}$; x_i & y_i are the normalized pixel coordinates and are given by

$$x_i = \frac{2_i}{N-1} - 1 \,\&\, y_i = \frac{2_j}{N-1} - 1 \quad (16)$$

4.4. Exact Legendre Moments

Legendre Moments (LM) are continuous and orthogonal moments, they can be used to represent an image with minimum amount of information redundancy. Many algorithms are developed for the computation of LM [71, 72, 73], but these methods focus mainly on 2D geometric moments. When they are applied to a digital image, a numerical approximation is necessary. Error due to approximation increases as the order of the moment increases. An accurate method for computing the Exact Legendre Moments (ELM) proposed by Hosney [74].

5. DISTANCE MEASURES

Distance measures are used for comparing the similarity of two images. There are different kinds of similarity measurements like Euclidean distance, histogram intersection, Bhattacharya distance and Mahalanobis distance for CBIR applications.

5.1. Euclidean Distance

Let p be the query image and t be the target image and let $p_q(z_q)$ and $p_t(z_t)$ be their respective probability densities. The Euclidean distance between the query image and the target image is given by equation

$$D_{Ece}(q_i, t_i) = \sum_{i=1}^{n} (q_i - t_i)^2 \tag{17}$$

In Euclidean distance, the least value of distance measure indicates the similarity [62, 63, 64].

5.2. Histogram Intersection

It is a distance measure for comparing histograms. It calculates the common part of the two histograms, and neglects the features occurring in a single histogram. The histogram intersection of two histograms [65] H and $H^{'}$ is calculated using equation

$$d_{\cap}(H, H^{'}) = \sum_{m=1}^{M} \min((H_{m^{'}}, H^{'}_{m}) \tag{18}$$

5.3. Bhattacharya Distance

The Bhattacharya Distance measures the similarity between two discrete or continuous probability distributions. A popular distance of similarity between two Gaussian distributions is the Bhattacharya distance. The Bhattacharya distance [66] between the query image q and the target image t in the database is given by equation

$$D_{Bhat}(q,t) = \frac{1}{g}(\mu_q - \mu_t)^T \left[\frac{\sum_q - \sum_t}{2}\right]^{-2} (\mu_q - \mu_t) + \frac{1}{2} l_n \frac{\left| (\sum_q - \sum_t \;/\, 2 \right|}{\sqrt{\left|\sum_q\right| \left|\sum_t\right|}} \tag{19}$$

Where μ_q and μ_t are the mean vectors $\sum_q$ and $\sum_t$ are the covariance matrices of the query image q and the target image t, respectively [67, 68].

5.4. Mahalanobis Distance

The Mahalanobis Distance is based on the correlations between variables, and is used to analyze various patterns. It is useful in determining the similarity between an unknown sample set and a known one. The unknown sample set is the query image, and the known set is the images in the database. The Mahalanobis distance between the query image q and the target image t is given by equation following equation [66].

$$D_{Maha}(q,t) = (\mu_q - \mu_t)^T \sum{}^{-2} (\mu_q - \mu_t) \tag{20}$$

6. MACHINE LEARNING TECHNIQUES

The high level semantic features are derived from the image database with the help of machine learning techniques. There are two types of machine learning techniques i.e. supervised machine learning technique and unsupervised machine learning technique.

6.1. Supervised Machine Learning Techniques

Neural networks, Decision trees, and Support Vector Machines (SVMs) are some of the supervised machine learning techniques, which learn the high level concepts from low-level image features. The supervised machine learning techniques perform the classification process with the help of the already categorized training data. For the training data, the input (low level image features) and the desired output is already known. Hence, given a query image, the low level features are extracted and it is given as input to any one of the machine learning algorithms which is already trained with the training data. The machine learning algorithm predicts the category of the query image which is nothing but the semantic concept of the query image. Hence instead of finding similarity between the query image and all the images in database, it is found between the query image and only the images belonging to the query image category. Also when the entire database is searched, the retrieval result contains images of various categories. But when the machine learning techniques are used, since the query image‘s category (semantic concept) is predicted, the retrieval results will contain the images belonging to that category alone.

6.1.1. Neural Network

Neural networks are also useful in concept learning. The low level features of the segmented regions of the training set images are fed into the neural network classifiers, to establish the link between the low level image features and high level semantics. The disadvantage of this method is that it requires a large amount of training data, and is computationally intensive [75, 76, 77]. When the query image feature vector is presented to the neural network, it gives its semantic concept.

6.1.2. Support Vector Machine

Support Vector Machines (SVMs) are supervised learning methods [78, 79] used for image classification. It views the given image database as two sets of vectors in an ‘ n ’ dimensional space and constructs a separating hyper plane that maximizes the margin between the images relevant to query and the images not relevant to the query. SVM is a kernel method and the kernel function used in SVM is very crucial in determining the performance.

The basic principle of SVMs is a maximum margin classifier. Using the kernel methods, the data can be first implicitly mapped to a high dimensional kernel space. The maximum margin classifier is determined in the kernel space and the corresponding decision function in the original space can be non-linear [80]. The non-linear data in the feature space is classified into linear data in kernel space by the SVMs. The aim of SVM classification method is to find an optimal hyper plane separating relevant and irrelevant vectors by maximizing the size of the margin (between both classes).

Image classification or categorization is a machine learning approach and can be treated as a step for speeding-up image retrieval in large databases and to improve retrieval accuracy. Similarly, in the absence of labelled data, unsupervised clustering is also found useful for increasing the retrieval speed as well as to improve retrieval accuracy. Image clustering inherently depends on a similarity measure, while image classification has been performed by different methods that neither require nor make use of similarity measures [81].

6.1.3. SVM-Binary Decision Tree

The SVM-BDT takes advantage of the efficient computation of the binary tree architecture, and the high classification accuracy of the SVMs. Here, (N-1) SVMs are needed to train an N class problem. For the construction of the SVM-BDT, first, the semantic template of each of the categories or classes is found. The Euclidean distance between the semantic templates of each of the N classes is the N × N distance matrix. Two classes that have the largest Euclidean distance are assigned to each of the two clustering groups. The semantic template of these two classes is the cluster center of the two groups. The cluster center is updated to the semantic template of the newly added class. All the classes are assigned to one of the two possible groups of classes. The process continues, until there is only one class per group. The SVM binary classifier is used to train the samples in each non leaf nodes of the decision tree [68]. During the testing time, the feature vector of the query image is given as input to the SVM-BDT, and only $[\log_2 N]$ classifiers are consulted during the testing time.

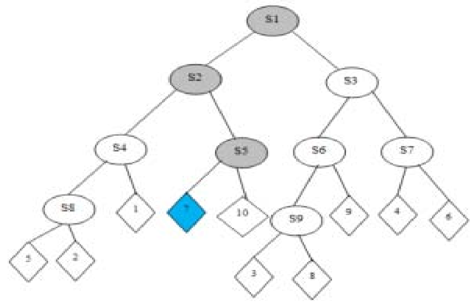


FIGURE 8: SVM-BDT for 10 Category Image Set.

The SVM-BDT predicts the label (semantic category) of the query image. Hence, the similarity distance between the query image and the predicted category images are computed and the images with least distance are retrieved. An example of a 10- class SVM-BDT is shown in Figure 8. For SVM-based image classification, recent work shows that the radial basis kernel function (RBF) works well, when the relation between the class labels and attributes is nonlinear [62].

6.1.4. Bayes Classifiers

Bayes classifiers are used in text retrieval systems. Since ten years, CBIR community is transposing them to image retrieval [85, 66]. Bayes binary classifiers use the class-conditional likelihood associated with class $cP(X \mid c)$ to compute the mapping function $g(x)$ of an input vector x.

$$g(x) = \arg\max_{c \in (-1,1)} P(X \mid c) P(c) \tag{21}$$

Because we have no prior assumption on the size of a class, we assume that $p(1) = p(-1) = \frac{1}{2}$ once $g(x)$ is computed the relevance function $f(x)$ may be expressed as follows.

$$f(x) = p(x \mid c) = g(x) \tag{22}$$

To estimate the probability density function, we use a kernelized version of Parzen windows:

$$p(x \mid c) = \frac{1}{\left|\{i(y_i = c\}\right| \sum_{i \in \{i \mid y_i \in c\}} k(x, x_i)} \tag{23}$$

where K(., .) is a kernel function.

6.1.5. k-Nearest Neighbors

This classification method has been used successfully in image processing and pattern recognition. For instance, in competition with neural networks, linear discriminant analysis (and others), kNearest Neighbors performed best results on pixel classification tasks [87]. The k-NN algorithm is one among the simplest of all machine learning algorithms. K-nearest neighbor algorithm (KNN) is also involved into our CBIR system. Based on the training result, KNN is applied for the query data images. KNN helps to classify the input data; also it fixes the code book which means the training result can be self-adapted. k-Nearest Neighbors classifiers attempt to directly estimate $f(x)$ using only the k nearest neighbors of x. In a small database ,a simple sequential scan is usually employed for k nearest – neighbor (KNN) search. But for large data set, efficient indexing algorithms are imperative. High dimensional data is increasingly in many common fields. As the number of dimensions increase, many clustering techniques begin to suffer from the curse of dimensionality, degrading the quality of the results.

6.2. Unsupervised Machine Learning Techniques

Unsupervised learning refers to the problem of trying to find hidden structure in the unlabeled data. It has no measurements of outcome, to guide the learning process. Image clustering is a typical unsupervised learning technique. It groups the sets of image data in such a way, that the similarity within a cluster should be maximized, and the similarity between different clusters must be minimized [82].

K-means clustering aims to partition the given n observations into k clusters. The mean of each cluster is found and the image is placed in a cluster, whose mean has the least Euclidean distance with the image feature vector. Parallel techniques for K-means are developed that can largely accelerate the algorithm [89], [90], [91]. In high dimensions, data becomes very sparse and distance measures become increasingly meaningless. Paper [88] reviewed the literature on parsimonious models and Gaussian models from the most complex to simplest which yields a method similar to the K Means approach. N Cut clustering is used to cluster the database images into different semantic classes. A set of n images is represented by a weighted undirected graph represents image.

CBIR based on Visual Contents

Sl. No.	Papers	Features	Approaches	Limitations
1	W. Niblack et al. [92]	Colour	Histogram and colour moments	Query image is an unknown image, then the retrieval performance is poor
2	Chad Carson et al. [93]		Region Histogram	Result in the mismatch of the retrieval process when the image‘s orientation, and position or scales are altered.
3	J. Sawhney & Hefner et al. [94]		Colour Histogram	Similarity measure is extended to retrieve the texture regions from a database of natural images.
4	Stricker & Orengo [95]		Colour Moment	Semantically relevant images will be retrieved with amount of time
5	Michel Orega et al. [96]	Shape	Fourier Transform	The user gives feedback and the query image information becomes a new cluster
6	F. Mokhtarian et al. [97]		Curvature Scale Space	Given a query image, the user has to select the region of interest from the query image
7	Sougata Mukherjea et al. [98]		Template Matching	Images from the history of the user access patterns, and the access frequencies of the images in the database.

8	Furnikaza Kanehara et al. [99]	Shape	Convex Parts	The feedback can be got from the user again and again, till the user is satisfied with the results.
9	Pentland et al. [100]		Elastic Deformation of Templates	The similarity distance is found between the query image and the images belonging to the predicted cluster alone.
10	J. R. Smith et al. [101]	Texture	Wavelet Transform	All the classes are assigned to one of the two possible groups of classes.
11	S. Michel et al. [102]		Edge Statistics	Due to the complex distribution of the image data, the k-means clustering often cannot separate images
12	B. S Manjunath et al. [103]		Gabor Filters	The machine learning predicts the category of the query image
13	George Tzaglarakis et al. [104]		Statistical	The query image belongs to the class for which the membership is very large
The Semantic Gap In Image Retrial				
Sl. No.	Papers	Techniques	Approaches	Limitations
14	S. F Chang et al. [109]	Semantic Template	Semantic Visual Template	Texture descriptors contain features derived from co-occurrence matrices
15	Yang et al. [105]	Relevance Feedback	Semantic Feedback Mechanism	Lower retrieval precision by introducing the semi-supervision to the non-linear Gaussian-shaped RBF relevance feedback
16	Rege et al. [106]		Multiuser relevance feedback (User centered semantic hierarchy)	System is said to be efficient if semantic gap is minimum.
17	Liu et al. [107]		Semantic manifold	Bridging the semantic gap between the low-level features and the high-level semantics
18	Janghy Yoon & Nikil Jayant et al. [108]	Relevance Feedback	Multimedia model feedback	Descriptors based on color representation might be effective with a data set containing black and white images.
19	I notes [110]	Manual	User Annoted region	Retrieval application is

				specialized for a certain, limited domain, the smaller the gap can be made by using domain knowledge.
20	Face book [111]		User Annoted region/object	The resulting segments can be described by shape features
21	Google user labella [112]		User Annoted whole image	Often, the performance of a retrieval system with feedback
22	Brad Shaw et al. [113]	Semiautomatic	Bayer Probability	Probabilistic retrieval form is the use of support vector machines
23	Ghoshot [114]		Co-occurrence model	Background complication and independent of image size and orientation
24	Ilaria et al. [115]		Graph based link	Cannot be used to differentiate objects with different sizes or shapes.
25	Huang et al. [116]	Automatic	Decision tree	Line contents of images can be used to represent texture of the image.
26	Feng & Chan [117]		Bootstrapping	Statistical approaches do not exploit the sensitivity of the human visual system to textures.
27	Gao et al. [118]		Latent Semantic analysis	The characterization consists of local autocorrelation of coefficients in each subband.
28	Mori et al. [119]		Hidden Markov model	descriptor can tackle not only rotation but also small non-rigid deformation
29	P. L Standchey et al. [120]	Object Ontology	Colour representation ontology	It is extremely difficult to describe high level semantic concepts with image features only
30	V. Mezaris [121]	Object Ontology	High level concept ontology	A query system integrating multiple query seeds
31	Huan Wang [122]		Multimodality ontology	Could accommodate any number of features in a modular and extensible way.

TABLE 1: Performance Comparison of Various CBIR Techniques.

$G = (v, E), V = \{1, 2, ..., n\}$ and edges $E = \{(i, j) / i, j \in V$

$E = \{(i, j) / i, j \in V$ are formed between every pair of nodes. The weight w_{ij} of an edge (i, j) is an image. The system displays the image clusters and adjusts the model of similarity measure according to user feedbacks [83].

Fuzzy clustering models provide a promising solution to the clustering problem. The Fuzzy c-means (FCM) clustering is the most widely used fuzzy clustering algorithm. This algorithm is based on an iterative optimization of a fuzzy objective function. The degree of membership of a data item to a cluster is between [0, 1]. For a given query image, the output of the FCM is the membership value of the image with each of the K classes. The query image belongs to the class for which the membership value is high [62, 63, 84].

Hence the unsupervised learning algorithms takes the uncategorized image data as input and clusters those data into a number of categories by finding the hidden structure in the unlabelled data. The clustering algorithms divide the given data into n clusters and give cluster centers of each cluster. When a query image features are given to the clustering algorithm, it finds the distance between the query image and the entire cluster centers.

7. COMPARISON OF VARIOUS CBIR METHODS

The field of Content Based Image Retrieval has been an active research area for several decades and has been paid more and more attention in recent years as a result of the dramatic and fast increase in the volume of digital images. Many novel techniques have been proposed to tackle the challenges. However; all current CBIR systems suffer from insufficient generalization performance and accuracy as they are not able to establish a robust link or between image features and high-level concepts. In this section Table I summarize performance comparison of various CBIR techniques with respect to various components such as visual contents and semantic gaps. Table II gives comparison of various CBIR components used in various techniques with appropriate image data sets.

Sl. No.	Papers	Data Set	Distance measures	Machine Learning
1	Rahman M.H. et al. [123]	Brodatz database	Normalized Euclidean distance	Random walk with relevance feedback
2	Saptadi et al. [61]	MRI images	Error tolerance distance	R-Tree Data Structure
3	Felci Rajam et al. [63]	COREL dataset	Euclidean, Bhattacharya-Mahalanobis	SVM Binary Decision Tree
4	Lining et al. [124]	COREL dataset	Mahalanobis distance	Generalized Biased Discriminant Analysis (GBDA)
5	Imtnan et al. [125]	Vistex database, Outex database	Kullback–Leibler(KL) divergence	Parametric Spectral Analysis
6	[Hatice et al.[126]	Follicular Lymphoma, Neuroblastoma	Correlation distance measure	SVM, Nearest neighbor search
7	Ja-Hwung et al. [127]	COREL dataset	Weighted KNN search	Navigation Pattern based Relevance Feedback
8	Yu-Gang et al. [128]	NUS-WIDE TRECVID	Canberra distance	Semantic graph
9	Kuldeep et al. [129]	Data set of MRI, CT-scan and X-ray	Euclidean Distance	Parallel implementation feature extraction and feature matching
10	Felci et al. [68]	Caltech	Euclidean, Bhattacharya, Mahalanobis	SVM-BDT, SCM
11	Wang Xing Yuan, [129]	Public sources	Weighted KNN search	Navigation Pattern based Relevance Feedback
12	Samuel et al., [130]	Oliva dataset, Caltech dataset	l^1- norm	Random walk with relevance feedback

TABLE 2: Comparison of Various CBIR Components used in Various Components.

8. CONCLUSION

This paper presents a brief survey on work related to the exciting fields of content-based image retrieval and provides a detailed review of the works carried out in this field. This paper also discusses the various methodologies used for extracting the salient low level features and various distance measures to find the similarity between images in reducing the semantic gap between the low level features and the high level semantic concepts. A discussion of various approaches of CBIR and comparison of various techniques with respect to data are also made.

9. FUTURE RESEARCH

This paper presents a comparative study of Content Based Image Retrieval Trends and the various approaches towards resolving some of the problems encountered in CBIR. One alternative is to use more sophisticated feature representations. Instead of using a purely data-driven evaluation using basic image features, higher-level information about regions could be used. Since an image epitome provides a composite description of shape and appearance, it is possible to achieve a better measure of homogeneity/heterogeneity of the segments. One of the steps towards resolving the semantic information problem, when possible, prior knowledge, especially application-dependent knowledge, should be incorporated into an evaluation method so that the evaluation method knows the preferred characteristics of a segment. Different methods can be applied to include prior knowledge about a preferred features.

10. REFERENCES

[1] T. Huang, Y. Rui, Image retrieval: Past, present, and future, in: Proceedings of the International Symposium on Multimedia Information Processing, 1997, pp. 1–23.

[2] Ch. Srinivasa Rao , S. Srinivas Kumar and B. Chandramohan," Content Based Image Retrieval using Exact Legendre Moments and Support Vector Machine", The International Journal of Multimedia and its Applications (IJMA), Vol.2, No.2, pp: 69-79, May 2010.

[3] S.K. Chang, T. Kunii, “Pictorial Database Applications”, IEEE Computer, Vol. 14, No. 11, 1981.

