

Segmentation using Six Sigma Threshold on Spectral Bands of Malignant Melanoma

Gopalakrishnan Sethumadhavan and Srinivasan Sankaran

Abstract— This paper presents a method for analyzing the variations of the RGB spectrum of lesion skin images using the novel segmentation process based on Six Sigma concept. This analysis further contemplates on the incidence and propagation of cancer. It is based on the underlying principles of Dr. W.A. Shewhart's Control Charts, which focuses on the fact that the variability does exist in all repetitive processes. The heterogeneous color variation within the skin is considered as an assignable cause and is due to the secretion of excess melanin. These variations possess greater magnitude as compared to the chance causes due to the color variations found in normal skins. The power of control chart that lies in its ability to separate out this assignable cause, which is one of inherent mnemonics of Malignant Melanoma, is exhibited. The proposed Six Sigma based segmentation identifies the normal skin region from the regions of lesions, besides its fuzzy border. Results show that the method produces a robust segmentation of regions of high color contrast.

Index Terms— Melanoma, Color, Threshold, Six Sigma, Assignable cause.

I. INTRODUCTION

Malignant melanoma is named after the cell from which it presumably arises, the melanocyte. Melanocytes are the skin cells, which produce the dark protective pigment called melanin, a natural sunscreen. Melanoma cells usually continue to produce melanin, which accounts for the cancers appearing in mixed shades of tan, brown and black (variegated coloring). Melanoma has a tendency to metastasize; hence early detection and treatment are essential.

Friedman et al have enumerated the mnemonic "ABCD" to describe early malignant melanoma [1,2]:

- A**symmetry – One half of the tumor does not match the other half.
- B**order Irregularity – The edges are ragged, notched, blurred.
- C**olor – Pigmentation is not uniform. Shades of tan, brown, and black are present. Dashes of red, white and blue add to the mottled appearance.
- D**iameter – Greater than 6 mm and growing.

Computer vision methods have been previously applied to the problem of skin tumor diagnosis [1,2,3,4,5,6,7,8,9,10,11].

Manuscript received December 3, 2008. The Management of SASTRA University and HCL Technologies Limited supported this work.

S.Gopalakrishnan is with the Department of Computer Applications, School of Computing, SASTRA University, Thanjavur, India-613 402.

S.Srinivasan is with H.C.L Technologies Limited, Chennai, India – 603 103

In general, these methods are used to find tumor boarders, segment out the tumor from the rest of the image, and extract features from the tumor.

Tim Lee et al. [12] tried to diagnose the malignant melanoma through its color images by applying multistage central weighted median filter. Based on the different categories it is reported that 80% of images falls in the good and fair categories. It was also observed that the separation between a lesion and the surrounding skin is fuzzy.

In the proposed approach, the image is divided into rectangular sub regions small enough to be considered as having only homogenous color. Then based on the control chart methodology the significant sub regions are identified and coalesced for region growing. This method brings an effective segmentation of regions of high color contrast.

II. BACKGROUND OF VARIABLE CONTROL CHARTS

A control chart is a graphical representation of the collected information. It detects the variation in processing and warns of any possible departure from the specified tolerance limits. The control limits on the chart are so placed as variation. There are many types of control charts designed for different control situations, each with its own advantages and disadvantages [4] and with its own field of application. The control charts that are most commonly used for measuring variables are \bar{X} and R charts.

\bar{X} and R charts are used in combination for the control process. \bar{X} chart shows the centering of the process, i.e., the variation in the averages of sample. R-Chart shows the uniformity or consistency of the process i.e., the variations in the ranges of samples [13,14].

There is a statistical relationship between the mean range for data from a normal distribution and σ , the standard deviation of that distribution. This relationship depends only on the sample size, n . The mean of R is $d_2\sigma$, where the value of d_2 is also a function of n . An estimator of σ is therefore R/d_2 [15]. Armed with this background \bar{X} and R control charts are explained below:

