World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:2, No:9, 2008

Extended Study on Removing Gaussian Noise in Mechanical Engineering Drawing Images using Median Filters

Low Khong Teck, Hasan S. M. Al-Khaffaf, Abdullah Zawawi Talib, and Tan Kian Lam

***Abstract*—**In this paper, an extended study is performed on the effect of different factors on the quality of vector data based on a previous study. In the noise factor, one kind of noise that appears in document images namely Gaussian noise is studied while the previous study involved only salt-and-pepper noise. High and low levels of noise are studied. For the noise cleaning methods, algorithms that were not covered in the previous study are used namely Median filters and its variants. For the vectorization factor, one of the best available commercial raster to vector software namely VPstudio is used to convert raster images into vector format. The performance of line detection will be judged based on objective performance evaluation method. The output of the performance evaluation is then analyzed statistically to highlight the factors that affect vector quality.

***Keywords*—**Performance Evaluation, Vectorization, Median Filter, Gaussian Noise.

I. Introduction

VECTORIZATION is a process that converts a raster image into a corresponding vector image. The quality of the detected vectors is based on how accurate the attributes of the lines in the raster image were recognized by the vectorization methods. The detected vectors may have some distortions due to many operations that have been operated upon the raster images. Among the operations that may affect the quality of the detected vector are the amount of noise, the noise removal method, and the vectorization software used. The previous study [1], involved several noise removal algorithms and several algorithms that have been proposed. Several vectorization software systems have been used for the vectorization process and salt-and-pepper noise were applied to the image document. However, in reality the common noise is Gaussian noise and Gaussian noise may embed into an image document during scanning, printing or image capturing process. It is not a trivial task to remove the noise without affecting the quality of the image documents. Hence many noise removal algorithms will be examined and we will study their effects on the image document after the cleaning process.

So, cleaning of Gaussian noise in document images are studied using Median filter and its variants. Vector Recovery Index (VRI) is the criteria used to measure the quality of the vectors compared to their ground truth data [2]. Besides, statistical analysis is also performed on the VRI values to study the factors that affect vector quality.

There are three objectives in this extended study. The first is to study the effects of Gaussian noise in document images with low level of noise and high level of noise. The second is to study and analyze different noise removal algorithms focusing on two algorithms which are Median and Centre Weighted Median (CWM). The third is to study the efficiency of VPstudio [3] software with respect to different noise levels. The study will focus on the ability/efficiency of the software to convert from raster to vector. The output of the software will be compared with ground truth data. A performance evaluation method is used to measure the quality of the VPstudio detected vectors.

II. Background and Related Work

Weighted median (WM) is expanded from the median filter by including weight into the computation. By adjusting the weight, WM is able to control the smoothing behavior, and indirectly it is a promising image improvement method.

Centre weighted median filter is expanded from weighted median filter, but this filter emphasizes on more weight on the centre value of the window. It is easier to implement and design compared to weighted median filters. In order to enhance CWM filter, another technique has been proposed, which is adaptive centre weighted median filters [4].

Given a pixel $X_{(i,j)}$, the equivalent outcome using CWM filters can be defined as follows:

$$Y_{(i,j)} = median\{X_{(i-s,j-t)}, w_c \Diamond X_{(i,j)} \mid (s,t) \in W, (s,t) \neq (0,0)\} \quad (1)$$

Where w_c is the center weight and $\Diamond$ denotes the replication operator. For a 3*3 window, W can be given by

$$W = \{(s,t) \mid -1 \leq s \leq +1, -1 \leq t \leq +1\} \quad (2)$$

Low Khong Teck is MSc. Graduate from the School of Computer Sciences, Universiti Sains Malaysia, 11800 USM Penang, Malaysia (e-mail: khongteck@hotmail.com).

Hasan S. M. Al-Khaffaf is a PhD candidate in the School of Computer Sciences, Universiti Sains Malaysia, 11800 USM Penang, Malaysia (e-mail: hasan@cs.usm.my).

Abdullah Zawawi Talib, is with School of Computer Sciences, Universiti Sains Malaysia, 11800 USM Penang, Malaysia (e-mail: azht@cs.usm.my).

Tan Kian Lam is MSc. Graduate from the School of Computer Sciences, Universiti Sains Malaysia, 11800 USM Penang, Malaysia (e-mail: andrewtan2000@hotmail.com).

An adaptive median filter regulates the window size by relying on the input presented to deal among others detailed preservation and noise repression.