[4] C.Gururaj, D.Jayadevappa, Satish Tunga, “ Novel Algorithm for Exudate Extraction from Fundus Images of the Eye for a Content Based Image Retrieval System”, 4th IEEE International Conference on Control System, Computing and Engineering (ICCSCE – 2014), ISBN- 978-1-4799-5685-2, 28th – 30th November 2014, pp 460 – 465, Penang, Malaysia, DOI: 10.1109/ICCSCE.2014.7072763

[5] J. Li, J.Z. Wang, G. Widerhold, “IRM: Integrated Region Matching for Image Retrieval”, 8th ACM International Conference on Multimedia, pp. 147-156, October 2000.

[6] H.J. Lin, et al. “A Study of Shape Based Image Retrieval”,24th IEEE International Conference on Distributed Computing Systems Workshops, 2004.

[7] Y. Rui, T.S. Huang, S.F. Chang SF, “Image Retrieval: Current Techniques, Promising Directions and Open Issues”, J. Vis. Commun. Image Rep., Vol. 10, No. 4, pp.39-62, 1999.

[8] P. Hiremath, J. Pujari, “Content Based Image Retrieval Using Color, Texture and Shape Features”, IEEE International Conference on Advanced Computing and Communications, ADCOM, 2007M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 989.

[9] M.M. Islam, D. Zhang, G. Lu, “Automatic Categorization of Image Regions Using Dominant Colour Based Vector Quantization”, Digital Image Computing: Techniques and Applications, Canberra, Australia, pp.191-198, December 1-3, 2008.

[10] N. Rasiwasia, N. Vasconcelos, P.J. Moreno, “Query by Semantic Example”, 15th International Conference on Image and Video Retrieval, Lecture Notesin Computer Science, Vol. 4071, Springer, Berlin, pp. 51-60, 2006.

[11] V.A. Tanase, “Content Based Image Retrieval Systems: A Survey”, Dept. of Computing Science, Utrecht University, Technical Report, 2000.

[12] W. Zhang, et al., “Shape Based Indexing in a Medical Image Database”, IEEE Workshop on Biomedical Image Analysis, 1998.

[13] Ville Viitaniemi, “Image Segmentation in Content-Based Image Retrieval”, Helsini University of Technology, Department of Electrical and Communications Engineering, Master Thesis, Finland 24th May 2002.

[14] H.B. Kekre, “Survey of CBIR Techniques and Semantics”, International Journal of Engineering Science and Technology (IJEST), ISSN: 0975-5462 Vol. 3 No. 5, May 2011.

[15] Kanchan Saxena, Vineet Richaria, Vijay Trivedi, “A Survey on Content Based Image Retrieval using BDIP, BVLC AND DCD”, Journal of Global Research in Computer Science, Vol.3, No. 9, Sep. 2012, ISSN-2229-371X.

[16] Z. Wang, Z. Chi, and D. Feng, “Fuzzy integral for leaf image retrieval,” Proc. IEEE Intl. Conference on Fuzzy Systems, 2002.

[17] A. Csillaghy, H. Hinterberger, and A.O. Benz, “Content Based Image Retrieval in Astronomy,” Information Retrieval, 3(3):229–241, 2000.

[18] T. H. Painter, J. Dozier, D. A. Roberts, R. E. Davis, and R. O. Green, “Retrieval of sub-pixel snow-covered area and grain size from imaging spectrometer data,” Remote Sensing of Environment, 85(1):64–77, 2003.

[19] M. Schroder, H. Rehrauer, K. Seidel, M. Datcu, “Interactive learning and probabilistic retrieval in remote sensing image archives,” IEEE Trans. Geosciences and Remote Sensing, 38(5):2288–2298, 2000.

[20] J.Z. Wang, J. Li, and G. Wiederhold, “Simplicity: Semantics-Sensitive Integrated Matching for Picture Libraries,” IEEE Trans. Pattern Analysis and Machine Intelligence, 23(9), 947–963, 2001.

[21] Airliners.net, http://www.airliners.net.

[22] H. Muller, W. Muller, D. M. Squire, S. Marchand-Maillet, and T. Pun, “Performance evaluation in content-based image retrieval: Overview and proposals,” Pattern Recognition Letters, 22(5):593-601, 2001.

[23] F. Long, H. Zhang, H. Dagan, and D. Feng, “Fundamentals of content based image retrieval,” in D. Feng, W. Siu, H. Zhang (Eds.): “Multimedia Information Retrieval and Management. Technological Fundamentals and Applications,” Multimedia Signal Processing Book, Chapter 1, Springer-Verlag, Berlin Heidelberg New York, 2003, pp.1-26.

[24] M. Ghahroudi, M. Sarshar, and R. Sabzevari, “A Novel Content-Based Image Retrieval Technique Using Tree Matching,” Proceedings of The World Congress on Engineering, pp1797-180, May 2008.

[25] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, and P. Yanker, “Query by image and video content: The QBIC system,” IEEE Computer, vol. 28, no 9, pp.23-32, Sep. 1995.

[26] D. Zhang, "Improving Image Retrieval Performance by Using Both Color and Texture Features," In Proc. of IEEE 3rd International Conference on Image and Graphics (ICIG04), Hong Kong, China, Dec.18-20, 2004, pp.172-175.

[27] J. Wang, J. Li, G. Wiederhold, “Simplicity: semantics-sensitive integrated matching for picture libraries,” IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 23, no. 9, pp. 947–963, Sep. 2001.

[28] W. Ma and B. Manjunath, “Natra: A Toolbox for Navigating Large Image Data bases,” Proc. IEEE Int'l Conf. Image Processing, Santa Barbara, 1997, pp. 568- 571.

[29] C. Carson, M. Thomas, S. Belongie, J.M. Hellerstein, and J. Malik, “Blob world: A System for Region-Based Image Indexing and Retrieval,” Proc. Visual Information Systems, pp. 509-516, June 1999.

[30] A. Natsev, R. Rastogi, and K. Shim, “WALRUS: A Similarity Retrieval Algorithm for Image Databases,” IEEE Trans. On Knowledge and Data Engineering, vol.16, pp. 301-318, Mar. 2004.

[31] M. Sudhamani, and C. Venugopal, “Image Retrieval from Databases: an Approach using Region Color and Indexing Technique,” International Journal of Computer Science and Network Security, vol.8 no.1, pp. 54-64, January 2008.

[32] S. Nandagopalan, B. Adiga, and N. Deepak, “A Universal Model for Content-Based Image Retrieval,” Proc. of world academy of science, engineering and technology, vol. 36, pp. 659-662, DEC. 2008.

[33] J. Li, J. Wang, and G. Wiederhold, “Integrated Region Matching for Image Retrieval,” In Proceedings of the 2000 ACM Multimedia Conference, Los Angeles, October 2000, pp. 147-156.

[34] R. Zhang, and Z. Zhang, “A Clustering Based Approach to Efficient Image Retrieval,” Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence (ICTAI’02), Washington, DC, Nov. 2002, pp. 339-346.

[35] R. Gonzales, R. E. Woods, “Digital Image Processing,” 2nd Ed., New Jersey Prentice Hall, 2002.

[36] T. Gevers and A. Smeulders, “Pictoseek: Combining color and shape invariant features for image retrieval,” IEEE Trans. Image Processing, vol. 9, no. 1, pp.102– 119, Nov. 2000

[37] A. Ouyang, and Y. Tan, “A novel multi-scale spatial-color descriptor for content-based image retrieval,” Proceedings of the 7th International Conference on Control, Automation, Robotics and Vision, Mexico, August 2002, vol. 3, pp. 1204-1209.

[38] H. Yu, M. Li, H. Zhang, and J. Feng, “Color texture moments for content-based image retrieval,” Proceedings of the International Conference on Image Processing, Rochester, New York, USA, September 22-25, 2002, vol. 3, pp. 929-932.

[39] T. Gevers, and H. Stokman, “Classifying color edges in video into shadow-geometry, highlight, or material transitions,” IEEE Transactions on Multimedia, vol. 5, no. 2, pp. 237-243, Sep. 2003.

[40] H. Guan, and S. Wada, “Flexible color texture retrieval method using multi-resolution mosaic for image classification,” Proceedings of the 6th International Conference on Signal Processing, vol. 1, pp. 612-615, Feb. 2002.

[41] H. Moghaddam, T. Khajoie, and A. Rouhi, “A new algorithm for image indexing and retrieval using wavelet correlogram,” Proceedings of the International Conference on Image Processing, vol. 3, pp. 497-500, May 2003.

[42] M. Lew, N. Sebe, C. Djeraba and R. Jain, “Content-Based Multimedia Information Retrieval: State of the Art and Challenges,” ACM Transactions on Multimedia Computing, Communications and Applications, vol. 2, no. 1, pp. 1-19, February 2006.

[43] J. Fuertes, M. Lucena, N. Perez, and J. Martinez, “A Scheme of Color Image Retrieval from Databases,” Pattern Recognition Letters, vol. 22, pp.323–337, June 2001.

[44] B. Manjunath and W. Ma, “Texture features for Browsing and retrieval of image data,” IEEE transactions on pattern analysis and machine intelligence, vol. 18. No. 8, pp. 837-842, August 1996

[45] R. Haralick, K. Shanmugam, and I. Dinstein. “Texture Features for Image Classification,” IEEE Trans. on Systems, Man and Cybernetics, SMC, vol.3, no 6, pp. 610–621, Nov. 1973.

[46] R. Datta, J. Li, and J. Wang, “Content-based image retrieval - approaches and trends of the new age,” ACM Computing Surveys, vol. 40, no. 2, Article 5, pp. 1-60, April 2008.

[47] M. Chantler, and J. Wu, “Rotation Invariant Classification of 3D Surface Textures using Photometric Stereo and Surface Magnitude Spectra,” Proceedings of British Machine Vision Conference, Vol.2, pp 486-495, Jan. 2002.

[48] R. Dubes, and A. Jain, “Random field models in image analysis,” Journal Applied Statistic, vol. 16, no. 2, pp.131-164, Nov. 1989.

[49] A. Pentland, “Fractal-Based Description of Natural Scenes,” IEEE Transaction on Pattern Analysis Machine Intelligence, vol. 6, no. 6, pp. 661-674, 1984.

[50] J. Mao, and A. Jain, “Texture Classification and Segmentation using Multi-Resolution Simultaneous Autoregressive Models,” Pattern Recognition, vol. 25, no. 2, pp. 173-188, 1992.

[51] C. Teh and T. Roland, “On image analysis by the methods of moments,” IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 10, no. 4, pp. 496-513, 1988.

[52] M. Hu. “Visual Pattern Recognition by Moment Invariants,” IEEE Transactions on Information Theory, IT, vol. 8, pp. 179-187, Feb. 1962.

[53] M. Bober, “MPEG-7 Visual Shape Descriptors”, IEEE Trans. on Circuits and Systems for Video Technology, vol. 11, no. 6, pp. 716-719, June 2001.

[54] http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm

[55] F. Mokhtarian, and R. Suomela, “Robust Image Corner Detection Through Curvature Scale Space”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1376-1381, December 1998

[56] G. Qian, S. Sural, Y. Gu, and S. Pramanik, “Similarity between Euclidean and cosine angle distance for nearest neighbor queries,” Proceedings of ACM Symposium on Applied Computing, vol. 12, no. 22, pp. 1232-1237, 2004.

[57] M. Swain, D. Ballard, “Color indexing,” International Journal of Computer Vision vol. 7, no 1, pp 11-32, Nov. 1991.

[58] J. Laaksonen, M. Koskela, S. Laakso, and E. Oja, "Picsom-content-based image retrieval with self-organizing maps," Pattern Recognition Letters, vol. 21, pp. 1199– 207, Feb. 2000.

[59] Yang Mingqiang, Kpalma Kidiyo, Ronsin Joseph. A survey of shape feature extraction techniques, Pattern Recognition, Peng-Yeng Yin (Ed.) pp. 43-90, 2008.

[60] Asmatullah Chaudhry, Javed Ullah, M. Arfan Jaffar, Jin Young Kim, Tran Anh Tuan. Human Activity Recognition System: Using Improved Crossbreed Features and Artificial Neural Network. Life Sci J, vol. 9, no. 4, pp. 5351-5356, 2012.

[61] Saptadi Nugroho and Darmawan Utomo. Rotation Invariant Indexing For Image Using Zernike Moments and R–Tree. TELKOMNIKA, Vol.9, No.2, pp. 335-340, 2011.

[62] Rahman M.M., Bhattacharya M.P. and Desai B.C. A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. IEEE Trans. Inform. Technol. Biomed., Vol. 11, No. 1, pp.58–69, 2007.

[63] Felci Rajam I. and Valli S. SRBIR: semantic region based image retrieval by extracting the dominant region and semantic learning. Journal of Computer Science, Vol. 7, No. 3, pp.400–408, 2011.

[64] Felci Rajam I. and Valli S. Region-based image retrieval using the semantic cluster matrix and adaptive learning. International Journal of Computational Science and Engineering, Vol. 7, No. 3, pp.239–252, 2012.

[65] Deselaers T, Keysers D, Ney H. Features for image retrieval: an experimental comparison. Inf. Retr. 11(2), pp. 77–107, 2007.

[66] Md. Mahmudur Rahman, Bipin C. Desai, and Prabir Bhattacharya. Supervised Machine Learning Based Medical Image Annotation and Retrieval in Image CLEF med 2005. Proceedings of the CLEF 2005, LNCS 4022, pp. 692–701, 2006.

[67] Felci Rajam I. and Valli S. SRBIR: semantic region based image retrieval by extracting the dominant region and semantic learning. Journal of Computer Science, Vol. 7, No. 3, pp.400–408, 2011.

[68] Felci Rajam I. and Valli S. Region-based image retrieval using the semantic cluster matrix and adaptive learning. International Journal of Computational Science and Engineering, Vol. 7, No.3, pp.239–252, 2012.

[69] M. Zhenjiang, "Zernike moment - based image shape analysis and its application", Pattern Recognition Letters, Vol. 21, pp. 169-177, 2000.

[70] A. Khotanzad, Y.H.Hong, "Invariant image recognition by Zernike moments", IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(5), pp. 489-498, 1990.

[71] J. D. Zhou, H. Z. Shu, L. M. Luo, W. X. Yu, "Two new algorithms for efficient computation of Legendre moments", Pattern Recognition, Vol.35(5), pp. 1143-1152, 2002.

[72] P. T. Yap and R. Paramesaran, "An Efficient Method for the Computation of Legendre Moments," IEEE Transactions on Pattern Recognition and Machine Intelligence, Vol.27, No.12, pp.1996-2002.

[73] S. X. Liao, M. Pawlak, "On image analysis by moments," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18(3), pp. 254-266, 1996.

[74] Khalid M. Hosney, "Exact Legendre moments computation for gray level images", Pattern Recognition, Vol. 40, pp. 3597-3605, 2007.

[75] Liu Y., Zhang D., Lu G. and Ma W.Y. A survey of content-based image retrieval with high-level semantics. Patt. Recog. 40, pp. 262-282, 2007.

[76] Aymn E.Khedr and Abd El-Ghany A. M. Mohmed. A proposed image processing framework to support early liver Cancer Diagnosis. Life Sci J, vol. 9, no. 4, pp. 3808-3813, 2012.

[77] Thenmozhi. S, Balasubramanie. P, Venkatesh. J, Aarthy. C. Detection of reflection in iris images using back propagation. Life Sci J, vol. 9, no. 3, pp. 2446-2450, 2012.

[78] C. Campbell, "Algorithmic Approaches to Training Support Vector Machines: A Survey, ESANN, D-Facto Publications, pp. 27- 36, 2000.

[79] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, Vol. 2, pp. 121-167, 1998.

[80] K. K. Seo, "An application of one-class support vector machines in content-based image retrieval," Expert Systems with Applications, Vol. 33(2), pp.491-498. 2007.

[81] R. Datta, D. Joshi, J. Li, and J. Z. Wang,"Image Retrieval: Ideas, Influences, and Trends of the New Age", ACM Computing Surveys, Vol. 40, No.2, Article 5, 2008.

[82] Liu Y., Zhang D., Lu G. and Ma W.Y. A survey of ontent-based image retrieval with high-level semantics. Patt. Recog. 40, pp. 262-282, 2007.

[83] Zakariya S. M., Rashid Ali and Nesar Ahmad. Content Based Image Retrieval by Combining Visual Features of an Image with A Threshold. Special Issue of IJCCT Vol. 2, No. 2, 3, 4, pp. 204-209, 2010.

[84] Celia B, Felci Rajam I. An efficient content based image retrieval framework using machine learning techniques. Proceedings of the Second international conference on Data Engineering and Management (ICDEM ‗10), Springer LNCS, Vol. 6411, pp 162- 169, 2010.

[85] N. Vasconcelos. Bayesian models for visual information retrieval. Ph.D thesis, Massachusetts Institute of Technology, 2000.

[86] N. Vasconcelos and M. Kunt. Content-based retrieval from image databases: current solutions and future directions. In International Conference in Image Processing (ICIP'01), volume 3, pages 6–9, Thessaloniki, Greece, October 2001.

[87] D. Michie, D. J. Spiegelhalter, and C. C. Taylor, editors. Machine Learning, Neural and Statistical Classification. Ellis Horwood, 1994.

[88] C.Fraley and A. Raftery, "Model based clustering discriminate analysis and density estimation", journal of American statistics association, 97, 611-631 ,2002.

[89] E. Dahlhaus, "Parallel algorithms for hierarchical clustering and applications to split decomposition and parity graph recognition," J. Algorithms, vol. 36, no. 2, pp. 205–240, 2000.

[90] C. Olson, "Parallel algorithms for hierarchical clustering," Parallel Comput., vol. 21, pp. 1313–1325, 1995.

[91] K. Stoffel and A. Belkoniene, "Parallel K-means clustering for large data sets," in Proc. EuroPar'99 Parallel Processing, 1999, pp.1451–1454.

[92] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin. "The qbic project: Quering images by content using color,

texture, and shape.” In Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases, 2□3 February ’93, San Jose, CA, pages 173–187, 1993.

[93] Chad Carson and Virginia E. Ogle. “Storage and retrieval of feature data for a very large online image collection”. IEEE Computer Society Bulletin of the Technical Committee on Data Engineering, 19(4):19–27, December 1996.

[94] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Nib□ lack. “Effcient color histogram indexing for quadratic form dis□ tance functions.” IEEE Transactions on Pattern Analysis and Ma□ chine Intelligence, 17(7):729-736, 1999

[95] M. Stricker and M. Orengo. “Similarity of color images”. In W. Niblack and R. Jain, editors, Storage and Retrieval for Image and Vid□ eo DatabasesIII (SPIE), volume 2420, pages 381{392, San Diego/La Jolla, CA, USA, February 1995.

[96] MichaelOrtega,Yong Rui, Kaushik Chakrabarti, Sharad Mehro□ tra, and Thomas S. Huang. “Supporting similarity queries in MARS”. In Proceedings of the 5th ACM International Multimedia Conference, Seattle, Washington, 8□14 Nov. ’97, pages 403–413, 1997.

[97] F. Mokhtarian, S. Abbasi, and J. Kittler. “Efficient and robust retrieval by shape content through curvature scale space”. In Smeulders and Jain, pages 35–42

[98] Sougata Mukherjea, Kyoji Hirata, and Yoshinori Hara. “To□ wards a multimedia world wide web information retrieval en□ gine.” In Sixth International WWW Conference, 7□11 April ’97, San□ ta Clara, CA, USA, 1997.

