Application of 'TissueQuant' Algorithm to Automate Quantification of ck7 Stain for the Diagnosis of Sinovial Sarcoma

Keerthana Prasad
MCIS
Manipal

Bernhard Zimmermann
MCIS
Manipal

Gopalakrishna Prabhu K
Department of Biomedical Engg.
MIT, Manipal

ABSTRACT
This paper presents the application of 'TissueQuant' algorithm which is a unique approach to assess color shade similarities to quantify the extent of staining to arrive at a diagnosis. This is done by assigning scores to different shades of a particular color based on how close the shade is to a reference color. This approach is applied to the quantification of ck7 stain for diagnosis of sinovial sarcoma. The images were obtained from the tissue microarray database maintained by Stanford university (website: http://tma.stanford.edu/cgi-bin/cx?n=SSTMA18-new). This algorithm is based on HSI color model. With the color of interest taken as the centre of a Gaussian function, appropriate weightage is given to different shades of the specific color. Using this technique as the basis, the specific stain is quantified. This information is then used for diagnosis of sinovial sarcoma. This approach is very useful in handling tissue microarray images due to its high throughput. The results were compared with the diagnosis provided by the database and were found to be very much in agreement with it.

General Terms
Image analysis, computer aided diagnosis

Keywords
TissueQuant, stain quantification, color quantification.

1. INTRODUCTION
Computational power has dramatically increased in the past decade which has led to developments in the area of digital microscopy. It has introduced more systems to automate tissue processing, staining, image acquisition and management of virtual slides. Currently, most of the pathological diagnosis is done based on evaluation by pathologists, which is subjective and qualitative. Researchers from both computer science and medical science recognize the importance of quantitative, image-based evaluation of microscopic slides. Visual interpretation of images is an integral part of medical diagnostic procedures. However, it is tedious, time consuming and error prone. Computerized image analysis has been widely used in medical research facilitating quick, repeatable and objective evaluation which makes it possible to obtain statistically significant results. The tissue microarray has enabled high-throughput pathology. Rather than the laborious review of individual slides and issues of assay reproducibility across large series of specimens, tissue microarrays allow the review of a single stain on a single slide containing tens to hundreds of samples. This is a paradigm shift in pathology, away from histomorphology and toward molecular characterization by immunohistochemistry. This platform allows large retrospective clinical studies of biomarkers for correlation with outcome and can equally well be applied toward high-throughput analysis of cell lines and xenografts. Tissue microarrays encourage novel approaches to assaying tissue with retained histomorphology and have enabled image analysis in pathology. The reduction of tissue to an analyte for high-throughput analysis has highlighted the importance of a high quality tissue and the impact of tissue handling and processing in the quality of data that can be obtained from analysis of tissue. This kind of analysis can be done by the computer relatively quickly with excellent reproducibility. The need to handle huge number of images from the tissue microarray assay which is a high throughput technique, automation of quantification of the stain and the diagnosis is highly desirable.

Many research groups have put efforts to apply digital image analysis for the quantification of a biological substance for study of diseases like breast cancer [1,2,3,4], lung diseases [5], skin diseases [6,7], nephropathy [8], tuberculosis [9], atheromatous lesions of human aorta and coronary arteries [10]. Digital image analysis has also been tried for quantification of stained specimen for medical research in [11,12,13,14]. Some groups performed analysis in gray scale, some others used threshold and area measurement. Some groups stored the possible color shades of region of interest in a color file with the help of user interaction and the same file was consulted to evaluate rest of the images in the dataset. Color quantization [15] and color clustering algorithms [16] have also been tried to classify the regions in an image into foreground or background. All these techniques to assess amount of stain and hence the biological substance under study, just determine the shades of a particular color of interest and select all pixels with those shades as foreground pixels. This approach does not take into account, the possibility of the substance being present in varying concentrations across the specimen. A technique which can grade the shades of the color of interest could be more useful to handle such variations. 'TissueQuant' algorithm [17] assigns lesser scores to paler shades and higher scores to deeper shades thus providing better quantification of the substance under study. Any other color shade is given a score of zero.

We present the application of 'TissueQuant' algorithm to assign scores to the quantification of ck7 stain for the diagnosis of sinovial sarcoma.

2. METHOD

2.1 TissueQuant Algorithm

The main use of this algorithm is to grade the shades of a color based on how much they differ from the pure color. The algorithm uses HSI color space. Different shades of a particular color of interest are given appropriate scores based on a Gaussian weighting functions to the three color components of the color space, with the chosen color at the centre. Using this technique as the basis, the specific stain is quantified.