A. \bar{X} - Chart

Let R_1, R_2, \dots, R_k , be the range of k samples. The average range is $\bar{R} = \frac{R_1 + R_2 + \dots + R_k}{k}$. Then an estimate of σ

can be computed as $\hat{\sigma} = \frac{R}{d_2}$. So, if we use $\bar{\bar{x}}$ (or a given

target) as an estimator of μ and \bar{R}/d_2 as an estimator of σ ,

then the parameters of the \bar{X} -chart are

$$UCL = \bar{\bar{x}} + \frac{3}{d_2\sqrt{n}}\bar{R}, \quad CenterLine = \bar{\bar{x}}, \text{ and}$$

$$LCL = \bar{\bar{x}} - \frac{3}{d_2\sqrt{n}}\bar{R}.$$

The simplest way to describe the limits is to define the

factor $A_2 = \frac{3}{d_2\sqrt{n}}$ and the construction of the \bar{X}

becomes

$$UCL = \bar{\bar{x}} + A_2\bar{R} \quad (1)$$

$$CenterLine = \bar{\bar{x}} \quad (2)$$

$$\text{and } LCL = \bar{\bar{x}} - A_2\bar{R} \quad (3)$$

The factor A_2 depends only on n [15].

B. R-Chart

This chart controls the process variability since the sample range is related to the process standard deviation. The center line of the R chart is the average range.

To compute the control limits we need an estimate of the true, but unknown standard deviation $W = R/\sigma$. This can be found from the distribution of $W = R/\sigma$ (assuming that the items that are measured follow a normal distribution). The standard deviation of W is d_3 , and is a known function of the sample size, n .

Therefore since $R = W\sigma$, the standard deviation of R is $\sigma_R = d_3 \sigma$. But since the true σ is unknown, we may

estimate σ_R by $\hat{\sigma}_R = \frac{d_3\bar{R}}{d_2}$. As a result, the parameters

of the R chart with the customary 3-sigma control limits are

$$UCL = \bar{R} + 3\hat{\sigma}_R = \bar{R} + 3d_3\frac{\bar{R}}{d_2}, \quad CenterLine = \bar{R},$$

$$\text{and } LCL = \bar{R} - 3\hat{\sigma}_R = \bar{R} - 3d_3\frac{\bar{R}}{d_2}.$$

As was the case with the control chart parameters for the subgroup averages, defining another set of factors will ease the computations, namely: $D_3 = 1 - 3d_3/d_2$ and $D_4 = 1 + 3d_3/d_2$. These yield

$$UCL = \bar{R}D_4 \quad (4)$$

$$CenterLine = \bar{R} \quad (5)$$

$$\text{and } LCL = \bar{R}D_3 \quad (6)$$

The factors D_3 and D_4 depend only on n [15].

III. IMAGES, SOFTWARE AND DATABASE

The images used in this research were digitized by researchers and were uploaded in various websites. These images have spatial resolution of 300 x 300 pixels, a brightness resolution of 256 levels per color plane (8 bits), and consisted of three-color planes (Red, Green and Blue) for a total of 24 bits per pixel. Further, they have a rich variation of color and are chosen to test the performance of Six Sigma thresholding.

A package was developed in Visual Basic 6.0 to fetch, store and process the images using M.S Access 2003 as backend and a DLL was developed in Visual C++ 6.0 for displaying the processed image by fetching the processed data from the database. The code verification is done with the software tool "Beyond Compare" on Intel Core 2 Duo (x86 architecture) under the Windows XP operating system.

IV. SIX SIGMA THRESHOLD

Many practical segmentation problems need more information than is contained in one spectral band. Color images are natural example, in which information is coded in three spectral bands. Usually multi-spectral threshold based segmentation approach is done by determining thresholds independently in each spectral band and are combined into a single segmented image. The methodology suggested in this paper focuses on the segmentation process based on the color variability. The equations presented in Section 2 are used to define different thresholds viz., **threshold-UCL- \bar{X}** , **threshold-LCL- \bar{X}** , **threshold-UCL-R**, and **threshold-LCL-R**.

A. Six Sigma thresholding on Images

RGB color plane, which is used as one of the representations of the pixel color values, is considered for analysis. The information extracted from control charts depends on the basis of selection of subgroups with maximum homogeneity. To facilitate this, the image of size $M \times N$ is sliced into small sub regions of sizes $m \times n$, where $m=n=4,5$ are considered as samples of size mn each. These MN/mn samples are numbered in column majoring.

For plotting control charts, $\pm 3\sigma$ limits are calculated and they are termed as control limits. Assignable causes of variation are identified by the way of finding the sub regions falling outside the control limits on either \bar{X} or R chart. The falling of all sub regions, within the control limits indicates that the spreading of epidermis is in control [10,11]. It really means that for all practical purposes it acts as if no assignable cause of variation is present [13,14].