III. Proposed Methodology

In this extended study, the methodology proposed in [1] is applied, and the study carried out in [1] is extended by using Gaussian noise instead of uniform salt-and-pepper noise. Fig. 1 shows the flowchart of the methodology [1]. The experiment starts with noise being added to a clean image as an input data. The noise is then removed using the noise removal algorithms. Then, vectorization by commercial software is carried out to obtain detected vectors which are saved in DXF file. The file needs to be converted into a simple file format which is VEC file format. Performance evaluation is then carried out by comparing the result obtained from the experiment with the ground truth data. VRI value is computed by performance evaluation tool and statistical analysis for VRI is then carried out.

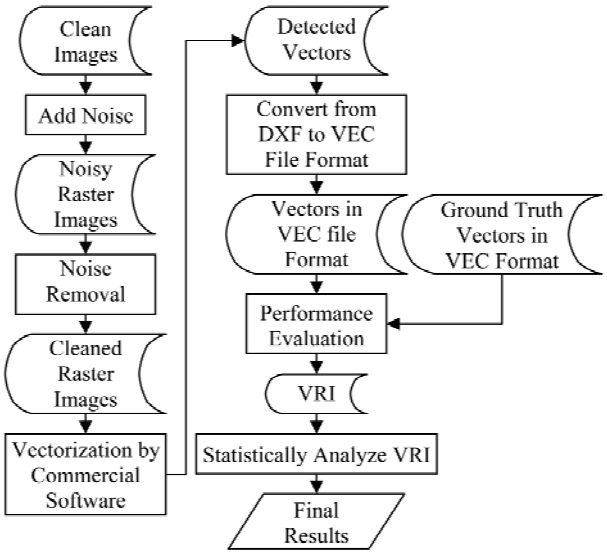

Fig. 1 Flowchart of the methodology. Taken from [1]

IV. Implementation

The implementation phase is important for this study in order to obtain the results of the experiment and analyze the results. The Median filters and Centre Weighted Median filters (CWM3 and CWM5) are implemented to filter Gaussian noise of mechanical engineering drawing images. The algorithms have been programmed in C# using Microsoft Visual Studio 2005. The operating system is Windows Vista running on AMD Athlon™ 64 X2 Dual Core processor 3800+ 2.0GHz and 2GB of RAM.

The program implementation starts with noisy input image. Then the image is filtered using the three types of noise removal algorithms used in this study namely Median filter, CWM 3, and CWM 5. The Median filter filters the noisy image with normal median filter algorithm. CWM 3 and CWM 5 filters filter the noisy image with centre weighted median 3 and 5, respectively. After these processes, clean images are obtained as final results. A 3 * 3 window is used in the implementation and it is shown in Fig. 2.

P[3]	P[2]	P[1]
P[4]	P	P[0]
P[5]	P[6]	P[7]

Fig. 2 Configuration of 3x3 window

V. Experiment

The first part of the study involves studying the performance of Median, CWM 3, and CWM 5 filters. In this part the noisy engineering drawing image is filtered with two levels of noise (10% and 40%) and the input data is a binary image. The image data used in the experiments are taken from GREC 07 [5].

Commercial vectorization software, VPstudio [3] is chosen for the vectorization process. The experiment is executed with the following limitations or constraints:

1) The vectorization software is a demo version of VPstudio V8.02 C6.
2) Performance evaluation tool is ArcEval2005.exe that will only work on arc entities [6]. Other graphical elements will be ignored.

The steps of the experiment using the software are as follows:

1) Experiment setting:
   a. Some parameters of the VPstudio software should be set prior to starting the experiment (i.e. before going through steps 2 to 5)
   b. Set the type of drawing to Mechanical drawing (MECH) from the Vectorize->Vectorize Parameter -> Load. Then load MECH.PCH.
2) Convert each raster image into vector form.
   a. Load raster image into VPstudio.
   b. Mirror the image vertically.
   c. Use Options->User coordinate System. Set User Units/Coordinate System to Pixels[px].
   d. Convert to vector form and save the output DXF files.
3) Convert all DXF files into VEC files.
4) Get performance of the vectorization process.
   a. Use ArcEval2005.exe as the performance evaluation tool.
5) Analyze the results of the performance evaluation.

VI. Result Analysis and Discussion

Table I below shows statistics of the VRI values. Mean is the average value of the VRI. From the statistical analysis of the VRI values, the mean value of 0.62173 is obtained. Median is the middle value from the sorted ascending VRI values. From

the statistical analysis of the VRI values, the median value of 0.64850 is obtained.

TABLE I
STATISTICS FOR VRI

N	30
Mean	0.62173
Median	0.64850
Min	0.373
Max	0.752

The minimum value of the VRI is 0.373 and the maximum value for VRI is 0.752. The mean value for the VRI which is 0.622 is below the satisfactory value of 0.8 as proposed by [6]. This could be due to the weak features of the original images as well as the amount of the noise added to the image. The mean value is close to the median suggesting normal distribution of the data.