The image is first represented in HSI color space. The Hue, Saturation and intensity components of every pixel with coordinates (i, j) are represented as H(i, j), S(i, j) and I(i, j) respectively. These components are studied in comparison with the reference color and are assigned appropriate scores. The formula used for obtaining the score is based on how close the hue, saturation and intensity components are to the hue (H_c) Saturation (S_c) and Intensity (I_c) components of the color of interest. The widths of the Gaussian weighting functions are decided by the different ranges for the hue (H_w), Saturation (S_w) and intensity (I_w) components. These values decide the range of shades of the color which should be given a score. The ck7 stain gives brown staining for the sinovial sarcoma positive cells. Hence we choose the centres and widths for hue, saturation and intensity for this color as reference and obtain the Gaussian weighting function to obtain the score for each of the pixels.

The score for the pixel at (i, j) is obtained from the weighting functions for the three components as follows. The score for how close the hue of pixel (i, j) is to the reference color is represented as $I_H(i, j)$, score for how close the saturation is to the reference color is represented as $I_S(i, j)$ and score for how close the intensity is to the reference color is represented as $I_I(i, j)$. To take care of the fact that colors at the end of the hue scale are similar to colors at the beginning of the scale, the function $I_H(i, j)$ is made to wrap around at the limits of the scale. The corresponding equation for a scale [0,1] is shown below. The cumulative score is given as the product of the three components represented as follows.

$$I(i,j) = I_H(i,j) \cdot I_s(i,j) \cdot I_I(i,j)$$

$$I_H(i,j) = e^{\frac{-(H(i,j)-H_C-1)^2}{H_W{}^2}} + e^{\frac{-(H(i,j)-H_C)^2}{H_W{}^2}} + e^{\frac{-(H(i,j)-H_C+1)^2}{H_W{}^2}}$$

$$I_S(i,j) = e^{\frac{-(S(i,j)-S_C)^2}{S_W{}^2}}$$

$$I_I(i,j) = e^{\frac{-(I(i,j)-I_C)^2}{I_W{}^2}}$$

2.2 Application to quantify ck7 stain

The scores for all the pixels are calculated using the 'TissueQuant' algorithm as described above. The mean of the pixel scores (Mean) is taken as one feature used for classification. The image with strong response to the ck7 stain in a very small area also represents a positive diagnosis. Hence, as another feature the mean of the highest pixel scores (Mean-Max) was studied. For the pictures we worked with, it seemed optimal to consider only the four highest scored pixels. These features are used for the diagnosis of sinovial sarcoma.

3. RESULTS

The proposed method has been applied to the images obtained from the website: http://tma.stanford.edu/cgi-bin/cx?n=SSTMA18-new. Seventy nine images were downloaded from this website and were studied for the diagnosis of sinovial sarcoma. The Gaussian weighting function was applied to provide a color score for each of the pixels in the image. The features namely mean and Mean-Max are calculated from these pixel scores. The results are summarized in the following figures.

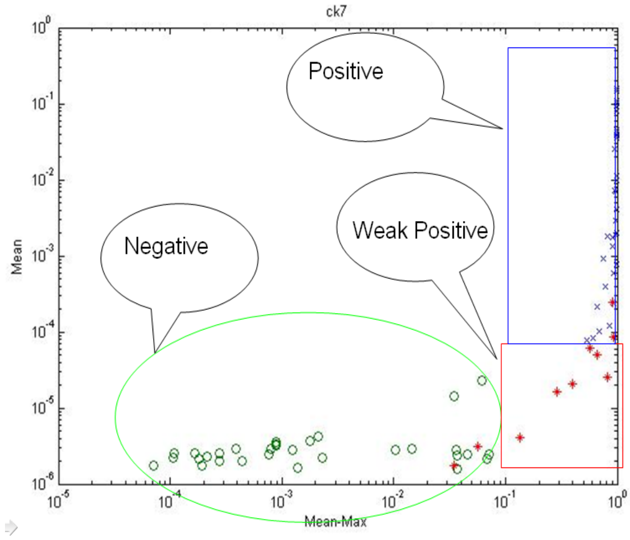

Fig 1: The negative cases are shown as green circles, the red stars are weak positive cases and the blue crosses in the above figure shows positive cases.

Table **1: Frequency table for the 79 cases**

Frequency table				
Mean-Max score	Neg	WkPos	Pos	Total
<30	23	0	0	23
30-100	8	3	0	11
100-400	0	5	0	5
400-950	0	7	14	21
>950	0	0	19	19
Total	31	15	33	79

Table 2: **Confidence percentage for the classification**

	Confidence (%)		
Mean-Max Score	Neg	WkPos	Pos
<30	100	0	0
30-100	73	27	0
100-400	0	100	0
400-950	0	33	67
>950	0	0	100

The classification can be efficiently done if we consider both the features namely Mean-Max and Mean, as depicted in Fig 1. The figure shows that the two features are able to properly classify the different cases. However, we can observe that Mean-Max score alone can also provide quite good results. The score obtained for the image is used to classify the given sample as positive, weak positive or negative. Table 1 shows the distribution of the cases into the different classes called negative, weak positive and positive. The confidence percentages of classification for these cases are tabulated in Table 2. It can be seen that the scores in the range 30-100 and 400-950 may call for human review but the rest of the cases are quite efficient in diagnosis.