The variation in the color propagation is analyzed by considering the sub regions, which falls outside the control limits. The pixels belong to these sub regions, which fall outside the control limits in each slicing as well as falls on these common intersections between various slicing, are identified. These pixels are termed as critical pixels and are due to the color variation (assignable cause). Figure 1 shows the schema for the segmentation using Six Sigma threshold.

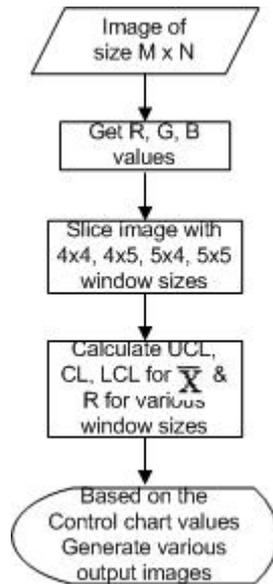


Figure 1. Schema for Segmentation using Six Sigma Threshold

B. Intersection regions

In this section, various possibilities of the intersection sub regions, which fall outside the control limits due to (4 x 4), (4 x 5), (5 x 4) and (5 x 5) slicing are presented. It is noted that the pixel (x_{i,j}) due to the Category 2 is well significant because of the uniqueness [16,17]. The lower the window size, higher will be the accuracy. The following are the different categories of intersection regions:

Category 1: (x_{i-1,j-1}, x_{k,l}) ∩ (x_{i,j}, x_{k+2,l+2}) ∩ (x_{i-1,j}, x_{k,l+2}) ∩ (x_{i,j-1}, x_{k+2,l}) = (x_{i,j}, x_{k,l}); i=6, 13, 26, 33, 46,53,... etc.; j=6, 13, 26, 33, 46,53,... etc ; k=i+2 and l=j+2;

Category 2: (x_{i,j}, x_{i+3,j+3}) ∩ (x_{i-4,j-4}, x_{i,j}) ∩ (x_{i,j-4}, x_{i+3,j}) ∩ (x_{i-4,j}, x_{i,j+3}) = (x_{i,j}); i=5, 16, 25, 36, 45, 56,... etc. and j=5, 16, 25, 36, 45, 56,... etc.;

Category 3: (x_{i,j-1}, x_{k+3,l}) ∩ (x_{i-4,j}, x_{k,l+2}) ∩ (x_{i,j}, x_{k+3,l+2}) ∩ (x_{i-4,j-1}, x_{k,l}) = (x_{i,j}, x_{k,l}); i=6, 13, 26, 33, 46, 53,... etc.; j=6, 13, 26, 33, 46,53,... etc ; k=i and l=j+2 or k=i+2 and l=j;

Category 4: (x_{i,j}, x_{k+3,l+2}) ∩ (x_{i-4,j-3}, x_{k,l}) ∩ (x_{i,j-3}, x_{k+3,l}) ∩ (x_{i-4,j}, x_{k,l+2}) = (x_{i,j}, x_{k,l}) i=5, 16, 25, 36, 45, 56,... etc.; j=9, 11, 29, 31, 49,51,... etc ; k=i and l=j+1 or k=i+1 and l=j;

Category 5: (x_{i-1,j}, x_{k,l+2}) ∩ (x_{i,j-3}, x_{k+2,l}) ∩ (x_{i-1,j+3}, x_{k,l}) ∩ (x_{i,j}, x_{k-2,l+2}) = (x_{i,j}, x_{k,l}); i=6, 13, 26, 33, 46,53,... etc.; j=9, 11, 29, 31, 49, 51,... etc ; k=i+2 and l=j+1 or k=i+1 and l=j+2;

Category 6: (x_{i,j-1}, x_{k+3,l}) ∩ (x_{i-4,j}, x_{k,l+2}) ∩ (x_{i,j}, x_{k+3,l+2}) ∩ (x_{i-4,j-1}, x_{k,l}) = (x_{i,j}, x_{k,l}) where i=5, 16, 25, 36, 45, 56,... etc. and; j=6, 13, 26, 33, 46,53,... etc.

where (x_{i,j}, x_{k,l}) is the set of pixels bounded by the

rectangle with vertices (i,j; k,j; i,l and k,l).

V. LESION SEGMENTATION USING SIX SIGMA THRESHOLD, RESULTS AND DISCUSSIONS

Images with lesions located near the center and surrounded by the normal skin are considered. Usually, the intensity of the pixels in the normal skin is uniformly distributed, but lesions are varying in color, size, and shape. However, in all cases the separation between the lesions and the surrounding skins are fuzzy.