[99] Fumikazu Kanehara, Shin’ichi Satoh, and Takashi Hamada. “A flexible image retrieval using explicit visual instruction.” In Proceedings of the Third International Conference on Document Analysis Recognition, Montreal, Canada, August ’95, pages 175– 178, 1995.

[100] Pentland A et al (1996) “Photobook: tools for content□based manipulation of image databases” International Journal of Computer Vision 18(3), 233□254

[101] J. R. Smith and S. F. Chang, “Transform features for texture classification and discrimination in large image databases”, in Proc. IEEE Int. Conf. on Image Proc., 1994

[102] S. Michel, B. Karoubi, J. Bigun, and S. Corsini. “Orientation radiograms for indexing and identification in image databases” In Proceedings Eusipco□96, European conference on signal processing, pages 1693–1696, 1996.

[103] B. S. Manjunath and W. Y. Ma. “Texture features for browsing and retrieval of large image data” IEEE Transactions on Pattern Analysis and Machine Intelligence, (Special Issue on Digital Libra□ ries), Vol. 18 (8), August 1996, pp. 837□842.

[104] George Tzagkarakis and Panagiotis Tsakalides. “A Statistical approach to texture image retrieval via alpha□stable modeling of wavelet decompositions”. in Proc. 5th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS □04), Lisbon, Portugal, April 21□23, 2004.

[105] C. Yang, M. Dong, and F. Fotouhi. “S□iras: An interactive se□ mantic image retrieval and annotation system”. International Journal on Semantic Web and Information Systems, 2(3):37{54, 2006.

[106] M. Rege, M. Dong, and F. Fotouhi. “Building a user□centered semantic hierarchy in image databases”. Multimedia Systems, 12(45):325{338, March 2007.

[107] Y. Y. Lin, T.L. Liu and H.T. Chen. “Semantic manifold learning for image retrieval.” In Proceedings of the ACM International Conference on Multimedia, 2005

[108] Janghyun Yoon and Nikil Jayant. “Relevance Feedback for semantics based Image retrieval”

[109] S.F. Chang, W. Chen, H. Sundaram, “Semantic visual templates: linking visual features to semantics”, International Conference on Image Processing (ICIP), Workshop on Content Based Video Search and Retrieval, vol. 3, October 1998, pp. 531–534.

[110] http://www.iath.virginia.eduinote

[111] http://www.facebook.com

[112] http://images.google.com/imagelabeler/

[113] Bradshaw B. “Semantic based image retrieval: a probabilistic approach” Proceedings of the ACM International Conference on Multimedia, Los Angeles, California, Oct. 30Nov.4, 2000; 167-176.

[114] Ghoshal A, Ircing P, Khudanpur S. “Hidden Markov models for automatic annotation and contentbased retrieval of images and video”. Proceedings of the 28th International ACM SIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil, Aug 2005; 1519: 544551.

[115] Ilaria Bartolini and Paolo Ciaccia. “Towards an Effective Semi Automatic Technique for Image Annotation”

[116] Huang J., Kumar SR., Zabih R. “An automatic hierarchical image classification”. Proceedings of the ACM International Conference on Multimedia, Bristol, UK, Sep. 1216, 1998; 219228.

[117] H. Feng and T.S. Chua. “A bootstrapping approach to annotating large image collection.” Proceedings of the Fifth ACM SIGMM international workshop on Multimedia information retrieval (MIR 03), pages 55{62, Berkeley, California, USA, November 2003.

[118] S. Gao, D.H. Wang, and C.H. Lee. “Automatic image Aannotation through multitopic text categorization”. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006), volume 2, pages II{377{II{380, Toulouse, France, May 2006.

[119] Y. Mori, H. Takahashi, and R. Oka. “Imagetoword transformation based on dividing and vector quantizing images with words”. In Proceedings of the First International Workshop on Multimedia Intelligent Storage and Retrieval Management, Orlando, Florida, October 1999.

[120] P.L. Stanchev, D. Green Jr., B. Dimitrov, “High level color similarity retrieval”, Int. J. Inf. Theories Appl. 10 (3) (2003) 363–369.

[121] V. Mezaris, I. Kompatsiaris, M.G. Strintzis, “An ontology approach to objectbased image retrieval”, Proceedings of the ICIP, vol. II, 2003, pp. 511–514.

[122] Huan Wang, Song Liu Liangtienchia. “Image retrieval with a multimodality ontology”. Multimedia System (2008).

[123] Rahman M.H., Pickering M.R., Frater M.R. Scale and Rotation Invariant Gabor Features for Texture Retrieval. IEEE International Conference on Digital Image Computing Techniques and Applications (DICTA), pp. 602-607, 2011.

[124] Lining Zhang, Lipo Wang and Weisi Lin. Generalized Biased Discriminant Analysis for Content-Based Image Retrieval. IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, Vol. 42, No. 1, pp. 282-290, 2012.

[125] Imtnan-Ul-Haque Qazi, OlivierAlata, Jean-Christophe Burie, Ahmed Moussa, ChristineFernandez-Maloigne. Choice of a pertinent color space for color texture characterization using parametric spectral analysis. Pattern Recognition 44, pp. 16–31, 2011.

[126] Hatice Cinar Akakin and Metin N. Gurcan. Content-Based Microscopic Image Retrieval System for Multi-Image queries. IEEE Transactions on Information Technology.

[127] Ja-Hwung Su, Wei-Jyun Huang, Philip S. Yu, Fellow, and Vincent S. Tseng. Efficient Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patterns. IEEE Transactions on Knowledge and Data Engineering, Vol. 23, No. 3, pp. 360-372, 2011.

[128] Yu-Gang Jiang, Qi Dai, Jun Wang, Chong-Wah Ngo, Xiangyang Xue and Shih-Fu Chang. Fast Semantic Diffusion for Large-Scale Context-Based Image and Video Annotation. IEEE Transactions on Image Processing , Vol. 21, No. 6, pp. 3080 – 3091, 2012.

[129] Wang Xing-Yuan, ChenZhi-feng, YunJiao-jiao. An effective method for color image retrieval based on texture. Computer Standards & Interfaces 34, pp. 31–35, 2012.

[130] Samuel Rota Bul_o, Massimo Rabbi and Marcello Pelillo. Content-Based Image Retrieval with Relevance Feedback using Random Walks. Pattern Recognition, Vol. 44, No. 9, pp. 2109-2122, 2011.

Robust Digital Watermarking Scheme of Anaglyphic 3D for RGB Color Images

Chunxing Wang cxwang@sdnu.edu.cn
College of Physics & Electronics,
Shandong normal university
No.88 of East Wenhua Road, Jinan, 250014, P.R. China

Fei Han hanfeilove2010@126.com
College of Physics & Electronics,
Shandong normal university
No.88 of East Wenhua Road, Jinan, 250014, P.R. China

Xiaomei Zhuang zhuangxiaomei@163.com
College of Physics & Electronics,
Shandong normal university
No.88 of East Wenhua Road, Jinan, 250014, P.R. China

Abstract

In this paper, digital watermarking technique using spread spectrum (SS) technology and adaptive DM (dither modulation) with the improved Watson perception model are applied for copyright protection of anaglyphic 3D images. The improved Watson perception model can well solve the problem that the slack do not change linearly as the amplitude scale. Experimental results show that the watermarking schemes provide resistance to Gaussian noise, salt and pepper noise, JPEG compression, constant luminance change and valumetric scaling; the scheme employing improved Watson perception model is better than the one using unimproved Watson perception model. Compared experiments with the works [4] and [19] were also carried out in experiments. On the other hand, the approach is not sensitive to the JPEG compression while the other based on QIM is not sensitive to constant luminance change and valumetric scaling.

Keywords: Digital Watermarking, RGB, Anaglyphic 3D Images, Spread Spectrum (SS) Watermarking Technique, Quantization Index Modulation (QIM), Robust Watermarking.

1. INTRODUCTION

The digitization, storage, transmission and reproduction of multimedia resources have become very convenient. However, the accompanying piracy problem emerges increasingly. Therefore, digital watermarking technology is used to address this issue effectively at home and abroad.

Spread spectrum (SS) technology in the 1950s originated in the communication system, and first used for military communications. It has the advantages of anti-interference, confidentiality, low density of power spectrum and high-precision measurement, so digital watermarking technology employing the spread spectrum principle has high robustness and security. The first digital watermarking algorithm (NEC algorithm) based on spread spectrum thought is proposed by Coxetal [1-8] in 1996. At present, watermarking algorithms based on spread spectrum have been applied to images, video, audio, text and other carriers, and the transformation also involves DWT, DCT and DFT.

Quantization index modulation (QIM) is proposed by Chen and Wornell [9] originally. And they have improved the QIM by using Costa's ideas, resulting in distortion-compensated QIM, (DC-QIM) [10]. Recently, the specific QIM methods are dither modulation (DM) and spread transform dither modulation (STDM) [11]. The achievable watermarking algorithm of DC-QIM is distortion-compensated dither modulation (DC-DM) [12]. Other watermarking algorithms using quantization are references [13-17]: quantize carrier data for embedding watermarking information; modify the carrier data according to quantized value and the watermarking to be embedded in order to achieve the embedding of watermarking information.

In this paper, spread spectrum watermarking algorithms and quantization index modulation watermarking algorithm are used for anaglyphic 3D images.

2. DIGITAI WATERMARKING SCHEME FOR ANAGLYPHIC 3D IMAGES BASED ON SPREAD SPECTRUM

The human eyes get stereoscopic sense for the nuances of the objects seen by the left and right eye. To have three-dimensional sense from a planar image, the image must contain the information of two images with a certain parallax. The two images with optic difference will be sent to our left and right eyes by appropriate means respectively.

Any pixel in the color image can be expressed and recorded using a set of RGB values. Generally, color images require information of at least three dimensions. The color image needs to be changed to a grayscale image when embedding watermarking image. After finishing embedding the watermarking image, the grayscale image should restore the original color image. Therefore, the paper has taken the following two decomposition methods:

(1)The RGB is converted to YIQ, and the Y component is the equivalent of grayscale data. Only the value of gray-scale image is an integer from 0 to 255, but Y is the real number of 0 to 1. At the same time, YIQ can restore the original RGB image.

(2) The RGB is converted to RGg, the gray-scale component g replaces the blue component B, and the red component R and the green component G keep unchanged. Because the gray level is derived by $g = p * R + q * G + t * B$, where p=0.2989，q=0.5870，t=0.1140, then $B = (g - p * R - q * G)/t$.

Shannon summarized channel capacity formula from the information theory; the formula can be written by:

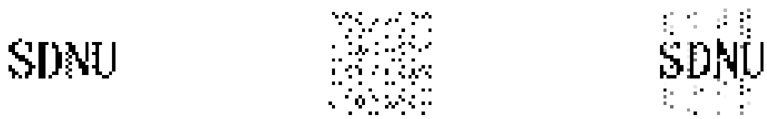
$$C = W \log_2 (1 + \frac{\delta_X^2}{\delta_n^2}) \tag{1}$$

Where C represents the channel capacity; W is the channel bandwidth; δ_X^2 denotes signal power and δ_n^2 means noise power. The formula shows the relationship between the ability of error-free transmission of information with channel SNR. The procedures of this watermarking algorithm for anaglyphic 3D images are as follows: watermarking is embedded in the maximal N frequency coefficients in the DCT domain; then modify the frequency coefficients to achieve the embedding of watermarking information, and we use the addition criterion [1-2]in this paper. The nature of watermarking algorithm based on SS determines that the original host image is needful for extracting the watermarking if the original meaningful watermarking information is embedded directly without pseudo-random sequence modulation. In order to achieve the blind extraction, the watermarking is modulated with pseudo-random sequence before embedding the watermarking image. After pseudo-random modulation, watermarking image is embedded into the DCT coefficients of 3D image synthesized by two images with certain optic differences.

Figure 1 shows the watermarked image by use of the decomposition method (1) above; the original watermarking and extracted watermarking image are shown in Figure 2. RGB1 and RGB2 are the two images with optic difference. The NC values of extracting watermarking resisting various attacks is listed in Table 1, the watermarked image with PSNR = 41.8847. Where the mean and variance of Gaussian noise is zero and 0.001 respectively; density of salt and pepper noise is taken as 0.001; quality factor of JPEG is 80%.

(a)RGB1 (b)RGB2 (c)Anaglyphic 3D image (d)Watermarked 3D image

FIGURE1: Watermarked Image of The Scheme using SS Technology.



(a)Original watermarking Image (b)Arnold scrambling watermarking (c)Extracted watermarking

FIGURE 2: Extracted Watermarking Images of The Scheme using SS Technique.

Attacks type	No attack	Gaussian noise	Salt & pepper noise	Shrink twice	JPEG compression	Median filtering	Luminance change	Valumetric scaling
NC	1	0.9113	0.9679	0.9950	0.9921	0.9798	0.9967	0.8429

TABLE 1: The NC Values For Extracted Watermarking After Attacks, PSNR=41.8847 dB.

3. WATERMARKING SCHEME FOR ANAGLYPHIC 3D IMAGES BASED ON ADAPTIVE DM WITH THE IMPROVED WATSON PERCEPTION MODEL

Quantization index modulation [9] means modulating one or a series of indexes by use of watermarking to be embedded, and then to quantify the carrier signal by means of corresponding quantizer or quantizer sequence. Quantization can be expressed by

$$y = \Delta * round(\frac{x}{\Delta}) \tag{2}$$

Where Δ denotes that the quantization step, $round(.)$ indicates rounding operation. For the binary watermarking, binary 0 and 1 correspond to two indexes, and the two indexes respond to two quantizers. The function of embedding watermarking can be expressed by

$$\vec{S}(\vec{x};\vec{m}) = q(\vec{X};\vec{m},\Delta) \tag{3}$$

Where Δ means the quantization step size; $q(\vec{X};\vec{m},\Delta)$ denotes the m-th quantizer function with the quantization step size Δ. In the processing of extracting the watermark, minimum distance decoding or maximum likelihood decoding [18] can be adopted.

3.1 Dither Modulation (DM)

DM is a special method of QIM. In order to embed information, dither value may be modulated by the watermarking having been embedded. All available watermarking information to be embedded will map to different dither values. Carrier signal after dither generate the synthesized signal by quantifying. In the case of the basic quantizer $q(\cdot)$, embedding function can be expressed by

$$S(k) = q(X(k) + d[k,b_k]) - d[k,b_k] \tag{4}$$

Here, $d[k,b_k]$ presents the k-th dither when the embedding watermarking bit is b_k; $q(\cdot)$ denotes the quantizer; $X(k)$ indicates the original signal; $S(k)$ means the signal after quantization index modulation. Suppose the embedding watermarking information is a binary sequence, the dither value $d(k,0)$ is a pseudorandom signal usually chosen with a uniform distribution between $[-\frac{\Delta}{2},\frac{\Delta}{2}]$, but the dither $d(k,1)$ should be selected according to the following formula:

$$\begin{cases} d[k,1]=d[k,0]+\frac{\Delta_k}{2}, d[k,0]<0 \\ d[k,1]=d[k,0]-\frac{\Delta_k}{2}, d[k,0]>0 \end{cases} \quad k=1,2,\ldots L \tag{5}$$

3.2 Watson Perceptual Model

Watson perception model, a visual fidelity model put forward by Watson in 1993 [14], based on the block discrete cosine transformation can estimate the perceptibility of images' change.

For Watson perception model, the original image is transformed by $8*8$ blocks, and the luminance masking threshold is shown as formula (6). Where, αT is a constant, usually the value is 0.649; $C_{0,0}$ indicates the mean value of DC coefficients of the original image; $C_0[0,0,k]$ means the DC coefficient of the k-th block; $t[i,j]$ is the sensitivity table defined by Watson perception model that reflecting sensitive degree of the human eyes to different frequency.

$$t_L[i,j,k]=t[i,j](\frac{C_0[0,0,k]}{C_{0,0}})^{\alpha T} \tag{6}$$

The expression of contrast masking threshold is shown in formula (7). The threshold estimates the slacks which mean the change range of each DCT block within the limits of JND.

$$s[i,j,k]=\max(t_L[i,j,k],\left|C_0[i,j,k]\right|^{0.7} t_L[i,j,k]^{0.3}) \tag{7}$$

3.3 Watermarking scheme for anaglyphic 3D images based on adaptive DM with the improved Watson perception model

In order to solve the problem that the slack do not change linearly as the amplitude scale, we modify luminance masking according to formula (8), and the improved slack shown as formula (9). When the image is scale β times, the luminance masking threshold and slack are respectively shown as formula (10) and (11). The operation result shows that the improved slack change linearly with the amplitude scale. Then the adaptive quantization step size can be set based on the improved slack.

$$t_L'[i,j,k]=t_L[i,j,k](\frac{C_{0,0}}{128})=t[i,j](\frac{C_0[0,0,k]}{C_{0,0}})^{\alpha T}(\frac{C_{0,0}}{128}) \tag{8}$$

$$s'[i,j,k]=\max(t_L'[i,j,k],\left|C_0[i,j,k]\right|^{0.7} t_L'[i,j,k]^{0.3}) \tag{9}$$

According to the modified luminance masking threshold $t_L'[i,j,k]$ and slack $s'[i,j,k]$, when the image is resized β times, the luminance masking threshold $\overline{t_L'}[i,j,k]$ and slack $\overline{s'}[i,j,k]$ are resized β times, too.

$$\begin{aligned} &\overline{t_L'}[i,j,k] \\ &=t_L[i,j,k](\frac{\beta C_{0,0}}{128}) \\ &=t[i,j](\frac{\beta C[0,0,k]}{\beta C_{0,0}})^{\alpha T}(\frac{\beta C_{0,0}}{128}) \\ &=\beta t_L'[i,j,k] \end{aligned} \tag{10}$$

$$
\begin{aligned}
&\overline{s'}[i,j,k] \\
&= \max(\overline{t'_L}[i,j,k], |(\beta C_0[i,j,k])|^{0.7}\,\overline{t'_L}[i,j,k]^{0.3}) \\
&= \max(\beta t'_L[i,j,k], \beta^{0.7}|C_0[i,j,k]|^{0.7}\beta^{0.3}t'_L[i,j,k]^{0.3}) \\
&= \beta \max(t'_L[i,j,k], C_0[i,j,k]^{0.7}t'_L[i,j,k]^{0.3}) \\
&= \beta s'[i,j,k]
\end{aligned} \tag{11}
$$

We employ the method of improved adaptive dither modulation and the decomposition method (2) mentioned above to embed the watermarking into the middle frequency coefficients in the DCT domain. The minimum distance detection is used to extract the watermarking. RGB1 and RGB2 are the two images with visual difference; Figure 3 shows the anaglyphic 3D image and watermarked anaglyphic 3D image; the original watermarking image and the extracted watermarking image are shown in Figure 4. PSNR and BER are as a function of image fidelity and accuracy of the extracted watermarking respectively. Table 2 shows the values of BER of extracted watermarking resistance to a variety of attacks when PSNR = 41.7612 dB. Where mean and standard deviation of Gaussian noise are zero and 0.01 respectively; density of salt and pepper noise is 0.001; quality factor of JPEG compression is 80%.