4. CONCLUSION

Through this paper, the application of 'TissueQuant' algorithm to assess the quantification of ck7 staining is described for the diagnosis of sinovial sarcoma. For histological and cytological diagnosis, the amount of tissue or the substance under study can also be quantified as described in this paper. Since the development of technology brings in high throughput, automation of the analysis could play a very important role. Tissue microarray provides such a high through put that handling them individually is quite laborious. It is hence desirable to limit the human intervention to cases which need some human judgment. This paper demonstrates the use of the color quantification algorithm to contribute to match up with the high throughput of the imaging technology, to bring in lot of advancements in the area of medical research.

5. REFERENCES

[1] Caldwell M., Liu A., Jain P., Sharma R., 2008. Simple quantification of multiplexed quantum dot staining in clinical tissue samples, 30th Annual International Conference of the IEEE EMBS 1, 1907-1910

[2] Aziz D., 1992. Quantitation of estrogen and progesterone receptors by immunocytochemical and image analyses. American Journal of Clinical Pathology 98, 105-111.

[3] Lehr H., Mankoff D., Corwin D., Santeusanio G., Gown A., 1997. Application of Photoshop-based image analysis to quantification of hormone receptor expression in breast cancer. J. Histochem. Cytocehm. 45, 1559-1565.

[4] De Solorzano C., Costes S., Callahan D., Pravin B., 2002. Application of quantitative digital image analysis to breast cancer research. Microscopy Research and Technique 59, 119-127.

[5] De Boer W., Hiemstra P., Sont J., De Heer E., 2001. Image analysis and quantification in lung tissue. Clinical and Experimental Allergy 31, 504-508.

[6] Herbin M., Venot A., Devaux J., Piette C., 1990. Color quantitation through image processing in dermatology. IEEE Transactions on Medical Imaging 9, 262-269.

[7] Alla S., Huddle A., Butler J., Bowman P., Clark J., Beyette F., 2011. Point-of-care device for quantification of bilirubin in skin tissue. IEEE Transactions on Biomedical Engineering 58, 3.

[8] Encarnacion M., Griffin M., Slezak J., Bergstralh E., 2004. Correlation of quantitative digital image analysis with the glomerular filtration rate in chronic allograft nephropathy. American Journal of Transplantation 4, 248.

[9] Forero M., FilipSroubek, Gabriel C., 2004. Identification of tuberculosis bacteria based on shape and color. Real-Time Imaging 10, 251-262.

[10] Niendorf A., Rath M., Wolf K., Peters S., 1990. Morphological detection and quantification of lipoprotein (a) deposition in atheromatous lesions of human aorta and coronary arteries. Virchows Archives of Pathological Anatomy 417, 105-111.

[11] Lehr H., van Deer Loos C., Teeling P., Gown A., 1999. Complete chromogen separation and analysis in double immunohistochemical stains using photoshop-based image analysis. J. Histochem. Cytocehm. 47, 119-126.

[12] Pauschinger M., Knopf D., Petschauer S., Doerner A., 1999. Dilated cardiomyopathy is associated with significant changes in collagen type I/III ratio. Circulation -American Heart Association Inc. 99, 2750-2756.

[13] Johansson A., Visse E., Widegren B., Sjogren H.,2001. Computerized image analysis as a tool to quantify infiltrating leukocytes: A comparison between high- and low- magnification images. J. Histochem. Cytocehm. 49, 1073-1080.

[14] Maximova A., Taffs R., Pomeroy K., Piccardo P.,2006. Computerized morphometric analysis of pathological prion protein deposition in scrapie-infected hamster brain. J. Histochem. Cytocehm. 54, 97-107.

[15] Park J., Hong W., Kim C., 2010. Color intensity method for hydrogel oxygen sensor array. IEEE Sensors Journal 10, 1855-1862.

[16] Masmoudi H., Hewitt S., Petrick N., Myers K., Gavrielides M., 2009. Automated quantitative assessment of her-2/neu immunohistochemical expression in breast cancer. IEEE Transactions on Medical Imaging 28, 916-925.

[17] Prasad K, Kumar B.P, Marx C.S, Prabhu G.K, 2012. Applications of 'TissueQuant' – a color intensity quantification tool for medical research, Computer methods and programs in biomedicine, 106, 27-36.