Figure 2. Melanoma image Sample 6 with region marking

To explain the working procedure, region of size 40 x 40 having high color variation within (identified with naked eye) is chosen from the skin lesion image. (Figure 2. shows the lesion image with region of interest). This region is sliced into 100 samples each with window size 4 x 4. These samples are numbered using column majoring [18] and are represented in the horizontal axis. Mean and range of red color of these samples are in the vertical axis. Figure 3 and Figure 4 show \bar{X} and R charts of the region of interest.

It was observed that 70% and 5% of the pixels are outside of the control limits in these respective charts. Applying this procedure on the whole of the image, Six Sigma thresholds are computed for R, G, B, and Average color with different slicing and are tabulated in Table 1.

Visualization of Six Sigma thresholding of an image f to an output image g is as follows:

$$g(i, j) = \begin{cases} \text{red} & \text{if } x(i, j) \geq \text{Threshold} - \bar{X} \\ \text{blue} & \text{if } x(i, j) \leq \text{Threshold} - \text{LCL} - \bar{X} \\ \text{Yellow} & \text{otherwise} \end{cases}$$

where x(i,j) = R/G/B/Average of the pixel (i,j) ∨ (i,j).

Figure 5 shows the corresponding visualizations of a sample image. The fine and fuzzy edges of the lesion are detected by the pixels, which fall inside the control limits. It was observed that the segmentation due to **threshold-UCL- \bar{X} and threshold-LCL-R** shows the normal and lesion skins respectively.

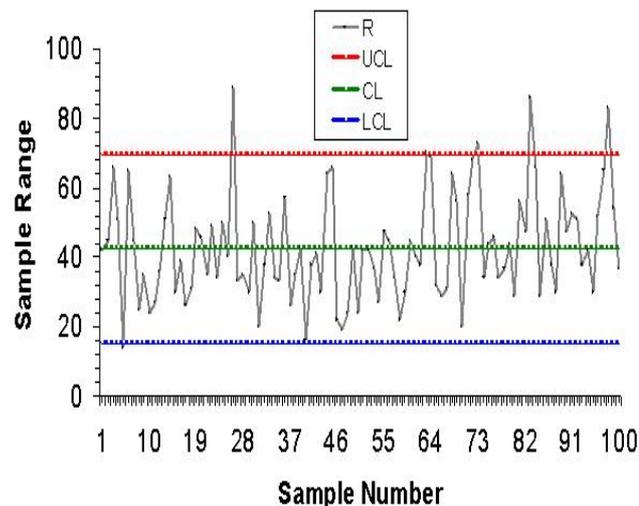
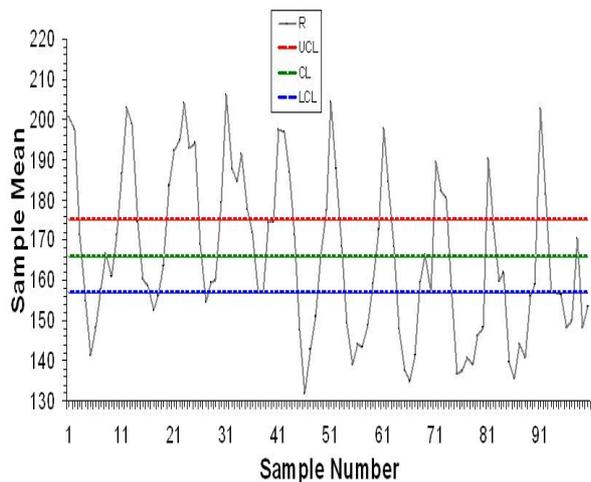


Figure 3. \bar{X} -Chart of the marked region of Figure 2 with (4 x 4) window size for Red color component

Figure 4. R-Chart of the marked region of Figure 2 with (4 x 4) window size for Red color component