Fig. 3 shows the frequency histogram of VRI. Based on this figure VPstudio can be said to be efficient and robust with different levels of noise as the higher frequencies for VRI are in the range of 0.5 to 0.8 (close to 1.0). This result has again confirmed that VPstudio is one of the most efficient vectorization software as in previous study by other researchers [1].

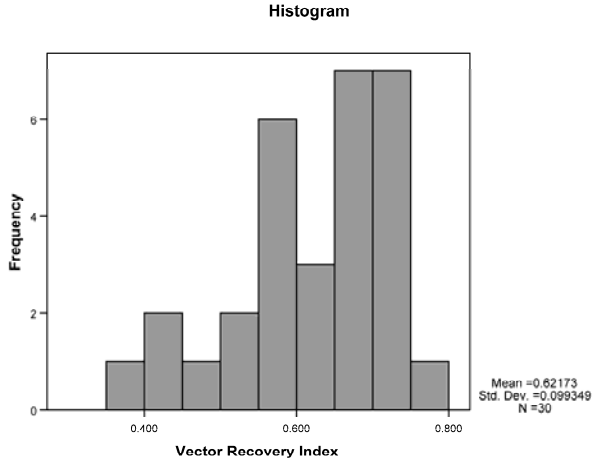

Fig. 3 Histogram of VRI

As shown in Fig. 4, Median filter performs better with high level of Gaussian noise compared to CWM. On the other hand, Centre Weighted Median algorithms (CWM 3 and CWM 5) perform better with low level of noise. Also as shown in Figure 9, for the image with low level of noise, CWM 5 performs better than CWM 3, but for the image with high level of noise, CWM 5 performs lower than CWM 3. CWM 5 performs well in 10% of noise compared to others, but this algorithm has the lowest performance for 40% noise.

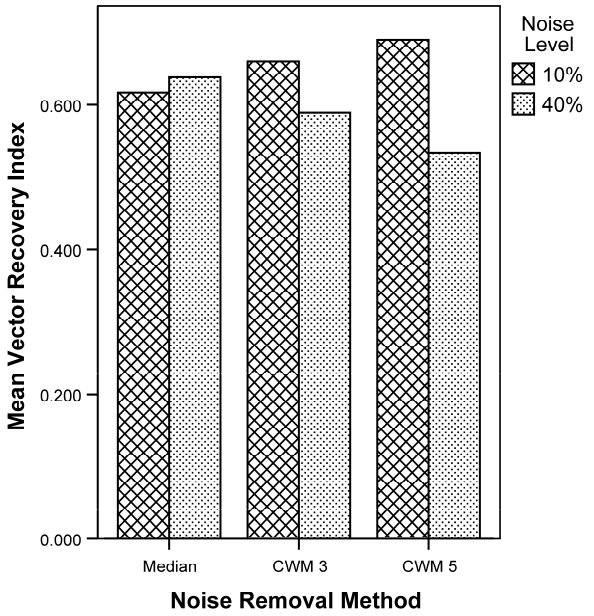

Fig. 4 Performance of noise removal algorithms with different noise levels

As a conclusion, Median filter can be said to be more efficient for high level of noise, but CWM is more efficient for low level of noise.

With 10% noise, CWM 5 filter shows better results compared to other algorithms. This is shown in Fig. 5. With 40% noise Median filter shows the best result compared to other filters. The performance of CWM 5 shows high sensitivity to noise level compared to other algorithms while CWM 3 shows moderate sensitivity to noise. Median filter shows more stability with low and high amount of noise.

With 10% noise the image detail is not corrupted much and the filters with small ability to remove noise such as CWM 5 show better performance compared to other filters that are sharper on removing noise (Median). The reason is that median filter distorts image detail when cleaning the image while CWM 5 preserves image detail. As a result CWM 5 shows better performance than median.

With 40% noise, algorithms that are sharper in cleaning noise have a high value of VRI. Median which can remove more noise compared to CWM 3 and CWM 5 shows the best performance with 40% noise, while CWM 3 which can remove more noise than CWM5 but less noise than median shows moderate performance (in VRI). CWM 5 which is less able to remove noise compared to the other two tested algorithms shows lower performance with 40% of noise.