(a)RGB1 (b)RGB2 (c)Anaglyphic 3D image (d)Watermarked 3D image

FIGURE 3: Watermarked Image of The Scheme using Adaptive DM.

(a)Original watermarking image (b)Extracted watermarking image

FIGURE 4: Extracted Watermarking Images of The Scheme using Adaptive DM.

Attacks type	**No attack**	**Gaussian noise**	**Salt & pepper noise**	**JPEG**	**Shrink twice**	**Low-pass filtering**	**Cutting**	**Luminance change**	**Valumetric scaling**
BER	0	0	0.0293	0	0.0195	0.0684	0.0205	0	0.0117

TABLE 2: The BER values for extracted watermarking against attacks, PSNR=41.7612 dB.

It can be seen from the figures and tables illustrated above have good performance in image fidelity and robustness against common signal processing or attacks. Therefore, the two schemes meet the requirements of digital watermarking scheme, which are practicable.

4. SIMULATION AND CONTRAST EXPERIMENTS

In order to denote the robustness of the watermarking schemes, we compare with robust performance of three schemes under the same attacks. Here, SS-3D is the watermarking scheme for anaglyphic 3D images based on the spread spectrum technique; DM-mW-3D means the watermarking scheme for anaglyphic 3D images based on adaptive DM with improved Watson perception model; DM-W-3D indicates the watermarking scheme for anaglyphic 3D images based DM with unimproved Watson perception model. In fairness, both of them employ decomposition method (1) above; adjust PSNR values of the watermarked images to 41.8 dB. Calculate the NC values of extracted watermark to measure the robustness performance.

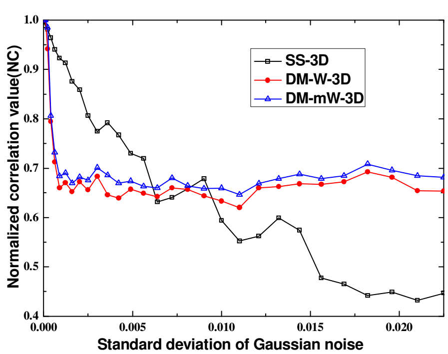
Under the mean of 0 of additive Gaussian noise. The graph of NC values in the two schemes with Gaussian noise is demonstrated in Figure 5 (a) With the increase of standard deviation of additive Gaussian noise, NC values in scheme SS-3D degrade sharply while the change of NC values of scheme DM-mW-3D is stable slightly. Figure 5 (a) shows that the watermarking scheme SS-3D is more sensitive to the Gaussian noise than scheme DM-mW-3D. Scheme DM-mW-3D is more robust against Gaussian noise than scheme SS-3D. In addition, it is clear that resistance to noise of the scheme DM-mW-3D is better than scheme DM-W-3D.

Figure 5 (b) shows the sensitivity of the two schemes to JPEG compression attack for a fixed PSNR of 41.8 dB. That is, we maintain fixed image fidelity. In this case, the sensitivity of watermarking schemes is tested. The performance of watermarking scheme DM-mW-3D is worse than the scheme SS-3D, but better than the scheme DM-W-3D. This shows that our improved scheme is necessary. Scheme SS-3D, for quality factors of less than 63%. Scheme SS-3D, the value of NC is greater than 0.95. There is still a good performance. However, NC values of the extracted watermarking in scheme DM-mW-3D descend sharply when the quality factors are less than 63%. The scheme SS-3D is not sensitive to JPEG compression noise. Because JPEG compression mainly affects the high frequency coefficients so the scheme SS-3D has larger embedded capacity in the medium-high frequency coefficients. Figure 5 (b) shows that the scheme DM-mW-3D for embedded capacity in the low frequency is larger than in high frequency.

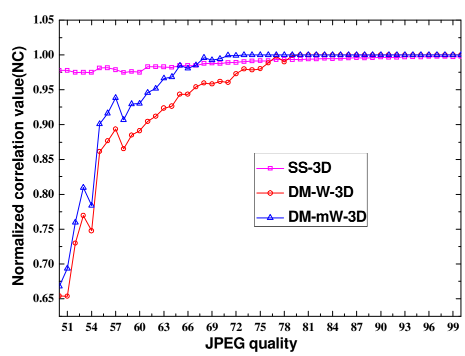
Figure 5(c) illustrates the sensitivity of the schemes to the addition/subtraction of a constant luminance value. Of course, the PSNR is fixed at 41.8 and NC is as a function of extracted watermarking. Watermarking scheme DM-mW-3D and DM-W-3D are significantly superior to scheme SS-3D, which is demonstrated in Figure 5 (c). Scheme SS-3D has a good performance only for variations of luminance change from -45 to 30, while scheme DM-mW-3D is not sensitive to constant luminance change. Compare the scheme SS-3D with DM-mW-3D, the scheme SS-3D attacking robustness of constant luminance changes less stable. This may be caused by the smaller quantization step length.（quantization step is 14）

The curve of NCs of valumetric scaling are shown in Figure 5 (d). It can be found that the scheme DM-mW-3D is better than the scheme DM-W-3D. NC values of extracted watermarking of scheme DM-mW-3D are greater than the NC values of scheme SS-3D. What's more, scheme DM-mW-3D has a good robustness reflecting in the valumetric scaling factors from 0.8 to 1.8. While scheme SS-3D is extremely sensitive to valumetric scaling, and the change of NC values for extracted the watermark is sharp.

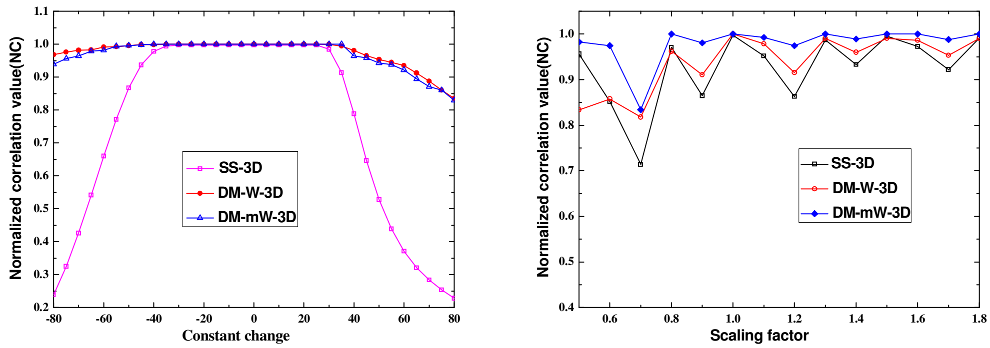
In short, overall performance of our watermarking scheme for anaglyphic 3D images based on adaptive DM with improved Watson perception model is superior to the watermarking scheme for anaglyphic 3D images based DM with unimproved Watson perception model. The data analysis about that are performed in the following paragraph.



(a) The curve of NCs of additive white Gaussian for fixed PSNR of 41.8 dB.



(b) The curve of NCs of JPEG quali noise for fixed PSNR of 41.8 dB



(c) The curve of NCs of constant luminance change for fixed PSNR of 41.8 dB.

(d) The curve of NCs of volumetric scaling for fixed PSNR of 41.8 dB.

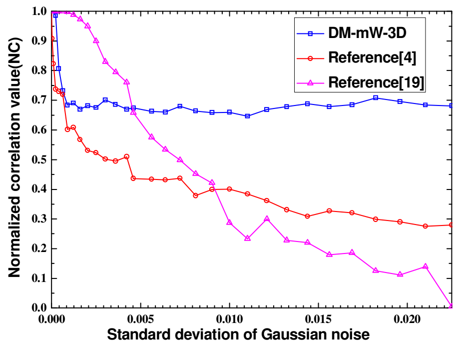
FIGURE 5: Robust Tests of Attacks.

In addition, the contrast experiment between the scheme DM-mW-3D of our paper and the reference [4] and [19] was made; the watermarking scheme of reference [19] is used for video, but the video is incompressible as image sequences, so the comparison experiment is feasible. In the experiments, simulation tests on Gaussian noise, constant luminance change, JPEG compression and valumetric scaling are carried out respectively when the same watermarking invisibility is guaranteed, namely the values of PSNR are all about 41.8 dB, and the results are shown in Figs. 6(a), 6(b), 6(c) and 6(d).

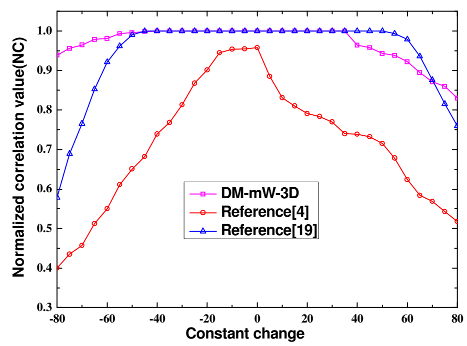
Figure 6 (a) shows that the resistance against Gaussian noise of the paper's scheme is superior to reference [4] and [19]. Reference [4] is most sensitive to Gaussian interference while reference [19] has strong resistance only under the low intensity of Gaussian interference. With the increase of the attack strength, NC values of extracting the watermarking fall sharply. It can be seen in the Figure 6 (b) that the solution of our paper can extract the watermarking certainly when attack parameters of constant luminance change are between -80 and 80 while reference [4] can hardly extract the watermarking and reference [19] can extract the watermarking only the parameters about the attack between -65and 65. There is no doubt that the effect of scheme DM-mW-3D is better than reference [19] .

Figure 6 (c) shows the results of JPEG compression attack, and the methods of our scheme and reference [19] perform better than reference [4]. It is regret that the performance of our scheme is not better than the one of reference [19] when strength of JPEG compression is larger.

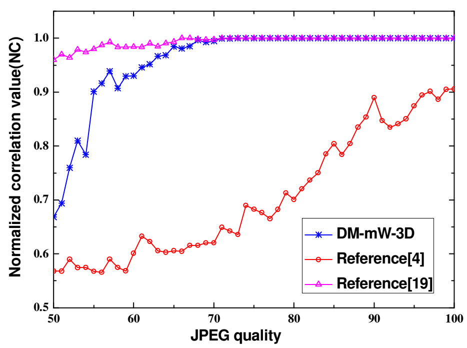
The curve of NCs of valumetric scaling are shown in Figure 6 (d). We found that the NC values of literature [4] fluctuate up and down with the increase of attack strength; that may be due to the image is saved as uint8 format and uint8 data format conversion is similar to the noise attack. In addition, NC values of reference [19] are all 1 when the scaling factors are between 0.5 and 1.8, so we increased the attack strength. It can also extract the watermarking when the scaling factor is about 0.3, and NC value is 0.3256 when the scaling factor is 0.2, namely the watermarking have not been extracted. The resistance against JPEG compression attack of the paper is not stronger than reference [19], but also it has good performance for the compression factor exceed to 63, that means the scheme is effective within restricted attack strength.



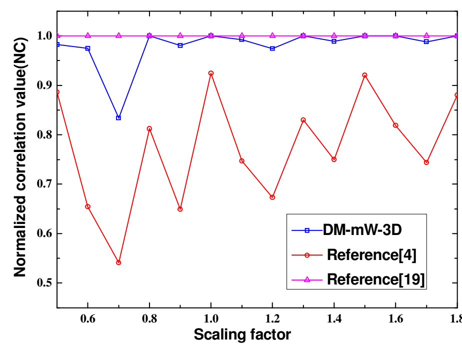
(a) The curve of NCs of additive white Gaussian noise for fixed PSNR of 41.8 dB.



(b) The curve of NCs of constant luminance change for fixed PSNR of 41.8 dB.



(c) The curve of NCs of JPEG quality for fixed PSNR of 41.8 dB.



(d) The curve of NCs of valumetric scaling for fixed PSNR of 41.8 dB

FIGURE 6: Comparison Experiment Results.

5. CONCLUSIONS

In this paper, we proposed two digital watermarking schemes for anaglyphic 3D images. The first one is watermarking scheme for anaglyphic 3D images based on SS. The results show that the watermarked anaglyphic 3D images have high fidelity, and the PSNR is up to 41.8847dB. Under the premise of such high fidelity, this scheme has good robustness which could be against common signal processing or attack. This scheme has certain contribution to the copyright protection of stereo image.

Secondly, another watermarking scheme for anaglyphic 3D images is based on adaptive DM with improved watson perception model. The modified gap can be used to calculate the adaptive quantization step size, and we apply this scheme to the watermarking technology for 3D images. The experimental results demonstrate that watermarked 3D images still have satisfactory fidelity and good robustness as well. BER values of extracted watermarking resistance to Gaussian noise, salt and pepper noise, JPEG compression, constant luminance change and valumetric scaling are less than 0.07.

Finally, We fixed the same image fidelity and used the same color basis decomposition method. The sensitivity of the three watermarking schemes is tested for Gaussian noise, JPEG compression, constant luminance adjustment, and valumetric scaling. The experimental results demonstrate that there is an improvement in the performance of the watermarking scheme for anaglyphic 3D images based on adaptive DM with improved Watson perception model. In this paper, the performance of the first watermarking scheme is stable to resist the JPEG compression attack. But the performance of the Gaussian noise and valumetric scaling is poor. This scheme is not sensitive to the intensity of the attack when the constant luminance change of the attack is relatively small. The performance of the second watermark scheme changed slowly and stably when we adjusted the constant luminance change and changed the attack intensity of valumetric scaling. There is an overall stability of the performance when subjected to Gaussian noise and JPEG compression.

The contrast experiment between the scheme DM-mW-3D of our paper and the reference [4] and [19] was made; the DM-mW-3D scheme can achieve satisfactory effect for Gaussian noise, constant luminance change and valumetric scaling attacks and the sensitivity for Gaussian of the paper is significantly lower than another two references. We have a plan of combining the two watermarking schemes. For example, the method in [20] is successful and typical. This is the future work we will carry out to improve the research.

6. ACKNOWLEDGEMENTS

The research is supported by the National Natural Science Foundation of China(Grant No. 11074151). We sincerely thank the editors for their kind help and valuable comments and suggestions which led to an improvement of this work.

7. REFERENCES

[1] I J Cox, J Kilian, T Leighton, T Shamoon. "Secure Spread Spectrum Watermarking for Images, Audio and Video." Proceedings of the 1996 IEEE International Conference in Image Processing. CA: IEEE, Los Alamitos, 1996, 3, pp. 243-246.

[2] I J Cox, J Kilian, T Leighton, T Shamoon. "Secure Spread Spectrum Watermarking for Multimedia." IEEE Transactions on Image Processing. 1997, 6(12), pp. 1673-1687.

[3] J W Huang, Y Q Shi. "Embedding Gray Level Images." The 2001 IEEE International Symposium on Circuits and Systems. May 6-9, 2001, 5, pp. 239-242.

[4] Y H Wu, X Guan, M S Kankanhalli. "Robust Invisible Watermarking of Volume Data Using 3D DCT." Proceedings of Computer Graphics International Conference Cgi,2001,1,pp. 359-362.

[5] J R Hernandez, M Amado, F P Gonzalez. "DCT-Domain Watermarking Techniques for Still Images: Detector Performance Analysis and a New Structure." IEEE Transactions on Image Processing, 2000, 9(1), pp. 55-68.

[6] T L Liu, H T Chen. "A Variational Approach for Digital Watermarking." International Conference on Image Processing. Piscataway: Institute of Electrical and Electronics Engineers Computer Society, 2000, 3, pp. 674-677.

[7] Y P Wang, M J Chen, P Y Cheng. "Robust Image Watermark with Wavelet Transform and Spread Spectrum Techniques." Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers. Pacific Grove: Institute of Electrical and Electronics Engineers Computer Society. 2000, 2, pp. 1846-1850.

[8] I Kitamura, S Kanai, T Kishinami. "Copyright Protection of Vector Map Using Digital Watermarking Method Based on Discrete Fourier Transform." IEEE 2001 International Geoscience and Remote Sensing Symposium. Piscataway: Institute of Electrical and Electronics Engineers Inc. 2001, 3, pp. 1191-1193.

[9] B. Chen and G. Wornell, "An information-theoretic approach to the design of robust digital watermarking systems," presented at the Int. Conf. Acoustics, Speech and Signal Processing, Phoenix, AZ, 1999, 3.

[10] B Chen, G W Wornell. "Preprocessed and Postprocessed Quantization Index Modulation Methods for Digital Watermarking." Security and Watermarking of Multimedia Contents II, Bellingham: Society of Photo-Optical Instrumentation Engineers, 2000, 3971, pp. 48-59.

[11] B Chen, G W Wornell. "Provably robust digital watermarking." Proceedings of the 1999 Multimedia Systems and Applications. Bellingham: Society of Photo-Optical Instrumentation Engineers, 1999, 3845, pp. 43-54.

[12] B Chen, G W Wornell. “Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding.” IEEE Transactions on Information Theory. 2001, 47(4), pp. 1423-1443.

[13] Y M Fang, JW Huang, Y Q Shi. “Image Watermarking Algorithm Applying CDMA. Proceedings of the 2003 International Symposium on Circuits and Systems.” Piscataway: Institute of Electrical and Electronics Engineers Inc, 2003, 2, pp. 948-951.

[14] M J Tsai, K Y Yu, Y Z Chen. “Joint Wavelet and Spatial Transformation for Digital Watermarking.” IEEE Transactions on Consumer Electronics, 2000, 46(1), pp. 241-245.

[15] M Ramkumar, A N Akansu. “A Robust Oblivious Watermarking Scheme. International Conference on Image Processing.” Vancouver: Institute of Electrical and Electronics Engineers Computer Society, 2000, 2, pp. 61-64.

[16] Z G Hu, P Xie and X M Zhang. “SVD Based Digital Watermarking Algorithm.” Computer Engineering, 2003, 29 (17), pp. 162-164.

[17] R Sun, H Sun, T R Yao. “SVD Based Semi-fragile Watermarking Method.” Journal of Circuits and Systems, 2002, 7 (3), pp. 62-65.

[18] Y Wang, J Xiao, Y H Wang. “Digital watermarking principle and technology.” Science Press, 2007, 3, pp. 87-88.

[19] Hui-Yu Huang, Cheng-Han Yang, and Wen-Hsing Hsu. “A Video Watermarking Technique Based on Pseudo-3-D DCT and Quantization Index Modulation.” IEEE transactions on information forensics and security, 2010, 5, (4), pp.625-637.

[20] C Fei, D Kundur, R H Kwong. “The Choice of Watermark Domain in the Presence of Compression.” Proceedings of the International Conference on Information Technology: Coding and Computing. Washington: IEEE Computer Society, 2001, pp. 79-84.