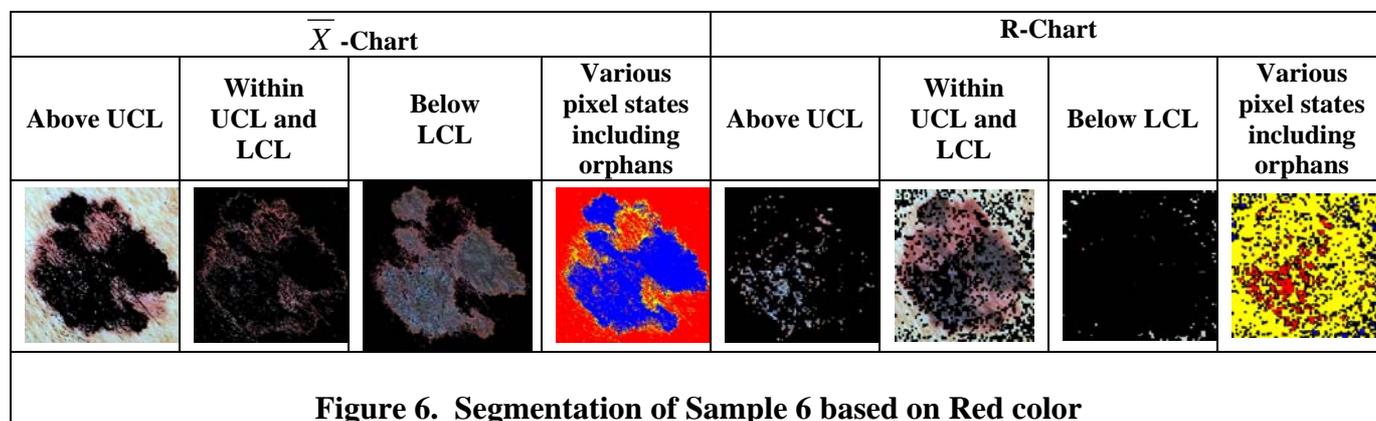
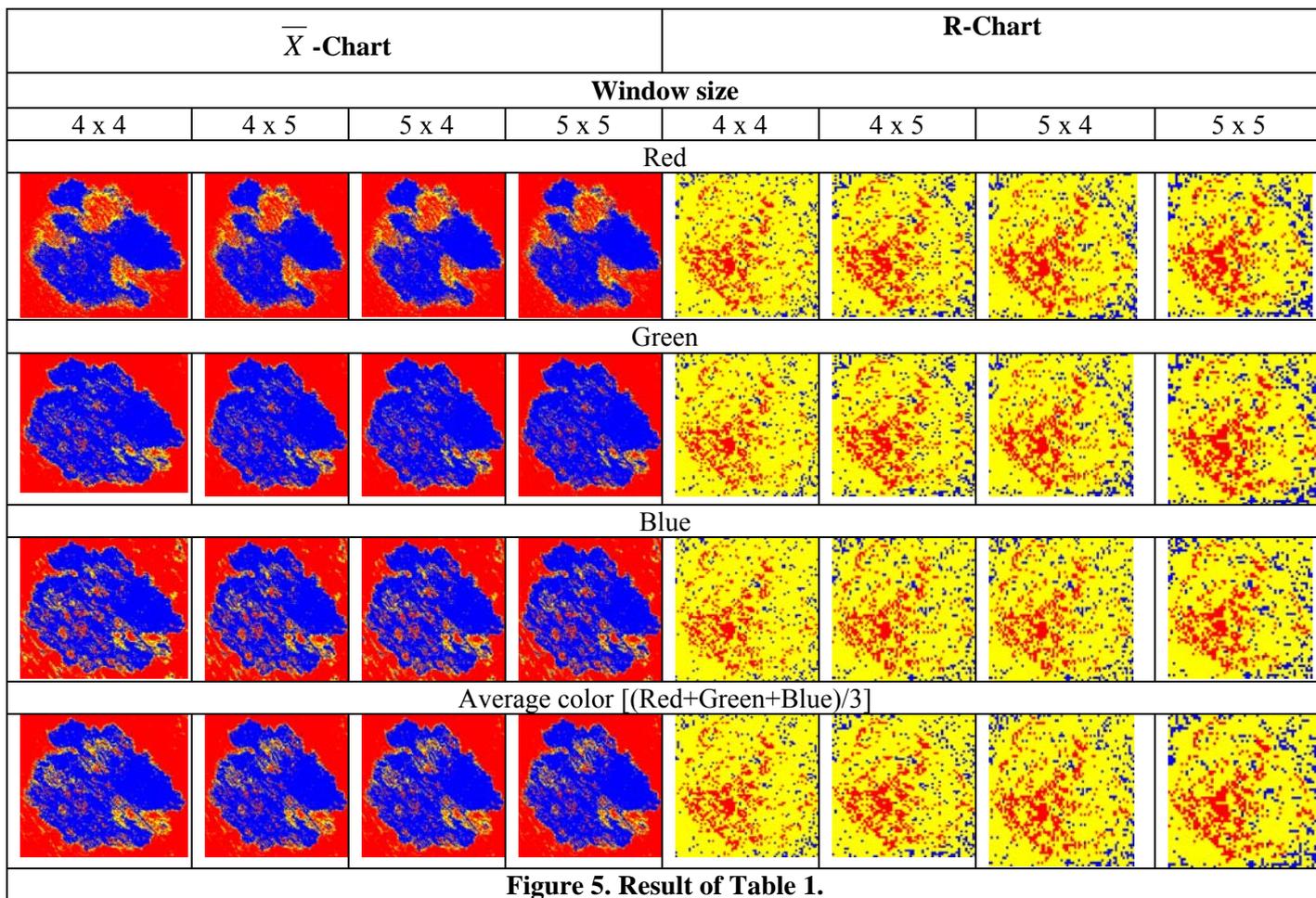
Table 1. Variable control charts data for various colors of different window sizes