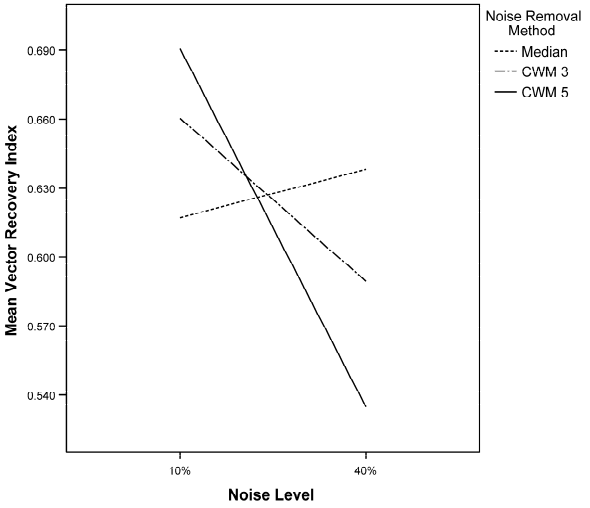

Fig. 5 Performance of noise removal algorithms with different levels of noise

VII. Conclusion and Future Work

The effect of the Gaussian noise with two levels of noise (10% (low) and 40% (high)) have been studied and they are used with three noise removal algorithms namely Median, CWM3 and CWM5 filters. The three noise removal algorithms have been studied and analyzed. The efficiency of a vectorization software namely VPstudio have also been studied.

From the result of the study, Median filter is deemed to have a low sensitivity to changing in noise level compared to the filters based on CWM. Median filter also exhibits the highest performance when the noise level is high. On the other hand, the filters based on Centre Weighted Median have the highest performance when the noise level is low. Based on the study, VPstudio is found to be efficient and robust with respect to different levels of noise.

For future work, more levels of noise can be added to the images in order to study the effect with a wider perspective. In this study, only two major noise removal algorithms are studied. So for future work, more noise removal algorithms should be included in the study. In fact for Gaussian noise, no study has been made on kFill [7], Enhanced kFill [8], Activity Detector [9], and other algorithms studied in [1,10]. More vectorization software can also be included, such as "Raster to Vector", "RasVector", and "VexaTractor" for a more comprehensive study on the vectorization factor.

With more noise removal algorithms and more vectorization software, it is possible to carry out a more statistically proven study by applying the Analysis of Variance (ANOVA).

Acknowledgment

The first author is supported by a grant from Ministry of Science Technology and Innovation (MOSTI) under project number 01-01-05-SF0147. The second author is supported by Universiti Sains Malaysia Fellowship.

References

[1] H.S.M. Al-Khaffaf , A.Z. Talib, and R. Abdul Salam, (2008) A Study on the effects of noise level, cleaning method, and vectorization software on the quality of vector data, Lecture Notes in Computer Science 5046, pp. 299-309.

[2] W.Y. Liu, and D. Dori, (1997) A protocol for performance evaluation of line detection algorithms, Machine Vision and Applications, 9(5-6): 240-250.

[3] VPstudio _ver_8. Raster to Vector Conversion Software, Softelec, Munich, Germany, [Online]. (Accessed. 10 Feb 2008) available for (http://www.softelec.com)

[4] S.-J Ko and Y.-H Lee, (1991) Center weighted median filters and their applications to image enhancement, IEEE Trans. Circuits Syst., 38: 984-993.

[5] F. Shafait, D. Keysers, T.M. Breuel, (2008) GREC 2007 Arc Segmentation Contest, Evaluation of Four Participating Algorithms. Lecture Notes in Computer Science 5046, pp. 310-320.

[6] L. Wenyin, Performance Evaluation tool (accessed on 2008) http://www.cs.cityu.edu.hk/~liuwy/ArcContest/

[7] W. Liu, (2004) Report of the Arc Segmentation Contest, in Graphics Recognition, Lecture Notes in Computer Science: 363—366.

[8] L. O'Gorman, (1992) Image and Document Processing Techniques for the RightPages Electronic Library System. Proc. 11th IAPR Int'l Conf. Pattern Recognition, 2: 260-263.

[9] K. Chinnasarn, Y. Rangsanseri, and P. Thitimajshima, (1998) Removing salt-and-pepper noise in text/graphics images, Proc. Asia-Pacific Conf. on Circuits and Systems. (APCCAS), Chiangmai, Thailand,: 459-462.

[10] P.Y. Simard and H. Malvar,(2004) An Efficient Binary Image Activity Detector Based on Connected Components, Proc. International Conference on Accoustic, Speech and Signal Processing (ICASSP), vol. 3: 229-232.

[11] H.S.M. Al-Khaffaf, A.Z. Talib, and R. Abdul Salam, (2006), Internal Report, Enhancing salt-and-pepper noise removal in binary images of engineering drawing. Artificial Intelligence Research Group, School of Computer Sciences, Universiti Sains Malaysia.