Efficient Image Compression Technique using JPEG2000 with Adaptive Threshold

Md. Masudur Rahman *masudur_2006@yahoo.com*
Mawlana Bhashani Science and Technology University
Santosh, Tangail-1902 (Bangladesh)

Mohammad Motiur Rahman *mm73rahman@gmail.com*
Mawlana Bhashani Science and Technology University
Santosh, Tangail-1902 (Bangladesh)

Abstract

Image compression is a technique to reduce the size of image which is helpful for transforms. Due to the limited communication bandwidth we have to need optimum compressed image with good visual quality. Although the JPEG2000 compression technique is ideal for image processing as it uses DWT (Discrete Wavelet Transform).But in this paper we proposed fast and efficient image compression scheme using JPEG2000 technique with adaptive subband threshold. Actually we used subband adaptive threshold in decomposition section which gives us more compression ratio and good visual quality other than existing compression techniques. The subband adaptive threshold that concentrates on denoising each subband (except lowest coefficient subbands) by minimizing insignificant coefficients and adapt with modified coefficients which are significant and more responsible for image reconstruction. Finally we use embedded block coding with optimized truncation (EBCOT) entropy coder that gives three different passes which gives more compressed image. This proposed method is compared to other existing approach and give superior result that satisfy the human visual quality and also these resulting compressed images are evaluated by the performance parameter PSNR.

Keywords: Image, JPEG2000, DWT, Adaptive Threshold, and EBCOT.

1. INTRODUCTION

A digital image is an array of pixels, where each pixel represents the intensity or brightness of the image at its respective location. Each pixel is a signed or unsigned integer and can be represented using bits. A gray scale image has one value per pixel location while a color image has three values per pixel location. Imaging is a very basic way for humans to convey information to one another and also express the emotion of a human voice. But it is very difficult to store and transmit the data if the size of the image is large. In this case image compression is performed. The only objective of image compression technique is to reduce the redundancy of the image data in order to store or transmit data in an efficient manner. Compression technique reduces the file size and allows to store more images in a given amount of disk or memory space[1][2].Generally two types of image compression technique are used for compressing the image those are lossy compression and lossless compression [3]. In case of lossy compression technique, the original signal cannot be exactly reconstructed from the compressed data, because of much of the detail in an image is discarded. Lossy compression provide high compression ratio but degrade the image quality. In a lossless compression, compressed data is used to recreate an exact replica of the original with no content loss to the compression process. Lossless compression provides good visualization but does not provide sufficient high compression ratio. That’s why we need the compression algorithm which will provide high compression ratio and also provide good visualization [4]. Generally for image compression, JPEG and JPEG2000 and some other compression algorithm are used. JPEG uses Discrete

Cosine Transform (DCT) which give high compression ratio but it gives lowest quality image. On the other side the JPEG2000 uses Discrete Wavelet Transform (DWT) that is based on sub-band technologies [5][6]. Comparatively it gives high compression ratio with good image quality. This paper described a modified technique based on JPEG200 that analyze the image as perform discrete wavelet transform in different level with adaptive threshold. This proposed technique provides better result than other existing techniques by improving compression ratio and image quality.

This paper is organized as follows: Section 2 discuss about the JPEG2000compression. In section 3 the proposed JPEG 2000 based compression with adaptive threshold is elaborated, while the experimental analysis and results are explained in section 4 followed by conclusion in section 5.

2. ABOUT JPEG2000

JPEG2000 is an international standard compression technology and it is a powerful new tool that provides power capabilities for designers and users of networked imaging applications. It is a compression standard enabling both lossless and lossy storage. This technique improves the quality and compression ratios, but also requires more computational power to process. JPEG2000 is a new wavelet based compression methodology [6] that provides many benefits over DCT compression method which was used in the JPEG format. The architecture of the basic JPEG2000 encoder and decoder are shown in Fig.1.

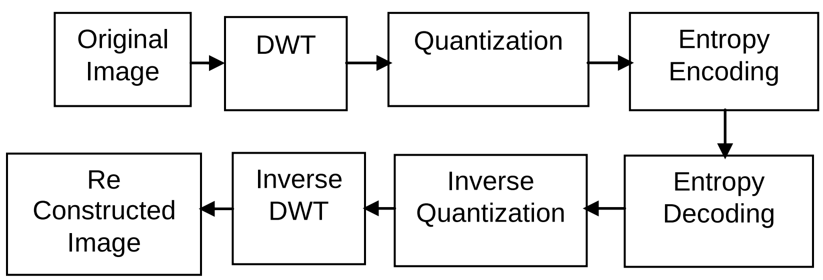


FIGURE 1: Basic Architecture of JPEG2000 Encoder and Decoder.

In JPEG2000, a DWT is used which transform the image into a series of wavelets that can be stored more efficiently than pixel blocks. After transformation, all coefficients are quantized using quantization. The quantized coefficients are entropy coded, and sent as output code stream (bit stream).

3. PROPOSED ARCHITECTURE

Digital Images are compressed through usage of various standards based techniques and algorithms. The lossless compression techniques are used in places where the quality and accuracy of image is of extreme important. Although the lossless compression not give high compression ratio, so we have given emphasize on compression to get the high ratio with respect to human visual quality and measured the quality by using the performance parameter like as PSNR. Finally to achieve a higher compression rate for lossless-compressed gray images, we propose a JPEG2000 based compression scheme incorporated with an adaptive threshold. The proposed method consists of four stages: Discrete Wavelet Transform, Adaptive Threshold, Quantization, and embedded block coding with optimized truncation coding (EBCOT) as shown in Fig.2 and briefly described as follows:

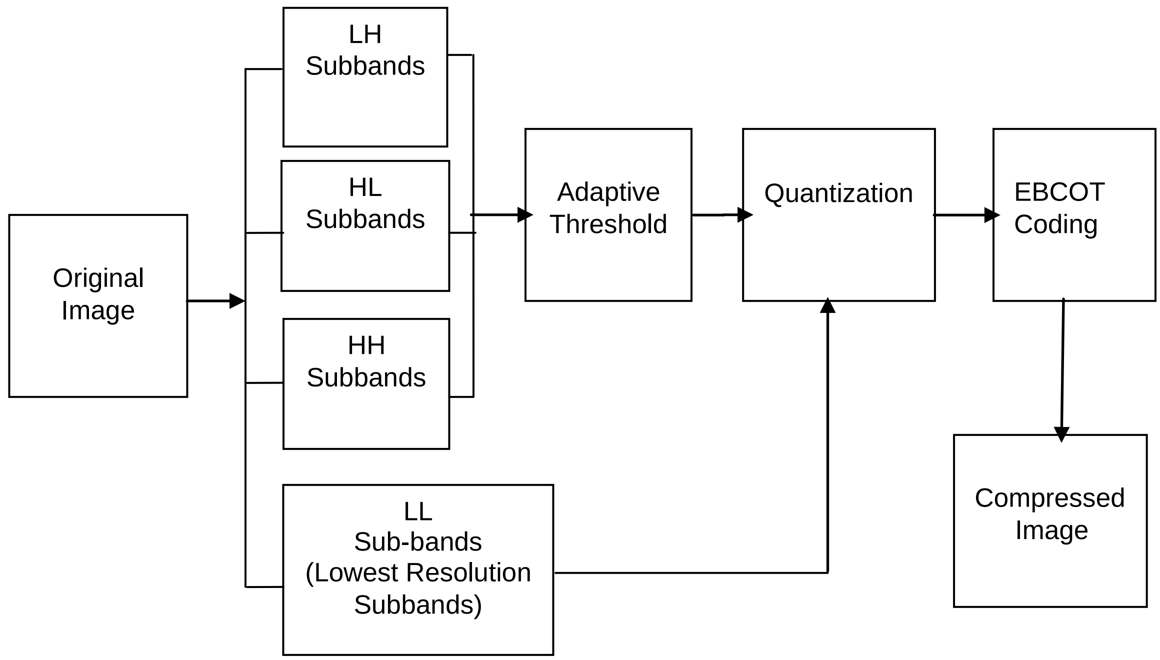


FIGURE 2: Proposed JPEG2000 Based Block Scheme with Adaptive Threshold.

3.1 Discrete Wavelet Transform

DWT can be used to reduce the image size without losing much of the resolution. The DWT [7] [8] apply it to a whole original image and provide a different level of decomposition with coefficients block of image and the block of transformed coefficients are classified into types HH, HL, LH, and LL. These types are described as Diagonal, horizontal, vertical, and image approximation. The DWT applied on an image is depicted by Fig.3.1

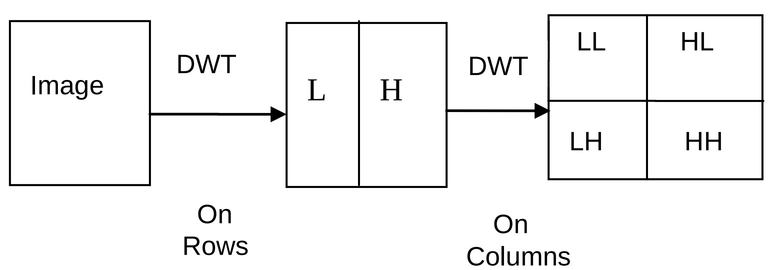


FIGURE 3.1: DWT Applied on Image.

Suppose we have taken an original image and then we apply low pass filter and high pass filter on the rows. Since we are applying the wavelet transform in both the dimensions, so we first apply on the rows and then columns [9] [10]. Whenever we apply low pass filter on rows to preserve the lowest discrete time frequency that is horizontal approximation and apply high pass filter to preserve the highest discrete time frequency that is horizontal detail. After getting the horizontal approximation we again apply low pass filter on column to get the approximate image (LL) and apply high pass filter on column to get the vertical details (LH). Similarly After getting the horizontal detail we again apply low pass filter on column to get the horizontal details (HL) and apply high pass filter on column to get the diagonal details (HH). The 2D discrete wavelet transform can be implemented with below figure.3.2:

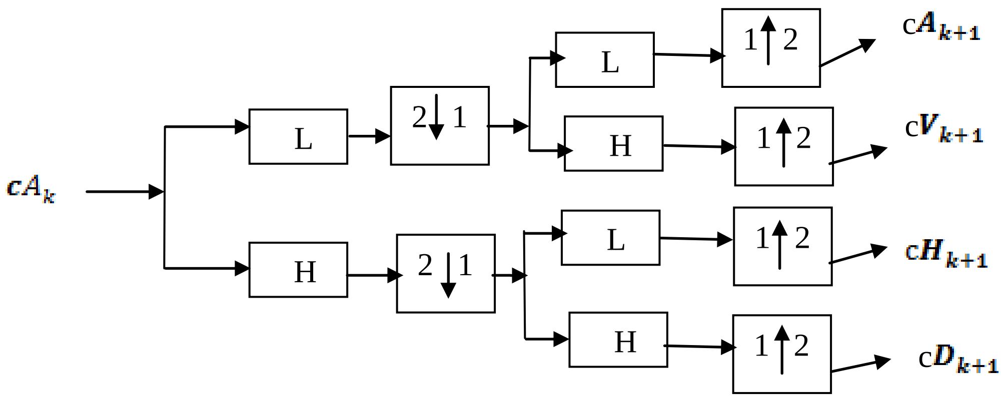


FIGURE 3.2: 2D DWT Implementation.

The low-pass and high-pass decomposition filters are represented as L and H, respectively. Applying and both in horizontal and vertical directions to the original image *I* (*i*, *j*), and next, recursively to the resulting gradually coefficients, we can obtain the following decomposition coefficients.

$$cA_{k+1} = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} L(2i-m)L(2j-n)A_{m,n} \qquad (1.1)$$

$$cV_{k+1} = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} L(2i-m)H(2j-n)A_{m,n} \qquad (1.2)$$

$$cH_{k+1} = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} H(2i-m)L(2j-n)A_{m,n} \qquad (1.3)$$

$$cD_{k+1} = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} H(2i-m)H(2j-n)A_{m,n} \qquad (1.4)$$

The sub-bands are labeled by using following:

(1.1). LL (cA_{k+1}) is the approximation image that indicate low-frequency component (LL) resulting from low-pass filtering in the vertical and horizontal directions.

(1.2). HL (cV_{k+1}) represents the vertical details that results from vertical low-pass filtering and horizontal high-pass filtering.

(1.3). LH (cH_{k+1}) represents the horizontal details that results from horizontal low-pass filtering and vertical high-pass filtering.

(1.4). HH (cD_{k+1}) represents the diagonal details that results from high-pass filtering in both directions.

3.2 Adaptive Threshold

This section describes the method for computing the threshold value *T,* which is adaptive to different sub-bands characteristics [11] [12] for each subbands (except loest resolution subband). After getting the all sub-band coefficients we just applied adaptive threshold into all sub-band coefficients except lowest resolution that's mean image approximation part. Actually this part carry significant bit for image reconstruction with noise free. Actually highest resolution parts that are diagonal, vertical and horizontal carry high value with noise. In that part an adaptive threshold is applied and adapts all coefficients with modified value. The new threshold is calculated by the following equation:

$$Tnew = \sigma\sqrt{2logM}$$

Where M is the number of pixels in the image and σ is the noise variance that is defined as:

$$\sigma = \frac{median|(Yi,j)|}{0.6745}$$

Here $Y(i, j)$ ∈HH sub-band coefficients those are obtained by applying the wavelet transform to the image.
Now we adapt all sub-bands (except loest resolution subband) using following equation:

$$\hat{Y}(i,j) = \begin{cases} (\,Y(i,j) - |Tnew|\,), & |\,Y(i,j)| > Tnew \\ 0, & |\,Y(i,j)| \le Tnew \end{cases}$$

Where $\hat{Y}(i,j)$ is estimated sub-band

3.3 Quantization
The quantization process is calculated by the following equation:

$$Q(i,j) = sign(\hat{Y}(i,j))\left\lfloor \frac{|\hat{Y}(i,j)|}{\Delta n} \right\rfloor$$

Where Q (i, j) is the quantized result at position (i, j) and $\hat{Y}(i,j)$ represent the original DWT coefficient at position (i, j) and Δn stands for the interval width for quantization

3.4 EBCOT Coder
After quantization, each sub-band is divided into rectangular blocks, called code-blocks. These code-blocks are encoded independently. The code-block is decomposed into planes and they are encoded from the most significant plane to the least significant bit-plane sequentially.

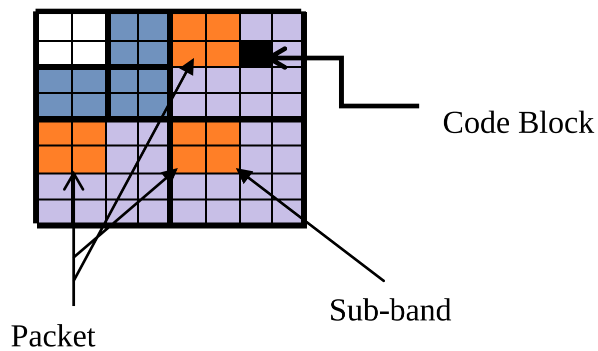


FIGURE 3.3: Partitioning Into Code Block.

The embedded block coding with optimized truncation (EBCOT) encodes each plane in three coding passes [13]. The three coding passes in the order in which they are performed on each plane are significant propagation pass, magnitude refinement pass, and cleanup pass. All three types of coding passes scan the samples of a code block in the same order.

The three passes are described below:

a) Significance propagation pass: During the significance propagation pass, a bit is coded if its location is not significant, but at least one of its eight connects neighbors is significant.

b) Magnitude refinement pass: During this pass, all bits that became significant in a previous biplane are coded. The magnitude refinement pass includes the bits from coefficients that are already significant.

c) Clean-up pass : The clean-up pass is the final pass in which all bits not encoded during the previous passes are encoded (i.e., coefficients that are insignificant and had the context value of zero during the significance propagation pass).The very first pass in a new code block is always a clean-up pass.

After every small block receives three coding passes in every bit-plane, the bit-stream will be generated. These bit-streams need to be reassembled to form the fine bit-stream. The second coding stage is for packaging the bit-streams from the first coding stage into data units called packets. The resulting packets are then assembled into the final bit stream.

4. EXPERIMENTAL RESULT AND ANALYSIS

In this papper, an efficient image compression technique based on JPEG2000 is studied. An image is taken to justify the effectiveness of the algorithm. The resulting compressed image is compared to differents existing techniques like JPEG. Also the proposed technique is compared to existing JPEG2000 that gives high compression ratio and it gives visualy better image quality which is evaluated by the performance parameter Peak Signal Noise Ratio (PSNR). The three 512 x 512 gray scale images considered for analysis that are “Lena image”, “Barbara image” and “Baboon image”.

The Peak Signal to noise Ratio PSNR is estimated by the following equation:

$$PSNR = 10\ log_{10}\frac{(255)^2}{MSE}$$

Where MSE refers to the mean squared error between the original image and the reconstructed image and MSE can be calculated by the following equation:

$$MSE = \left[\frac{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(X(i,j) - Y(i,j))^2}{M * N}\right]$$

Where X means original Image and Y means processed Image. M×N is size of Image and x means row and y means columns.

By using the above formulae in the proposed technique the following parameters are calculated for the Lena image, Barbara image and Baboon image and resulting compressed images that are given in the following below:

FIGURE 4.1: Original Images (Lena Image, Barbara Image and Baboon Image).

FIGURE 4.2: Compressed Images using JPEG (Lena Image, Barbara Image and Baboon Image).

FIGURE 4.3: Compressed Images using JPEG2000 (Lena Image, Barbara Image and Baboon Image).

FIGURE 4.4: Compressed images using Proposed method (Lena image, Barbara image and Baboon image).

The experimental results with the proposed compression method compared to differents existing techniques have been arranged in the Table 1, Table 2 and Table 3 that are given in the following below:

Compression Techniques	File sizes	Compression Ratio	PSNR
Original Image (Lina.bmp)	257.00 KB	----	----
JPEG	18.48 KB	13.90 : 1	24.42
JPEG2000	8.03 KB	31.99 : 1	40.53
Proposed Method	7.07 KB	36.35 : 1	40.97

TABLE 1.1: Comparison between Proposed method and other Compression Methods for the ''Lena image''.

Compression Techniques	File sizes	Compression Ratio	PSNR
Original Image (Barbara.bmp)	257.00 KB	----	----
JPEG	19.89 KB	12.92 : 1	16.42
JPEG2000	9.18 KB	27.97 : 1	35.03
Proposed Method	8.07 KB	31.81 : 1	36.02

TABLE 1.2: Comparison between Proposed method and other Compression Methods for the ‘‘Barbara image’’.

Compression Techniques	File sizes	Compression Ratio	PSNR
Original Image (Baboon.bmp)	257.00 KB	----	----
JPEG	23.92 KB	10.74 : 1	14.68
JPEG2000	10.71 KB	23.99 : 1	27.37
Proposed Method	9.42 KB	27.26 : 1	27.89

TABLE 1.3: Comparison between Proposed method and other Compression Methods for the ‘‘Baboon image’’.

From the comparison table we see that the experimental results demonstrate that the proposed compression technique gives better performance compared to other compression techniques.

5. CONCLUSION

In this paper, we improve the JPEG 2000 image compression technique by adding adaptive threshold. The effectiveness of the proposed technique has been justified using a set of still images. From the experimental results it is clear that the proposed method perform better compression than JPEG and JPEG2000 compression techniques. The experimental result also shows that the proposed method produce the output which are more cleaner and smoother and at the same time kept significant details, resulting in a clearer and appealing vision.

6. REFERENCES

[1] A. Skodras, C. Christopoulos, and T. Ebrahim, “ The JPEG2000 still image compression standard, ” IEEE signal processing image, Vol. 18, PP.36-58, Sept.2001.