Window size	\bar{X} -Chart						R-Chart					
	Threshold			Number of pixels			Threshold			Number of Pixels		
	UCL	CL	LCL	Above UCL	Within UCL & LCL	Below LCL	UCL	CL	LCL	Above UCL	Within UCL & LCL	Below LCL
Red												
4 x 4	162.86	151.05	139.24	47503	9568	32929	91.19	55.70	20.22	11312	73488	5200
4 x 5	161.93		140.17	47952	8761	33287	95.78	60.43	25.07	12560	69960	7480
5 x 4	162.01		140.09	47503	9210	33287	96.50	60.88	25.26	12300	70320	7380
5 x 5	161.10		141.01	47952	8398	33650	101.16	65.64	30.13	12875	67400	9725
Green												
4 x 4	151.89	139.75	127.60	39052	6716	44232	93.76	57.27	20.79	11456	73712	4832
4 x 5	150.63		128.87	39271	6141	44588	95.78	60.43	25.07	12560	69960	7480
5 x 4	150.71		128.79	39271	6141	44588	96.50	60.88	25.26	12300	70320	7380
5 x 5	149.79		129.70	39476	5603	44921	101.16	65.64	30.13	12875	67400	9725
Blue												
4 x 4	154.28	141.65	129.03	38259	10059	41682	97.46	59.54	21.61	10768	74992	4240
4 x 5	153.39		129.91	38644	9674	41682	103.36	65.21	27.06	11700	71460	6840
5 x 4	153.49		129.82	38644	9674	41682	104.21	65.75	27.28	11580	71620	6800
5 x 5	152.58		130.73	39036	8880	42084	110.00	71.38	32.76	12225	69625	8150
Average Color [(Red+Green+Blue)/3]												
4 x 4	156.18	144.15	132.12	40558	9332	40110	92.90	56.75	20.60	11616	73600	4784
4 x 5	155.28		133.02	40885	8549	40566	97.97	61.81	25.65	12480	70780	6740
5 x 4	155.35		132.95	40767	8805	40428	98.65	62.24	25.83	12360	70620	7020
5 x 5	154.45		133.85	41088	8062	40850	103.71	67.30	30.89	12775	68200	9025

Final step of the process is to identify the contributing pixels with respect to various slicing and based on Six Sigma threshold. Pixels that fall in the different control limits in various slicing are identified as per categories presented in Section IV B.

Pixels, which fall on any of these categories, are termed as Orphan pixels. These pixels are activated with black color and are considered to be the error incurred in this method. Figure 6 shows the segmentation of a sample of the result based on the red color. It is observed that the healthy and lesion regions can be clearly identified through Six Sigma segmentation. The segmentation due to R chart is the

indicator of the propagation of cancer. Further, it was also noted that the fine and fuzzy border detection are quite accurate which also indicates the propagation of cancer to the adjacent regions. It was also noted that the orphan pixels are at a maximum of 2%. The proposed method is tested on a set of 35 malignant lesions and the results of the segmentation with respect to RGB of 9 lesions are shown in Figure 7. The results of various analysis and output images are available at <http://www.sastra.edu/mnews/six%20sigma%20segmentation%20on%20melanoma.rar>



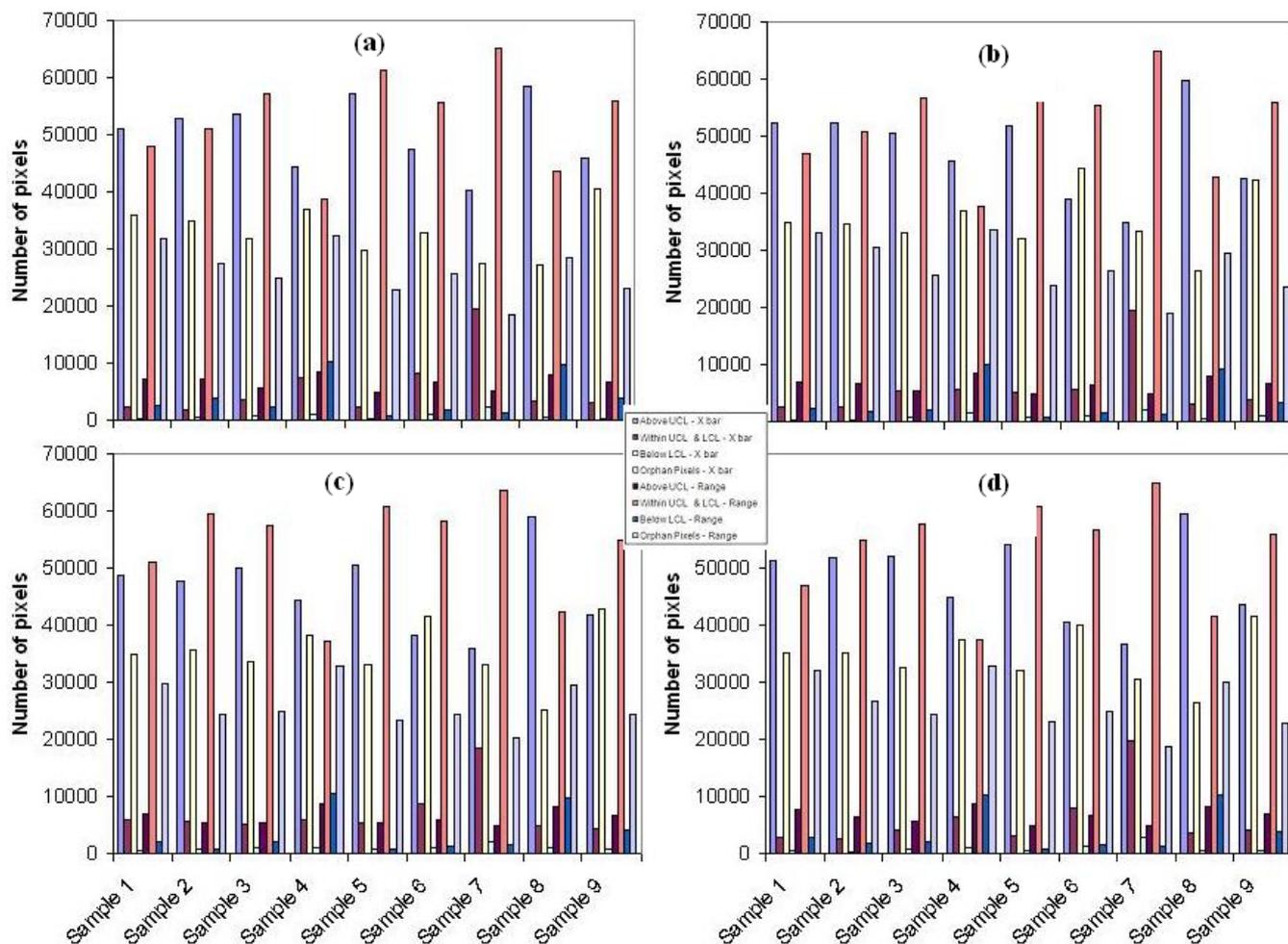
VI. CONCLUSIONS AND FUTURE WORK

This research is aimed at identifying the lesion and fuzzy border of the malignant melanoma based on the color information of images through the segmentation called “Six Sigma”. The results obtained from this method are very promising, as the identification of lesion correlates well with all the samples. The proposed segmentation method may not perform well on images with significant amount of hair. For such images, a preprocessor that can eliminate hairs such as DullRazor™ [19] may improve the results.

The methodology proposed in this study need not be confined to melanoma analysis alone. It opens a new approach on segmentation of images of the other fields too. The next phase of this work, which is already an ongoing research, will be the automatic detection of the degree of fuzziness found in the melanoma.