[2] M. D. Adams and F. Kossentini, "JasPer: a software-based JPEG2000 codec implementation," IEEE Int. Conf. Image Processing, vol. 2 significance state values Of the three coding pass prediction pp. 53-56, Sep. 2000.

[3] M. J. Weinberger, G. Seroussi and G. Sapiro, “The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS”, IEEE Trans. on Image Processing, Vol. 2, pp. 1309- 1324, Aug. 2000.

[4]. S. Bhavani, K. Thanushkodi, “A Survey on Coding Algorithms in Medical Image Compression”, International Journal on Computer Science and Engineering, Vol. 02, No. 05, pp. 1429-1434, 2010

[5] B. Sriram and S. Thiyagarajan, “Hybrid Transformation Technique For Image Compression”, Journal of Theoretical and Applied Information Technology,Vol. 41, No.2, 2012.

[6] Amita Rakshit, Tapaswini Pattnaik, “A Hybrid JPEG & JPEG 2000 Image Compression Scheme for Gray Images,” International Journal of Scientific Engineering and Technology - 1 June 2014 Volume No.3 Issue No.6, pp : 771-774.

[7] Swapna,Subudhiray,Abhishek,Kr.Srivastav"Implementation of Hybrid DWT-DCT Algorithm for Image Compression: A Review", IJREAS Vol. 2, Issue 2 February 2012

[8] M.I.H. Bhuiyan, M.O. Ahmad, and M.N.S. Swamy, "A New Wavelet-based Method for Despeckling SAR Image," IEEE, 2005.

[9] A. K. Katharotiya, S. Patel and M. Goyani, "Comparative Analysis between DCT & DWT Techniques of Image Compression", Journal of Information Engineering and Applications, Vol 1, No.2, 2011

[10] S. Sridhar , P. Rajesh Kumar , K.V.Ramanaiah, "Wavelet Transform Techniques for Image Compression – An Evaluation" , I.J. Image, Graphics and Signal Processing, 2014, 2, 54-67 Published Online January 2014 in MECS.

[11] S.G. Chang, B. Yu, and M. Vetterli, "Adaptive Wavelet Thresholding for Image Denoising and Compression," IEEE Trans. on Image Processing, Sept. 2000, vol. 9, No.9, pp. 15321546.

[12] S.Sudha, G.R. Suresh, and R.Sukanesh, "Wavelet Based Image Denoising using Adaptive Thresholding," International Conference on Computational Intelligence and Multimedia Applicatio, 2007, pp.296-300.

[13] Tsung-Han Tsai, Che-Wei Chang, and Chih-Lun Fang, "Efficient Bit-Plane Coding Scheme for EBCOT Algorithm in JPEG2000", Innovative Computing, Information and Control (ICICIC), 2009 Fourth International Conference on.

A Novel System to Monitor Illegal Sand Mining using Contour Mapping and Color based Image Segmentation

Akash Deep Singh *akash.singh011235@gmail.com*
Electronics and Instrumentation Engineering Department
Bits-Pilani KK Birla Goa Campus
Goa, 403726, India

Abhishek Kumar Annamraju *abhishek4273@gmail.com*
Electrical and Electronics Engineering Department
Bits-Pilani KK Birla Goa Campus
Goa, 403726, India

Devprakash Harihar Satpathy *devprakash.bits@gmail.com*
Electronics and Instrumentation Engineering Department
Bits-Pilani KK Birla Goa Campus
Goa, 403726, India

Adhesh Shrivastava *adhesh.shrivastava@gmail.com*
Electrical and Electronics Engineering Department
Bits-Pilani KK Birla Goa Campus
Goa, 403726, India

Abstract

Developing nations face the issue of illegal and excessive land mining which has adverse effects on the environment. A robust and cost effective system is presented in this paper to monitor the mining process. This system includes a novel vehicle detection approach for detecting vehicles from static images and calculating the amount of sand being carried to prevent the malpractices of sand smuggling. Different from traditional methods, which use machine learning to detect vehicles, this method introduces a new contour mapping model to find important “vehicle edges” for identifying vehicles The sand detection algorithm uses color based segmentation since sand can have various colors under different weather and lighting conditions The proposed new color segmentation model has excellent capabilities to identify sand pixels from background, even though the pixels are lighted under varying illuminations. The detected amount of sand is checked against the maximum set threshold value specific to the recognized vehicle. Experimental results show that the integration of Hough features and color based image segmentation is powerful. The average accuracy rate of the system is 94.9%.

Keywords: Illegal Sand Mining, Contour Detection, Hough Transform, Color based Segmentation.

1. INTRODUCTION

Sand mining is a process of extracting sand from beaches, dunes, or open pits. Excessive mining has a serious environmental effect such as soil erosion, loss of biodiversity, contamination of groundwater, and formation of sinkholes [1]. The author in [2] gives a detailed explanation about the ill effects of mining on water resources and air quality. Developed nations have taken measures to keep these practices regulated using stricter government policies.

The issue of Illicit sand mining is one of the major problems in developing nations. Weaker government policies for mining leads to illegal sand mining in rural areas. It is difficult to stop such malpractices on a large scale without the support of the ruling government. Increasing corruption adds more trouble to curbing the issue. The progress of technology has provided ways to monitor this issue. To put a check on such malpractices this paper presents an application developed using computer vision methods. The proposed system recognizes vehicle type and segments the visible amount of sand present on the vehicle which is then compared to the maximum set threshold capacity specific to the vehicle. Thereby monitoring the illegal transportation of sand beyond the vehicle's set capacity. Section 2 provides an overview of the application, section 3 includes proposed algorithm, section 4 presents the results and the paper is summarized in section 5.

2. APPLICATION OVERVIEW

The application developed can classify a vehicle from static images and detect the amount of sand it is carrying. This application steps in at a point where the sand is being made ready for transportation after the mining process is done. We present a novel system to detect vehicles from static image and calculate the amount of sand it is carrying. This application involves acquisition of rear and side view images, which are then processed separately to extract data. The entire application is a real time based two-step process. In the first step the system takes the side view image and finds the number of tires and classifies the vehicle on the basis of number of tires. This information is transferred to the database to fetch data containing the amount of sand the vehicle can carry. For most of the cases the vehicles are weighed on a weighing machine to keep monitoring the amount of sand being mined and transported. But there can be situations where the weighing systems may be tampered with. As a solution, we take the acquired images and segment out the amount of sand visible above the container. The area of the extracted part is calculated and is matched against the set values from the database specific to the type of vehicle. The parameters for every vehicle includes,

- Maximum allowable weight of sand that can be carried
- Maximum allowable visible sand area over the vehicle's side panels as explained in Figure 1

FIGURE 1: Explains the positioning of side panels and the visible area of sand concept.

To ensure that the application is user friendly, the freedom of obtaining images from any camera with resolution greater than 5 Megapixels is provided. The system is developed using Open Computer Vision Libraries. The working platform must have minimum requirements as shown in Table 1, thus the entire application can be ported onto any smart-phone. The presented system is simple, cost effective and computationally fast.

Processor	>1.5GHz
Camera resolution	>5Megapixels
Web Connectivity	To transmit captured images for off-site processing

TABLE 1: Minimum System requirements for the proposed application.

The system carries out two different sets of operations on the input image. First part is to detect tires from the static side view image. For dealing with static images, Wu et al. [3] used wavelet transform to extract texture features to detect vehicles on roads. Then, each positive detection is verified using a PCA (principal component analysis) classifier. In addition, Sun et al. [4] used Gabor filters to extract textures and each vehicle candidate is verified using a SVM (support vector machines) classifier. In addition to textures, “symmetry” is an important feature used for vehicle detection. In [5], Broggi et al. described a detection system to search for areas with a high vertical symmetry as vehicle candidates. However, this is prone to greater false detections. Furthermore, in [6], Bertozzi et al. used corner features to build four templates of vehicles for detection. In [7], Tzomakas and Seelen found that the area shadow underneath a vehicle can be used to detect vehicles. In [8], Ratan et al. created an algorithm to detect vehicles’ wheels as features to find possible vehicle position and then used a Diverse Density for verification. In addition, Bensrhari [9] and Aizawa [10] used stereo-vision methods and 3-D vehicle models to detect vehicles and obstacles.

To overcome these complications and to make the applications work with lesser resources we present a vehicle detection system which is customized to this particular situation. Since the user has the liberty to capture the image, the side view and the back view are the inputs given by the user to the system. After this the images are processed using a preprocessing algorithm which equalizes the light conditions of the image. The image is then processed through edge detection and Hough detection algorithms to detect and classify the vehicle on the basis of number of tires and the container size both of which are predefined within the system.

The second part involves extraction of sand from the images. Image segmentation and object detection is one of the major fields of Computer Vision. One of most common approaches for object detection is using vision-based techniques to analyze images or videos. However, due to the variations of object’s colors, sizes, orientations, shapes, and poses, developing a robust and effective system of vision-based detection is very challenging. The algorithm used in our approach includes a pre-processing technique, followed by K-Means clustering segmentation and then a thresholding operation. Narkhede in [11] gives a clear review of state-of-the-art image segmentation techniques. Dutta et.al [12] proposed a system for color based segmentation using the homogeneous similarities in adjoining pixels. The classification done is based on the TSK-Fuzzy [13] logic followed by several (IF-THEN) logics, which makes it computationally expensive. To keep the system simple and efficient the proposed system uses K-Means Clustering algorithm [14] for segmentation.

3. PROPOSED SYSTEM

A two-stage algorithm for vehicle and sand detection is developed by combining an image-processing of the edge information and color based segmentation. Both the image-processing systems have been designed for real-time implementation with minimal resource usage. This section is divided into two sub-sections, namely vehicle classification and segmentation of sand. Both the sections include in-depth explanation of the algorithms used with illustrations. Every step explained in the section is accompanied with proper advantage of it as well as its effect in functioning of the entire system.

3.1 Vehicle Classification

One of the main aims of this paper is to detect a vehicle from static images. Rather than using any Machine learning methods, a simpler approach of using edges and contours is used for detection. Since the final implementation is to be carried out on smartphones the algorithm needs to use minimal resources while providing best results. Figure 2 shows the flow diagram of the process of vehicle detection.

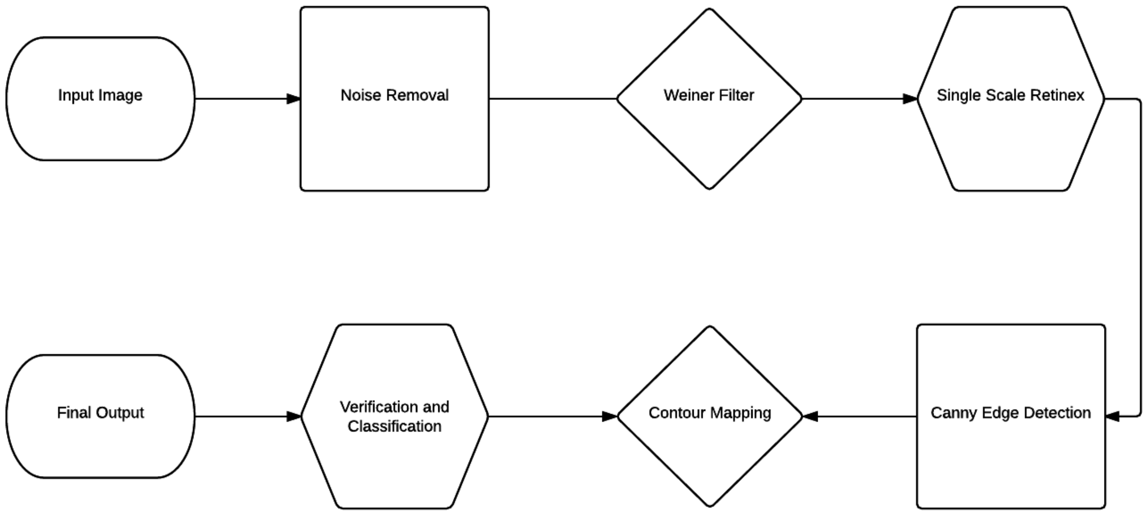


FIGURE 2: Algorithm of the Process.

The major problem of the input image is the varied light conditions. To resolve this a Multi Scale Retinex algorithm followed by Gaussian smoothing R model [15] is used, where the smoothness of region *u* is measured by the Dirichlet integral *|Du|*^2. The Multiscale Scale Retinex algorithm solves the problem of variable light conditions since the image is affected by daylight and presence of shadows. Retinex is an image enhancement algorithm that is used to improve the contrast, brightness and sharpness of an image primarily through dynamic range compression. The algorithm also simultaneously provides color constant output and thus it removes the effects caused by different illuminants on a scene. The original algorithm is based on a model of human visions lightness and color constancy developed by Edward Land. Jobson et al. extended the last version of Lands model. The smoothing requirement is usually expressed by the positivity of the kernel:

$$G_h(x) = \frac{e^{\frac{-|x|}{4h^2}}}{4\pi h^2}$$

The paradigm of such kernels is of course the gaussian kernel *x*. In that case, G_h has standard deviation *h* and it is easily seen that:

$$u - G_h * u = -h^2 \Delta u + o(h^2)$$

A similar result is actually valid for any positive radial kernel with bounded variance, so one can keep the gaussian example without loss of generality. The preceding estimate is valid if *h* is small enough. However since the noise reduction properties in this case depend upon the fact that the neighborhood involved in the smoothing is large enough, so that the noise gets reduced by averaging. So in the following we assume that *h* = *k*, where *k* stands for the number of samples of the function *u* and of the noise in an interval of length *h*, *k* must be much larger than 1 to assure the noise reduction. The effect of a gaussian smoothing on the noise can be evaluated at a reference pixel *i* = 0. At this pixel,

$$G_h * n(0) = \sum_{i \epsilon I} \int_{P_i} G_h n(x) dx = \sum_{i \epsilon I} \epsilon^2 G_h(i) n_i$$

where we recall that *n(x)* is been interpolated as a piecewise function, the P_i square pixels centered in *i* have size 2 and $G_h(i)$ denotes the mean value of the function G_h on the pixel *i*. Denoting by *Var(X)* the variance of a random variable *X*, the additivity of variances of independent centered random variables yields:

$$Var(G_h * n(0)) = \sum_i \epsilon^4 G_h(i)^2 \sigma^2$$

$$\cong \sigma^2 \epsilon^2 \int G_h(x)^2 dx = \frac{\epsilon^2 \sigma^2}{8 \pi h^2}$$

After the process of equalizing the brightness of the image the output image becomes blurry. The Wiener filter can remove the additive noise and invert the blurring. A Low pass filter is not suitable in this case since it is very sensitive to additive noise. The Wiener filter is proposed to optimize the trade-off between de-noising and inverse filtering. Noise removal in previous steps may lead to blurring of edges. Experiments show that a symmetric low-pass filter of size 7x7 with standard deviation of 0.5 efficiently restores the texture pattern. This pre-processing provides a de-noised pre-segmented image. To summarize (and convert to 2D), given a system:

$$y(h, m) = h(n, m) * x(n, m) + v(n, m)$$

where * denotes convolution and *x* is the (unknown) true image *h* is the impulse response of a linear, time-invariant filter *v* is additive unknown noise independent of *x*, and *y* is the observed image. We find a deconvolution filter *g* to estimate *x*:

$$x'(n, m) = g(n, m) * y(n, m)$$

where x_0 is an estimate of *x* that minimizes the mean square error. In the frequency domain, the transfer function of *g*, *G* is:

$$G(w_1, w_2) = \frac{H^*(w_1, w_2) S(w_1, w_2)}{\mid H(w_1, w_2) \mid^2 S(w_1, w_2) + N(w_1, w_2)}$$

where *G* is the Fourier transform of *g*, *H* is the Fourier transform of *h*, *S* is the mean power spectral density of the input *x*, and *N* is the mean power spectral density of the noise *v*. The equation for *G* can be rewritten as:

$$G(w_1, w_2) = \frac{\mid H(w_1, w_2) \mid^2}{H(w_1, w_2)[\mid H(w_1, w_2) \mid^2 + \frac{N(w_1, w_2)}{S(w_1, w_2)}]}$$

So the Wiener filter has the inverse filter for *H*, but also a frequency-dependent term that attenuates the gain based on the signal to noise ratio.

Edges characterize boundaries and therefore carry fundamental information of an image. Edges in any image consist of areas with strong intensity contrasts i.e. a jump in intensity from one pixel to the next. As shown edge detection in an image significantly reduces the amount of data to process, while preserving the important structural properties in an image.

One of the best available algorithms present is the canny edge detection algorithm. To further enhance the edge information a Sobel filter is applied. The Sobel operator [16] performs a 2-D spatial gradient measurement on an image. Then, the approximate absolute gradient magnitude at each point is found. In the Sobel operator a pair of 3x3 convolution masks are used, where one estimates the gradient in the x-direction (columns) and the other in the y-direction (rows).From these two images, we find edge gradient and direction for each pixel as follow

$$Edge\_gradient(G) = \sqrt{G_x^2 + G_y^2} \qquad Angle(\theta) = \arctan \frac{G_y}{G_x}$$

After getting gradient magnitude and direction, a full scan of image is done to remove any unwanted pixels which may not constitute the edge. For this, at every pixel, pixel is checked if it is a local maximum in its neighborhood in the direction of gradient. The next step removes any background edges which act as noise for the result. For this, we need two threshold values. Any edges with intensity gradient more than the maximum threshold are sure to be edges and those below minimum are sure to be non-edges, so are discarded. Contour data are an integral part of the application. Contours can be explained simply as curves joining all the continuous points (along the boundary), having same color or intensity. It helps in segregating the real data points from unnecessary background noise. They further enhance the quality of shapes that are to be detected as can be seen in Figure 3.

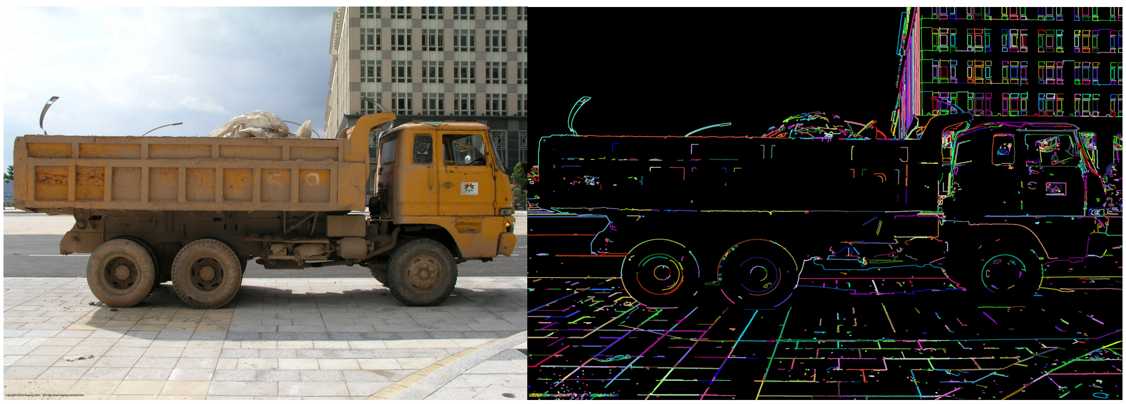


FIGURE 3: Contour Map.

Once the edge information is extracted, the contour mapping of the image is done to further remove any unnecessary edges and to map the vehicle. Using the contour map the vehicle dimensions can be extracted and classification can be carried out.

Hough Transform uses parametric representation for the family of all circles and transform each figure point in the obvious way [17]. The classical Hough transform was developed to identify lines in the image, but later the Hough transform extended to identify the positions of arbitrary shapes, most commonly circles or ellipses. Due to imperfections in either the image data or the edge detector, however several points or pixels on the desired curves are missed. Also there are spatial deviations between the ideal circle and noisy edge points are obtained from the edge detector. For these reasons, it is often non-trivial to group the extracted edge features to an appropriate set of circles. Our procedure addresses this problem by making it possible to perform groupings of edge points into object candidates by performing an explicit voting procedure over a set of parameterized image objects as can be seen from Figure 4. A circle is fully defined with three parameters: the center coordinates (a, b) and the radius (R):

$$x = a + R\cos\theta$$

$$y = b + R\sin\theta$$

As Θ varies from 0 to 360, a complete circle of radius R is created. So the transform function looks for triplets of (x, y, R) from the image. Therefore, we need to construct a 3D accumulator for Hough transform, which would be highly ineffective. So, we use a trickier method, Hough Gradient Method which uses the gradient information of edges. However, the method also includes hidden circles which are meant to be perceived as circles.

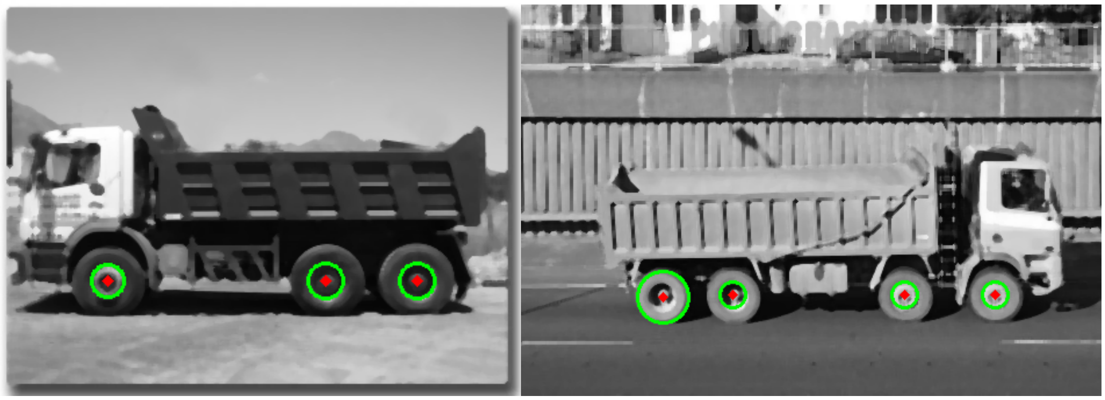


FIGURE 4: Tire Detection.

In general, without image de-noising, the algorithms tends to extract too many circular features. So, to be more successful, the preprocessing applied is a crucial step.

3.2 Segmentation of Sand

The first step is *Image acquisition*. Two images of the vehicle carrying sand is captured. These images include the side and rear view of the vehicle. The purpose of obtaining these two views is to ensure a dual check process on the quantity estimation of the sand carried by the vehicle. The next step is *Downsampling* the images. The 2D area of the sand in both the views is matched against a set threshold value. This threshold value is subject to the type of the truck and the view. To keep a common base for testing the area the image is down sampled to a width of 500 pixels and height of 500 pixels. The threshold values are obtained from a set of training images and the process is explained later in the paper. The whole reason for down sampling is to create an access image that is a miniaturized duplicate of the optical resolution of master scan. If the image signal and the image noise had similar properties, averaging neighboring pixels in order to reduce the resolution would not improve the signal-to-noise ratio. However, signal and noise have different properties. But in this paper the major reason is to increase signal to noise ratio [18].

The step followed by *Downsampling* the images is *Application of morphological operations*. The segmentation algorithm used in this application is K-Means clustering. To ensure perfect clustering, sharpness of the image has to be reduced and components must emerge out as more connected. This is facilitated by the two morphological operators, namely, dilate and erode. The importance of this step can be understood from Figure 5. The sub-images from Figure 5.(a) to Figure 5.(e) are experimental outputs with different combinations of morphological operations. Figure 5.(a) is the output of clustering without any morphological pre-processing operation, Figure 5.(b) was the output when only dilation operation was applied, Figure 5.(c) was the output when only erosion operation was applied. A region of interest highlighted in the sub-images (a) to (c) of Figure 5, depicts the inefficient clustering outcome. Even the order of using the two operators is important. In the same figure, it can be observed in (e) that smooth clustering output is obtained when dilate operation is followed by erode, while on the other hand, irregular clustering output is obtained when erode operation is followed by dilate operation.

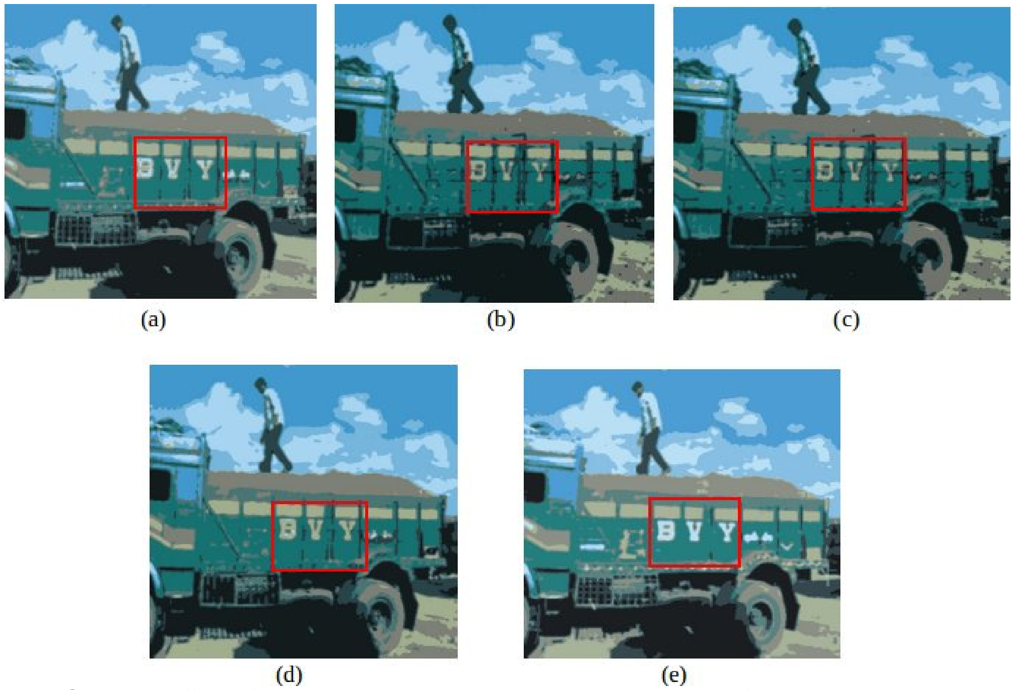


FIGURE 5: Effect of morphological pre-processing on the output of clustering algorithm.

The kernel size used for both the morphological operations is of width 3 px and height 3 px. Another important parameter related to these operations is the number of iterations. In the application the value for this parameters is 1, as it can be observed from Figure 6 that as the number of iterations increase, the clustering of image deteriorates. In figure 6, the number of iterations is 1 in (a), 3 in (b), 5 in (c) and 7 in (d).

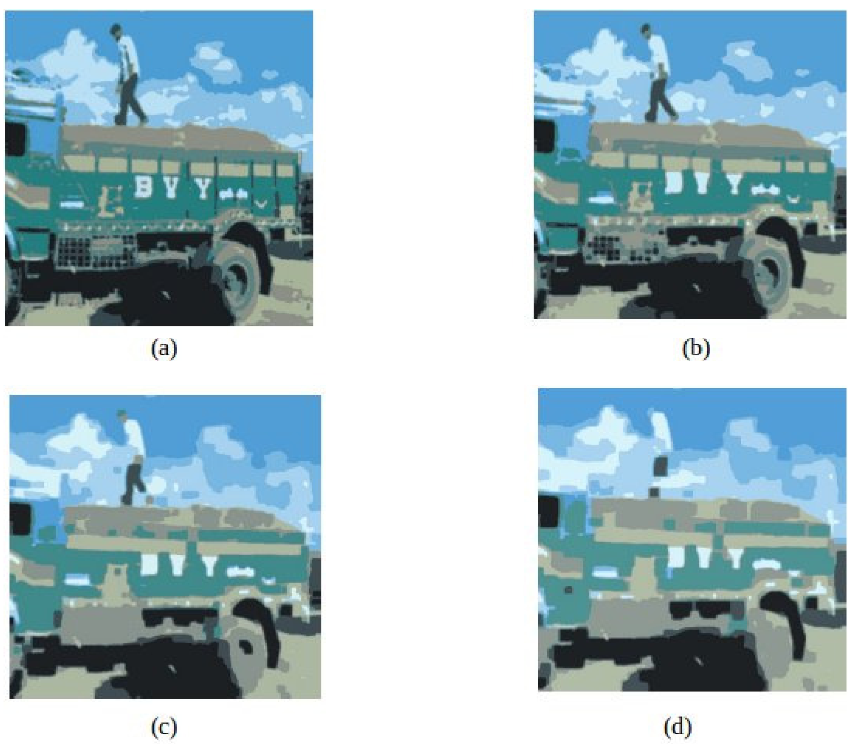


FIGURE 6: Effect of the parameter iterations of morphological operations on the output of clustering algorithm.

After performing the morphological operations, the next step is Noise removal. This step is to make sure that the noise, which the morphological operations are not able to remove, does not affect the application's output. The noise is removed by applying a mean filter of kernel width 5 px and height 5 px. Experiments were conducted on the filter's kernel size value and the results are shown in Figure 7. The optimal kernel for the application was chosen to be of width 5 px and height 5 px. The experiments had symmetric kernels, it was 3 in Figure 7.(a), 5 in 7.(b), 7 in 7.(c), 9 in 7.(d) and 11 in 7.(e). It can be clearly seen that the output with filter kernel size 3 had noise and the ones with kernel size greater than 5 are excessively blurred with loss of important data.

FIGURE 7: Effect of kernel size of pre-processing filter on the output of clustering algorithm.

The next step is the most important step, *Clustering*. The algorithm used in this application is K-Means clustering algorithm. This step has two important parameters, namely cluster size(c) and number of iterations (i) and the values selected for the two parameters are 13 and 15 respectively. It can be observed from Figure 8 that as the cluster size increases the calculation time increases. The results in Figure 9 guide us to the fact that optimal value of c is 13, as till c = 4 the extraction of sand pixels is very poor. The values of c from 5 to 12 there is a lot of false segmentation, optimal value appears at c = 13, and this has been verified experimentally using various test images. Similarly the optimal value of i occurs at 15 as it can be observed from Figure 10.

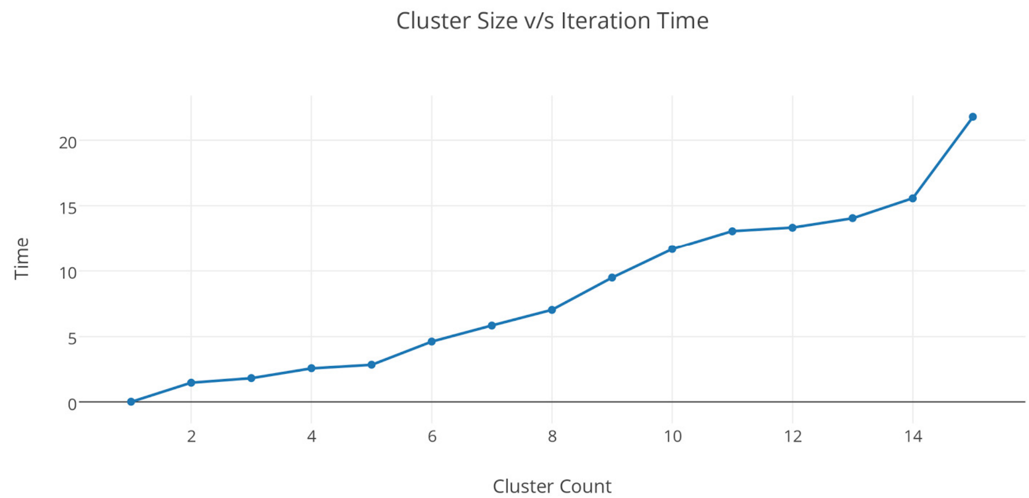


FIGURE 8: Variation in processing time with the parameter cluster size of K-Means clustering algorithm.

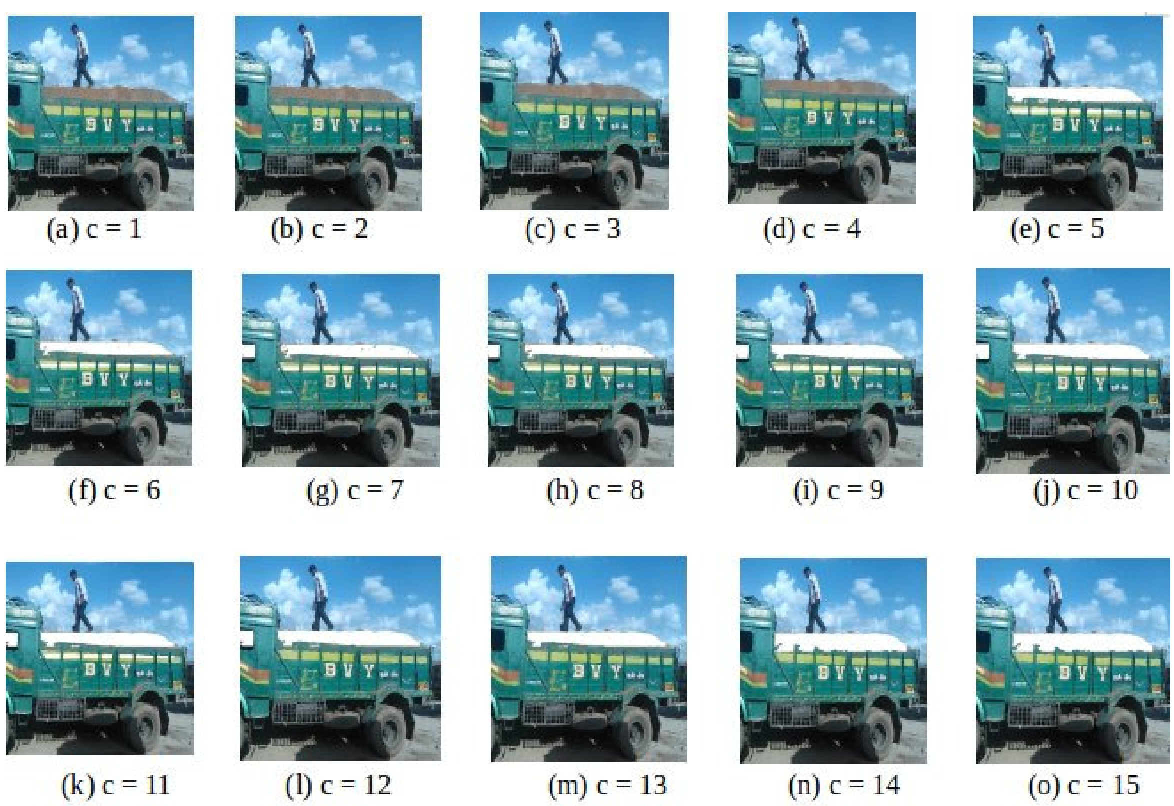


FIGURE 9: Effect of the parameter cluster size of K-Means clustering algorithm on the segmentation process.

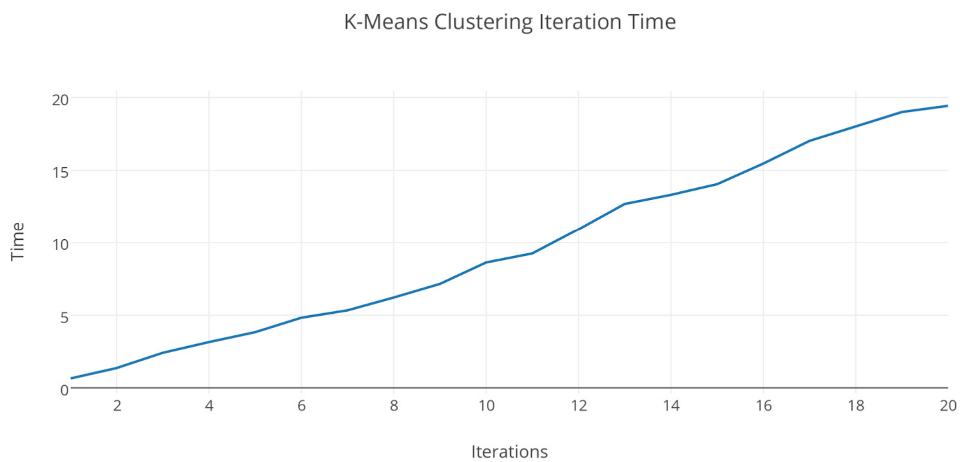


FIGURE 10: Variation in processing time with the parameter cluster size of K-Means clustering algorithm

The step after applying clustering algorithm is changing the image representation. The representation of image until this step was RGB format. Setting up color based thresholds in RGB format results in false detections most of the times. For this reason the image is transformed into HSV format. The step following is Thresholding. In this step a pre-obtained threshold level for the H, S, and V values is put against the transformed test image. The purpose is to extract all the sand pixels that are over the vehicle. To ensure that the sand pixels other than the required, i.e., over the vehicle's tires or that on the ground level are detected, the thresholding is carried over a specific region of interest. The region of interest (rectangular) for a side view has properties shown in Table 2.

Top left x coordinate	0
Top left y coordinate	105
Region of interest width	450
Region of interest height	145

TABLE 2: Parameters of Region of interest defined for side view images.

Similarly the region of interest for a rear view image of the truck has properties shown in Table 3.

Top left x coordinate	50
Top left y coordinate	100
Region of interest width	400
Region of interest height	150

TABLE 3: Parameters of Region of interest defined for rear view images.

The values are obtained over experimentation. The threshold values have been set after understanding the different gradients in the color of sand at different regions, weather conditions and lighting conditions. The set value functions in such a way that all the pixels with H between 0 and 160, S between 0 and 50, V between 0 and 200 are extracted out and labelled as sand pixels.

The final step in the proposed approach is Correction Algorithm. The aim of this step is to remove the undesired and false detections by exploiting the connected nature of the sand present in the image. An all-region sweep stores the data of the pixels which are most illuminated as per the clustering algorithm. The sweep is divided on the basis of rows. A particular row is selected and in that row, any pixel in the threshold range is labelled as white and the rest as black. For any row if *[white/(white+black)] <= 0.48,* then all the pixels in that row labelled as sand pixels are rejected.

This approach removes 93% of the false positive patches and provides a stable detection as it can be seen in Figure 11. In the Figure 11.(a), the region of false detection is highlighted, and Figure 11.(b) is processed out by applying the Correction algorithm.