VII ACKNOWLEDGEMENTS

The authors would like to acknowledge the support rendered by the Managements of SASTRA and HCL Technologies in the way of providing necessary infrastructure and financial support for this research. The websites www.dermatlas.org and <http://skincancer.dermis.net> are also gratefully acknowledged.



**Figure 7. Number of pixels common to (4x4); (4x5);(5x4) and (5x5) window sizes
 (a) Red color (b) Green color (c) Blue color and (d) Average color**

REFERENCES

- [1]. Friedman R, Rigel D, Kopf A. "Early detection of malignant melanoma: The role of physician examination and self-examination of the skin," CA 35 pp 130-151, 1985.
- [2]. Friedman R.J and Riegel D.S. "The clinical features of malignant melanoma," Dermatologic Clin., vol. 3, pp. 271-283, 1985.
- [3]. Dhawan A.P, Sicsu A. "Segmentation of images of skin lesions using colour and texture information of surface pigmentation," Computerized Medical Imaging and Graphics. 16: 163-177; 1992.
- [4]. Ganster H., Pinz A., R'ohrer R., Wilding E., Binder M., Kittler H. Automated Melanoma Recognition, IEEE Trans. on Medical Imaging 20 , pp. 233-239, 2001.
- [5]. Golston JE, Stoecker WV, Moss RH. "Automatic detection of irregular borders in melanoma and other skin tumors," Computerized Medical Imaging and Graphics. 16: 199-203; 1992.
- [6]. Stoecker W.V, Moss R.H. "Digital imaging in dermatology," Editorial, Computerized Medical Imaging and Graphics. 16 145-150; 1992.
- [7]. Umbaugh S.E. "Computer Vision in Medicine: Colour metrics and image segmentation methods for skin cancer diagnosis." Ph.D. Dissertation, Electrical Engineering Department, University of Missouri-Rolla. UMI Dissertation Services. AnnArbor, MI; 1990.
- [8]. Umbaugh SE, Moss RH, Stoecker WV. "An automatic colour segmentation algorithm with application to identification of skin tumor borders," Computerized Medical Imaging and Graphics.
- [9]. Umbaugh SE, Moss RH, Stoecker WV. "Applying artificial intelligence to the identification of variegated colouring in skin tumors," IEEE Eng. Med. Biol. 8: 43-52, 1989.
- [10]. Umbaugh SE, Moss RH; Stoecker, W.V. "Automatic colour segmentation of images with application to detection of variegated colouring in skin tumors," IEEE Eng. Med. Biol. 10: 57-62; 1991.
- [11]. Umbaugh, S.E "Computer vision in medicine: Colour metrics and image segmentation methods for skin cancer diagnosis," Ph.D. dissertation, Elect. Eng. Dep., Univ. Missouri-Rolla, Rolla, MO: UMI Dissertation Services, Ann Arbor, MI, 1990
- [12]. Tim Lee, Vincent Ng., David McLean, Andrew Coldman, Richard Gallagher, and Joanna Sale. " A Multi-Stage Segmentation Method for Images of Skin Lesions", IEEE Pacific Rim Conf. Communications, Computers and Signal Processing, Victoria, BC, Canada, 602-605; 1995.
- [13]. Douglas C Montgomery. "Introduction to Statistical Quality Control", Second Edition, John Wiley & Sons, Inc., 1991.
- [14]. Statistical Quality Control, Eugene L.Grant and Richard S.Leavenworth, 6th editions, McGraw-Hill International Edition, 1988.
- [15]. ASTM Manual on Presentation of Data, American Society for Testing and Materials, Philadelphia, 1945.
- [16]. Thiyagarajan M and Gopalakrishnan S, and Jagannathan B. "Pattern Identification of fingerprints using colour spread process index," ACCST Research Journal Volume 1, NO.2, April 2003.
- [17]. Thiyagarajan M and Gopalakrishnan S. "Micro level analysis for incidence and propagation of skin cancer through Quality Control Tool," Proceedings of XXXVIII National Convention of Computer Society of India 2003.
- [18]. Trembley. J.P & Sorenson. P.G, "An Introduction to Data Structures with Applications", Second Edition, Tata McGraw Hill, 1981.
- [19]. Lee T.K. Ng V., Gallagher R., Coldman A., and McLean D. (1997) "Dullrazor – A Software Approach to Hair Removal from Images" Computers in Biology and Medicine, 27(6):533-543.