(a) (b)

FIGURE 11: Advantage of the Correction Algorithm.

The entire segmentation algorithm is shown in Figure 12 as a flow-chart format.

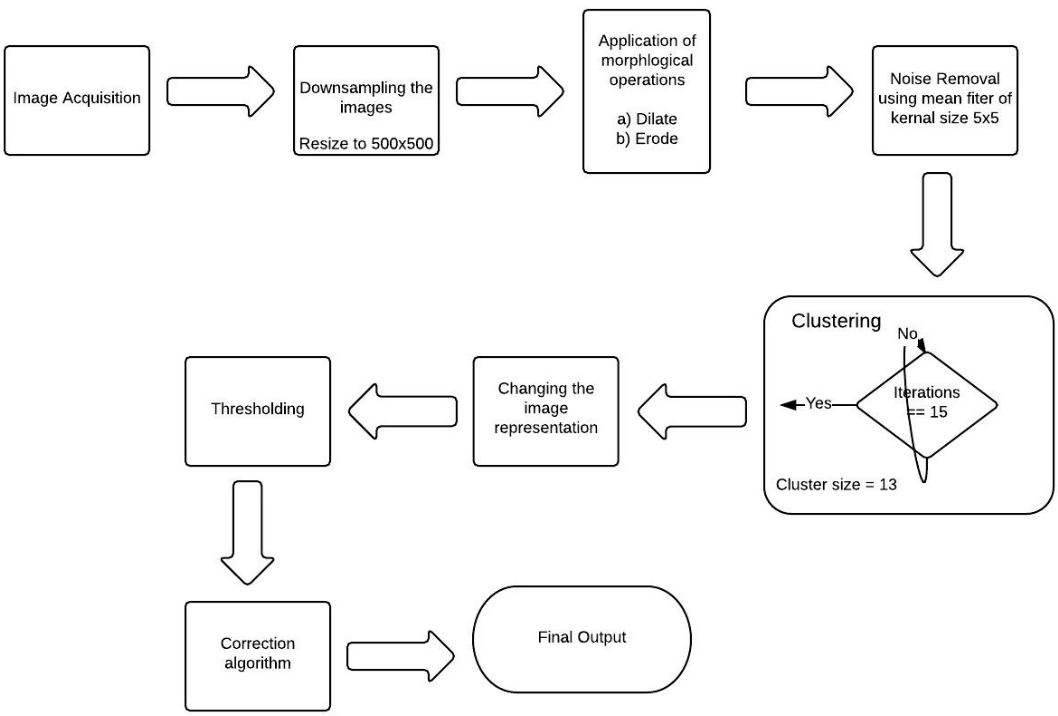


FIGURE 12: Sand Segmentation Process.

4. RESULTS

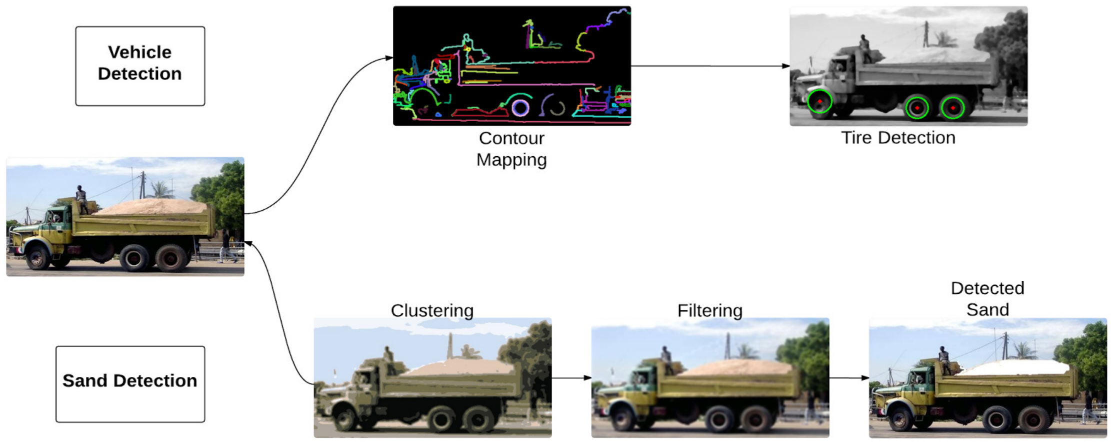


FIGURE 13: Illustrating system working with an example.

In order to analyze the performance of our proposed method, various static images captured under different weather conditions and lighting conditions were used. The approach presented in this paper supports direct measurement of detection and tracking, facilitates iterative algorithm development, and provides important diagnostic feedback. The flowchart in Figure 13 presents the overall process from capturing the image to the end output. The process begins by determining the vehicle class according to its sand carrying capacity. As shown in Figure 14 after detection the trucks are classified into one of the following classes: Large, Medium or Small. After the process of classification the input image is fed into the sand detection algorithm.

FIGURE 14: Vehicle Classification.

The robust sand detection results are shown in Figure 15. The preset threshold values determine whether the vehicle is overloaded or not. Experimental results suggest that in sub-images 15.(a) and 15.(c) the vehicle is carrying sand beyond the maximum capacity while in 15.(b) and 15.(d) it is well within the limits.

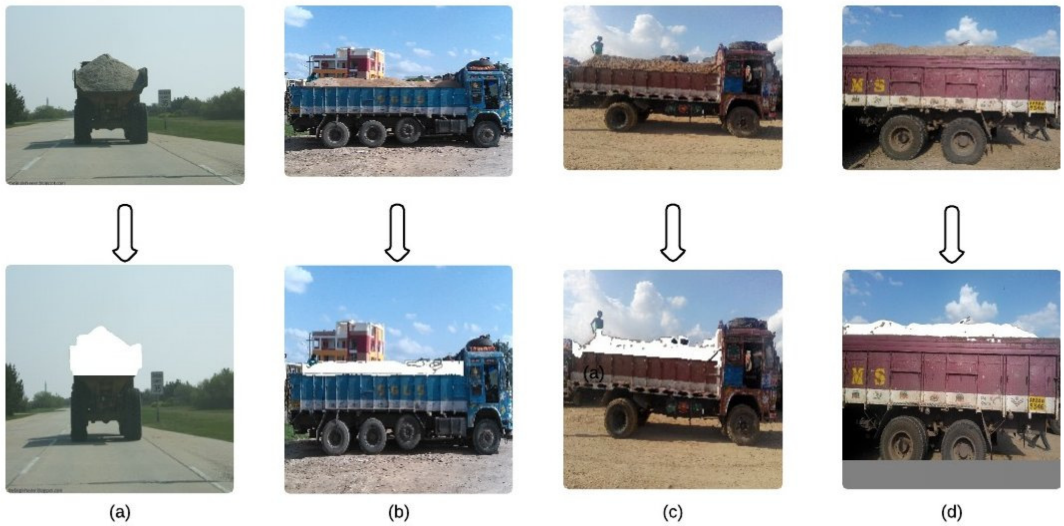


FIGURE 15: Results of Sand Detection

5. COMPARITIVE ANALYSIS

This paper presents a fast and reliable vehicle recognition system and classification system. Many systems exist which make use of infrared sensors to generate heat maps and detect objects [19][20]. Different machine learning methods tend to fail in poor visibility conditions since their positive training dataset generally do not consider such cases. A well trained classifier also fails in cases of high occlusion. To overcome these problems, we present a vehicle detection method which does not rely on machine learning techniques. [21] presents a scale invariant method for vehicle detection. They use SIFT features to train a classifier which requires a well-built dataset. [22] gives a fast feature extraction method to generate features for training. However all these methods require time to generate a well-trained classifier. The quality of such a system depends highly on the training dataset. In this paper we present a method that overcomes these hassles. The method we present detects tires as the targets of vehicles and is effective in any perspective.

The technique used in the application for segmenting sand is a pixel based segmentation technique. The main aim is to segment out sand regions from the vehicle in a noisy environment. While significant progress has been made in texture segmentation [23–26] and color segmentation [27–29] separately, the combined texture and color segmentation problem is considerably more challenging [30–32]. The most popular method for image segmentation is k-means clustering which provides a regional segmentation within an image [33][34].However the results from existing system tend to error prone and the amount of uncertainty is variable with different settings. The algorithm presented in this paper takes into account such variability and gives near perfect results of the lighting conditions.

Combining these two modules we have developed our original vehicle detection methods which is compatible with any platform be it PCs or smartphones and can be integrated without the hassles of machines learning.

6. CONCLUSION

In this paper a vehicle classification and sand detection algorithm is presented. The system presented in this paper provides a layer of abstraction to the user which simplifies the whole process of implementing and using the application. The user is provided with a black box which can take, as input, the side and front view of the vehicle and give the vehicle type and the amount of sand present as output. The algorithm developed here is error free even though a lot of variability in the capturing conditions act as hindrances. Since it can well record different changes of vehicle appearances, desired vehicles can be very accurately and effectively detected from static images. However there is still a lot of scope for development. The algorithm presented can be used to develop applications with minimal hardware requirements. The customizability of such an application is endless and such systems can be employed to several other industries.

7. ACKNOWLEDGEMENT

We would like to express our deepest gratitude towards WHISHWORKS, Hyderabad, for providing us the dataset for testing our algorithms and for their humble cooperation throughout this research.

8. REFERENCES

[1] Ojos Negros Research Group. “Sand Mining Impacts” USENET: http://threeissues.sdsu.edu/three_issues_sandminingfacts01.html. [April. 29, 2015].

[2] M. Naveen Saviour. (2012). “Environmental impact of soil and sand mining: A review,” International Journal of Science, Environment and Technology. [Online]. 1(3), pp. 125-134. Available: http://www.ijset.net/journal/27.pdf [April. 29, 2015].

[3] J. Wu, X. Zhang, and J. Zhou, “Vehicle detection in static road images with PCA and wavelet-based classifier,” in Proc. IEEE Intelligent Transportation Systems Conf., Oakland, CA, Aug. 25–29, 2001, pp. 740–744.

[4] Z. Sun, G. Bebis, and R. Miller, “On-road vehicle detection using Gabor filters and support vector machines,” presented at the IEEE Int. Conf. Digital Signal Processing, Santorini, Greece, Jul. 2002.

[5] A. Broggi, P. Cerri, and P. C. Antonello, “Multi-resolution vehicle detection using artificial vision,” in Proc. IEEE Intelligent Vehicles Symp., Jun. 2004, pp. 310–314.

[6] M. Bertozzi, A. Broggi, and S. Castelluccio, “A real-time oriented system for vehicle detection,” Journal of Systems Architecture, pp. 317–325, 1997.

[7] C. Tzomakas and W. Seelen, “Vehicle detection in traffic scenes using shadow,” Tech. Rep. 98-06 Inst. fur neuroinformatik, Ruhtuniversitat, Germany, 1998.

[8] A. L. Ratan, W. E. L. Grimson, and W. M. Wells, “Object detection and localization by dynamic template warping,” International Journal of Computer Vision, vol. 36, no. 2, pp. 131–148, 2000.

[9] A. Bensrhair et al., “Stereo vision-based feature extraction for vehicle detection,” in Proc. IEEE Intelligent Vehicles Symp., Jun. 2002, vol. 2, pp. 465–470.

[10] T. Aizawa et al., “Road surface estimation against vehicles’ existence for stereo-based vehicle detection,” in Proc. IEEE 5th Int. Conf. Intelligent Transportation Systems, Sep. 2002, pp. 43–48.

[11] H.P. Narkhede. (2013). “Review of Image Segmentation Techniques,” International Journal of Science and Modern Engineering (IJISME). [Online]. 1(8), pp. 54-61. Available: http://www.ijisme.org/attachments/File/v1i8/H0399071813.pdf, [April. 29, 2015].

[12] S. Dutta, B.B. Chaudhuri, “Homogeneous Region based Color Image Segmentation”, in Proc. World Congress on Engineering and Computer Science, Oct. 2002, Vol II

[13] Fuzzy models, “Fuzzy Rule-based Models”, USENET: http://www.csee.wvu.edu/classes/cpe521/presentations/Sugeno-TSKmodel.pdf, 1997. [April. 29, 2015].

[14] Shi Na, Guan Yong and Liu Xumin, “Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm,” in Third International Symposium on Intelligent Information Technology and Security Informatics, 2010, pp. 63-67.

[15] M. Lindenbaum and M. Fischer and A. M. Bruckstein.(1994). “On Gabor Contribution to Image Enhancement”. Pattern Recognition, 27(1), pp. 1-8.

[16] Irwin Sobel, “History and definition of the Sobel operator”, USENET: http://www.researchgate.net/profile/Irwin_Sobel/publication/210198558_A_3x3_isotropic_gradient_operator_for_image_processing/links/0a85e52eeba907a814000000.pdf, 2004, [April. 29, 2015].

[17] Richard O. Duda, Peter E. Hart. “Use of Hough Transformation to detect lines and curves in pictures”,USENET:http://www.cse.unr.edu/~bebis/CS474/Handouts/HoughTransformPaper.pdf, 1971. [April. 29, 2015].

[18] Ian T. Young, Jan J. Gerbrands, Lucas J. Van Vliet. “Fundamentals of Image Processing,” USENET:http://www.researchgate.net/publication/2890160_Fundamentals_Of_Image_Processing,1998. [April. 29, 2015].

[19] A. El Maadi and X. Maldague, “Outdoor infrared video surveillance: a novel dynamic technique for the subtraction of a changing background of IR images,” Infrared Physics & Technology, vol. 49, no. 3, pp. 261–265, 2007.

[20] Y. Chen, X. Liu, and Q. Huang, “Real-time detection of rapid moving infrared target on variation background,” Infrared Physics & Technology, vol. 51, no. 3, pp. 146–151, 2008.).

[21] Dorko, G.; Schmid, C., "Selection of scale-invariant parts for object class recognition," Proceedings. Ninth IEEE International Conference on Computer Vision, 2003., pp.634,639 vol.1, 13-16 Oct. 2003

[22] Son, T.T.; Mita, S., "Car detection using multi-feature selection for varying poses," Intelligent Vehicles Symposium, 2009 IEEE, pp.507,512, 3-5 June 2009

[23] A. Kundu, J.-L. Chen, “Texture classification using QMF bank-based sub-band decomposition,” CVGIP, GMIP, v. 54, p. 369–384, Sept. 1992.

[24] T. Chang, C.-C.J. Kuo, “Texture analysis and classification with tree-structured wavelet transform,” IEEE Tr. IP, v. 2, p. 429–441, Oct. 1993.

[25] M. Unser, “Texture classification and segmentation using wavelet frames,” IEEE Tr. IP, v. 4, p. 1549–1560, Nov. 1995.

[26] T. Randen, J.H. Husoy, “Texture segmentation using filters with optimized energy separation,” IEEE Tr. IP, v. 8, p. 571– 582, Apr. 1999.

[27] T.N. Pappas, “An adaptive clustering algorithm for image segmentation,” IEEE Tr. SP, v. 40, p. 901–914, Apr. 1992.

[28] M.M. Chang, M.I. Sezan, A.M. Tekalp, “Adaptive Bayesian segmentation of color images,” JEI, p. 404–414, Oct. 1994.

[29] D. Comaniciu, P. Meer, “Robust analysis of feature spaces: Color image segmentation,” CVPR, June 1997, p. 750–755.

[30] W.Y. Ma, Y.Deng, B.S. Manjunath, “Tools for texture/color based search of images,” Human Vision and Electronic Imaging II, Feb. 1997, Proc. SPIE, Vol. 3016, p. 496–507.

[31] S. Belongie, et al., “Color- and texture-based image segmentation using EM and its application to content based image retrieval,” ICCV, 1998, p. 675–682.

[32] Y. Deng, B.S. Manjunath, “Unsupervised segmentation of color-texture regions in images and video,” IEEE Tr. PAMI, v. 23, p. 800–810, Aug. 2001.

[33] J.L Marroquin,F. Girosi,,”Some Extentions of the K-Means Algorithm For Image Segmentation and Pattern Classification”, Technical Report, MIT Artificial Intelligence Laborartory, 1993.

[34] M.Luo, Y.F.Ma, H.J. Zhang,”ASpecial Constrained K-Means approach to Image Segmentation”,proc. The 2003 Joint Conference of Fourth International Conference on Informations Communications and Signal Processing and the Fourth Pacific Rim Conference on Multimedia, Vol.2, pp.738-742, 2003.

INSTRUCTIONS TO CONTRIBUTORS

The *International Journal of Image Processing (IJIP)* aims to be an effective forum for interchange of high quality theoretical and applied research in the Image Processing domain from basic research to application development. It emphasizes on efficient and effective image technologies, and provides a central forum for a deeper understanding in the discipline by encouraging the quantitative comparison and performance evaluation of the emerging components of image processing.

We welcome scientists, researchers, engineers and vendors from different disciplines to exchange ideas, identify problems, investigate relevant issues, share common interests, explore new approaches, and initiate possible collaborative research and system development.

To build its International reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for IJIP.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Started with Volume 9, 2015, IJIP will be appearing with more focused issues. Besides normal publications, IJIP intends to organize special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

We are open to contributions, proposals for any topic as well as for editors and reviewers. We understand that it is through the effort of volunteers that CSC Journals continues to grow and flourish.

LIST OF TOPICS
The realm of International Journal of Image Processing (IJIP) extends, but not limited, to the following:

- Architecture of imaging and vision systems
- Character and handwritten text recognition
- Chemistry of photosensitive materials
- Coding and transmission
- Color imaging
- Data fusion from multiple sensor inputs
- Document image understanding
- Holography
- Image capturing, databases
- Image processing applications
- Image representation, sensing
- Implementation and architectures
- Materials for electro-photography
- New visual services over ATM/packet network
- Object modeling and knowledge acquisition
- Autonomous vehicles
- Chemical and spectral sensitization
- Coating technologies
- Cognitive aspects of image understanding
- Communication of visual data
- Display and printing
- Generation and display
- Image analysis and interpretation
- Image generation, manipulation, permanence
- Image processing: coding analysis and recognition
- Imaging systems and image scanning
- Latent image
- Network architecture for real-time video transport
- Non-impact printing technologies
- Photoconductors

- Photographic emulsions
- Prepress and printing technologies
- Remote image sensing
- Storage and transmission
- Photopolymers
- Protocols for packet video
- Retrieval and multimedia
- Video coding algorithms and technologies for ATM/p

CALL FOR PAPERS

Volume: 9 - **Issue: 4**

i. Submission Deadline : June 30, 2015

ii. Author Notification: July 31, 2015

iii. Issue Publication: August 2015

CONTACT INFORMATION

Computer Science Journals Sdn BhD

B-5-8 Plaza Mont Kiara, Mont Kiara

50480, Kuala Lumpur, MALAYSIA

Phone: 006 03 6204 5627

Fax: 006 03 6204 5628

Email: cscpress@cscjournals.org

CSC PUBLISHERS © 2015
COMPUTER SCIENCE JOURNALS SDN BHD
B-5-8 PLAZA MONT KIARA
MONT KIARA
50480, KUALA LUMPUR
MALAYSIA

PHONE: 006 03 6204 5627
FAX: 006 03 6204 5628
EMAIL: cscpress@cscjournals.org